Institut für Insektenbiotechnologie

Professur für Insektenbiotechnologie im Pflanzenschutz

Justus-Liebig-Universität Gießen

Genetic variability and population dynamics of the Spotted Wing Drosophila, *Drosophila suzukii,* in Germany

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades (Dr. rer. nat.)

im Fachbereich Agrarwissenschaften, Ökotrophologie und Umweltmanagement

der Justus-Liebig-Universität Gießen

vorgelegt von

Sarah Petermann

Gießen, 2021

Mit Genehmigung des Fachbereichs Agrarwissenschaften, Ökotrophologie und Umweltmanagement der Justus-Liebig-Universität Gießen

1. Gutachter	**Professor Dr. Marc F. Schetelig**	
	Institut für Insektenbiotechnologie	
	Fachbereich Agrarwissenschaften, Ökotrophologie und Umweltmanagement	
	Justus-Liebig-Universität Gießen	

1. Gutachter **Professor Dr. Marc F. Schetelig**
Institut für Insektenbiotechnologie
Fachbereich Agrarwissenschaften, Ökotrophologie und
Umweltmanagement
Justus-Liebig-Universität Gießen

2. Gutachter **Professor Dr. Nikola-Michael Prpic-Schäper**
Institut für Allgemeine Zoologie und Entwicklungsbiologie
Fachbereich Biologie und Chemie
Justus-Liebig-Universität Gießen

Prüfer: **Professor Dr. Steffen Pauls**
Institut für Insektenbiotechnologie
Fachbereich Agrarwissenschaften, Ökotrophologie und
Umweltmanagement
Justus-Liebig-Universität Gießen

Prüfer: **Professor Dr. Volkmar Wolters**
Institut für Tierökologie und Spezielle Zoologie
Fachbereich Biologie und Chemie
Justus-Liebig-Universität Gießen

Vorsitz: **Professor Dr. Gunter P. Eckert**
Justus-Liebig-Universität Gießen

Tag der Disputation: 10.03.2022

Table of Contents

Summary

Drosophila suzukii (Matsumura), the Spotted Wing Drosophila (SWD), is native to Southeast Asia and rapidly invaded America and Europe in the past 20 years. SWD is a crop pest of soft-skinned fruits with a wide range of host plants, threatening the fruit industry worldwide and causing enormous economic losses. One step to control this invasive pest species is to understand its population dynamics and structure.

In this work, I present the population genetics and development of SWD in Germany from 2017-2019 using microsatellite markers over eleven different sample sites. It is the first study that examines the genetic changes in SWD populations over three years compared to multiple international SWD laboratory strains. It was shown that SWD populations in Germany are highly homogenous. Differences between populations or years were not observed, indicating that populations are well adapted, migrate freely, and potential reinvasions from outside Germany either do not take place or are negligible.

The results of this work help to understand SWD biology and population development. The high genetic variability and migration between populations could allow for a fast establishment of this pest species. This is especially problematic concerning the ongoing spread of this invasive species and could bear a potential for developing pesticide resistance. This could have a significant effect on pest control strategies for SWD in the future.

Zusammenfassung

Die Kirschessigfliege *Drosophila suzukii*, oder auch Spotted Wing Drosophila (SWD), ist beheimatet in Südostasien und fiel in den letzten zwanzig Jahren in großen Teilen von Amerika und Europa ein. *D. suzukii* ist ein Schädling von weichschaligen Früchten und ist verantwortlich für immense wirtschaftliche Verluste im Obstanbau weltweit. Um diesen invasiven Schädling besser kontrollieren zu können, ist das Verständnis seiner Populationsdynamik und Struktur notwendig.

In dieser Arbeit stelle ich die räumliche und zeitliche Entwicklung von *Drosophila suzukii* in Deutschland vor. Dazu habe ich Populationen an elf verschiedenen Standorten in den Jahren 2017 bis 2019 mit Hilfe von molekularen Markern untersucht. Dies ist die erste Studie die genetische Veränderung in *D. suzukii* Populationen über einen Zeitraum von drei Jahren untersucht und mit internationalen Laborpopulationen vergleicht. Ich konnte zeigen, dass deutsche Populationen sehr homogen sind. Unterschiede zwischen Jahren oder Standorten konnten nicht festgestellt werden. Dies deutet darauf hin, dass die Populationen gut angepasst sind, frei migrieren und dass potentielle Re-Invasionen von außerhalb Deutschlands entweder nicht stattfinden oder keinen Einfluss auf die vorhandenen Populationen ausüben.

Die Ergebnisse dieser Arbeit helfen die Biologie und die Populationsentwicklung von *D. suzukii* besser zu verstehen. Die hohe genetische Variabilität und Migration zwischen Populationen könnten eine Rolle in der schnellen Anpassung dieses Schädlings spielen. Dies ist insbesondere problematisch in Hinblick auf die andauernde Ausbreitung dieser invasiven Art und könnte auch das Potenzial zur Entwicklung von Pestizidresistenzen haben. Dies wiederum könnte in Zukunft erheblichen Einfluss auf Strategien zur Bekämpfung dieses Schädlings nehmen.

1. Introduction

1.1. Invasive Pest Species and their Impact on Biodiversity and Economy

An invasive or non-indigenous species is a non-native animal, plant, or microorganism that was introduced and established into a new area, often as a result of human activities. Many invasive species are considered pest species as they often have a negative impact on the new environment or pose a health risk (Sala et al., 2000). Others are introduced intentionally like potatoes (Hawkes and Francisco-Ortega, 1993), tomatoes (Bergougnoux, 2014), ornamental plants (Dehnen-Schmutz, 2011), or animals (Kumschick et al., 2016), which were introduced to Europe from America and Asia. Those can even have economic benefits. Of greater concern are those species that escape or invade unnoticed. Climate change is one important driver for invasion success and it can facilitate the establishment of non-native species (Altizer et al., 2013). The geographical range of a species is naturally limited by climate and other abiotic or biotic factors, but rising temperatures have a profound effect on these ranges by altering fundamental characteristics of the environment. These changes allow species to expand their geographical range into new ecosystems (Dukes and Mooney, 1999). But climate change is only a part of the problem. The movement of people and goods on a global scale eases the worldwide introduction of non-native species (Hulme, 2009). Alien species are transported passively to new regions by airplanes, ships, or automobiles, either by contaminating goods or attached to the vehicle. The increasing number of invasive species in Europe shows a pattern that is consistent with the increase in trade and travel (Jeschke and Strayer, 2005; Keller et al., 2011).

Invasion is a multi-step process that starts with the introduction of a species into new areas, which were previously not inhabited by this species. Afterward, the introduced non-native species has to establish a self-sustaining population to be considered invasive. Therefore it has to survive and sexually reproduce and it has to spread further from its new habitat (Keller et al., 2011). At this stage, most species already have a measurable impact on the environment, economy, or health (Nentwig, 2008). In addition, the establishment of a species in a novel area depends on several factors. Again, climate plays an important part in the survival of a species. Other factors are the availability of food and the presence or absence of predators and competitors. Without natural control agents in a habitat, there is little to no selection pressure for a species. Lastly, the success of an invasive species depends on whether or not it has traits that are advantageous for the colonization of new areas. Generalists for example have, at least in theory, better chances of survival and propagation than specialists. Beneficial traits do also include a short life cycle and high reproductive rates (Jeschke and Strayer, 2006; Keller et al., 2007). *Harmonia axyridis*, the seven-spot ladybird, is an example of how well an invasive species can thrive in a new environment due to greater competitive ability. First used as a classical biological control agent for aphids in the US and Europe, it is now considered a pest species because it is in

direct competition with native ladybird species, it is feeding on soft fruit crops, negatively impacting vinery and of general annoyance when overwintering in masses inside houses (Kovach, 2004; Majerus et al., 2006).

The impact of invasive species can be quite diverse, ranging from negative effects on biodiversity to economic risks and health. The database DAISIE[1] estimates that there are several thousand non-native species invasive in Europe, including but not limited to 1,522 terrestrial invertebrates of which almost 86% are insects (Drake, 2017; Keller et al., 2011; Roques et al., 2009). The overall costs for production losses, lower quality of goods, and higher production expenses caused by invasive pest species in Europe are assumed to range between 12.5 and 20 billion EUR per year (Kettunen et al., 2009). This topic has consequently developed into a pressing issue for the European Union, which is suggesting and implementing new policies to reduce importation, release, and establishment of non-native species (Shine et al., 2010). A study showed that addressing invasive alien species would cost the EU 40 to 190 million EUR per year (Shine et al., 2010). Regarding negative impacts on ecosystems, invasive species are thought to be one of the main reasons for the loss of biodiversity, alongside other man-made causes like climate change and pollution (Wilcove et al., 1998). They influence their surroundings either directly by feeding on native plants or animals or indirectly as they constitute as (food) competitors or by transferring pathogens and parasites. There are numerous examples for this such as the two wasp species *Vespula germanica* and *Vespula vulgaris* which are now established in New Zealand, having competitive effects on native invertebrate species and birds (Beggs, 2001) or the zebra mussel competing with native clams, leading to local extinctions of the native species (Sousa et al., 2011; Strayer et al., 1999).

1.2. The Spotted Wing Drosophila, *Drosophila suzukii*

Drosophila suzukii (Matsumura) or Spotted Wing Drosophila (SWD) belongs to the genus *Drosophila*, subgenus Sophophora. While this genus harbors several harmless species that feed and breed on decaying fruits, including *Drosophila melanogaster*, SWD, however, is a serious fruit crop pest with substantial economic impact. Economic losses are reported mainly for the US and in Europe for Italy and Switzerland (Bolda et al., 2010; De Ros et al., 2013; De Ros et al., 2015; Knapp et al., 2020; Mazzi et al., 2017). Bolda et al. (2010) calculated that the revenue losses in raspberries and blackberries in California totaled approximately $63.2 million in 2008 alone. The total economic damage in Trentino (Italy) was estimated to value almost 3.3 MEUR per year (De Ros et al., 2013). SWD is placed in the so-called 'oriental lineage' but the relationships between the drosophilids in the

---

[1] www.europe-aliens.org

subgroup are not completely resolved yet (Kopp and True, 2002; Schawaroch, 2002). So far, *D. biarmipes* and *D. subpulchrella* are considered to be the most likely sister taxa of *D. suzukii* (Chiu et al., 2013; Yang et al., 2012). The first high-quality genome sequence was published in 2013 by Chiu et al. The genome is comparable in size to other *Drosophila* species (Boulesteix et al., 2006; Chiu et al., 2013; Moriyama et al., 1998).

Morphology: A complete and detailed guide for identification was given by Hauser (2011) and Vlach (2010). A side-by-side comparison of *D. melanogaster* and *D. suzukii*, as well as the most prominent male and female characteristics, is given in Figure 1. Females and males both are 2 – 3 mm in length, females being usually slightly bigger, and have a brownish body with black stripes on the abdomen (Hauser 2011). The most striking characteristic of SWD is the dark spot males display on the top of each wing, which gives this species its common name Spotted Wing Drosophila (SWD). In addition, males show two black sex combs on each foreleg. Female flies possess a serrated ovipositor that functions like a saw, with which they pierce the skin of ripening fruits (Hauser, 2011). Species identification can be difficult in regions where distinguishing features are present also in other *Drosophila* species like *D. subpulchrella*. However, this is not relevant for Germany since SWD is currently the only species of the 'melanogaster' group present in Europe that is characterized by these morphological traits (Lavrinienko et al., 2017). In addition, it takes up to two days till the male wing spot is visible or males can occasionally lack the wing spot, which can lead to misidentification as well. A genetic identification by DNA barcoding is possible and allows a reliable identification even in immature specimens (Freda and Braverman, 2013; Kim et al., 2014).
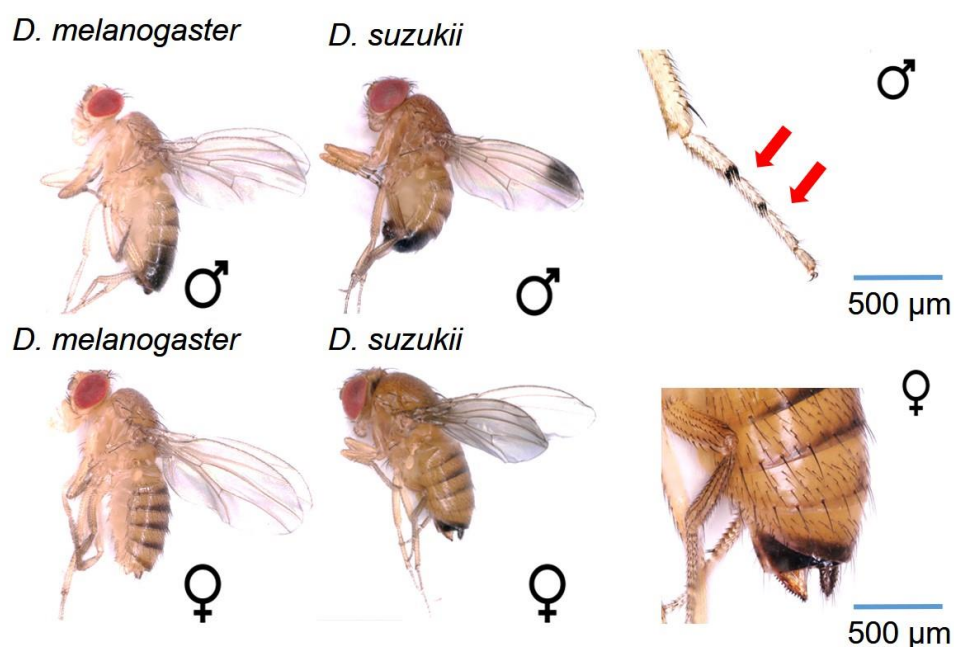
Figure 1: Morphology of *D. suzukii*. Comparison of morphological traits of *D. melanogaster* and *D. suzukii* (left). Close up on male sex combs and female ovipositor of *D. suzukii* as markers for identification of this species (right).

Life Cycle: Similar to other drosophilids SWD has a high reproductive capacity with a short generation time. The entire life cycle from egg to adult depends on the climate conditions but is usually completed in one or two weeks (Kanzawa, 1939). A complete life cycle with details of each life stage is shown in Figure 2. Females lay eggs on ripening or ripe fruits. The number of eggs on one fruit can vary but ranges from only a few per plant to several on one single berry. SWD can infest a wide range of host plants, including grapes, blackberries, strawberries, and cherries. The color, skin firmness, and odor of the host plants play an important role in host choice preferences (Cloonan et al., 2019; Little et al., 2019; Takahara and Takahashi, 2017). The female uses its serrated ovipositor to open the fruit skin and to lay eggs directly in the fruit pulp. The eggs are only 0.4 to 0.6 mm in length, white, and with two aeropyles at one end. Under optimal developmental temperatures (22 °C) a female can lay up to 195 eggs during its lifetime (Tochen et al., 2014). For temperatures above 28 °C low levels of reproduction or no reproduction were found (Tochen et al., 2014). The hatching larvae feed inside the fruit on the pulp. Larvae grow through three larval stages, they are white to yellowish with black mouthparts. The infested fruit collapses, leaving unmarkable fruits. The larvae mature in 3 – 13 days and pupate in the fruit. This stage lasts 4 – 15 days. In addition to the damage caused by the feeding larvae, the oviposition scar can cause secondary infections with bacteria and fungi, which may cause rotting (Lewis et al., 2019; Walsh et al., 2011). The pupae are oval with a dark brown color and 3.5 mm long. Adult SWD have a lifespan of 20 – 56 days, but overwintering adults show a much longer lifespan with up to 200 days recorded (Kanzawa, 1934, 1939). Usually,

SWD overwinters as adults in cultivated landscapes or forests as well as uncultivated woodlands with wild fruit crops, arising in spring (Dalton et al., 2011; Jakobs et al., 2015).
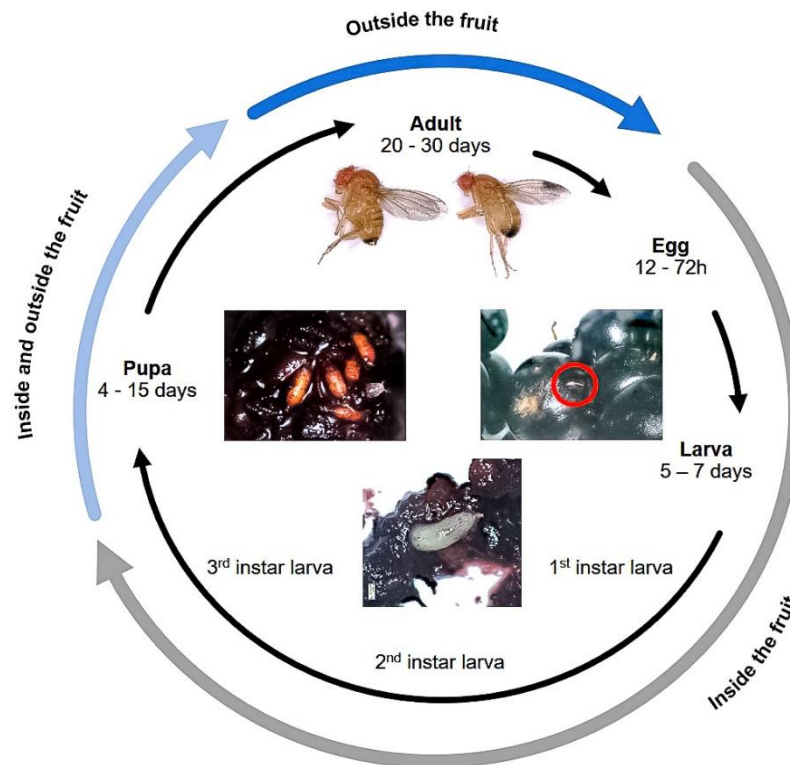
Figure 2: Life cycle of *D. suzukii*. The life cycle from egg to adult individual with estimated time and localization of each stage. Figure based on Rendon et al. (2019).

Distribution and Invasion History: Native to East Asia (Hauser, 2011; Walsh et al., 2011), SWD is now invasive in most of the US, Canada, and Europe. First introductions were recorded on the Hawaiian Islands in 1980 (Hauser, 2011). An introduction to Europe followed later in 2008, starting in Spain and Italy (Calabria et al., 2012), spreading further to France (Cini et al., 2012), Switzerland (Baroffio and Fischer, 2011), Germany (Vogt et al., 2012), Romania (Chireceanu et al., 2015) and Slovenia (Seljak, 2011). Canada reported first occurrences in 2010, with a rapid spread across British Columbia, Alberta, Ontario, and Quebec (Lee et al., 2012). The map in Figure 3 presents the currently known worldwide distribution of SWD.
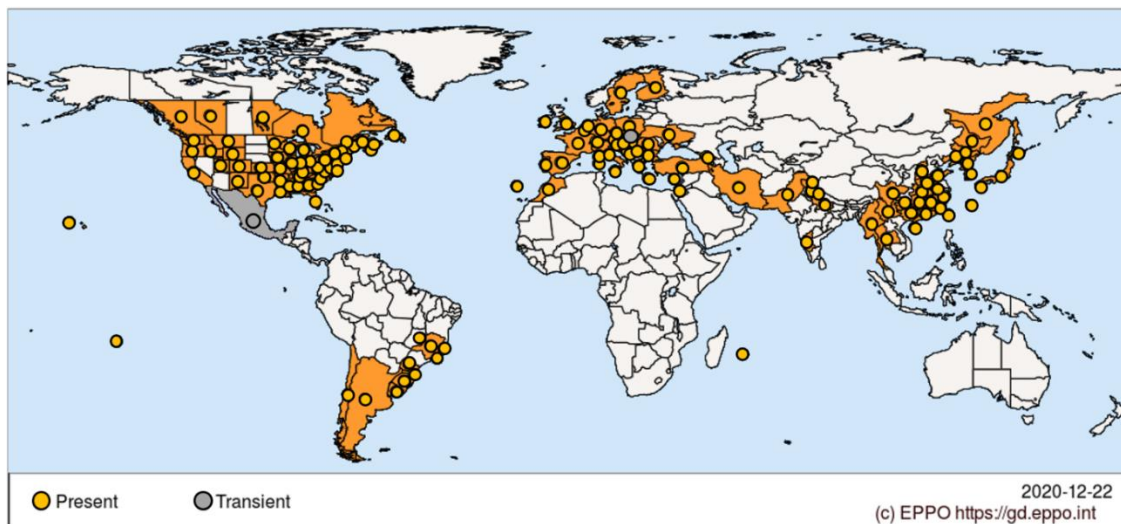
Figure 3: Worldwide distribution of _D. suzukii_. Shown is a map with the currently known worldwide distribution of SWD. Orange = SWD is present, grey = SWD is transient, records only in single years. The map originates from https://gd.eppo.int (date: 2020-12-22).

The rapid spread across continents is associated with the global trade of fresh fruits and potential host plants, which makes passive human-mediated diffusion the most likely cause of the spread of SWD (Cini et al., 2014; Cini et al., 2012; Westphal et al., 2008). The high reproductive rate and short generation time with several generations per year, contribute towards a rapid spread. Since larvae and eggs are difficult to detect due to them hiding inside the fruits, they might likely spread even further and expand their range in Europe and America. While an expansion in more arid and hot climatic zones seems to be unlikely, a spread to northern areas with colder climates is possible, since overwintering in SWD is common and often affiliated with human habitation. A first estimation of Calabria et al. (2012) showed that SWD might be capable of expanding approximately 1400 km per year.

1.3. Pest Control in _Drosophila suzukii_

An essential part of tackling an invasive pest like SWD is to understand its biology and to answer questions on the threat it poses for individual fruit crops (Bellamy et al., 2013; Naranjo-Lázaro et al., 2014). Several studies have already been conducted or are still ongoing on this subject, mainly focused on traps and pesticide applications (Bruck et al., 2011; Cha et al., 2013; Iglesias et al., 2014; Van Timmeren and Isaacs, 2013). Usually, the monitoring of a pest foregoes other measures. An estimate can be made on how severe an infestation is and for early detection in potentially newly-invaded areas. Therefore, monitoring of fly populations is continuously conducted in Germany (Briem et al., 2015; Briem et al., 2018). Adult SWD are monitored in the field by using traps baited with different attractants like wine and apple cider vinegar (Briem et al., 2015; Cha et al., 2013) or sugar water mixed with baker's yeast (Iglesias et al., 2014; Walsh et al., 2011). Monitoring

programs like the one managed by the Julius-Kühn-Institute can provide valuable information on best practice control methods for fruit growers (Briem et al., 2018).

One of the first measures after monitoring includes the removal of infested fruits as well as the removal of other plants at the crop site that serve as potential hosts. This method is extremely time-consuming and not applicable in large plantations. In viticulture, it turned out to be effective to remove excessive foliage from the vine stock. That way the fruit is exposed to sunlight and SWD has fewer options to withdraw in the more shadowy and therefore cooler parts of the plant (Knapp et al., 2019; Leach et al., 2018). This method is tailored for grapevine and only applicable to crops that are manually maintained. An alternative method is the netting of crop plants to prevent SWD from reaching the fruits (Leach et al., 2016). However, nets have to be installed before fruits begin to ripen, it can cost a substantial amount of money even if it is reused for several years and it has to be checked for holes and entry sites repeatedly (Del Fava et al., 2017). Studies showed that badly constructed and handled nets might even increase the damage to the crops (Augel et al., 2020).

Currently, the most effective and commonly used control methods are chemical insecticides, even though their usage has several disadvantages. First, the range of used insecticides is small and the efficacy is limited since highly efficient chemicals are restricted by the EU (Cini et al., 2012). Jarausch et al. (2017) tested the activity of different products in all life stages of SWD and found that products with an estimated efficiency of less than 95% are not sufficient enough to manage this pest. Insecticides in use include spinosyns, organophosphates, pyrethroids, and neonicotinoids, which is problematic because these insecticides are either severely restricted by the European Commission (neonicotinoids) or their approval ends soon and will probably not be renewed (spinosyns). In organic crop production, they are not allowed at all except for spinosyns in bait or the approved alternative variants are not as effective (Walsh et al., 2011). Chemical control methods in SWD have to be applied repeatedly at the ripening stage, which can entail serious consequences. There is a much higher chance of pesticide residues in fruits and the development of pesticide resistance in the insect (Rota-Stabelli et al., 2013). Furthermore, chemical insecticides often have negative effects on beneficial insects (Ndakidemi et al., 2016).

In contrast, other practices such as mass trapping and biological control methods using imported or native parasitoids could be central for future pest management strategies (Gabarra et al., 2015; Stacconi et al., 2015). In its native range in Asia, hymenopteran parasitoids of the genera *Leptopilina*, *Trichopria,* and *Asobara* have been reported to target SWD larvae or pupae. Unfortunately, parasitoids native to Europe or the US are not as effective as the Asian species, probably due to the lack of co-evolution together with the host (Chabert et al., 2012; Stacconi et al., 2015). Further, SWD produces up to five times

more hemocytes than *D. melanogaster*, which makes it more resistant to parasitism and less likely for native parasitoids to switch to SWD as a host (Kacsoh and Schlenke, 2012). In a more recent approach, viruses, bacteria, and entomopathogenic fungi such as *Metarhizium anisopliae* have been tested as potential biological control agents (Naranjo-Lázaro et al., 2014; Siozios et al., 2013; Unckless, 2011).

Another option for integrated pest management and an environmentally friendly alternative to pesticides could be the Sterile Insect Technique (SIT) that has proven highly effective in agricultural insect species (Augustinos et al., 2017; Benedict and Robinson, 2003; Krafsur, 1998; Wyss, 2000). For SIT programs, sterilized male individuals are released into the environment and lead to infertile mating. That method could be used in the future to reduce the population size of SWD. New methods and strains for SIT pest control programs against SWD are already in development (Kalajdzic and Schetelig, 2017; Yan et al., 2020). Therefore, while SWD control is still challenging, biological control methods, including SIT, remain a beneficial option for sustainable pest control.

To complement and improve efforts of pest control applications, an understanding of a pest's biology and genetic evolution is necessary. Knowledge about invasion history and population genetics can identify introduction pathways, improve testing, and help integrated pest management strategies in ultimately avoiding multiple (re)introductions (Estoup and Guillemaud, 2010). The reconstruction of invasion pathways is particularly crucial to uncover and understand potential patterns in the spread of an invasive species. This can be challenging, considering the sheer amount of fruit transported through global trading. For example, Germany imported 51,776 tons of fresh fruits in the year 2017 alone (FAOSTAT, 2021). Studies on population genetics have already been proven beneficial for biological control methods like the sterile insect technique (SIT) (Lanzavecchia et al., 2014). Since gene flow can vary between natural populations, locally adapted and isolated populations can occur. In these populations, SIT can be impaired by mating barriers, making them less effective. Population genetics can be used to trace changes in strain efficacy and it can help to improve the competitiveness of laboratory strains that are sterilized and released and can be used in monitoring programs to differentiate released laboratory insects from wild ones (Aketarawong et al., 2011; Aketarawong et al., 2014; Azrag et al., 2016; Zygouridis et al., 2014).

1.4. Population Genetics and the Detection of Genetic Diversity

Population genetics is a field in biology that studies the genetic arrangement of populations and the changes that occur over time due to evolutionary influences like mutation, natural selection, or gene flow (Bruederle et al., 2001; Lowe et al., 2017). Population genetics studies and theories find application in many areas and not only in conservation biology. Data on genetic distance can be used to infer evolutionary history, molecular methods are used in forensics and it can be used to trace invasive species or transmission routes of infectious diseases (Balding and Nichols, 1995; Gillespie, 2004; Hill, 2014). Ronald Fisher, John Haldane, and Sewall Wright are thought to be the founders of today's population genetics by combining natural selection and evolution with Mendelian genetics in complex mathematical models (Ewens, 2012). Populations of a species are shaped through evolutionary forces and exhibit some sort of genetic structure. Only if gene flow is not restricted by barriers of any kind or shape (mating or actual geographical barriers), populations would appear to show no genetic structure (Barton and Hewitt, 1985). A high gene flow between populations would ultimately lead to a homogenization of allele frequencies and therefore prevent local adaptation and speciation (Barton and Hewitt, 1985; Slarkin, 1985). On the other hand, a population without any gene flow, which does not receive migrant individuals from other populations, can only introduce novel alleles due to mutations (Levin, 2001). Another important evolutionary mechanism that shapes genetic structure in a population is genetic drift. Rare alleles can become much more frequent and even fixed, while other allele variants might disappear and thereby reduce the population's genetic variation (Masel, 2011; Whitlock, 2000). The comparison of genetic structure in a metapopulation to its subpopulations can allow an estimate of the current evolutionary potential of this population(s), which is of great importance for conversation biology and biodiversity (Pannell and Charlesworth, 2000). If gene flow is restricted between local populations, more subpopulations of smaller size can occur with less genetic diversity. Those populations are more likely to suffer negative consequences from genetic drift and are more likely to mate with related individuals (Hensen and Oberprieler, 2005; Pannell and Charlesworth, 1999). The genetic content of individuals in a species is usually not identical. DNA sequences differ to a certain degree, which forms the genetic diversity in a species or a population (Ellegren and Galtier, 2016). In theory, genetic diversity describes the presence or absence of allele variants and their ratio in a population or subpopulation (Hughes et al., 2008). Random mutations generate new allele variants in every generation but the rate at which mutations occur is not equal among individuals or populations and contributes to variation in genetic diversity (Lynch, 2010). Other factors like genetic drift, inbreeding, or assortative mating influence the variation as well (Ellegren and Galtier, 2016). Information on genetic diversity and genetic structure in populations is key in optimizing conservation and utilization strategies (Amos and Harwood, 1998; Hughes et al., 2008).

Traditionally, diversity between individuals or populations could only be accessed through the observation of phenotypic traits and pedigree information. The amount of informative content gained by morphological markers is however limited, as the number of phenotypes can be constrained, environmental factors, age, and gender can shape morphology and it requires an experienced evaluation (Tanksley, 1983). The development of molecular marker systems made phenotypic characterization lose its importance in diversity studies relatively fast. These new tools allowed access to a vast collection of novel molecular markers systems, which removed the limitations of phenotypic markers (Al-Samarai and Al-Kazaz, 2015; Idrees and Irshad, 2014).

1.4.1. Marker Systems for Population Genetics

A vast number of different molecular marker techniques are available for population genetic studies. The choice of which of these techniques is used depends on the research question and the objects studied. Genetic markers are either protein- or DNA-based and expose differences in DNA or protein sequences, usually with trade-offs between accuracy and handiness (Sunnucks, 2000). In general, DNA markers are preferred over protein markers, because DNA can be extracted from low-quality samples, and most importantly, it can be processed using Polymerase Chain Reaction (PCR) (Bruford and Wayne, 1993; Estoup and Angers, 1998; Sunnucks, 2000). Another problem with protein markers is, that the number of polymorphisms observed at a given locus is often low (Richardson et al., 1988). Favorable characteristics of genetic marker systems also include rapid development and comparability. Different systems that are often used include sequence-related amplified polymorphism (SRAP) (Li and Quiros, 2001), restriction fragment length polymorphism (RFLP) (Burr et al., 1983), amplified fragment length polymorphism (AFLP) (Vos et al., 1995), random amplified DNA polymorphism (RAPD) (Williams et al., 1990), variable number tandem repeat (VNTR) (Johansson et al., 2004), single nucleotide polymorphisms (SNPs) (Syvänen, 2001) and simple sequence repeats or microsatellites (SSR) (Litt and Luty, 1989). All of the mentioned marker systems have particular strengths and weaknesses and the results of a population genetics study are limited by the marker system used (Behura, 2006). Co-dominant markers like microsatellites or RFLPs can differentiate heterozygotes from homozygotes, while dominant markers systems like RAPD cannot distinguish between hetero- and homozygotes (Sunnucks, 2000). Therefore, dominant markers cannot be used to detect polymorphic loci or to calculate the number of alleles at each locus, which themselves can already be used to interpret genetic diversity. Unlike RAPD or AFLP, microsatellite markers deliver this information and are the most frequently used markers for diversity studies as well as for parentage analysis (Schlötterer, 2004; Sunnucks, 2000).

Microsatellite Markers (Simple Sequence Repeats - SSRs): Microsatellites, also known as simple sequence repeats (SSR) or short tandem repeats (STR), are repetitive DNA elements and widely used as genetic markers in eukaryotes (Bruford and Wayne, 1993; Schlötterer, 2000). The repeated nucleotide motifs are usually between 2 and 10 bp long and repeated at least once but usually several times (Schlötterer, 2000) (Figure 4). This type of polymorphic marker allows genetic differentiation of closely related individuals in a single population. Individuals vary in the number of repeats, which results from additions or deletions of repeat units caused by strand-slippage during DNA replication and recombination (Vieira et al., 2016). They are a subcategory of tandem repeats, distributed throughout the genome with high mutation rates estimated at 10^{-3} to 10^{-6} (Gemayel et al., 2012) which makes them highly polymorphic. These traits make them useful for population genetic studies and since they are inherited from both parents, they are used for parental analysis as well. The high mutation rates result in a high number of alleles per locus, which makes SSRs more informative than other molecular markers, especially at the subpopulation level. Factors influencing repeat instability (the mutation rate) are the length and the 'purity' of the repeat. Studies found that longer and purer repeats tend to have higher mutation rates than shorter or imperfect repeats (Legendre et al., 2007; Petes et al., 1997). Microsatellites are referred to as pure if they contain only exact copies of a repeat motif. Imperfect or less pure microsatellites contain at least a single variant in the repeat motif. Another factor influencing microsatellite stability is the base composition of the repeat. DNA slippage occurs more often in repeats with polyC or polyG and less in repeats with a polyA or polyT motif (Gragg et al., 2002).
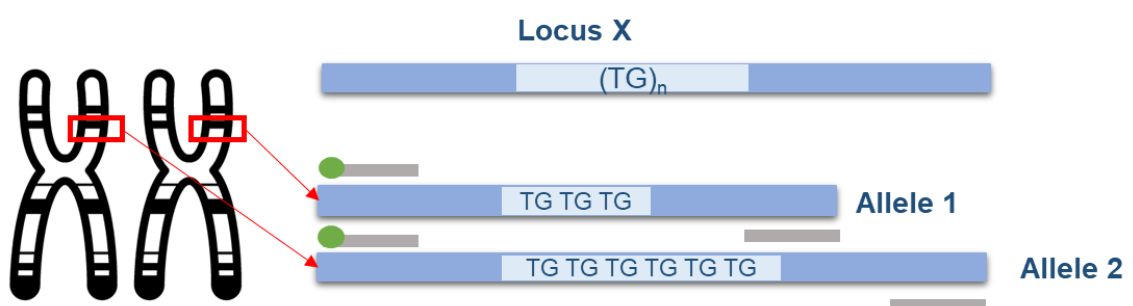
Figure 4: Schematic illustration of a microsatellite locus. Shown is a microsatellite locus X with a $(TG)_n$ repeat motif (light blue) in an individual. Both alleles vary in size due to different numbers of repeated motifs. For use as a genetic marker, the microsatellite locus is flanked by primer binding sites (illustrated by grey bars) of a 5'-labeled (green dot) primer and the unlabeled reverse primer.

SSRs have been extensively used over the past 20 years because they are considered highly informative, they are co-dominant, and hence allow the differentiation between

homozygous and heterozygous individuals, they can be transferable between related taxa, they are relatively cost-efficient and have a high reproducibility (Vieira et al., 2016). In addition, they are PCR based, which enables high-throughput analyses and multiplexing and they can be amplified from low quality and quantity of DNA (Kumar et al., 2018). One disadvantage of microsatellite markers is the relatively complex development (Zane et al., 2002). Another drawback is that the underlying mutation model is still uncertain (Balloux and Lugon-Moulin, 2002). This is problematic since the high mutation rate of microsatellite markers makes it necessary to understand the mutational patterns. An understanding of the process is important to correctly interpret the information obtained with these markers and to use relevant statistics (Balloux and Lugon-Moulin, 2002). Several theoretical models are aiming to describe the mutation mechanism in microsatellites but none of these models seem to be capable of describing the mutation behavior of all microsatellite loci (Balloux and Lugon-Moulin, 2002). The most fitting mutation models are the infinite allele model (IAM) (Kimura and Crow, 1964), the stepwise-mutation model (SMM) (Kimura and Ohta, 1978), and the two-phase model (TPM) (Di Rienzo et al., 1994; Valdes et al., 1993). IAM describes how each mutation creates a novel allele, which has not been present in the population. This model accounts for small changes in the repeat number and alleles of similar size. The same alleles share the same ancestry (Kimura and Crow, 1964). The SMM states that each mutation creates a new allele either by adding or deleting a single repeat unit of the microsatellite with an equal probability. Alleles with great differences in sizes will be less related than alleles of similar size (Kimura and Ohta, 1978). The TPM is a variation of the SMM and simply accounts for larger mutation events (Di Rienzo et al., 1994; Valdes et al., 1993).

Nevertheless, thanks to reduced analysis costs and sample usage, the possibility of high throughput analysis, PCR amplification, and multiplexing using fluorescent labels as well as non-overlapping PCR product sizes has made SSR markers a method of choice in many laboratories (Guichoux et al., 2011; Vieira et al., 2016).

1.4.2. Population Genetic Statistics for Microsatellites

Genetic and allelic diversity plays a vital role in the invasion success of a species, and the assessment of these traits helps to understand the genetic relationship among populations. Population genetics itself and the use of microsatellite markers, in particular, rely greatly on mathematical models and statistics. Data quality is of extreme importance and preliminary analysis of the marker system is advisable. Aside from missing data or unusual patterns, testing for null alleles, Hardy-Weinberg equilibrium (HWE), and Polymorphism Information Content (PIC) is a necessary first step in every marker analysis. There are two main approaches on how to estimate null alleles in a dataset. The first is a maximum likelihood

method that compares observed and expected homozygote frequencies in the population under the assumption of HWE (Brookfield, 1996) and the second method estimates null alleles from progeny where parent genotypes are known (Dakin and Avise, 2004). It was shown that both approaches are equally reliable (Oddou-Muratorio et al., 2009). HWE states that genotypes occur in predictable frequencies and remain constant in the absence of evolutionary forces, a state that is rarely met in nature. Deviations from HWE occur due to mutations, genetic drift, gene flow, or inbreeding (Chen, 2010). Testing for HWE can be used as a data quality check by discarding loci that significantly deviate from controls which is difficult if there is no control available. Even if loci are not discarded, it can help to compare samples and to explain unusual results later on. One way of testing deviations from HWE is to use a goodness-of-fit test, often referred to as Pearson's χ^2 test (Li, 1955). Another method are exact tests, which perform better if sample sizes are small but they are relatively computation-intensive (Haldane, 1954; Levene, 1949).

Especially for new marker systems, it is important to evaluate the so-called Polymorphism Information Content (PIC). This is a value that helps to determine if a locus is 'useful' for analysis (Guo and Elston, 1999). It is used to measure the ability of a marker to detect polymorphism. It is an alternative to the measure of heterozygosity, which is just another parameter used to evaluate the quality of a marker. It is defined as the probability that the genotype of a given offspring will allow a conclusion which of the two alleles of the parent it received (Guo and Elston, 1999).

After the described tests for data quality are passed, the actual population data can be analyzed. Populations can be described with two different approaches, namely frequency-based and distance-based analysis. A very basic first step is to use frequency-based statistical procedures including the calculation of allele frequencies at each locus for every population, the number of alleles (N_a), the effective number of alleles (N_e), the observed (H_o), and expected (H_e) heterozygosity, private alleles and the Fixation index (F). Another way to describe the relationship between populations is to measure the genetic distance. Distance-based methods calculate the pairwise distance between samples. While they are easy to use, it is difficult to include any further information like sampling locations. Also, they are heavily dependent on the chosen distance measure, which can lead to varying outcomes (Pritchard et al., 2000) and there is a huge variety in different distance measures available: D_A distance (Nei et al., 1983), D_{ST} (Nei, 1972) and F_{ST} distance (Latter, 1972), which can be used for almost all allele frequency data including microsatellites. Other measures like $(\delta\mu)^2$ (Goldstein et al., 1995), and D_{SW} (Shriver et al., 1995) can only be used for microsatellite data. The information gained by distance matrices can be further used to construct phylogenetic trees. Suitable programs perform evolutionary analysis of allele frequency data and compute phylogenetic trees using the neighbor-joining (NJ) algorithm

(Saitou and Nei, 1987) or unweighted pair-group method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973).

An alternative way of exploring and visualizing dissimilarities or similarities between populations is to use Principal Coordinates Analysis (PCoA) or Principal Components Analysis (PCA) (Hirst and Jackson, 2007). Both start with a distance matrix and assign each data point to a location in a two-dimensional space (Legendre and Legendre, 1983). The main difference between PCoA and PCA is that PCoA is used for dissimilarity, PCA is used for similarity. Both produce a set of uncorrelated axes, each with an eigenvalue which indicates the amount of variation captured, to summarize the data. Ideally, a PCoA has only two or three axes with high eigenvalues that capture more than 50% of the variation. Every other axis has small eigenvalues (Jackson et al., 1989; Legendre and Legendre, 1983; Peres-Neto et al., 2003). In a PCoA plot data points are ordinated closer together, if they are similar and further away if they show more dissimilarities. PCA and PCoA will be similar if the data set is relatively small and has no missing data. However, Mohammadi and Prasanna (2003) found that PCoA treats missing data better than PCA and recommend PCoA over PCA when there is missing data in a dataset.

Population differentiation can also be detected by using an analysis called AMOVA. It stands for Analysis of Molecular Variance. AMOVA is estimating population differentiation directly from the molecular data and not from allele frequencies (Excoffier et al., 1992). AMOVA estimates how much of the differentiation between samples is due to differences between individuals, between individuals within a population, or between populations (Meirmans, 2012). This helps to interpret whether or not there is a genetic structure between populations. As an example, if the AMOVA results found 90% of the variation within populations and only 10% among populations, then the populations are not structured. If it is the other way around with 90% variation among populations, then the result implies significant genetic differentiation between populations and they can be considered as structured populations. AMOVA handles molecular data as a vector (Excoffier et al., 1992) and squared Euclidean distances are calculated for all pairwise groups of vectors, creating a matrix that is then further divided into smaller matrices that correspond to divisions within the population (Excoffier et al., 1992).

Another common way of analyzing populations is the F-statistics, since it calculates the expected level of heterozygosity at different population levels (Frydenberg et al., 2002; Hu et al., 2020; Morand et al., 2002; Razifard et al., 2020; Weir and Cockerham, 1984; Weir and Hill, 2002; Weissensteiner et al., 2020; Wondji et al., 2002). The foundation for this was laid by Fisher, Haldane and Wright, which made it possible to mathematically describe the genetic layout of a population. One of Wright's most influential studies was the development

of a fixation index. It allowed to account for the effect of inbreeding in a population and is defined as:

$$F = 1 - H_o/H_e$$

H_o is the observed heterozygosity and H_e is the heterozygosity expected under HWE (Wright, 1922). It describes the likelihood that two alleles at the same locus are identical by parentage. Wright expanded this model to be able to apply it to subpopulations in the total population (Wright, 1951), which led to the hierarchical F-statistics namely F_{IS}, F_{IT} and F_{ST}. In these statistics *I* stands for Individuals, *S* for Subpopulation and *T* for total population (Figure 5) and are calculated as followed:

$$F_{IS} = \frac{H_S - H_I}{H_S}; \ F_{IT} = \frac{H_T - H_I}{H_T}; \ F_{ST} = \frac{H_T - H_S}{H_T}$$

H_I is based on the observed heterozygosity in a population, H_S is based on the expected heterozygosity in populations, and H_T is based on the overall expected heterozygosity. The most common of the three values is the F_{ST} value. It is the proportion of total genetic variance in a subpopulation compared to the total genetic variation.
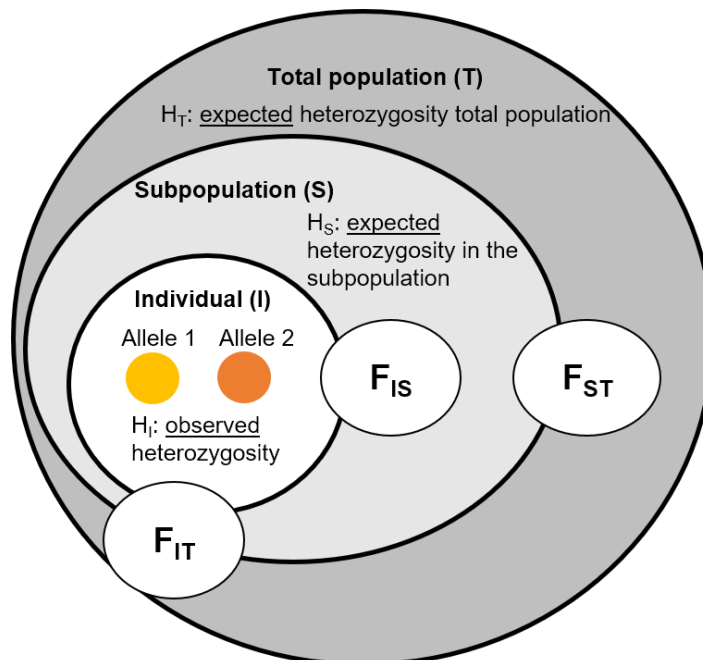
Figure 5: Illustration of classical F-statistics. Shown is the connection between different F-values in F-statistics. Dark grey = Total population or Metapopulation (T), light grey = Subpopulation (S), white = Individual in a subpopulation (I), yellow and orange circle = the two allele variants present in a single individual (I). H_I is based on the observed heterozygosity in a population, H_S is based on the expected heterozygosity in populations and H_T is based on the overall expected heterozygosity. F_{IS} is the inbreeding coefficient within individuals relative to the subpopulation, F_{IT} is the inbreeding coefficient within individuals relative to the total and F_{ST} is the inbreeding coefficient within subpopulations relative to the total.

Clustering is an important part of characterizing populations (Kopelman et al., 2015). A Bayesian clustering approach can be used to identify possible subpopulations in a metapopulation and to ascribe individuals to these populations based on their genotypes (Falush et al., 2003; Pritchard et al., 2000). Clustering methods define K groups/clusters (e.g. populations) for N objects (e.g. individual samples). Each group or cluster contains objects that are similar to each other and share certain characteristics, but the groups themselves are different from each other. When analyzing structured populations, the parameter K is of great importance, since it describes the number of subpopulations (Verity and Nichols, 2016). Unfortunately, the exact value of K cannot be calculated, instead, heuristic estimators are used to estimate the number of K (Verity and Nichols, 2016). The most likely K value can be detected according to Evanno and Pritchard, assessed through analysis of ΔK, the Dirichlet parameter alpha (α) or LnP(D)/L(K) distribution plots (Evanno et al., 2005; Hubisz et al., 2009). In any case, these K groups are characterized by allele frequency data (François et al., 2006; Kopelman et al., 2015; Pritchard et al., 2000). Each object (in this study an individual sample) is assigned to either one population or several populations if the genotype is admixed. The admixture model is used if the origin and degree of isolation in the sampled populations are unknown. It assumes that allele frequencies are correlated and each sample contains a portion of the genome of each 'original' population (Falush et al., 2003; Pritchard et al., 2000). Model-based clustering methods assume that this is the case, that all data in the dataset is admixed. This is an advantage compared to the more 'traditional' distance-based methods. Further, a model-based method assumes that all samples are randomly drawn and uses certain characteristics for clusters, and tries to optimize the match between the data and the characteristics (Pritchard et al., 2000; Yeung et al., 2001).

Demographic processes like migration, growth, or reduction of population size shape populations and their genetic diversity over time. Genetic assignment methods are used to describe the relationship between individuals of different (sub)populations. Individuals are either excluded or assigned to their most likely natal population. The information gained from this method is, if or if not an individual originates from a certain population, which enables the potential to estimate migration and movement between populations (Paetkau et al., 2004; Rannala and Mountain, 1997). The goal of this kind of method is to estimate a migration rate in a population system with ongoing gene flow (Paetkau et al., 2004). A variety of different methods and assignment criteria can be used to perform this statistic. Different genetic assignment criteria used for likelihood estimations are the genetic distance (Cornuet et al., 1999), allele frequencies (Paetkau et al., 1995) and the Bayesian Criterion (Rannala and Mountain, 1997).

Other demographic processes aside from migration are population bottlenecks (Figure 6) and founder events (Figure 7) that have led to reduced genetic diversity in many species. Domesticated species like dogs often show evidence for such effects in their population history (Marsden et al., 2016). Decreased diversity based on bottleneck events or founder effects is also found in natural populations such as monkeys (Hernandez et al., 2007). A bottleneck occurs when the size of a population is (drastically) reduced. The causes for such an event can be manifold, for example, the destruction of natural habitats, natural disasters like earthquakes and fires or diseases. As a result, a smaller population with smaller genetic diversity passes on genes to their offspring. The genetic diversity will remain low unless gene flow from another population occurs. A reduced genetic variation is problematic for a population, since it may not have the ability to adapt to new selection pressures (Ellstrand and Elam, 1993; Ouborg et al., 1991; Williams, 2001).
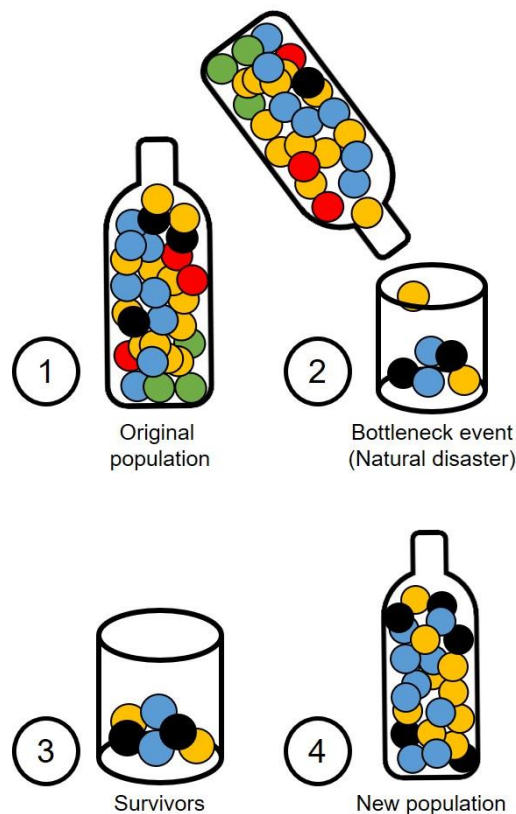
Figure 6: Bottleneck event in a population. Shown is the schematic overview of a population bottleneck. Different allele variants are represented as colored circles. The original population (1) undergoes a bottleneck event (2), for example, a forest fire. Only a small number of individuals with smaller genetic diversity survive (3), they start reproducing and generate a new population (4). In this example, the red and the green allele variants get lost, while the blue, yellow, and black variants are passed on to the next generation. Figure based on 'Population genetics: Figure 3,' by OpenStax College, Biology, CC BY 3.0.

The Founder effect (Figure 7) is similar to a population bottleneck but the main difference is that it occurs when a population is started by only a few members of the species (Barton and Charlesworth, 1984). Invasive species can undergo such an event when they are introduced to the newly invaded area. A program like Bottleneck v.1.2.2 (Piry et al., 1999) is used to determine if recent demographic events like expansion or mitigation in population size took place. The observed number of alleles (k) is used to calculate the distribution of the expected heterozygosity for each population and each locus. To do so, it is assumed that a mutation-drift equilibrium exists. This mutation-drift equilibrium describes a state at which the rate at which variation is lost through drift is equal to the rate at which mutation creates new variation (Piry et al., 1999). The observed heterozygosity (H_o) is compared to the expected heterozygosity (H_e) to calculate if there is an excess or deficit of heterozygosity and the distribution of the expected heterozygosity for each population and each locus is obtained by using one of three mutation models, the IAM, SMM and the TPM that were discussed above (Piry et al., 1999).
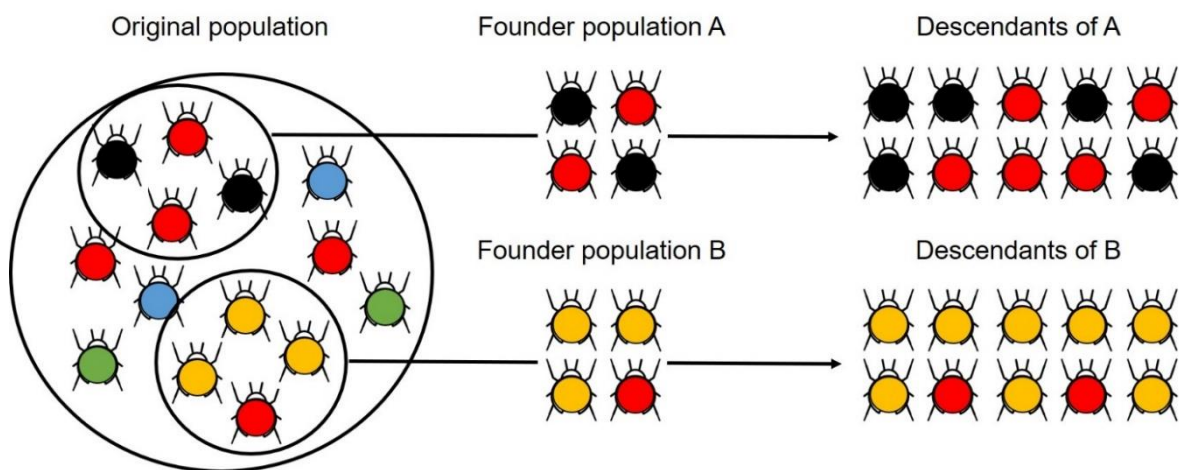
Figure 7: Founder effect on a population. Shown is a schematic illustration of how a founder effect affects the genetic diversity in a population. Different allele versions are represented by colored circles. A founder effect occurs when a small number of individuals get separated from the original population and establish a new colony. During this process, genetic diversity can easily get lost. In this example, only the red, black and yellow versions of alleles are carried over to the new population. Alleles are inherited by chance, depending on which individuals from the original population are present in the founder population. The blue and the green version are lost during the founding process but remain in the original population. Figure based on an image from (McCrone and Lauring, 2018).

2. Aims and Objectives

Drosophila suzukii (SWD) is a severe invasive pest species in America and Europe. It has drawn much attention as a crop pest of soft-skinned fruits since it causes substantial economic losses for fruit growers worldwide. Most studies regarding invasive species focus on how to prevent further expansions or how to reduce possible damages to crops, livestock, or human health. Knowledge about invasion history and population genetics can help to complement and improve these efforts. The goal of this study is to determine intra- and interspecific genetic diversity between different populations across Germany since little is known about the genetic diversity of SWD populations in this country. Moreover, it is of interest, if different genetically defined populations exist and if they exhibit a geographical pattern. Additional information about population development and behavior can be provided by comparing population genetic data over several years. Regarding pest control, it would be beneficial to know, if an annual reinvasion from other countries takes place or if SWD overwinters locally and reemerges in spring. Previous studies have shown that adult SWDs migrate over large distances to find suitable overwintering habitats and that they recover fast with a high survival potential (Dalton et al., 2011; Hamby et al., 2016; Stephens et al., 2015). Based on these findings, it is likely that individuals in Germany reemerge from local overwintering habitats and if differences between years occur that they turn out to be small. The comparison over several years could allow an estimate of the onward risk potential of SWD. It is known that invasive species often show a high genetic divergence, which allows them to adapt better to new environments. Hence, eleven *D. suzukii* populations from different areas in Germany were studied for three years using a set of microsatellite markers designed and tested for SWD by Fraimout et al. (2015). This is the first study that provides insights into the population genetics of SWD on a large-scale level in Germany as well as its development over a multi-year period. In addition, different laboratory strains from Europe and America were used as outgroups and to test the marker system upfront. A new laboratory strain that derives from one of the German field populations is used to illustrate the effect of laboratory breeding and the change of genetic markers over time.

3. Methods

3.1. Sampling and Identification of German SWD Populations

For this thesis, several sampling areas in Germany were studied. Sampling areas had heterogeneous vegetation, some collections sites were more rural and some others were closer to cities. Location sites were mapped using geographic coordinates with Simplemappr (Shorthouse, 2010), assuring that each year fruits from the same locations were collected. The German samples' names result from the vehicle registration plate of the respective district and the sampling year. Fruits were sampled by private persons or research and governmental institutes. Collectors were provided with a carton box to send samples back, including Drosophila food vials with a foam plug, sponge cloth, and padding material as well as a short instruction for the sampling procedure. Berries were filled into the provided food vials on the day of shipping, the sponge cloth was added to soak leaking fruit juice and the foam plug was used to prevent oxygen shortage/hypoxia in the vial. The packages were not sent during longer heat periods (>27°C), and only between Monday and Wednesday. To prevent sampling offspring of only one female fly, berries were sampled from different plants, if possible, or at different time points (Table 9 for more details).

Upon arrival in the laboratory in Giessen, berries were split into small groups and transferred to modified cages, which were made from plastic buckets (diameter 10 cm, Eimer-Welt.de, Hamburg) with two 4 cm big holes in the site, covered with a gaze. A small food vial containing a wet sponge cloth was clued to a petri dish to prevent low humidity and easy cleaning of the cage if necessary. Paper tissues or sponge cloth were used as padding material at the bottom to soak leaking fruit juice (Figure 8). The cages were cleaned at least once a week to remove fruit juice or fungus. They were kept at RT until no more larvae or pupae emerged from the fruits or until berries got moldy. If possible, pupae and larvae were transferred to fresh food vials before emerging to adults, to prevent adults from laying eggs in the cage and biasing sampling.

Adults were identified based on Hauser 2011 and pictures were taken with a Keyence VHX-5000 (Keyence Corporation, Osaka, Japan). For males, the main criteria were the spot on both wings, sex combs, and banding on the abdomen. For females banding on the abdomen and the ovipositor were the main characteristics (Figure 1). Animals without clear identification as well as *Drosophila melanogaster* were not used for further experiments and were discarded.
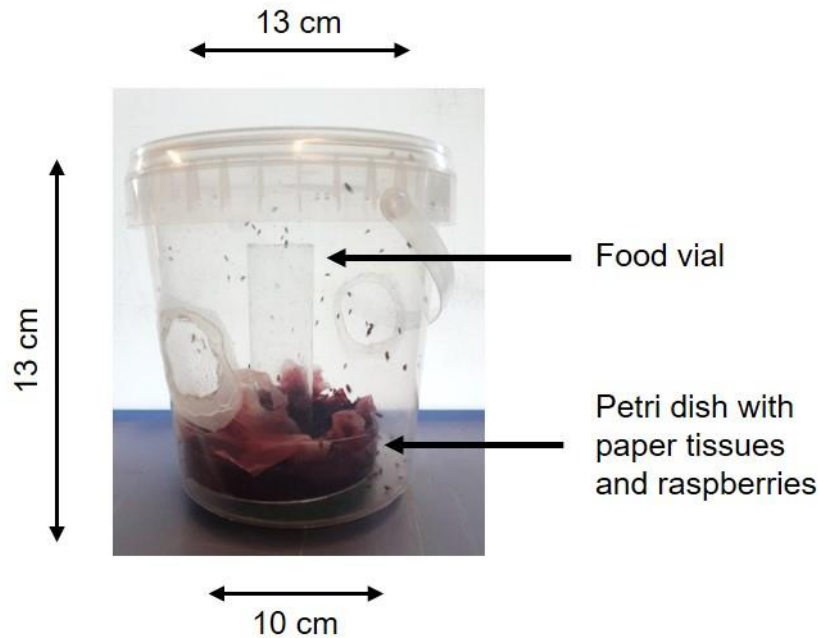
Figure 8: Fly cage for *D. suzukii* sampling. Invested fruits were put in a cage made from a plastic bucket. A wet sponge cloth in a drosophila food vial was used to raise the humidity in the bucket if needed. Paper tissues were used to absorb fruit juice. Two small holes on the side were cut into the bucket and covered with gaze to ensure a sufficient oxygen supply.

3.2. Laboratory Strains

Laboratory strains from the USA (LS_USA), Ontario (LS_Canada), Italy (LS_Italy), Kriftel (LS_Frankfurt), Valsugana (LS_Valsugana), and Strasbourg (LS_France) were used. LS_USA was established in 2010 from a field collection in North Carolina (Stockton et al., 2020), LS_Canada was started in 2012 (Jakobs et al., 2015; Renkema et al., 2015), LS_Italy was kept in the laboratory since 2014 and LS_Frankfurt was established in 2016 (Lee and Vilcinskas, 2017). LS_France was kindly provided by Eric Marois and originated from Strasbourg (France). Alberto Grassi provided LS_Valsugana from Valsugana in Italy. Both were collected in 2018 and kept in the laboratory since.

In addition, a new laboratory strain was established from the sampling location Bad Homburg (LS_HG). The samples named LS_HG18 and LS_HG19 originated from these flies and were collected for analysis after one (LS_HG18) and two years (LS_HG19) in culture, respectively. All *D. suzukii* laboratory strains were maintained on standard Drosophila medium at 25°C and 55% humidity with a 12 h-photoperiod and transferred to fresh media every week.

3.3. Molecular Techniques

Detailed protocols with step-by-step instructions for all molecular techniques used in this thesis can be found under B Laboratory Protocols. These protocols are adjusted to current (2021) laboratory equipment.

3.3.1. DNA Extraction

Total genomic DNA was extracted from whole insects. Samples were placed in lysis tubes with 1.4 mm ceramic spheres (Lysing Matrix D bulk, MP Biomedicals, Solon, OH), 200 µl sterile filtered homogenization buffer was added (1 M Tris-HCl (pH 7.5), 5 M NaCl, 0.5 M EDTA (pH 8), 0.3 M spermine tetra-HCl, 1 M spermidine tri-HCl, 1 g sucrose), and then homogenized at 6000 rpm for 40 sec in a Fast Prep-24™ (MP Biomedicals, Solon, OH). Afterward, 200 µl lysis buffer (1 M Tris-HCl (pH 9.0), 0.5 M EDTA (pH 8.0), 10% SDS, 1 g sucrose) was added and incubated at 70°C for 10 min. 60 µl of 8 M KOAc was added, and tubes were stored on ice for 30 min. After transferring the suspension to a new tube, samples were centrifuged at 12000 rpm for 10 min at 4°C. DNA was precipitated with two volumes of ice-cold 100% EtOH and stored at -20°C overnight. DNA was pelleted by centrifugation at 12000 rpm for 40 min at 4°C. The pellet was washed with 30 µl of 70% ice-cold ethanol for 10 min while centrifuging at 12000 rpm at 4°C. Pellets were air-dried and resuspended in 50 µl H_2O for further use in Multiplex PCR or in 1x TE buffer (1 M Tris-HCl (pH 7.5), 500 mM EDTA (pH 8)) for long term storage.

3.3.2. Determination of DNA Concentration

DNA concentration of plasmids and purified PCR products was determined using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). For the quantification 2 µl reference and 2 µl of the sample were pipetted. Absorbance was measured at 260nm and the purity of samples was determined using the 260/280nm ratio. This ratio should be between 1.7 and 1.8, lower or higher values can indicate contamination with protein or RNA.

The concentration of PCR products generated by multiplex PCR for fragment length analysis was assessed using gel electrophoresis on 3% agarose gels, stained with SYBR® Safe DNA Gel Stain (Invitrogen, Massachusetts, USA), and visualized under UV light. The intensity of undiluted sample DNA was compared to that of the Quick-Load® Purple 100 bp DNA Ladder (New England Biolabs, NEB, Massachusetts, USA), which was used as a reference. If the intensity of the 500 bp band (97 ng, according to the manufacturer) is similar to a 2 µl undiluted sample, then the sample concentration is 48.5 ng/µl (97 ng divided by 2 µl). Because a multiplex PCR exhibits multiple bands it was important to check the band intensity between samples to guaranty similar concentrations.

3.3.3. Polymerase Chain Reaction (PCR)

Amplification of a specific DNA sequence was done by performing a Polymerase Chain Reaction (PCR). PCR was performed under the following conditions with either Platinum polymerase (5 U/µl) (Invitrogen) (Table 1, Table 2) or Q5 Polymerase (2,000 U/ml) (NEB) (Table 3, Table 4). Primers used in this thesis are listed in Table A.2.4.1 and Table A.2.4.2. PCR primer pairs were between 18-25 bp long and had a melting temperature between 57 - 62°C. Primers were designed with Primer 3 implemented in Geneious Prime 2019.2 software and tested with IDT OligoAnalyzer 3.1 (IDT, Iowa, USA) as well. Primers were shipped dry and adjusted to 100 µM with 1x TE buffer. Afterward, primers were adjusted to 10 µM with HPLC-H_2O. A BioRad C1000 touch (BioRad, California; USA) was used as a thermal cycler.

Table 1: Reaction setup for PCR using Platinum polymerase. Shown are the components of a single reaction and the amount needed for either a 25 µl or a 50 µl reaction.

Component	25 µl rxn	50 µl rxn
Water, nuclease-free	to 25 µl	to 50 µl
10x PCR Buffer, -Mg	2.5 µl	5 µl
50 mM $MgCl_2$	0.75 µl	1.5 µl
10 mM dNTP Mix	0.5 µl	1 µl
10 µM forward Primer	0.5 µl	1 µl
10 µM reverse Primer	0.5 µl	1 µl
Template DNA	varies	varies
Platinum Taq DNA Polymerase	0.1 µl	0.2 µl

Table 2: Cycling parameter for PCR using Platinum polymerase. Shown is each step of the PCR cycle with the corresponding temperature, time, and the number of cycles.

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	2 min	1
Denaturation	95°C	30 sec	35
Annealing	Tm-5	30 sec	
Extension	72°C	1:30 min	
Final Extension	72°C	5 min	1

Table 3: Reaction setup for PCR using Q5 polymerase. Shown are the components of a single reaction and the amount needed for either a 25 µl or a 50 µl reaction.

Component	25 µl rxn	50 µl rxn
Water, nuclease-free	to 25 µl	to 50 µl
5x Q5 Reaction Buffer	5 µl	10 µl
10 mM dNTP Mix	0.5 µl	1 µl
10 µM forward Primer	1.25 µl	2.5 µl
10 µM reverse Primer	1.25 µl	2.5 µl
Template DNA	varies	varies
Q5 High-Fidelity DNA Polymerase	0.1 µl	0.2 µl

Table 4: Cycling parameter for PCR using Q5 polymerase. Shown is each step of the PCR cycle with the corresponding temperature, time, and the number of cycles.

Step	Temperature	Time	Number of Cycles
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	35
Annealing	Tm-5	20 sec	
Extension	72°C	2:30 min	
Final Extension	72°C	5 min	1

3.3.3.1. Colony PCR

A colony PCR was performed to determine the presence or absence of insert DNA in plasmids. For this, the primer pairs designed for PCR of the target sequence were used (Table A.2.4.1 and Table A.2.4.2). Template DNA for the PCR originated from picked *E. coli* colonies. PCR reactions were performed with Dream Taq DNA Polymerase (5 U/µl) (Thermo Fisher Scientific, MA, USA) in 50 µl volume, 0.2 mM dNTPs, and 10x Dream Taq Buffer (Table 5). Cycling parameters are given in Table 6.

Table 5: Reaction setup for PCR using DreamTaq polymerase. Shown are the components of a single reaction and the amount needed for either a 25 µl or a 50 µl reaction.

Component	25 µl rxn	50 µl rxn
Water, nuclease-free	to 25 µl	37.75 µl
10x DreamTaq Buffer	2.5 µl	5 µl
2 mM dNTP Mix	2.5 µl	5 µl
10 µM forward Primer	0.5 µl	1 µl
10 µM reverse Primer	0.5 µl	1 µl
Template	Bacterial clone	Bacterial clone
DreamTaq DNA	0.125 µl	0.25 µl

Table 6: Cycling parameter for PCR using DreamTaq polymerase. Shown is each step of the PCR cycle with the corresponding temperature, time, and the number of cycles.

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	3 min	1
Denaturation	95°C	30 sec	35
Annealing	Tm-5	30 sec	
Extension	72°C	1 min	
Final Extension	72°C	5 min	1

3.3.3.2.　Multiplex PCR for Microsatellite Analysis

Primers for FLA were designed and published by Fraimout (2015). Each forward primer (Metabion, Planegg, Germany) was marked with a fluorescent dye, while the corresponding reverse primer (Eurofins Genomics GmbH, Ebersberg, Germany) was unmarked (Table A.2.4.2).

DNA amplification for fragment analysis was performed using the Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany) in 10 µL final reaction volume, containing 1x QIAGEN Multiplex PCR Master Mix, 0.2 µM primer mix, 0.5x Q-Solution, and 100 ng of genomic DNA (Table 7). The PCR cycling protocol was: 95°C, 5 min; 32 cycles of 95°C for 30 s, 57°C for 90 s, 72°C for 3 min; final elongation at 72°C for 30 min, the latter is advised to be used for analysis on capillary sequencers (Table 8).

Table 7: Reaction setup for PCR using Qiagen Multiplex PCR Kit. Shown are the components of a single reaction and the amount needed for either a 10 µl or a 50 µl reaction.

Component	10 µl rxn	50 µl rxn
Water, nuclease-free	to 10 µl	to 50 µl
2x Multiplex PCR Master Mix	5 µl	25 µl
5x Q-Solution	1 µl	5 µl
10x primer mix, 2 µM each primer	1 µl	5 µl
Template DNA	100 ng	100 ng

Table 8: Cycling parameter for PCR using Qiagen Multiplex PCR Kit. Shown is the step of the PCR cycle with the corresponding temperature, time, and the number of cycles.

Step	Temperature	Time	Number of Cycles
Initial PCR	95°C	5 min	1
Denaturation	95°C	30 sec	32
Annealing	57°C	90 sec	
Extension	72°C	3 min	
Final Extension	72°C	30 min	1

3.3.4. PCR Purification

PCR products were either purified with the Zymo Clean and Concentrator-5 Kit or the Zymo Clean and Concentrator-25 Kit (Zymo Research), which provide PCR purification of up to 5 µg or 25 µg DNA, respectively. An alternative method was to excise gel bands from an agarose gel and purify them with the Zymo Gel DNA Recovery Kit (Zymo Research). DNA for FLA was eluted in H_2O, DNA for Sanger sequencing was eluted using 1x TE buffer.

3.3.5. Gel Electrophoresis

Agarose gels were used to separate and analyze DNA fragments. For the preparation of an agarose gel, 1x TAE buffer, agarose, and SYBR® Safe DNA Gel Stain (Invitrogen, California, USA) were used. A 1.5% agarose gel was used to analyze PCR fragment length. To determine the concentration of multiplex PCR products for fragment length analysis, a 3% agarose gel was used. The DNA samples were loaded into the wells with Purple Gel Loading Dye (6X) (NEB). In addition, 3 µl of the Quick-Load® Purple 100 bp DNA Ladder (NEB) was loaded into a separate well on each gel, which is a molecular weight DNA marker and functioned as a reference. Agarose gels were run in 1x TAE buffer at 90 – 120 V for 40 to 60 min. For visualization of the DNA fragments, the VersaDoc and Quantity One Software (4.6.9) were used.

3.3.6. Preparation of *E. coli* Competent Cells

E. coli competent cells were prepared using the Mix & Go *E. coli* Transformation Kit & Buffer Set from Zymo Research (Zymo Research). Approximately 10 µl of competent cells were transferred in 4 to 5 ml of LB medium. Cells were then grown at 37°C at 110 rpm for one day. In the evening 50 ml ZymoBroth medium were inoculated with 500 µl of the respective *E. coli* strain and grown at 20°C at 110 rpm overnight. On the next day, cells were harvested by centrifugation at 4°C and 1,600 g for 10 min. The cell pellet was resuspended in 5 ml wash buffer and re-pelleted. Supernatant had to be removed completely and cells were resuspended in 5 ml Competent buffer. A final volume of 50 µl solution was transferred into precooled 1.5 ml reaction tubes and immediately frozen in liquid nitrogen. Competent cells were stored at -80°C.

3.3.7. Cloning

Purified PCR products were cloned into the pCR4-TOPO TA vector or the Zero Blunt TOPO PCR Cloning vector (both Thermo Fisher Scientific, MA, USA), depending on the polymerase used in PCR. A ligation mixture of 1.5 µl was prepared by mixing 1 µl purified PCR product, 0,25 µl Salt-solution and 0,25 µl vector and incubated for 30 min at RT. An aliquot of competent *E. coli* cells was thawed on ice. A total of 1.1 µl ligation mixture was added to the cells and incubated on ice for 30 min. The cells were heated at 42°C in a water bath for 40 sec and placed back on ice for 2 min. 250 µl SOC medium was added to the cells and incubated for 60 min at 37°C at 220 rpm. After the incubation 200 µl were plated on LB agar plates containing the appropriate antibiotic. The remaining 50 µl were stored at 4°C and kept in case the LB agar plates were overgrown. The plates were incubated at 37°C overnight.

3.3.8. Overnight Culture

For overnight cultures, 5 ml or 50 ml of LB medium with an appropriate antibiotic were inoculated with either a single colony of an LB agar plate or from a cryo-stock. The cultures were grown at 37°C for 16 h. 5 ml cultures were shaken at a speed of 220 rpm and 50 ml cultures at a speed of 180 rpm.

3.3.9. Plasmid Isolation

Plasmid isolation was performed with the NucleoSpin Plasmid Kit (Macherey-Nagel). Cells of 4 ml overnight culture were used according to the vendor's protocol. The plasmid DNA was eluted in 50 µl elution buffer.

3.3.10. Restriction Enzyme Digestion

This method was performed as an alternative to the colony PCR to ensure a successful transformation. Restriction was carried out by using EcoRI-HF and CutSmart buffer (NEB, Thermo Fisher Scientific). The digestion was performed at 37°C for 60 min. The results of the digestion were checked by gel electrophoresis on a 1.5% agarose gel.

3.3.11. Sanger Sequencing

The sequencing reaction for purified PCR products and cloned fragments was carried out by Macrogen (Amsterdam, Netherlands). For sequencing, 500 ng of purified plasmid DNA were used, or 75 ng of purified PCR product, mixed with 2.5 µl of the sequencing primer of the corresponding plasmid or PCR product and filled to a total volume of 10 µl with HPLC-H_2O. Sequence data in .ab1-format was analyzed with Geneious Prime 2019.2 (Biomatters Ltd, Auckland, New Zealand). Regions with bad quality at the 5' and 3'-end were trimmed and sequences were checked for vector contaminations with the 'Trim Ends' function implemented in Geneious Prime.

3.3.12. Fragment Length Analysis (FLA)

Samples were sent for fragment analysis on an ABI 3730 Genetic Analyzer to StarSEQ (Mainz, Germany). For each sample, a total volume of 10 µl was pipetted in a 96 well plate and sealed using pierceable heat seal foil (BioRad) sealed with a BioRad PX1 plate sealer. GeneScan™-500LIZ™ (ThermoFisher Scientific, Waltham, Massachusetts) was used as an internal size standard. Fragments were sized with the Geneious Prime 2019.2 software. If no sample amplification was obtained after three attempts, the locus was classified as missing data.

3.4. Bioinformatic Methods and Statistical Analysis

Detailed protocols with step-by-step instructions for all bioinformatics methods used in this thesis can be found under C Bioinformatic Protocols for Microsatellite Analysis.

3.4.1. Online-Tools and Databases

Location sites were mapped using geographic coordinates with Simplemappr[2] (Shorthouse, 2010). Coordinates from each sample site were taken to assure that each year fruits from the same locations were collected.

BLAST (Altschul et al., 1990), the Basic Local Alignment Search Tool, was used to compare nucleotide sequences with sequences from databases by finding similarities between those sequences. It is available at the website of the National Center for Biotechnology Information (NCBI[3]).

The Spotted Wing Fly Base[4] (Chiu et al., 2013) was used to analyze and compare nucleotide sequences of SWD with the *D. suzukii* genome or to search for specific genes of interest. In contrast to the BLAST function in NCBI, nucleotide sequences in the Spotted Wing Fly Base are specifically compared to the SWD genome.

3.4.2. Geneious

The Geneious Prime 2019.2 software is a bioinformatics software platform that was used for Sanger sequence analysis, including trimming, pairwise and multiple alignments, mapping, chromatogram analysis, annotation, primer design, and microsatellite analysis with the microsatellite external plugin. The latter allows streamlined microsatellite genotyping. ABI fragment analysis files can be imported and it is possible to visualize traces, fit ladders, call peaks, bin them and produce a table of genotypes to export for further analysis in GenAlex software v.6.41 (Peakall and Smouse, 2012). Before starting the analysis, the locus information has to be set, which is based on the characteristics of the microsatellites used. It includes the dye used (FAM, HEX, ROX, TAMRA), the expected number of peaks (usually two for a diploid organism), the repeat unit (e.g. dinucleotide repeat), and the microsatellite range (e.g. 160 bp to 320 bp, this information was obtained from the original paper Fraimout 2015). First, the ladder has to be called correctly. Since StarSeq used the GeneScan™ 500 LIZ™ Size Standard (ThermoFisher Scientific, Waltham, Massachusetts) for the FLA, this ladder had to be called and recognized in all ABI files. For binning the '3rd Order Least Squares' sizing algorithm was chosen. It was chosen

---

[2] https://www.simplemappr.net/
[3] https://blast.ncbi.nlm.nih.gov/Blast.cgi
[4] http://spottedwingflybase.org

over the alternative 'Local Southern' and 'Cubic Spine Interpolation' algorithms because it uses regression analysis to build a best-fit-size-calling curve, it compensates for any fragment that may run anomalously and results in the least amount of deviation for all the fragments, including size standard and samples. 3rd order was chosen over 2nd order because it uses a higher polynomial degree and captures more of the peak structure. It also provides more flexibility when generating best-fit curves for sizing samples with anomalously migrating fragments.

3.4.3. Population Genetics Software

GenAlex software v.6.41 (Peakall and Smouse, 2012) is a plugin for Microsoft Excel and offers population genetic analysis tools. It was used for basic frequency-based and distance-based analysis of microsatellite data. Frequency-based statistical procedures included the calculation of allele frequencies at each locus for every population, the number of alleles (N_a), the effective number of alleles (N_e), the observed (H_o) and expected (H_e) heterozygosity, private alleles, and Fixation index (F) together with the mean over loci or populations and the standard error. GenAlex was further used to output the classical Wright's F statistics (Wright 1946, 1951, 1965). For the allelic pattern a graphical output is available that is summarized by the number of alleles (N_a) across loci, N_a with a frequency higher than 5%, the effective number of alleles (N_e), the number of private alleles (A_P), number of locally common alleles with a frequency higher than 5% found in less than 25% and/or less than 50% of populations and the expected heterozygosity (H_e). The pairwise population Nei's Genetic Distance/Identity and pairwise F_{ST} between populations is given by frequency-based statistics too. A major part of the distance-based statistical options in GenAlex is the Principal Coordinates Analysis (PCoA). PCoA was used to plot the dissimilarities within a multivariate dataset. GenAlex was further used to calculate deviations from Hardy-Weinberg equilibrium (HWE) by performing a Chi-Square test of HWE to test if H_o is consistent with the expectations under HWE. The null hypothesis (H0) states that populations are randomly mating and the alternative hypothesis (H1) states that populations are not randomly mating.

FreeNA (Chapuis and Estoup, 2007) was used to estimate null allele frequencies for each locus analyzed following the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). It was used during the first experiments to exclude loci that show a certain degree of null alleles. Cervus 3.0 software was used to calculate the Polymorphism Information Content (PIC) (Kalinowski et al., 2007). The PIC calculation is included in the summary statistics in the 'Allele Frequency Analysis' module.

A hierarchical analysis of molecular variance (AMOVA) was obtained using Arlequin v.3.5.2.2 (Excoffier and Lischer, 2010). The input file was generated using the export function in GenAlex. The option 'Locus by Locus AMOVA' was chosen in the settings tab of the program. In the same tab, the option 'Compute distance matrix' from the dropdown menu was used and the number of permutations was set to 1000. It was further used to test for Linkage disequilibrium (LD), which refers to the nonrandom association of alleles at different loci. Detecting LD does not confirm either linkage or a lack of equilibrium in a locus or population (Slatkin, 2008). For each pair of loci in each population, the genotypic linkage disequilibrium was tested using a Chi-Square test (Di Rienzo et al., 1994) and a Markov chain method with 1000 iterations. This test assumes that populations are in Hardy-Weinberg equilibrium.

STRUCTURE 2.3.4 (Falush et al., 2003; Pritchard et al., 2000) was used to investigate the number of genetically distinct clusters (K) in a dataset. The input data file was obtained by using the export function in GenAlex. Each analysis was run with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions and a burn-in period of 100,000 repetitions using 20 iterations of K = 1-20. The analysis was further run using the admixture ancestry model. Results from STRUCTURE were summarized using STRUCTURE HARVESTER[5] (Web v0.6.94 July 2014) (Earl and vonHoldt, 2012). It was used to detect the most likely K value according to Evanno and Pritchard, assessed through analysis of ΔK, the Dirichlet parameter alpha (α) and LnP(D)/L(K) distribution plots (Evanno et al., 2005; Hubisz et al., 2009). Data for the most likely K was then loaded into the Program Pophelper to visualize the final result. Pophelper Structure Web App v1.0.10[6] (Francis, 2017) was used to align assignment clusters across replicate runs and visualize the results. STRUCTURE was set to run 20 iterations per K and this tool was used to merge and align all repetitions, therefore generating one evaluable graph.

An unrooted Neighbor-Joining (NJ) dendrogram was constructed with PoptreeW[7] (Takezaki et al., 2014) based on Da distance (Nei et al., 1983), which is a genetic dissimilarity coefficient that is based on mutation and drift. It is defined as $D_A = 1 - \frac{1}{r}\sum_j^r \sum_j^{m_j} \sqrt{x_{ij} y_{ij}}$, where r is the number of loci used, m_j is the number of alleles at the j-th locus and x_{ij} and y_{ij} are the frequencies of the i-th allele at the j-th locus in populations X and Y (Nei et al., 1983; Takezaki et al., 2014). A test for robustness was carried out by using a bootstrap value of 10,000.

---

[5] http://taylor0.biology.ucla.edu/structureHarvester/

[6] http://pophelper.com/

[7] http://www.med.kagawa-u.ac.jp/~genomelb/takezaki/poptreew/

GENECLASS2 (Piry et al., 2004) was run to determine the probability of each individual originating from the sample area or another reference population. Genetic assignment methods resolve population structures and relationships on an individual level (Estoup and Angers, 1998). Since this method gives an estimate of where individuals originate from, it is possible to detect immigrant individuals and to predict the real-time dispersal of a species/population (Piry et al., 2004; Rannala and Mountain, 1997). The standard Bayesian criterion of Rannala and Mountain (1997) and the Monte Carlo resampling method of Paetkau et al. (2004) was used with an alpha value of 0.05. Results were based on 10,000 simulated genotypes for each population and a threshold probability value of 0.05.

Bottleneck v.1.2.2 (Piry et al., 1999) was used to determine if recent demographic events like expansion or mitigation in population size took place. The two-phase model (TPM) and the stricter Stepwise Mutation Model (SMM) were used with model options for the TPM of 80% single-step mutations, a variance among multiple steps of 12 and with 5,000 iterations. Wilcoxon's signed-rank test was used to assess the probability of heterozygosity.

SigmaPlot for Windows Version 14.0 (© 2017 Systat Software, Inc.), a tool for statistics and data analysis, was only used in some cases to verify or to reconstruct results obtained from other programs. Measures of diversity were analyzed using repeated-measures analysis of variance (ANOVA) to determine differences among populations and years. Pairwise comparisons were obtained from posthoc Bonferroni correction.

A sample size estimation was conducted to determine the minimum number of observations required for the experiments. Under the assumption, that the population standard deviation of allele size is at most seven (based on preliminary exploratory data analyses) and under the requirement that a 90%-confidence interval (CI) for the population mean of allele size has a length of at most 10 with a probability of 80% ('precision power'), a sample size of at least 19 was necessary. Based on this sample size estimation, we chose to use 20 individuals from each location and year except for HB17, where only 10 individuals were available.

4. Results

4.1. Fly Sampling

Adult *D. suzukii* were collected between June and September in 2017, 2018, and 2019 from eleven different locations by collecting fruit samples infested with *D. suzukii* eggs, larvae, or pupae (Figure 9). Collected fruits included elderberries, cherries, raspberries, strawberries, and blackberries. Strawberries were the only fruits that were not infested in either of the three years. If possible, to prevent a biased result by sampling, related individuals and fruits of different shrubs or locations in a 20 km radius were collected (Table 9). As an alternative, fruits were collected at different time points, for example, once in June and later in September. In the cases of Kassel and Bremen, this was not possible (Table 9). Names for German samples were chosen from the vehicle registration plate of the respective district, the sampling year, and a personal 3-digit ID for each individual. Thus, the first individual collected in Dortmund in the year 2017 would be named DO17001.

Location acronym	Location	Coordinates	Collection Month 2017	Collection Month 2018	Collection Month 2019
DO	Dortmund	51.300499 7.674342	August	July	August
FF	Frankfurt (Oder)	52.325507 14.497815		August	August
FR	Sasbach	48.678752 8.030775	August	August	September
HB	Bremen	52.861452 9.370046	September		
HG	Bad Homburg	50.229450 8.622310	August	August	August
HH	Hamburg	53.480480 10.222784			August
HOH	Hofheim	50.110740 10.511780	August	June	July
KS	Kassel	51.370993 9.416376	August	August	August
LB	Gerlingen	48.792326 9.05524	August	June	July
PM	Derwitz	52.397132 12.823140	September	September	August
R	Regensburg	49.077134 12.208671	September	August	August

Figure 9: Sample sites and sampling time points of *D. suzukii* wild populations in Germany. Population location acronym, full location name, coordinates, and sampling year and month are listed in the table (right). The map (left) was generated by SimpleMappr and coordinates are listed in the table.

For all experiments on the eight laboratory strains, 20 specimens from each strain were used (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_France, LS_Valsugana, LS_HG18, and LS_HG19). The two 'oldest' laboratory strains are LS_USA and LS_Canada, followed by LS_Italy and LS_Frankfurt. LS_USA is the oldest laboratory strain and was established in 2010 from a field collection in North Carolina (Stockton et al., 2020). In 2012,

the laboratory strain LS_Canada was established (Jakobs et al., 2015; Renkema et al., 2015), LS_Italy was kept in the laboratory since 2014. The youngest of the four laboratory strains is LS_Frankfurt, established in 2016 (Lee and Vilcinskas 2017). LS_France was kindly provided by Eric Marois and originated from Strasbourg (France). Alberto Grassi provided LS_Valsugana from Valsugana in Italy. Both strains were collected in 2018 and kept in the laboratory. The laboratory strains LS_HG18 and LS_HG19 originated from flies collected in 2017 in Bad Homburg (HG), reared in the laboratory, and analyzed again after one (LS_HG18) and two years (LS_HG19) in culture, respectively. With Dr. Gerrit Eichner[8], a sample size estimation was conducted to determine the minimum number of observations required for the experiment. Under the assumption that the population standard deviation of allele size is at most seven (based on preliminary exploratory data analyses) and under the requirement that a 90%-confidence interval (CI) for the population mean of allele size has a length of at most 10 with a probability of 80% ('precision power'), a sample size of at least 19 was necessary. Based on this sample size estimation, 20 individuals from each location and year were used except for HB17, where only ten individuals were available. The final data set included 550 individuals from Germany and 160 individuals from laboratory strains, totaling 28 different German populations from 11 sample sites and over three years and eight laboratory populations (Table 9).

---

[8] Justus-Liebig-University Giessen, Mathematical Institute, Arndtstrasse 2, 35392 Giessen, Germany

Table 9: Number of SWD individuals sampled and used for analysis per year and population. Shown is the number of SWD individuals used for analysis, including the laboratory strains.

Population	Sampling areas	Total number of sampled SWD			Number of SWD used for analysis			
		2017	2018	2019	2017	2018	2019	Σ
DO	2	41	26	37	20	20	20	60
FF	3	0	46	37	0	20	20	40
FR	4	34	25	33	20	20	20	60
HB	1	10	0	0	10	0	0	10
HG	1	47	34	48	20	20	20	60
HH	2	0	0	39	0	0	20	20
HOH	3	43	33	35	20	20	20	60
KS	1	24	28	32	20	20	20	60
LB	4	39	57	41	20	20	20	60
PM	3	48	32	52	20	20	20	60
R	4	39	29	44	20	20	20	60
LS_France	0	0	20	0	0	20	0	20
LS_Valsugana	0	0	20	0	0	20	0	20
LS_Italy	0	20	0	0	20	0	0	20
LS_Frankfurt	0	20	0	0	20	0	0	20
LS_Canada	0	20	0	0	20	0	0	20
LS_USA	0	20	0	0	20	0	0	20
LS_HG18	0	0	20	0	0	20	0	20
LS_HG19	0	0	0	20	0	0	20	20
Total number of individuals used in the analysis								**710**

4.2. Initial Verification of the Marker System

The SSR markers were tested before the actual experiments to establish a working microsatellite marker system for German and laboratory SWD strains and to better understand the microsatellite marker characteristics. In Fraimout 2015, microsatellite loci were multiplexed because multiplex PCR allows amplifying several different DNA sequences simultaneously by using multiple primers (Hayden et al., 2008). Multiplex primers must be optimized to produce the desired amplicons in good quality and quantity. Primers should have a similar annealing temperature, amplicon sizes have to be different to form distinct bands, and amplicons that overlap in size must be distinguished using fluorescently-labeled primers. This approach offers considerable cost and labor savings (Hayden et al., 2008). Based on these considerations, the primer combinations used in Fraimout (2015) were tested on their compatibility and finally split into four primer sets (Table 10). Unfortunately, this information is not given in the publication of Fraimout (2015) but was personally communicated with Mr. Fraimout via E-Mail.

Table 10: Four different multiplex primer sets used in Fraimout (2015). The table contains the names of the primers that were split into four different multiplex primer sets. Boldface indicates primer pairs that were used for initial testing in this study.

Kit 1

Set 1	DS05	DS09	DS12	DS15	DS16	DS33		
Set 2	DS08	**DS17**	DS19	DS25	DS38	DS39		

Kit 2

Set 3	**DS14**	**DS34**	**DS32**	**DS35**	**DS22**	**DS20**	**DS21**	**DS23**
Set 4	**DS06**	**DS28**	**DS26**	**DS36**	**DS11**	**DS27**	**DS07**	**DS10**

Since Kit 2 contains four more primer pairs than Kit 1, unlabeled primers (A.2.4. Oligonucleotide Primers) were ordered and tested for all loci in Kit 2 Set 3 and Kit 2 Set 4 and in addition primers for DS17 since its amplicon size would fit into Kit 2 Set 3. One individual from the laboratory strain LS_USA was selected and genomic DNA extracted to serve as a template for the following experiments. First, unlabeled primers were used to generate DNA sequences that were cloned and sequenced. This allowed the identification of the genomic location and more detailed characterization of each microsatellite marker. The Sanger sequencing confirmed the repeat motifs given in the publication of Fraimout (2015). From the 17 tested markers, ten loci were identified as homozygous according to the sequencing result (Table 12). Two loci, DS20 and DS21, showed a second repeat motif (CA) that was not mentioned in the original paper of Fraimout (2015). A BLAST search in the Spotted Wing Fly Base indicated that all SSR markers are located in non-coding genome regions (Table 12).

In another experiment, the multiplex PCR protocol was tested. First, the standard multiplex PCR protocol was used as described in the Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany). This PCR resulted in rather faint bands (Figure 10). Different settings like the primer concentration, extension time, and the additional use of Q-solution included in the Qiagen Multiplex PCR Kit (Table 11) were altered to improve the clarity of the bands. Different annealing temperatures in the range of 61°C to 55.2°C resulted in similar results, while bands became faint below a temperature of 55°C (Figure 10). Q-Solution and a three-minute extension time improved the PCR reaction further, while a lower primer concentration did not influence the result (Figure 10).

Table 11: Variations in the multiplex protocol setup. Four changes were tested to improve the multiplex PCR reaction: the additional use of Q-Solution, a lower primer concentration, longer extension time, and different annealing temperatures. Changes to the protocol are marked in blue. If not stated otherwise, the parameters were according to the original publication Fraimout (2015).

Q-Solution	Reaction setup: additional use of Q-Solution	
	20 µl rxn	**Component**
	to 20 µl	Water, nuclease-free
	10 µl	2X Multiplex PCR Master Mix
	2 µl	5X Q-Solution
	2 µl	10X primer mix
	100 ng	Template DNA

Primer	Reaction setup: lower primer concentration	
	20 µl rxn	**Component**
	to 20 µl	Water, nuclease-free
	10 µl	2X Multiplex PCR Master Mix
	1 µl	10X primer mix
	100 ng	Template DNA

Extension time	Cycling parameter with 3 min extension time		
	Initial activation	95°C	5 min
	Denaturation	95°C	30 sec
	Annealing	57°C	90 sec
	Extension	72°C	3 min
	Number of cycles: 32		
	Final Extension	72°C	30 min

Annealing temperature	Cycling parameter different annealing temperatures		
	Initial PCR activation	95°C	5 min
	Denaturation	95°C	30 sec
	Annealing	54.1 - 61°C	90 sec
	Extension	72°C	1 min
	Number of cycles: 32		
	Final Extension	72°C	30 min
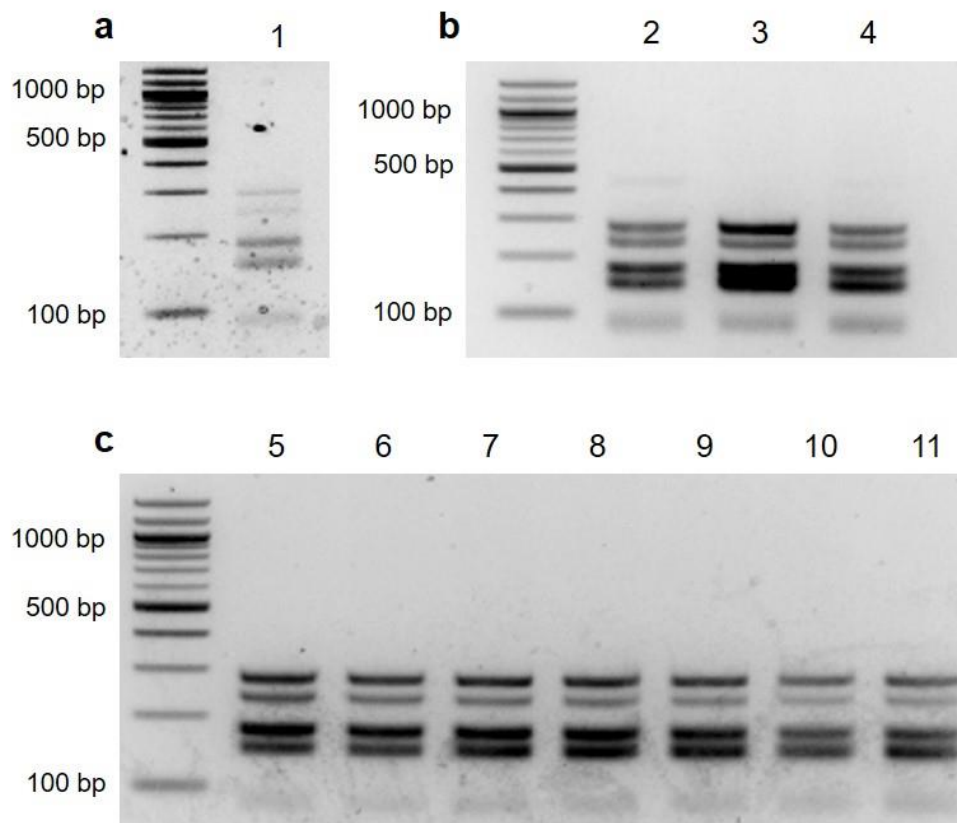
Figure 10: Agarose gel analysis of different multiplex PCR test conditions. a) 1 = standard protocol suggested by the manufacturers protocol. b) protocol from Fraimout (2015) with different changes, 2 = reduced primer concentration, 3 = additional use of Q-Solution in the PCR reaction, 4 = three-minute-long extension time. c) PCR protocol according to Fraimout (2015) with different annealing temperatures. 5 = 61°C, 6 = 59.8°C, 7 = 58.6°C, 8 = 57.1°C, 9 = 55.2°C, 10 = 54.6°C, 11 = 54.1°C. Size ladder used in this image: Quick-Load® 100 bp DNA Ladder (NEB).

Based on these results, it was decided to use the protocol written down in 3.3.3.2 (Multiplex PCR for Microsatellite Analysis) and to order oligonucleotides with fluorescent dyes for Kit2 Set 3 and Set 4 and DS17. These primers were then used to do a first Fragment Length Analysis (FLA) at StarSeq (Mainz, Germany). To test for different allele size variants, each locus was sequenced several times. The locus was classified as a homozygote if no alternative allele variant was identified after ten sequencing reactions. The FLA later confirmed the zygosity and the differences in repeat length that were identified by Sanger-sequencing (Table 12). The three loci, DS17, DS26, and DS32, were excluded from further experiments because the FLA chromatogram was of poor quality even after several replications. The remaining 14 microsatellite markers were used for FLA in the German and laboratory samples.

Table 12: Initial testing of 17 microsatellite markers. Sanger sequencing results show 17 microsatellite markers, compared to FLA results from one individual from LS_USA. For each locus, the published repeat motif was confirmed by Sanger sequencing, the repeat length in bp was checked for different size variants, sequences were checked for noticeable problems, Sanger and FLA results were compared, and the sequence information was used to define the genomic location of each marker in the SWD genome.

Marker	Correct repeat motif	Repeat length in bp (Sanger Seq)		Note	FLA in bp		Genomic location (Chiu et al., 2013)
					Allele 1	Allele 2	
DS06	yes	10			151	151	scaffold1: 11,483,889..11,484,042
DS07	yes	22	24		188	190	scaffold1: 3,237,190..22,186,765
DS10	yes	18	24		288	294	scaffold200: 81,328..142,031
DS11	yes	14			244	244	scaffold3366: 2,800..3,063
DS14	yes	18	20		197	199	scaffold6: 146,952..4,165,965
DS17	yes	20		Null allele	x	x	scaffold6: 3,127,060..3,127,160
DS20	yes	8		2nd repeat (CA)	222	222	scaffold3: 2,408,346..9,382,492
DS21	yes	14	18	2nd repeat (TG)	326	330	scaffold14: 284,142..810,706
DS22	yes	18			326	326	scaffold16: 113,346..113,499
DS23	yes	16			257	257	scaffold27: 697,440..731,721
DS26	yes	8		Bad quality FLA	x	x	scaffold400: 83,177..83,267
DS27	yes	14	20		88	94	scaffold400: 83,177..83,267
DS28	yes	20	24		156	160	scaffold443: 36,041..51,998
DS32	yes	12		Bad quality FLA	343	343	scaffold72: 219,191..229,192
DS34	yes	18			252	252	scaffold93: 172,600..204,768
DS35	yes	18			222	222	scaffold356: 99,650..99,866
DS36	yes	18	20		181	183	scaffold200: 81,328..142,031

4.3. Null Alleles, Polymorphisms and Genetic Variability at 14 Microsatellite Loci

Fifteen microsatellite markers were tested for null alleles. Only one locus (DS17) was suggested to have a null allele frequency higher than 10% in almost all populations (Figure 11). This 10% value is described as a threshold from which null allele frequencies are problematic when testing for selection (Fraimout et al., 2015). Loci with higher frequency values than 10% often show an excess of homozygotes, leading to an overestimation of inbreeding. Another locus with null allele frequencies higher than 0.1 in three out of 36 tested populations was DS21 (Figure 11). Also, the number of values close to 0.1 was higher than in the other tested loci for DS21. For the remaining 13 loci, null alleles were suggested in two populations, HB17 and LS_USA (Figure 11).
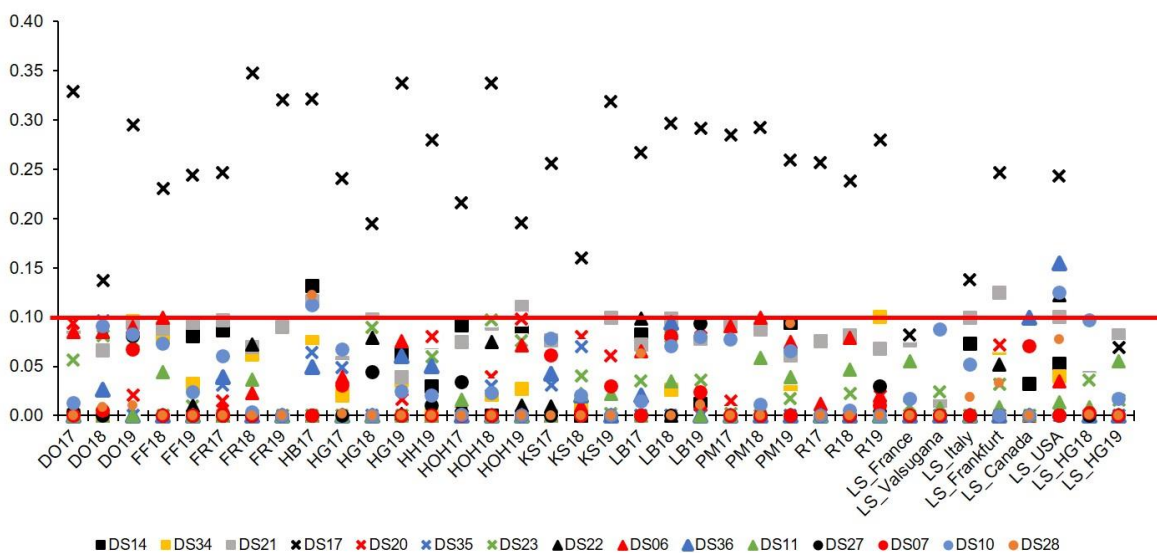
Figure 11: Estimate of null allele frequencies. The null allele frequency for each locus in every population and year is shown. The different loci are color- and form-coded. The red line marks the 10% null allele frequency mark.

First, the genetic variability and the information content of the microsatellite markers were evaluated. This allows an estimation of how well a marker will perform in a population genetic study and if it should be included in an experiment. The results for variability indices in the 14 SSR marker are shown in Table 13. The table contains the number of different alleles (N_a), the number of effective alleles (N_e = number of equally frequent alleles it would take to achieve a given level of gene diversity), observed heterozygosity (H_o = actual observed proportion of heterozygotes in a locus), expected heterozygosity (H_e = proportion of genotypes expected to be heterozygous under HWE), the mean inbreeding coefficient (F_{IS} = genetic differences of the subpopulation contained in an individual), and information on the polymorphism information content (PIC). PIC was obtained as an index for gene abundance. The level of diversity reflects genetic variation in loci (PIC > 0.5 = high

polymorphism, $0.5 > PIC > 0.25$ = moderate polymorphism, $PIC < 0.25$ = low polymorphism). PIC across loci ranged from 0.474 (DS11) to 0.836 (DS07). All loci except for DS11 showed high polymorphism (Table 13), confirming that this set of markers is suitable for population genetic studies. Locus DS11 was the least variable microsatellite marker with the lowest values for all tested characteristics, except for observed heterozygosity and the fixation index. Negative F values were detected in five loci, indicating heterozygote excess (outbreeding). Nine of the 14 loci had a positive F_{IS} value, indicating heterozygote deficiency (inbreeding) compared to expectations under HWE. The differences between observed and expected heterozygosity were negligible in all loci except for locus DS21. H_e was expected to be as high as 0.66, but H_o was only 0.21, suggesting that more homozygotes were present in DS21 than expected under HWE. This fits the result from FreeNa (Figure 11), since DS21 was the only locus that showed higher null allele frequencies compared to other loci. Deviation from HWE was tested for all years in each locus. A significant difference ($p > 0.05$) from HWE was observed in 18 of 42 year-locus combinations (Table 14).

Under the assumption that populations are in Hardy-Weinberg equilibrium, significant LD was found for all pairs of loci in at least one German population or one laboratory strain (A.1.3 Results for Linkage Disequilibrium for Each Population). All populations showed significant linkage amongst several loci. The European laboratory strains showed the most linked loci per locus. In LS_Frankfurt and LS_Italy, the number of linked loci were six in DS36 and DS23, respectively. In LS_France, six loci were linked to DS36, DS07, and DS28 (A.1.3 Results for Linkage Disequilibrium for Each Population).

Table 13: Characteristics of 14 microsatellite markers analyzed in this study. Mean and SE for all populations and years for each locus. PIC = Polymorphism Information Content; N_a = No. of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity; F_{IS} = mean inbreeding coefficient.

		DS14	DS34	DS21	DS20	DS35	DS23	DS22	DS06	DS36	DS11	DS27	DS07	DS10	DS28
PIC		0.689	0.621	0.714	0.750	0.718	0.818	0.666	0.664	0.608	0.474	0.722	0.836	0.767	0.768
N_a	Mean	6.31	4.33	4.97	5.72	5.81	6.97	4.67	4.81	4.78	3.53	6.28	7.03	6.72	6.08
	SE	0.29	0.2	0.19	0.27	0.25	0.3	0.19	0.19	0.21	0.14	0.23	0.27	0.26	0.27
N_e	Mean	3.32	2.62	3.24	3.63	3.46	4.13	3.07	3.19	2.52	1.95	3.63	4.89	4.02	4.09
	SE	0.16	0.13	0.14	0.21	0.13	0.22	0.14	0.17	0.11	0.06	0.14	0.22	0.17	0.21
H_o	Mean	0.65	0.55	0.21	0.64	0.78	0.64	0.65	0.54	0.6	0.47	0.75	0.87	0.62	0.72
	SE	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.03
H_e	Mean	0.67	0.57	0.66	0.67	0.68	0.72	0.64	0.64	0.58	0.47	0.7	0.77	0.73	0.72
	SE	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.01	0.02	0.02
F_{IS}	Mean	0.03	0.04	0.67	0.04	-0.13	0.12	-0.02	0.13	-0.04	0	-0.06	-0.12	0.15	0
	SE	0.03	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.02	0.03	0.03	0.03

Table 14: Deviation from Hardy-Weinberg equilibrium in 42 year-locus combinations. Chi-Square test of HWE to test if H_0 is consistent with the expectations under HWE. Deviations from HWE over all populations in each year for each locus. d.f.= degree of freedom; ChiSq = Chi-Square value; ns = not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Year	Locus	d.f.	ChiSq	Probability	Significance
2017	DS14	36	54.765	0.023	*
	DS34	10	11.001	0.357	ns
	DS21	21	532.075	0.000	***
	DS20	21	46.905	0.001	***
	DS35	28	59.694	0.000	***
	DS23	45	93.524	0.000	***
	DS22	28	191.725	0.000	***
	DS06	21	190.454	0.000	***
	DS36	28	31.412	0.299	ns
	DS11	15	8.151	0.918	ns
	DS27	36	24.837	0.920	ns
	DS07	66	65.852	0.482	ns
	DS10	55	71.142	0.070	ns
	DS28	28	30.047	0.361	ns
2018	DS14	36	24.794	0.921	ns
	DS34	15	43.173	0.000	***
	DS21	36	710.716	0.000	***
	DS20	21	45.550	0.001	**
	DS35	28	32.138	0.269	ns
	DS23	45	113.234	0.000	***
	DS22	15	22.823	0.088	ns
	DS06	21	39.081	0.010	**
	DS36	15	15.562	0.412	ns
	DS11	10	19.232	0.037	*
	DS27	28	41.972	0.044	*
	DS07	45	62.302	0.045	*
	DS10	36	70.386	0.001	***
	DS28	28	24.121	0.675	ns
2019	DS14	36	72.714	0.000	***
	DS34	15	41.022	0.000	***
	DS21	21	639.496	0.000	***
	DS20	21	51.809	0.000	***
	DS35	28	43.199	0.033	*
	DS23	55	269.671	0.000	***
	DS22	15	19.221	0.204	ns
	DS06	21	119.937	0.000	***
	DS36	36	44.475	0.157	ns
	DS11	10	7.712	0.657	ns
	DS27	28	23.988	0.682	ns
	DS07	66	74.566	0.220	ns
	DS10	78	136.419	0.000	***
	DS28	28	34.582	0.182	ns

4.4. Laboratory Strains

4.4.1. Genetic and Allelic Diversity among Laboratory Strains

Estimates for genetic and allelic diversity for the eight laboratory strains were calculated separately from the German populations because laboratory strains are generally kept in an artificial environment leading to inbred populations and artificial selection that could bias the results for the German populations. Nevertheless, the use of laboratory strains allows to test the experimental design upfront and those strains can be used as outgroups. Testing the workflow and the microsatellite markers constituted an essential part of this study. Samples from the laboratory strains were available in high numbers, while genomic DNA from the German samples was restricted. Laboratory strains, therefore, added the opportunity to test optimal PCR conditions, PCR purification protocols, and the FLA workflow. It also allowed the pre-sequencing of the microsatellite markers important for the analysis. This was a prerequisite to conduct because the sequence information was not given in the original publication by Fraimout (2015). In addition, the results for the laboratory strains were more predictable than the German samples and allowed the evaluation of the discriminative power of the microsatellite marker system in use. Moreover, a new laboratory strain (LS_HG) was established. It was interesting to study its development as a new laboratory population compared to the German field collections and the established laboratory strains.

Genetic and allelic diversity plays a vital role in the invasion success of a species, and the assessment of these traits helps to understand the genetic relationship among populations. N_a and N_e describe the diversity in populations at the allele level. N_a is the average number of alleles per locus (sum of all detected alleles in all loci, divided by the total number of loci). N_e is the number of equally frequent alleles it would take to achieve a given level of gene diversity (H_e). H_o and H_e describe the diversity on a genotype level. While H_o is the observed proportion of heterozygotes in a population, H_e is the proportion of genotypes expected to be heterozygous under HWE. Low heterozygosity indicates effects of small population sizes like bottleneck events with low genetic variability, and high heterozygosity shows high genetic variability. If the observed heterozygosity is lower than expected ($H_o < H_e$), it can be assumed that inbreeding takes place. If H_o is higher than H_e, it indicates the mixing of different populations. On an individual level, H_e can be interpreted as the expected probability that an individual is heterozygous (at a given locus). Based on the experimental parameters, the most diverse laboratory strain was LS_HG18, and the least diverse strain was LS_Canada. These results correlate well with the time the strains were kept in the laboratory, with LS_HG18 being the newest (one year in culture) and LS_Canada one of the oldest strains (approx. eight years in culture). The Fixation index (F_{IS}) of an individual (I) relative to the subpopulation (S), is the average coefficient of inbreeding in a population and ranges from -1 to +1. The F_{IS} value was positive for most

48

strains except for LS_France and LS_Valsugana, indicating inbreeding (Figure 12, Table 16 for more details). Values close to zero are expected under random mating. Positive values indicate inbreeding or null alleles, and negative values indicate an excess of heterozygotes due to assortative mating or selection for heterozygotes.
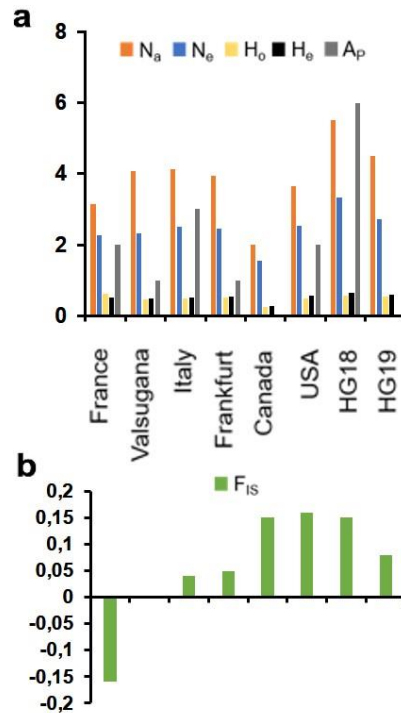
Figure 12: Level of genetic diversity across studied laboratory strains. a) N_a = Number of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity, A_P = Number of private alleles unique to a single population. b) F_{IS} = mean inbreeding coefficient.

The number of private alleles in the tested laboratory strains was relatively low, suggesting that most alleles are shared between strains. The laboratory strains LS_Canada and LS_HG19 were the only two laboratory strains without any private alleles (Table 15). The new laboratory strain LS_HG18 had the most private alleles compared to the other strains with six alleles in total (Figure 12, Table 15).

Table 15: Private allele list for laboratory strains. Given is the name of the laboratory strain, the locus that contains the private allele, the allele length in bp, and its frequency in the population.

Population	Locus	Allele	Frequency
LS_France	DS21	316	0.050
LS_France	DS23	249	0.150
LS_Valsugana	DS28	158	0.050
LS_Italy	DS10	282	0.050
LS_Italy	DS10	284	0,050
LS_Italy	DS27	96	0.025
LS_Frankfurt	DS23	273	0.075
LS_USA	DS06	145	0.025
LS_USA	DS21	332	0.025
LS_HG18	DS14	211	0.150
LS_HG18	DS10	302	0.050
LS_HG18	DS20	226	0.075
LS_HG18	DS23	259	0.075
LS_HG18	DS34	254	0.225
LS_HG18	DS35	226	0.025

Table 16: Genetic and allelic characteristics of eight laboratory strains. Shown is the allelic diversity of eight laboratory strains based on 14 microsatellite markers analyzed in this study. Mean and SE over all laboratory strains and loci. N = sample size; N_a = Number of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity F = mean inbreeding coefficient.

Population		N	N_a	N_e	H_o	H_e	F
LS_France	Mean	20	3.14	2.26	0.61	0.52	-0.16
	SE		0.33	0.2	0.06	0.04	0.04
LS_Valsugana	Mean	20	4.07	2.31	0.46	0.48	0
	SE		0.45	0.3	0.07	0.06	0.08
LS_Italy	Mean	20	4.14	2.51	0.48	0.51	0.04
	SE		0.48	0.31	0.07	0.07	0.06
LS_Frankfurt	Mean	20	3.93	2.45	0.52	0.53	0.05
	SE		0.35	0.26	0.07	0.05	0.08
LS_Canada	Mean	20	2	1.56	0.25	0.28	0.15
	SE		0.23	0.15	0.06	0.06	0.11
LS_USA	Mean	20	3.64	2.54	0.49	0.58	0.16
	SE		0.25	0.17	0.05	0.03	0.09
LS_HG_2018	Mean	20	5.5	3.34	0.56	0.66	0.15
	SE		0.34	0.33	0.06	0.04	0.08
LS_HG_2019	Mean	20	4.5	2.73	0.53	0.59	0.08
	SE		0.39	0.24	0.07	0.04	0.1

4.4.2. Genetic Distance and Relationship among Laboratory Strains

AMOVA – Genetic Variation within and among populations

Population differentiation can be detected by performing an Analysis of Molecular Variance (AMOVA), estimating population differentiation directly from molecular data and not from allele frequencies. AMOVA estimates how much of the differentiation between samples is due to differences between individuals, individuals within a population, or populations. AMOVA in laboratory strains indicated that most variation (72%) occurred within populations, while 28% of the variation was detected among populations (Table 17).

Table 17: Summary of AMOVA result for laboratory populations. d.f.= degree of freedom; SS = sum of squares; VC = variance components; % PV = percentage of total variation.

Source of variation	d.f.	SS	VC	% PV
Among populations	7	429.013	1.439	28%
Within populations	312	1160.950	3.721	72%
Total	319	1589.963	5.160	100%

Pairwise F_{ST} – patterns of differentiation among pairs of populations

The F_{ST} value was used to measure genetic differentiation between populations and is the inbreeding coefficient within subpopulations (S) relative to the total (T). It is often referred to as a fixation index but is a genetic distance value and should not be confused with the F_{IS} value. It compares the proportion of the genetic variation within a population relative to the total genetic variance and is derived from the variances of allele frequencies. Pairwise F_{ST} is calculated similarly to the 'normal' F_{ST} used in Wright's F-Statistics. Instead of a single and average statistic over loci and populations, pairwise F_{ST} provides insights into genetic relationships between populations by revealing possible differentiation patterns among pairs of populations. It does measure the heterozygote deficiency and can estimate how genetically distant populations are compared to each other. F_{ST} values for laboratory strains reveal a minor differentiation ($F_{ST} < 0,1$) only between LS_HG19 and LS_HG18 and between LS_HG19 and LS_France, which are 7.14% of the pairwise values. A moderate differentiation (0.1 – 0.25) was detected for 75% of pairwise F_{ST} values and 17.86% of the F_{ST} values showed a strong differentiation ($F_{ST} > 0,25$) (Figure 13). The pairwise F_{ST} for laboratory strain LS_HG revealed more differentiation in the second year of inbreeding (2019) than in 2018. While the differentiation was moderate in 2018, it was moderate to strong in 2019 (Figure 13).
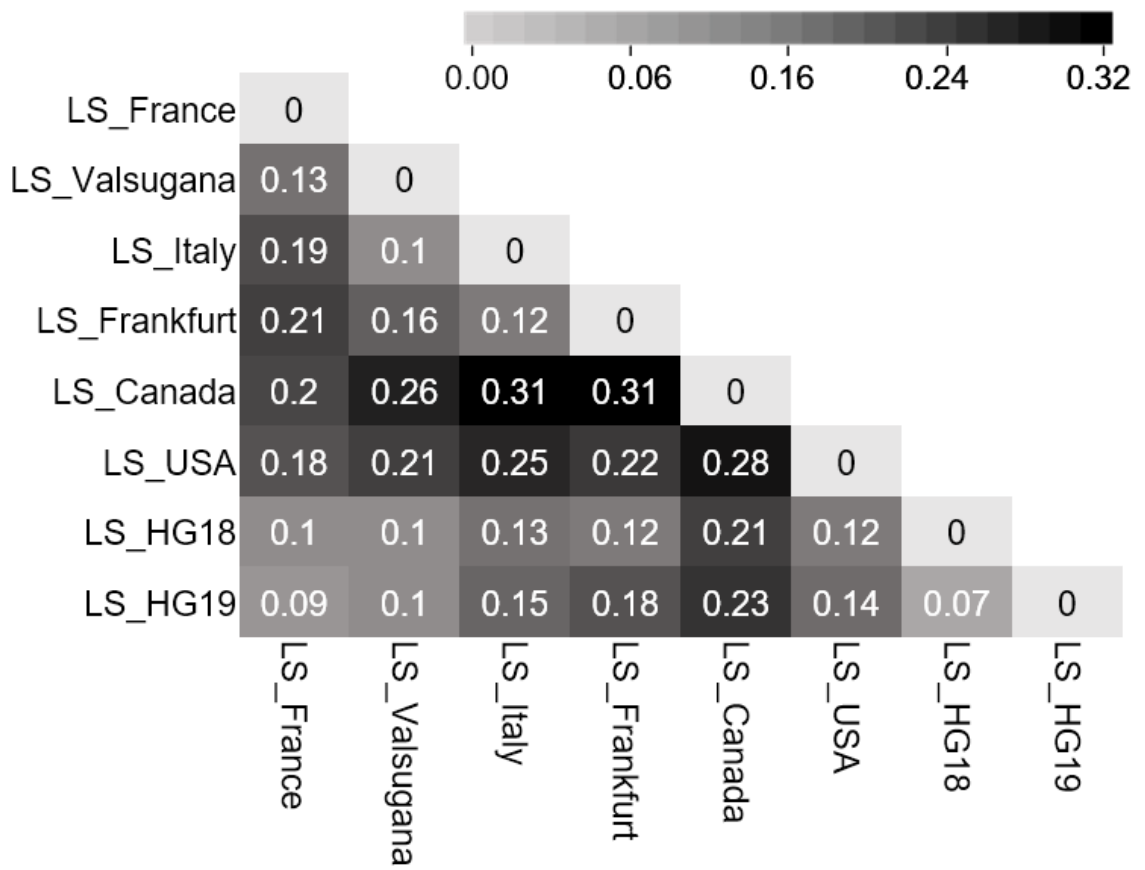
Figure 13: Pairwise F_{ST} among laboratory strains. F$_{ST}$ values for eight laboratory strains are shown in pairwise comparisons. Values are color coded (light grey = F $_{ST}$ < 0.1: minor differentiation; grey = 0.1 < F$_{ST}$ < 0.25: moderate differentiation; black = F $_{ST}$ > 0.25: strong differentiation).

Nei's Genetic Distance/Identity and Principle Coordinate Analysis (PCoA) – the genetic relationship between populations

Genetic distance is a measure for dissimilarities or not relatedness between genotypes. Nei's Genetic Distance is a dissimilarity matrix that shows the distance between a pair of objects (in this case, populations) and ranges from 0 to infinity (Nei, 1972). A genetic identity measure does the exact opposite by showing the similarities between populations and ranges from 0 (no similarities detected) to 1 (identical) (Nei, 1972). However, both matrices will lead to the same interpretation by explaining the genetic relationship between a pair of populations. A pairwise unbiased Nei's genetic identity calculation showed that LS_Italy and LS_Valsugana, and LS_HG18 and LS_HG19 share the most genetic material with 77% and 75%, respectively. The least identical laboratory strains are LS_USA and LS_Italy (26%). These findings agree with the pairwise F_{ST} results. Also, the majority (67%) of pairwise comparisons reveal a moderate identity with values over 50% (Figure 14).
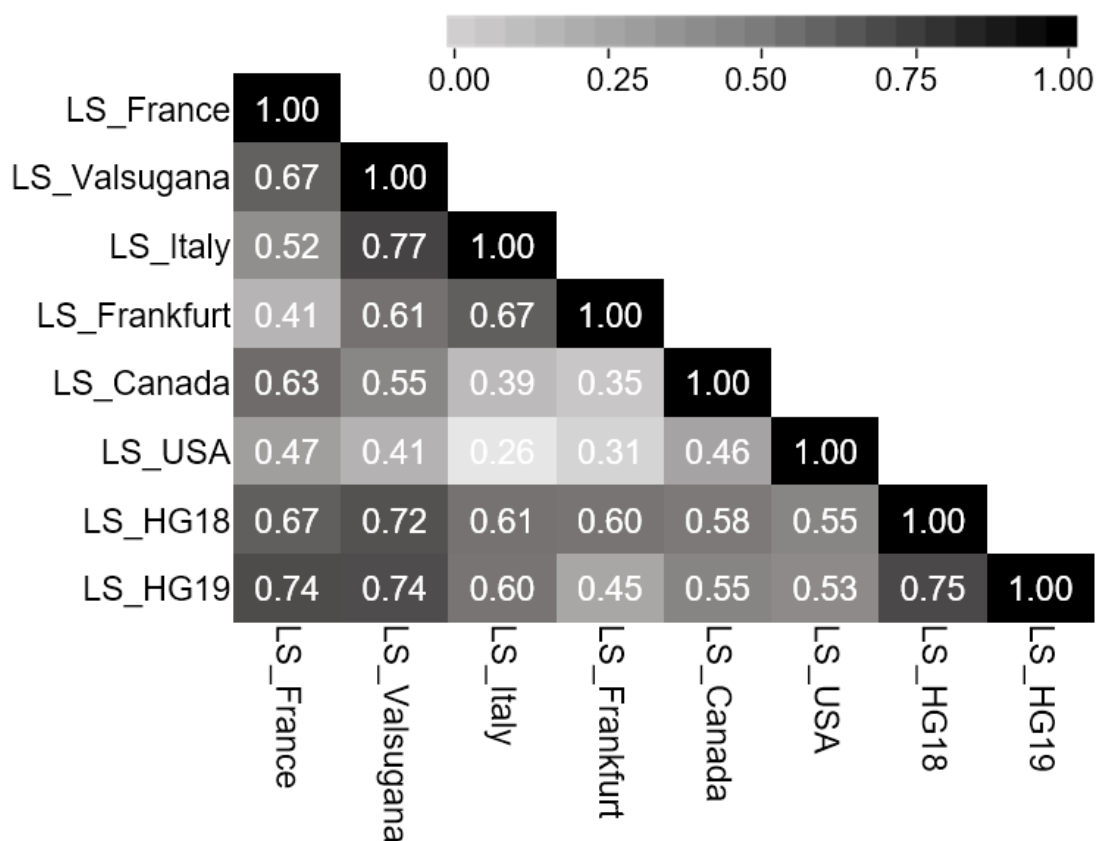
Figure 14: Pairwise population matrix of Nei's Genetic Identity among laboratory strains. Genetic identity values for eight laboratory strains are shown in pairwise comparisons below diagonal. Values are color coded (light grey = Identity < 0.3: low identity; grey = 0.3 < Identity < 0.7: moderate Identity; black = Identity > 0.7: high identity).

Another way to detect possible genetic connections between populations is to perform a Principal Coordinate Analysis (PCoA) (Figure 15). It uses Nei's Genetic Distance as a distance matrix and identifies a set of axes representing the variability in the data set. Each axis has a value, which indicates the amount of variation captured in that axis. Samples are represented as points in this graph, and the distance between the points is related to the (dis)similarities of the samples. Points (here populations) closer to each other are more similar than those that are further away. PCoA was performed to evaluate the genetic relationship between populations based on a distance measure approach. Overall, the separation between laboratory strains is distinct. The first axis accounts for 42.03% of the detected variation, and the second axis accounts for 21.21% of the variation, which captures a robust amount of the present variation. LS_USA and LS_Frankfurt show the most significant allocation from the other populations (Figure 15), fitting Nei's Genetic Identity measure.
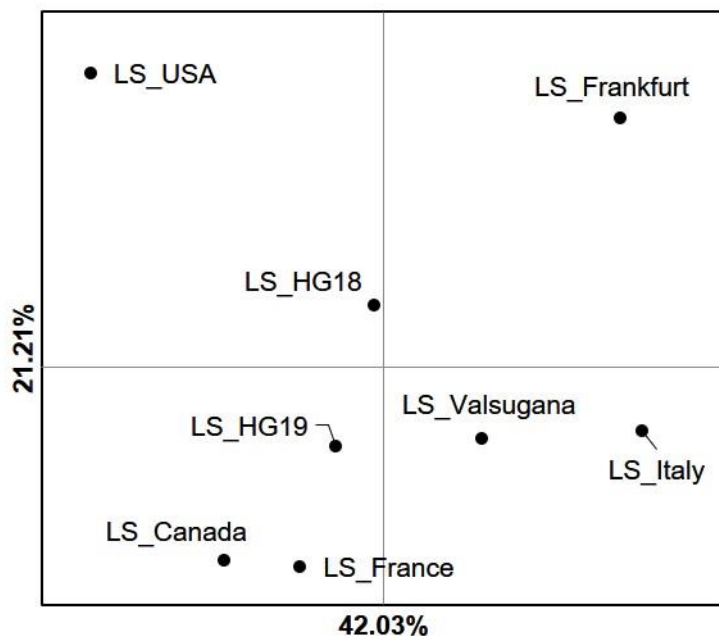
Figure 15: PCoA at population level for laboratory strains generated from 14 microsatellite markers. PCoA for samples of all eight laboratory strains was computed to visualize genetic dissimilarities in the laboratory dataset. The axes account for a portion of the detected variation.

Genetic population structure with an NJ population tree

A neighbor-joining tree based on Nei's genetic distance was constructed to visualize the relationship between the laboratory strains. The genetic distance between the laboratory populations is represented as the total branch length and the bootstrap values support this tree. It corresponds to findings from Nei's Genetic Identity, PCoA, and F_{ST} (Figure 16). Overall, the NJ tree showed that LS_USA and LS_Canada are the most distinct laboratory strains with the highest distance to the European laboratory strains LS_Italy and LS_Frankfurt. LS_France and LS_Valsugana grouped as well as LS_HG18 and LS_HG19. Generally speaking, the NJ tree captured the strains' relationship as follows: it clustered the North American, the older European (LS_Italy and LS_Frankfurt), the newer European (LS_Valsugana and LS_France), and the novel laboratory strains from LS_HG.
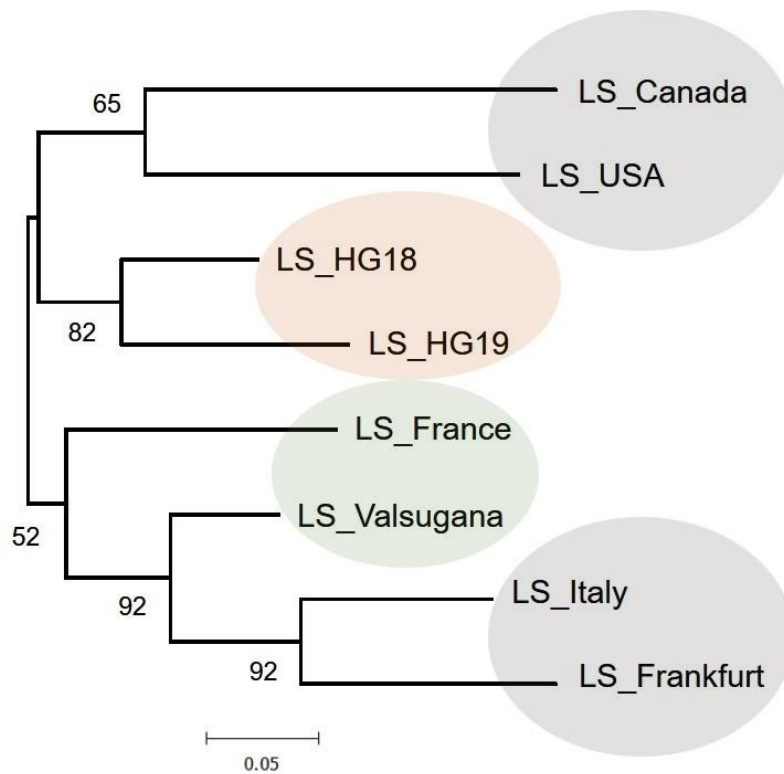
Figure 16: Unrooted neighbor-joining tree based on Nei's genetic distance for laboratory strains. Allelic frequencies were obtained with 14 microsatellite markers for the eight *D. suzukii* laboratory populations. Genetic distance between populations is represented as the length of the lines. Bootstrap values are given on the nodes. Circles represent population origin with grey = long-established laboratory strains (2010-2016), green = laboratory strains from France (LS_France, 2018) and Italy (LS_Valsugana, 2018), and orange = laboratory strain from Bad Homburg (2017).

4.5. German Populations

4.5.1. Genetic and Allelic Diversity among Populations and Years

Statistics based on gene and allele frequency data were performed to gain information on genetic variation in the German populations. Diversity measures can provide information about population size fluctuations or the amount of gene flow between populations (Slatkin, 1985). Diversity plays a substantial role in the fitness and adaptive capacity of a population (Markert et al., 2010). It plays an essential role in enabling a population to tolerate and survive different biotic and abiotic stress factors. It can even ensure a population's capability to evolve under changing environmental conditions (Stojnić et al., 2019).

As with the laboratory strains, the number of private alleles and their frequency was estimated. The number of private alleles (A_P) was low throughout the study period, suggesting that most alleles are shared between populations and years (Figure 17, Table 18). Only seven out of 28 populations showed private alleles, six with a frequency of 0.025 and one (FF19) with a frequency of 0.05. FR17 was the only population with two private alleles. The rest had only one private allele (Figure 17, Table 18).

Table 18: Private Allele list for German populations. Given is the name of the population, the locus that contains the private allele, the allele as the length in bp, and its frequency in the population.

Population	Locus	Allele	Frequency
FF19	Locus DS23	251	0.05
FR17	Locus DS06	161	0.025
FR17	Locus DS07	288	0.025
HG17	Locus DS27	82	0.025
HOH17	Locus DS07	198	0.025
LB17	Locus DS22	330	0.025
PM18	Locus DS07	194	0.025

Genetic and allelic diversity in German populations was relatively high in all years and over all locations. There was no significant difference between years or sample sites (Figure 17, Table 19). The mean number of observed alleles (N_a) over the loci for each year reached from 6.23 to 5.93 and was overall similar. The mean number of effective alleles (N_e, the number of equally frequent alleles it would take to achieve a given level of gene diversity) ranged between 3.19 (R17) and 4.05 (KS19) across all tested locations and years, supporting our findings that there is no significant difference in allelic diversity between years or sample sites. Observed heterozygosity (H_o, the observed ratio of heterozygotes) was again similar between locations and sampling years. The highest value was measured for KS19 and HG19 with 0.71 and the lowest value for LB in 2019 with 0.58. Values for expected heterozygosity (H_e, the proportion of heterozygous genotypes expected under

Hardy-Weinberg equilibrium) were similar in all three years, with R17 and PM18 showing the lowest value with 0.65, and KS having the highest value in 2019 with 0.74. H_O was higher than H_e in four populations (FF19, PM17, R17, and R19), indicating heterozygote excess. It coincided in one case (R18) and was lower than expected in all remaining populations, indicating a heterozygote deficiency. The F_{IS} value represents the average deviation from HWE across all loci in a population. It is the average coefficient of inbreeding in a population and ranges from -1 to +1. F_{IS} for German populations was positive in 25 populations and negative in three populations (R17, R18, and FF19). Values close to zero are expected under random mating. Positive values indicate inbreeding or null alleles, and negative values indicate an excess of heterozygotes due to assortative mating or selection for heterozygotes. In 2018 all populations had a positive F_{IS} value (Figure 17, Table 19). Positive values can indicate inbreeding and heterozygote deficiency or, to be precise, an excess of homozygotes due to the presence of null alleles, inbreeding, or population subdivision.
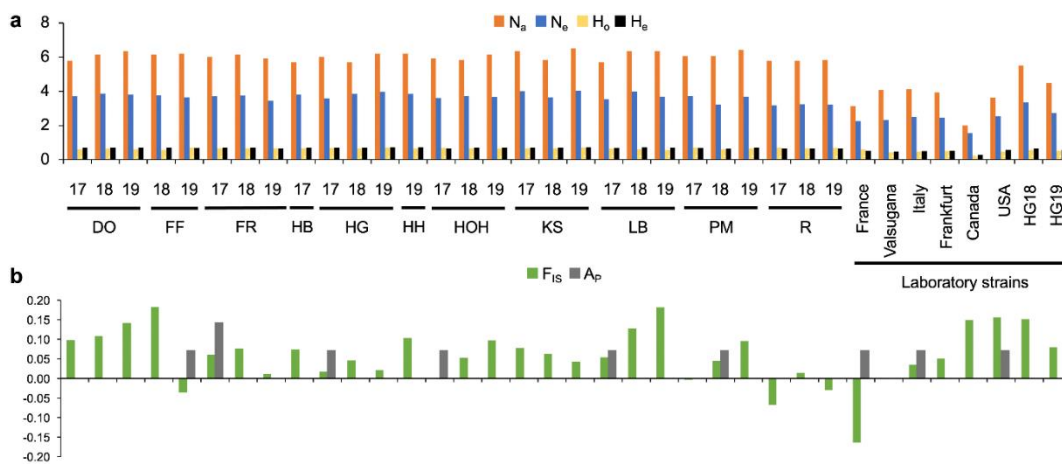
Figure 17: Level of genetic diversity across studied populations over three years. a) N_a = No. of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity. b) A_P = No. of private alleles unique to a single population, F_{IS} = mean inbreeding coefficient.

In comparison to the laboratory strains, it occurs that the overall genetic and allelic diversity in the tested laboratory strains was lower than in the field collection from Germany with 42% fewer different alleles, 38% less effective alleles, 29% less observed heterozygosity, and 31% less expected heterozygosity (Figure 17) making the laboratory strains less diverse than the German field collections.

Table 19: Genetic and allelic diversity measured for German populations over three years. Shown is the allelic diversity of German populations collected between 2017 and 2019 based on 14 microsatellite markers. Mean and SE over all populations and loci. N = sample size; N_a = No. of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity F = mean inbreeding coefficient.

Population	N	N_a	N_e	H_o	H_e	F
DO17	20	5.79 (SE = 0.33)	3.70 (SE = 0.30)	0.62 (SE = 0.05)	0.70 (SE = 0.04)	0.10 (SE = 0.06)
DO18	20	6.14 (SE = 0.40)	3.87 (SE = 0.29)	0.64 (SE = 0.04)	0.72 (SE = 0.02)	0.11 (SE = 0.05)
DO19	20	6.36 (SE = 0.36)	3.81 (SE = 0.27)	0.61 (SE = 0.05)	0.72 (SE = 0.02)	0.14 (SE = 0.07)
FF18	20	6.14 (SE = 0.31)	3.76 (SE = 0.31)	0.59 (SE = 0.06)	0.71 (SE = 0.02)	0.18 (SE = 0.07)
FF19	20	6.21 (SE = 0.46)	3.65 (SE = 0.38)	0.70 (SE = 0.06)	0.68 (SE = 0.04)	-0.04 (SE = 0.06)
FR17	20	6.00 (SE = 0.33)	3.71 (SE = 0.28)	0.67 (SE = 0.06)	0.71 (SE = 0.02)	0.06 (SE = 0.08)
FR18	20	6.14 (SE = 0.38)	3.74 (SE = 0.27)	0.66 (SE = 0.06)	0.71 (SE = 0.03)	0.08 (SE = 0.07)
FR19	20	5.93 (SE = 0.37)	3.46 (SE = 0.27)	0.67 (SE = 0.07)	0.68 (SE = 0.03)	0.01 (SE = 0.08)
HB17	10	5.71 (SE = 0.41)	3.83 (SE = 0.34)	0.66 (SE = 0.07)	0.71 (SE = 0.03)	0.07 (SE = 0.08)
HG17	20	6.00 (SE = 0.38)	3.6 (SE = 0.28)	0.68 (SE = 0.04)	0.70 (SE = 0.02)	0.02 (SE = 0.06)
HG18	20	5.71 (SE = 0.41)	3.87 (SE = 0.38)	0.66 (SE = 0.04)	0.70 (SE = 0.04)	0.05 (SE = 0.05)
HG19	20	6.21 (SE = 0.33)	3.98 (SE = 0.31)	0.71 (SE = 0.06)	0.73 (SE = 0.02)	0.02 (SE = 0.07)
HH19	20	6.21 (SE = 0.41)	3.85 (SE = 0.28)	0.65 (SE = 0.06)	0.72 (SE = 0.02)	0.10 (SE = 0.07)
HOH17	20	5.93 (SE = 0.45)	3.61 (SE = 0.40)	0.69 (SE = 0.05)	0.68 (SE = 0.03)	0 (SE = 0.06)
HOH18	20	5.86 (SE = 0.36)	3.73 (SE = 0.27)	0.68 (SE = 0.05)	0.71 (SE = 0.02)	0.05 (SE = 0.07)
HOH19	20	6.14 (SE = 0.48)	3.67 (SE = 0.25)	0.64 (SE = 0.06)	0.71 (SE = 0.02)	0.10 (SE = 0.08)
KS17	20	6.36 (SE = 0.41)	4.01 (SE = 0.33)	0.66 (SE = 0.04)	0.72 (SE = 0.03)	0.08 (SE = 0.05)
KS18	20	5.86 (SE = 0.25)	3.65 (SE = 0.26)	0.66 (SE = 0.05)	0.71 (SE = 0.02)	0.06 (SE = 0.07)
KS19	20	6.50 (SE = 0.36)	4.05 (SE = 0.26)	0.71 (SE = 0.05)	0.74 (SE = 0.02)	0.04 (SE = 0.06)
LB17	20	5.71 (SE = 0.30)	3.55 (SE = 0.28)	0.65 (SE = 0.05)	0.69 (SE = 0.03)	0.05 (SE = 0.06)
LB18	20	6.36 (SE = 0.44)	3.99 (SE = 0.31)	0.63 (SE = 0.06)	0.73 (SE = 0.02)	0.13 (SE = 0.07)
LB19	20	6.36 (SE = 0.45)	3.70 (SE = 0.28)	0.58 (SE = 0.05)	0.70 (SE = 0.03)	0.18 (SE = 0.07)
PM17	20	6.07 (SE = 0.44)	3.72 (SE = 0.34)	0.70 (SE = 0.06)	0.69 (SE = 0.03)	0 (SE = 0.07)

Table 19 (continued from page 59): Genetic and allelic diversity measured for German populations over three years. Shown is the allelic diversity of German populations collected between 2017 and 2019 based on 14 microsatellite markers. Mean and SE over all populations and loci. N = sample size; N_a = No. of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity F = mean inbreeding coefficient.

Population	N	N_a	N_e	H_o	H_e	F
PM18	20	6.07 (SE = 0.46)	3.24 (SE = 0.35)	0.62 (SE = 0.05)	0.65 (SE = 0.03)	0.04 (SE = 0.06)
PM19	20	6.43 (SE = 0.44)	3.69 (SE = 0.7)	0.64 (SE = 0.05)	0.71 (SE = 0.02)	0.10 (SE = 0.07)
R17	20	5.79 (SE = 0.39)	3.19 (SE = 0.29)	0.70 (SE = 0.06)	0.65 (SE = 0.03)	-0.07 (SE = 0.07)
R18	20	5.79 (SE = 0.35)	3.24 (SE = 0.29)	0.66 (SE = 0.06)	0.66 (SE = 0.03)	0.01 (SE = 0.07)
R19	20	5.86 (SE = 0.44)	3.23 (SE = 0.25)	0.68 (SE = 0.06)	0.66 (SE = 0.03)	-0.03 (SE = 0.08)

4.5.2. Genetic Distance and Relationship among Populations and Years

AMOVA – Genetic Variation within and among populations

Analysis of Molecular Variance (AMOVA) and Principle Coordinated Analysis (PCoA) provide information on population distribution, mating behavior, and potential population borders. The analysis of molecular variance for laboratory strains was calculated separately (4.4.2) since they are expected to be highly inbred. AMOVA indicated that 98% of the variation in German populations occurred within populations, while only 2% appeared among populations in 2017 and 2019. In 2018, 3% of variation originated among populations and 97% from within populations (Table 20). This data suggests a substantial gene flow and low intrapopulation differentiation, indicating that populations are connected and not isolated. It is noteworthy that the among-population variation in the tested laboratory strains was significantly higher (28%) (Table 17) than the among-population variation in the German populations (2% and 3%) (Table 20), indicating that the laboratory strains are more structured than the German populations.

Table 20: Summary of AMOVA result for German populations across three years. d.f. = degree of freedom; SS = sum of squares; VC = variance components; % PV = percentage of total variation.

Year	Source of variation	d.f.	SS	VC	% PV
2017	Among populations	8	65.41	0.0869	1.74
	Within populations	331	1623.60	4.9051	98.26
	Total	339	1689.01	4.9920	100
2018	Among populations	8	94.39	0.1711	3.34
	Within populations	351	1738.67	4.9535	96.66
	Total	359	1833.06	5.1246	100
2019	Among populations	9	81.07	0.0989	1.92
	Within populations	390	1970.07	5.0515	98.08
	Total	399	2051.14	5.1504	100

Pairwise F$_{ST}$ – patterns of differentiation among pairs of populations

The pairwise F$_{ST}$ revealed a minor differentiation (F$_{ST}$ < 0.1) between all populations and years. In comparison, values in experiments, including laboratory strains as outgroups ranged from 0.06 to 0.31, with 54.64% of all F$_{ST}$ values showing a moderate (0.1 – 0.25), 43.17% a minor (F$_{ST}$ < 0.1), and 2.19% a strong differentiation (F$_{ST}$ > 0.25). The pairwise population F$_{ST}$ ranged from 0.01 to 0.03 in 2017, from 0.01 to 0.04 in 2018, and again from 0.01 to 0.04 in 2019. The pairwise F$_{ST}$ for laboratory strains revealed more differentiation in the laboratory strains than in the German samples. A special case is LS_HG18, where the differentiation compared to the German populations was moderate. The moderate to strong differentiation in LS_HG19 indicates that the two years of inbreeding resulted in stronger differentiation compared to its origin (Figure 18).
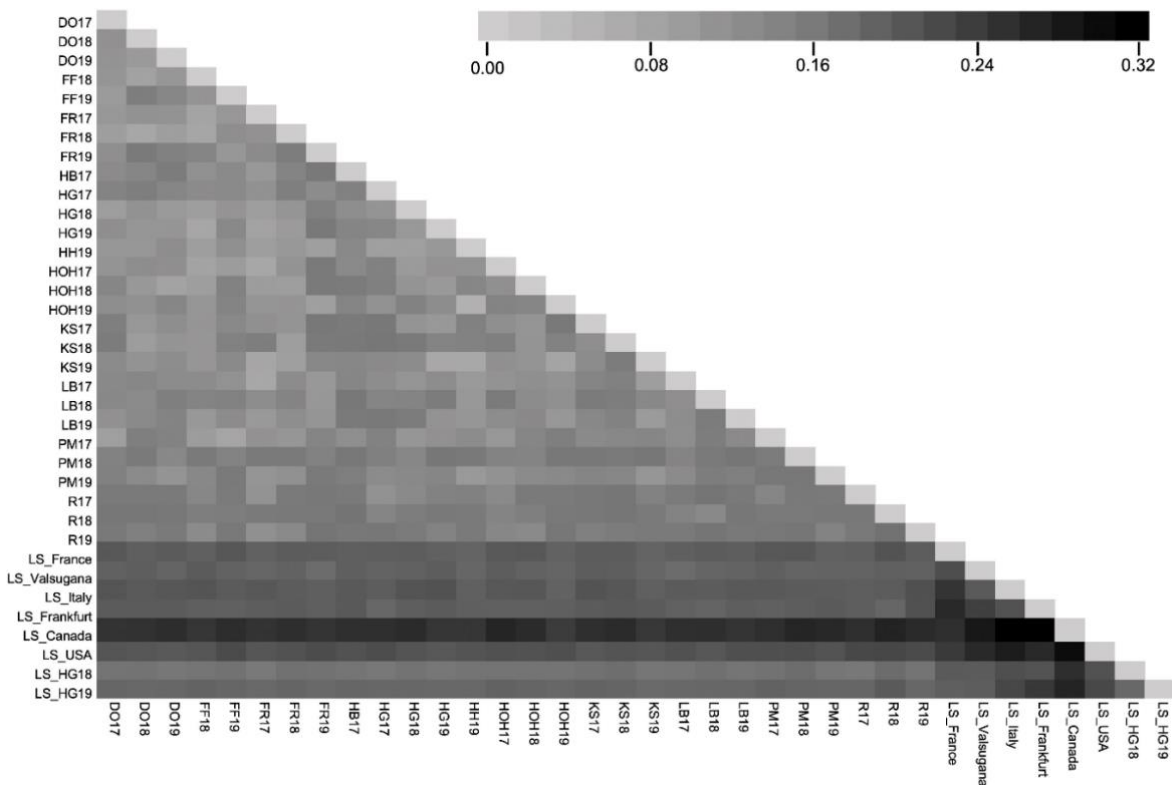
Figure 18: Pairwise F$_{ST}$ among studied populations over three years. Eight laboratory strains (LS_France, LS_Valsugana, LS_Italy, LS_Frankfurt, LS_USA, LS_Canada, LS_HG18, and LS_HG19) and 28 populations from Germany, collected over three years, are shown in pairwise comparisons. Values are color coded (light grey = FST < 0,1: minor differentiation; grey = 0,1 < FST < 0,25: moderate differentiation; black = FST > 0,25: strong differentiation).

Nei's Genetic Distance/Identity and Principle Coordinate Analysis (PCoA) – genetic relationship between populations

A genetic identity measure was conducted to explain the genetic relationship between a pair of populations. Nei's Identity ranges from 0 (no similarities detected) to 1 (identical) (Nei, 1972). A pairwise unbiased Nei's genetic identity calculation showed that most German populations share their genetic material. All German field populations without exceptions shared more than 80% identity, and 37.57% of all populations shared even more than 90%. The laboratory strains were less similar, with most values close to or below 70% identity. Minimal identity was found between the two oldest laboratory strains LS_USA or LS_Canada, compared to the German populations with values around 50%. LS_HG18, on the other hand, still shared most of its identity (~80%) with the German populations and less with the other laboratory strains (60 – 70%). However, the level of identity to the German populations declined in LS_HG19 (~70%), proving that laboratory cultivation over time changes the genetic background of a strain (Figure 19).
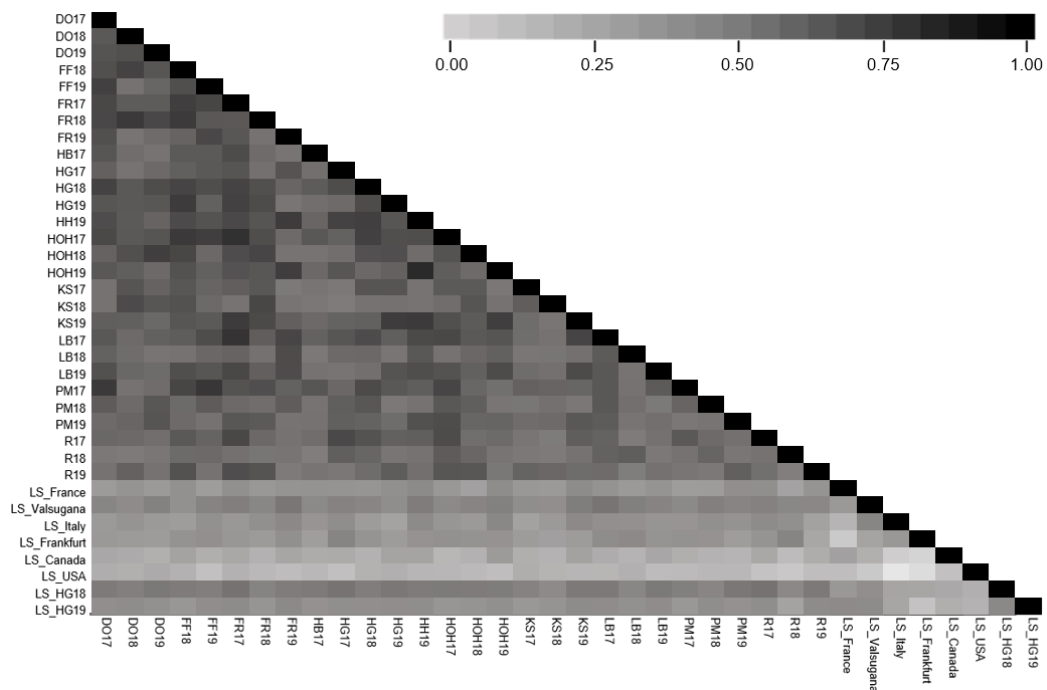
Figure 19: Pairwise population matrix of Nei's Genetic Identity among German populations and laboratory strains. Genetic identity values for German populations over three years and eight laboratory strains are shown in pairwise comparisons below diagonal. Values are color coded (light grey = Identity < 0.3: low identity; grey = 0.3 < Identity < 0.7: moderate Identity; black = Identity > 0.7: high identity).

The PCoA confirmed what the other statistical tests already implied. A high dimensionality was observed for German populations, independent of the sampling year. The first two axes in 2017 represented only 24.07% and 17.97% of the differentiation, with similar values in 2018 and 2019 (Figure 20). For comparison, in the laboratory strains, two axes were sufficient to cover more than 50% of the differentiation (Figure 15). Overall, the separation between German populations was not distinct. The tested laboratory strains in comparison were noticeably separated from the German field collections. The relatively high dimensionality of the data suggests that neither location nor year can accurately differentiate populations, which agrees with the results obtained with AMOVA. Strain LS_HG19 did show more similarity to the laboratory strains than to the field collections, confirming the impact of laboratory cultivation over time. The laboratory strains showed that the used marker system in our experiment could discriminate between populations but that there were no apparent dissimilarities between German sample sites or years.
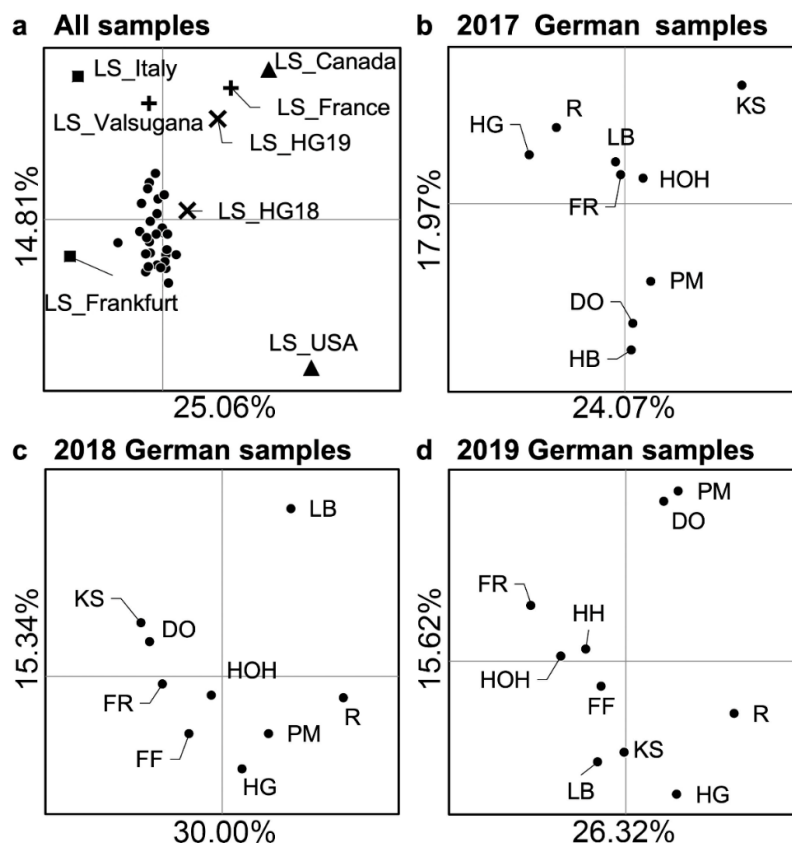
Figure 20: PCoA at population level for 2017-2019 generated from 14 microsatellite markers. **a)** Samples of all populations in Germany from 2017 to 2019 (●) and laboratory outgroups (▲ 8- to 10-year-old laboratory strains, ■ laboratory strains that were established between 2014 and 2016, ✖ laboratory strain established from the wild population HG17, ✚ laboratory strains LS_France and LS_Valsugana which were established in the year 2018) **b)** PCoA for German populations sampled in 2017 **c)** The lower left shows the result for the year 2018 **d)** PCoA for German populations sampled in 2019.

Genetic population structure with STRUCTURE and population tree

A Bayesian model can be used to infer (genetic) population structure, which detects clusters of genetically similar individuals within subpopulations. In a first attempt, we excluded all laboratory strains, but STRUCTURE could not identify population structure in the German dataset. The exclusion of laboratory strains was done to detect any possible population structure solely in the German field populations. Adding laboratory strains LS_USA, LS_Canada, LS_Italy, LS_France, and LS_Valsugana for comparison resulted in two likely lineages (K=2) throughout all three years and populations (Figure 21). There were two values for K detected, which were K=2 and K=3, with the more likely one being K=2, according to the analysis of ΔK. For the analysis including LS_HG18 and LS_HG19, three lineages (K=3) were identified according to the analysis of ΔK. K=4 was also a possible K value but not the most likely one. In both cases, the two most likely K are displayed in Figure 21 to overcome the problem of underestimating the 'true' value of K (Janes et al., 2017). Results indicate that German populations are not significantly different from each other, the genetic structure is not pronounced, gene flow is not restricted or a single genetic cluster explains the distribution of genetic variation in the sampled German populations. Only the additional usage of distinct laboratory strains resulted in the detection of population structure.
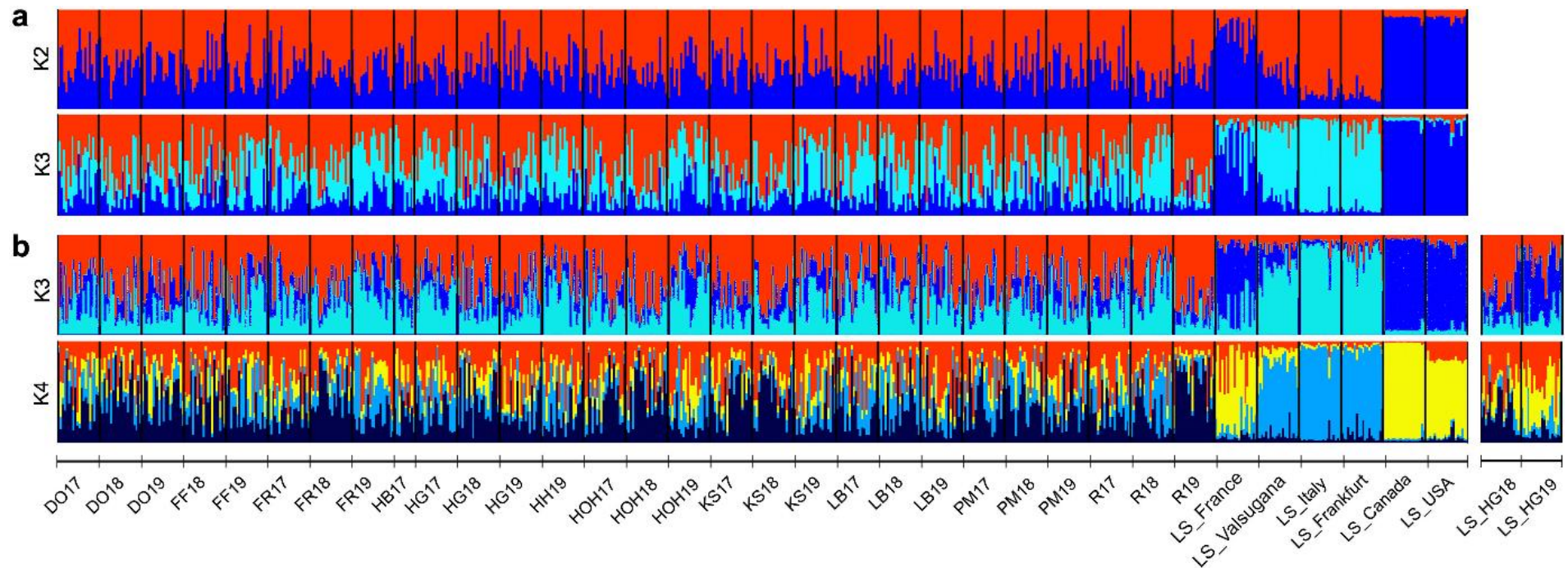
Figure 21: Structural analysis on *D. suzukii* populations in Germany. a) 34 populations with a total of 670 individuals were analyzed: six laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_France, and LS_Valsugana), nine populations from 2017, nine populations from 2018, and ten populations from 2019; the possible number of clusters are shown for K = 2 and K = 3. b) 36 populations with a total of 710 individuals are shown. In addition to the 34 populations in a), laboratory strains LS_HG18 and LS_HG19 were added to the calculation. Shown are the two clusters K = 3 and K = 4.

A neighbor-joining tree based on Nei's genetic distance corresponds to findings in STRUCTURE and F$_{ST}$. While German populations are not grouped or structured, the differences compared to laboratory strains are visible. While LS_HG18 is already separated from the German field collections after one year in culture, the separation gets even more prominent in LS_HG19 after two years in culture (Figure 22).
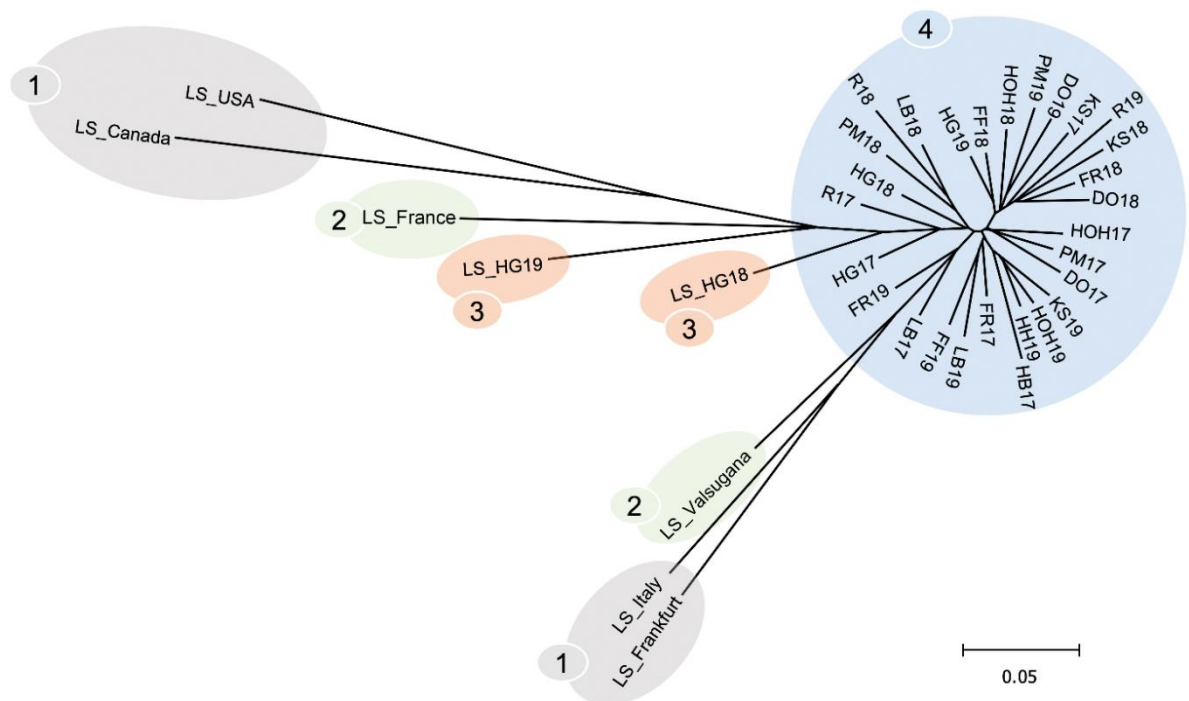
Figure 22: Unrooted neighbor-joining tree based on Nei's genetic distance. Allelic frequencies were obtained with 14 microsatellite markers for the 36 *D. suzukii* populations. Circles represent population origin with 1 = long-established laboratory strains (2010-2016), 2 = laboratory strains from France (LS_France, 2018) and Italy (LS_Valsugana, 2018), 3 = laboratory strain from Bad Homburg (2017) and 4 = German populations collected over three years (2017-2019) in the field

Bottleneck – an estimate of recent demographic events

A BOTTLENECK analysis was used to test the hypothesis of a recent expansion or reduction in each of the 28 populations from Germany (Table 21). The laboratory strains were not included, because they do not contain information about demographic events in the German field populations. Also, it has to be expected that the result would be heavily biased by laboratory rearing conditions. Probability values for the Stepwise Mutation Model (SMM) and the Two-Phase Model (TPM) were calculated. The SMM states that each mutation creates a new allele either by adding or deleting a single repeat unit of the microsatellite with an equal probability in both directions. Alleles of different sizes will be less related than alleles of similar size (Kimura & Otha 1978). The TPM is a spinoff of SMM and accounts for a proportion of more significant mutation events (Valdes et al. 1993 and di Renzo et al. 1994). The Infinite Allele Model (IAM) was not used, because microsatellite loci are generally thought to follow the SMM and thus the TPM.

Table 21: Wilcoxon test for recent demographic events in BOTTLENECK. Reported are p values of Wilcoxon sign-ranked tests for each German population. Given is the probability for heterozygote deficiency and excess for the TPM and SMM. Boldface indicates significant p values < 0.05. TPM = two-phase mutational model, SMM = stepwise mutation model.

	TPM		SMM	
	Heterozygote deficiency	Heterozygote excess	Heterozygote deficiency	Heterozygote excess
DO17	0.879	0.134	0.596	0.428
DO18	0.961	**0.045**	0.313	0.708
DO19	0.665	0.357	0.163	0.852
FF18	0.665	0.357	0.059	0.948
FF19	0.428	0.596	**0.021**	0.982
FR17	0.923	0.086	0.148	0.866
FR18	0.821	0.195	0.097	0.913
FR19	0.358	0.665	0.076	0.932
HB17	0.548	0.476	0.195	0.821
HG17	0.940	0.067	0.291	0.729
HG18	0.961	**0.045**	0.852	0.163
HG19	0.892	0.121	0.357	0.665
HH19	0.821	0.195	0.163	0.852
HOH17	0.619	0.404	0.134	0.879
HOH18	0.975	**0.029**	0.313	0.708
HOH19	0.642	0.380	0.163	0.852
KS17	0.852	0.163	0.357	0.665
KS18	0.821	0.195	0.357	0.665
KS19	0.804	0.213	0.271	0.749
LB17	0.955	0.052	0.271	0.743
LB18	0.955	0.052	0.452	0.572
LB19	0.335	0.687	0.108	0.903
PM17	0.500	0.524	0.134	0.879
PM18	0.148	0.866	**0.010**	0.992
PM19	0.380	0.642	0.068	0.940
R17	0.059	0.948	**0.002**	0.998
R18	0.148	0.886	**0.002**	0.999
R19	0.163	0.852	**0.021**	0.982

For the Wilcoxon test, and under the assumption of the SMM, no bottleneck event was detected. In contrast, significant heterozygote deficit was identified in FF19 (p = 0.021), PM18 (p = 0.010), R17 (p = 0.002), R18 (p = 0.002) and R19 (p = 0.021), suggesting that a recent expansion took place. However, under the TPM model, none of the above-mentioned expansion signals could be confirmed, instead, a bottleneck event was more likely in DO18 (p = 0.045), HOH18 (p = 0.029) and HG18 (p = 0.045) (Table 21).

Migration rate – identification of gene flow between populations

Migration rate values were calculated with GENECLASS 2.0 and showed that migration occurred in all populations over three years except for R18 and R19 (Table 22, Table 23, Table 24). Results indicate that less migration took place from Regensburg to the other population in both 2018 and 2019. In contrast, in 2017, migration was detected for seven out of eight populations with m > 0.1. Although migration from R to the other populations decreased over time, migrant flow occurred from all other populations to R. In 2017, the migrant flow was identified from KS to all other populations while only a low migration rate was detected from the other populations to KS. Overall migration occurred at a lower rate in 2018 than in 2017 or 2019.

Table 22: Mean assignment rates of individuals from 2017 estimated by using GENECLASS 2.0. Assignment rates of individuals into (rows) and from (columns) each population in 2017. Boldface indicates migration rate values (m) above 0.1. Grey indicates migration rate values (m) below 0.1.

2017	DO	FR	HB	HG	HOH	KS	LB	PM	R
DO	0.612	0.308	0.322	0.216	0.201	0.247	0.219	0.305	0.122
FR	0.237	0.631	0.389	0.207	0.228	0.251	0.215	0.216	0.160
HB	0.196	0.277	0.673	0.137	0.127	0.248	0.183	0.220	0.177
HG	0.225	0.337	0.336	0.588	0.231	0.230	0.207	0.258	0.214
HOH	0.337	0.476	0.411	0.275	0.640	0.346	0.259	0.385	0.177
KS	0.060	0.127	0.170	0.077	0.100	0.634	0.069	0.162	0.032
LB	0.194	0.427	0.357	0.212	0.171	0.217	0.599	0.245	0.127
PM	0.321	0.328	0.383	0.200	0.278	0.344	0.183	0.592	0.154
R	0.224	0.392	0.399	0.398	0.299	0.316	0.189	0.317	0.644

Table 23: Mean assignment rates of individuals from 2018 estimated by using GENECLASS 2.0. Assignment rates of individuals into (rows) and from (columns) each population in 2018. Boldface indicates migration rate values (m) above 0.1. Grey indicates migration rate values (m) below 0.1.

2018	DO	FF	FR	HG	HOH	KS	LB	PM	R
DO	0.587	0.237	0.297	0.162	0.174	0.177	0.178	0.105	0.034
FF	0.298	0.556	0.290	0.196	0.192	0.172	0.188	0.107	0.091
FR	0.340	0.307	0.588	0.172	0.207	0.231	0.180	0.086	0.053
HG	0.326	0.295	0.270	0.603	0.271	0.108	0.275	0.183	0.148
HOH	0.305	0.300	0.316	0.215	0.615	0.180	0.223	0.136	0.114
KS	0.330	0.290	0.318	0.128	0.187	0.599	0.141	0.084	0.048
LB	0.172	0.184	0.183	0.143	0.201	0.104	0.589	0.120	0.141
PM	0.356	0.293	0.313	0.329	0.350	0.159	0.345	0.611	0.225
R	0.163	0.310	0.210	0.219	0.272	0.127	0.466	0.289	0.597

Table 24: Mean assignment rates of individuals from 2019 estimated by using GENECLASS 2.0. Assignment rates of individuals into (rows) and from (columns) each population in 2019. Boldface indicates migration rate values (m) above 0.1. Grey indicates migration rate values (m) below 0.1.

2019	DO	FF	FR	HG	HH	HOH	KS	LB	PM	R
DO	**0.580**	0.077	**0.103**	**0.222**	**0.159**	**0.133**	**0.202**	**0.145**	**0.206**	0.038
FF	**0.188**	**0.655**	**0.239**	**0.289**	**0.288**	**0.246**	**0.387**	**0.285**	**0.160**	0.079
FR	**0.258**	**0.376**	**0.606**	**0.356**	**0.384**	**0.342**	**0.387**	**0.362**	**0.193**	0.071
HG	**0.238**	**0.170**	**0.115**	**0.617**	**0.236**	**0.222**	**0.351**	**0.226**	**0.185**	**0.100**
HH	**0.146**	**0.167**	**0.174**	**0.291**	**0.598**	**0.308**	**0.359**	**0.227**	**0.162**	0.074
HOH	**0.213**	**0.202**	**0.233**	**0.299**	**0.428**	**0.599**	**0.429**	**0.234**	**0.201**	0.075
KS	**0.145**	**0.126**	0.094	**0.250**	**0.276**	**0.235**	**0.630**	**0.180**	**0.114**	0.055
LB	**0.164**	**0.193**	**0.188**	**0.326**	**0.276**	**0.283**	**0.359**	**0.571**	**0.149**	0.092
PM	**0.235**	0.096	0.090	**0.158**	**0.267**	**0.136**	**0.222**	**0.102**	**0.587**	0.074
R	**0.269**	**0.168**	**0.108**	**0.438**	**0.394**	**0.207**	**0.316**	**0.223**	**0.293**	**0.654**

5. Discussion

5.1. Fly Sampling

Sampling plays a crucial role in the analysis of genetic variation and population structure. Therefore, the sampling method and how it might affect the results will be discussed.

In this study, a population is represented by 20 individuals from one sample site. Unbiased population properties can only be estimated, if the sampling is random, especially regarding kinship (Wang, 2018). In this context, closely related individuals are defined as full and half-siblings. If those close relatives appear at a higher proportion in a population than expected under random sampling, accurate results cannot be expected. Such related samples bias (population) genetic studies is of great concern, especially in human genetics and medicine (Ott, 1992). In animal population studies, it was shown that allele frequency estimates were biased when testing juvenile brown trout individuals sampled from a constrained region since they were represented by a small number of families (Hansen et al., 1997). Another example shows that an excessive number of closely related samples influences Bayesian clustering algorithms and results in a misinterpreted number of K populations (Rodríguez-Ramilo et al., 2014; Rodríguez-Ramilo and Wang, 2012). To minimize the risk of sampling related individuals, SWD individuals were collected as described in 3.1. Samples were taken, if possible, from different plants, from different shrubs and trees in a 20 km radius or at different time points, and were separated in different cages upon arrival in the laboratory. Even with all these precautions, it cannot be excluded that some samples are related. If it is the case that sampled individuals were closely related, this should be visible in the results. A population of closely related individuals would show changes in the allele frequency and expected heterozygosity compared to other populations. Such a population might have exceptionally high numbers and frequencies of private alleles, and distinction and differentiation to other populations might be more prominent (Bonin et al., 2004; Waples and Anderson, 2017). The results do not show any of these signs. The positive F_{IS} value in 15 out of 28 populations would be the only value that implies relatedness between individuals within populations. However, a positive F_{IS} does not necessarily mean inbreeding. It can be related to null alleles as well (Waples, 2018). In comparison, the laboratory strains are inbred, and sampled individuals are indeed related. Even here, the F_{IS} value was positive in only four strains, three strains showed signs of random mating, and one showed an excess of heterozygotes. This indicates that F_{IS} alone is not enough to exclude random mating and might be misleading as a standalone value. Overall, it could be hypothesized that a German field population composed of closely related individuals would be more similar to one of the laboratory strains. However, this is not the case. Another argument against closely related samples in at least one population is that all field samples are similar in genetic variation, diversity, and structure. Still, it is unlikely that only relatives were sampled across all sample sites and years.

Another critical factor in population genetic studies is the sample size estimate. It is generally assumed that large sample sizes are imperative to estimate genetic diversity and differentiation (Holsinger and Weir, 2009; Kalinowski, 2005; Nazareno et al., 2017). While smaller sample sizes are prone to bias genetic diversity estimates, large sample sizes per population are more expensive in labor (Nazareno and Jump, 2012). Reducing the sample size per population, on the other hand, would allow the analysis of more populations for the same costs. Also, the number of genetic markers (here microsatellite markers) used in population genetic studies is critical. To strengthen the discriminative power of SSR markers, more markers should be used, and an increased number of microsatellite markers might also reduce the number of required samples per population (Willing et al., 2012). The problem is that the number of used SSR markers is often limited due to costs and the required workload needed (Landguth et al., 2012; Li et al., 2020). Ultimately, obtaining an accurate estimate of allele frequencies and diversity is more important than detecting all allele variants, because extremely rare alleles are not very informative (Hale et al., 2012). These factors were considered when defining a feasible amount of microsatellite markers for this study. Further, it had to be determined how many and which populations should be analyzed. Most important for this consideration was the availability of samples over all three years and the geographic distance between sample sites. To reflect the status of SWD populations in Germany, samples from the north should be represented in equal proportions as south German samples. This is also the reasoning behind keeping HB17, even though only ten samples from one year were available. Other sample sites from the south or the middle of Germany did not make it into the final experiments because the sampling locations were either too close geographically speaking or the region was already represented by another population, which performed better during sampling. The sample size estimate was then carried out based on these initial decisions. Under the assumption that the population standard deviation of allele size is at most seven (based on preliminary exploratory data analyses) and under the requirement that a 90%-confidence interval (CI) for the population mean of allele size has a length of at most 10 with a probability of 80% ('precision power'), a sample size of at least 19 was necessary. Based on this sample size estimation, I used 20 individuals from each location and year except for HB17, where only ten individuals were available.

5.2. Polymorphisms and Genetic Variability at 14 Microsatellite Loci in SWD

Overall, the microsatellite markers used in this study showed genetic diversity and a high PIC. Based on these results and compared to other studies (Fraimout et al., 2017; Tait et al., 2017), the markers are well suited for a population genetic study. Given the results from Tait et al. (2017) and Fraimout (2015, 2017), the values for polymorphism and genetic variability are similar. Comparisons to other population genetic studies in SWD are difficult since other SSR markers or other marker systems like mitochondrial markers were used (Bahder et al., 2015; Lavrinienko et al., 2017). Significant differences (p > 0.05) from HWE were observed in 18 of 42 year-locus combinations for the tested loci. In a population that fits HWE, no change in allele or genotype frequency will occur and remain constant over generations, given the absence of any evolutionary influences like gene flow, genetic drift, mate choice, bottleneck, or mutation (Mayo, 2008). HWE describes an idealized population that rarely applies in nature. Nevertheless, it does allow the measurement of genetic variation as deviations from the equilibrium. A possible explanation for the deviation from HWE is that the markers and populations are subject to evolutionary forces, and therefore, violate the rules for HWE (Chen, 2010). Another reason for the disequilibrium could be the presence of null alleles. Per definition, a microsatellite null allele is an allele at a microsatellite locus that does not amplify to detectable levels in a PCR (Chapuis and Estoup, 2007). Microsatellite markers are prone to variation in the nucleotide sequence of the flanking regions, which can prevent primer annealing to the template DNA (Chapuis and Estoup, 2007). A polymorphism in the primer binding region can result in null alleles (Callen et al., 1993). Other causes for null alleles are slippage during PCR or the better amplification of short alleles (Lai et al., 2003). Insects tend to have a high frequency of null alleles (Chapuis and Estoup, 2007). Null allele analysis with FreeNa pointed to only two loci with a null allele frequency > 10% (DS21 and DS17). This 10% value is often mentioned as a point at which null allele frequencies are problematic when testing for selection (Fraimout et al., 2015). Loci with higher frequency values than 10% show an excess of homozygotes, leading to an overestimation of inbreeding. However, some loci and populations showed null allele frequencies close to the 10% mark, including DS21, meaning some loci may bear undetected null alleles. The microsatellite marker DS21 was kept in the analysis because null allele frequencies higher than 1.0 were only detected in three (HB17, HOH19, LS_USA) out of 36 tested populations. Nevertheless, 14 populations had frequencies greater than 0.9 and thus were close to the 1.0 mark. Microsatellite loci with null alleles can be included in population genetic analysis, but they have to be treated with caution, and analysis needs a correction for potential bias because they can falsely reduce the population differentiation (Slatkin, 1995). Thus, genetic distance measures may incorrectly increase, but it is not clear to which extend these values may be influenced through null alleles (Paetkau et al., 1995; Slatkin, 1995). Further, null alleles were suggested to be present in two populations, HB17

and LS_USA. The result for LS_USA as a long-established laboratory strain is not unexpected. Inbreeding and artificial selection over the years might be the reasons for the higher number of detected microsatellite null alleles (Chybicki and Burczyk, 2009; Van Oosterhout et al., 2006). These detected null alleles are most likely homozygous for the specific locus. It is more surprising that not more laboratory strains show similar potential null alleles, but this might be due to the discriminative power and high variability in the used markers (Fraimout et al., 2015). In the case of the German field population HB17, the sample size might influence the result. It was the only population with a sample size of 10 individuals. The low sample size may negatively affect the null allele frequency, or samples were related and showed reduced heterozygosity because they share the homozygous allele variant (Chybicki and Burczyk, 2009).

When microsatellite loci are characterized, a test for Linkage disequilibrium is often conducted. Two microsatellite loci are in LD when their different alleles do not occur randomly to each other (Schwab, 2008). LD is an indicator of the population genetic forces that structure a genome (Slatkin, 2008) and is influenced by many factors, including selection, recombination and mutation rates, genetic drift, and population structure (Slatkin, 2008). LD can also reflect changes in populations. For example, it can help to track migration patterns (Slatkin, 2008). The number of 'founding' chromosomes carried over to new continents is often small, and variation induced by recombination is limited. Therefore, LD is often observed in populations that developed in recent times (Schwab, 2008). The high number of linked loci in German populations could arise from that effect. On the other hand, in the original publication of Fraimout et al. (2015), significant LD was found in all pairs of loci putatively located on the same chromosome arm. Since those microsatellite markers were used for this study, the physical linkage is presumably a likely cause for the detected LD. Another reason for the high number of linked loci when testing for LD might be that the test assumes that populations are in HWE (Schwab, 2008). However, as it was already noted, loci and populations show deviations from HWE. This might be the reason why even loci on different chromosomes were considered to be linked. Due to the close physical distance of the used microsatellite markers and the violation of HWE, the results for LD should be treated with caution.

5.3. Genetic Diversity and Relationship among Laboratory Strains

The results for genetic diversity reflect well the time the strains were kept in the laboratory, with LS_HG18 being the newest and most diverse strain (one year in culture) and LS_Canada as one of the longest established and least diverse strains (approx. eight years in culture). The number of private alleles in the tested laboratory strains was relatively high compared to the German field populations, suggesting a restricted gene flow between strains. Private alleles are unique to a single population but can occur at any frequency. An estimate of private alleles can be helpful because it can contribute to the identification of gene flow between populations (Slatkin, 1985). Populations with several private alleles are less likely to share their genetic material with other populations, indicating restricted gene flow.

These first impressions were strengthened when analyzing the genetic distance and relationship among the laboratory strains. AMOVA in laboratory strains indicates that most variation (72%) occurred within populations, while 28% of the variation was detected among populations. The results for pairwise F_{ST}, Nei's Genetic Distance/Identity, PCoA, and NJ confirmed this finding. Worth mentioning is that laboratory strains that are longer in captivity, namely LS_USA and LS_Canada, are more distinct from the 'younger' strains and laboratory strains originating from Europe are more similar to each other than to their North American counterparts. This is an expected result for the laboratory strains, considering the artificial nature of those strains. Proper stock-keeping implies a clean separation of strains and mixture between strains is only possible by accident or when the experimenter intends a crossing.

Furthermore, using the results from the laboratory strains, it was possible to demonstrate the possible consequences of using only laboratory strains in experiments. Experiments meant to produce data for field applications, like pest control in SWD, should be performed with freshly sampled flies or genetically refreshed strains since SWD laboratory strains differ from wild populations and can change reasonably within a short period. The newly established laboratory strain LS_HG illustrates the effects of laboratory inbreeding well, including the change of genetic markers and a decline in allelic diversity. The pairwise F_{ST} for laboratory strain LS_HG revealed more differentiation in the second year of inbreeding (2019) than in 2018. While the differentiation was moderate in 2018, it was moderate to strong in 2019, which means that the two years of inbreeding resulted in more substantial differentiation compared to its origin. The present data cannot conclude whether the observed decline is due to a random selection of individuals during stock keeping or due to laboratory 'adaptation'. However, over two years, the strain got more similar to other, older laboratory strains. This effect is important to consider during scientific experiments that rely on laboratory strains in field applications (Hamby et al., 2016; Kinjo et al., 2014; Lee et al.,

2011; Shearer et al., 2016). Laboratory strains enable transparent and easily reproducible experiments. Results from those studies are taken as the threshold for similar experiments and as a portrayal of natural processes. However, strains that arise from human cultivation undergo artificial evolutionary genetic changes (Knudsen et al., 2020). Therefore, for evaluations of the efficiency of naturally occurring predators, parasitoids, bacteria, or viruses for biological pest control, invasion history, or behavior, fresh field collections or genetically refreshed strains should be the first choice over highly inbred laboratory strains. It should be noted that this finding is only relevant for field applications but does not influence genetic monitoring of laboratory strains to keep genetically pure animals as it is common for example for rodents (Cohan et al., 2019; Guénet and Benavides, 2010).

5.4. Spatial and Temporal Genetic Variation *of Drosophila suzukii* in Germany

Invasive species like SWD constitute a threat to agriculture, economy, and biodiversity (Mooney and Cleland, 2001). While the development and improvement of control methods are undoubtedly extremely important to tackle a pest species, understanding population movement, genetic structure, and diversity are also crucial for developing pest management strategies (Fraimout et al., 2017). This study focused on the spatial and temporal genetic variation of SWD populations in Germany for three years (2017-2019) to help understand the genetic population development of this invasive pest species. So far, population genetic studies in Europe did not analyze population genetics in several consecutive years. This type of data is only available for the SWD populations in the USA (Bahder et al., 2015).

Our analysis of 14 microsatellite markers revealed that levels of genetic diversity in Germany are comparable with other European countries (Lavrinienko et al., 2017; Tait et al., 2017) and that genetic differentiation of sampled SWD is displayed among individuals within a single population but not among populations from different sample sites. None of the populations showed a gain or loss of genetic information over the years. The results suggest a substantial gene flow and a more homogeneous gene pool across different geographical populations. That gene flow between populations is present can also be confirmed by the lack of private or low-frequency alleles. Populations that share most of the alleles and show only a small number of private alleles with low frequencies are more likely to have experienced gene flow in recent generations. The F_{IS} value was positive in most populations. This is usually interpreted as a sign of inbreeding, but precautions were taken to avoid sampling related individuals in this study. As discussed above, another factor that can influence the F_{IS} value is the presence of null alleles. Since null alleles can cause a reduction in heterozygosity, the F_{IS} value increases. This might be the most plausible reason for positive F_{IS} values in the studied populations.

Results from this study suggest that SWD is a well-established, uniform population in Germany that might not be altered much by multiple invasions or reinvasions. Based on the low differentiation between populations and years, there are either no reinvasions, or they do not impact local populations. This is supported by the STRUCTURE, NJ, and PCoA results, which did not group German populations into distinct subpopulations. Methodological issues can be excluded because the marker system established by Fraimout et al. (2015) and used in this study detected genetic differences between our laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_Valsugana, LS_France, and LS_HG18 or LS_HG19). To measure the genetic differentiation between populations, an analysis of the F_{ST} was done, comparing the proportion of the genetic variation contained within a population relative to the total genetic variance. Following the other genetic distance analyses, the pairwise F_{ST} reveals only a minor differentiation ($F_{ST} < 0.1$) between all German populations and years. In contrast, when including laboratory strains as outgroups, most pairwise comparisons indicate a moderate differentiation. It could be criticized that the alternative to Wright's F-statistics, the so-called R-statistics (Slatkin, 1995), was not used. Slatkin's R_{ST} can be calculated using allele size variances, while Wright's F_{ST} is calculated from the variances of allele frequencies. Wright's F and Slatkin's R both have their advantages as well as drawbacks. R_{ST} assumes a stepwise mutation model and is thought to accurately reflect the mutation pattern of microsatellites (Balloux and Lugon-Moulin, 2002), but it has a high variance. The main drawback of F_{ST} is its sensitivity to the mutation rate when migration is low (Balloux and Lugon-Moulin, 2002). Nevertheless, to provide a less biased estimate of gene flow when sample sizes are moderate, F_{ST} should be performed instead of R_{ST} due to its high associated variance (Gaggiotti et al., 1999). Not only is the choice of F_{ST} or R_{ST} often highly discussed, but the interpretation can be difficult as well. In the case of F_{ST}, values can range from 0 to 1, where 0 means complete sharing of genetic material and 1 means no sharing at all. Populations with a value of 1 are completely isolated from one another and do not share any alleles. In reality, it is rarely larger than 0.5. Wright himself proposed that values of $F = 0.25$ are considered to show significant differentiation, values between 0.15 and 0.25 show moderate differentiation, and values smaller than 0.05 indicate no or negligible differentiation between subpopulations (Hartl and Clark, 1997; Wright, 1978). However, such a strict interpretation might not be accurate, as Wright (1978) explained. He pointed out that differentiation cannot be disregarded just because F_{ST} is smaller than 0.05. The effect of polymorphism can drastically lower F_{ST} outcomes, and even low values may indicate crucial genetic differentiation (Charlesworth, 1998; Nagylaki, 1998; Wright, 1978). Nevertheless, this does not seem to be true for the sampled and analyzed SWD populations.

Keeping these results in mind, the observed lack of admixture might be interpreted as evidence that colonization may have involved a single founder population rather than individuals from different origins. This is in contrast to findings from Ukraine, which found evidence for multiple sources of SWD invasions into Europe (Lavrinienko et al., 2017). On the other hand, our results are following results demonstrating that European SWD populations were genetically more homogenous with lower levels of genetic diversity compared to populations from North America (Adrion et al., 2014; Fraimout et al., 2017).

Other factors leading to genetic differentiation of local populations are mutation, genetic drift, and natural selection due to local (environmental) adaptation (Slatkin, 1987). Geographic barriers like mountains, lakes, and rivers can lead to genetic differences in populations. Germany is only streaked by an orogenic belt of relatively low mountains and hills, the Central German Uplands, but not by high mountain ranges. More extreme geographic landforms are not found in Germany, except for the Alps in the southernmost reach of the country. Other potential barriers include the two important waterways Rhine and Danube, a range of tributaries, islands along with the northwest coast and northeastern Baltic coast, and numerous lakes within German borders. Even though the characteristics of geographic barriers seem negligible, especially for a flying insect, it was speculated that differences between SWD populations might be present in the data. It was shown that the Rhine River and Lake Constance can act as geographic barriers for other animals like the small bank vole *Clethrionomys glareolus* (Gerlach and Musolf, 2000). The same study showed that more recent artificial fragmentations of landscapes, for example, highways, also have an important impact on gene flow and genetic population structure (Gerlach and Musolf, 2000). Habitat fragmentation is a known problem as it can cause small population sizes, obliteration of metapopulation structure, inbreeding, and a decrease in fertility due to low gene flow (Gonzalez et al., 1998; Hartl et al., 1992). Based on the data obtained in this study, especially the low variation detected among populations, it seems that neither geographic properties nor artificial fragmentation act as barriers for SWD. At least, they do not isolate populations from each other. As an insect capable of flying, SWD has apparent advantages over smaller mammals like the bank vole, which are greatly influenced by geographic and artificial barriers. SWD can move from high to low elevation and travel long distances by flight (Tait et al., 2018). Nevertheless, distinct SWD populations from the island versus the mainland can be found in Italy (Tait et al., 2017). In this respect, it would be interesting to analyze samples from a German island in the North Sea or Baltic Sea in future experiments.

Another factor influencing populations on a genetic level is climate and the short-termed localized variation in weather. The consequences and effects of long-termed climate changes on genetic diversity have been studied across many taxa and geographical

landscapes, including invasive alien species (Galatowitsch et al., 2009; Hellmann and Pineda-Krch, 2007; Mainka and Howard, 2010). If generation times are short and adaptive capacity is enhanced, genetic diversity can be affected rapidly (Avolio et al., 2013; Proft et al., 2021). A prediction of what type of evolutionary responses can be expected or whether it is likely to help or hinder species expansions remains a challenging question, foremost because it is difficult to perform experiments over a long enough period and secondly because climate and weather itself are difficult to predict (Moran and Alexander, 2014; Moran et al., 2017). In the case of SWD, a study in the US found a significantly lower genetic diversity and a bottleneck event in a SWD population from eastern Washington compared to a coastal California population. It concluded that it is plausible that the more extreme climate in Washington could be the issue for this result (Bahder et al., 2015). The climate in Germany is temperate throughout the country, even though temperature extremes, especially in the summer, are occurring more frequently (Carney and Kantz, 2020; Estrella and Menzel, 2013). Still, climate, temperature, and humidity differ to a certain extent between sampling locations and years, and it could have been expected to influence the German SWD population, but we could not detect these differences in our data. These findings again match the results from Tait et al. (2017), who did not detect differences between the populations from the much colder climate of Trentino compared to the rest of Italy.

Another factor that can be interesting regarding population structure is the comparison of rural and urbanized areas. That urbanization can substantially affect biodiversity and population genetic variation and differentiation is well known (Johnson and Munshi-South, 2017). Human-induced environmental changes have an unprecedented influence on the adaptation of invasive organisms (McDonnell and Hahs, 2015; McDonnell and Pickett, 1990). Invasive species are most likely to thrive in these changing conditions since they can increase the availability of food resources and overwintering habitats (Millennium ecosystem assessment, 2005; Santana Marques et al., 2020; Weaver et al., 2011). This seems to be true for the tested SWD populations in Germany. No differences were found between sample sites close to cities like Dortmund (DO) and populations in more rural areas like Derwitz (PM), indicating that urbanization does not adversely affect SWD populations.

Considering the invasion history and success of SWD in other parts of the world (Asplen et al., 2015), it is not surprising that the results obtained in this study indicate a well-established SWD metapopulation in Germany. The term 'established' in the context of invasion biology refers to self-maintaining populations of non-native species (Hayes and Barry, 2008). Reproduction-related characteristics are often positively associated with invasion success (Allen et al., 2017). These traits are also referred to as life-history traits and are thought to underlay introduction success and determine a species' growth rate

(Allen et al., 2017). They include the amount of offspring, the frequency of reproduction, the age at which sexual maturity is reached, and the reproductive duration (Sæther et al., 2013). SWD has a relatively high reproductive capacity, with a single female laying hundreds of eggs during its life with up to 10 generations a year (Dalton et al., 2011), which plays an important role in its invasion success (Allen et al., 2017). Invasion success is also facilitated by the similarity between the habitats of native and introduced ranges (Kolar and Lodge, 2001), and SWD fits the European and German ecosystem, not only in means of climate but also due to a constrained number of natural predators and parasitoids present (Chabert et al., 2012; Stacconi et al., 2015). The 'Enemy Release Hypothesis' (ERH) suggests that the establishment and population growth of an invasive species is heavily dependent on the absence or reduced effectiveness of naturally occurring enemies (Keane and Crawley, 2002; Liu and Stiling, 2006). In terms of parasitoids, only three larval endoparasitoids, namely *Ganaspis xanthopoda* (Hymenoptera: Figitidae), *Asobara tabida* (Hymenoptera: Braconidae), and *Asobara japonica*, have been successfully reared on SWD, all three from the native range of SWD in Japan (Mitsui et al., 2007). To date, only three generalist parasitoids, *Pachycrepoideus vindemiae* (Hymenoptera: Pteromalidae) and *Leptopilina heterotoma* (Hymenoptera: Figitidae) and *Trichopria drosophilae* (Hymenoptera: Diapriidae), were found to parasitize SWD with low effectiveness in Europe (Chabert et al., 2012; Stacconi et al., 2015). Other potential European larval or pupal parasitoids are either unable to develop or rarely oviposit in SWD due to a strong immune response in the host (Chabert et al., 2012).

One critical topic regarding invasive alien species is the movement of people and goods on a global scale that eases the worldwide introduction and passive spread of non-native invasive species (Hulme, 2009, 2014). The increasing number of invasive species in Europe shows a consistent pattern with the increases in trade and travel (Jeschke and Strayer, 2005, 2006; Keller et al., 2011). All invasive species, independent from their origin, have in common that passive transport enables their successful spread (Banks et al., 2015). Anthropogenic transport of goods is an important aspect for the dissemination of flies since it facilitates the gene flow between locations, and international trade via air and sea transport provides new pathways for the spread of insect pests in general (Hulme, 2009). Even though it seems to be impossible to reconstruct the exact routes in detail due to the enormous amount of imported fresh produce, it is reasonable to assume that transportation of host fruits and plants lead to an extensive movement of SWD or other pests not only across Germany but all over Europe (Cini et al., 2014). Germany is importing large amounts of host plants and crops from all around the world. A look at the importation routes shows that most fresh fruits are imported from within Europe, namely Spain and Italy, and from the USA and South America (source: International Trade Center, www.trademap.org). In that respect, distinct populations that originated from different countries or even continents could

be possible, but we did not find proof for this. With the data obtained in this study, it is not possible to reconstruct invasion routes from outside Germany, but it would be an interesting aspect for further experiments. However, our collection did not include SWD samples from countries like Spain or the US, and, therefore, it is not clear where the German population might have originated from. However, Fraimout et al. (2017) did find evidence that the German sample site in their experiment most likely originated from an admixture of SWD populations from Asia and the eastern US. In general, a better comprehension of genetic structure, population dynamics, and the reconstruction of invasion routes could improve pest control at a regional scale.

6. Conclusion and Outlook

The results obtained using microsatellite markers suggest that the sampled SWD populations across Germany contain the same level of genetic diversity. The results suggest substantial gene flow and a more homogeneous gene pool across different geographical populations, but no changes were detected over a three-year sampling period.

This study indicates that SWD is a well-established species in Germany. This would ultimately imply that it might be difficult to eradicate this invasive pest at all. Based on the data presented here, it is impossible to evaluate if SWD is already well adapted to its 'new' habitat in Germany or if there is no need for further adaptation due to a lack of natural enemies or competitors. One way or another, it is plausible to assume that Central and Western European habitats seem to provide a thriving environment for SWD. Furthermore, since genetic characteristics affect the success of an invasive species, SWD might even harbor a high potential for further spread (Allen et al., 2017; Sæther et al., 2013). Understanding the biology of an invasive species can help define management strategies, mitigate its spread and the damage it causes and predict further outbreaks (Fraimout et al., 2015). For SWD, the development of pesticide resistance would be the worst-case scenario since there are no valid alternatives. The main problem remains that only a few approved pesticides affect SWD successfully (Haviland and Beers, 2012; Shawer et al., 2018). The pesticides currently used are under strict regulations from the EU or banned from the market entirely, while alternative methods are not promising yet (Chabert et al., 2012). However, since only a few alternatives are available, the risk of SWD developing resistance against those chemicals increases with time, particularly concerning the high genetic diversity we found in the analyzed German populations. There is an urgent need for the development of alternative pest control methods besides chemical pesticides.

Improved trapping and monitoring have to be the first measures to prevent a further spread or frequent reintroductions. Then, integrated pest management, which incorporates

biological control methods, could be a valuable tool to fight SWD spread. This includes the usage of natural parasitoids and predators as it has been already investigated for SWD but with admittedly little success (Chabert et al., 2012; Stacconi et al., 2015). The incorporation of other methods like exclusion netting can be a valuable option as well. Augel et al. (2020) found that Exnet systems can successfully prevent SWD infestation but they are accompanied by several problems. The installation of Exnet systems involves an estimated annual investment of 410 to 1,620 €/ha (Augel et al., 2020). Other disadvantages are that Exnets require proper maintenance to be of use against SWD, Exnets do not necessarily prevent infestations with other pests like spider mites, and they also exclude beneficial insects, resulting in pollination problems that in turn have to be countered by placing bumblebees or honeybees within the Exnet (Augel et al., 2020; Boehnke et al., 2019; Kuesel et al., 2019; Leach et al., 2016). Another method that could be interesting for pest control in SWD is the Male Annihilation Technique (MAT) that is used in the suppression of tephritid pest species (Vargas et al., 2014). In MAT, traps are used which are baited with a male lure in combination with an insecticide (Vargas et al., 2003). The male proportion of a population is reduced, causing suppression or even eradication of the whole population (Steiner et al., 1965; Steiner and Lee, 1955; Vargas et al., 2003). Similar lures are not available for SWD, fermented food baits such as vinegar or yeast solutions are used instead (Cini et al., 2012; Cloonan et al., 2019; Walsh et al., 2011). This type of trap has the undesired side effect of trapping also nontarget insects (Cha et al., 2013). A lure that is more specific to SWD, ideally not even for males but for female SWD, could be an important tool for pest control (Cha et al., 2013).

Another option for integrated pest management and environmentally friendly alternatives to pesticides could be the Sterile Insect Technique (SIT) that has proven highly effective in agricultural insect species (Augustinos et al., 2017; Benedict and Robinson, 2003; Krafsur, 1998; Wyss, 2000). For SIT programs, sterilized male individuals are released into the environment and lead to infertile mating that could reduce the population size of SWD in the future. The benefits of SIT application in SWD would be its species-specificity and a safe application at any time in the season without any risk on human or environmental health (Sassù et al., 2020). An important aspect for successful mass rearing is the development of protocols guaranteeing a cost-effective and stable production of insects (Gast, 1968). First efforts have already been made to develop efficient mass rearing methods for SWD. Sassù et al. (2019) evaluated the efficiency of two oviposition systems and found that cages equipped with a wax panel resulted in more eggs, higher viability, and emergence rate (Sassù et al., 2019). In contrast, cages with a netted oviposition system did not perform as well (Sassù et al., 2019). One of the most critical parts of mass rearing protocols is the larval diet since it has a great influence on operational costs and the quality of the insects (Parker et al., 2021). The diet for SWD is mainly based on expensive brewer's yeast as a protein

source (Hardin et al., 2015; Lewis and Hamby, 2019; Spitaler et al., 2020). To circumvent this problem, Nikolouli et al. (2021) were experimenting with *Enterobacter* sp. AA26 as a cheaper replacement for brewer's yeast in the SWD diet. In medfly, using *Enterobacter* sp. AA26 dry biomass as a yeast replacement in the diet was successful (Kyritsis et al., 2019), while Nikolouli et al. (2021) showed that a full replacement in the SWD diet resulted in decreased fitness and fertility. However, SWD fed with a diet that replaced the yeast only partially performed much better and without severe effects. Nikolouli et al. (2021) suggest that halving the yeast quantity is still sufficient to produce fit adults and could reduce the costs for a potential SWD mass production. Those findings bear great potential for mass production of SWD for SIT. In addition, efficient and successful SIT programs rely on additional aspects besides mass rearing protocols and larval diet. Several current SIT programs consider the genetic background refreshing as a vital tool for mating success and efficacy of the release programs (Estes et al., 2012; Parreño et al., 2014; Zygouridis et al., 2014). Due to our findings of genetic uniformity in wild German SWD populations, we hypothesize that this could be beneficial for the mating success of a single suitable mass-reared SWD strain to different wild-type populations during SIT programs. Therefore, while SWD control is still challenging, biological control methods, including the SIT, remain a beneficial option for sustainable pest control.

References

Adrion, J.R., Kousathanas, A., Pascual, M., Burrack, H.J., Haddad, N.M., Bergland, A.O., Machado, H., Sackton, T.B., Schlenke, T.A., Watada, M., *et al.* (2014). *Drosophila suzukii*: the genetic footprint of a recent, worldwide invasion. Mol Biol Evol *31*, 3148-3163.

Aketarawong, N., Chinvinijkul, S., Orankanok, W., Guglielmino, C.R., Franz, G., Malacrida, A.R., and Thanaphum, S. (2011). The utility of microsatellite DNA markers for the evaluation of area-wide integrated pest management using SIT for the fruit fly, *Bactrocera dorsalis* (Hendel), control programs in Thailand. Genetica *139*, 129-140.

Aketarawong, N., Isasawin, S., and Thanaphum, S. (2014). Evidence of weak genetic structure and recent gene flow between *Bactrocera dorsalis* s.s. and *B. papayae*, across Southern Thailand and West Malaysia, supporting a single target pest for SIT applications. Bmc Genet *15*.

Al-Samarai, F.R., and Al-Kazaz, A.A. (2015). Molecular markers: An introduction and applications. European Journal of Molecular Biotechnology, 118-130.

Allen, W.L., Street, S.E., and Capellini, I. (2017). Fast life history traits promote invasion success in amphibians and reptiles. Ecology Letters *20*, 222-230.

Altizer, S., Ostfeld, R.S., Johnson, P.T., Kutz, S., and Harvell, C.D. (2013). Climate change and infectious diseases: from evidence to a predictive framework. Science *341*, 514-519.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

Amos, W., and Harwood, J. (1998). Factors affecting levels of genetic diversity in natural populations. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences *353*, 177-186.

Asplen, M.K., Anfora, G., Biondi, A., Choi, D.-S., Chu, D., Daane, K.M., Gibert, P., Gutierrez, A.P., Hoelmer, K.A., and Hutchison, W.D. (2015). Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. Journal of Pest Science *88*, 469-494.

Augel, C., Boehnke, B., Wichura, A., Lindstaedt, J., Wiebusch, J., Engel, A., Benz, S., Saltzmann, J., Eberhardt, G., and Vogt, H. (2020). Demonstration project exclusion netting for managing spotted wing drosophila in fruit crops—Results 2017–2019. Paper presented at: Proceedings of the Ecofruit 19th International Conference on Organic Fruit Growing, Hohenheim, Germany.

Augustinos, A., Targovska, A., Cancio-Martinez, E., Schorn, E., Franz, G., Cáceres, C., Zacharopoulou, A., and Bourtzis, K. (2017). *Ceratitis capitata* genetic sexing strains: laboratory evaluation of strains from mass-rearing facilities worldwide. Entomologia Experimentalis et Applicata *164*, 305-317.

Avolio, M.L., Beaulieu, J.M., and Smith, M.D. (2013). Genetic diversity of a dominant C 4 grass is altered with increased precipitation variability. Oecologia *171*, 571-581.

Azrag, R.S., Ibrahim, K., Malcolm, C., El Rayah, E., and El-Sayed, B. (2016). Laboratory rearing of *Anopheles arabiensis*: impact on genetic variability and implications for Sterile Insect Technique (SIT) based mosquito control in northern Sudan. Malaria journal *15*, 1-8.

Bahder, B.W., Bahder, L.D., Hamby, K.A., Walsh, D.B., and Zalom, F.G. (2015). Microsatellite variation of two Pacific coast *Drosophila suzukii* (Diptera: Drosophilidae) populations. Environmental entomology *44*, 1449-1453.

Balding, D.J., and Nichols, R.A. (1995). A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica *96*, 3-12.

Balloux, F., and Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. Mol Ecol *11*, 155-165.

Banks, N.C., Paini, D.R., Bayliss, K.L., and Hodda, M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. Ecology letters *18*, 188-199.

Baroffio, C., and Fischer, S. (2011). Neue Bedrohung für Obstplantagen und Beerenpflanzen: die Kirschessigfliege. UFA Revue *11*, 46-47.

Barton, N.H., and Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. Annual review of ecology and systematics *15*, 133-164.

Barton, N.H., and Hewitt, G.M. (1985). Analysis of hybrid zones. Annual review of Ecology and Systematics *16*, 113-148.

Beggs, J. (2001). The ecological consequences of social wasps (*Vespula* spp.) invading an ecosystem that has an abundant carbohydrate resource. Biol Conserv *99*, 17-28.

Behura, S.K. (2006). Molecular marker systems in insects: current trends and future avenues. Mol Ecol *15*, 3087-3113.

Bellamy, D.E., Sisterson, M.S., and Walse, S.S. (2013). Quantifying host potentials: indexing postharvest fresh fruits for spotted wing drosophila, *Drosophila suzukii*. PloS one *8*, e61227.

Benedict, M.Q., and Robinson, A.S. (2003). The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends in parasitology *19*, 349-355.

Bergougnoux, V. (2014). The history of tomato: from domestication to biopharming. Biotechnology advances *32*, 170-189.

Boehnke, B., Köppler, K., Augel, C., Wichura, A., Lindstaedt, J., Wiebusch, J., Engel, A., Benz, S., and Vogt, H. (2019). Demonstration project "Exclusion netting for managing spotted wing Drosophila in fruit crops". Results 2017. IOBC-WPRS Bull *144*, 78-84.

Bolda, M.P., Goodhue, R.E., and Zalom, F.G. (2010). Spotted wing drosophila: potential economic impact of a newly established pest. Agricultural and Resource Economics Update *13*, 5-8.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. Mol Ecol *13*, 3261-3273.

Boulesteix, M., Weiss, M., and Biémont, C. (2006). Differences in genome size between closely related species: the *Drosophila melanogaster* species subgroup. Molecular biology and evolution *23*, 162-167.

Briem, F., Breuer, M., Köppler, K., and Vogt, H. (2015). Phenology and occurrence of spotted wing Drosophila in Germany and case studies for its control in berry crops. IOBC-WPRS Bull *109*, 233-237.

Briem, F., Dominic, A.R., Sinn, C., Golla, B., Hoffmann, C., Englert, C., Herz, A., and Vogt, H. (2018). Explorative Datenanalyse zum Auftreten der Kirschessigfliege, Drosophila suzukii, anhand von Monitoringdaten aus DrosoMon. Paper presented at: 61 Deutsche Pflanzenschutztagung (Hohenheim, Germany).

Brookfield, J. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol *5*, 453-455.

Bruck, D.J., Bolda, M., Tanigoshi, L., Klick, J., Kleiber, J., DeFrancesco, J., Gerdeman, B., and Spitler, H. (2011). Laboratory and field comparisons of insecticides to reduce infestation of *Drosophila suzukii* in berry crops. Pest management science *67*, 1375-1385.

Bruederle, L.P., Rogers, D.L., Krutovskii, K.V., and Politov, D.V. (2001). Population genetics and evolutionary implications. Whitebark Pine Communities: Ecology and Restoration Islands Press Washington, DC, USA, 137-153.

Bruford, M.W., and Wayne, R.K. (1993). Microsatellites and their application to population genetic studies. Current opinion in genetics & development *3*, 939-943.

Burr, B., Evola, S., Burr, F., and Beckmann, J. (1983). The application of restriction fragment length polymorphism to plant breeding. In Genetic engineering, J. Setlow, and H. A, eds. (Springer, Boston, MA), pp. 45-59.

Calabria, G., Máca, J., Bächli, G., Serra, L., and Pascual, M. (2012). First records of the potential pest species *Drosophila suzukii* (Diptera: Drosophilidae) in Europe. Journal of Applied entomology *136*, 139-147.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C., and Sutherland, G.R. (1993). Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. American journal of human genetics *52*, 922.

Carney, M., and Kantz, H. (2020). Robust regional clustering and modeling of nonstationary summer temperature extremes across Germany. Advances in Statistical Climatology, Meteorology and Oceanography *6*, 61-77.

Cha, D.H., Hesler, S.P., Cowles, R.S., Vogt, H., Loeb, G.M., and Landolt, P.J. (2013). Comparison of a synthetic chemical lure and standard fermented baits for trapping *Drosophila suzukii* (Diptera: Drosophilidae). Environmental entomology *42*, 1052-1060.

Chabert, S., Allemand, R., Poyet, M., Eslin, P., and Gibert, P. (2012). Ability of European parasitoids (Hymenoptera) to control a new invasive Asiatic pest, *Drosophila suzukii*. Biological Control *63*, 40-47.

Chapuis, M.-P., and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. Molecular biology and evolution *24*, 621-631.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. Molecular biology and evolution *15*, 538-543.

Chen, J.J. (2010). The Hardy-Weinberg principle and its applications in modern population genetics. Frontiers in Biology *5*, 348-353.

Chireceanu, C., Chiriloaie, A., and Teodoru, A. (2015). First record of the spotted wing Drosophila *Drosophila suzukii* (Diptera: Drosophilidae) in Romania. Romanian Journal for Plant Protection *8*, 86-95.

Chiu, J.C., Jiang, X., Zhao, L., Hamm, C.A., Cridland, J.M., Saelao, P., Hamby, K.A., Lee, E.K., Kwok, R.S., and Zhang, G. (2013). Genome of *Drosophila suzukii*, the spotted wing drosophila. G3: Genes, Genomes, Genetics *3*, 2257-2271.

Chybicki, I.J., and Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. Journal of Heredity *100*, 106-113.

Cini, A., Anfora, G., Escudero-Colomar, L., Grassi, A., Santosuosso, U., Seljak, G., and Papini, A. (2014). Tracking the invasion of the alien fruit pest *Drosophila suzukii* in Europe. Journal of Pest Science *87*, 559-566.

Cini, A., Ioriatti, C., and Anfora, G. (2012). A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. Bulletin of Insectology *65*, 149-160.

Cloonan, K.R., Hernández-Cumplido, J., De Sousa, A.L.V., Ramalho, D.G., Burrack, H.J., Della Rosa, L., Diepenbrock, L.M., Ballman, E., Drummond, F.A., and Gut, L.J. (2019). Laboratory and field evaluation of host-related foraging odor-cue combinations to attract *Drosophila suzukii* (Diptera: Drosophilidae). Journal of economic entomology *112*, 2850-2860.

Cohan, R.A., Inanlou, D.N., Aref, M.H.S., Zeinali, S., and Farhoudi, R. (2019). Microsatellite Marker Analysis for Laboratory Mice Profiling. Advanced biomedical research *8*.

Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A., and Solignac, M. (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics *153*, 1989-2000.

Dakin, E., and Avise, J. (2004). Microsatellite null alleles in parentage analysis. Heredity *93*, 504-509.

Dalton, D.T., Walton, V.M., Shearer, P.W., Walsh, D.B., Caprile, J., and Isaacs, R. (2011). Laboratory survival of *Drosophila suzukii* under simulated winter conditions of the Pacific Northwest and seasonal field trapping in five primary regions of small and stone fruit production in the United States. Pest management science *67*, 1368-1374.

De Ros, G., Anfora, G., Grassi, A., and Ioriatti, C. (2013). The potential economic impact of *Drosophila suzukii* on small fruits production in Trentino (Italy). IOBC-WPRS Bull *91*, 317-321.

De Ros, G., Conci, S., Pantezzi, T., and Savini, G. (2015). The economic impact of invasive pest *Drosophila suzukii* on berry production in the Province of Trento, Italy. Journal of Berry Research *5*, 89-96.

Dehnen-Schmutz, K. (2011). Determining non-invasiveness in ornamental plants to build green lists. Journal of Applied Ecology *48*, 1374-1380.

Del Fava, E., Ioriatti, C., and Melegaro, A. (2017). Cost–benefit analysis of controlling the spotted wing drosophila (*Drosophila suzukii* (Matsumura)) spread and infestation of soft fruits in Trentino, Northern Italy. Pest management science *73*, 2318-2327.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological) *39*, 1-22.

Di Rienzo, A., Peterson, A., Garza, J., Valdes, A., Slatkin, M., and Freimer, N. (1994). Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences *91*, 3166-3170.

Drake, J.A. (2017). Handbook of alien species in Europe (Springer).

Dukes, J.S., and Mooney, H.A. (1999). Does global change increase the success of biological invaders? Trends in Ecology & Evolution *14*, 135-139.

Earl, D.A., and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources *4*, 359-361.

Ellegren, H., and Galtier, N. (2016). Determinants of genetic diversity. Nature Reviews Genetics *17*, 422-433.

Ellstrand, N.C., and Elam, D.R. (1993). Population genetic consequences of small population size: implications for plant conservation. Annual review of Ecology and Systematics *24*, 217-242.

Estes, A., Nestel, D., Belcari, A., Jessup, A., Rempoulakis, P., and Economopoulos, A. (2012). A basis for the renewal of sterile insect technique for the olive fly, *Bactrocera oleae* (Rossi). Journal of Applied Entomology *136*, 1-16.

Estoup, A., and Angers, B. (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical consideration. Advances in molecular ecology.

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

Estrella, N., and Menzel, A. (2013). Recent and future climate extremes arising from changes to the bivariate distribution of temperature and precipitation in Bavaria, Germany. International Journal of Climatology *33*, 1687-1695.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol *14*, 2611-2620.

Ewens, W.J. (2012). Mathematical population genetics 1: theoretical introduction, Vol 27 (Springer Science & Business Media).

Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources *10*, 564-567.

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

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

FAOSTAT (2021). Food and agricultural trade dataset of the Food and Agriculture Organization of the United Nations, pp. Germany, Crops and livestock products, fresh fruits.

Fraimout, A., Debat, V., Fellous, S., Hufbauer, R.A., Foucaud, J., Pudlo, P., Marin, J.-M., Price, D.K., Cattel, J., and Chen, X. (2017). Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. Molecular biology and evolution *34*, 980-996.

Fraimout, A., Loiseau, A., Price, D.K., Xuéreb, A., Martin, J.-F., Vitalis, R., Fellous, S., Debat, V., and Estoup, A. (2015). New set of microsatellite markers for the spotted-wing *Drosophila suzukii* (Diptera: Drosophilidae): a promising molecular tool for inferring the invasion history of this major insect pest. Eur J Entomol *112*, 855.

Francis, R.M. (2017). pophelper: an R package and web app to analyse and visualize population structure. Molecular ecology resources *17*, 27-32.

François, O., Ancelet, S., and Guillot, G. (2006). Bayesian clustering using hidden Markov random fields in spatial population genetics. Genetics *174*, 805-816.

Freda, P., and Braverman, J. (2013). *Drosophila suzukii*, or spotted wing Drosophila, recorded in Southeastern Pennsylvania, USA. Entomological News *123*, 71-75.

Frydenberg, J., Pertoldi, C., Dahlgaard, J., and Loeschcke, V. (2002). Genetic variation in original and colonizing *Drosophila buzzatii* populations analysed by microsatellite loci isolated with a new PCR screening method. Mol Ecol *11*, 181-190.

Gabarra, R., Riudavets, J., Rodríguez, G.A., Pujade-Villar, J., and Arnó, J. (2015). Prospects for the biological control of *Drosophila suzukii*. BioControl *60*, 331-339.

Gaggiotti, O., Lange, O., Rassmann, K., and Gliddon, C. (1999). A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Mol Ecol *8*, 1513-1520.

Galatowitsch, S., Frelich, L., and Phillips-Mao, L. (2009). Regional climate change adaptation strategies for biodiversity conservation in a midcontinental region of North America. Biol Conserv *142*, 2012-2022.

Gast, R. (1968). Mass rearing of insects: its concept, methods and problems. International Atomic Energy Agency, IEAE (Ed) Radiation, Radioisotopes, and Rearing Methods in the Control of Insect Pests Int Ato Energy Agency, Vienna, p59-67.

Gemayel, R., Cho, J., Boeynaems, S., and Verstrepen, K.J. (2012). Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes *3*, 461-480.

Gerlach, G., and Musolf, K. (2000). Fragmentation of landscape as a cause for genetic subdivision in bank voles. Conservation biology *14*, 1066-1074.

Gillespie, J.H. (2004). Population genetics: a concise guide (JHU Press).

Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. Proceedings of the National Academy of Sciences *92*, 6723-6727.

Gonzalez, A., Lawton, J.H., Gilbert, F., Blackburn, T.M., and Evans-Freke, I. (1998). Metapopulation dynamics, abundance, and distribution in a microecosystem. Science *281*, 2045-2047.

Gragg, H., Harfe, B.D., and Jinks-Robertson, S. (2002). Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. Molecular and Cellular Biology *22*, 8756-8762.

Guénet, J.-L., and Benavides, F.J. (2010). Genetic Monitoring of Laboratory Rodents. Molecular Diagnostics, 461-469.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., and Salin, F. (2011). Current trends in microsatellite genotyping. Molecular ecology resources *11*, 591-611.

Guo, X., and Elston, R. (1999). Linkage information content of polymorphic genetic markers. Human Heredity *49*, 112-118.

Haldane, J.S. (1954). An exact test for randomness of mating. Journal of Genetics *52*, 631-635.

Hale, M.L., Burg, T.M., and Steeves, T.E. (2012). Sampling for microsatellite -based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PloS one *7*, e45170.

Hamby, K.A., Bellamy, D.E., Chiu, J.C., Lee, J.C., Walton, V.M., Wiman, N.G., York, R.M., and Biondi, A. (2016). Biotic and abiotic factors impacting development, behavior, phenology, and reproductive biology of *Drosophila suzukii*. Journal of pest science *89*, 605-619.

Hansen, M.M., Nielsen, E.E., and Mensberg, K.L. (1997). The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout *Salmo trutta* L. Mol Ecol *6*, 469-474.

Hardin, J.A., Kraus, D.A., and Burrack, H.J. (2015). Diet quality mitigates intraspecific larval competition in *Drosophila suzukii*. Entomologia Experimentalis et Applicata *156*, 59-65.

Hartl, D.L., and Clark, A.G. (1997). Principles of population genetics. In Principles of population genetics (Sunderland, MA: Sinauer associates), pp. 542-542.

Hartl, G.B., Markowski, J., Świątecki, A., Janiszewski, T., and Willing, R. (1992). Studies on the European hare. 43. Genetic diversity in the Polish brown hare *Lepus europaeus* Pallas, 1778: implications for conservation and management. Acta Theriologica *37*, 15-25.

Hauser, M. (2011). A historic account of the invasion of *Drosophila suzukii* (Matsumura)(Diptera: Drosophilidae) in the continental United States, with remarks on their identification. Pest management science *67*, 1352-1357.

Haviland, D.R., and Beers, E.H. (2012). Chemical control programs for *Drosophila suzukii* that comply with international limitations on pesticide residues for exported sweet cherries. Journal of Integrated Pest Management *3*, F1-F6.

Hawkes, J.G., and Francisco-Ortega, J. (1993). The early history of the potato in Europe. Euphytica *70*, 1-7.

Hayden, M.J., Nguyen, T.M., Waterman, A., and Chalmers, K.J. (2008). Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. BMC genomics *9*, 1-12.

Hayes, K.R., and Barry, S.C. (2008). Are there any consistent predictors of invasion success? Biological invasions *10*, 483-506.

Hellmann, J.J., and Pineda-Krch, M. (2007). Constraints and reinforcement on adaptation under climate change: Selection of genetically correlated traits. Biol Conserv *137*, 599-609.

Hensen, I., and Oberprieler, C. (2005). Effects of population size on genetic diversity and seed production in the rare *Dictamnus albus* (Rutaceae) in central Germany. Conservation genetics *6*, 63-73.

Hernandez, R.D., Hubisz, M.J., Wheeler, D.A., Smith, D.G., Ferguson, B., Rogers, J., Nazareth, L., Indap, A., Bourquin, T., and McPherson, J. (2007). Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science *316*, 240-243.

Hill, W.G. (2014). Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. Genetics *196*, 1-16.

Hirst, C.N., and Jackson, D.A. (2007). Reconstructing community relationships: the impact of sampling error, ordination approach, and gradient length. Diversity and distributions *13*, 361-371.

Holsinger, K.E., and Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nature Reviews Genetics *10*, 639-650.

Hu, Y., Thapa, A., Fan, H., Ma, T., Wu, Q., Ma, S., Zhang, D., Wang, B., Li, M., and Yan, L. (2020). Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. Science advances *6*, eaax5751.

Hubisz, M.J., Falush, D., Stephens, M., and Pritchard, J.K. (2009). Inferring weak population structure with the assistance of sample group information. Molecular ecology resources *9*, 1322-1332.

Hughes, A.R., Inouye, B.D., Johnson, M.T., Underwood, N., and Vellend, M. (2008). Ecological consequences of genetic diversity. Ecology letters *11*, 609-623.

Hulme, P.E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology *46*, 10-18.

Hulme, P.E. (2014). Invasive species challenge the global response to emerging diseases. Trends in Parasitology *30*, 267-270.

Idrees, M., and Irshad, M. (2014). Molecular markers in plants for analysis of genetic diversity: a review. European academic research *2*, 1513-1540.

Iglesias, L.E., Nyoike, T.W., and Liburd, O.E. (2014). Effect of trap design, bait type, and age on captures of *Drosophila suzukii* (Diptera: Drosophilidae) in berry crops. Journal of economic entomology *107*, 1508-1518.

Jackson, D.A., Somers, K.M., and Harvey, H.H. (1989). Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence? The American Naturalist *133*, 436-453.

Jakobs, R., Gariepy, T.D., and Sinclair, B.J. (2015). Adult plasticity of cold tolerance in a continental-temperate population of *Drosophila suzukii*. Journal of insect physiology *79*, 1-9.

Janes, J.K., Miller, J.M., Dupuis, J.R., Malenfant, R.M., Gorrell, J.C., Cullingham, C.I., and Andrew, R.L. (2017). The K= 2 conundrum. Mol Ecol *26*, 3594-3602.

Jarausch, B., Müller, T., Gramm, T., and Hoffmann, C. (2017). Comparative evaluation of insecticide efficacy tests against *Drosophila suzukii* on grape berries in laboratory, semi-field and field trials. Vitis *56*, 133-140.

Jeschke, J.M., and Strayer, D.L. (2005). Invasion success of vertebrates in Europe and North America. Proceedings of the National Academy of Sciences *102*, 7198-7202.

Jeschke, J.M., and Strayer, D.L. (2006). Determinants of vertebrate invasion success in Europe and North America. Global Change Biology *12*, 1608-1619.

Johansson, A., Farlow, J., Larsson, P.r., Dukerich, M., Chambers, E., Byström, M., Fox, J., Chu, M., Forsman, M., and Sjöstedt, A. (2004). Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. Journal of bacteriology *186*, 5808-5818.

Johnson, M.T., and Munshi-South, J. (2017). Evolution of life in urban environments. Science *358*.

Kacsoh, B.Z., and Schlenke, T.A. (2012). High hemocyte load is associated with increased resistance against parasitoids in *Drosophila suzukii*, a relative of *D. melanogaster*. PloS one *7*, e34721.

Kalajdzic, P., and Schetelig, M.F. (2017). CRISPR/Cas-mediated gene editing using purified protein in *Drosophila suzukii*. Entomologia Experimentalis et Applicata *164*, 350-362.

Kalinowski, S. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity *94*, 33-36.

Kalinowski, S.T., Taper, M.L., and Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol *16*, 1099-1106.

Kanzawa, T. (1934). Research into the fruit-fly *Drosophila suzukii* Matsumura, Vol 48 (Yamanashi Prefecture Agricultural Experiment Station Report).

Kanzawa, T. (1939). Studies on *Drosophila suzukii* mats (Kofu: Yamanashi Prefecture Agricultural Experiment Station).

Keane, R.M., and Crawley, M.J. (2002). Exotic plant invasions and the enemy release hypothesis. Trends in ecology & evolution *17*, 164-170.

Keller, R.P., Drake, J.M., and Lodge, D.M. (2007). Fecundity as a basis for risk assessment of nonindigenous freshwater molluscs. Conservation Biology *21*, 191-200.

Keller, R.P., Geist, J., Jeschke, J.M., and Kühn, I. (2011). Invasive species in Europe: ecology, status, and policy. Environmental Sciences Europe *23*, 1-17.

Kettunen, M., Genovesi, P., Gollasch, S., Pagad, S., Starfinger, U., ten Brink, P., and Shine, C. (2009). Technical support to EU strategy on invasive alien species (IAS). Institute for European Environmental Policy (IEEP), Brussels *44*.

Kim, S.S., Tripodi, A., Johnson, D., and Szalanski, A. (2014). Molecular diagnostics of *Drosophila suzukii* (Diptera: Drosophilidae) using PCR-RFLP. Journal of Economic Entomology *107*, 1292-1294.

Kimura, M., and Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. Genetics *49*, 725.

Kimura, M., and Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences *75*, 2868-2872.

Kinjo, H., Kunimi, Y., and Nakai, M. (2014). Effects of temperature on the reproduction and development of *Drosophila suzukii* (Diptera: Drosophilidae). Applied Entomology and Zoology *49*, 297-304.

Knapp, L., Mazzi, D., and Finger, R. (2019). Management strategies against *Drosophila suzukii*: insights into Swiss grape growers choices. Pest management science *75*, 2820-2829.

Knapp, L., Mazzi, D., and Finger, R. (2020). The economic impact of *Drosophila suzukii*: perceived costs and revenue losses of Swiss cherry, plum and grape growers. Pest Management Science *77*, 978-1000.

Knudsen, K.E., Reid, W.R., Barbour, T.M., Bowes, L.M., Duncan, J., Philpott, E., Potter, S., and Scott, M.J. (2020). Genetic variation and potential for resistance development to the tTA overexpression lethal system in insects. G3: Genes, Genomes, Genetics *10*, 1271-1281.

Kolar, C.S., and Lodge, D.M. (2001). Progress in invasion biology: predicting invaders. Trends in ecology & evolution *16*, 199-204.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources *15*, 1179-1191.

Kopp, A., and True, J.R. (2002). Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Systematic biology *51*, 786-805.

Kovach, J. (2004). Impact of multicolored Asian lady beetle as a pest of fruit and people. American Entomologist *50*, 159-161.

Krafsur, E. (1998). Sterile insect technique for suppressing and eradicating insect population: 55 years and counting. Journal of Agricultural Entomology *15*, 303.

Kuesel, R., Scott Hicks, D., Archer, K., Sciligo, A., Bessin, R., and Gonthier, D. (2019). Effects of fine-mesh exclusion netting on pests of blackberry. Insects *10*, 249.

Kumar, M., Chaudhary, V., Sharma, R., Sirohi, U., and Singh, J. (2018). Advances in biochemical and molecular marker techniques and their applications in genetic studies of orchid: A review. Int J Chem Stud *6*, 806-822.

Kumschick, S., Blackburn, T.M., and Richardson, D.M. (2016). Managing alien bird species: Time to move beyond" 100 of the worst" lists? Bird Conservation International *26*, 154-163.

Kyritsis, G.A., Augustinos, A.A., Ntougias, S., Papadopoulos, N.T., Bourtzis, K., and Cáceres, C. (2019). Enterobacter sp. AA26 gut symbiont as a protein source for Mediterranean fruit fly mass-rearing and sterile insect technique applications. BMC microbiology *19*, 1-15.

Lai, Y., Shinde, D., Arnheim, N., and Sun, F. (2003). The mutation process of microsatellites during the polymerase chain reaction. Journal of Computational Biology *10*, 143-155.

Landguth, E.L., Fedy, B.C., Oyler-McCance, S.J., Garey, A.L., Emel, S.L., Mumma, M., Wagner, H.H., Fortin, M.J., and Cushman, S.A. (2012). Effects of sample size, number of markers, and allelic richness on the detection of spatial genetic pattern. Molecular ecology resources *12*, 276-284.

Lanzavecchia, S.B., Juri, M., Bonomi, A., Gomulski, L., Scannapieco, A.C., Segura, D.F., Malacrida, A., Cladera, J.L., and Gasperi, G. (2014). Microsatellite markers from the 'South American fruit fly' *Anastrepha fraterculus*: a valuable tool for population genetic analysis and SIT applications. Bmc Genet *15*, 1-8.

Latter, B. (1972). Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. Genetics *70*, 475-490.

Lavrinienko, A., Kesäniemi, J., Watts, P.C., Serga, S., Pascual, M., Mestres, F., and Kozeretska, I. (2017). First record of the invasive pest *Drosophila suzukii* in Ukraine indicates multiple sources of invasion. Journal of pest science *90*, 421-429.

Leach, H., Moses, J., Hanson, E., Fanning, P., and Isaacs, R. (2018). Rapid harvest schedules and fruit removal as non-chemical approaches for managing spotted wing Drosophila. Journal of pest science *91*, 219-226.

Leach, H., Van Timmeren, S., and Isaacs, R. (2016). Exclusion netting delays and reduces *Drosophila suzukii* (Diptera: Drosophilidae) infestation in raspberries. Journal of economic entomology *109*, 2151-2158.

Lee, J.C., Bruck, D.J., Curry, H., Edwards, D., Haviland, D.R., Van Steenwyk, R.A., and Yorgey, B.M. (2011). The susceptibility of small fruits and cherries to the spotted-wing drosophila, *Drosophila suzukii*. Pest management science *67*, 1358-1367.

Lee, J.C., Burrack, H.J., Barrantes, L.D., Beers, E.H., Dreves, A.J., Hamby, K.A., Haviland, D.R., Isaacs, R., Richardson, T.A., and Shearer, P.W. (2012). Evaluation of monitoring traps for *Drosophila suzukii* (Diptera: Drosophilidae) in North America. Journal of economic entomology *105*, 1350-1357.

Lee, K.-Z., and Vilcinskas, A. (2017). Analysis of virus susceptibility in the invasive insect pest *Drosophila suzukii*. Journal of Invertebrate Pathology *148*, 138-141.

Legendre, M., Pochet, N., Pak, T., and Verstrepen, K.J. (2007). Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome research *17*, 1787-1796.

Legendre, P., and Legendre, L. (1983). Numerical ecology (Amsterdam: Elsevier).

Levene, H. (1949). On a matching problem arising in genetics. Annals of mathematical statistics *20*, 91-94.

Levin, S.A. (2001). Encyclopedia of biodiversity, Vol 5 (Academic Press).

Lewis, M., Koivunen, E., Swett, C., and Hamby, K. (2019). Associations between *Drosophila suzukii* (Diptera: Drosophilidae) and fungi in raspberries. Environmental entomology *48*, 68-79.

Lewis, M.T., and Hamby, K.A. (2019). Differential impacts of yeasts on feeding behavior and development in larval *Drosophila suzukii* (Diptera: Drosophilidae). Scientific reports *9*, 1-12.

Li, C.C. (1955). Population genetics (Chicago: University of Chicago Press).

Li, G., and Quiros, C.F. (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theoretical and applied genetics *103*, 455-461.

Li, H., Qu, W., Obrycki, J.J., Meng, L., Zhou, X., Chu, D., and Li, B. (2020). Optimizing Sample Size for Population Genomic Study in a Global Invasive Lady Beetle, *Harmonia Axyridis*. Insects *11*, 290.

Litt, M., and Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American journal of human genetics *44*, 397.

Little, C.M., Rizzato, A.R., Charbonneau, L., Chapman, T., and Hillier, N.K. (2019). Color preference of the spotted wing Drosophila, *Drosophila suzukii*. Scientific reports *9*, 1-12.

Liu, H., and Stiling, P. (2006). Testing the enemy release hypothesis: a review and meta-analysis. Biological invasions *8*, 1535-1545.

Lowe, W.H., Kovach, R.P., and Allendorf, F.W. (2017). Population genetics and demography unite ecology and evolution. Trends in Ecology & Evolution *32*, 141-152.

Lynch, M. (2010). Evolution of the mutation rate. Trends in Genetics *26*, 345-352.

Mainka, S.A., and Howard, G.W. (2010). Climate change and invasive species: double jeopardy. Integrative Zoology *5*, 102-111.

Majerus, M., Strawson, V., and Roy, H. (2006). The potential impacts of the arrival of the harlequin ladybird, *Harmonia axyridis* (Pallas)(Coleoptera: Coccinellidae), in Britain. Ecological Entomology *31*, 207-215.

Markert, J.A., Champlin, D.M., Gutjahr-Gobell, R., Grear, J.S., Kuhn, A., McGreevy, T.J., Roth, A., Bagley, M.J., and Nacci, D.E. (2010). Population genetic diversity and fitness in multiple environments. BMC evolutionary biology *10*, 1-13.

Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., and Lohmueller, K.E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proceedings of the National Academy of Sciences *113*, 152-157.

Masel, J. (2011). Genetic drift. Current Biology *21*, R837-R838.

Mayo, O. (2008). A century of Hardy–Weinberg equilibrium. Twin Research and Human Genetics *11*, 249-256.

Mazzi, D., Bravin, E., Meraner, M., Finger, R., and Kuske, S. (2017). Economic impact of the introduction and establishment of *Drosophila suzukii* on sweet cherry production in Switzerland. Insects *8*, 18.

McCrone, J.T., and Lauring, A.S. (2018). Genetic bottlenecks in intraspecies virus transmission. Current opinion in virology *28*, 20-25.

McDonnell, M.J., and Hahs, A.K. (2015). Adaptation and adaptedness of organisms to urban environments. Annual review of ecology, evolution, and systematics *46*.

McDonnell, M.J., and Pickett, S.T. (1990). Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology *71*, 1232-1237.

Meirmans, P.G. (2012). AMOVA-based clustering of population genetic data. Journal of Heredity *103*, 744-750.

Millennium ecosystem assessment, M. (2005). Ecosystems and human well-being, Vol 5 (Island press Washington, DC).

Mitsui, H., Van Achterberg, K., Nordlander, G., and Kimura, M.T. (2007). Geographical distributions and host associations of larval parasitoids of frugivorous Drosophilidae in Japan. Journal of Natural History *41*, 1731-1738.

Mohammadi, S.A., and Prasanna, B. (2003). Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop science *43*, 1235-1248.

Mooney, H.A., and Cleland, E.E. (2001). The evolutionary impact of invasive species. Proceedings of the National Academy of Sciences *98*, 5446-5451.

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

Moran, E.V., Reid, A., and Levine, J.M. (2017). Population genetics and adaptation to climate along elevation gradients in invasive *Solidago canadensis*. PloS one *12*, e0185539.

Morand, M.E., Brachet, S., Rossignol, P., Dufour, J., and Frascaria-Lacoste, N. (2002). A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. Mol Ecol *11*, 377-385.

Moriyama, E.N., Petrov, D.A., and Hartl, D.L. (1998). Genome size and intron size in Drosophila. Molecular biology and evolution *15*, 770-773.

Nagylaki, T. (1998). Fixation indices in subdivided populations. Genetics *148*, 1325-1332.

Naranjo-Lázaro, J.M., Mellín-Rosas, M.A., González-Padilla, V.D., Sánchez-González, J.A., Moreno-Carrillo, G., and Arredondo-Bernal, H.C. (2014). Susceptibility of *Drosophila suzukii* Matsumura (Diptera: Drosophilidae) to entomophatogenic fungi. Southwestern Entomologist *39*, 201-203.

Nazareno, A.G., Bemmels, J.B., Dick, C.W., and Lohmann, L.G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources *17*, 1136-1147.

Nazareno, A.G., and Jump, A.S. (2012). Species–genetic diversity correlations in habitat fragmentation can be biased by small sample sizes. Mol Ecol *21*, 2847-2849.

Ndakidemi, B., Mtei, K., and Ndakidemi, P.A. (2016). Impacts of synthetic and botanical pesticides on beneficial insects. Agricultural Sciences *7*, 364.

Nei, M. (1972). Genetic distance between populations. The American Naturalist *106*, 283-292.

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. Journal of molecular evolution *19*, 153-170.

Nentwig, W. (2008). Biological invasions: why it matters. In Biological invasions (Berlin: Springer), pp. 1-6.

Nikolouli, K., Sassù, F., Ntougias, S., Stauffer, C., Cáceres, C., and Bourtzis, K. (2021). Enterobacter sp. AA26 as a Protein Source in the Larval Diet of *Drosophila suzukii*. Insects *12*, 923.

Oddou-Muratorio, S., Vendramin, G.G., Buiteveld, J., and Fady, B. (2009). Population estimators or progeny tests: what is the best method to assess null allele frequencies at SSR loci? Conservation Genetics *10*, 1343.

Ott, J. (1992). Strategies for characterizing highly polymorphic markers in human gene mapping. American journal of human genetics *51*, 283.

Ouborg, N., Van Treuren, R., and Van Damme, J. (1991). The significance of genetic erosion in the process of extinction. Oecologia *86*, 359-367.

Paetkau, D., Calvert, W., Stirling, I., and Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol *4*, 347-354.

Paetkau, D., Slade, R., Burden, M., and Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol *13*, 55-65.

Pannell, J.R., and Charlesworth, B. (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution *53*, 664-676.

Pannell, J.R., and Charlesworth, B. (2000). Effects of metapopulation processes on measures of genetic diversity. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences *355*, 1851-1864.

Parker, A., Mamai, W., and Maiga, H. (2021). Mass-rearing for the sterile insect technique. In Sterile Insect Technique (CRC Press), pp. 283-316.

Parreño, M.A., Scannapieco, A.C., Remis, M.I., Juri, M., Vera, M.T., Segura, D.F., Cladera, J.L., and Lanzavecchia, S.B. (2014). Dynamics of genetic variability in *Anastrepha fraterculus* (Diptera: Tephritidae) during adaptation to laboratory rearing conditions. Bmc Genet *15*, 1-8.

Peakall, R., and Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28: 2537e2539.

Peres-Neto, P.R., Jackson, D.A., and Somers, K.M. (2003). Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. Ecology *84*, 2347-2363.

Petes, T.D., Greenwell, P.W., and Dominska, M. (1997). Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. Genetics *146*, 491-498.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., and Estoup, A. (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. Journal of heredity *95*, 536-539.

Piry, S., Luikart, G., and Cornuet, J.-M. (1999). BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. Journal of Heredity *90*, 502-503.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945-959.

Proft, K.M., Bateman, B.L., Johnson, C.N., Jones, M.E., Pauza, M., and Burridge, C.P. (2021). The effects of weather variability on patterns of genetic diversity in Tasmanian bettongs. Mol Ecol *30*, 1777-1790.

Rannala, B., and Mountain, J.L. (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences *94*, 9197-9201.

Razifard, H., Ramos, A., Della Valle, A.L., Bodary, C., Goetz, E., Manser, E.J., Li, X., Zhang, L., Visa, S., and Tieman, D. (2020). Genomic evidence for complex domestication history of the cultivated tomato in Latin America. Molecular biology and evolution *37*, 1118-1132.

Rendon, D., Mermer, S., Brewer, L.J., Dalton, D.T., Da Silva, C.B., Lee, J.C.-T., Nieri, R., Park, K., Pfab, F., and Tait, G. (2019). Cultural Control Strategies to Manage Spotted-wing Drosophila (Oregon State University Extension Service).

Renkema, J.M., Telfer, Z., Gariepy, T., and Hallett, R.H. (2015). *Dalotia coriaria* as a predator of *Drosophila suzukii*: Functional responses, reduced fruit infestation and molecular diagnostics. Biological Control *89*, 1-10.

Richardson, B., Baverstock, P., and Adams, M. (1988). Allozyme electrophoresis: A handbook for animal systematics and population studies (Academic press).

Rodríguez-Ramilo, S.T., Toro, M.A., Wang, J., and Fernández, J. (2014). Improving the inference of population genetic structure in the presence of related individuals. Genetics research *96*.

Rodríguez-Ramilo, S.T., and Wang, J. (2012). The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. Molecular Ecology Resources *12*, 873-884.

Roques, A., Rabitsch, W., Rasplus, J.-Y., Lopez-Vaamonde, C., Nentwig, W., and Kenis, M. (2009). Alien terrestrial invertebrates of Europe. In Handbook of alien species in Europe (Springer), pp. 63-79.

Rota-Stabelli, O., Blaxter, M., and Anfora, G. (2013). *Drosophila suzukii*. Current Biology *23*, R8-R9.

Sæther, B.-E., Coulson, T., Grøtan, V., Engen, S., Altwegg, R., Armitage, K.B., Barbraud, C., Becker, P.H., Blumstein, D.T., and Dobson, F.S. (2013). How life history influences population dynamics in fluctuating environments. The American Naturalist *182*, 743-759.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution *4*, 406-425.

Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., and Kinzig, A. (2000). Global biodiversity scenarios for the year 2100. Science *287*, 1770-1774.

Santana Marques, P., Resende Manna, L., Clara Frauendorf, T., Zandonà, E., Mazzoni, R., and El-Sabaawi, R. (2020). Urbanization can increase the invasive potential of alien species. Journal of Animal Ecology *89*, 2345-2355.

Sassù, F., Nikolouli, K., Caravantes, S., Taret, G., Pereira, R., Vreysen, M.J., Stauffer, C., and Cáceres, C. (2019). Mass-Rearing of *Drosophila Suzukii* for sterile insect technique application: Evaluation of two oviposition systems. Insects *10*, 448.

Sassù, F., Nikolouli, K., Stauffer, C., Bourtzis, K., and Cáceres, C. (2020). Sterile Insect Technique and Incompatible Insect Technique for the Integrated *Drosophila suzukii* Management. Drosophila suzukii Management (ed Garcia, FRM), 169-194.

Schawaroch, V. (2002). Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). Biological Journal of the Linnean Society *76*, 21-37.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma *109*, 365-371.

Schlötterer, C. (2004). The evolution of molecular markers—just a matter of fashion? Nature reviews genetics *5*, 63-69.

Schwab, M. (2008). Encyclopedia of cancer (Springer Science & Business Media).

Seljak, G. (2011). Spotted wing drosophila-*Drosophila suzukii* (Matsumura). SAD, Revija za Sadjarstvo, Vinogradništvo in Vinarstvo *22*.

Shawer, R., Tonina, L., Tirello, P., Duso, C., and Mori, N. (2018). Laboratory and field trials to identify effective chemical control strategies for integrated management of *Drosophila suzukii* in European cherry orchards. Crop protection *103*, 73-80.

Shearer, P.W., West, J.D., Walton, V.M., Brown, P.H., Svetec, N., and Chiu, J.C. (2016). Seasonal cues induce phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. BMC ecology *16*, 1-18.

Shine, C., Kettunen, M., Genovesi, P., Essl, F., Gollasch, S., Rabitsch, W., Scalera, R., Starfinger, U., and ten Brink, P. (2010). Assessment to support continued development of the EU Strategy to combat invasive alien species. Final Report for EC, Institute for European Environmental Policy (IEEP), Brussels.

Shorthouse, D.P. (2010). SimpleMappr, an online tool to produce publication-quality point maps. Availabe online: http://www.simplemappr.net (accessed on Sep 17, 2020).

Shriver, M.D., Jin, L., Boerwinkle, E., Deka, R., Ferrell, R.E., and Chakraborty, R. (1995). A novel measure of genetic distance for highly polymorphic tandem repeat loci. Molecular biology and evolution *12*, 914-920.

Siozios, S., Cestaro, A., Kaur, R., Pertot, I., Rota-Stabelli, O., and Anfora, G. (2013). Draft genome sequence of the Wolbachia endosymbiont of *Drosophila suzukii*. Genome announcements *1*, e00032-00013.

Slarkin, M. (1985). Gene flow in natural populations. Annual review of ecology and systematics *16*, 393-430.

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

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. Science *236*, 787-792.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. Genetics *139*, 457-462.

Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics *9*, 477-485.

Sneath, P.H., and Sokal, R.R. (1973). Numerical taxonomy. The principles and practice of numerical classification (San Francisco).

Sousa, R., Pilotto, F., and Aldridge, D.C. (2011). Fouling of European freshwater bivalves (Unionidae) by the invasive zebra mussel (*Dreissena polymorpha*). Freshwater Biology *56*, 867-876.

Spitaler, U., Bianchi, F., Eisenstecken, D., Castellan, I., Angeli, S., Dordevic, N., Robatscher, P., Vogel, R.F., Koschier, E.H., and Schmidt, S. (2020). Yeast species affects feeding and fitness of *Drosophila suzukii* adults. Journal of Pest Science *93*, 1295-1309.

Stacconi, M.V.R., Buffington, M., Daane, K.M., Dalton, D.T., Grassi, A., Kaçar, G., Miller, B., Miller, J.C., Baser, N., and Ioriatti, C. (2015). Host stage preference, efficacy and fecundity of parasitoids attacking *Drosophila suzukii* in newly invaded areas. Biological Control *84*, 28-35.

Steiner, L., Mitchell, W., Harris, E., Kozuma, T., and Fujimoto, M. (1965). Oriental fruit fly eradication by male annihilation. Journal of Economic Entomology *58*, 961-964.

Steiner, L.F., and Lee, R. (1955). Large-area tests of a male-annihilation method for oriental fruit fly control. Journal of economic Entomology *48*, 311-317.

Stephens, A., Asplen, M., Hutchison, W.D., and Venette, R.C. (2015). Cold hardiness of winter-acclimated *Drosophila suzukii* (Diptera: Drosophilidae) adults. Environmental entomology *44*, 1619-1626.

Stockton, D.G., Wallingford, A.K., Brind'amore, G., Diepenbrock, L., Burrack, H., Leach, H., Isaacs, R., Iglesias, L.E., Liburd, O., and Drummond, F. (2020). Seasonal polyphenism of spotted-wing Drosophila is affected by variation in local abiotic conditions within its invaded range, likely influencing survival and regional population dynamics. Ecology and evolution *10*, 7669-7685.

Stojnić, S., V Avramidou, E., Fussi, B., Westergren, M., Orlović, S., Matović, B., Trudić, B., Kraigher, H., A Aravanopoulos, F., and Konnert, M. (2019). Assessment of genetic diversity and population genetic structure of Norway spruce (*Picea abies* (L.) Karsten) at Its southern lineage in Europe. Implications for conservation of forest genetic resources. Forests *10*, 258.

Strayer, D.L., Caraco, N.F., Cole, J.J., Findlay, S., and Pace, M.L. (1999). Transformation of freshwater ecosystems by bivalves: a case study of zebra mussels in the Hudson River. BioScience *49*, 19-27.

Sunnucks, P. (2000). Efficient genetic markers for population biology. Trends in ecology & evolution *15*, 199-203.

Syvänen, A.-C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics *2*, 930-942.

Tait, G., Grassi, A., Pfab, F., Crava, C.M., Dalton, D.T., Magarey, R., Ometto, L., Vezzulli, S., Rossi-Stacconi, M.V., and Gottardello, A. (2018). Large-scale spatial dynamics of *Drosophila suzukii* in Trentino, Italy. Journal of Pest Science *91*, 1213-1224.

Tait, G., Vezzulli, S., Sassù, F., Antonini, G., Biondi, A., Baser, N., Sollai, G., Cini, A., Tonina, L., and Ometto, L. (2017). Genetic variability in Italian populations of *Drosophila suzukii*. Bmc Genet *18*, 1-10.

Takahara, B., and Takahashi, K.H. (2017). Associative learning of color and firmness of oviposition substrates in *Drosophila suzukii*. Entomologia Experimentalis et Applicata *162*, 13-18.

Takezaki, N., Nei, M., and Tamura, K. (2014). POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. Molecular Biology and Evolution *31*, 1622-1624.

Tanksley, S.D. (1983). Molecular markers in plant breeding. Plant Molecular Biology Reporter *1*, 3-8.

Tochen, S., Dalton, D.T., Wiman, N., Hamm, C., Shearer, P.W., and Walton, V.M. (2014). Temperature-related development and population parameters for *Drosophila suzukii* (Diptera: Drosophilidae) on cherry and blueberry. Environmental entomology *43*, 501-510.

Unckless, R.L. (2011). A DNA virus of Drosophila. PloS one *6*, e26564.

Valdes, A.M., Slatkin, M., and Freimer, N.B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics *133*, 737-749.

Van Oosterhout, C., Weetman, D., and Hutchinson, W. (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Mol Ecol Notes *6*, 255-256.

Van Timmeren, S., and Isaacs, R. (2013). Control of spotted wing drosophila, *Drosophila suzukii*, by specific insecticides and by conventional and organic crop protection programs. Crop Protection *54*, 126-133.

Vargas, R.I., Leblanc, L., Pinero, J.C., and Hoffman, K.M. (2014). Male annihilation, past, present, and future. In Trapping and the detection, control, and regulation of tephritid fruit flies (Springer), pp. 493-511.

Vargas, R.I., Miller, N.W., and Stark, J.D. (2003). Field trials of spinosad as a replacement for naled, DDVP, and malathion in methyl eugenol and cue-lure bucket traps to attract and kill male oriental fruit flies and melon flies (Diptera: Tephritidae) in Hawaii. Journal of economic entomology *96*, 1780-1785.

Verity, R., and Nichols, R.A. (2016). Estimating the number of subpopulations (K) in structured populations. Genetics *203*, 1827-1839.

Vieira, M.L.C., Santini, L., Diniz, A.L., and Munhoz, C.d.F. (2016). Microsatellite markers: what they mean and why they are so useful. Genetics and molecular biology *39*, 312-328.

Vlach, J. (2010). Identifying *Drosophila suzukii*. Oregon Department of Agriculture, Salem, OR.

Vogt, H., Baufeld, P., Gross, J., Köppler, K., and Hoffmann, C. (2012). *Drosophila suzukii*: a new threat feature for the European fruit and viticulture-report for the international conference in Trient, 2, December 2011. Journal für Kulturpflanzen *64*, 68-72.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T.v.d., Hornes, M., Friters, A., Pot, J., Paleman, J., and Kuiper, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic acids research *23*, 4407-4414.

Walsh, D.B., Bolda, M.P., Goodhue, R.E., Dreves, A.J., Lee, J., Bruck, D.J., Walton, V.M., O'Neal, S.D., and Zalom, F.G. (2011). *Drosophila suzukii* (Diptera: Drosophilidae): invasive pest of ripening soft fruit expanding its geographic range and damage potential. Journal of Integrated Pest Management *2*, G1-G7.

Wang, J. (2018). Effects of sampling close relatives on some elementary population genetics analyses. Molecular Ecology Resources *18*, 41-54.

Waples, R.S. (2018). Null Alleles and F_{IS} x F_{ST} Correlations. Journal of Heredity *109*, 457-461.

Waples, R.S., and Anderson, E.C. (2017). Purging putative siblings from population genetic data sets: a cautionary view. Mol Ecol *26*, 1211-1224.

Weaver, J., Conway, T., and Fortin, M. (2011). Invasive Species Undeterred by Increasing Urbanization and Climate Change. Paper presented at: AGU Fall Meeting Abstracts.

Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. evolution, 1358-1370.

Weir, B.S., and Hill, W.G. (2002). Estimating F-statistics. Annual review of genetics *36*, 721-750.

Weissensteiner, M.H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V., Pophaly, S.D., Sedlazeck, F.J., and Suh, A. (2020). Discovery and population genomics of structural variation in a songbird genus. Nature communications *11*, 1-11.

Westphal, M.I., Browne, M., MacKinnon, K., and Noble, I. (2008). The link between international trade and the global distribution of invasive alien species. Biological invasions *10*, 391-398.

Whitlock, M.C. (2000). Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. Evolution *54*, 1855-1861.

Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A., and Losos, E. (1998). Quantifying threats to imperiled species in the United States. BioScience *48*, 607-615.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research *18*, 6531-6535.

Williams, S.L. (2001). Reduced genetic diversity in eelgrass transplantations affects both population growth and individual fitness. Ecological Applications *11*, 1472-1488.

Willing, E.-M., Dreyer, C., and Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. PloS one *7*, e42649.

Wondji, C., Simard, F., and Fontenille, D. (2002). Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. Insect molecular biology *11*, 11-19.

Wright, S. (1922). Coefficients of inbreeding and relationship. The American Naturalist *56*, 330-338.

Wright, S. (1951). The genetical structure of species. Ann Eugenics *15*, 323-354.

Wright, S. (1978). The relation of livestock breeding to theories of evolution. Journal of Animal Science *46*, 1192-1200.

Wyss, J.H. (2000). Screwworm eradication in the Americas. Annals of the New York Academy of Sciences *916*, 186-193.

Yan, Y., Jaffri, S.A., Schwirz, J., Stein, C., and Schetelig, M.F. (2020). Identification and characterization of four *Drosophila suzukii* cellularization genes and their promoters. Bmc Genet *21*, 1-10.

Yang, Y., Hou, Z.-C., Qian, Y.-H., Kang, H., and Zeng, Q.-T. (2012). Increasing the data size to accurately reconstruct the phylogenetic relationships between nine subgroups of the *Drosophila melanogaster* species group (Drosophilidae, Diptera). Molecular Phylogenetics and Evolution *62*, 214-223.

Yeung, K.Y., Fraley, C., Murua, A., Raftery, A.E., and Ruzzo, W.L. (2001). Model-based clustering and data transformations for gene expression data. Bioinformatics *17*, 977-987.

Zane, L., Bargelloni, L., and Patarnello, T. (2002). Strategies for microsatellite isolation: a review. Mol Ecol *11*, 1-16.

Zygouridis, N., Argov, Y., Nemny-Lavy, E., Augustinos, A., Nestel, D., and Mathiopoulos, K. (2014). Genetic changes during laboratory domestication of an olive fly SIT strain. Journal of applied entomology *138*, 423-432.

Index of Abbreviations

°C	degree celsius
%	percent
bp	base pair
cm	centimeter
D. melanogaster	*Drosophila melanogaster*
D. suzukii	*Drosophila suzukii*
ddH2O	double-distilled water
DNA	deoxyribonucleic acid
dNTP	desoxyribonucleosidtriphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
et al.	et alii/ aliae/ alia
EtOH	ethanol
fwd	forward
g	relative centrifugal force
h	hour
HWE	Hardy-Weinberg equilibrium
l	liter
LB-Medium	lysogeny broth medium
LD	linkage disequilibrium
M	Mol/l
mg	milligram
min	minute
ul	microliter
ml	milliliter
mM	millimolar
ng	nanogram
p	p-value
PCR	polymerase chain reaction
pH	potentia hydrogenii
pmol	picomole
rev	reverse

RT	room temperature
SSR	simple sequence repeats
SWD	Spotted Wing Drosophila
TAE	tris-acetate-EDTA
TB	terrific broth
TBE	tris-borate-EDTA
Tris	tris(hydroxymethyl)aminomethane
U	unit (enzyme)
VE	deionized water

Abbreviations and descriptions for population genetics and statistics

- SMM – Stepwise Mutation Model
- IAM – Infinite Allele Model
- $F_{ST} = H_T - H_S / H_T$; where H_S represents the expected level of heterozygosity in a subpopulation, and H_t is the expected level of heterozygosity if all subpopulations were pooled together (i.e. in the total population).
- d-F_{ST} - The average number of different alleles between individuals within a population
- $R_{ST} = S - S_w / S$; where S_w and S are the average sum of squares of the difference in allele size within a subpopulation and for the entire population, respectively
- d-R_{ST} - the average squared difference in allele size (measured as microsatellite product length) between individuals within a population
- M - the ratio of number of alleles (k) to range in allele size (r) for any given population
- m – absolute number of migrants

Moreover, standard SI units were used

A   Appendix

A.1   Additional Data

A.1.1 Publication 'Spatial and temporal genetic variation of *Drosophila suzukii* in Germany'

ORIGINAL PAPER

Check for updates

Spatial and temporal genetic variation of *Drosophila suzukii* in Germany

Sarah Petermann[1] · Sabine Otto[1] · Gerrit Eichner[2] · Marc F. Schetelig[1]

Abstract

Native to Southeast Asia, the spotted wing drosophila (SWD), *Drosophila suzukii* Matsumura, rapidly invaded America and Europe in the past 20 years. As a crop pest of soft-skinned fruits with a wide range of host plants, it threatens the fruit industry worldwide, causing enormous economic losses. To control this invasive pest species, an understanding of its population dynamics and structure is necessary. Here, we report the population genetics and development of SWD in Germany from 2017–19 using microsatellite markers over 11 different sample sites. It is the first study that examines SWD's genetic changes over 3 years compared to multiple international SWD laboratory strains. Results show that SWD populations in Germany are highly homogenous without differences between populations or years, which indicates that populations are well adapted, migrate freely, and multiple invasions from outside Germany either did not take place or are negligible. Such high genetic variability and migration between populations could allow for a fast establishment of the pest species. This is especially problematic with regard to the ongoing spread of this invasive species and could bear a potential for developing pesticide resistance, which could increase the impact of the SWD further in the future.

Keywords Invasive species · Microsatellite markers · Population structure · Gene flow · Migration

Key message

- *Drosophila suzukii* (Spotted Wing Drosophila) is established and migrates freely in Germany.
- Homogenous populations across Germany were found over the 3-year sampling period.
- Multiple reinvasions of *Drosophila suzukii* either do not take place or are negligible.

✉ Sabine Otto
   sabine.otto@agrar.uni-giessen.de

✉ Marc F. Schetelig
   marc.schetelig@agrar.uni-giessen.de

[1]   Department of Insect Biotechnology in Plant Protection,
      Institute for Insect Biotechnology, Justus-Liebig-University
      Giessen, Winchester Str. 2, 35394 Giessen, Germany

[2]   Mathematical Institute, Justus-Liebig-University Giessen,
      Arndtstrasse 2, 35392 Giessen, Germany

Introduction

The spotted wing drosophila (SWD), *Drosophila suzukii* Matsumura, became a severe invasive pest species in America and Europe. First descriptions of *D. suzukii* in Japan date back to 1931, recorded by Matsumura (Matsumura 1931). At this time, SWD was present throughout Japan, Korea, and China (Hauser 2011). Through its rapid spread across North America, Africa and Europe in the last few years, SWD has drawn much attention as a crop pest in affected countries (Hauser 2011; Calabria et al. 2012; Asplen et al. 2015; Boughdad et al. 2021). SWD spread all over the continent and belongs to the commonly found drosophilids in South and Central Europe (Cini et al. 2012; Calabria et al. 2012). The first specimens of SWD from Germany were caught in 2011 in Rhineland-Palatinate, Baden-Württemberg, and at Lake Constance (Bavaria and Baden-Württemberg) (Vogt et al. 2012).

In contrast to the vinegar fly *Drosophila melanogaster*, SWD is a crop pest for soft-skinned fruits like cherries, blueberries, blackberries, grapes, and strawberries. Female flies possess a serrated ovipositor with which they pierce the skin of ripening fruits. Hatching larvae feed on fruit pulp and

🖄 Springer

110

make the fruits unmarketable, causing substantial economic losses (Bolda et al. 2010; Mazzi et al. 2017). Bolda et al. (2010) calculated that the revenue losses in raspberries and blackberries in California totaled approximately $63.2 million in 2008 alone. Besides the larval damage, probing and rupturing of the fruit skin by the female ovipositor can entail several other problems like fungal infections that cause additional damage to the fruits (Rombaut et al. 2017; Ioriatti et al. 2018). An essential part of tackling an invasive pest like SWD is understanding its biology and answering, e.g., questions on the threat, SWD poses for individual fruit crops (Lee et al. 2011; Bellamy et al. 2013). Therefore, monitoring of fly populations is continuously conducted in Germany, and institutions provide growers with information on best practice control methods (Briem et al. 2015, 2018). Many approaches to develop SWD control have focused on live traps and pesticide applications so far. In contrast, other practices such as mass trapping and biological control methods using imported or native parasitoids could be central for future pest management strategies (Gabarra et al. 2015; Rossi Stacconi et al. 2015).

Efforts of pest control applications can be complemented and improved by understanding genetic development. Knowledge about invasion history and the genetic differences between populations can identify introduction pathways, improve testing, and help integrated pest management strategies in ultimately avoiding multiple (re)introductions (Estoup and Guillemaud 2010). Stabelli et al. (2020) reported that SWD populations from separate geographic areas exhibit different genotypic and phenotypic traits pointing at the importance of understanding intraspecific variability in pest species for pest control. The reconstruction of invasion pathways is particularly crucial to uncover and understand potential patterns in the spread of an invasive species. It is thought that pest organisms are distributed beyond their native ranges through trade by contaminations of traded goods or as stowaways (Chapman et al. 2017). Information on transport and introduction of non-native species is of great importance, because the management of invasive pest species is most effective during the beginning of an invasion, if, e.g., biosecurity measures would be adopted at transport routes (Rout et al. 2011; Chapman et al. 2016). In a recent study, it was shown that the latest invasion of SWD in Argentina could be ascribed to fruit trade from freshly invaded areas in North and South America and not from its native range in Asia (de la Vega et al. 2020). Data from Ukraine also implies that SWD invasion in Europe can be ascribed to multiple sources, along with possible recurrent introductions (Lavrinienko et al. 2017). The results from these two studies show how the identification of invasion routes is an important issue, especially from a pest management perspective, and they point out the problematic link between global trade and pest invasiveness.

The reconstruction of invasion routes for international trade can be challenging, though, considering the sheer amount of fruit transports through global trading. For example, Germany imported 51,776 tons of fresh fruits in the year 2017 alone (FAO 2021). Studies on population genetics have proven beneficial for biological control methods like the sterile insect technique (SIT) (Lanzavecchia et al. 2014). SIT is a biological method for pest control in which sterilized male insects are released in a field population to reduce reproduction by infertile mating. Since gene flow can vary between natural populations, locally adapted and isolated populations can occur. In these populations, SIT can be impaired by mating barriers, making them less effective. Population genetics can be used to trace changes in strain efficacy, which is an assertion of how well a strain would perform in the field. It can also help to improve the competitiveness of laboratory strains that are sterilized and released and can be used in monitoring programs to differentiate released laboratory insects from wild ones (Aketarawong et al. 2011, 2014; Zygouridis et al. 2014; Azrag et al. 2016).

One way to characterize genetic relationships between populations is by using microsatellite or simple sequence repeat (SSR) markers. Microsatellites are repetitive DNA sequences that have high mutation rates, which can lead to high polymorphism within populations and a rapid genetic differentiation between distinct populations (Schlötterer 2000; Selkoe and Toonen 2006). Variations in the number of repetitions generate different alleles. This long-established method in population genetics is used regularly due to its cost-effectiveness and informative content (Jarne and Lagoda 1996; Schlötterer 2004). For *D. suzukii*, a set of 28 microsatellite markers was previously developed and used to characterize genetic variation within Hawaiian and French populations (Fraimout et al. 2015). A subset of those markers was then utilized to get an insight into genetic variability in Italian populations of *D. suzukii* (Tait et al. 2017). Those populations showed extensive genetic similarity, except for a population from Sicily, confirming isolation relative to the mainland. Whereas the technique has been widely used to describe population variation and genetic differences in SWD, only few research has investigated differences that occur over consecutive years when this pest species begins to become established. Bahder et al. 2015 evaluated the genetic variation in SWD populations from California and Washington collected across three years and showed that the population from Washington had undergone a significant bottleneck compared to the coastal California population. These studies show that the use of microsatellite markers for population structure and genetic diversity analysis is an important part of understanding the biology of an invasive species. Because little is known about the genetic diversity of SWD populations in Germany, the goal of this study was to determine intrapopulation and interpopulation genetic

111

diversity between different populations across the whole country. Moreover, it was of interest, if diverse genetically defined populations exist and if they exhibit a geographical pattern. On the one hand, we expected some differences between northern and southern locations, since the North and South of Germany are two distinct major geographical regions with different geographic aspects and different climates with a more maritime climate in the North and an increasingly subcontinental climate toward the South. On the other hand, based on the results of Tait et al. (2017) from Italy, we expected low differences between sample sites, since the distances between locations are relatively low. An additional comparison of data over several years can provide information about population development and dynamics. In respect to pest control, it would be beneficial to know, if an annual reinvasion from other European countries or warmer regions in Germany takes place or if SWD overwinters locally and reemerges in spring. It is known that SWD overwinter as adults and can recover quickly (Dalton et al. 2011; Hamby et al. 2014; Stephens et al. 2015). Thus, we expected that a reemergence from locally overwintering individuals is more likely and, if differences between years occur, that they turn out to be small. Invasive species often show a high genetic divergence between their invasive and native populations, which allows them to adapt faster to new environments (Guo et al. 2018; Xia et al. 2020). High genetic diversity in Germany would implicate a well-established population, while reduced genetic diversity, for example through a bottleneck effect, in contrast, would correspond to reduced adaptability or at least to a recent invasion at that specific locality (Schrader et al. 2014). Therefore, we investigated D. suzukii populations for three years using 14 microsatellite markers designed by Fraimout et al. (2015). This is the first study to provide insights into population genetics of D. suzukii on a large-scale level in Germany as well as a comparison over a multi-year period. Besides, different laboratory strains from Europe and America were included in the comparison and a new laboratory strain that derived from one of the German field populations was established to illustrate the effect of laboratory breeding and change of genetic markers over time. As a result, this strain functioned as a link between field collections and the laboratory strains.

Material and methods

Drosophila suzukii collection and identification

Adult D. suzukii from 11 locations in Germany were collected between June and September in 2017, 2018, and 2019 by collecting fruit samples infested with D. suzukii eggs, larvae, or pupae. Location sites were mapped using geographic coordinates with Simplemappr (Shorthouse 2010), assuring that each year fruits from the same locations were collected (Fig. 1). If possible, different kind of fruits from various shrubs or trees were sampled within a radius of approximately 20 km at each location, to prevent sampling of related individuals. Blackberries showed the highest amount of

Location acronym	Location	Coordinates	Collection Month 2017	Collection Month 2018	Collection Month 2019
DO	Dortmund	51.300499 7.674342	August	July	August
FF	Frankfurt (Oder)	52.325507 14.497815		August	August
FR	Sasbach	48.678752 8.030775	August	August	September
HB	Bremen	52.861452 9.370046	September		
HG	Bad Homburg	50.229450 8.622310	August	August	August
HH	Hamburg	53.480480 10.222784			August
HOH	Hofheim	50.110740 10.511780	August	June	July
KS	Kassel	51.370993 9.416376	August	August	August
LB	Gerlingen	48.792326 9.05524	August	June	July
PM	Derwitz	52.397132 12.823140	September	September	August
R	Regensburg	49.077134 12.208671	September	August	August

Fig. 1 Sample sites and sampling time points of D. suzukii wild populations in Germany. Population location acronym, location, coordinates and sampling year and month are listed in the table (left). The map (right side) was generated by SimpleMappr and coordinates listed in the table

infestation followed by raspberries and cherries. Elderberries were collected only in a few cases since the degree of infestation was lower than in cherries. Strawberries were not infested. The number of sampled fruits varied between years and locations (Table 1). The German samples' names result from the vehicle registration plate of the respective district and the sampling year. Fruits were kept in the laboratory until adult flies emerged. Adult *D. suzukii* were identified according to Hauser (2011) and documented using a Keyence VHX-5000 (Keyence Corporation, Osaka, Japan). No samples were available from FF in 2017, HB in 2018 and 2019, and HH in 2017 and 2018 (Fig. 1, Table 1). From each geographical population and year, 20 individuals were tested, except for HB17, with only ten individuals available.

In addition, 160 specimens from eight laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_France, LS_Valsugana, LS_HG18, and LS_HG19) were used as outgroups. LS_USA was established in 2010 from a field collection in North Carolina (Stockton et al. 2020), LS_Canada was started in 2012 (Jakobs et al. 2015; Renkema et al. 2015), LS_Italy was kept in the laboratory since 2014 and LS_Frankfurt was established in 2016 (Lee and Vilcinskas 2017). LS_France was kindly provided by Eric Marois and originated from Strasbourg (France). Alberto

Grassi provided LS_Valsugana from Valsugana in Italy. Both were collected in 2018 and kept in the laboratory since. The laboratory strains LS_HG18 and LS_HG19 originated from flies collected from the location Bad Homburg (HG) in 2017 that were sampled for analysis after one (LS_HG18) and two years (LS_HG19) in culture, respectively. *D. suzukii* laboratory strains were maintained on standard *Drosophila* medium at 25 °C and 55% humidity with a 12 h-photoperiod and transferred to fresh media every week. A sample size estimation was conducted to determine the minimum number of observations required for our experiment. Under the assumption, that the population standard deviation of allele size is at most seven (based on preliminary exploratory data analyses) and under the requirement that a 90%-confidence interval (CI) for the population mean of allele size has a length of at most 10 with a probability of 80% ("precision power"), a sample size of at least 19 was necessary. Based on this sample size estimation, we chose to use 20 individuals from each location and year with the exception of HB17, where only 10 individuals were available. The final data set included 550 individuals from Germany and 160 individuals from laboratory strains (Table 1). German populations and laboratory outgroups were analyzed using 14 microsatellite markers.

Table 1 Number of SWD individuals sampled and used for analysis per year and population. Shown is the number of SWD individuals used for analysis across the years 2017 to 2019, including the laboratory strains

Population	Sampling areas	Total number of sampled SWD			Number of SWD used for analysis			
		2017	2018	2019	2017	2018	2019	Σ
DO	2	41	26	37	20	20	20	60
FF	3	0	46	37	0	20	20	40
FR	4	34	25	33	20	20	20	60
HB	1	10	0	0	10	0	0	10
HG	1	47	34	48	20	20	20	60
HH	2	0	0	39	0	0	20	20
HOH	3	43	33	35	20	20	20	60
KS	1	24	28	32	20	20	20	60
LB	4	39	57	41	20	20	20	60
PM	3	48	32	52	20	20	20	60
R	4	39	29	44	20	20	20	60
LS_France	0	0	20	0	0	20	0	20
LS_Valsugana	0	0	20	0	0	20	0	20
LS_Italy	0	20	0	0	20	0	0	20
LS_Frankfurt	0	20	0	0	20	0	0	20
LS_Canada	0	20	0	0	20	0	0	20
LS_USA	0	20	0	0	20	0	0	20
LS_HG18	0	0	20	0	0	20	0	20
LS_HG19	0	0	0	20	0	0	20	20
Total number of individuals used in the analysis								710

113

DNA extraction and microsatellite analysis

Total genomic DNA was extracted from 20 or 10 individuals of each geographical population and laboratory strain in single reactions (Table 1). Samples were placed in lysis tubes with 1.4 mm ceramic spheres (Lysing Matrix D bulk, MP Biomedicals, Solon, OH), 200 µl sterile filtered homogenization buffer added (1 M Tris–HCl (pH 7.5), 5 M NaCl, 0.5 M EDTA (pH 8), 0.3 M spermine tetra-HCl, 1 M spermidine tri-HCl, 1 g sucrose), and then homogenized at 6000 rpm for 40 s in a Fast Prep-24™ (MP Biomedicals, Solon, OH). Afterward, 200 µl lysis buffer (1 M Tris–HCl (pH 9.0), 0.5 M EDTA (pH 8.0), 10% SDS, 1 g sucrose) was added and incubated at 70 °C for 10 min. 60 µl of 8 M KOAc was added, and tubes were stored on ice for 30 min. After transferring the suspension to a new tube, samples were centrifuged at 12,000 rpm for 10 min at 4 °C. DNA was precipitated with two volumes of ice-cold 100% EtOH and stored at -20 °C overnight. DNA was pelleted by centrifugation at 12,000 rpm for 40 min at 4 °C. The pellet was washed with 30 µl of 70% ice-cold ethanol for 10 min while centrifuging at 12,000 rpm at 4 °C. Pellets were air-dried and resuspended in 50 µl H_2O. DNA concentration was measured using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). Samples were stored at -20 °C.

SSR markers were previously characterized (Fraimout et al. 2015). Of the 28 published SSRs, 17 dinucleotide markers were selected for initial testing with genomic DNA from LS_USA. Each of the published primer pairs was tested beforehand in single PCR reactions using the Platinum Taq DNA polymerase (Thermo Fisher Scientific, MA, USA) following the manufacturer's protocol with a primer annealing temperature of 57 °C. PCR products were purified with Zymo Clean and Concentrator-5 spin columns (Zymo Research, Orange, CA). Purified PCR products were cloned into the pCR4-TOPO TA vector (Thermo Fisher Scientific, MA, USA) using standard protocols. The sequencing reaction was carried out by Macrogen (Amsterdam, Netherlands) and analyzed with Geneious Prime 2019.2 software. Three of 17 tested loci were not included in further experiments due to amplification problems and null alleles, leaving 14 markers for analysis (Online Resource Table S1).

DNA amplification for fragment analysis on the population samples was then performed using the Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany) in 10 µL final reaction volume, containing 1X QIAGEN Multiplex PCR Master Mix, 0.2 µM primer mix, 0.5X Q-Solution, and 100 ng of genomic DNA. The PCR cycling protocol was: 95 °C, 5 min; 32 cycles of 95 °C for 30 s, 57 °C for 90 s, 72 °C for 3 min; final elongation at 72 °C for 30 min, the latter is advised to be used for analysis on capillary sequencers. PCR products were purified using the Zymo Clean and Concentrator-5 Kit (Zymo Research, Orange, CA). Concentration was assessed using electrophoresis on 3% agarose gels, stained with SYBR Safe, and visualized under UV light. Samples were sent for fragment analysis on an ABI 3730 Genetic Analyzer to StarSEQ (Mainz, Germany). GeneScan™-500LIZ™ was used as an internal size standard. Fragments were sized with the Geneious Prime 2019.2 software. If no sample amplification was obtained after three attempts, the locus was classified as missing data.

Statistics on genetic diversity

Population genetic indexes like the number of alleles (N_a), the effective number of alleles (N_e), the inbreeding coefficient (F_{IS}) and the F_{ST} index (where I stands for individuals, S for subpopulations and T for the total population), number of private alleles (A_P), observed and expected heterozygosity and deviations from Hardy–Weinberg-Equilibrium were calculated with GenAlex software v.6.41 (Peakall and Smouse 2012). Polymorphism information content (PIC) was calculated using the Cervus 3.0 software (Kalinowski et al. 2007). Measures of diversity were analyzed using repeated-measures analysis of variance (ANOVA) to determine differences among populations and years. Pairwise comparisons were obtained from posthoc Bonferroni correction. All standard statistical tests were carried out using SigmaPlot 14 software (Systat Software, Inc.).

Statistics—genetic structure

In general, the genetic structure refers to patterns of genetic diversity across multiple populations and subpopulations and provides information on distribution, mating behavior, and potential species and population borders. The genetic structure of D. suzukii populations in Germany was analyzed using a Bayesian approach, a population tree based on allele frequency data, analysis of molecular variance (AMOVA), and Principal coordinate analysis (PCoA). Structure 2.3.4 (Pritchard et al. 2000; Falush et al. 2003) was used to investigate the number of genetically distinct clusters (K) in a data set. Each analysis was run with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions and a burn-in period of 100,000 repetitions using 20 iterations of K = 1–20. The analysis was run using the admixture ancestry model, based on correlated allele frequencies. Structure Harvester (Web v0.6.94 July 2014) (Earl and von Holdt 2012) was used to detect the most likely K value according to Evanno and Pritchard, assessed through analysis of ΔK, the Dirichlet parameter alpha (α) and LnP(D) distribution plots (Evanno et al. 2005; Hubisz et al. 2009). Pophelper (Francis 2017) was used to align assignment clusters across replicate runs and visualize the results.

An unrooted neighbor-joining (NJ) dendrogram was constructed with PoptreeW (Takezaki et al. 2014) based on Da

distance (Nei et al. 1983), which is a genetic dissimilarity coefficient that is based on mutation and drift. It is defined as $D_A = 1 - \frac{1}{r}\sum_j^r\sum_j^{m_j}\sqrt{x_{ij}y_{ij}}$, where r is the number of loci used, m_j is the number of alleles at the j-th locus, and x_{ij} and y_{ij} are the frequencies of the i-th allele at the j-th locus in populations X and Y (Nei et al. 1983; Takezaki et al. 2014). A test for robustness was carried out by using a bootstrap value of 10,000. An analysis of molecular variance (AMOVA) was performed using Arlequin v.3.5.2.2 (Excoffier and Lischer 2010). PCoA implemented in GenAIex software v.6.41 (Peakall and Smouse 2012) based on Nei's genetic distance was used to visualize the genetic relationship between populations.

Geneclass2 (Piry et al. 2004) was run to determine the probability of each individual to originate from the sample area or another reference population. The standard Bayesian criterion of Rannala and Mountain (1997) and the Monte Carlo resampling method of Paetkau et al. (2004) were used with an alpha value of 0.05. Results were based on 10,000 simulated genotypes for each population and a threshold probability value of 0.05. Tables with migration rate values are shown in Online Resource S8.

Bottleneck v.1.2.2 (Piry and Luikart 1999) was used to determine if recent demographic events like expansion or mitigation in population size took place. The two-phase model (TPM) and the stricter stepwise mutation model (SMM) were used with model options for the TPM of 80% single-step mutations, a variance among multiple steps of 12 and with 5,000 iterations. Wilcoxon's signed-rank test was used to assess the probability of heterozygosity.

Results

Polymorphisms and genetic variability at 14 selected microsatellite loci in *Drosophila suzukii*

In the first step, we evaluated the genetic variability and the information content of the 14 microsatellite markers used in this study. In total, 115, 107, and 112 alleles were identified in 2017, 2018, and 2019, respectively, and all three years shared 81.82% of the identified alleles. The detailed results for variability indices in the 14 SSR markers are shown in Online Resource S2. The analyzed loci showed a consistently high level of genetic variability throughout the years and populations. Deviation from HWE was tested for all year-locus combinations. A significant difference ($p < 0.05$) from HWE was observed in 24 of 42 year-locus combinations (Online Resource S3). A reason for the disequilibrium could be the presence of null alleles. Per definition, a microsatellite null allele is an allele at a microsatellite locus that does not amplify to detectable levels in a PCR test. The

result is an excess of homozygotes or to be precise a deficiency of observed heterozygotes in a population. Other reasons for a deviation from HWE can be migration, selection, phenotypic assortative mating, inbreeding, genetic drift and small population size.

Polymorphism information content (PIC) was obtained as index for gene abundance. The level of diversity reflects genetic variation in loci (PIC > 0.5 = high polymorphism, 0.5 > PIC > 0.25 = moderate polymorphism, PIC < 0.25 = low polymorphism). PIC across loci ranged from 0.474 (DS11) to 0.836 (DS07). All loci except for DS11 showed high polymorphism (Online Resource S2), confirming that this set of markers is suitable for a population genetic study.

Genetic and allelic diversity among populations and years

Genetic and allelic diversity in German populations was similar and relatively high in all years and over all locations. There was no significant difference between years or sample sites by ANOVA after correction according to the Bonferroni method ($p > 0.05$) (Fig. 2 and Online Resource S4). The mean number of observed alleles (N_a) over the loci for each year reached from 6.23 in 2019 to 5.93 in 2017 and was overall similar (ANOVA for years: $p = 0.147$, for populations: $p = 0.414$). The mean number of effective alleles (N_e, the number of equally frequent alleles it would take to achieve a given level of gene diversity) was similar in the years 2017, 2018, and 2019, respectively, supporting our findings that there is no significant difference in genetic diversity between years or sample sites (ANOVA for years: $p = 0.823$, for populations: $p = 0.491$). Observed heterozygosity (H_o, the observed ratio of heterozygotes) ranged between 0.57 (LB) and 0.71 (KS) in 2019 to values in 2018 between 0.59 in FF to 0.67 in HOH (ANOVA for years: $p = 0.313$, for populations: $p = 0.306$). Values for expected heterozygosity (H_e, the proportion of heterozygous genotypes expected under Hardy–Weinberg equilibrium) were similar in all three years with R showing the overall lowest values in 2017 and 2019 and KS having the highest values during these two years (ANOVA for years: $p = 0.566$, for populations: $p = 0.200$). The number of private alleles (an allele that is present in only one population but at any frequency) was low throughout the study period, suggesting that most alleles were shared between populations and years (Fig. 2 and Online Resource S5). The F_{IS} value or inbreeding coefficient of an individual (I) relative to the subpopulation (S) showed a range of -0.07 (R17) to 0.18 (FF18 and LB19) (ANOVA for years: $p = 0.203$, for populations: $p = 0.134$). Negative F_{IS} values would reveal a heterozygote excess, while positive values indicate a heterozygote deficit. It was positive for 25 of the German populations, except for R in 2017 (0.07) and 2019 (0.03) as well as for FF in 2019
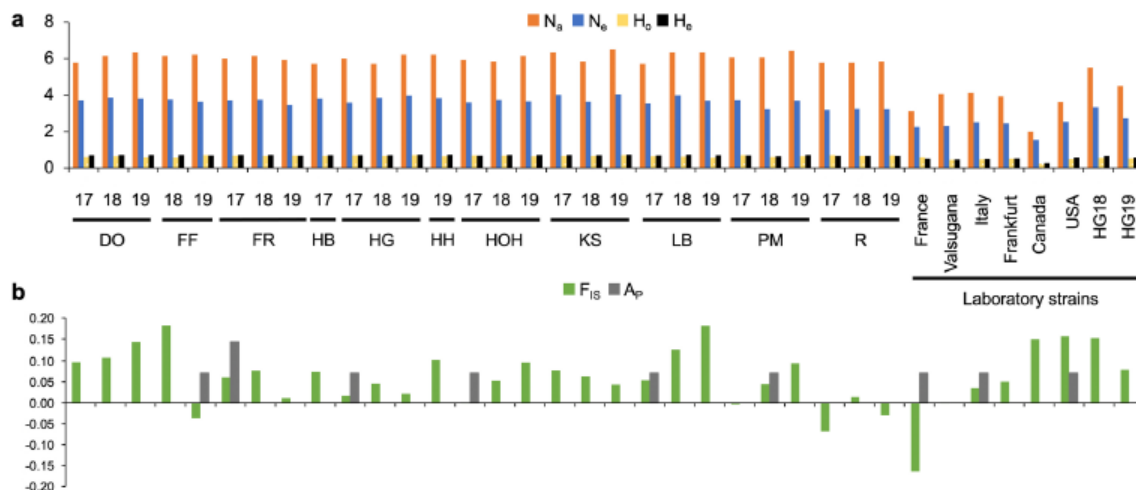
Fig. 2 Level of genetic diversity across studied populations over three years. **a** N_a = No. of different alleles; N_e = Number of effective alleles; H_o = observed heterozygosity; H_e = expected heterozygosity. **b** A_P = No. of private alleles unique to a single population, F_{IS} = mean inbreeding coefficient

(0.04), all three showing a heterozygosity excess, which can be a consequence of a genetic bottleneck. In 2018, all populations had a positive F_{IS} value (Fig. 2 and Online Resource S4). Positive values can indicate inbreeding and heterozygote deficiency or to be precise an excess of homozygotes due to the presence of null alleles, inbreeding or population subdivision.

Genetic and allelic diversity among laboratory strains

Since the eight laboratory strains were thought to be highly inbred, estimates for genetic and allelic diversity were calculated separately from the German populations to not bias the results. Nevertheless, using laboratory strains as outgroups can provide valuable information, especially for testing the marker system and the effect of artificial breeding in the laboratory. The overall genetic and allelic diversity in the tested laboratory strains was lower than in the field collection from Germany with 42% less different alleles, 38% less effective alleles, 29% less observed heterozygosity and 31% less expected heterozygosity (Fig. 2 and Online Resource S4) making the laboratory strains less diverse than the German field collections. The most diverse laboratory strain was LS_HG18, and the least diverse strain was LS_Canada. These results reflect well the time the strains were kept in the laboratory with LS_HG18 being the newest strain (one year in culture) and LS_Canada being one of the longer established ones (approx. eight years in culture). The F_{IS} value was positive for most strains, which could indicate inbreeding, except for LS_France and LS_Valsugana with

negative values, which can be interpreted as an excess of heterozygotes.

Genetic distance and relationship among populations and years

Laboratory strains were excluded from the analysis of molecular variance since they were expected to be highly inbred. AMOVA indicated that 98% of variation occurred within populations, while only 2% appeared among populations in 2017 and 2019. In 2018, 3% of variation originated among populations and 97% from within populations (Table 2). This data suggests that the tested populations were similar to each other and do not show differences between sampling years.

To detect possible genetic relationships between field and laboratory populations, PCoA was performed (Fig. 3). Overall, the separation between German populations was not distinct. The tested laboratory strains were noticeably separated from the German field collections but also from LS_HG18. LS_HG19, on the other hand, those two did show more similarity to the laboratory strains than to the field collections, confirming the impact of laboratory cultivation over time. The laboratory strains showed that the used marker system in our experiment was able to discriminate between populations but that there were no apparent dissimilarities between German sample sites or years. This data suggests that tested German populations were similar over all three years and sample sites, which is in agreement with the results obtained with AMOVA.

Table 2 Summary of AMOVA result for German populations across three years

Year	Source of variation	d.f	SS	VC	% PV
2017	Among populations	8	65.41	0.08689	1.74
	Within populations	331	1623.60	4.9051	98.26
	Total	339	1689.01	4.9920	100
2018	Among populations	8	94.39	0.1711	3.34
	Within populations	351	1738.67	4.9535	96.66
	Total	359	1833.06	5.1246	100
2019	Among populations	9	81.07	0.09891	1.92
	Within populations	390	1970.07	5.05147	98.08
	Total	399	2051.14	5.15038	100

d.f. = degree of freedom; SS = sum of squares; VC = variance components; % PV = percentage of total variation

F_{ST} was used as a measure of genetic differentiation between populations. It compares the proportion of the genetic variation contained within a population relative to the total genetic variance and is derived from the variances of allele frequencies. To provide a less biased estimate of gene flow when sample sizes are moderate, F_{ST} was estimated instead of R_{ST} (Gaggiotti et al. 1999), which is analogous to F_{ST}, but is based on the stepwise mutation model (SMM) and can be estimated from the differences of allele sizes. The results revealed a minor differentiation ($F_{ST} < 0.05$) between all populations and years (Fig. 4 and Online Resource S7). In comparison, in experiments including the long-term laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_France and LS_Valsugana) 54.64% of all F_{ST} values showed a moderate (0.1–0.25), 43.17% a minor ($F_{ST} < 0.1$), and 2.19% a strong differentiation ($F_{ST} > 0.25$) (Fig. 4 and Online Resource S7). The pairwise F_{ST} for laboratory strain LS_HG revealed that more differentiation was detected in the second year of inbreeding (2019) than in 2018. While the differentiation was moderate in 2018, it was moderate to strong in 2019, which means that the two years of inbreeding resulted in more substantial differentiation compared to its origin (Fig. 4 and Online Resource S7).

Fig. 3 PCoA at population level for 2017–2019 generated from 14 microsatellite markers. **a** Samples of all populations in Germany from 2017 to 2019 (●) and laboratory outgroups (▲ 8- to 10-year-old laboratory strains, ■ laboratory strains that were established between 2014 and 2016, ✖ laboratory strain established from the wild population HG17, ✚ laboratory strains LS_France and LS_Valsugana which were established in the year 2018) **b** PCoA for German populations sampled in 2017 **c** The lower left shows the result for the year 2018 **d** PCoA for German populations sampled in 2019

117
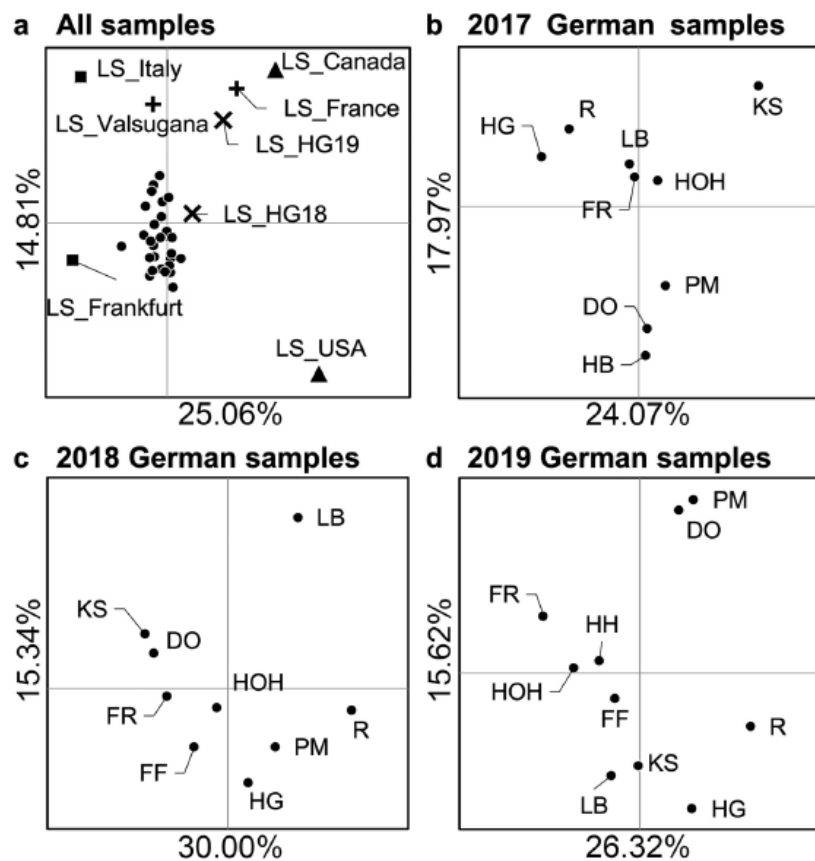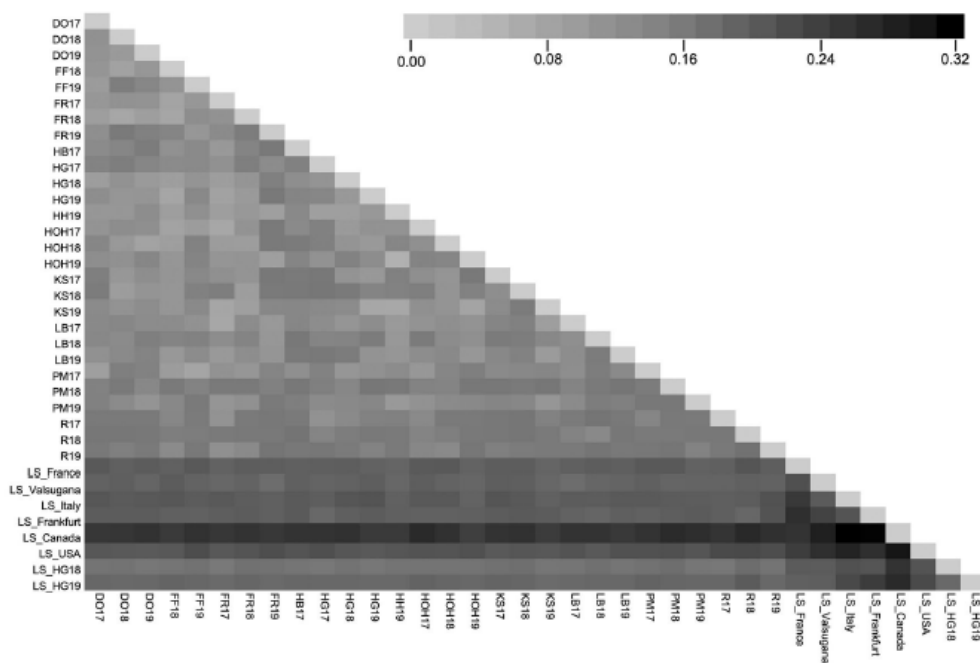
Fig. 4 Pairwise F_{ST} among studied populations over three years. Eight laboratory strains (LS_France, LS_Valsugana, LS_Italy, LS_Frankfurt, LS_USA, LS_Canada, LS_HG18, and LS_HG19) and 28 populations from Germany, collected over three years, are shown in pairwise comparisons. Values are color coded (light gray = F_{ST} < 0.1: minor differentiation; gray = 0.1 < F_{ST} < 0.25: moderate differentiation; black = F_{ST} > 0.25: strong differentiation)

To infer (genetic) population structure, a Bayesian model was used, which detects clusters of genetically similar individuals within subpopulations. First, we excluded all laboratory strains, but STRUCTURE was unable to identify population structure in the German data set. Adding laboratory strains LS_USA, LS_Canada, LS_Italy, LS_France, and LS_Valsugana for comparison revealed two lineages (K = 2) throughout all three years, and populations (Fig. 5). The two most reasonable values for K were K = 2 and K = 3, with the more likely one being K = 2, according to the analysis of ΔK. LS_HG18 and LS_HG19 were excluded from this analysis after we saw that including them changes the STRUCTURE outcome. For the analysis including LS_HG18 and LS_HG19, three lineages (K = 3) were identified, with K = 4 being also a possible K value but not the most likely one, according to the analysis of ΔK. In both cases, the two most likely K are displayed (Fig. 5) to overcome the problem of underestimating the "true" value of K (Janes et al. 2017). These results indicate that German populations are not structured, since only the additional usage of distinct laboratory strains resulted in the detection of population structure. This finding indicates that most likely a single genetic cluster explains the distribution of genetic variation in the sampled German populations.

A neighbor-joining tree based on Nei's genetic distance corresponds to findings in STRUCTURE and F_{ST} (Fig. 6). The German populations were not grouped in any way, and the differences to laboratory strains were visible. While LS_HG18 was already separated from the German field collections after one year in culture, the separation got even more prominent in LS_HG19 after two years in culture.

A bottleneck analysis was used to test the hypothesis of a recent expansion or bottleneck in each of the 28 populations from Germany, not including the laboratory strains since they were irrelevant for the detection of recent reduction or expansion of the population size (Online Resource S9). Under the stepwise mutation model (SMM), no bottleneck event was detected. In contrast, significant heterozygote deficit was identified in FF19 ($P = 0.021$), PM18 ($P = 0.010$), R17 ($P = 0.002$), R18 ($P = 0.002$) and R19 ($P = 0.021$) suggesting that a recent expansion took place. However, under the two-phase model (TPM model), none of the abovementioned expansion signals could be confirmed. Instead, a bottleneck event was more likely in HOH18 ($P = 0.029$), HG18 ($P = 0.045$) and DO18 ($P = 0.045$).
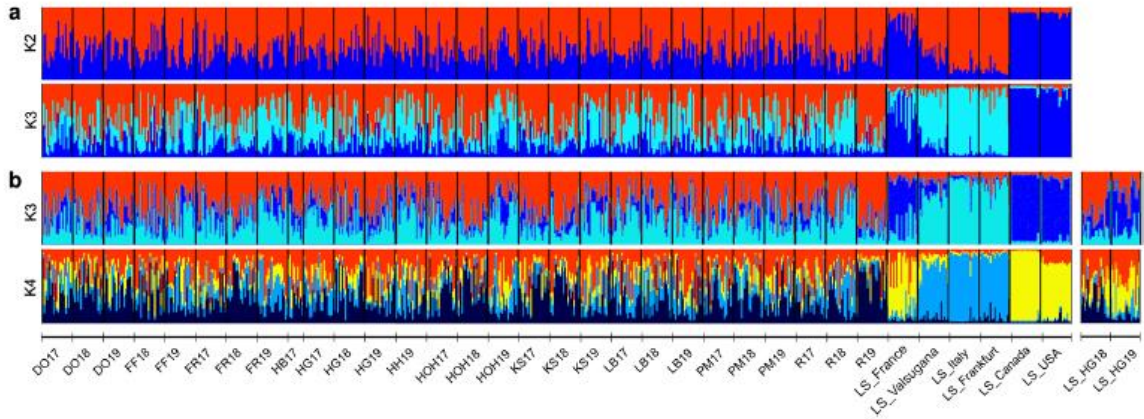
Fig. 5 Structural analysis on *D. suzukii* populations in Germany. **a** 34 populations with a total of 670 individuals were analyzed: six laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_France, and LS_Valsugana), nine populations from 2017, nine populations from 2018 and ten populations from 2019; the possible number of clusters are shown for K=2 and K=3. **b** 36 populations with a total of 710 individuals are shown. In addition to the 34 populations in A), laboratory strains LS_HG18 and LS_HG19 were added to the calculation. Shown are the two clusters K=3 and K=4
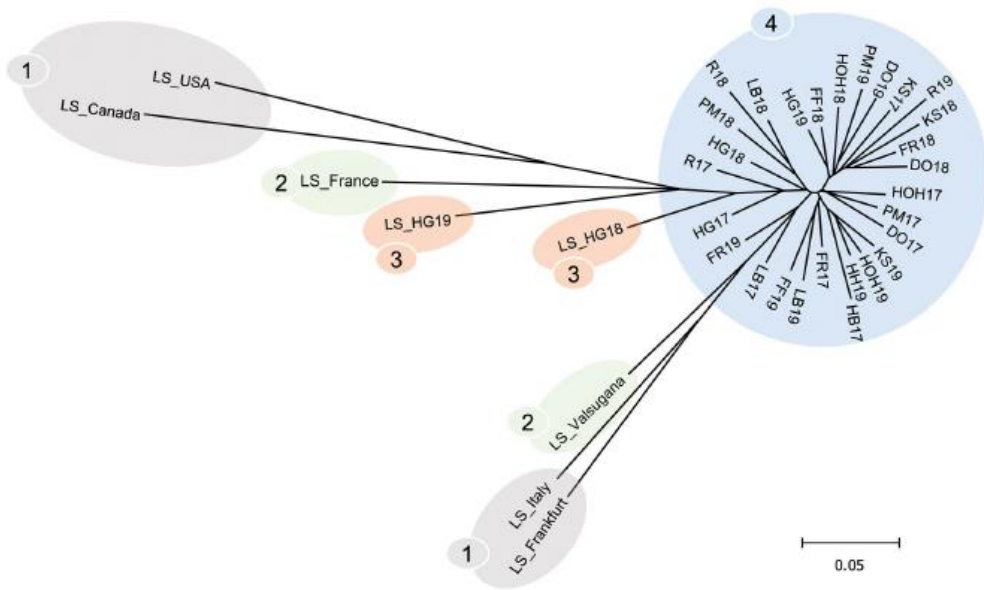
Fig. 6 Unrooted neighbor-joining tree based on Nei's genetic distance. Allelic frequencies were obtained with 14 microsatellite markers for the 36 *D. suzukii* populations. Circles represent population origin with 1 = long-established laboratory strains (2010–2016), 2 = laboratory strains from France (LS_France, 2018) and Italy (LS_Valsugana, 2018), 3 = laboratory strain from Bad Homburg (2017) and 4 = German populations collected over three years (2017–2019) in the field

119

Discussion

Invasive species like the SWD constitute a threat to agriculture, economy, and biodiversity (Mooney and Cleland 2001). Thus, understanding population movement, genetic structure, and diversity is one crucial component for the development of pest management strategies. This study addressed the spatial and temporal genetic variation of SWD populations in Germany for three years (2017–2019).

Our analysis of microsatellite markers revealed that levels of genetic diversity in Germany are comparable with other European countries (Tait 2017) and that genetic differentiation of sampled SWD is reflected among individuals within populations but not among populations from different sample sites. None of the populations showed a gain or loss of genetic information over the years. The results suggest a substantial gene flow and a more homogeneous gene pool across different geographical populations. Also, no differences were found between sample sites close to cities like Dortmund (DO) and populations in more rural areas like Derwitz (PM). The mean inbreeding coefficient (F_{IS}) was positive in most populations. This is usually interpreted as a sign of inbreeding, but in this study, precautions were taken during the sampling to avoid this. Another factor that can influence the F_{IS} value is the presence of null alleles. Since null alleles can cause a reduction in heterozygosity, the F_{IS} value increases. However, we did not include microsatellite markers that showed the possibility of null alleles in our final data set.

Furthermore, methodological issues can also be excluded, since the marker system established by Fraimout et al. (2015) and used in this study, was detecting genetic differences between our laboratory strains (LS_USA, LS_Canada, LS_Italy, LS_Frankfurt, LS_Valsugana, LS_France and LS_HG18 or LS_HG19). This study suggests that SWD is a well-established, uniform population in Germany, that might not be altered much by additional invasions. Based on the low differentiation between populations and years, there are either no reinvasions taking place or they do not have an impact on local populations. This is supported by our STRUCTURE, NJ, and PCoA results, which did not group German populations into distinct subpopulations. The lack of admixture suggests that colonization may have involved a single founder population rather than individuals from different origins. This is in contrast to findings from Ukraine, which found evidence for multiple sources of SWD invasions into Europe (Lavrinienko et al. 2017). On the other hand, our results are in accordance with results demonstrating European populations of SWD were genetically more homogenous with lower levels of genetic diversity compared to populations from North America (Adrion et al. 2014; Fraimout et al. 2017).

Based on our presented data, it seems likely that SWD established itself in Germany, which is not a surprise considering its invasion history in other parts of the world (Asplen et al. 2015). SWD has a relatively high reproductive capacity with a single female laying hundreds of eggs during its life with up to 10 generations a year (Walsh et al. 2011), which plays an important role in its invasion success. Additionally, SWD fits the European and German ecosystem, where there is only a constrained number of natural predators and parasitoids (Chabert et al. 2012; Rossi et al. 2015). Geographic barriers like mountains, lakes, and rivers can lead to genetic differences in populations. Germany is only streaked by an orogenic belt of relatively low mountains and hills, the Central German Uplands, but not by high mountain ranges. Based on the low variation among populations, it seems that these geographic properties do not act as barriers or at least they are not isolating populations. Also, SWD can move from high to low elevations, travel long distances by flight and distinct populations from island versus mainland can be found (Tait et al. 2017, 2018). In this respect, it would be interesting to analyze samples from a German island in the North Sea or Baltic Sea in the future. Another factor that can influence populations is the climate. Even though climate, temperature, and humidity differ to a certain extent between sampling locations and years, we were not able to detect these differences in our data. These findings again match the results from Tait et al. (2017), who did not detect differences between the populations from the much colder climate of Trentino compared to the rest of Italy.

All invasive species, independent from their origin, have in common that passive transport enables their successful spread (Banks et al. 2015). Anthropogenic transport of goods is an important aspect for the dissemination of flies since it facilitates the gene flow between locations and international trade via air and sea transport provide new pathways for the spread of insect pest in general (Hulme 2009). Even though it seems to be impossible to reconstruct the exact routes in detail due to the enormous amount of imported fresh produce, it is reasonable to assume that transportation of host fruits and plants lead to an extensive movement of SWD or other pests not only across Germany but all over Europe (Cini et al. 2014). Germany is importing large amounts of host plants and crops from all around the world. A look at the importation routes shows that most fresh fruits are imported from within Europe, namely Spain and Italy but also from the USA, and South America (source: International Trade Center, www.trademap.org). In that respect, distinct populations could be possible, but we did not find proof for this. With the data obtained in this study, it is not possible to reconstruct invasion routes from outside Germany, but would be an interesting aspect for further experiments. However, our collection did not include SWD samples from countries like Spain or the US and, therefore, it is not clear

where the German population might have originated from. However, Fraimout et al. (2017) did find evidence that the German sample site in their experiment most likely originated from an admixture of SWD populations from Asia and the eastern US. In general, a better comprehension of genetic structure, population dynamics and the reconstruction of invasion routes could improve the pest control at a regional scale.

Another option for integrated pest management and environmentally friendly alternatives to pesticides could be the Sterile Insect Technique (SIT) that has proven highly effective in agricultural insect species (Krafsur 1998; Benedict and Robinson 2003; Wyss 2006; Augustinos et al. 2017). For SIT programs, sterilized male individuals are released into the environment and lead to infertile matings that could reduce the population size of SWD in the future. Several current SIT programs consider the genetic background refreshing as a vital tool for mating success and efficacy of the release programs (Estes et al. 2012; Zygouridis et al. 2014; Parreño et al. 2014). Due to our findings of genetic uniformity in wild German SWD populations, we hypothesize, that this could be beneficial for the mating success of a single suitable mass-reared SWD strain to different wildtype populations during a SIT program. Therefore, while SWD control is still challenging, biological control methods, including the SIT, remain a beneficial option for sustainable pest control.

Furthermore, our study demonstrates that experiments that are meant to produce data for field applications should be performed with freshly sampled flies because SWD laboratory strains are different from wild populations and can change reasonably within two years. Our newly established laboratory strain LS_HG illustrated the effects of laboratory inbreeding, change of genetic markers, and a decline of allelic diversity. Whether this decline is due to a random selection of individuals during stock keeping or due to laboratory "adaptation" cannot be concluded from our data. However, over two years, the strain got more similar to other laboratory strains. This effect is important to consider during scientific experiments that rely on the use of laboratory strains (Lee et al. 2011; Kinjo et al. 2014; Shearer et al. 2016; Hamby et al. 2016). These strains are often kept over several years or even decades without any changes. This way, transparent and easily reproducible experiments can be performed, and results from those studies are taken as the threshold for similar experiments and as a portrayal of natural processes. However, strains that arise from human cultivation undergo evolutionary genetic changes (Knudsen et al. 2020). Therefore, for evaluations of the efficiency of naturally occurring predators, parasitoids, bacteria or viruses for biological pest control, invasion history, or behavior, fresh field collections or genetically refreshed strains should be the first choice over highly inbred laboratory strains.

Author contributions

SP performed research. SP, SO, and MFS designed experiments, SP, GE, and SO analyzed data and SP, SO, and MFS wrote the paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or vertebrates performed by any of the authors.

Consent for publication We consent with the publication of all submitted data.

Availability of data and material All data discussed in the paper will be made available to readers.

121

References

Adrion JR, Kousathanas A, Pascual M et al (2014) *Drosophila suzukii*: The genetic footprint of a recent, worldwide invasion. Mol Biol Evol 31:3148–3163. https://doi.org/10.1093/molbev/msu246

Aketarawong N, Chinvinijkul S, Orankanok W et al (2011) The utility of microsatellite DNA markers for the evaluation of area-wide integrated pest management using SIT for the fruit fly, *Bactrocera dorsalis* (Hendel), control programs in Thailand. Genetica 139:129–140. https://doi.org/10.1007/s10709-010-9510-8

Aketarawong N, Isasawin S, Thanaphum S (2014) Evidence of weak genetic structure and recent gene flow between Bactrocera dorsalis s.s. and *B. papaya*, across Southern Thailand and West Malaysia, supporting a single target pest for SIT applications. BMC Genet 15:70. https://doi.org/10.1186/1471-2156-15-70

Asplen MK, Anfora G, Biondi A et al (2015) Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci 88:469–494. https://doi.org/10.1007/s10340-015-0681-z

Augustinos AA, Targovska A, Cancio-Martinez E et al (2017) *Ceratitis capitata* genetic sexing strains: laboratory evaluation of strains from mass-rearing facilities worldwide. Entomol Exp Appl 164:305–317. https://doi.org/10.1111/eea.12612

Azrag RS, Ibrahim K, Malcolm C et al (2016) Laboratory rearing of *Anopheles arabiensis*: impact on genetic variability and implications for Sterile Insect Technique (SIT) based mosquito control in northern Sudan. Malar J 15:432. https://doi.org/10.1186/s12936-016-1484-2

Bahder BW, Bahder LD, Hamby KA et al (2015) Microsatellite variation of two Pacific Coast *Drosophila suzukii* (Diptera: Drosophilidae) Populations. Environ Entomol 44:1449–1453. https://doi.org/10.1093/ee/nvv117

Banks NC, Paini DR, Bayliss KL, Hodda M (2015) The role of global trade and transport network topology in the human-mediated dispersal of alien species. Ecol Lett 18:188–199. https://doi.org/10.1111/ele.12397

Bellamy DE, Sisterson MS, Walse SS (2013) Quantifying host potentials: indexing postharvest fresh fruits for spotted wing Drosophila, *Drosophila suzukii*. PLoS ONE 8:e61227–e61227. https://doi.org/10.1371/journal.pone.0061227

Benedict MQ, Robinson AS (2003) The first releases of transgenic mosquitoes: an argument for the sterile insect technique. Trends Parasitol 19:349–355. https://doi.org/10.1016/S1471-4922(03)00144-2

Bolda MP, Goodhue RE, Zalom FG (2010) Spotted wing drosophila: potential economic impact of a newly established pest. Agric Resour Econom Update 13:5–8

Boughdad A, Haddi K, El Bouazzati A et al (2021) First record of the invasive spotted wing Drosophila infesting berry crops in Africa. J Pest Sci 94:261–271. https://doi.org/10.1007/s10340-020-01280-0

Briem F, Breuer M, Köppler K, Vogt H (2015) Phenology and occurrence of spotted wing Drosophila in Germany and case studies for its control in berry crops. IOBC-WPRS Bull 109:233–237. http://www.iobcwprs.org/pub/bulletins/bulletin_2015_109_table_of_contents_abstracts.pdf

Briem F, Dominic AR, Golla B et al (2018) Explorative data analysis of *Drosophila suzukii* trap catches from a seven-year monitoring program in Southwest Germany. Insects 9:125. https://doi.org/10.3390/insects9040125

Calabria G, Máca J, Bächli G et al (2012) First records of the potential pest species *Drosophila suzukii* (Diptera: Drosophilidae) in Europe. J Appl Entomol 136:139–147. https://doi.org/10.1111/j.1439-0418.2010.01583.x

Chabert S, Allemand R, Poyet M et al (2012) Ability of European parasitoids (Hymenoptera) to control a new invasive Asiatic pest, *Drosophila suzukii*. Biol Control 63:40–47. https://doi.org/10.1016/j.biocontrol.2012.05.005

Chapman D, Purse BV, Roy HE, Bullock JM (2017) Global trade networks determine the distribution of invasive non-native species. Glob Ecol Biogeogr 26:907–917. https://doi.org/10.1111/geb.12599

Chapman D, Makra L, Albertini R et al (2016) Modelling the introduction and spread of non-native species: international trade and climate change drive ragweed invasion. Glob Chang Biol 22:3067–3079. https://doi.org/10.1111/gcb.13220

Cini A, Anfora G, Escudero-Colomar LA et al (2014) Tracking the invasion of the alien fruit pest Drosophila suzukii in Europe. J Pest Sci 87:559–566. https://doi.org/10.1007/s10340-014-0617-z

Cini A, Ioriatti C, Anfora G (2012) A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. Bull Insectol 65:149–160

Dalton DT, Walton VM, Shearer PW et al (2011) Laboratory survival of *Drosophila suzukii* under simulated winter conditions of the Pacific Northwest and seasonal field trapping in five primary regions of small and stone fruit production in the United States. Pest Manag Sci 67:1368–1374. https://doi.org/10.1002/ps.2280

de la Vega GJ, Corley JC, Soliani C (2020) Genetic assessment of the invasion history of *Drosophila suzukii* in Argentina. J Pest Sci 93:63–75. https://doi.org/10.1007/s10340-019-01149-x

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Estes AM, Nestel D, Belcari A et al (2012) A basis for the renewal of sterile insect technique for the olive fly, *Bactrocera oleae* (Rossi). J Appl Entomol 136:1–16. https://doi.org/10.1111/j.1439-0418.2011.01620.x

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: Why, how and so what? Mol Ecol 19:4113–4130. https://doi.org/10.1111/j.1365-294X.2010.04773.x

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer H (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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Food and Agriculture Organization of the United Nations (2021) FAOSTAT. http://www.fao.org/faostat/en/#data/TP. Accessed 15 Feb 2021

Fraimout A, Debat V, Fellous S et al (2017) Deciphering the routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. Mol Biol Evol 34:980–996. https://doi.org/10.1093/molbev/msx050

Fraimout A, Loiseau A, Price DK et al (2015) New set of microsatellite markers for the spotted-wing *Drosophila suzukii* (Diptera: Drosophilidae): a promising molecular tool for inferring the invasion history of this major insect pest. Eur J Entomol 112:855–859. https://doi.org/10.14411/eje.2015.079

Francis RM (2017) pophelper: an R package and web app to analyse and visualize population structure. In: Molecular Ecology Resources. Blackwell Publishing Ltd, pp 27–32. Doi: https://doi.org/10.1111/1755-0998.12509

Gabarra R, Riudavets J, Rodríguez GA et al (2015) Prospects for the biological control of *Drosophila suzukii*. Biocontrol 60:331–339. https://doi.org/10.1007/s10526-014-9646-z

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Mol Ecol 8:1513–1520. https://doi.org/10.1046/j.1365-294x.1999.00730.x

Guo W, Lambertini C, Pyšek P et al (2018) Living in two worlds: evolutionary mechanisms act differently in the native and introduced ranges of an invasive plant. Ecol Evol 8:2440–2452. https://doi.org/10.1002/ece3.3869

Hamby KA, Bolda MP, Sheehan ME, Zalom FG (2014) Seasonal monitoring for Drosophila suzukii (Diptera: Drosophilidae) in California Commercial Raspberries. Environ Entomol 43:1008–1018. https://doi.org/10.1603/EN13245

Hamby KA, Bellamy DE, Chiu JC et al (2016) Biotic and abiotic factors impacting development, behavior, phenology, and reproductive biology of Drosophila suzukii. J Pest Sci 89(3):605–619. https://doi.org/10.1007/s10340-016-0756-5

Hauser M (2011) A historic account of the invasion of Drosophila suzukii (Matsumura) (Diptera: Drosophilidae) in the continental United States, with remarks on their identification. Pest Manag Sci 67:1352–1357. https://doi.org/10.1002/ps.2265

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

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

Ioriatti C, Guzzon R, Anfora G et al (2018) Drosophila suzukii (Diptera: Drosophilidae) contributes to the development of sour rot in grape. J Econ Entomol 111:283–292. https://doi.org/10.1093/jee/tox292

Jakobs R, Gariepy TD, Sinclair BJ (2015) Adult plasticity of cold tolerance in a continental-temperate population of Drosophila suzukii. J Insect Physiol 79:1–9. https://doi.org/10.1016/j.jinsphys.2015.05.003

Janes JK, Miller JM, Dupuis JR et al (2017) The K = 2 conundrum. Mol Ecol. https://doi.org/10.1111/mec.14187

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424–429. https://doi.org/10.1016/0169-5347(96)10049-5

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kinjo H, Kunimi Y, Nakai M (2014) Effects of temperature on the reproduction and development of Drosophila suzukii (Diptera: Drosophilidae). Appl Entomol Zool 49:297–304. https://doi.org/10.1007/s13355-014-0249-z

Knudsen KE, Reid WR, Barbour TM et al (2020) Genetic variation and potential for resistance development to the TTA overexpression lethal system in insects. G3 Genes Genomes Genetics 10:1271–1281. https://doi.org/10.1534/g3.120.400990

Krafsur ES (1998) Sterile insect technique for suppressing and eradicating insect population: 55 years and counting. Entomology Publications, p 417. https://lib.dr.iastate.edu/ent_pubs/417

Lanzavecchia SB, Juri M, Bonomi A et al (2014) Microsatellite markers from the "South American fruit fly" Anastrepha fraterculus: a valuable tool for population genetic analysis and SIT applications. BMC Genet 15:S13. https://doi.org/10.1186/1471-2156-15-S2-S13

Lavrinienko A, Kesäniemi J, Watts PC et al (2017) First record of the invasive pest Drosophila suzukii in Ukraine indicates multiple sources of invasion. J Pest Sci 90(2):421–429. https://doi.org/10.1007/s10340-016-0810-3

Lee JC, Bruck DJ, Curry H et al (2011) The susceptibility of small fruits and cherries to the spotted-wing drosophila, Drosophila suzukii. Pest Manag Sci 67:1358–1367. https://doi.org/10.1002/ps.2225

Lee KZ, Vilcinskas A (2017) Analysis of virus susceptibility in the invasive insect pest Drosophila suzukii. J Invertebr Pathol 148:138–141. https://doi.org/10.1016/j.jip.2017.06.010

Matsumura S (1931) 6000 illustrated insects of Japan-Empire (in Japanese). Tokohshoin, Tokyo, p 1497

Mazzi D, Bravin E, Meraner M et al (2017) Economic impact of the introduction and establishment of Drosophila suzukii on sweet cherry production in Switzerland. Insects 8:18. https://doi.org/10.3390/insects8010018

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. Proc Natl Acad Sci U S A 98:5446–5451. https://doi.org/10.1073/pnas.091093398

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data—II. Gene frequency data. J Mol Evol 19:153–170. https://doi.org/10.1007/BF02300753

Nei M, Takezaki N (1983) Estimation of genetic distances and phylogenetic trees from DNA analysis. Proc 5th World Cong Genet Appl Livstock Prod 21:405–412

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65. https://doi.org/10.1046/j.1365-294X.2004.02008.x

Parreño MA, Scannapieco AC, Remis MI et al (2014) Dynamics of genetic variability in Anastrepha fraterculus (Diptera: Tephritidae) during adaptation to laboratory rearing conditions. BMC Genet 15:S14. https://doi.org/10.1186/1471-2156-15-S2-S14

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Piry S, Luikart GCJ (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503. https://doi.org/10.1093/jhered/90.4.502.

Piry S, Alapetite A, Cornuet JM et al (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536–539. https://doi.org/10.1093/jhered/esh074

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1534/genetics.116.195164

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci U S A 94:9197–9201. https://doi.org/10.1073/pnas.94.17.9197

Renkema JM, Telfer Z, Gariepy T, Hallett RH (2015) Dalotia coriaria as a predator of Drosophila suzukii: Functional responses, reduced fruit infestation and molecular diagnostics. Biol Control 89:1–10. https://doi.org/10.1016/j.biocontrol.2015.04.024

Rombaut A, Guilhot R, Xuéreb A et al (2017) Invasive Drosophila suzukii facilitates Drosophila melanogaster infestation and sour rot outbreaks in the vineyards. R Soc Open Sci 4:170117. https://doi.org/10.1098/rsos.170117

Rossi Stacconi MV, Buffington M, Daane KM et al (2015) Host stage preference, efficacy and fecundity of parasitoids attacking Drosophila suzukii in newly invaded areas. Biol Control 84:28–35. https://doi.org/10.1016/j.biocontrol.2015.02.003

Rota-Stabelli O, Ometto L, Tait G et al (2020) Distinct genotypes and phenotypes in European and American strains of Drosophila suzukii: implications for biology and management of an invasive organism. J Pest Sci 93(1):77–89. https://doi.org/10.1007/s10340-019-01172-y

Rout TM, Moore JL, Possingham HP, McCarthy MA (2011) Allocating biosecurity resources between preventing, detecting, and eradicating island invasions. Ecol Econ 71:54–62. https://doi.org/10.1016/j.ecolecon.2011.09.009

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371. https://doi.org/10.1007/s004120000089

Schlötterer C (2004) The evolution of molecular markers—Just a matter of fashion? Nat Rev Genet 5:63–69. https://doi.org/10.1038/nrg1249

Schrader L, Kim JW, Ence D et al (2014) Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun 5:1–10. https://doi.org/10.1038/ncomms6495

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Shearer PW, West JD, Walton VM et al (2016) Seasonal cues induce phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. BMC Ecol 16:1–18. https://doi.org/10.1186/s12898-016-0070-3

Shorthouse DP (2010) SimpleMappr, an online tool to produce publication-quality point maps. http://www.simplemappr.net. Accessed 18 July 2020.

Stephens AR, Asplen MK, Hutchison WD, Venette RC (2015) Cold Hardiness of Winter-Acclimated *Drosophila suzukii* (Diptera: Drosophilidae) Adults. Environ Entomol 44:1619–1626. https://doi.org/10.1093/ee/nvv134

Stockton DG, Wallingford AK, Brind'amore G et al (2020) Seasonal polyphenism of spotted-wing Drosophila is affected by variation in local abiotic conditions within its invaded range, likely influencing survival and regional population dynamics. Ecol Evol 10:7669–7685. https://doi.org/10.1002/ece3.6491

Tait G, Grassi A, Pfab F et al (2018) Large-scale spatial dynamics of *Drosophila suzukii* in Trentino, Italy. J Pest Sci 91:1213–1224. https://doi.org/10.1007/s10340-018-0985-x

Tait G, Vezzulli S, Sassù F et al (2017) Genetic variability in Italian populations of *Drosophila suzukii*. BMC Genet 18:87. https://doi.org/10.1186/s12863-017-0558-7

Takezaki N, Nei M, Tamura K (2014) POPTREEW: web version of POPTREE for constructing population trees from Allele frequency data and computing some other quantities. Mol Biol Evol 31:1622–1624. https://doi.org/10.1093/molbev/msu093

Vogt H, Baufeld P, Gross J et al (2012) *Drosophila suzukii:* a new threat feature for the European fruit and viticulture-report for the international conference in Trient, 2, December 2011. J für Kulturpflanzen 64:68–72

Walsh DB, Bolda MP, Goodhue RE et al (2011) *Drosophila suzukii* (Diptera: Drosophilidae): invasive pest of ripening soft fruit expanding its geographic range and damage potential. J Integr Pest Manag 2:G1–G7. https://doi.org/10.1603/ipm10010

Wyss JH (2006) Screwworm eradication in the Americas. Ann N Y Acad Sci 916:186–193. https://doi.org/10.1111/j.1749-6632.2000.tb05289.x

Xia L, Geng Q, An S (2020) Rapid genetic divergence of an invasive species, *Spartina alterniflora*, in China. Front Genet 11:284. https://doi.org/10.3389/fgene.2020.00284

Zygouridis NE, Argov Y, Nemny-Lavy E et al (2014) Genetic changes during laboratory domestication of an olive fly SIT strain. J Appl Entomol 138:423–432. https://doi.org/10.1111/jen.12042

124

A.1.2 Sanger Sequencing Results for Tested Microsatellite Markers

Shown are the Sanger sequencing results for each microsatellite marker tested in this study. Letters in boldface indicate the primer sequence and letters marked in red show the repeat motif as it is described in the publication of Fraimout (2015).

DS21

5'-**GAGACGCGATGGTACCGTTA**CCTATTTATCCCTTTGTGAGCCCAATCTTCGA
ATACGCTTTCACCTCTCTCCCGCTGTTAGATTACATATGTGTGTGTGTGTGTGTGCTTCGATGGCC
TTTTTGGTCAATTGTCCTTTTTGTTGTTGTTCTTTCCGTGGGTCAAACACACACACACACACTT
AGGTGTTTTTTTGGTGCTCGGCCAAGGGAACAACAAATACATGCGACCCTCGGCAGCGCTGCGACT
GCGCTGCCCGCGTCCACCTGCAGGGGGTTGCATTTTTAGGCCACACCCACATGCGAATGCG**ACGCT
TGCACTCGATTGG-3'**

DS27

5'-**CCAGCGACTGCAGAAGTGAC**GTCAATGAGCGCGAGGCCAAATTGCCACAGTGTGTG
TGTGTGTGA**GTTGTGTTGTGGAGGATTGC-3'**

DS10

5'-**CGAGACTGTGCGAACGAGAG**AGAAGGGACGAAGCCTGTGTGTGCTTCTTTT
GTCCTGCTCCCTCTGTGTATAACGGTAAATGGCAAATGGTACATATGTACATATATACACTTCGAA
GGCAAAGTAGAGGAAAATGTATGCAACATTTTACAATGCAAAAGCCATTTACTTGGGCAACGAAGA
AAAGCCAAGGGAGATGGAAAGCGGAGAACCTAAGAGAGAAAGAGAAAGAGGACCTTTGCGGGTACG
TGTGTGTGTGTGTGTGTGTGTGTA**TGTGAGGCAGTCAGCATATG-3'**

DS28

5'-**TTAAGCTGACCTCCTCCTCG**AGCGGCTAGCGTCGACGGCGACGTCGGCAGA
GGCGGCGGCAGCGGCAGCGGCGAATTGCGGTCGAAACCGTCCCTGGGCGAATCTATCTGAATCTGT
GTGTGTGTGTGTGTG**CCTTGTATCTGTGCGAGTGC-3'**

DS35

5'-**TCCGTATTCCGTATCCGTGT**TCCCCTTTTCCGTATCCTGGTAACATCCCATTG
GCCCACCTCTCACGCAGTTGCCCCGCGGGAGGTTGCGAAAACTTTGCGTAATACCAAAAAGTGCTG
GCAGCACATTTCGAGTGCAAACTTTATATTAGAGTTGAAGTGCCACACACACTCGCATCGCACACA
CACACACACACACACT**CTGCCACACTGCCATACTCC-3'**

DS34

5′–**AACAACGACGCAGAAGCTCA**AGGCGAGCAGCAACAACAAAAGTCGGCGGC
ATATAATACAACACGTGAAACGTGACAAATTCCAAAGCGAAACGGCGAAGTCAACTGGGCGGCGGC
TGCGAATGAGAGAGAGAGAGAGAGAGAGAAAAGCTGATAGCCGCCGCTTGGCAACAACAATCCCAT
CACGAAGCCAAACACACATGCACACACACACTCACCGCTCGGCCGACATTTTGTTTTT**ATTACAGA
GCGCAACAGTCG**–3′

DS23

5′–**TGCCACTAAGCTCACACGGT**GCTTTTCGCTTCTTATTTTATTTTTTTACTTAGGA
TGGGGATGGGGTTTTCCTCTTTTCCGAGCGATTTACGAGTCTGGCCTGGGTTCTACGTTTTCGCGT
CCCGGGGCAAAGGCATCCCCTGCATTTTACAACATTTGTATTTATTTGGCGTCATGAGTTTAAGAT
GTTGTTGATATATGTTTACCGTTGGGTGACACACACACACACACGAA**TACACAGCAAGTGGCAACT
G**–3′

DS22

5′–**TACAGATACGCCGTCGGATT**TCTTTATTATACGATTTTGATTCGAAAATAATTC
AAACGAATTTCGGGGCTGCCTGGCTTTTTGAGAGCGCGTGCGCAACAGCCAAGCAGACAAAGGAAAA
CCCGCTGCGGGTTTTTCGCCCAGTGTAGGCCACATAAAGCGTAGCACGGTAGCCGAGCTACAGGTTC
GAGTGCGCCAGTGTGTGTGTGTGTGTGTGTATGGACGATGACGCTGGTAAAAAAATAATGAATAAAAG
AAAACCTGTACACAAAAAACCAGACAAATAAACCAAATACAAAAGCCTTCTCCTCG**AGGTCCGTCG
TCTTGGTCTT**–3′

DS20

5′–**CAGCCATATGCAATGCACTG**GCCGGAAACGGAAATGCACGCTGCCATTTTA
GCATGCCGCCCGCACACACACACACACAAAAGAGAGAGAAAGAGAGGAAGCGGGGCATACAAATGAGG
CGTTGAATAATTGAAGACATAAATTCGCACTTGCACATACCAAAAAAAAAGTTTTTTTATTGTTGC
CGGCTTCTGGGCTCTCAG**CTCTCGACTTCCGCTGGATAT**–3′

DS14

5′–**AAGAACCGCAACGAGCAA**AACTCAATTGCGGCAAAGTAACTAAAATTGCTTT
GTCAGGGTAAAAGAGAATTTTCCATCAGGGGGAGGGCGGTAAATATTTAGCCCCCGCCCCCAAAAA
TACAAAAGAGGGGCGTGGCAGAAAGCCAAACTCCTGTGAACAGCACATGTGTGTGTGTGTGTGTGT
GC**GTCGTGTCGCTGGATAATTC**–3′

DS07

5′–**AAGGCTGGAGTGGCAACAA**ACATTAAGAGCAGAAACATTTGCATATTCCGTA
TATTTGGCTTTTGCCTTAATTTGTTTGCACATATAAGCAAACTCCCACACACACACATACACACAC
AATCCCAACAAAGTGACTTTGCATGTTTAACAGCTCAGCAGATAAGAACC**CAGCCGAACAGAACCT
TAGC**–3′

DS06

5'–**CGGTTCGAGTGCTTGTTAGA**AAGAGCGCGAGTGCGGTGGTAAAGAGATGCC
AAATGGATGTGAATGGACGTATAATGTGTGAGTGTGTGTGTGAGAGATCCTGAAACTCGTTTTGTC
GCTAATTGGATGCT**GAAGGTGTCCTCCACGTGT**–3'

DS11

5'–**CGGTGACTCGTGCAGTTGTA**TGGTTGTATTTTAAGAAGCTGACCACACACTTA
CACCCAGCTCACCCACAAGTGCATATGCAACACCCCCAAGCACACACACACACTCACACACACACT
GGCAAACCGCATGCAGCTGCAACAGCTTCCGTTTACCACCACCCCACTTGGCCACCCAAAAAGTTT
TCGTTTTGCTTTCTTGGACTTTGCCTAATGCACATTTTTGCATT**TTGCTCTAGACAGAGTCGGC**–
3'

DS36

5'–**TTGGCAACGTGTGTGAAGCTG**CGAAATCTCAGCTGTAATTGTTGAATAATTCAGT
TAACAGGCGGGATAATGAGACGCCGAAAATGAAAGCTCTCCTTGAGTTTTAGAGTGTGTGTGTGTG
TGTGTGTTTTTGGCCATCTAACGGCAATTTATTGATGAACAATG**AGGCAGCATTGCAGTGTCTC**–3'

DS26 (excluded)

5'–**CCTGTGTGCATCTCAGTGTT**TGTGTTTTGGTGTGTGTTTGTTGTGCGGTGTGTGTGTTG**T**
CATGTGCTGGAGTGCTGTA–3'

DS17 (excluded)

5'–**CATCTCAGGCCACGAATG**CCAGTGTGTGTGTGTGTGTGTGTGTGGAAAGTATCTGGA
ACTGGCTCTGGCGGGGCTTCATCTTT**CTGCACTCGAGAATCTGGAG**–3'

DS32 (excluded)

5'–**CGGCGTGTTGCAGTTATTC**ATAATCGTTTGTCAGCGCTGAGGGCTCTCTCTG
TGTGTGTGTGGGGGGGGGGGTCCTCTGGCGTGTGTGTATACACGTGCTTATCCACAGCAAACATGCG
GCCACCCATACTAGCGCACAGAGCGGCCGCAGTATGCGGCAAAGCAATTTCGGTAAACGCTTTTTG
GGCAGAACGGGGCGTATACGCAATACATTATATTATATTCCGCACTGACCAGAAACTGTTGTAAGC
CCACGCGGGAGCAATAAATAATTGAAATCTGGCATATTACCGAATGCTGTGAATTGTTGAATTCAA
CAATTGCCA**TGTCATGTCGACCAGTGCAT**–3'

A.1.3 Results for Linkage Disequilibrium for Each Population

<u>DO17</u>

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	2	2	1	1	2	0	2	0	0	1	2	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	+	-
34	-	*	+	-	-	-	-	-	-	-	-	-	-	-
21	-	+	*	-	-	-	-	-	-	-	-	-	-	+
20	-	-	-	*	+	-	+	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	+	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	+	-	-	-	-	-	-	-	-	-	-	-	*	+
28	-	-	-	-	-	-	-	-	+	-	-	-	+	*

<u>DO18</u>

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	0	4	3	4	1	4	4	0	1	0	3	2	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	+	-	-	+	-	-	-	-	+	+	-
20	-	-	+	*	+	-	-	+	-	-	-	-	-	-
35	-	-	-	+	*	-	+	-	-	-	-	+	+	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	+	-	+	+	*	-	-	-	-	-	-	+
06	-	-	-	+	-	-	-	*	-	+	-	+	-	+
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	+	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	+	-	+	-	-	+	-	-	-	*	-	-
10	-	-	+	-	+	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	+	+	-	-	-	-	-	*

DO19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	0	0	0	5	2	0	0	0	3	2	0	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	+	-	-	-	-	-	-	-	-
34	-	*	-	-	-	+	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	-	-	-	-	-	-	-	-
23	+	+	-	-	-	*	+	-	-	-	+	+	-	-
22	-	-	-	-	-	+	*	-	-	-	+	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	+	+	-	-	-	*	-	-	+
07	-	-	-	-	-	+	-	-	-	-	-	*	-	+
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	+	+	-	*

FF18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	2	1	1	2	1	2	3	3	0	1	1	1	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	+	-	-	-	-	+	-
21	-	-	*	-	-	+	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	+	-	-	-	-	-	-
35	-	-	-	-	*	-	+	+	-	-	-	-	-	-
23	-	-	+	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	+	-	*	-	-	-	+	-	-	-
06	-	-	-	+	+	-	-	*	-	-	-	-	-	+
36	-	+	-	-	-	-	-	-	*	-	-	+	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	+	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	+	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	+	+	-	-	-	-	*

FF19
Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	0	0	2	2	2	1	1	1	0	1	2	1	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	+	+	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	+
23	-	-	-	+	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	+	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	+
07	-	-	-	-	-	-	-	-	+	-	-	*	-	+
10	-	-	-	-	-	-	-	+	-	-	-	-	*	-
28	-	-	-	-	+	-	-	-	-	-	+	+	-	*

FR17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	3	2	1	3	2	3	0	0	2	1	1	1	0	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	+	-	-	+	-	-	-	-	+	-	-	-	-
34	+	*	-	-	-	-	-	-	+	-	-	-	-	-
21	-	-	*	-	-	+	-	-	-	-	-	-	-	-
20	-	-	-	*	-	+	-	-	-	-	-	-	-	+
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	+	+	-	*	-	-	-	-	-	-	-	+
22	-	-	-	-	-	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	+	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	+
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	+	-	+	-	-	-	-	+	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	1	0	1	1	1	1	0	1	1	1	1	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	+	-	-	-	-
34	-	*	-	-	+	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	+
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	+	-	-	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	+
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	+	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	+	-	-
07	-	-	-	-	-	-	-	-	-	-	+	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	+
28	-	-	+	-	-	-	-	+	-	-	-	-	+	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	2	5	1	2	5	3	2	2	0	1	1	2	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	+	-	-	-	-	-	-	-	-	-	-	-
34	-	*	+	-	-	+	-	-	-	-	-	-	-	-
21	+	+	*	-	-	+	-	+	-	-	-	-	+	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	-	-	+	*	+	-	-	-	-	-	-	-	-
23	-	+	+	-	+	*	+	+	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	+	-	-	-	-	+
06	-	-	+	-	-	+	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	+	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	+	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	+	-	-	-	-	-	-	-	+	-	*	-
28	-	-	-	-	-	-	+	-	-	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	3	2	0	0	1	3	2	2	1	2	2	2	1	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	+	-	+	-	-	+	-
34	-	*	-	-	+	-	-	-	-	-	-	+	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	+	-	-	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	-	-	+	+	-	+	-	-
22	-	-	-	-	-	-	*	+	-	-	+	-	-	-
06	+	-	-	-	-	-	+	*	-	-	-	-	-	-
36	-	-	-	-	-	+	-	-	*	-	-	-	-	-
11	+	-	-	-	-	+	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	+	-	-	-	*	-	-	+
07	-	+	-	-	-	+	-	-	-	-	-	*	-	-
10	+	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	+	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	0	1	1	1	2	2	3	0	1	0	1	2	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	+	-	-	-	-	-	-	+
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	+	-	-	-	-	-	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	+
22	+	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	+	-	-	-	-	*	-	+	-	-	+	-
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	+	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	-	-	-	*	-	-
10	-	-	-	-	-	-	-	+	-	-	-	+	*	-
28	+	-	-	-	-	+	-	-	-	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	0	1	0	1	0	0	0	1	1	0	2	0	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	+
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	+	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	-	-	-	-	-	-	+	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	+	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	+	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	+	-	-	-	-	-	-	-	-	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	2	1	0	3	0	2	1	4	0	2	3	0	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	+	-	-	-	+	-	-	-	-	-
34	-	*	+	-	-	-	-	+	-	-	-	-	-	-
21	-	+	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	+	-	-	-	*	-	-	-	+	-	-	-	-	+
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	*	-	-	-	+	+	-	-
06	-	+	-	-	-	-	-	*	-	-	-	-	-	-
36	+	-	-	-	+	-	-	-	*	-	-	+	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	+	-	-	-	*	+	-	-
07	-	-	-	-	-	-	+	-	+	-	+	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	+	-	-	-	+	-	-	-	-	*

HH19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	1	1	0	0	1	0	1	0	1	2	2	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	+	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	+
21	-	-	*	-	-	-	+	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	+	-
35	-	-	-	-	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	+	-	-	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	+	-	-
07	-	-	-	-	-	-	-	-	+	-	+	*	-	-
10	+	-	-	+	-	-	-	-	-	-	-	-	*	-
28	-	+	-	-	-	-	-	-	-	-	-	-	-	*

HOH17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	2	0	3	0	3	2	0	2	2	0	2	1	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	+	-	-	-	-	-	-	-	-	-	-	-	-
34	+	*	-	-	-	-	-	-	+	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	+	-	-	-	+	-	+	-	-
35	-	-	-	-	*	-	-	-	-	-	-	-	-	-
23	-	-	-	+	-	*	+	-	-	-	-	-	+	-
22	-	-	-	-	-	+	*	-	-	+	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	+	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	+	-	-	+	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	+	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	+	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

HOH18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	2	3	2	3	3	3	2	2	2	3	1	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	+	-	-	-	-	-	-	-	-
34	-	*	+	-	-	-	-	-	-	-	-	-	-	-
21	-	+	*	-	-	+	-	-	-	-	-	-	-	-
20	-	-	-	*	+	-	-	+	-	+	-	-	-	-
35	-	-	-	+	*	-	-	-	+	-	-	-	-	-
23	+	-	+	-	-	*	-	-	-	+	-	-	-	-
22	-	-	-	-	-	-	*	+	-	-	-	+	-	+
06	-	-	-	+	-	-	+	*	-	-	-	+	-	-
36	-	-	-	-	+	-	-	-	*	-	-	-	+	-
11	-	-	-	+	-	+	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	+	-	+
07	-	-	-	-	-	-	+	+	-	-	+	*	-	-
10	-	-	-	-	-	-	-	-	+	-	-	-	*	-
28	-	-	-	-	-	-	+	-	-	-	+	-	-	*

HOH19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	0	1	0	2	3	1	1	2	3	0	1	1	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	+	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	+	-	-	+	-	-	-	-	-
23	-	-	-	-	+	*	+	+	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	+	-	*	-	-	-	-	-	-
36	-	-	-	-	+	-	-	-	*	-	-	+	-	-
11	-	-	+	-	-	-	-	-	-	*	-	-	+	+
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	+	-	-	*	-
28	-	-	-	-	-	-	-	-	-	+	-	-	-	*

KS17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	0	2	0	1	0	1	1	0	2	1	2	2	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	+	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	+	-	+	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	-	-	-	-	-	-	-	+
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	+	-	-	-	-	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	+	-
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	+	-	-	-	-	-	-	*	+	-	-	-
27	-	-	-	-	-	-	-	-	-	+	*	-	-	-
07	-	-	+	-	-	-	-	-	-	-	-	*	+	-
10	-	-	-	-	-	-	-	+	-	-	-	+	*	-
28	-	-	-	-	+	-	-	-	-	-	-	-	-	*

KS18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	1	0	0	1	1	3	2	1	2	2	3	1	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	+	-	+	-	-
34	-	*	-	-	-	-	-	-	+	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	-	-	-	-	-	-	+	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	+	-	-	+
06	-	-	-	-	-	-	-	*	-	-	-	+	-	+
36	-	+	-	-	-	-	-	-	*	-	-	-	-	-
11	+	-	-	-	-	-	-	-	-	*	+	-	-	-
27	-	-	-	-	-	-	+	-	-	+	*	-	-	-
07	+	-	-	-	-	-	-	+	-	-	-	*	-	+
10	-	-	-	-	+	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	+	+	-	-	-	+	-	*

KS19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	1	0	1	2	1	2	1	3	0	1	1	1	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	+	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	+	-	-	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	+	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	+	-
36	-	+	-	-	+	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	+	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	+	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

LB17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	0	1	4	2	0	5	3	2	1	3	5	3	4

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	+	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	+	-	-	-	-	-	-	-
20	-	-	-	*	+	-	+	-	-	-	+	-	-	+
35	-	-	-	+	*	-	-	-	-	-	-	+	-	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	+	+	-	-	*	+	-	-	-	+	-	+
06	-	-	-	-	-	-	+	*	-	-	+	+	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	+	-
27	-	-	-	+	-	-	+	-	-	-	*	-	+	-
07	-	-	-	-	+	-	+	+	+	-	-	*	-	+
10	+	-	-	-	-	-	-	-	-	+	+	-	*	-
28	-	-	-	+	-	-	+	-	+	-	-	+	-	*

LB18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	2	1	1	2	1	2	2	1	2	0	2	1	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	+	-	+	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	+	-	-	-	-	-	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	+	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	+	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	+	-	-	-	-	*	-	+	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	+	-	*	-	-	-	+
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	+	-
10	-	-	-	-	-	-	-	-	-	-	-	+	*	-
28	-	-	-	-	-	-	-	-	-	+	-	-	-	*

LB19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	2	1	1	1	1	2	0	2	3	1	4	1	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	+	-	-	-	-
34	-	*	-	-	-	-	-	-	+	-	-	+	-	-
21	-	-	*	-	-	-	-	-	-	+	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	+
35	-	-	-	-	*	-	-	-	-	-	-	+	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	+	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	+	-	-	-	-	-	-	*	-	-	+	-	-
11	+	-	+	-	-	-	+	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	+
07	-	+	-	-	+	-	-	-	+	-	-	*	+	-
10	-	-	-	-	-	-	-	-	-	-	-	+	*	-
28	-	-	-	+	-	-	-	-	-	-	+	-	-	*

PM17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	0	1	1	1	0	0	1	1	1	0	1	2	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	+	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	+	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	*	-	-	-	-	-	-	-
06	+	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	+	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	+	-	-	-	-	-	-	+	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

PM18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	3	0	3	3	1	1	1	2	1	2	1	0	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	+	+	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	+	-	+	+	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	+	-	-	*	+	-	-	-	+	-	-	-	-	-
35	+	-	-	+	*	-	-	-	-	-	+	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	+	-	-	-	-
36	-	+	-	+	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	+	-	*	-	-	-	-
27	-	+	-	-	+	-	-	-	-	-	*	-	-	-
07	-	+	-	-	-	-	-	-	-	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

PM19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	2	0	1	2	1	1	1	2	0	1	0	3	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	+	-
34	-	*	-	-	-	-	-	-	+	-	-	-	+	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	+	-	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	+
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	+	-
36	-	+	-	-	-	-	-	-	*	-	+	-	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	+	-	*	-	-	-
07	-	-	-	-	-	-	-	-	-	-	-	*	-	-
10	+	+	-	-	-	-	-	+	-	-	-	-	*	-
28	-	-	-	-	+	-	-	-	-	-	-	-	-	*

R17

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	2	0	1	2	4	2	1	1	1	3	0	1	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	+	-	-	+	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	+	-	-	-
35	-	-	-	-	*	+	+	-	-	-	-	-	-	-
23	-	+	-	-	+	*	+	-	-	-	+	-	-	-
22	-	-	-	-	+	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	+	-
36	-	+	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	*	+	-	-	-
27	-	-	-	+	-	+	-	-	-	+	*	-	-	-
07	-	-	-	-	-	-	-	-	-	-	-	*	-	-
10	-	-	-	-	-	-	-	+	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	3	0	2	1	3	5	1	2	0	0	1	2	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	+	-	-	-	+	-	-	-	-	-	-	-	-
34	+	*	-	-	-	-	+	-	+	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	+	-	+	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	+	-	-	-	-	*	+	-	-	-	-	-	+	-
22	-	+	-	+	-	+	*	-	-	-	-	-	+	+
06	-	-	-	-	-	-	-	*	-	-	-	+	-	-
36	-	+	-	-	-	-	-	-	*	-	-	-	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	+	-	-	-	*	-	-
10	-	-	-	-	-	+	+	-	-	-	-	-	*	-
28	-	-	-	-	-	-	+	-	+	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	3	0	4	2	1	4	0	2	1	2	1	0	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	+	-	-	+	-	-	+	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	+	-	*	+	-	+	-	-	-	+	-	-	-
35	-	-	-	+	*	-	-	-	-	-	+	-	-	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	+	-	+	-	+	*	-	+	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	+	-	*	-	-	+	-	-
11	-	+	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	+	+	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	1	2	4	5	2	2	4	6	5	5	6	0	6

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	+
21	-	-	*	-	-	-	-	+	-	+	-	-	-	-
20	-	-	-	*	+	+	+	-	-	-	-	-	-	+
35	-	-	-	+	*	-	-	-	+	-	+	+	-	+
23	-	-	-	+	-	*	+	-	-	-	-	-	-	-
22	-	-	-	+	-	+	*	-	-	-	-	-	-	-
06	-	-	+	-	-	-	-	*	+	+	-	+	-	-
36	-	-	-	-	+	-	-	+	*	+	+	+	-	+
11	-	-	+	-	-	-	-	+	+	*	+	+	-	-
27	-	-	-	-	+	-	-	-	+	+	*	+	-	+
07	-	-	-	-	+	-	-	+	+	+	+	*	-	+
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	+	-	+	+	-	-	-	+	-	+	+	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	1	0	0	1	0	1	0	1	0	0	1	0	0

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	+	-	-	-	-	-	-	-	-	-	-	-	-
34	+	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	+	-	-	-	-	-	-	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	+	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	*

LS_Italy

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	4	4	5	5	2	6	4	0	0	2	5	3	1	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	+	+	+	-	-	-	-	-	-	+	-	-
34	-	*	+	-	-	+	+	-	-	-	-	+	-	-
21	+	+	*	-	-	+	+	-	-	-	+	-	-	-
20	+	-	-	*	-	+	+	-	-	-	+	-	+	-
35	+	-	-	-	*	+	-	-	-	-	-	-	-	-
23	-	+	+	+	+	*	+	-	-	-	+	-	-	-
22	-	+	+	+	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	*	+	-	-	+
27	-	-	+	+	-	+	-	-	-	+	*	-	-	+
07	+	+	-	-	-	-	-	-	-	-	-	*	-	+
10	-	-	-	+	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	+	+	+	-	*

LS_Frankfurt

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	1	3	0	4	3	2	3	0	6	3	3	4	3	5

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	+	-
34	-	*	-	-	+	+	-	-	-	-	-	+	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	+	+	+	-	+	-	-	-	-	-
35	-	+	-	+	*	-	-	-	+	-	-	-	-	-
23	-	+	-	+	-	*	-	-	-	-	-	-	-	-
22	-	-	-	+	-	-	*	-	-	-	-	+	-	+
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	+	+	-	-	-	*	+	+	+	-	+
11	-	-	-	-	-	-	-	-	+	*	+	-	+	-
27	-	-	-	-	-	-	-	-	+	+	*	-	-	+
07	-	+	-	-	-	-	+	-	+	-	-	*	-	+
10	+	-	-	-	-	-	-	-	-	+	-	-	*	+
28	-	-	-	-	-	-	+	-	+	-	+	+	+	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	2	0	0	0	2	2	0	1	3	0	1	3	3	3

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	+	-	-	-	+	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	*	-	-	-	-	-	-	-	-	-	-
35	-	-	-	-	*	-	-	-	+	-	-	-	-	+
23	-	-	-	-	-	*	-	-	-	-	-	+	+	-
22	-	-	-	-	-	-	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	+	-	-
36	+	-	-	-	+	-	-	-	*	-	-	-	-	+
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	*	-	-	+
07	-	-	-	-	-	+	-	+	-	-	-	*	+	-
10	+	-	-	-	-	+	-	-	-	-	-	+	*	-
28	-	-	-	-	+	-	-	-	+	-	+	-	-	*

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	0	1	2	1	3	2	0	1	2	0	1	0	1

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	-	-	-	-	-	-
21	-	-	*	-	-	-	+	-	-	-	-	-	-	-
20	-	-	-	*	+	+	-	-	-	-	-	-	-	-
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	+	-	*	+	-	-	+	-	-	-	-
22	-	-	+	-	-	+	*	-	-	-	-	-	-	-
06	-	-	-	-	-	-	-	*	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	*	-	-	+	-	-
11	-	-	-	-	-	+	-	-	-	*	-	-	-	+
27	-	-	-	-	-	-	-	-	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	+	-	-	*	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	*	-
28	-	-	-	-	-	-	-	-	-	+	-	-	-	*

LS_HG18

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	3	2	3	1	0	1	3	1	1	2	0	1	4

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	+	+	-	-	-	+
21	-	-	*	-	-	-	-	+	-	-	-	-	+	-
20	-	-	-	*	+	-	-	-	-	-	+	-	-	+
35	-	-	-	+	*	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	*	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	*	-	-	-	-	-	-	+
06	-	-	+	-	-	-	-	*	-	-	+	-	-	+
36	-	+	-	-	-	-	-	-	*	-	-	-	-	-
11	-	+	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	-	+	-	-	-	+	-	-	*	-	-	-
07	-	-	-	-	-	-	-	-	-	-	-	*	-	-
10	-	-	+	-	-	-	-	-	-	-	-	-	*	-
28	-	+	-	+	-	-	+	+	-	-	-	-	-	*

LS_HG19

Table of the number of linked loci per locus

Locus	14	34	21	20	35	23	22	06	36	11	27	07	10	28
	0	2	1	3	5	1	3	4	1	0	5	1	2	2

Table of significant linkage disequilibrium (significance level=0.0500):

	14	34	21	20	35	23	22	06	36	11	27	07	10	28
14	*	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	*	-	-	-	-	-	-	+	-	-	-	-	+
21	-	-	*	-	-	-	-	-	-	-	+	-	-	-
20	-	-	-	*	+	-	-	+	-	-	+	-	-	-
35	-	-	-	+	*	-	-	+	-	-	+	+	+	-
23	-	-	-	-	-	*	+	-	-	-	-	-	-	-
22	-	-	-	-	-	+	*	+	-	-	-	-	-	+
06	-	-	-	+	+	-	+	*	-	-	+	-	-	-
36	-	+	-	-	-	-	-	-	*	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	*	-	-	-	-
27	-	-	+	+	+	-	-	+	-	-	*	-	+	-
07	-	-	-	-	+	-	-	-	-	-	-	*	-	-
10	-	-	-	-	+	-	-	-	-	-	+	-	*	-
28	-	+	-	-	-	-	+	-	-	-	-	-	-	*

A.2   Materials

A.2.1. Chemicals

Table A.2.1.1: Chemicals used in this study.

Substance	Supplier	Catalog number
Agar-Agar, bacteriological	Roth	2267.5
dNTP Mix (2 mM each)	life technology	R0242
Quick-Load® Purple 100 bp DNA Ladder	New England Biolabs	N0550L
SYBR® Safe DNA Gel Stain	Invitrogen	S33102
Purple Gel Loading Dye (6X)	New England Biolabs	B7025
Tris	Roth	AE15.3
NaCl	Roth	9265.2
EDTA	Sigma-Aldrich	E5134
Spermine tetra-HCl	Sigma-Aldrich	S1141
Spermidine tri-HCl	Sigma-Aldrich	85580
Sucrose	Roth	4621.1
KoAc	Roth	P1190
SDS	Roth	0183.1
Ethanol	Roth	9065.2
Water Molecular Biology Reagent	Sigma	W4502-1L
Fadenagar	Brecht	00262-0500
KCl	Roth	6781.1
Bierhefe	Ramspeck	210099K
Malzin	CSM	4002715.72898.5
Propionsäure	Roth	6026.1

A.2.2. Consumables

Table A.2.2.1: Consumables used in this study.

Material	Supplier	Catalog number
Lysing Matrix D ceramic beads	Fisher Scientific	11442420
Gloves TouchNTuff®	Ansell	92-600
Parafilm® M	Roth	H666.1
KIMTECH® Science Präzisionstücher	Roth	AA64.1
Sterilfilter PVDF 0,22µM	Roth	P666.1
Pierceable Foil Heat Seal	BioRad	1814040
PCR SingleCap 8er-SoftStrips 0.2 ml	Biozym	710970
96 well PCR plates	Biozym	
2,0ml Mikroröhre PCR-PT	Sarstedt	72693465
Drosophila Zuchtbehälter 175 ml	Greiner Bio-one/Th. Geyer/ VWR/neolab	960177
Foam stopper for drosophila vials, big	Greiner Bio-one/Th. Geyer/ VWR/neolab	332070
Drosophila vial PS, 50ml	NerbePlus	11-881-0051
Foam stopper for drosophila vial, small	NerbePlus	11-881-1000
Faltenfilter Papier	Roth	CA16.1
Cell Spreader	Heathrow Scientific	HS8151
Kultur-Röhrchen, PP-Röhrchen 95x18mm	Greiner bio-one (Kobe)	9401337
SafeSeal SurPhob Filterspitzen > 10 µl	Biozym	VT0200X
SafeSeal SurPhob Filterspitzen > 20 µl	Biozym	VT0220X
SafeSeal SurPhob Filterspitzen > 200 µl	Biozym	VT0240X
SafeSeal SurPhob Filterspitzen > 1250 µl	Biozym	VT0270X
SafeSeal Reagiergef. 2 ml	Sarstedt	72695500
SafeSeal Reagiergef. 1.5 ml	Sarstedt	72706

A.2.3. Devices

Table A.2.3.1: Devices used in this study.

Machine	Model	Company
Autoclave	5075 ELV	Tuttnauer
Autoclave	3850 EL	Tuttnauer
Balance	ABT 220-5DM	Kern
Balance	Excellence XA 1502 S	Mettler Toledo
Biosafety cabinet	ESCO Class II biosafety cabinet	biomedis
Centrifuge	Mikro 220R	Hettich
Centrifuge	Rotina 420R	Hettich
Dishwasher	Compact Desinfektor G7783 CD Mielabor	Miele
Electrophoresis Power Supply	EV 231	Consort
Gel Documentation Station	VersaDoc Imaging System 4000 MP	BioRad
Gel Electrophoresis Chamber		BioRad
Homogenizer	Fast Prep-24™	MP Biomedicals
Hotplate Stirrer	Hotplate Stirrer Model L-81	Labinco
Hotplate Stirrer	VMS-A	VWR
Ice Machine	AF 80	Scotsman
Incubator	Heraeus Oven	Thermo
Microplate Reader	Epoch	BioTek instruments
Microwave	Grill Hot Air	Sharp
Multichannel Pipette	Rainin Pipet-Lite XLS 2-20 µl	Mettler Toledo
	Rainin Pipet-Lite XLS 20-200 µl	Mettler Toledo
PCR Cycler	C1000 Thermal Cycler	BioRad
pH-Meter	Seven Multi	Mettler Toledo
Pipette	Eppendorf Research Plus 0.1-2.5 µl	Eppendorf
	Eppendorf Research Plus 2-20 µl	Eppendorf
	Eppendorf Research Plus 10-100 µl	Eppendorf
	Eppendorf Research Plus 20-200 µl	Eppendorf
	Eppendorf Research Plus 100-1000 µl	Eppendorf
Plate sealer	BioRad PX1	BioRad
Purified Water System	TKA-GenPure	Thermo
Refrigerator	MPR 1411PE	Panasonic
Shake Incubator	Multitron II	Infors HAT
Shaker	Rocker 25	Labnet
Vortex	VV3	VWR
Water bath	Microprocessor control MPC	Huber
Microscope	VHX 5000	Keyence

A.2.4. Oligonucleotide Primers

Table A.2.4.1: Oligonucleotide primers used for PCR. Shown is the application the primer was used for in the experiment, the name and the sequence of the primer, the melting temperature in °C, the primer ID that is used in the laboratory primer list, and the manufacturer.

Application	Name	Sequence	Tm (°C)	Number	Manufacturer
Positive control	ARP1 fwd	GATTCGCCATGCCTCACAG	57 – 61°C	P544	IDT
	ARP1 rev	CTTGATGTTAGCTGACACAAGG	57 – 61°C	P545	IDT
Sequencing	M13F	TGTAAAACGACGGCCAGT	53.7 °C	MFS13	IDT
	M13R	AGGAAACAGCTATGACCAT	52.4°C	MFS14	IDT

Table A.2.4.2: Oligonucleotide primers used for Multiplex-PCR. Given are the primer name, the fluorescent dye used to label the 5' end of the primer, the primer sequence, the corresponding unmodified reverse primer, and its sequence, as well as the size range and the repeated motif of the microsatellite given by Fraimout et al. (2015). All labeled primers were manufactured by Metabion International AG (Planegg, Germany). The unmodified reverse primer was manufactured by IDT (Leuven, Belgium). Columns marked grey were excluded from further experiments.

Primer	5'Modification	Primer sequence 5'--> 3'	Unmodified rev Primer	Rev Primer sequence	Size range in bp	Repeat motif
DS06	6-Fam	CGGTTCGAGTGCTTGTTAGA	P1079	ACACGTGGAGGACACCTTC	130-190	(TG)11
DS26	Hex	CCTGTGTGCATCTCAGTGTT	P1101	TACAGCACTCCAGCACATGA	60-130	(CA)10
DS36	Hex	TTGGCAACGTGTGAAGCTG	P1111	GAGACACTGCAATGCTGCCT	160-220	(GT)13
DS11	Hex	CGGTGACTCGTGCAGTTGTA	P1083	GCCGACTCTGTCTAGAGCAA	230-290	(CA)11
DS28	Rox	TTAAGCTGACCTCCTCCTCG	P973	GCACTCGCACAGATACAAGG	140-195	(TG)11
DS07	Tamra	AAGGCTGGAGTGGCAACAA	P961	GCTAAGGTTCTGTTCGGCTG	160-210	(CA)13
DS10	Tamra	CGAGACTGTGCGAACGAGAG	P1081	CATATGCTGACTGCCTCACA	270-330	(GT)11
DS27	Tamra	CCAGCGACTGCAGAAGTGAC	P1103	GCAATCCTCCACAACACAAC	80-130	(GT)14
DS14	6-Fam	AAGAACCGCAACGAGCAA	P967	GAATTATCCAGCGACACGAC	180-220	(TG)10
DS20	Tamra	CAGCCATATGCAATGCACTG	P1091	ATATCCAGCGGAAGTCGAGA	210-270	(AG)12
DS21	Hex	GAGACGCGATGGTACCGTTA	P1093	CCAATCGAGTGCAAGCGT	310-370	(AC)11
DS22	Rox	TACAGATACGCCGTCGGATT	P1095	AAGACCAAGACGACGGACCT	290-360	(GT)11
DS23	Rox	TGCCACTAAGCTCACACGGT	P1097	CAGTTGCCACTTGCTGTGTA	237-300	(AC)10
DS32	6-Fam	CGGCGTGTTGCAGTTATTC	P975	ATGCACTGGTCGACATGACA	330-380	(TG)15
DS34	6-Fam	AACAACGACGCAGAAGCTCA	P1107	CGACTGTTGCGCTCTGTAAT	240-300	(GA)12
DS35	Rox	TCCGTATTCCGTATCCGTGT	P1109	GGAGTATGGCAGTGTGGCAG	198-240	(CA)11
DS17	6-Fam	CATCTCAGGCCACGAATG	P1087	CTCCAGATTCTCGAGTGCAG	80-130	(GT)10

A.2.5. Restriction Enzymes

Table A.2.5.1: Restriction enzymes.

Enzyme	Supplier	Catalog number
EcoR I-HF	New England Biolabs	R3101S

A.2.6. Selection Media and Plates

Table A.2.6.1: Components for *E. coli* selective media and agar plates.

Medium	Supplier	Catalog Number
LB-Medium	Roth	6673.1
LB-Agar 1,5%	Roth	2266.2
Ampicillin	Roth	HP62.1

A.2.7. Buffers

Table A.2.7.1: Composition of buffers used in this project.

Homogenization buffer	
200 µl	1 M Tris-HCl (pH 7.5)
240 µl	5 M NaCl
400 µl	0.5 M EDTA (pH 8)
83.3 µl	0.3 M Spermine tetra-HCl
25 µl	1 M Spermidine tri-HCl
1 g	Sucrose
Add 20 ml	H_2O
Lysis buffer	
6 ml	1 M Tris-HCl (pH 9)
4 ml	0.5 M EDTA (pH 8)
1,25 ml	10% SDS
1 g	sucrose
Add 20ml	H_2O
TAE buffer	
242 g	Tris base
57.1 ml	Pure acetic acid
100 ml	0.5 M EDTA (pH 8)
Add 1 L	H_2O
TE buffer	
10 ml	1 M Tris-Cl (pH 7.5)
2 ml	500 mM EDTA (pH 8)

A.2.8. Stocks

Table A.2.8.1: Antibiotics and their applied and stock concentrations.

Reagent	Stock Concentration	Solvent	Applied Concentration
Ampicillin	100 mg/ml	De-ionized water	100 u/ml
Kanamycin	100 mg/ml	De-ionized water	100 u/ml

A.2.9. Kits

Table A.2.9.1: Kits used in this study.

Kit	Supplier	Catalog Number
Dream Taq DNA polymerase	Thermo Scientific	EP0701
Mix & Go! E. coli Transformation Kit and Buffer Set	Zymo Research	T3001
Multiplex PCR plus Kit	Qiagen	206151
NucleoSpin Plasmid Kit	Macherey-Nagel	740588250
Platinum Taq DNA Polymerase	Invitrogen	10966-018
Q5 High-Fidelity DNA polymerase	New England Biolabs	M0491S
TOPO™ TA Cloning™ Kit	Invitrogen	450030
Zero Blunt TOPO PCR Cloning Kit	Invitrogen	450031
Zymo Clean & Concentrator-25	Zymo Research	D4034
Zymo Clean & Concentrator-5	Zymo Research	D4014
Zymo Clean Gel DNA Recovery	Zymo Research	D4008

A.2.10. Plasmids and Vector Maps
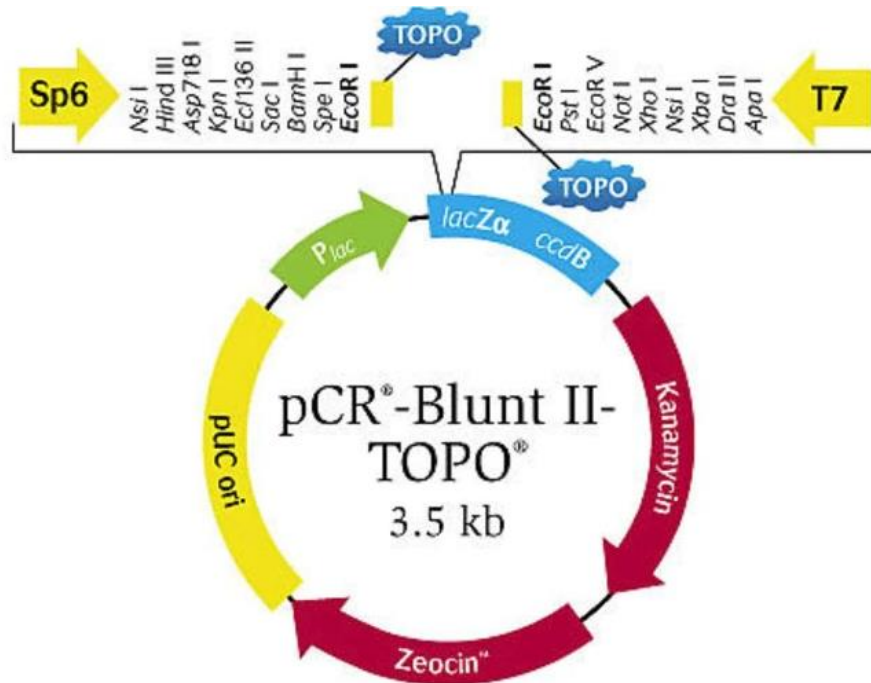
pCR®-Blunt II-TOPO® vector
Zero Blunt® TOPO® PCR Cloning kits contain pCR™-Blunt II-TOPO vector for use with proofreading enzyme and ccdB gene positive selection.

Figure A.2.10.1: pCR®-Blunt II-TOPO® vector map. Allows Zero Blunt TOPO for Sequencing and the direct insertion of blunt-ended PCR products amplified with proofreading thermostable polymerases into a plasmid vector (https://tools.thermofisher.com/content/sfs/vectors/pcrbluntiitopo_map.pdf).
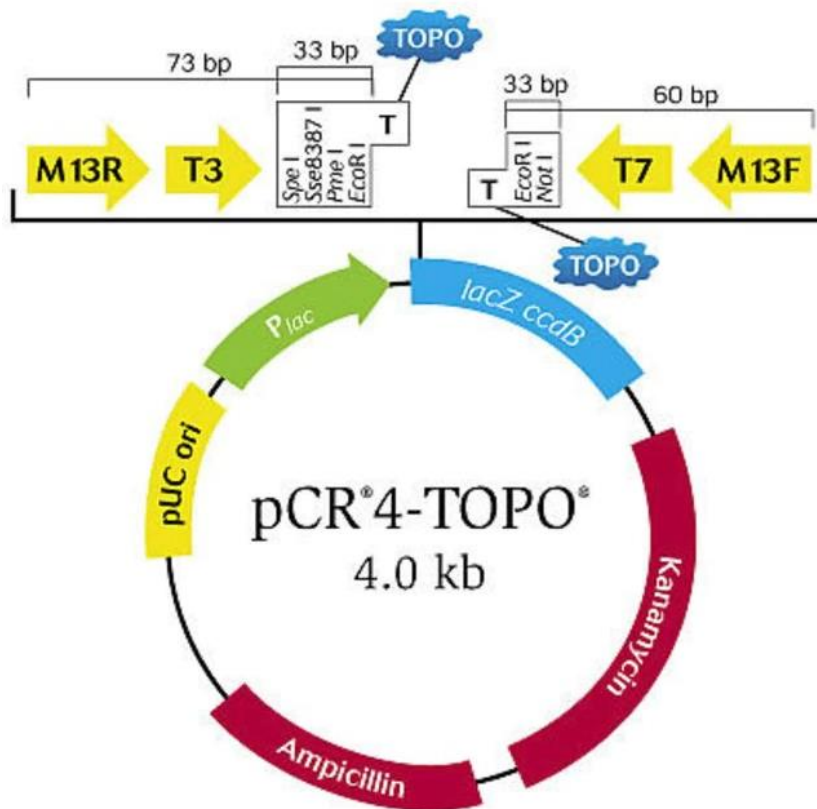
pCR®4-TOPO® vector
TOPO® TA Cloning® Kits for Sequencing contains pCR™4 TOPO® TA vector with shortened distance between sequencing primer sites and the insert site, allowing for less vector sequencing and more insert sequence.

Figure A.2.10.2: pCR®4-TOPO® vector map (Invitrogen). Allows TOPO TA Cloning for Sequencing and the direct ligation of Taq-amplified PCR products (https://tools.thermofisher.com/content/sfs/vectors/pcr4topo_map.pdf).

B   Laboratory Protocols

B.1 Workflow for Fragment Length Analysis (FLA)

1. Extraction of total genomic DNA from single individuals using the Maryland protocol (see B.2 Extraction of Genomic DNA – Maryland Protocol.
2. Multiplex PCR with the QIAGEN Multiplex Plus Kit (see B.4.4 QIAGEN® Multiplex PCR Plus Kit). From now on the samples should not be exposed to light whenever possible (fluorescent-labeled primers are light-sensitive!) and samples should not be stored at - 20°C, rather at +4°C.
3. PCR purification with Zymo Clean and Concentrator-5 (see B.5 Protocol for PCR Purification).
4. Make a 3% Agarose gel to determine the concentration of purified multiplex PCR products (see B.6 Protocol for Gel Electrophoresis). A determination using a spectrophotometer is not possible due to the fluorescent-labeled forward primers. Based on the band intensity, samples might have to be diluted with HPLC H_2O. All samples sent to StarSeq should have the same concentration.
5. Pipette a total volume of 10 µl into a 96 well plate. This can be a 10 µl purified multiplex PCR product or a diluted sample. No additional reagents are needed, everything else is added by StarSeq.
   Note: StarSeq is charging for 48 or 96 samples, everything below 48 will still be charged the same price as the 48 samples and everything between 48 and 96 samples will be charged the same price as 96 samples.

   StarSeq requires the following additional information and files:
   A **signed** order confirmation (Figure B.1.1). This has to be printed out and added to the envelope with the samples. You can add it to the E-Mail as well.

Figure B.1.1: Order confirmation for StarSeq. This form has to be signed and send to StarSeq together with the samples. It should also be sent via E-Mail together with the gel pictures and the table containing the sample names.

An Excel sheet with pipetting scheme and sample names. You receive this Excel sheet from StarSeq (Figure B.1.2). Provide this file with the E-Mail, since it will be used to enter the sample names in the sequencer at StarSeq.

Figure B.1.2: Excel sheet provided from StarSeq. This table has to be sent via E-Mail to StarSeq. It has to contain the pipetting scheme on the 96 well plates and the sample names. It contains additional information like how the samples should be loaded if only 48 samples are on the plate and if the samples are ready to load or ready to mix and which dye set was used. For this project, the 'ready to mix' option had to be used and 'Dye Set DS-33'.

A file with the agarose gel pictures. The samples should be labeled. Information about the used ladder should be added (Figure B.1.3).

#	Sample Name	Position on 96 well plate
1.	K2MP3_HOH18004	A1
2.	K2MP3_HOH18005	B1
3.	K2MP3_HOH18006	C1
4.	K2MP3_HOH18007	D1
5.	K2MP3_HOH18008	E1
6.	K2MP3_HOH18009	F1
7.	K2MP3_HOH18010	G1
8.	K2MP3_HOH18011	H1
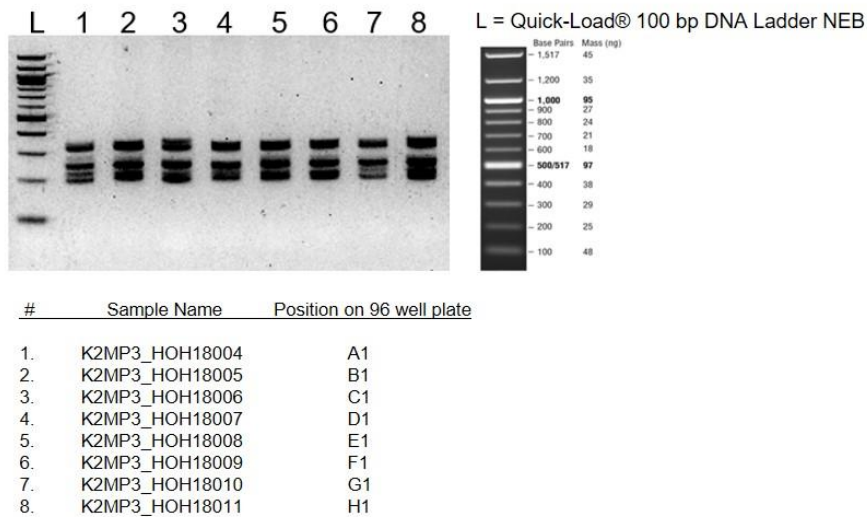
Figure B.1.3: Picture of a 3% agarose gel for multiplex PCR. The gel picture has to be labeled with ladder and sample names. The exact description of the used ladder is necessary. The sample names and the position of each sample on the 96 well plates are not required but should be added for the StarSeq employer preparing the samples. This can be either sent by E-Mail or/and with the samples.

6. Seal the well plate with a BioRad PX1 PCR Plate Sealer and BioRad Pierceable Foil Heat Seal (Catalog number 1814040) (BioRad, California; USA) and wrap it in tinfoil. The plate can be stored at +4°C for a short period, for example over the weekend.
7. Use a patted envelop, add postage stamps, and send the samples by mail.
8. You will receive FLA results in 1-2 weeks. Continue processing FLA results with bioinformatics tools (see C Bioinformatic Protocols for Microsatellite Analysis).

B.2 Extraction of Genomic DNA – Maryland Protocol

1. Put the insect in a tube filled with ceramic beads and add 200 µl DNA homogenization buffer. Recommended are tubes with a screw cap
2. Homogenize the sample in the homogenizer Fast Prep-24™. with the program 6,000 rpm for 40 sec. If the sample look foamy, centrifuge for a few seconds until no foam remains
3. Add 200 µl DNA lysis buffer and mix gently by inverting. Incubate at 70°C for 10 min
4. Centrifuge the tubes to get rid of vapor and allow the tubes to cool to RT
5. Add 60 µl 8 M KoAc and vortex briefly to mix. Place tubes on ice for 30 min (or longer)
6. Transfer the whole suspension to a new tube
7. Centrifuge at 12,000 rpm for 10 min at 4°C. Carefully transfer 300 µl of the suspension to a new tube and precipitate the DNA with two volumes of ice-cold 100% ethanol (- 20°C), mix the tube well by inverting
8. Allow the DNA to precipitate at -20°C for at least 1 h. Precipitating longer (for example overnight) does not harm and can be beneficial. Centrifuge at 12,000 rpm (or max) at 4°C for at least 40 min to pellet the DNA. Longer centrifugation does not harm the sample and can result in higher DNA concentrations
9. Remove the supernatant and wash the DNA pellet with 300 µl 70% Ethanol (can be ice-cold but not necessarily important) for 10 min while centrifuging at 12,000 rpm (or max) at 4°C
10. Remove the supernatant and air-dry the pellet. Resuspend the pellet in 50 µl 1xTE (pH 7.5) for longer storage (-20°C) and H_2O for direct use (+4°C). **For further use in FLA only use H_2O**! TE can falsify the result!
11. Check DNA concentration and adjust to 50 ng / µl

For the composition of buffers used in the protocol see A.2.7. Buffers.

B.3 Spectrophotometric DNA Quantification

When using the microplate reader, always be extremely careful with the equipment. Do not use Ethanol to clean the microplate! Only use ddH_2O and soft tissues!

1. Switch on the computer and log into the 'Arbeitsgruppe Schetelig' account.
2. Switch on the plate reader (Figure B.3.1).
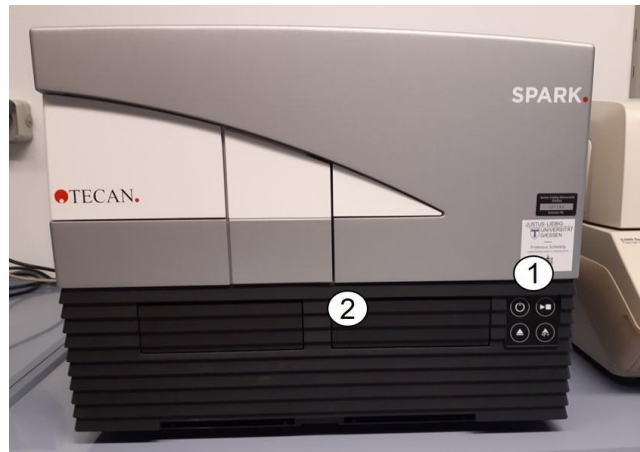
Figure B.3.1: Picture of Tecan Spark Multimode Plate Reader. For spectrophotometric DNA and RNA quantification, the machine has to be switched on (1). When everything is ready and set up for analysis, the door (2) will open automatically and the guiding rail for the NanoQuant PlateTM will move into position.

3. Open the 'Tecan SparkControl Dashbord' app (Figure B.3.2).

Figure B.3.2: Starting the Tecan Spark app and saving results. To start the application, click on the Tecan Spark App shortcut on the desktop (1). The program will start (2). The Shortcut for Analysis results in which the results are automatically saved as Excel sheets can be found on the desktop as well (3).

4. In the SparkControl app open 'Nucleic Acid Quantification' (Figure B.3.3).
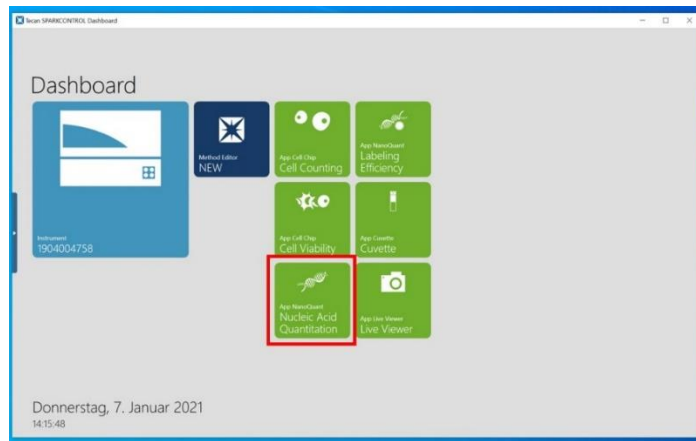
Figure B.3.3: Tecan Spark program. For DNA and RNA quantification click on the button 'Nucleic Acid Quantification' (red square).

5. Tell the program which spots on the plate will contain a sample and should therefore be measured. In addition, you can name your samples (Figure B.3.4).
6. Make sure that the right mode is activated. For genomic DNA the option 'dsDNA' has to be used. Other options are 'RNA' and 'ssDNA'.
7. Click on 'Individual Blanking' and then on 'Start'.
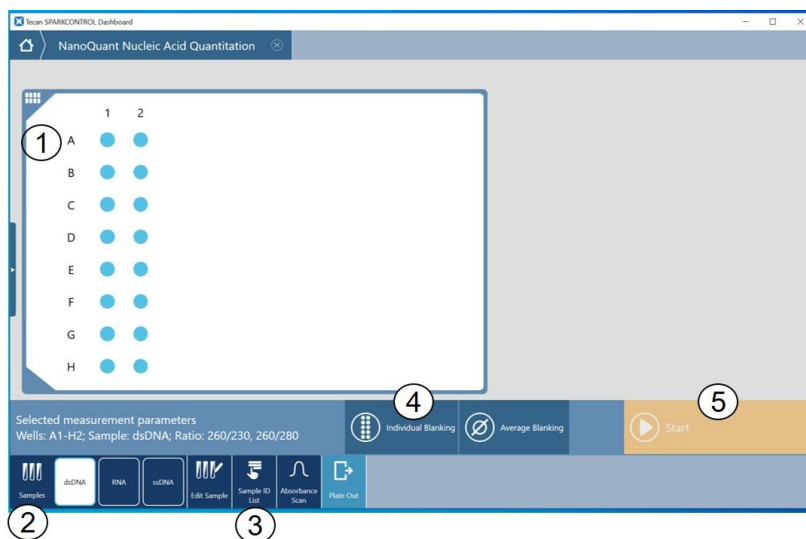
Figure B.3.4: User interface for DNA/RNA quantification. Mark or unmark all wells that are used (1). Blue means the plate reader will analyze this spot on the plate. Unmarked spots are grey and will not be analyzed. Tell the machine which type of sample to measure (2). You can choose between 'dsDNA', 'RNA', and 'ssDNA'. You can name your samples before analysis (3). The Excel sheet with the results will then contain the names of your samples. It is advised to use the 'Individual Blanking' option (4) before analysis. Start blanking by clicking on 'Start' (5).

8. The door of the microplate reader will open automatically.
9. Prepare the plate. Open the plate carefully and clean the surface with ddH$_2$O on very soft tissue. Do not scratch the surface (Figure B.3.5).

Figure B.3.5: NanoQuant Plate™. The NanoQuant™ Plate for Tecan Spark Multimode Plate Reader allows the analysis of up to 16 samples in parallel. Samples have to be pipetted on one of the 16 spots. The plate is optimized for a 2 µl sample volume.

10. For each sample that will be measured, blanking with a single 2 µl droplet has to be done upfront. Add 2 µl of buffer or water to the microplate for each sample. If 1xTE was used after DNA extraction, then 1xTE has to be used for blanking. If elution buffer from a kit was used to elute the DNA, then the same elution buffer should be used for blanking.
11. Close the lid of the microplate carefully. Do not let the lid snap, place it gently on the microplate.
12. Put the closed plate into the plate reader.
13. Press 'Start' to start the blanking.
14. After the machine is done, it will open the door and you can remove the plate from the slider.
15. Clean the plate as described in step 9.
16. Add 2 µl of each sample to the spots on the plates, close the lid and place it back in the machine.
17. Click 'Start'.
18. The slider will slide into the machine automatically and the door closes.
19. After the measurement is completed the door will open automatically and the plate can be removed again.
20. Clean the plate again and place it back in the metal case.
21. The Excel file that opened is saved automatically in the folder 'Tecan Excel Daten' (shortcut on the desktop).

Interpretation of the result

DNA and RNA absorb light at 260 nm, which is the measurement of total nucleic acid. Nucleic acid samples are also measured at 280 nm, which is the absorbance peak for proteins. The ratio of these two measurements allows an interpretation of the purity of the nucleic acid, with a value near 2 indicating a highly pure DNA or RNA sample.

B.4 Protocols for Polymerase Chain Reaction (PCR)

B.4.1 Invitrogen™ Platinum™ *Taq* DNA Polymerase

Catalog number: 10966-018 (120 rxns)

This protocol shows the PCR procedure for a single 50 µl reaction. For multiple reactions prepare a master mix of components common to all reactions to minimize pipetting error. Prepare sufficient master mix for the number of reactions plus one extra. Dispense appropriate volumes into each 0.2 ml PCR tube before adding template DNA and primers.

1. Thaw, mix and briefly centrifuge each component except for the enzyme. The enzyme remains in the -20°C freezer until it is needed.
2. Add the following components to each PCR tube:

Component	50 µl rxn
Water, nuclease-free	to 50 µl
10X PCR Buffer, -Mg	5 µl
50 mM MgCl2	1.5 µl
10 mM dNTP Mix	1 µl
10 µM forward Primer	1 µl
10 µM reverse Primer	1 µl
Template DNA	<500ng / rxn

3. Mix and briefly centrifuge.
4. Add enzyme to each tube for a final reaction volume of 50 µl.

Component	50 µl rxn
Platinum Polymerase	0.2 µl

5. Incubate the reactions in a thermal cycler. For a PCR machine with heated lid use 105°C (Figure B.4.1).

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	2 min	1
Denaturation	95°C	30 sec	35
Annealing	Tm-5	30 sec	
Extension	72°C	1:30 min	
Final Extension	72°C	5 min	1
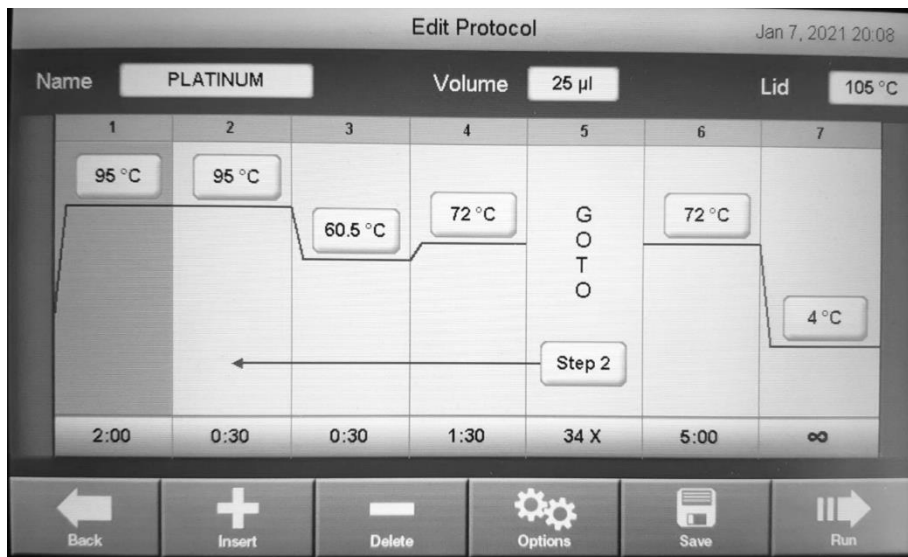
Figure B.4.1: Cycling parameter for Platinum™ *Taq* DNA Polymerase. The picture was taken from the BioRad C1000 touch (BioRad, California; USA) thermal cycler used in this project. It shows the different steps in the protocol as well as temperature, time and number of cycles. Step 7 in this image is the additional cooling at the end. It can also be set to 12°C, depending on the machine. This option should not be strained for too long, since it can harm the machine.

6. Use the PCR product immediately or store it at -20°C.

B.4.2 Thermo Scientific Dream *Taq* DNA Polymerase

Catalog number: #EP0701 (200 U Polymerase and 1.25 ml 10x Dream Taq buffer)

This protocol shows the PCR procedure for a single 50 µl reaction. For multiple reactions prepare a master mix of components common to all reactions to minimize pipetting error. Prepare sufficient master mix for the number of reactions plus one extra. Dispense appropriate volumes into each 0.2 ml PCR tube before adding template DNA and primers.

1. Thaw, mix and briefly centrifuge each component except for the enzyme. The enzyme remains in the -20°C freezer until it is needed.
2. Add the following components to each PCR tube:

Component	50 µl rxn
Water, nuclease-free	to 50 µl
10X PCR Buffer, +Mg	5 µl
2 mM dNTP Mix	5 µl
10 µM forward Primer	0.5 µl
10 µM reverse Primer	0.5 µl
Template DNA	10 pg – 1 µg

3. Mix and briefly centrifuge.
4. Add enzyme to each tube for a final reaction volume of 50 µl.

Component	50 µl rxn
Dream Taq Polymerase	0.25 µl

5. Incubate the reactions in a thermal cycler. For a PCR machine with heated lid use 105°C (Figure B.4.2).

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	3 min	1
Denaturation	95°C	30 sec	35
Annealing	Tm-5	30 sec	
Extension	72°C	1 min	
Final Extension	72°C	5 min	1
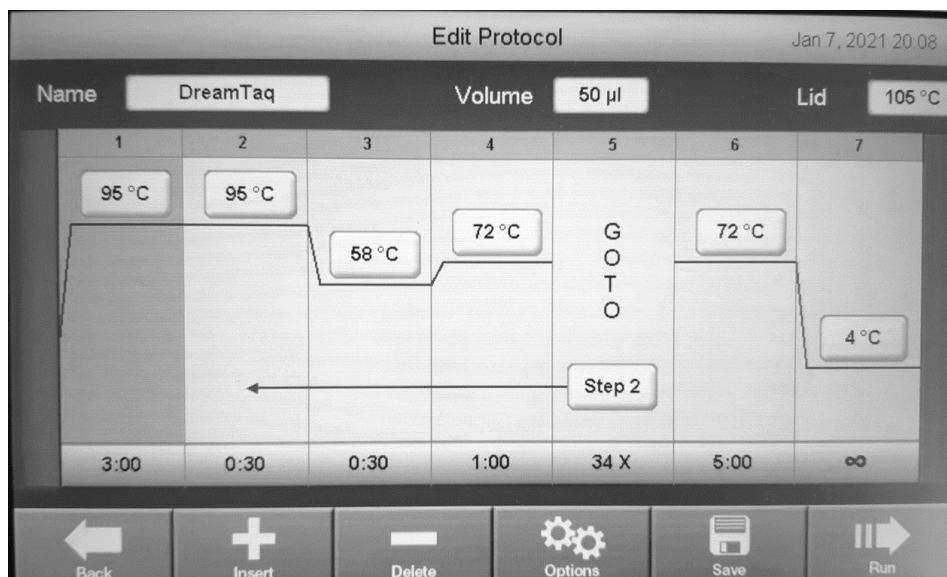
Figure B.4.2: Cycling parameter for Dream Taq DNA polymerase. The picture was taken from the BioRad C1000 touch (BioRad, California; USA) thermal cycler used in this project. It shows the different steps in the protocol as well as temperature, time and number of cycles. Step 7 in this image is the additional cooling at the end. It can also be set to 12°C, depending on the machine. This option should not be strained for too long, since it can harm the machine.

6. Use the PCR product immediately or store it at -20°C.

B.4.3 Q5™ High-Fidelity DNA Polymerase

Catalog number: M0491S

This protocol shows the PCR procedure for a single 50 µl reaction. For multiple reactions prepare a master mix of components common to all reactions to minimize pipetting error. Prepare sufficient master mix for the number of reactions plus one extra. Dispense appropriate volumes into each 0.2 ml PCR tube before adding template DNA and primers.

1. Thaw, mix and briefly centrifuge each component except for the enzyme. The enzyme remains in the -20°C freezer until it is needed.
2. Add the following components to each PCR tube:

Component	50 µl rxn
Water, nuclease-free	to 50 µl
5X PCR Buffer	10 µl
10 mM dNTP Mix	1 µl
10 µM forward Primer	2.5 µl
10 µM reverse Primer	2.5 µl
Template DNA	<1,000 ng

163

3. Mix and briefly centrifuge
4. Add enzyme to each tube for a final reaction volume of 50 µl.

Component	50 µl rxn
Dream Taq Polymerase	0.2 µl

5. Incubate the reactions in a thermal cycler. For a PCR machine with heated lid use 105°C (Figure B.4.3).

Step	Temperature	Time	Number of Cycles
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	35
Annealing	Tm-5	20 sec	
Extension	72°C	2.30 min	
Final Extension	72°C	5 min	1
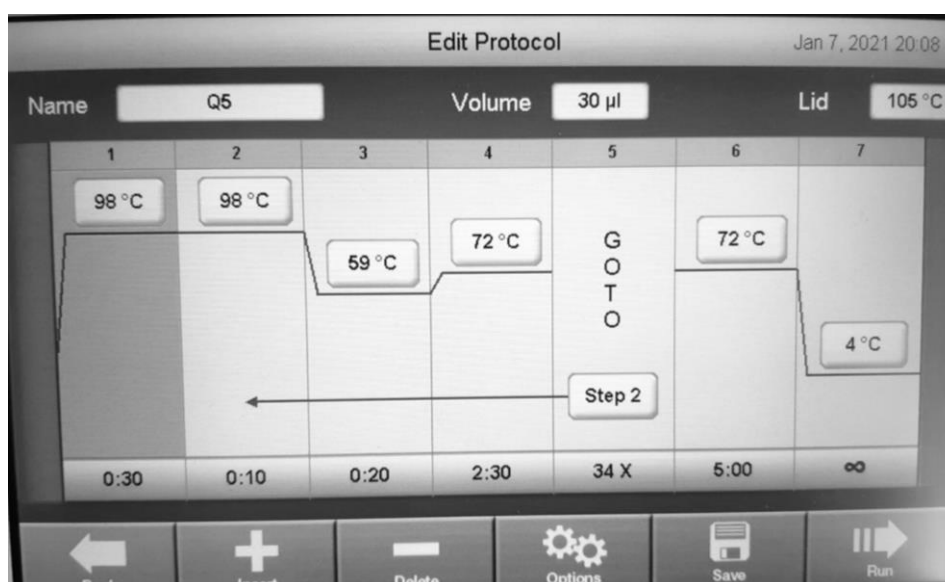
Figure B.4.3: Cycling parameter for Q5™ High-Fidelity DNA polymerase. The picture was taken from the BioRad C1000 touch (BioRad, California; USA) thermal cycler used in this project. It shows the different steps in the protocol as well as temperature, time and number of cycles. Step 7 in this image is the additional cooling at the end. It can also be set to 12°C, depending on the machine. This option should not be strained for too long, since it can harm the machine.

6. Use the PCR product immediately or store it at -20°C.

B.4.4 QIAGEN® Multiplex PCR Plus Kit

Catalog number: 206151

This protocol shows the PCR procedure for a single 10 µl reaction. For multiple reactions prepare a master mix of components common to all reactions to minimize pipetting error. Prepare sufficient master mix for the number of reactions plus one extra. Dispense appropriate volumes into each 0.2 ml PCR tube before adding template DNA and primers.

1. Thaw, mix and briefly centrifuge each component.
2. Add the following components to each PCR tube:

Component	10 µl rxn
Water, nuclease-free	to 10 µl
2X PCR Multiplex Master Mix	5 µl
5x Q-Solution	1 µl
10x primer mix, 2 µM each	5 µl

3. Mix and briefly centrifuge
4. Add template DNA (100 ng) to the individual PCR tubes
5. Incubate the reactions in a thermal cycler. For a PCR machine with heated lid use 105°C (Figure B.4.4).

Step	Temperature	Time	Number of Cycles
Initial denaturation	95°C	5 min	1
Denaturation	95°C	30 sec	30
Annealing	57°C	90 sec	
Extension	72°C	3 min	
Final Extension	72°C	30 min	1
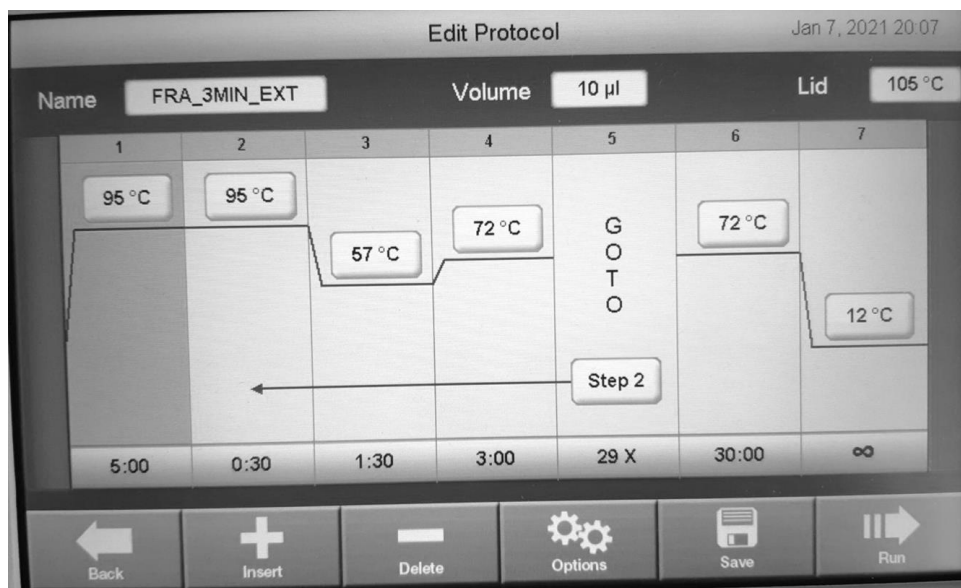
Figure B.4.4: Cycling parameter for QIAGEN® Multiplex PCR Plus Kit. The picture was taken from the BioRad C1000 touch (BioRad, California; USA) thermal cycler used in this project. It shows the different steps in the protocol as well as temperature, time and number of cycles. Step 7 in this image is the additional cooling at the end. It can also be set to 12°C, depending on the machine. This option should not be strained for too long, since it can harm the machine.

6. Use the PCR product immediately or store it at -20°C. Wrap samples in tin foil and/or keep them in the dark.

Preparing the 10x Primer mix for QIAGEN Multiplex PCR

To achieve 100 µM stock concentration for primer stocks of 5' fluorescent-labeled forward oligonucleotides and the complementary unlabeled reverse primer, 1x TE was added, following the instructions from the manufacturers. Labeled forward primers are light-sensitive! After preparing the stock solution, tubes were wrapped in tin foil and stored in a cardboard box at - 20°C. The QIAGEN Multiplex Plus Kit requires a 10x primer mix (2 µM each) for successful amplification.

Add the following components for a 100 µl aliquot of a 2 µM primer mix containing a total of 16 primers (8 x 5'labeled forward primer and 8 x unlabeled reverse primer):

Component	2 µM primer mix
Each 100 µM primer stock	2 µl
HPLC-H_2O	68 µl

Primer combinations for multiplex primer mix suggested by Fraimout for his microsatellite markers (personal communication):

Kit1

Multiplex PCR 1: DS05, DS09, DS12, DS15, DS16, DS33

Multiplex PCR 2: DS08, DS17, DS19, DS25, DS38, DS39

Kit2

Multiplex PCR 3: DS14, DS34, DS32, DS35, DS22, DS20, DS21, DS23

Multiplex PCR 4: DS06, DS28, DS26, DS36, DS11, DS27, DS07, DS10

Fluorescent oligonucleotides were ordered for Kit2 Multiplex PCR 3 and Multiplex PCR 4 and used in the final experiments.

B.5 Protocol for PCR Purification

Zymo Clean and Concentrator™ -5 and Clean and Concentrator™ -25

Catalog number DCC-5 (D4014), DCC-25 (D4034)

DCC-5 and DCC-25 have the same protocol. The only difference between the kits is the column. DCC-5 can purify up to 5 µg DNA, DCC-25 can purify up to 25 µg DNA.

All centrifugation steps should be performed between 10,000 – 16,000 x g.

Heat a thermal block to 60°C for the elution buffer or the water in step 8. Heat another thermal block to 40°C for the incubation in step 9.

Do not touch the matrix of the column with the pipette tip!

Use a 1.5 ml microcentrifuge tube and add 2 – 7 volumes of DNA Binding Buffer to each volume of DNA sample. For 50 µl PCR product (Platinum, Q5, Dream Taq) use 250 µl DNA Binding Buffer, for 10 µl PCR product from multiplexing use 50 µl DNA Binding Buffer.

Application	DNA Binding Buffer : Sample	Example
Plasmid	2 : 1	200 µl : 100 µl
PCR product	5 : 1	500 µl : 100 µl
ssDNA	7 : 1	700 µl : 100 µl

1. Transfer the mixture to a provided Zymo-Spin™ column in a collection tube.
2. Centrifuge for 30 sec and discard flow-through.
3. Add 200 µl DNA Wash Buffer to the column, centrifuge for 30 sec, and discard flow-through.
4. Repeat the wash step.
5. After discarding the flow-through from the last wash step, place the column back into the (same) tube and centrifuge one more time for 30 sec. The additional centrifugation step should get rid of any wash buffer (and Ethanol) residues on the column.
6. Place the column in a fresh 1.5 ml microcentrifuge tube (not included)
7. Add >25 µl DNA Elution buffer or water (60°C in the thermal block) directly on the column matrix but do not touch with the tip of the pipette. **Use HPLC H$_2$O for FLA!** Elution buffer can falsify the result.
8. Incubate the tube for 2 – 3 min at 40°C.
9. Centrifuge for 30 sec to elute the DNA.

B.6 Protocol for Gel Electrophoresis

This protocol is written for a large 1% agarose gel. Other concentrations are achieved simply by adjusting the mass of the agarose in a given volume.

Pouring the gel:

1. Place a well comb in a large gel tray. The gel tray has to be tightly fixed in the gel caster to prevent the leaking of the liquid agarose.
2. Measure 1 g of agarose.
3. Mix the agarose with 100 ml of 1xTAE in a flask. A small gel in contrast would only fit 50 ml 1xTAE.
4. Microwave the flask until the agarose is completely dissolved. Be careful to not overboil the solution. It is better to microwave for 30 – 45 sec, carefully whirl the flask and continue microwaving for another 30 – 45 sec until the agarose is dissolved.
5. Let the agarose cool down a little bit and add 10 µl SYBR® Safe DNA Gel Stain (Invitrogen). For a small gel, you would need 5 µl SYBR® Safe. Mix by carefully swirling the flask.
6. Pour the gel in the gel tray with a well comb from step 1. Pour slowly to avoid spilling and bubbles. If you still have bubbles in your gel, use a pipette tip to burst them.
7. Let the gel sit at RT until it has solidified.

The gel can be left on the bench for a few hours without drying out. If it will not be used for a longer period, it can be stored in a plastic bag at 4°C. Remove the gel caster and place the gel together with the gel tray in a plastic bag. Add a few ml 1x TAE. Close the bag and make sure it is not leaking! That way it stays fresh for one or two more days. Nevertheless, it is advisable to always use a freshly made gel and to store it only if necessary.

Loading samples and running the gel:

1. Place the gel tray in a gel chamber filled with fresh 1xTAE. The gel has to be covered with buffer.
2. Add loading buffer to each sample. Mix 5 µl of the DNA sample with 1 µl Purple Gel Loading Dye (6X) (NEB). If you need more sample on your gel, for example, if you want to cut the bands for purification then you first have to make sure that you used a comb with big enough wells. Secondly, you simply use the desired amount of DNA sample and adjust the amount of used gel loading dye. You can either prepare this mixture in a small reaction tube or use parafilm. To do so, simply pipette the loading dye on the parafilm and mix it with your DNA sample by pipetting up and down. It is important to not release all of the sample on the parafilm, only use the first pressure point of the pipette. Another important point is, to space the loading dye droplets far enough from each other to prevent contamination between samples. The parafilm method will only work well with small loading volumes. If you use this method, you will have to mix the sample and loading dye and immediately load it on the gel.
3. Load a molecular weight ladder into the first lane of the gel. This ladder does not need to be mixed with the loading dye. Be careful to not stab the pipette tip into the gel. Release the ladder/sample slowly and steadily in the well. If all of the sample is loaded in the well, carefully raise the pipette out of the buffer
4. Run the gel at 90 – 120 V. This depends on the gel concentration and on how well it should look. For a low concentrated agarose gel (e.g. 1%) 90 V is more than enough. A higher voltage will let it run faster but the result will look not as nice and you have to be careful to not lose smaller fragments. Higher concentrated gels (e.g. 3%) will need a higher voltage to start with. A typical runtime is about 45 – 60 min. The dye run line should be no more than 80% of the way down the gel. If uncertain, stop the run after 30 min and check under UV. You can leave the gel in the gel tray since this is usually UV-transparent, and simply put the tray back into the gel chamber if additional runtime is needed.

5. When connecting the electrodes make sure to connect them the right way. Electrodes and cables are usually color-coded. Black is negative and red is positive. The DNA is negatively charged, so it will run to the positive electrode.
6. Turn on the power supply.
7. After runtime is finished turn the power device off and disconnect the electrodes. Remove the gel from the chamber.

Visualizing the DNA on the gel:

1. Start the computer and log into the 'AG Schetelig' account. Open the program 'INSTAS GelDoc'.
2. Open the door of the UV chamber by sliding it to the side (Figure B.6.1).
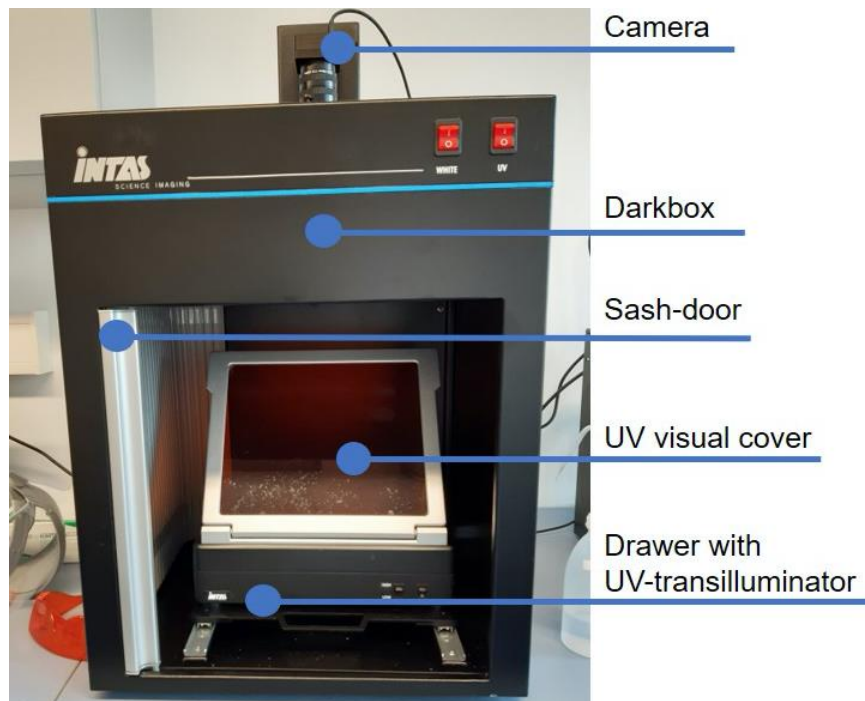
Figure B.6.1: Picture of the INTAS GelDoc. The GelDoc includes a Camera, the so-called 'Darkbox' contains a UV-transilluminator on a drawer and a UV cover. The Dark box can be closed with a sash-door.

3. Pull the UV table out and make sure that the protective cover is in an upright position. Otherwise, you will not be able to take a picture.
4. If the gel tray is UV-transparent, you can place it on a UV table to visualize DNA fragments on the gel.
5. Push the UV table back into the chamber.
6. Turn on the switch on the UV table. The UV intensity can be switched to 'Low' or 'High'. Push the switch to change the intensity.
7. Close the door.
8. Switch on the big red switch for UV on the top of the chamber.
9. Adjust the camera focus (if needed) on the camera itself on top of the chamber.
10. Adjust the settings in the program to optimize the gel picture (Figure B.6.2).
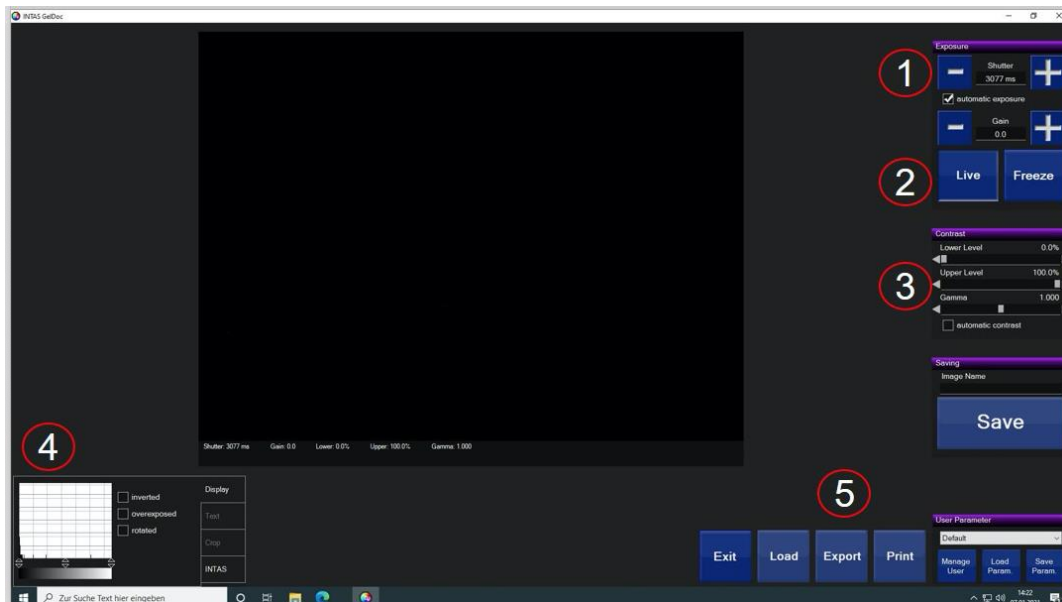
Figure B.6.2: INTAS GelDoc Capture Software. The GelDoc includes a Camera and a program for processing, saving, and exporting pictures on a Windows computer. The user can switch between a 'Live' view and 'Freeze' (2). The user can further change settings for 'Exposure' (1) and 'Shutter' (3) These settings have to be modified in the 'Live' mode. 'Freeze' has to be used, when taking the picture. After the picture is taken, the user can modify it with the options on the left side (4), including the 'Inverting' option. Clicking on the 'Save' button will save the image in the current format and last used folder on the computer. The button 'Export' (5) lets the user choose a file name, file format, and location.

11. Save and export the image in the desired data format.

B.7 Protocol for TOPO Cloning

The TOPO® TA Cloning® Kits for Sequencing are used to clone Taq polymerase-generated PCR products for sequencing. The pCR™ Blunt II-TOPO® vector is designed to clone blunt-ended PCR products generated by thermostable proofreading polymerases such as Q5™ High-Fidelity DNA polymerase. The same protocol can be used for both vectors. The used PCR product has to be purified upfront.

1. Thaw the PCR product, if needed.
2. Mix the following reagents in a 1.5 microcentrifuge tube:

Reagent	Volume
Purified PCR product	1 µl
Salt Solution	0.25 µl
TOPO vector	0.25 µl

3. Mix the solution by swirling the pipette tip. Do not vortex! Do not mix by pipetting up and down!
4. Incubate for 30 min at RT.
5. Place the reaction on ice for 30 min.
6. If SOC medium is stored at -20°C, take it out and thaw it at RT.
7. Thaw One Shot Top 10 or XL1 Blue cells on ice (50 µl aliquots prepared for direct use and stored in the -80°C freezer, these cells are chemically competent cells). This will take only a few minutes. Do not thaw at RT.
8. Now you will transform your TOPO construct into competent *E. coli*. Add 1.1 µl of the TOPO cloning reaction to a vial of bacterial cells.
9. Place the tube on ice for 30 min. Longer incubation on ice does not seem to influence the transformation efficacy negatively.
10. Warm a water bath to 42°C. Check the temperature!
11. Heat-shock the cells at 42°C in the water bath for 40 – 45 sec. Set an alarm because more than 45 sec will harm the cells! Use a floater.
12. Immediately transfer the tubes to ice for 2 min.
13. Add 250 µl SOC medium (RT or heated to 37°C).
14. Cap the tube tightly and shake the tube horizontally (use a floater) in a 37°C shaker for 1 h.
15. Prepare the selective agar plates in the meantime. Always check what antibiotics are required for the vector used and what antibiotics were used for the agar plates. The currently used TOPO vectors (2020/2021) can be plated on ampicillin and kanamycin agar plates. Incubate plates at 37°C. Label the side of the bottom part (not the lid) with: Date, Name, vector name, bacterial cells name, PCR product name/ID, amount of plated solution.
16. Check the clean bench that you use for plating the cells. Turn on the fan and clean it with water and/or Ethanol. If you do not have a clean bench for GMOs use a Bunsen burner! Do not work in a clean bench that is not suited for GMOs!
17. Plate only 200 µl of each transformation on a pre-warmed selective agar plate. This will leave 50 µl transformation reaction but those can be used, if the 200 µl plated transformation results in an overgrown plate. Store the 50 µl reaction at 4°C until the next day. If it is not needed, discard it.
18. Incubate the plates overnight at 37°C and check the growth the next morning. If only a few colonies grew, leave the plate at 37°C for some additional hours. If the plate is overgrown, you can use the remaining 50 µl solution from the day before and plate them on a fresh agar plate (repeat steps 15 to 18).
19. Use directly or store plates at 4°C. To do so, wrap the plate with parafilm.

B.8 Protocol for Colony PCR and Overnight Culture

Since the TOPO vectors are highly efficient, this step is not necessary but it is a convenient method for determining the presence or absence of insert DNA in plasmids. An alternative is to skip the colony PCR and do restriction digestion (see B.10 Restriction Digestion). For restriction digestion prepare overnight cultures directly and isolate plasmid DNA afterwards with a so-called 'MiniPrep' kit. After isolating the plasmid DNA, you can do restriction digestion.

1. Prepare a selective agar plate by warming it to 37°C in an incubator. This plate is called 'Master plate'. Be careful to use a plate with the right antibiotic. Draw a checkboard pattern on the back of the plate (bottom part, not the lid) and label it with the name of the transformation reaction from the TOPO cloning. Each box in this pattern will later contain an individual bacterial colony from the transformation.
2. Prepare a Dream Taq DNA master mix as described in A4.2 Thermo Scientific Dream Taq DNA polymerase. Primers designed to specifically target the insert DNA can be used to determine if the construct contains the DNA fragment. An alternative is to use a primer that targets the vector DNA flanking the insert.

   For the next step work under a clean bench or with a Bunsen burner.

3. Use a sterile pipette tip to pick up individual colonies and dip them into each (empty) PCR reaction tube. This will serve as a DNA template for the PCR.
4. Dip the same tip on the 'Master plate'. Use one box of the checkboard pattern on the back for one colony.
5. Add the Dream Taq master mix to the PCR tubes and perform the PCR in a thermal cycler. Use the primer Tm specific for the primers used in the colony PCR.
6. Place the Master plate back into an incubator at 37°C until colonies are visible. It is advisable to perform the colony PCR in the early morning so that you can continue in the late afternoon.
7. To check the fragment size of the PCR, use an agarose gel. See B.6 Protocol for Gel Electrophoresis for more information. You will proceed with those bacterial colonies that show the expected fragment size on the gel.
8. Set up overnight cultures for the colonies selected by colony PCR. Use 5 ml selective culture media tubes from the 4°C fridge. These are also called 'Minis'. The media tubes contain LB media and an antibiotic. Use the media that contains the same antibiotic as the plates used for cloning and the Master plate. The tubes are labeled with prepared stickers (on the fridge door in a plastic pocket). Those stickers are consecutively numbered. Stick to the numbering, since you have to fill out forms with details in the IBBP win file folder. You find an Excel sheet with the name 'Primer-Mini-Sequencing-Vectors_AGSchetelig' in 'Y:\AGSchetelig\Science' (effective January 2021). Always use the latest version of this document. Open the file and look for a spreadsheet called 'Minis 1500-XXXX'. According to the Mini number, fill out the additional details like description, date created, created by, storage, further use, Disposal, or glycerol stock (Figure B.8.1). Add your name, date, and if possible some more detail to the insert, cells, or vector on the Mini as well.

	A	B	C	D	E	F	G
1		Please do not change the order of the first 6 columns. Use these columns for GVO documentation in "Formblatt Z"					
2	Mini No.	Description/Label	Date created	created by (name)	Storage	Further use	Zerstörung/Entsorgung Disposal
644	6124	puc19_XL1Blue_Dsuz_glob3_5'RACE_clone004	06.10.2020	Sarah Petermann	4°C liquid culture		07.10.2020
645	6125	puc19_XL1Blue_Dsuz_glob2_3'RACE003_clone001	06.10.2020	Sarah Petermann	4°C liquid culture		07.10.2020
646	6126	puc19_XL1Blue_Dsuz_glob2_3'RACE003_clone002	06.10.2020	Sarah Petermann	4°C liquid culture		07.10.2020
647	6127	puc19_XL1Blue_Dsuz_glob2_3'RACE003_clone003	06.10.2020	Sarah Petermann	4°C liquid culture		07.10.2020

Figure B.8.1: Excel list with details for Overnight cultures. This image shows a part of the 'Mini list' in the file 'Primer-Mini-Sequencing-Vectors_AGSchetelig'. Next to each Mini number, the details for the Mini have to be given. Shown here is the Description/Label that contains information on the vector and the bacterial cells used for transformation and information on the insert. Other important information is the creation and disposal date, who created the Mini, and if it is stored or further used.

9. Get the Master plate from the incubator.
10. Pick a verified colony with a sterile pipette tip. Do this either at a clean bench or use a Bunsen burner.
11. Open a Mini tube with one hand and put the pipette tip in the media with the other hand. Close the tube.
12. Let the Mini incubate overnight at 37°C in a shaking incubator.
13. Check the Mini the next morning, if bacteria grew. If bacteria grew, the media should be cloudy.
14. Proceed with the plasmid isolation and purification ('Miniprep').

B.9 Protocol for Plasmid DNA Isolation and Purification (Mini Kit)

NucleSpin® Plasmid (NoLid), Macherey Nagel (REF 740588.250 for 250 preps)

This kit is used for the isolation and purification of plasmid DNA. It follows the overnight culture (see B.8 Protocol for Colony PCR and Overnight Culture). First, check if all buffers are prepared according to the manufacturer's protocol. All centrifugation steps are performed at RT and 11,000 x g.

Harvest bacterial cells

1. Prepare 2 ml reaction tubes.
2. Add 1.8 ml ON culture to the tube (2 ml can easily spill).
3. Pellet the cells in a benchtop centrifuge for 30 sec.
4. Discard the supernatant.
5. Repeat these first steps to use all of the ON cultures.

Cell lysis

6. Add 250 µl Buffer A1.
7. Resuspend the cell pellet completely by vortexing. No cell clumps should remain.
8. Add 250 µl Buffer A2. Mix gently by inverting the tube slowly. Do not vortex to avoid shearing of the DNA.
9. Incubate for 5 min at RT.
10. Add 300 µl Buffer A3. Mix by inverting the tube slowly until blue samples turn colorless.

Clarify lysate

11. Centrifuge the tube for 5 min. Repeat this step if the supernatant is not clear.

Bind DNA on Column

12. Place a NucleoSpin® Column in a collection tube (2 ml) and pipette a maximum of 750 µl of the supernatant onto the column.
13. Centrifuge for 1 min. Discard flow-through and place the column back in the collection tube. Repeat this step if you need to load the remaining lysate.

Wash silica membrane in column

14. Add 500 µl Buffer AW and centrifuge for 1 min.
15. Discard flow through.
16. Add 600 µl Buffer A4 (check if it is supplemented with ethanol!). Centrifuge for 1 min and discard flow through.

Dry the silica membrane

17. Centrifuge for 2 min. Discard the collection tube.

Elute DNA

18. Place the column in a 1.5 ml microcentrifuge tube (which is not provided in the kit) and add 50 µl Buffer AE (TE buffer or water work too).
19. Incubate for 1 min at RT.
20. Centrifuge for 1 min.
21. You can now discard the column.
22. Measure the concentration of the plasmid DNA with a spectrophotometer (see B.3 Spectrophotometric DNA Quantification).

B.10 Restriction Digestion

The restriction digestion follows the overnight cultures and the 'Miniprep' with the NucleoSpin Plasmid Isolation Kit (Macherey-Nagel). If you decide to use restriction digestion to check for the correct insert in the plasmid, you skip the colony PCR and the Master plate. Follow the protocol for ON cultures above and pick the colonies directly from the original agarose plate.

1. Thaw your plasmid DNA.
2. Heat the thermal block to 37°C.
3. Mix the following reagents in a 1.5 microcentrifuge tube:

Reagent	Volume
H_2O	14 µl
Cut Smart buffer	2 µl
Plasmid DNA	3 µl
EcoRI	1 µl

4. Incubate for 1 h at 37°C in the thermal block.
5. Use an agarose gel to check the result (see B.6 Protocol for Gel Electrophoresis).

B.11 Sanger Sequencing with Macrogen

For sequencing, 500 ng of purified plasmid DNA are recommended, mixed with 2.5 µl of the sequencing primer of the corresponding plasmid and filled to a total volume of 10 µl with HPLC-H_2O.

1. Get Macrogen sequencing labels. Labels are prepaid and consecutively numbered. Each label consists of two parts, a bigger label that needs to go on the sample and a smaller label. This smaller one should be clued into the laboratory notebook.
2. Mix the following reagents in a 1.5 microcentrifuge tube:

Reagent	Volume
H_2O	up to 10 µl
Primer	2.5 µl
Plasmid DNA	500 ng

Or if you want to sequence purified PCR products:

Reagent	Volume
H_2O	up to 10 µl
Primer	2.5 µl
PCR product	50-75 ng

3. Close the lid tightly. Wrap the big Macrogen sequencing label around the tube. Make sure that the information on the label is still well visible. The best way is to make a little 'flag' (Figure B.11.1). The smaller label belongs in the laboratory notebook.

Figure B.11.1: Sequencing label. The best and easiest way to stick the label is to make a 'flag' (left picture). The QR-Code must be still visible. The picture was taken from Macrogen (https://dna.macrogen-europe.com/eng/support/ces/ezseq_intro.jsp).

4. Put your samples in a UPS bag (prepaid), close the bag, and put it in the mailbox labeled 'Macrogen, UPS' outside TIG. You can write down the tracking number in case the bag gets lost.
5. Fill out the form in the Excel sheet with the name 'Primer-Mini-Sequencing-Vectors_AGSchetelig' in 'Y:\AGSchetelig\Science' (effective January 2021). Always use the latest version of this document. Open the file and look for a spreadsheet called 'Sequencing 1E29ZAEXXX'. The long number-letter combination at the end is unique for each batch of labels, so check the number on your used label and fill out the correct spreadsheet. Look for the matching label numbers you used and fill out the additional details like description, date sent, primer name, and created by (Figure B.11.2).

	A	B	C	D	E	F	G
1		Reaction Information		Sample Information			
2	Sticker	ample Name-Commen	Primer Name	Sample [c]	name of vector	Primer [c]	Date sent
3	1E29ZAE239	M6147	MFS14				14.10.2020
4	1E29ZAE238	M6148	MFS13				14.10.2020

Figure B.11.2: Sequencing documentation. Each batch of sequencing labels and each label in this batch have a unique ID. It is necessary to document each sequencing reaction by adding information like the sample name, the primer used for sequencing, or the date the sample was sent to Macrogen.

C Bioinformatic Protocols for Microsatellite Analysis

C.1 Simplemappr

Simplemappr (https://www.simplemappr.net) is a website that allows the creation of free point maps by using coordinates. If in doubt, you can check the coordinates with Google Maps (https://www.google.de/maps).

1. Start by zooming into the needed part of the map. For example, if you only have locations from Germany, zoom into Germany, otherwise, you will only see really small dots on the world map (Figure C.1.1).
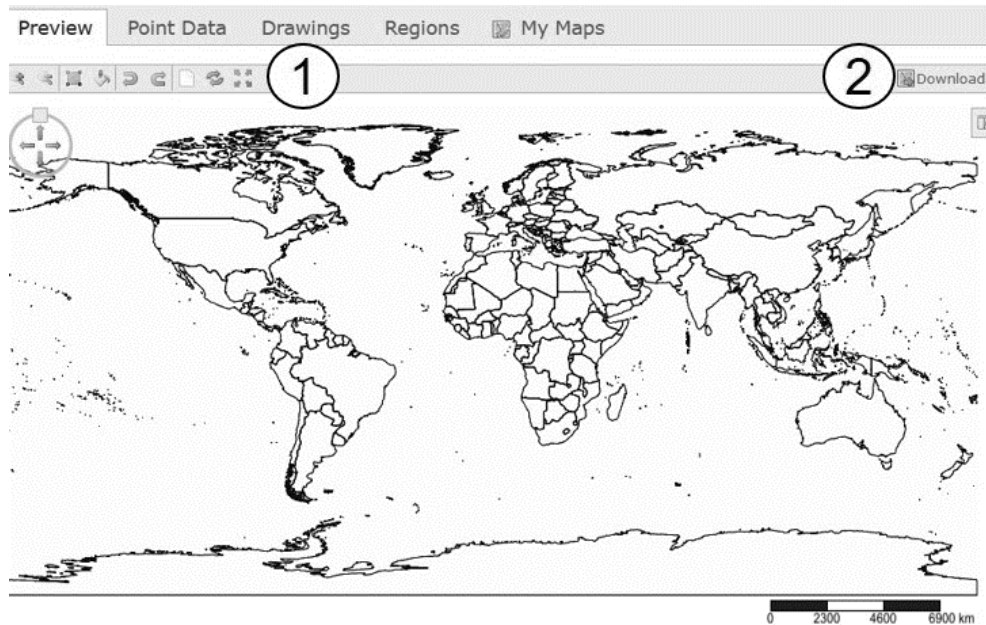
Figure C.1.1: Image of the Simplemappr website. Shown is the Homepage of Simplemappr.net. In default settings, it shows the world map with country borders. You can zoom in and out, crop the map, fill regions with color, undo and redo in the bar on the top (1). You can download the finished map with a click on the button in the top right corner (2).

2. Choose the settings for your map
3. Add the coordinates under 'Point data'. Type geographic coordinates on separate lines in decimal degrees (DD) or DD°MM'SS" as latitude, longitude separated by a space, comma, or semicolon. You can change. You can also change the shape, size, and color of the data points (Figure C.1.2).

Figure C.1.2: Point data and settings. Under 'Point data' the user can add coordinates (1). Each coordinate is entered as a 'Layer' (2). By default, the website provides three layers but the user can add or delete layers manually. In addition, the user can choose between different options on how the points in the map should be depicted. The size of the marker can be changed (3), the color (4), and the shape (5).

4. Save your map or make a screenshot. Be aware that saving sometimes changes the ratio of your map, so better check your saved map directly because closing the site will delete your data and you have to start from the beginning. This seems to be a bug.

C.2 Geneious for Microsatellite Analysis

The Microsatellites Plugin for Geneious Prime is a tool that imports ABI fragment analysis files and allows the user to visualize traces, fit ladders, call peaks, predict bins, display alleles in tabular format, and export data.

Calling the ladder

1. You need to know which ladder was used for sizing. In this project, the 'GeneScan 500 LIZ size standard' was used (Figure C.2.1). Check on the manufacturer's website how the ladder should look like. Only if the right ladder gets recognized, Geneious can calculate the size of the microsatellite fragments correctly.
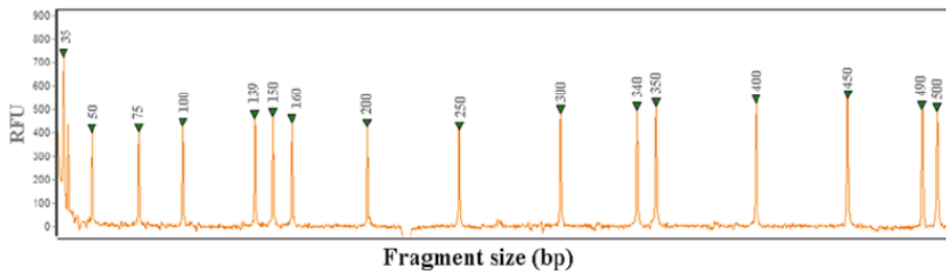
Figure C.2.1: Pattern of peaks in GeneScan 500 LIZ size standard. This image is taken from Flores-Rentería and Krohn (2013). It shows the 'GeneScan 500 LIZ size standard' available from Thermo Fisher Scientific (Catalog number: 4322682). The height of each peak corresponds to its relative fluorescence intensity (RFU). It has 16 single-strand-labeled fragments.

2. Open an ABI fragment analysis file. If you do not see the orange LIZ trace, make sure that the LIZ channel on the right is marked and uncheck all other dyes. The program tells you that it cannot fit the ladder (Figure C.2.2).
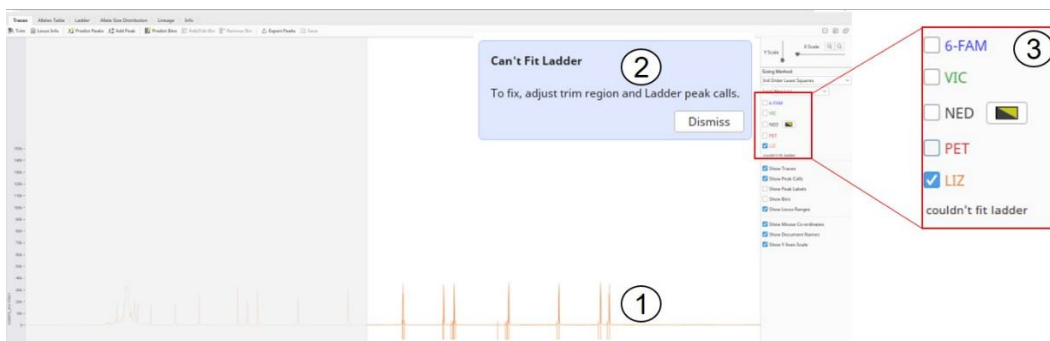
Figure C.2.2: Screenshot from Geneious. A new and unmodified ABI fragment analysis file opened in Geneious Prime. It shows the original peak call from the program (1). The program is usually unable to call the correct peaks and will let you know by showing a message in the top right corner (2). When calling the peaks for the ladder, make sure to only check the LIZ channel in the menu on the right (3).

3. Peak calls are shown by the vertical line below the trace. The program does that automatically but the calls look random (Figure C.2.3). The problem is that the algorithm thinks a peak is present every time there is a slight decrease followed by an increase in the peak height in relative fluorescence units (RFUs).
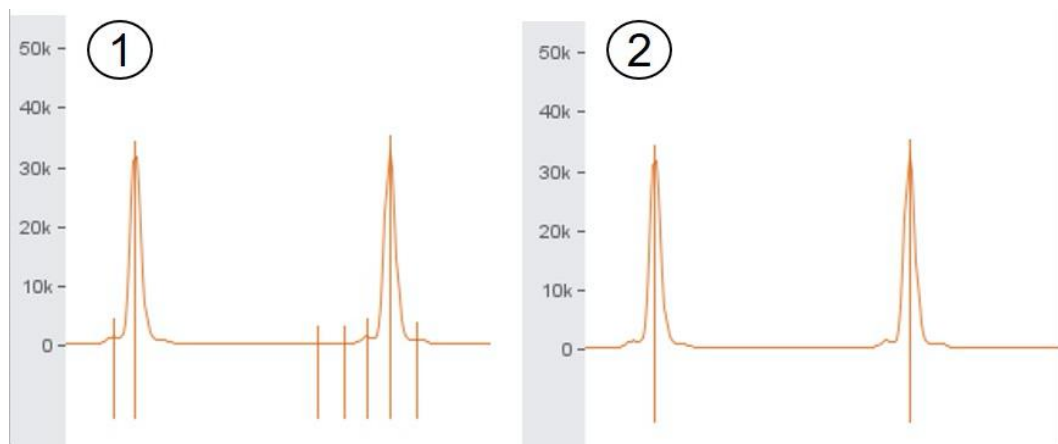
Figure C.2.3: Calling the ladder. Geneious will automatically call peaks for the ladder (1). Those are usually not correct and the peaks have to be called manually by deleting or adding peaks (2).

4. You have to delete all of those wrong peaks manually and make sure that only the correct ones remain. You remove wrong peaks by marking them with the mouse and pressing 'Del'/'Entf' on the keyboard. You add a call by simply right-clicking on the correct position and choosing 'Add peak' and the dye used (Figure C.2.4). A manually made peak might not be as precise as an automated peak call but the program is using algorithms that correct for such mistakes.

Figure C.2.4: Adding a peak to the ladder manually. Geneious allows the user to add peaks manually to a trace. This can be made by right-clicking on the correct position and clicking on 'Add peak'. In this case, LIZ is checked as the only dye but it works the same way for the other dyes as well.

5. The program will recognize the ladder at some point and tell you which ladder it thinks was used (Figure C.2.5). If it is not correct, check the peaks one more time and compare them to the manufacturer´s specification.

Figure C.2.5: Calling the correct ladder. After the peaks were called, Geneious will recognize the right size standard. The name of the ladder appears under the dye channel selection. In this example, Geneious recognized 'GeneScan 500' (red square).

6. Save the file.

Setting locus information

Once the ladder is correctly called, set the locus information.

1. Click on 'Locus Info' in the bar on the top and a window will pop up (Figure C.2.6).
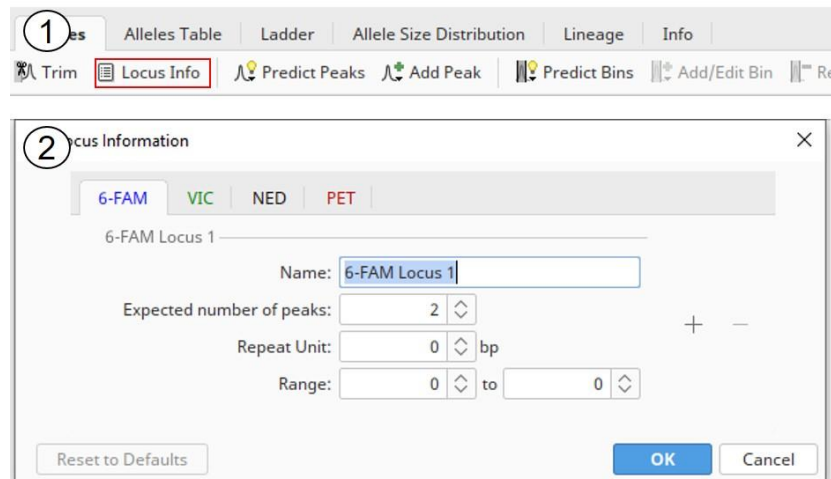
Figure C.2.6: Menu for locus settings. To set the locus information for the microsatellite markers used in the experiment, click on 'Locus Info' in the toolbar (1), marked here with a red square. A window opens that allows adding information on the microsatellite markers (2). You have to add the name of the marker, the expected number of peaks, repeat unit, and range for each available dye (FAM, VIC, NED, PET).

2. Add information for each marker used. You should find the information for the markers in the original publication. In this case, all the needed information was included in the publication of Fraimout (2015).
3. The name can be chosen freely but it helps to stick to the original publication.
4. In a diploid organism, two peaks are expected. This information is important for the analysis. If only one peak is predicted in a locus, the program will assume a homozygote and if two peaks are detected, the program assumes a heterozygote. If more peaks are detected, the program will give an alert and the user has to check the locus manually.

5. In this project only dinucleotide repeats were used, so the repeat unit would be two.
6. The range sets the limit for the locus. If a locus is known to have shown variations between 320 bp and 400 bp in length, then these values have to be filled in.

Calling peaks for the microsatellite loci

Now that the ladder is called correctly and the locus information is set, the actual fragment length of each microsatellite can be called. This step is similar to the procedure with the ladder.

1. Chose a sizing method. In this project '3rd Order Least Squares' was chosen (Figure C.2.7). This sizing algorithm was chosen over the alternative algorithms because it uses regression analysis to build a best-fit-size-calling curve, it compensates for any fragment that may run anomalously and results in the least amount of deviation for all the fragments, including size standard and samples. 3rd order was chosen over 2nd order because it uses a higher polynomial degree and captures more of the peak structure. It also provides more flexibility when generating best-fit curves for sizing samples with anomalously migrating fragments. For all samples, the same algorithm has to be used.
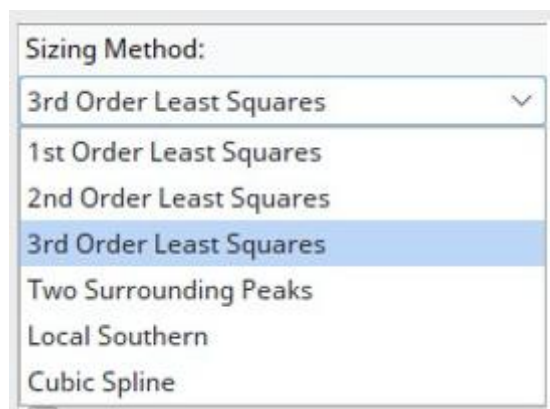
Figure C.2.7: Sizing methods in Geneious. The user can choose between six different sizing algorithms.

2. Mark one dye that you want to analyze, for example, 'FAM'. Only the FAM trace is now visible (Figure C.2.8).

Figure C.2.8: Calling peaks for a FAM locus. For analysis, only one dye after another should be edited. Check or uncheck dyes in the dye selection (1). The marked dye (here FAM) is visible (2). All other dyes are not visible for now but they are unaffected by the changes you may or may not make in the FAM channel.

3. As it was for the ladder, the program will try to call peaks. This works better than for the ladder. As you see in the example in Figure C.2.8, the program correctly called two peaks.
4. You will have to check every single locus in every sample manually. There is no guaranty that the program calls peaks correctly all of the time.
5. Try to analyze as many samples at one locus as possible at once. You can select more than one sample in the file selection. This will help to compare the results and you might find it easier to detect artifacts in a single sample or to find null alleles. In that case, the different samples are listed one above the other (Figure C.2.9).
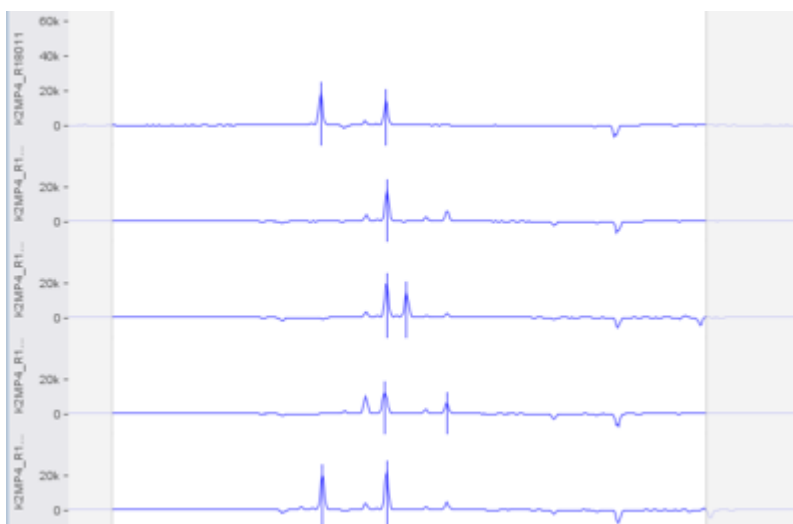
Figure C.2.9: Calling peaks for a FAM locus in more than one sample. In this example, five different samples are selected. The sample names are all given on the left side, next to the RUFs. Each sample can be modified independently by clicking on the corresponding track.

6. Repeat for every locus and dye channel in every sample.
7. Now that all loci are set and the incorrect peak calls were corrected you can move on and perform the binning.
8. Save as often as possible, at least after every locus.

Binning

1. Select all samples you modified and one dye channel, for example, 'FAM'.
2. Click the 'Predict Bins' button and just click OK. Do the same for the other dyes but not for LIZ, since this is the ladder.
3. Based on the observed peaks and their size using a specific sizing algorithm, the program creates bins.
4. Save now.
5. Open the Alleles Table, turn on all or just one of the dyes and see if there are any un-binned peaks (Figure C.2.10).

Figure C.2.10: The Allele Table. The Allele Table contains information on allele size for every locus in every sample. In this example, the two selected samples appear to have no peak in the VIC-locus DS26. This has to be checked manually. This table also shows that both samples are homozygous for FAM-locus DS06 and heterozygous for VIC-locus DS36.

6. Go to the affected sample and turn on the dye that has the un-binned peak.
7. Select the bin the peak should be in and choose 'Edit Bin'.
8. Extend the range of the bin by dragging it with the mouse to include the peak. If this is appropriate depends on how far away bin and peak are. If it is just a little bit outside the bin, then it is okay to move the bin. Otherwise, consider discarding this sample for now and repeating the whole FLA one more time (Figure C.2.11).
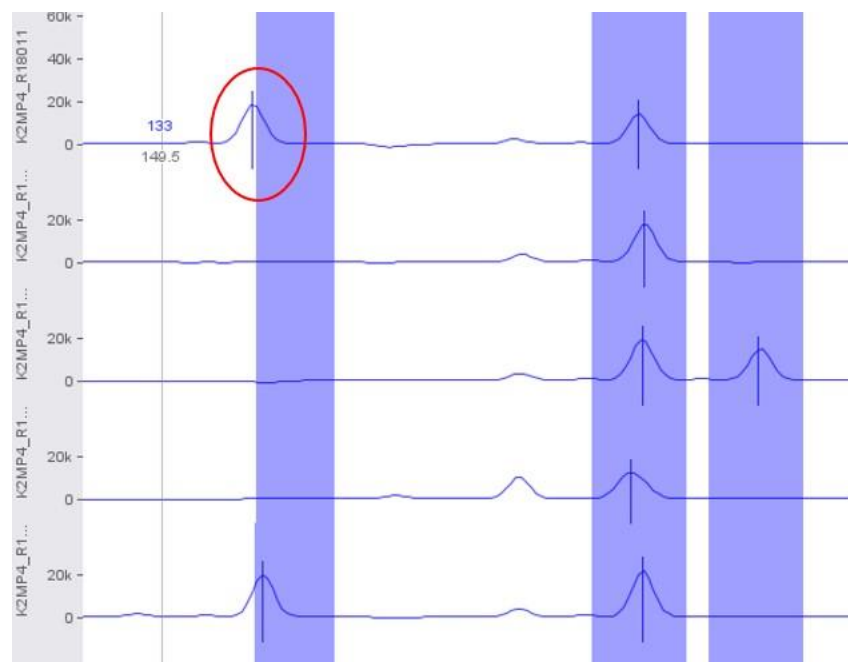
Figure C.2.11: Automated binning in a FAM locus. The binning in this example was done automatically by the program. The peak marked with a red circle is slightly outside the calculated bin. Extending the bin to the left so that the peak is included would be fine in this case.

9. Check the Alleles table again for un-binned peaks.
10. In the case of the warning 'no peaks', no alleles were amplified in those samples. Check those as well. It is possible that peaks are present but were not called. If no peaks are detected, even if you repeat the analysis, then you might be dealing with a null allele.
11. Export the Alleles table to a CSV file. This can be opened in Excel. You can export with or without warnings. If you export without warnings, it will leave those fields blank, otherwise, they'll contain text such as 'No peaks'.

Artifacts in FLA

Artifacts can be caused by air bubbles, crystallized polymer, or a voltage surge to the electrophoresis instrument. They can look just like a correct peak but are not reproducible. Knowing how common artifacts can look like will help to identify them and to tell them apart from real peaks (Figure C.2.12).
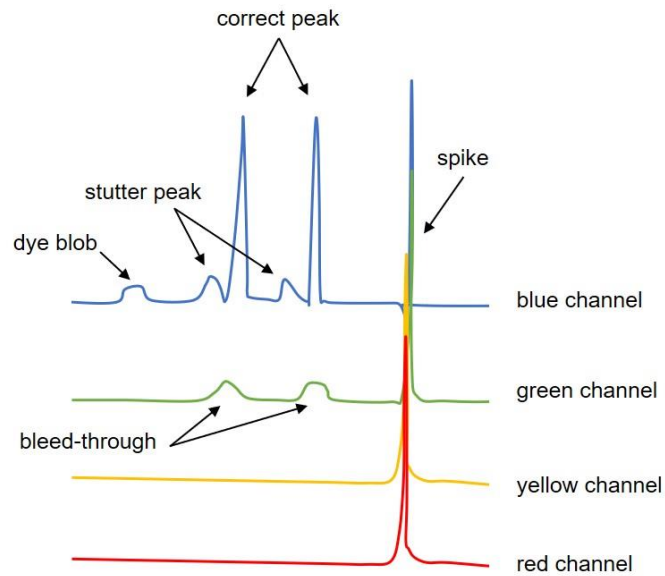
Figure C.2.12: Common artifacts in FLA. Shown is a schematic illustration of how the most common artifacts look like compared to correct peaks (STR alleles). Identifying them in a real sample takes some practice since the intensity of a true peak is not always as high as it is depicted here. Shown are a dye blob and a stutter in the blue channel, a Pull-up or bleed-through in the green channel, and a spike in all four channels.

C.3 GenAlex

Installing GenAlex on Windows

GenAlex is available as an Excel add-in on the website http://biology.anu.edu.au/GenAlEx/.

1. Download and save the zipped file to a dedicated folder.
2. Use the extract option to unzip the file.
3. To install it launch Excel, open the options menu via the 'File' menu.
4. Click on 'Add-ins' in the options menu and then on 'Manage' (Figure C.3.1). A window will pop up in which you can browse the folder that contains the GenAlex add-in. Confirm that the add-in can be used after restarting Excel.
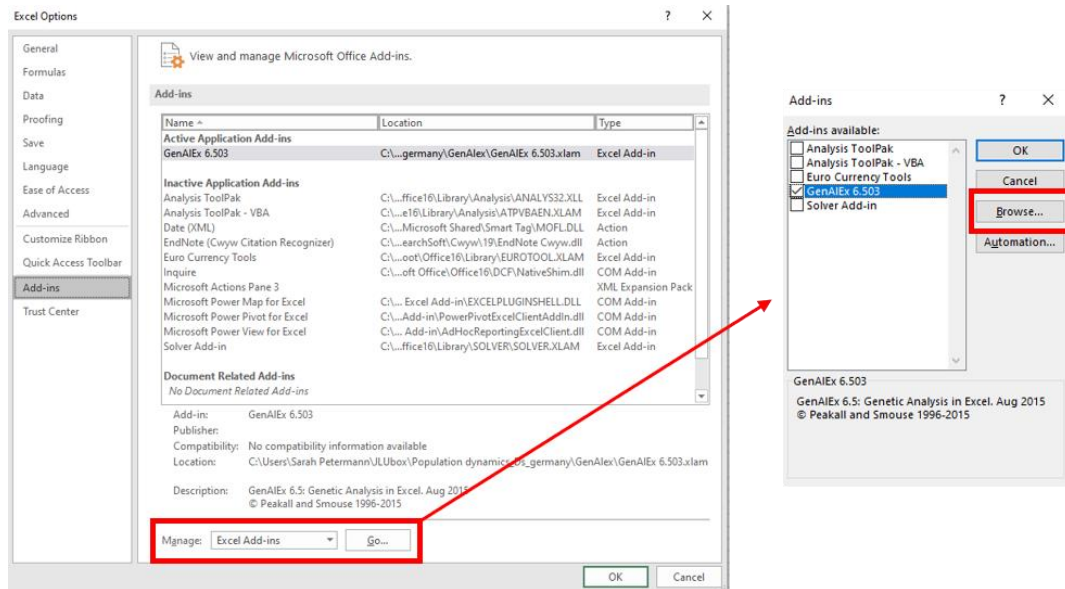
Figure C.3.1: Installing the GenAlex add-in in Excel. In the Option menu click on 'Add-ins'. If GenAlex is not installed yet, the window 'Active Application Add-ins' will be empty. Click on 'Manage' and a window called 'Add-ins' will pop up. Click on 'Browse' to select the folder containing the GenAlex download file.

5. If GenAlex was successfully installed, a loading screen will pop up when starting Excel.
6. The Excel toolbar should contain a tab called GenAlex (Figure C.3.2).
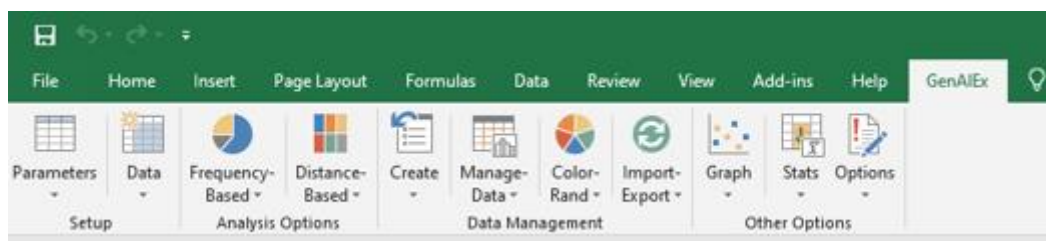
Figure C.3.2: Excel toolbar. To use GenAlex click on the tab in the Excel toolbar. It contains all tools provided by GenAlex.

Getting started

1. Use the Allele table CSV file from Geneious (see B2. Geneious for microsatellite analysis). Click on 'Import' in the Excel toolbar.
2. You can also copy and paste the allele table directly from Geneious into Excel.
3. In either case, you need to set the data parameter correctly. An example together with an explanation is given in Figure C.3.3. It is important to notice that missing data points have to be assigned to a value. A blank will not be accepted by GenAlex. Instead use a negative value, like '-9' and GenAlex will automatically recognize that all data points with the value '-9' have to be treated as missing data.

Figure C.3.3: Data format of codominant microsatellite data for GenAlex. This is an example of how the table has to be formatted for GenAlex to be able to run the analysis of codominant microsatellite data correctly. This is a cropped image, so the total number of samples (B1) does not add up with the size of each population (D1 – J1). GenAlex does not analyze data with discrepancies between the total number of samples (B1), the number of populations (C1), and the size of each population (D1 – J1). All information given in the table is necessary, except for the title of the worksheet in A2 and the locus names starting in C3.

4. Tell GenAlex to test the raw data by clicking on Data and then on Check Raw Data (Figure C.3.4).

188
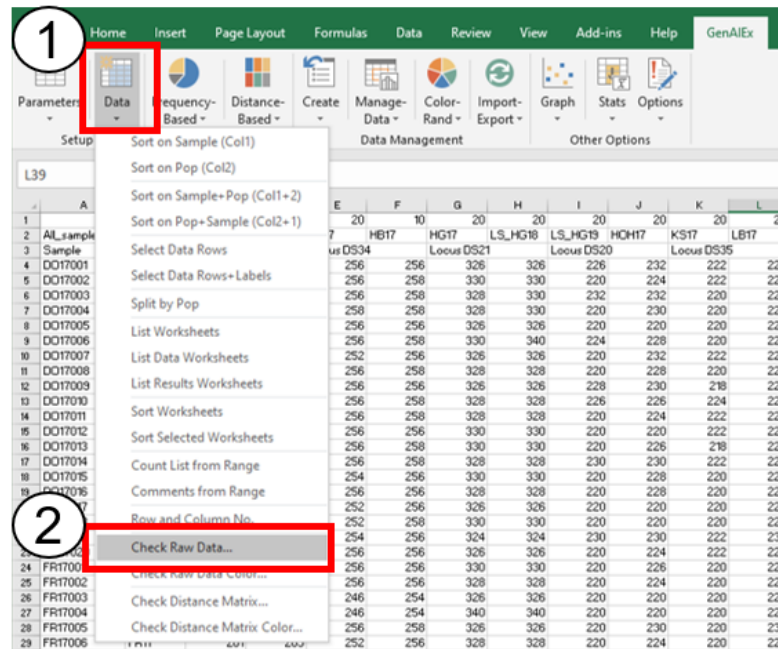
Figure C.3.4: Automatic raw data check by GenAlex. To let GenAlex check the raw data is especially interesting when there are many data points. Click on 'Data' (1) and then on 'Check Raw Data' (2). GenAlex will check if the data format is correct or if missing data points are present in the file.

Frequency-based analysis - HWE

Start with the calculations for HWE (Figure C.3.5) but be aware that the official manual advises using other programs like Arlequin for the calculations of HWE. You can still use it but better check the result with Arlequin or similar programs.

1. Click on the 'Frequency Based Analysis Option' in the toolbar.
2. Click on 'Disequil'.
3. Click on 'HWE' (Figure C.3.5).
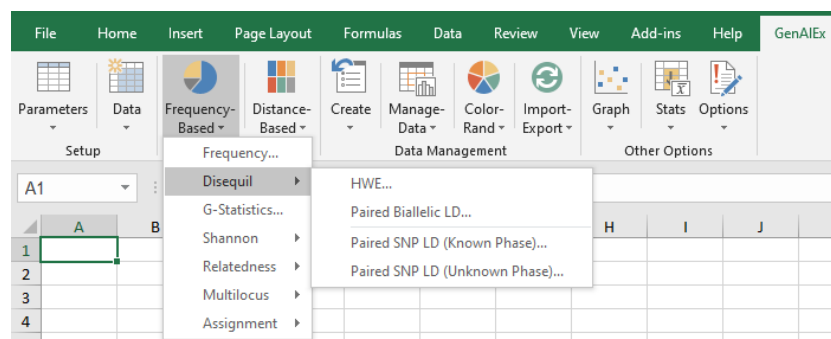
Figure C.3.5: Frequency Based Analysis Options in GenAlex. To calculate the deviations from HWE open the 'Frequency Based Analysis Options' in the toolbar and click on 'Disequil', then on 'HWE…'.

4. Set the HWE parameters (Figure C.3.6). The number of loci, samples. populations and the population sizes should already be filled out. The program takes the information from the data you put in, that's why it is so important to stick to the file format in Figure C.3.4.

189

Figure C.3.6 HWE Data Parameter in GenAlex. The information in the boxes is filled out correctly if the advised data/file format was applied. Click 'OK' to proceed.

5. Click 'OK'.
6. A window pops up, in which you can choose how the data should be presented. This is up to your likings. You can get a summary with only the most important data or you choose the 'Step by Step' option with more additional information. You can also decide if you only want to have a graphical output or the values or both (Figure C.3.7).

Figure C.3.7: Hardy-Weinberg Options in GenAlex. You can choose between different options on how the data should be presented. You can mark or unmark the options that suit you most. Click 'OK' to proceed.

7. GenAlex will now calculate HWE and possible deviations. A new sheet with the results will open automatically.

Frequency-based analysis - Frequency

The 'Frequency…' option does contain most of the analysis tools you will need in GenAlex. The procedure is similar to the HWE calculations.

1. Click on the 'Frequency Based Analysis Option' in the toolbar.
2. Click on 'Frequency…' (Figure C.3.8)

190
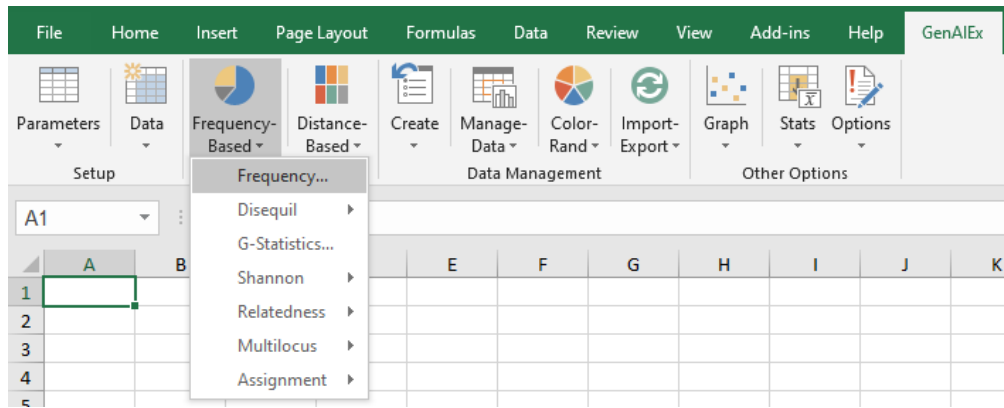
Figure C.3.8: Frequency Options in GenAlex. Like for HWE you first click on 'Frequency Based Analysis Option' in the toolbar. Then click on 'Frequency...', the first option in the list.

3. A window pops up, called 'Allele Frequency Data Parameters' (Figure C.3.9). As for HWE, the boxes should be filled with values, if the data/file format was correctly applied.
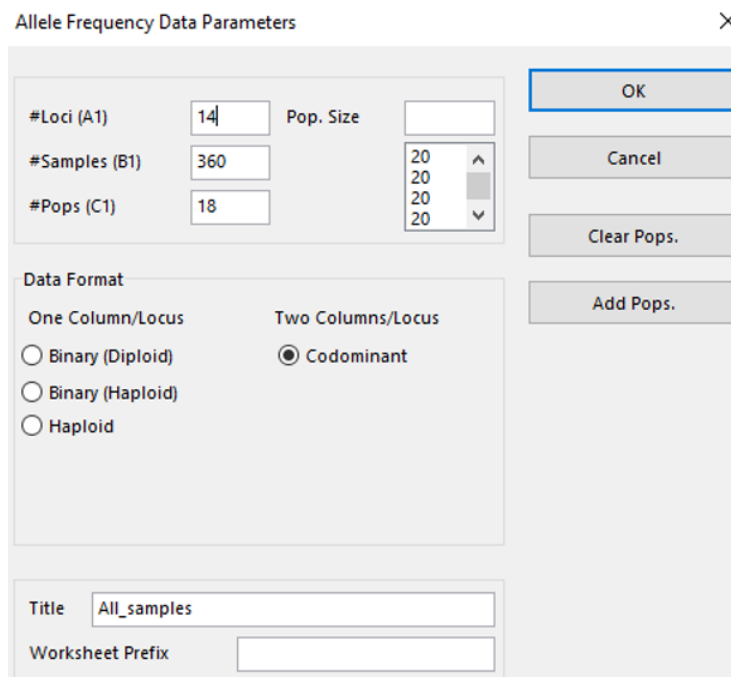
Figure C.3.9: Allele Frequency Data Parameters. The information in the boxes is filled out correctly if the advised data/file format was applied. Click 'OK' to proceed.

4. Click 'OK'.
5. The 'Codominant Frequency Options' open (Figure C.3.10).
6. Mark or unmark the options you would like to be calculated. To get a better overview of your data you might want to test all of it and see what the program calculates. Also, when using the 'Step by Step' option, it can provide some valuable information, especially if it is the first time you analyze such data.
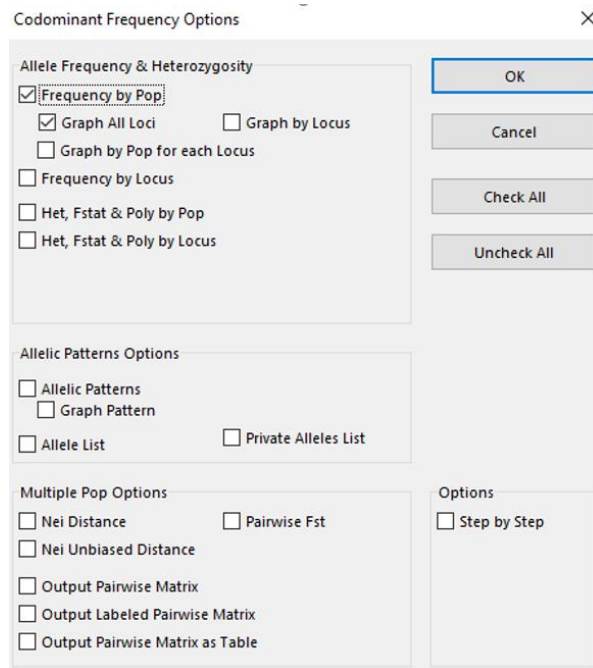
Figure C.3.10: Codominant Frequency Options. You can choose between different options on how the data should be analyzed. You can mark or unmark the options that suit you most. The first options all analyze the allele frequency and heterozygosity of the data. You can choose between calculations by population or locus and between graphic output or plain values. The next options are analyzing allelic patterns, again you can have values and/or graphs, a list with all alleles, and a list with private alleles. The 'Multiple Pop Options' contain the calculations for Nei Distance, which is needed for PCoA and the Pairwise F_{ST}. Both are important for the interpretation of population dynamics and genetics. Click OK to proceed.

7. Click 'OK'.
8. GenAlex will now analyze the data based on the options you choose. New sheets with different results will open automatically.

Distance-based analysis – AMOVA

1. Click on the 'Distance Based Analysis Option' in the toolbar.
2. Click on 'AMOVA' (Figure C.3.11).
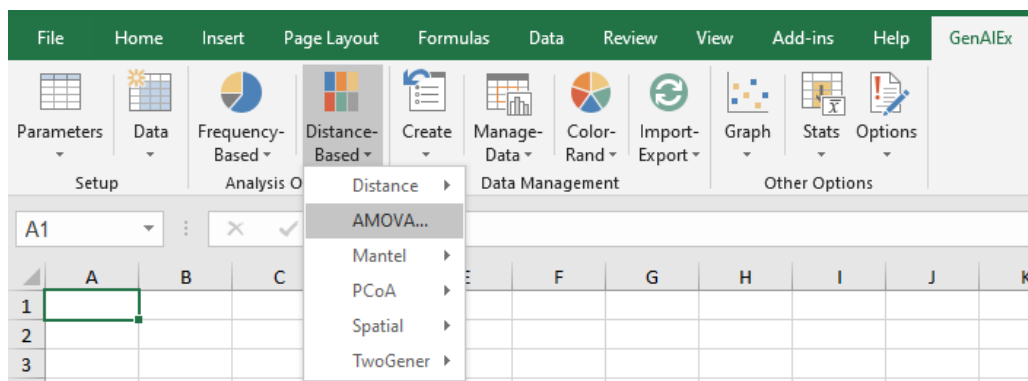
Figure C.3.11: Calculating AMOVA in Distance Based Options. First, click on 'Distance Based Analysis Option' in the toolbar. Then click on 'AMOVA', the second option in the list.

3. First, the 'AMOVA Data Parameters' have to be set (Figure C.3.12). The program fills the boxes automatically.

Figure C.3.12: AMOVA Data Parameters. The information in the boxes is filled out correctly if the advised data/file format was applied. Click 'OK' to proceed.

4. Select the 'AMOVA Genetic Distance Options' (Figure C.3.13).

Figure C.3.13: AMOVA Genetic Distance Options. The 'AMOVA Genetic Distance Options' gives you the option to choose between different distance calculations but if you use microsatellite markers the only useful option is 'Codom-Allelic' (the only option that is marked under 'Distance Calculations'). Choose if you want AMOVA for total only or each locus separately. Click 'OK' to proceed.

5. Set more details in 'AMOVA Options' (Figure C.3.14).

Figure C.3.14: AMOVA Options. The 'AMOVA Options' will first let you decide the number of permutations. A pie graph can be made as an output together with the AMOVA table. You also have an option to suppress within the individual analysis in AMOVA. Click 'OK' to proceed.

6. Click 'OK'.
7. GenAlex will now analyze the data based on the options you choose. New sheets with different results will open automatically.

Distance-based analysis – PCoA

For this, you need Nei Distance calculated with the 'Frequency…' option in the 'Frequency Based Analysis Options'. Be on the actual Excel sheet that has Nei Distance, otherwise, it will not work.

1. Open the Excel sheet containing Nei Genetic Distance.
2. Click on the 'Distance Based Analysis Options' in the toolbar.
3. Click on 'PCoA' and then on 'Analysis…' (Figure C.3.15).
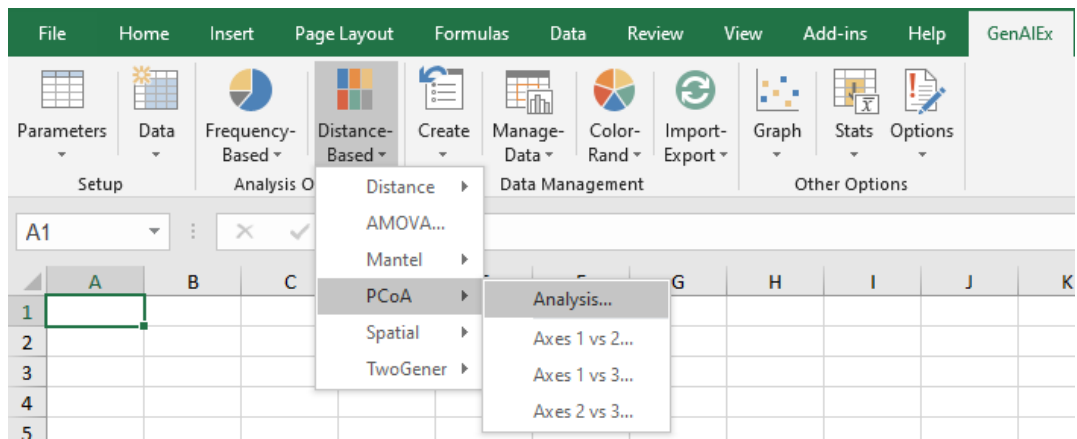
Figure C.3.15: PCoA for microsatellite data. Open the 'Distance Based Analysis Options', click on 'PCoA' and 'Analysis'. This will open the 'PCoA Parameters'.

4. The 'PCoA Parameters' open (Figure C.3.16).
5. Click on one of the PCoA methods.
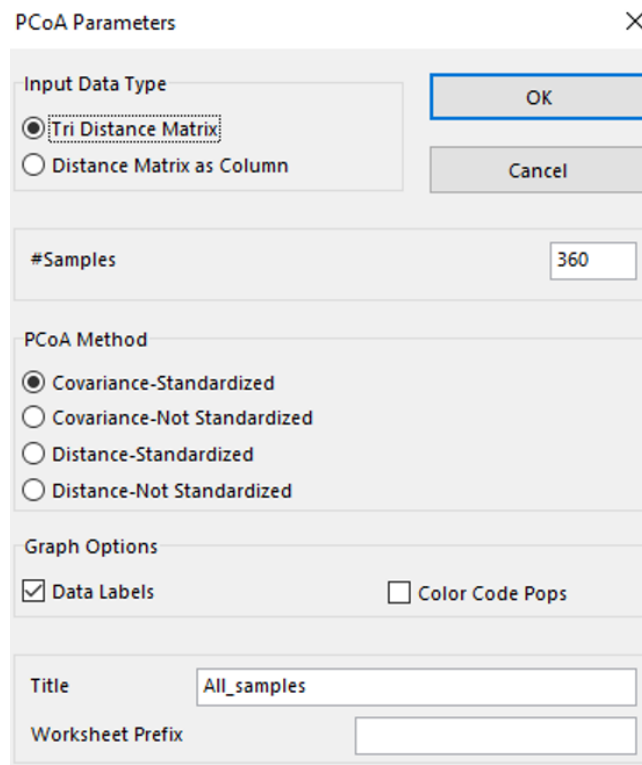
Figure C.3.16: PCoA Parameters. Make sure that 'Tri Distance Matrix' is marked and select one of the different PCoA methods. Click 'OK' to proceed.

6. Click 'OK'
7. GenAlex will now analyze the data based on the options you choose. New sheets with different results will open automatically.

Export data

All graphs and tables generated during analysis can be adjusted, modified, and copy and pasted as every other table in Excel. The 'Export' function in GenAlex can convert the raw data into other file formats for a variety of programs (Figure C.3.17). Just click on the format you need and save the file.
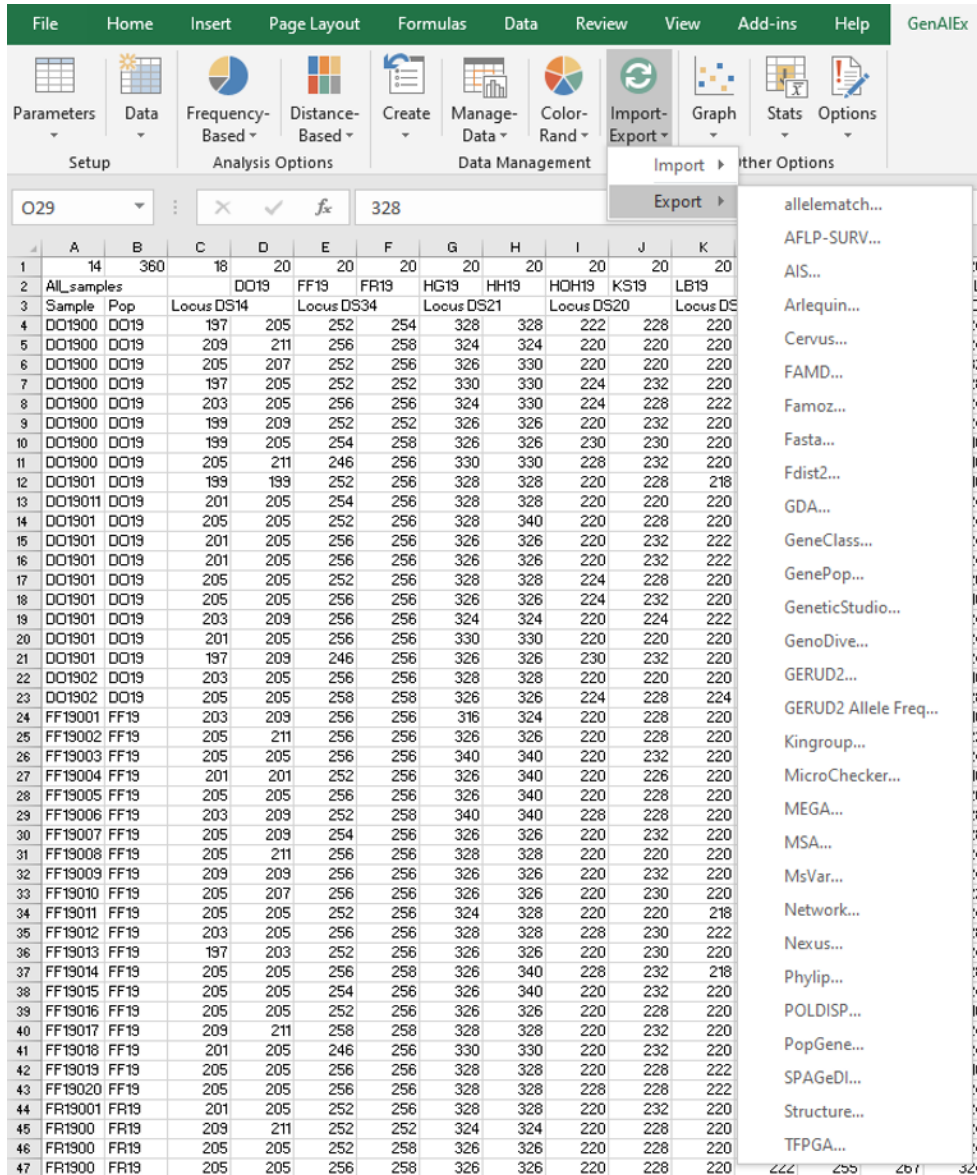
Figure C.3.17: Exporting/Converting raw data with GenAlex. Convert the microsatellite raw data in GenAlex in other file formats by clicking on 'Import/Export' in the toolbar. The user can choose between 29 different formats. The formats are named after the programs they are used for.

C.4 FreeNA

FreeNA estimates microsatellite marker null allele frequency for each locus and population. It can also estimate the unbiased F_{ST} and calculate the Cavalli-Sforza and Edwards (1967) genetic distance. This only runs on Windows computers.

As an Input file, simply use GenAlex 'Export' function to get a data file in the Genepop format. A genotype '0000' (for 2 digits allele coding) or '000000' (for 3 digits allele coding) represents missing data. A null homozygote should have the genotype '9999' (for 2 digits allele coding) or '999999' (for 3 digits allele coding).

1. To run FreeNA put the executable file in the same directory as the data file.
2. Open FreeNA by double-clicking on FreeNA.exe.
3. Following the menu, enter the names of your input and output files.
4. Fix the number of replicates for the computation of the bootstrap 95% confidence intervals automatically performed for the F_{ST} statistics.
5. The output of the program is provided as five different files:

   a. 'your_output_file_name.r' gives the estimate of null allele frequency for each population and locus.
   b. 'your_output_file_name.fr' contains allele frequencies and genotype numbers.
   c. 'your_output_file_name.gFst' contains the F_{ST} values.
   d. 'your_output_file_name.pFst' contains all pairwise F_{ST}.
   e. 'your_output_file_name.dc' contains the value of the Cavalli-Sforza and Edwards' (1967) genetic distance for each pair of analyzed populations.

C.5 Arlequin

Arlequin 3.5 may only run on Windows. To install Arlequin download the Arlequin zip file to a directory. Extract all files in the directory of your choice. Double click on the executable file to start the program. A detailed manual for the program Arlequin 3.5 is available at http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf.

1. Open Arlequin.
2. Start with the 'Arlequin configurations' (Figure C.5.1). Check or uncheck the configurations as you like. You have to specify a text editor that will be used by Arlequin. The normal preinstalled 'Windows Notepad' is sufficient. The Arlequin manual recommends TextPad (http://www.textpad.com). Rcmd requires R but is optional.
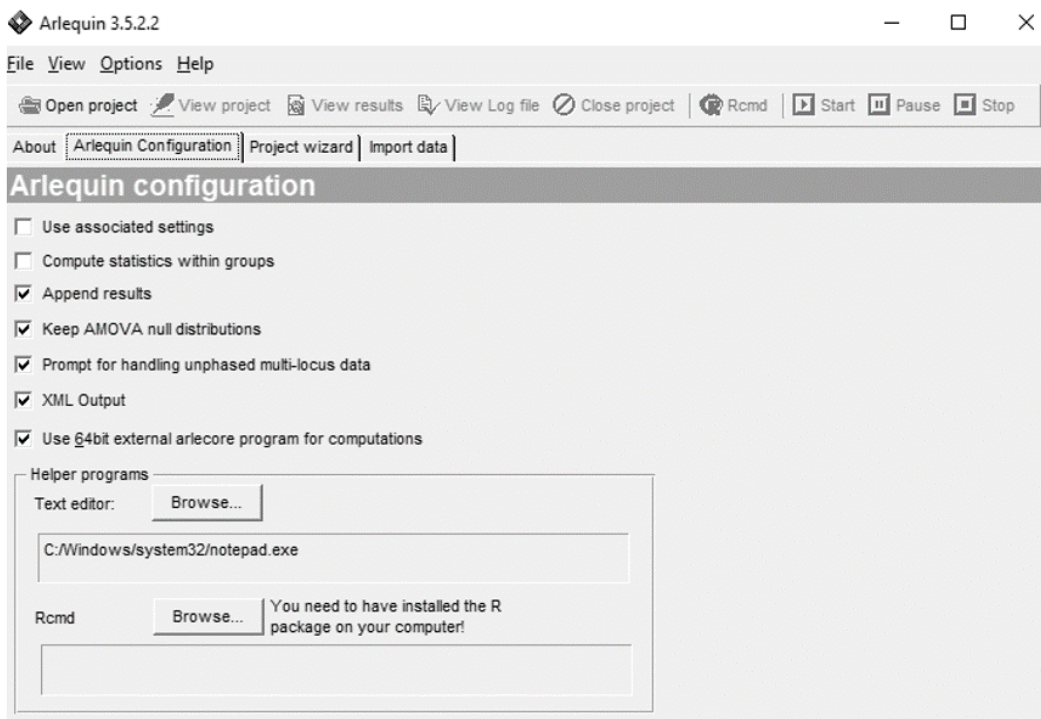
Figure C.5.1: Arlequin Configuration. Use the 'Arlequin configuration' window to set Arlequin options as you wish. Specify the text editor Arlequin should use in the 'Helper program' section of the window. Rcmd is optional.

3. Prepare an input data file. You can use the 'Export' function from GenAlex and create a new project with the 'Project wizard' implemented in Arlequin (Figure C.5.2). The 'Browse' button lets you specify the name and directory of the project. The file should have the extension '.arp'. Specify which type of data you are using, if the data is genotypic or haplotypic if the gametic phase is known or unknown and if the data contains recessive alleles under the 'Data type' options. Specify the number of samples and the character for missing data as well as the locus separator under 'Control' options. 'Optional sections' give the user the possibility to include a list of haplotypes, a distance matrix, and genetic structure.
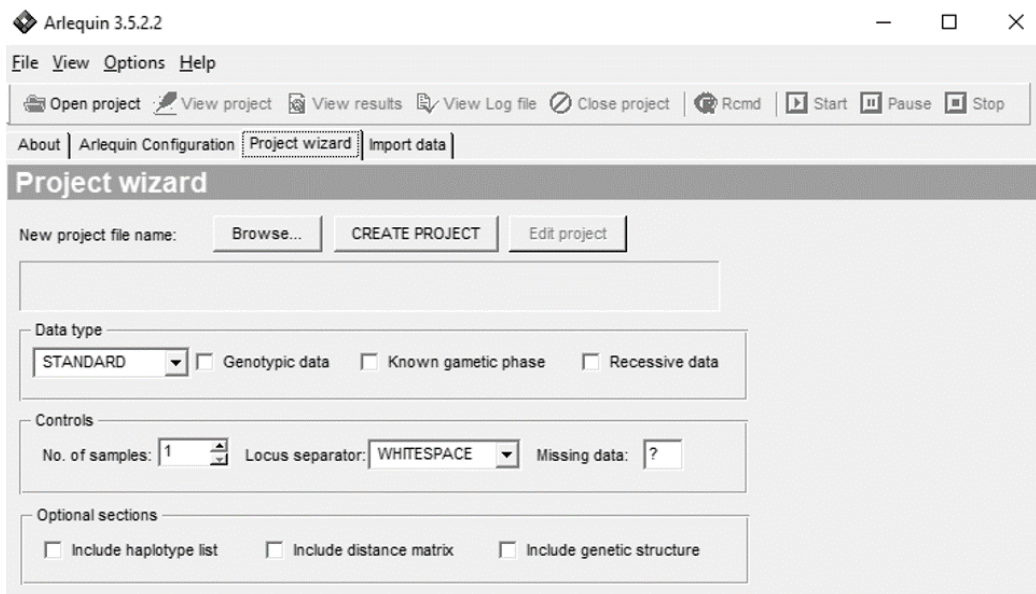
Figure C.5.2: Arlequin Project wizard. Use the Arlequin project wizard to quickly create the outline of a project.

4. Click on 'Create project' once all other options are specified. 'Edit project' allows the user to edit a saved project at any time.
5. Import the data by clicking on 'Import' next to the 'Project wizard' tab (Figure C.5.3).
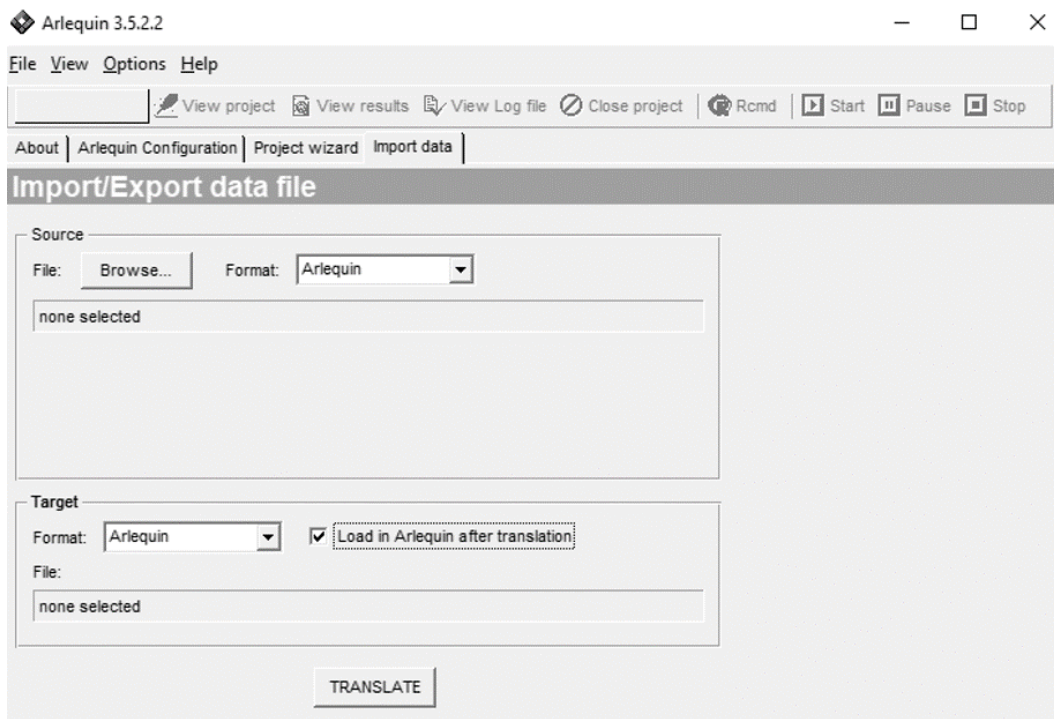
Figure C.5.3: Import a file into Arlequin. Browse for the file generated with the Export function in GenAlex. The Translate button is only used if you do not already have the right file format.

6. Click on 'Browse' and open the file generated with GenAlex.
7. The project will be loaded and the 'Project' tab becomes active (Figure C.5.4).
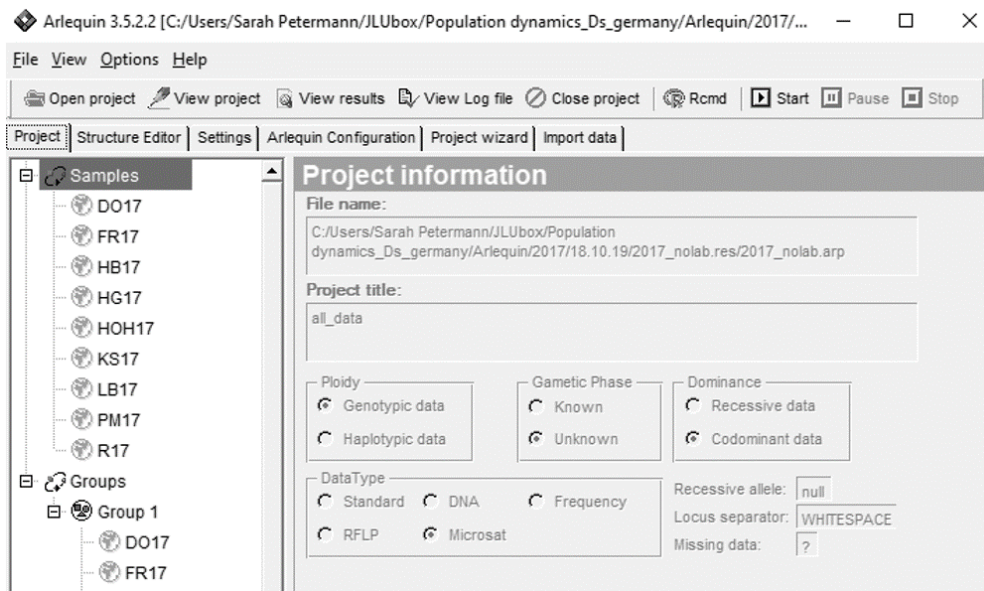
Figure C.5.4: Project information dialog. Once a project is loaded, the project tab dialog tab is available. It shows information about the project, like the data type and sample names.

8. Use the Settings tab to choose the tasks Arlequin should perform (Figure C.5.5). You can check and uncheck tasks by clicking on them. A task that Arlequin will perform is marked with a purple circle, tasks that will not be performed have a grey circle.
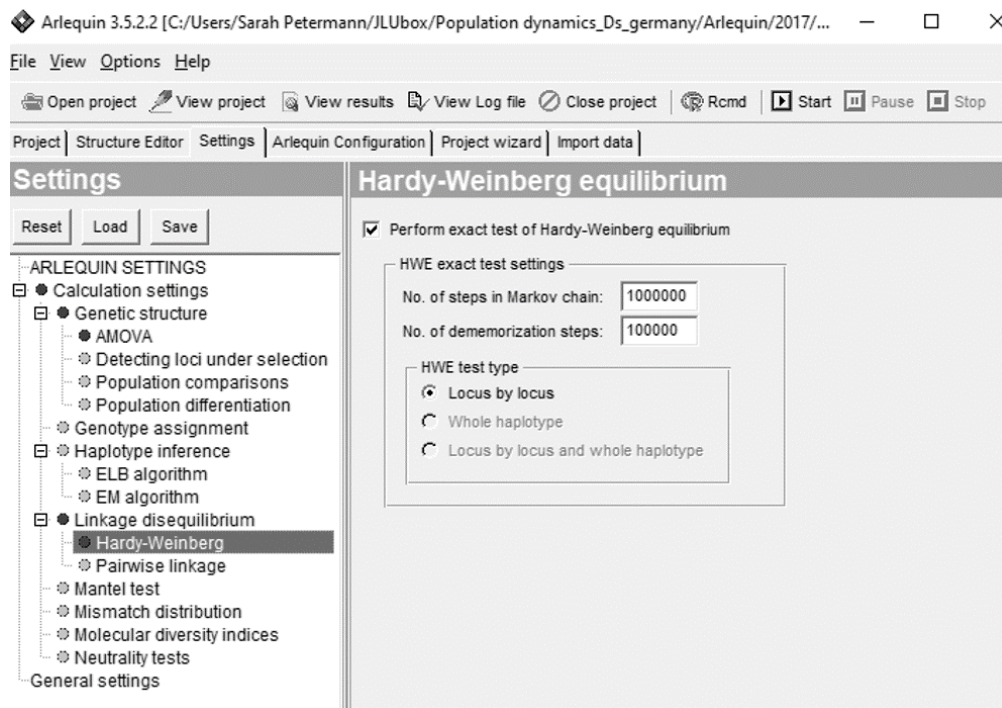
Figure C.5.5: Calculation settings in Arlequin. The left panel allows the user to choose which tasks to perform, the specific options for each task can be seen in the right panel.

9. Click on 'START' in the upper toolbar once all tasks are set.
10. Output files have the same name as the project but have an extension ('.res').

C.6 STRUCTURE

STRUCTURE 2.3.4 can be run on different platforms. (Windows, Mac OS X, Linux, and Sun). The executable installation file for Windows can be downloaded from the web page https://web.stanford.edu/group/pritchardlab/structure.html. Download the file to a directory and unzip the file. Double click on the executable file to start the installation. After the installation is finished, double-click on the STRUCTURE icon to start the program. The data input file can be generated using the Export function in GenAlex.

Starting a new project

1. If you are starting a new project, click on 'File' and on 'New Project' (Figure C.6.1).
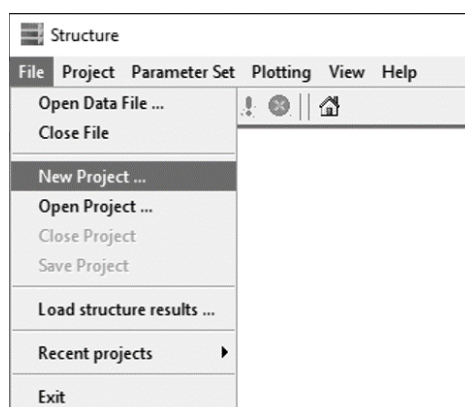
Figure C.6.1: Starting a new project with STRUCTURE. Click on File and then on New Project. You can also view data files but this will not allow you to analyze them. If you already made a project, then click on Open Project and you can browse your directories for files.
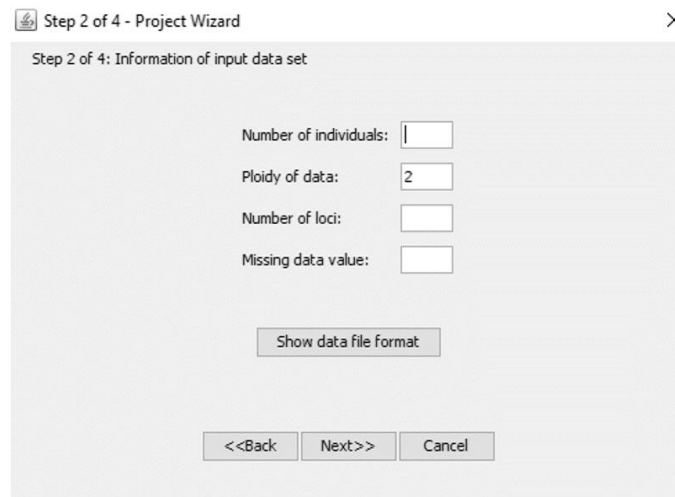
2. A project wizard opens that guides you through four steps to enter your data information.
3. In 'Step 1 of 4' specify the name, project directory, and data file (Figure C.6.2).
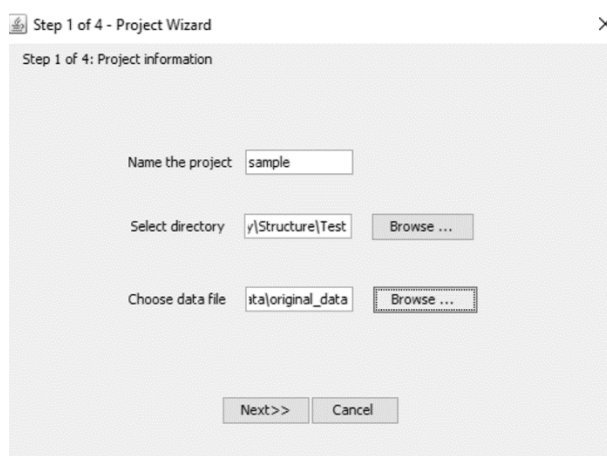
Figure C.6.2: Step 1 of 4 in the Project Wizard. Specify the name, directory of the project, and the data file for the new project. The input file can be generated using GenAlex.

4. Click 'Next'.
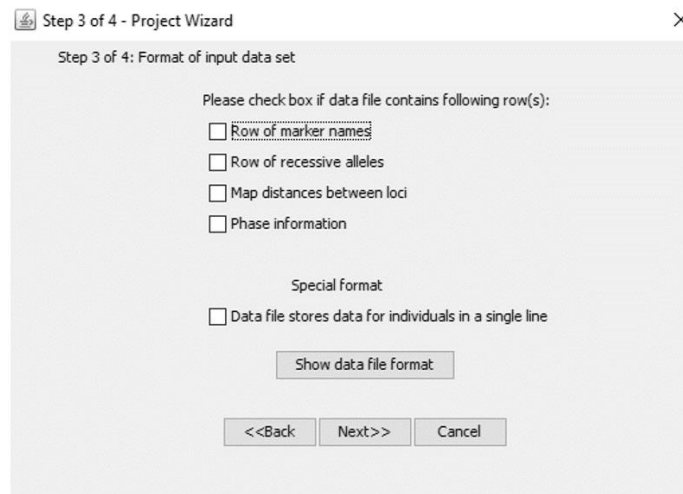
5. In 'Step 2 of 4' specify the data file in more detail. Give the number of individuals, the ploidy of the data (2 for diploid), number of loci, and the missing data value (you can use the value '-9') (Figure C.6.3).

Figure C.6.3: Step 2 of 4 in the Project Wizard. Specify additional information on your data. The number of individuals, the ploidy, the number of loci, and the missing data value.

6. In 'Step 3 of 4' you have to specify if there are any additional extra rows in your data. This can be a row with marker names (Figure C.6.4). Also, if you used GenAlex to format your data then check the box 'Data file stores data for individuals in a single row'.

Figure C.6.4: Step 3 of 4 in the Project Wizard. Check or uncheck boxes, if your data file contains optional rows. When using GenAlex check the box 'Data file stores data for individuals in a single row' on the bottom. You can see a summary of the length and number of lines in your file by clicking on 'Show data file format'.

7. 'Step 4 of 4': Specify any additional columns in your data (Figure C.6.5). If the input file was formatted using GenAlex, check the boxes 'ID for each individual', 'Putative population origin for each individual', and 'Sampling location information'.
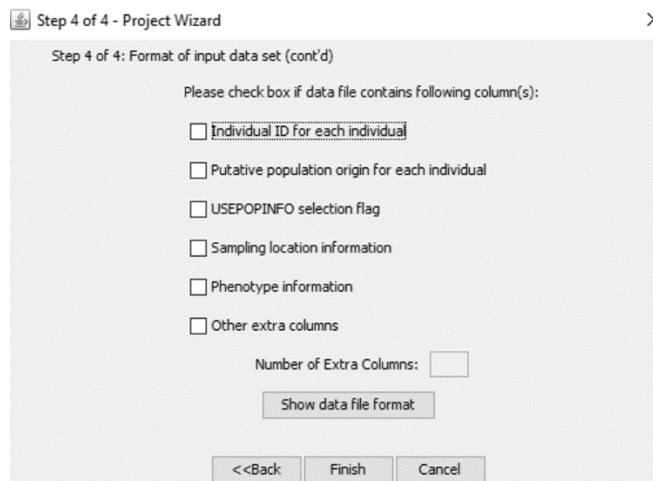
Figure C.6.5: Step 4 of 4 in the Project Wizard. Check or uncheck boxes, if your data file contains optional columns. When using GenAlex check the first, second, and fourth boxes. If the data contains any additional columns not listed in the options, type the number of additional columns into the box called 'Number of Extra Columns'. You can see a summary of the data format by clicking on 'Show data file format'.

8. Click on 'Finish'.

Configuring a parameter set

1. Go to the pull-down menu 'Parameter set' (Figure C.6.6).
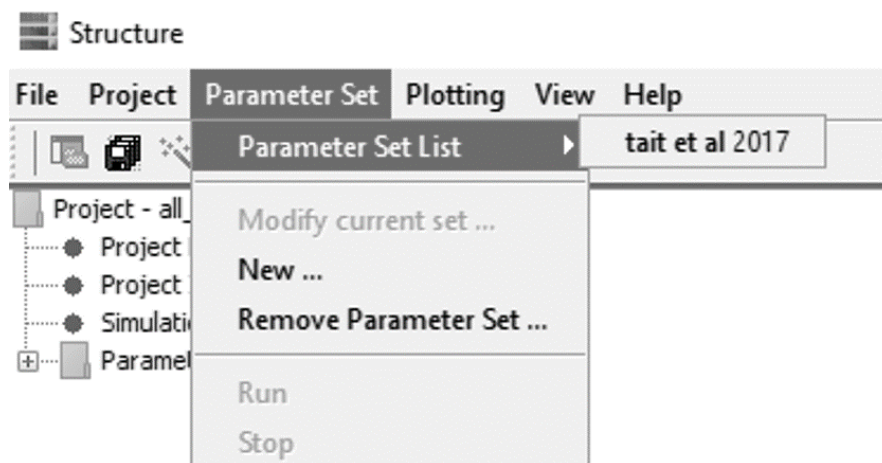
Figure C.6.6: Configuring a new parameter set. You can load, modify, delete or create a new parameter set by going to the pull-down menu under 'Parameter Set'. Existing parameter sets will be listed under 'Parameter Set List.

2. Click on 'New'.
3. Start by setting the run length (Figure C.6.7).
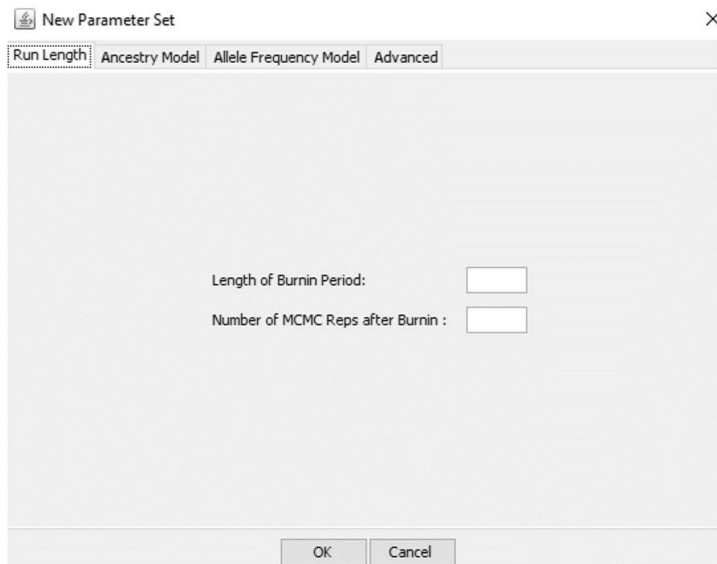
Figure C.6.7: Specifying the run length. The user can choose the number of Markov Chain Monte Carlo (MCMC) repetitions and the length of a burn-in period. The number of MCMC repetitions after burn-in defines how long a simulation is run after burn-in to get accurate parameter estimates.

4. Select if the admixture model should be applied or not (Figure C.6.8). It is used if the origin and degree of isolation in the sampled populations are unknown. It assumes that allele frequencies are correlated and each sample contains a portion of the genome of each original population.
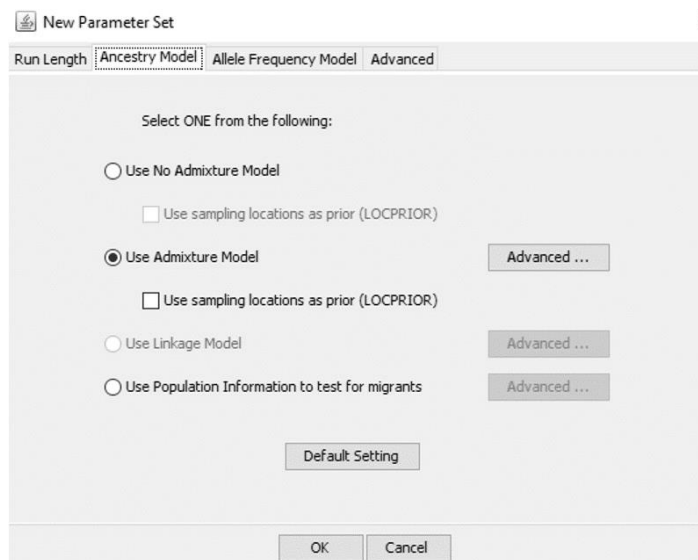
Figure C.6.8: Specifying the ancestry model. The user can select if the admixture model should be applied or not. A test for migrants in the population can be applied by checking the corresponding box.

5. Set the parameters in the 'Allele Frequency Model' tab (Figure C.6.9). The correlated frequencies model is better suited to detect subtle population structure.
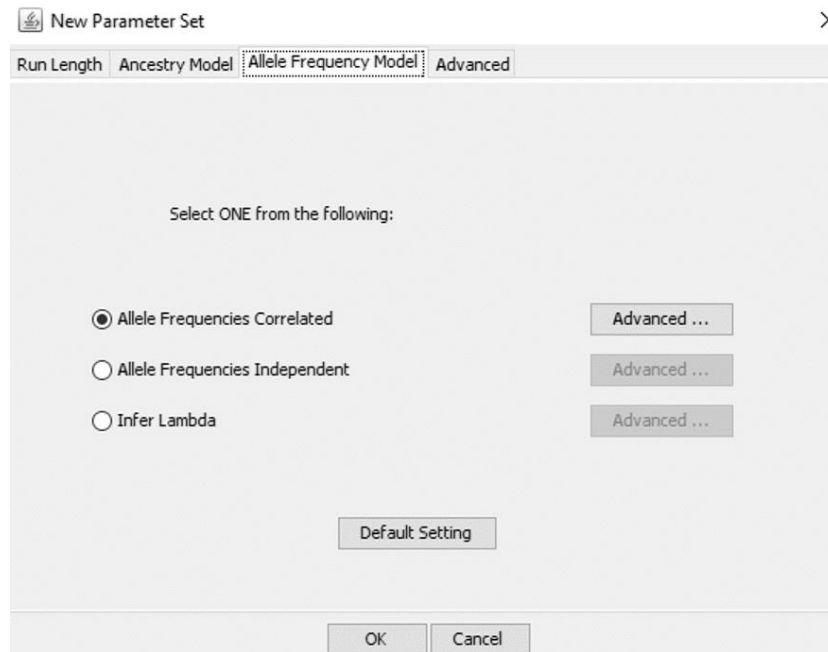
Figure C.6.9: Specifying the allele frequency model. The user can select between correlated allele frequencies, independent allele frequencies, and inferring lambda.

6. With the 'Advanced' options, the function to calculate posterior probabilities can be turned off, which will speed up the calculation time (Figure C.6.10).
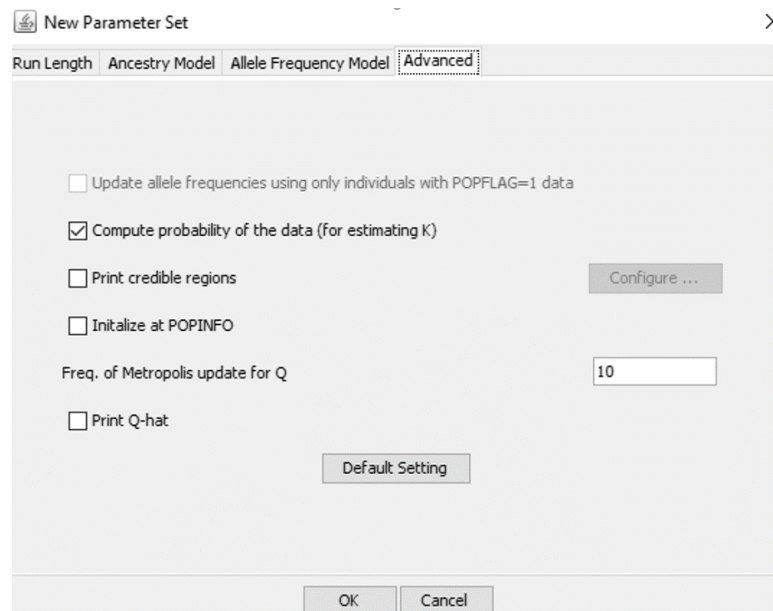
Figure C.6.10: Advanced options. Turning of the calculation of posterior probabilities will speed up the program. This can be used, when testing a data set the first time. But it should be used later on to get reliable results.

7. Once a parameter set is ready, click on 'Run' in the pull-down menu under 'Parameter set'.
8. The program asks you to set the number of populations (K).
9. You can see text data in the console at the bottom and a summary of simulation configuration in the window on the right (Figure C.6.11).
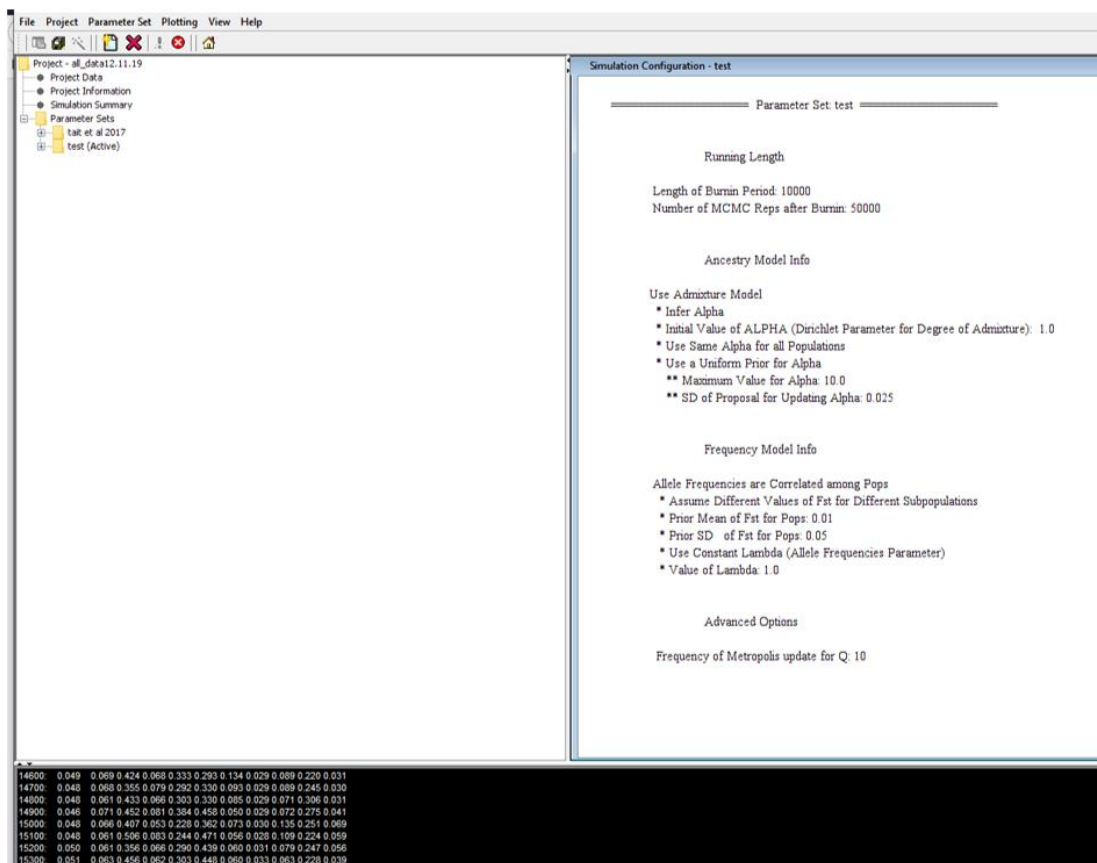
Figure C.6.11: Running a simulation. The console at the bottom prints real-time summary statistics in text form. Simulation configurations are visible in the right window.

Note: Running a simulation will take a long time depending on the settings and the data. STRUCTURE can crash easily. It is a so-called 'silent crash', so you might not even notice it. For me, it worked best to start the run and to not touch the computer until the run was over. Running overnight worked well, since a single run can easily take six or more hours. Just make sure that the computer does not shut down after a certain time, also disable energy-saving options.

10. Once a run is finished you are presented with the results. All results are present in the directory you chose at the beginning. You can have a look at the results in the right window (Figure C.6.12). To look at a bar plot, click on the 'Bar plot' button in the results window.
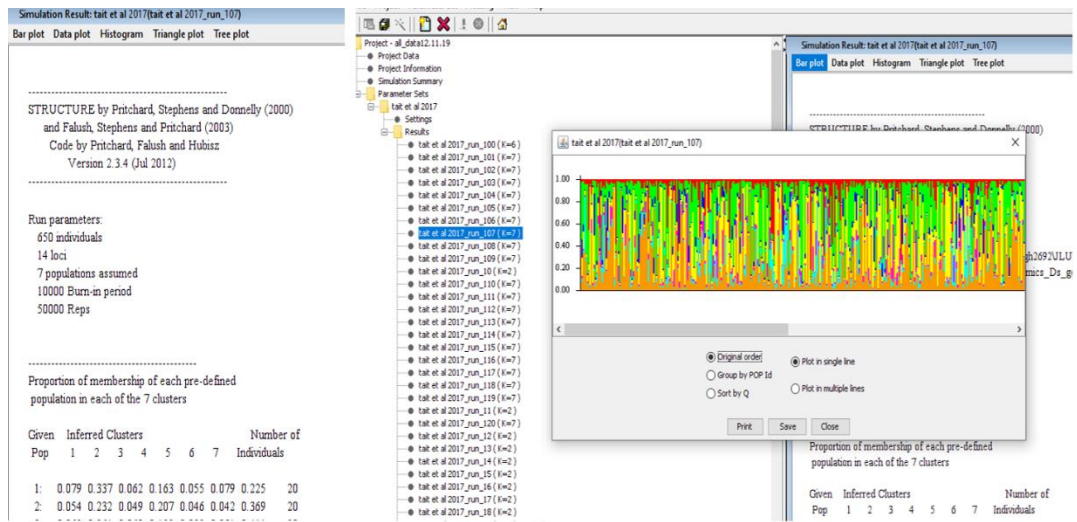
Figure C.6.12: Simulation results. Results for each run contain a summary of parameter sets. To see a bar plot, click on the Bar plot button at the top of the window.

The vast amount of information generated by STRUCTURE makes it difficult to analyze the results. That is why it is helpful to use programs like StructureHarvester and Pophelper afterward. StructureHarvester can summarize the results and help to make decisions on how to proceed with the data. Pophelper will help to make one comprehensive bar plot from the many bar plots generated in STRUCTURE.

C.7 StructureHarvester

StructureHarvester is an online web tool freely accessible at http://taylor0.biology.ucla.edu/structureHarvester/. Use Firefox, Safari, or Chrome.

1. Run STRUCTURE and open the results folder in your STRUCTURE directory.
2. Zip all results in one zip archive.
3. Click on the 'Browse' button on the StructureHarvester website and select your archive.
4. Click on the button 'Harvest!'.
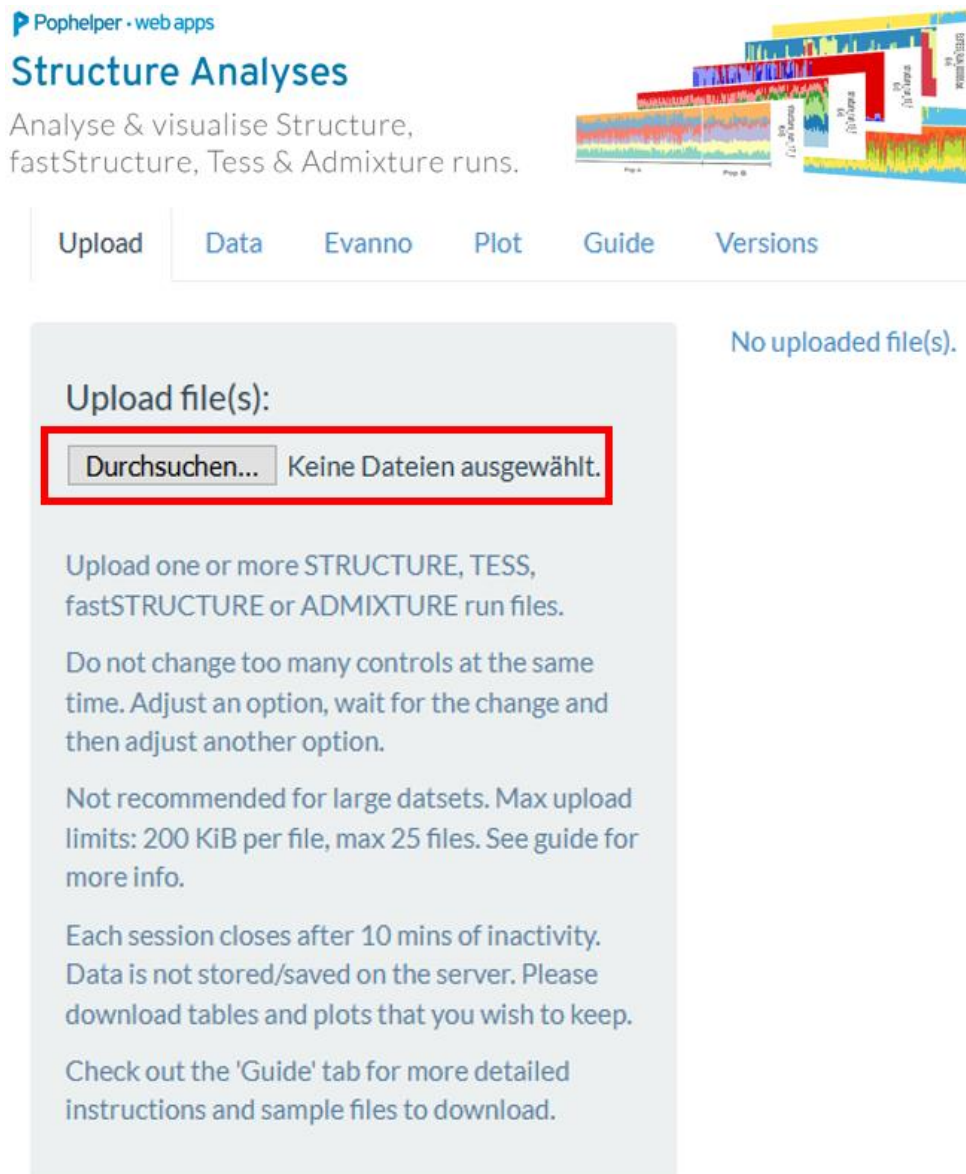
You will get the results online. You can download everything at once or as separate files. Based on the result, decide with which data you want to proceed in Pophelper.

C.8 Pophelper

Pophelper is a web app (http://pophelper.com/) that allows the user to analyze and visualize Structure runs.

1. Upload one or more STRUCTURE files (Figure C.8.1).

Figure C.8.1: Structure analysis with Pophelper. To upload STRUCTURE files, press the 'Browse' button (red square). The upload is limited to 25 files, so it does make sense to make a preselection with StructureHarvester. Uploaded files are presented on the right side.

2. A summary table should now be displayed.
3. The 'Tabulated data' lists all uploaded STRUCTURE runs sorted by loci, individuals, and K (Figure C.8.2).
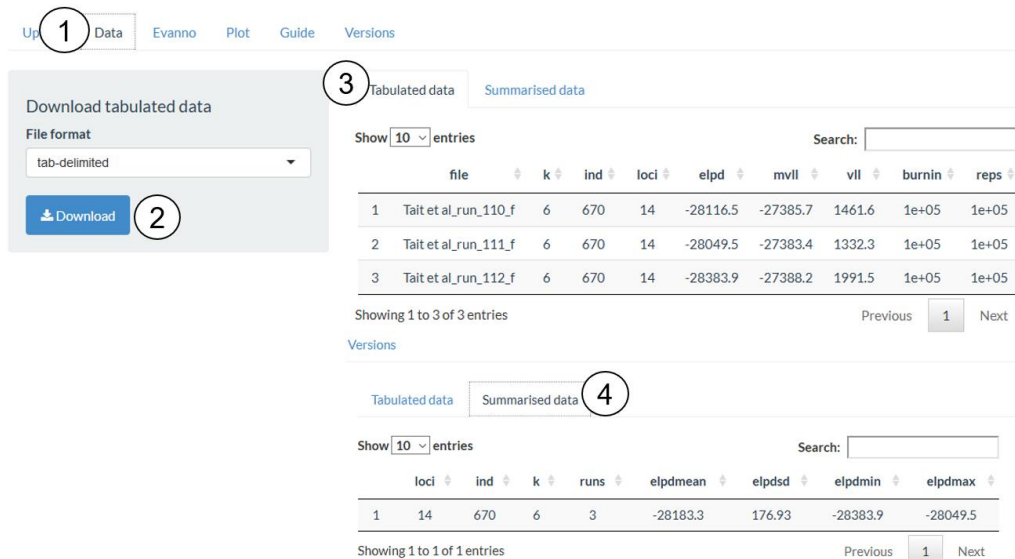
Figure C.8.2: Tabulated data list and summary table. Clicking on 'Data' (1) will open a table with data information. You can download the table either tab-delimited, comma-separated, or semicolon-separated (2). The data can either be displayed in a tabulated format (3) or a summarized version (4). The table contains information on the number of populations (K), number of individuals, and loci, as well as on the statistics.

4. To download, click on the 'Download' button on the left side.
5. The Evanno method is run in the Evanno tab. It estimates the number of K. If you used StructureHarvester, you do not need to do this because it was already calculated but you can compare the result.
6. Use the 'Standard plot' option in the 'Plot' tab to plot the files in a bar graph.
7. Select one or more of the uploaded files by clicking on it. The selected file(s) are plotted on the right site. If more than one file is selected, the bar plots are plotted one below the other (Figure C.8.3).
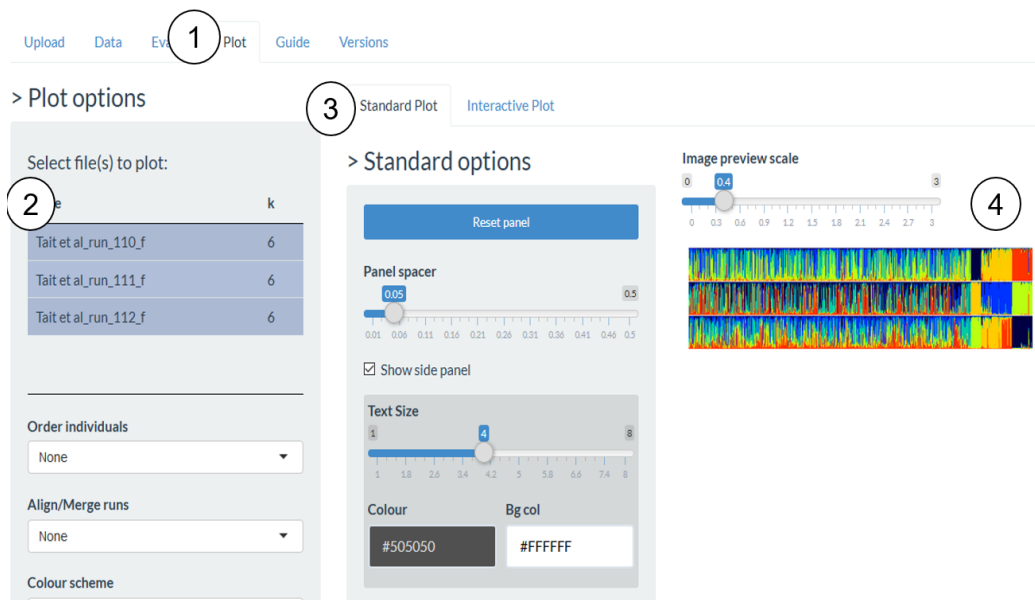
Figure C.8.3: Plotting data as bar plots. Clicking on 'Plot' (1) will open the 'Plot options'. Select the files you want to plot by clicking on them (2). Make sure the 'Standard Plot' option is selected (3). Each selected data file is displayed as a barplot on the right (4).

8. If you selected more than one file, a drop-down menu is available that allows you to align clusters or to merge runs. This does only work when all selected files have the same format, the same number of individuals, the same number of loci, and the same K. 'Align clusters' will plot the selected files with aligned clusters. 'Merge runs' will collapse all selected files to one bar plot. You can also choose between different color schemes that are either pre-defined or custom generated (Figure C.8.4).
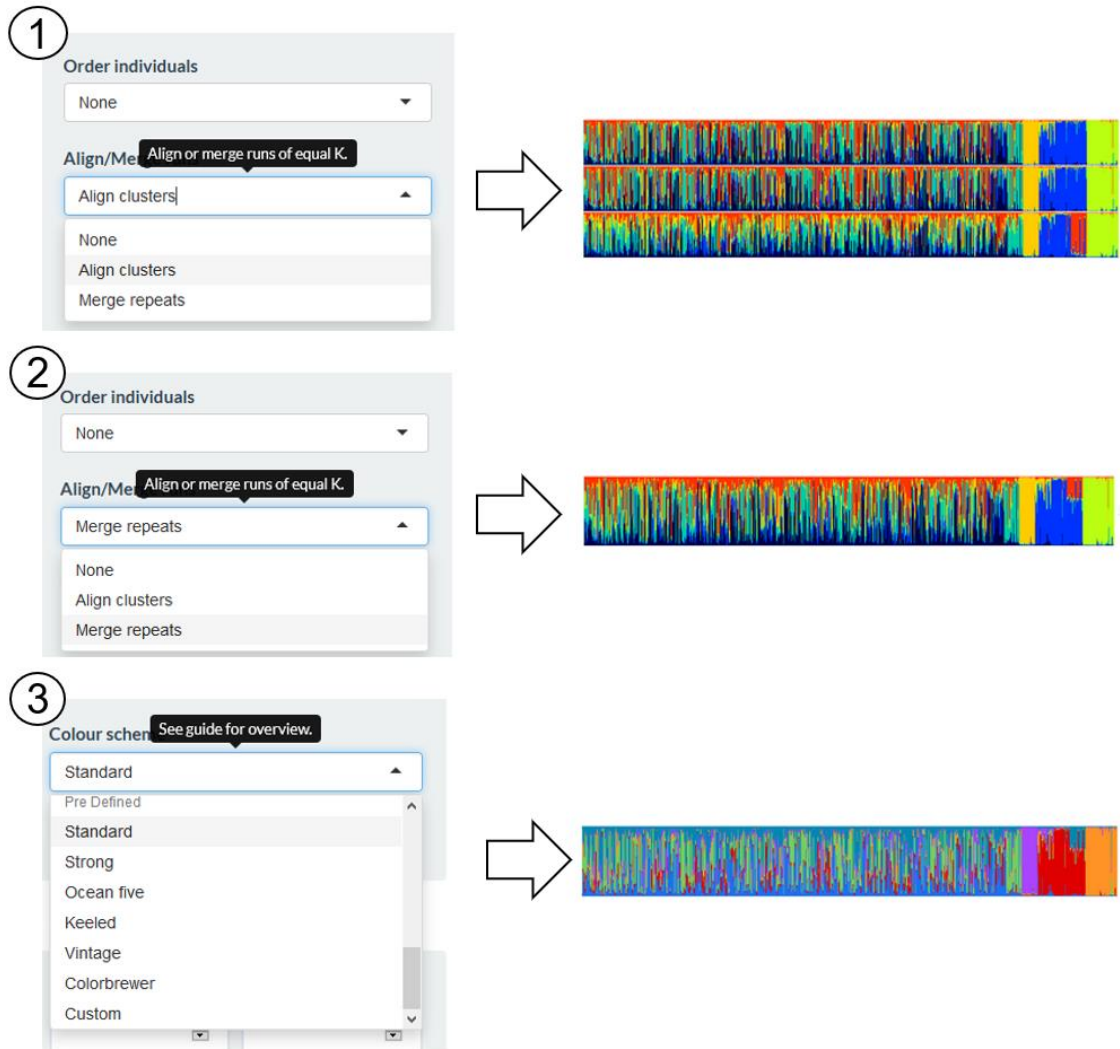
Figure C.8.4: Plot options include aligning, merging, and different color schemes. 'Align cluster' will plot the selected files with aligned clusters (1). 'Merge runs' will collapse all selected files to one bar plot (2). In addition, there are several different color schemes available to color the barplot (3).

9. Population labels can be uploaded as a tab-delimited text file, copy-pasted, or manually typed in. One label is needed for each individual, e.g. if you have 20 individuals from population A you have to type A in the first 20 rows (Figure C.8.5).
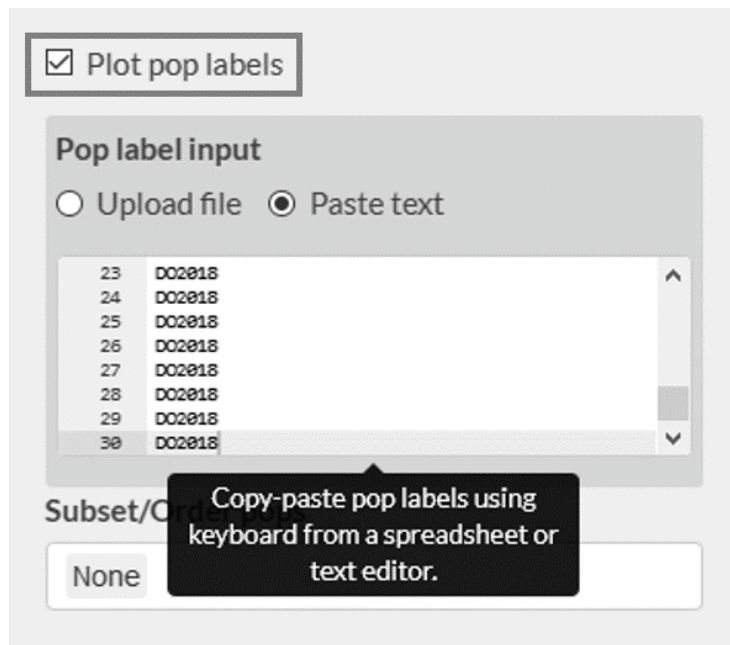
Figure C.8.5: Adding population labels to the bar plot. To add population labels, check the box 'Plot pop labels' and either upload a text file or copy-paste the population names.

10. Change standard options like the file name and K value. Those are visible in the side panel and can be turned off or adjusted. Population label options include height, label spacer, label text, and label markers (points or lines) as well as label points (Figure C.8.6).
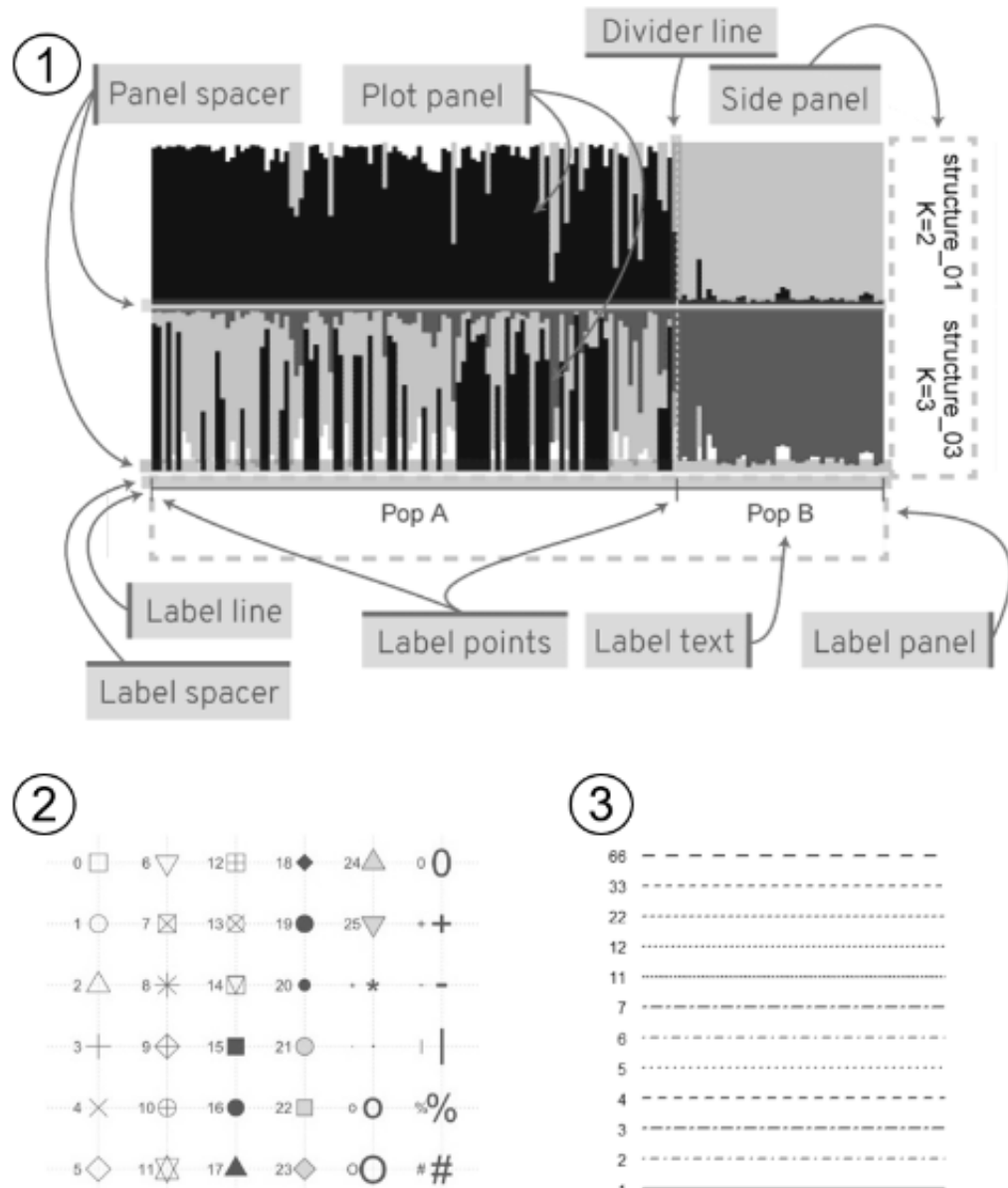
Figure C.8.6: Customization options for bar plots. Shown are the customizable parts of a bar plot (1). Label points can be changed by choosing the corresponding number (2), for example, the number '1' will generate a circle as a label point and number '2' will generate a triangle. The line type for the divider line can be changed as well by using a number code (3). The graphics in this image are from the manual from the Pophelper website (http://pophelper.com/).

11. Download the bar plot.

C.9 PoptreeW

PoptreeW is the web version of Poptree2. It is used for the construction of population genetic trees based on allele frequency data.

1. Open http://www.med.kagawa-u.ac.jp/~genomelb/takezaki/poptreew/. This site provides access to PoptreeW as well as to additional guides and references and to Poptree2, which is the program for Windows and Linux (Figure C.9.1).
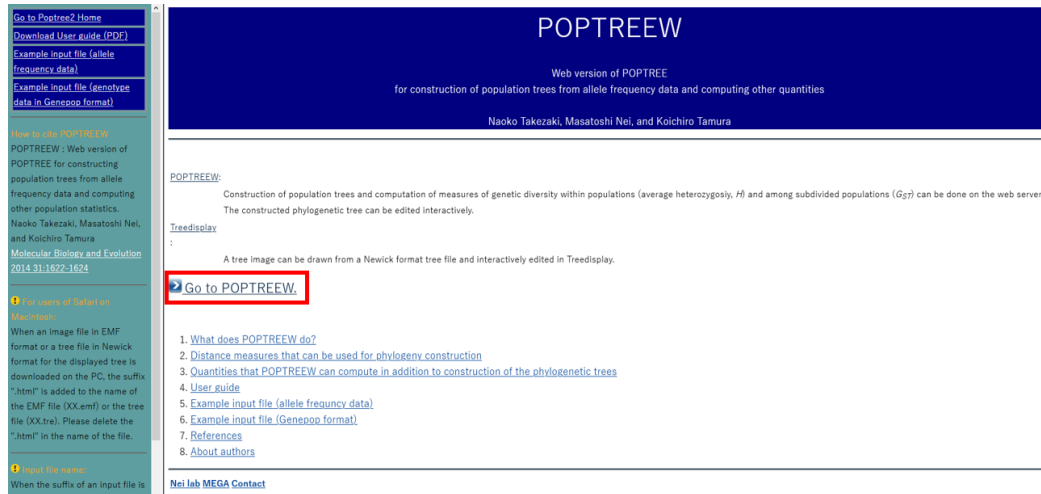
Figure C.9.1: PoptreeW website for phylogenetic trees from allele frequency data. Shown is a screenshot of the PoptreeW website. The site includes a guide and several links for references and additional information on the web app and the original program Poptree2. Click on 'Go to POPTREEW' (red square) to start.

2. Click on 'Go to POPTREEW'.
3. Click on 'POPTREE' to generate a new tree. Click on 'Tree display' to draw a tree from a Newick format tree (Figure C.9.2). Click 'Next' to proceed.

Figure C.9.2: PoptreeW and Tree display option. Start by choosing if you would like to generate a phylogenetic tree or to draw a tree from a Newick format tree.

4. Upload the input file. Allele frequency data in Genepop format can be used as input. To get the Genepop format use the 'Export' function in GenAlex (see C.3 GenAlex).
5. Select the computational methods. You can either construct a phylogenetic tree or let the program compute the heterozygosity and G_{ST} but you cannot do both at once. When using the phylogenetic tree option, then you need to choose a distance measure from the drop-down menu. Choose between an NJ or UPGMA tree. Check the box 'Bootstrap'

214

test, if you would like to perform a bootstrap test for the tree. Specify the number of replications (Figure C.9.3).
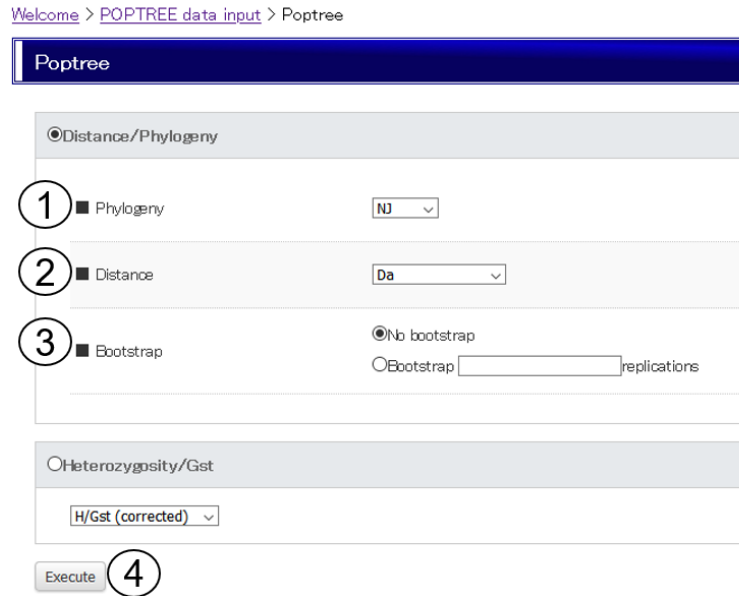
Figure C.9.3: Population tree options. Choose between NJ and UPGMA tree (1), the used distance measure (2), and if you would like to perform a bootstrap test and the repetitions (3). Click the 'Execute' button to proceed (4).

6. Click 'Execute'.
7. The 'Tree View' window opens and you can edit the tree after your likings by clicking on 'Edit' (Figure C.9.4).
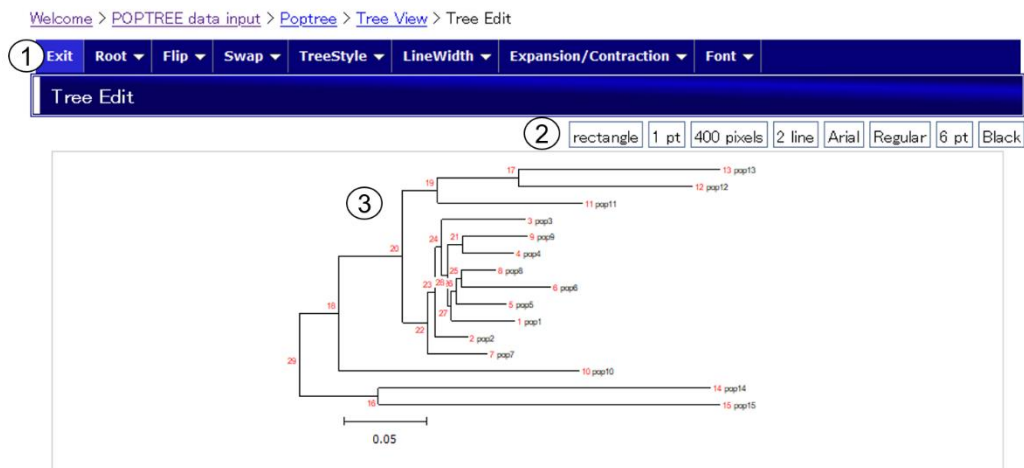
Figure C.9.4: Modification options for population trees. You can edit the tree with different options, for example setting a new root, swapping branches, or changing the font and line width (1). The current details of the tree are given in the top right corner (2). The tree is displayed in the big window below the toolbar (3).

8. Download the tree as an image or in Newick format. Latter allows you to load it in other programs if you would like to edit the tree further.

C.10 Geneclass2

The executable installation file for Windows can be downloaded from the web page http://www1.montpellier.inra.fr/URLB/GeneClass2. Download the file to a directory and unzip the file. Double click on the executable file to start the program. The data input file should be in 'Genpop' format and can be generated using the 'Export' function in GenAlex (see C.3 GenAlex).

Detection of first-generation migrants

To detect migrants in a data set, you need one file that contains both the populations you want to check for migrants and the potential source populations.

1. Open a data file (Figure C.10.1).
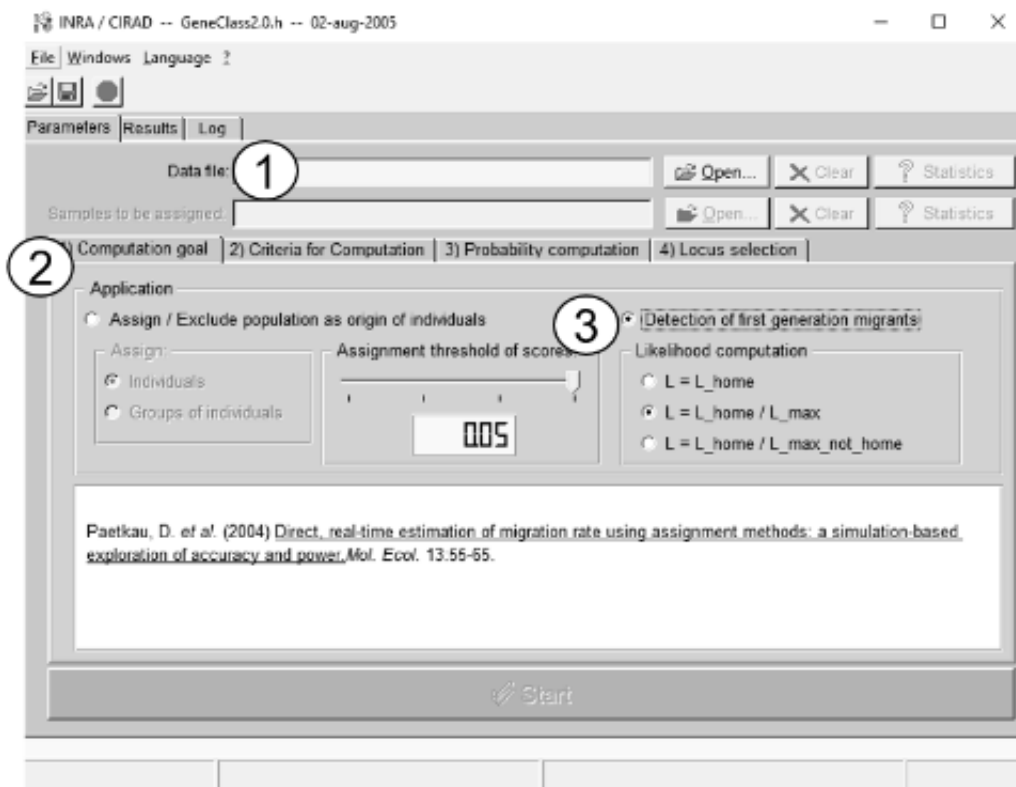
Figure C.10.1: Detection of first-generation migrants. Load a data file in the program (1). Set the first options in the 'Computation goal' tab (2) by clicking on the 'Detection of first-generation migrants' option (3). If you choose this option, you have to select a likelihood computation method as well.

2. Check the 'Detection of first-generation migrants' option (Figure C.10.1).
3. Choose a type of likelihood computation used for migrant detection (Figure C.10.1).
4. Open the 'Criteria for Computation' tab (Figure C.10.2).
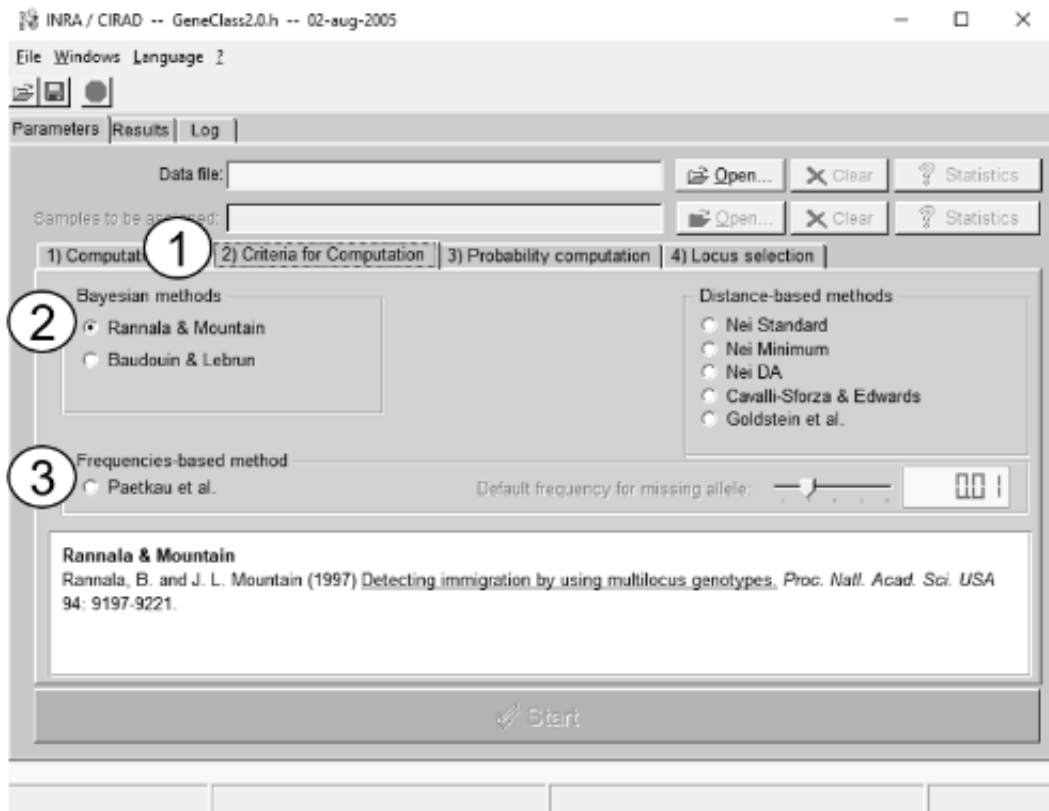
Figure C.10.2: Setting the criteria for computation. Open the 'Criteria for Computation' tab (1). Choose between the three method options, Bayesian, Frequency-based, or Frequency-based (2). When selecting the 'Frequency-based method', you have to set the default frequency for missing data as well (3).

5. Select criteria for likelihood computation. You can choose between Bayesian, Distance-based, and Frequency-based methods (Figure C.10.2).
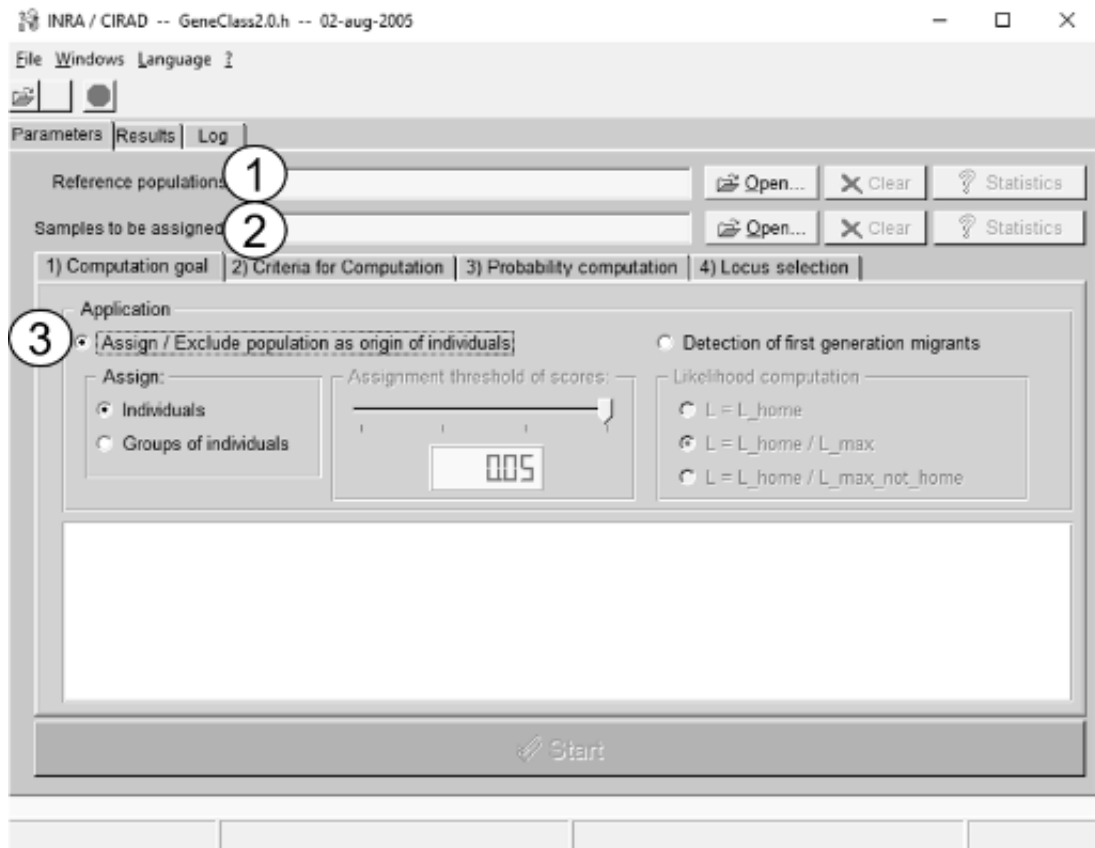6. Open the 'Probability computation' tab (Figure C.10.3).
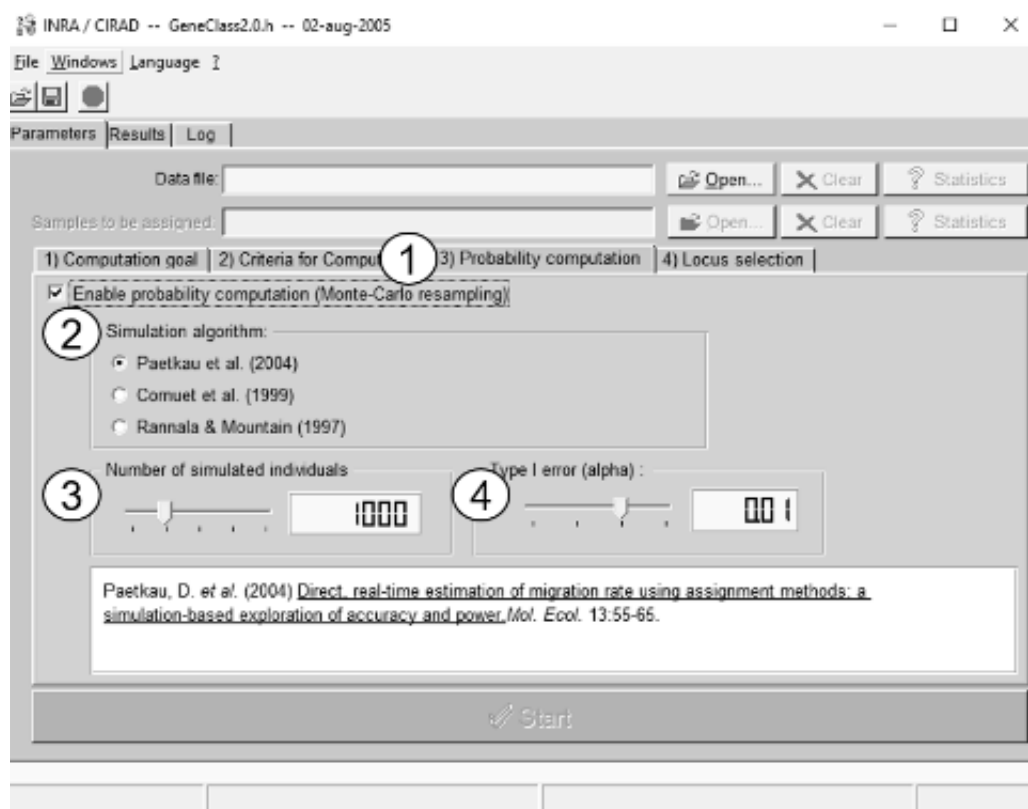
Figure C.10.3: Setting the probability computation options. Open the 'Probability computation' tab (1). Click on 'Enable probability computation' to compute the probability that an individual is a resident and not a first-generation migrant (2). Use the sliders to select the number of simulated individuals (3) and the alpha value (4).

7. You can click on 'Enable probability computation' to compute the probability that an individual is a resident and not a first-generation migrant (Figure C.10.3).
8. Select a resampling algorithm (Figure C.10.3).
9. You can deselect loci in the 'Locus selection' tab.
10. Click the 'Start' button.
11. The running parameters are displayed, together with a progress bar at the bottom.
12. Results are displayed in a table and potential migrants are labeled red, depending on the threshold you set. The most likely population is labeled green.
13. Export the results in CSV format.

Assignment of individuals

You need two files for this computation, one reference file and one file containing the individuals that have to be assigned.

1. Load the reference file as 'Reference population' and the other data file as 'Samples to be assigned' (Figure C.10.4).
2. Select 'Assign in the Application' options and decide if you like to assign individuals or groups.

Figure C.10.4: Assignment of individuals. Load a reference file (1) and a second file with samples to be assigned (2). Select the Assign option in the Application window (3).

3. Open the 'Criteria for Computation' tab and select a criterion. Choose between Bayesian, Distance-based, and Frequency-based methods (Figure C.10.2).
4. To calculate the probability that an individual belongs to each reference, click on 'Enable probability computation' (Figure C.10.3).
5. Select a resampling algorithm (Figure C.10.3).
6. You can deselect loci in the 'Locus selection' tab.
7. Click the 'Start' button.
8. The running parameters are displayed, together with a progress bar at the bottom.
9. Results are displayed in a table. If a given individual's probability in a reference population is lower than the threshold, the value is greyed out.
10. Export the results in CSV format.

C.11 Bottleneck

Download the executable Windows file for the program Bottleneck from the website http://www1.montpellier.inra.fr/CBGP/software/Bottleneck/bottleneck.html. Download the file to a directory and unzip the file. Double click on the executable file to start the program. The data input file should be in Genpop format and can be generated using the 'Export' function in GenAlex (see B3 GenAlex).

1. Load the data file (Figure C.11.1).
2. Select a mutation model.
3. Set the number of iterations.
4. Select the statistical test you would like to perform.
5. Click on 'GO!' To start the calculation.
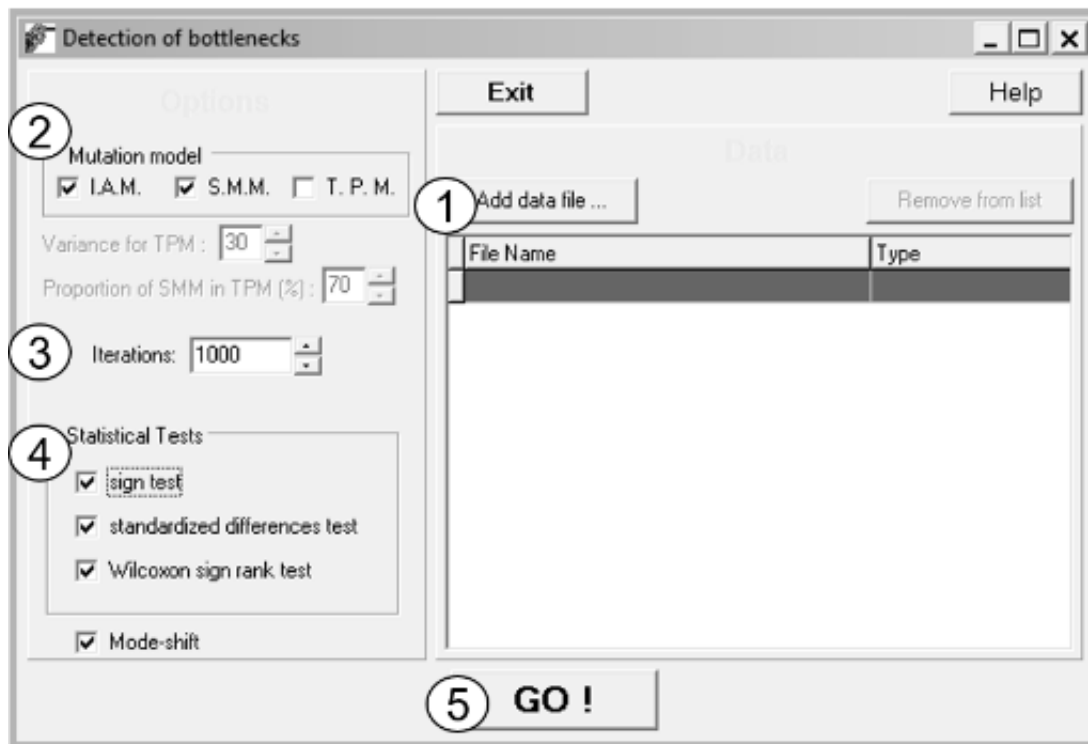
Figure C.11.1: Detection of bottleneck events with Bottleneck. Load a data file (1). Select the mutation model, you can run more than one at the same time (2). Select the number of Iterations (3) and choose statistical tests (4). Start the calculations by clicking on 'GO!'.

C.12 Cervus

Download the executable Windows file for the program Cervus from the website http://www.fieldgenetics.com/pages/home.jsp. Download the file to a directory and unzip the file. Double click on the executable file to start the program. The data input file should be in crv format and can be generated using the Export function in GenAlex (see B3 GenAlex).

1. Load a data file.
2. Click on 'Analysis' and choose the option 'Allele Frequency Analysis' (Figure C.12.1). This will provide you with allele frequency results like the number of alleles, expected heterozygosity, and other general frequency-based results. More important is, that Cervus calculates a reliable PIC value for the used microsatellite markers.
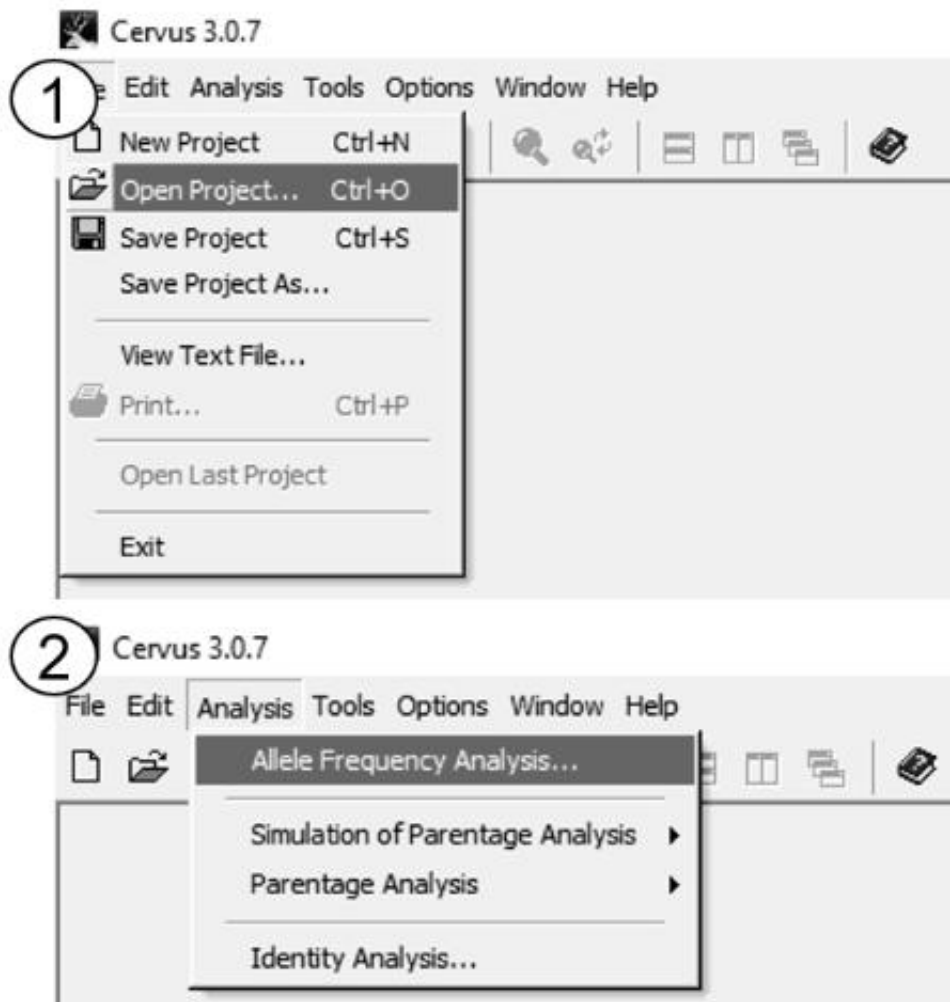
Figure C.12.1 Calculation of the PIC value with Cervus. Load a data file (1). Click on 'Analysis' and click on 'Allele Frequency Analysis' (2)

Curriculum Vitae

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.

Comment on Data Usage in This Thesis

Results shown and discussed in this thesis have been published previously in the Journal of Pest Science (https://doi.org/10.1007/s10340-021-01356-5). Thus, style and wording of some aspects can have some similarities to the publication. The publication is also attached in the Appendix in chapter A.1.1 Publication 'Spatial and temporal genetic variation of *Drosophila* suzukii in Germany'.

The raw data used in this thesis as well as in the publication can be viewed at the research data repository of the Justus-Liebig-University named JLUdata (http://dx.doi.org/10.22029/jlupub-179). The folder contains three Excel files. The first one is named '001_Fragment_Length_Analysis_raw_data' and it was generated by using the microsatellite external plugin of Geneious Prime 2019.2. The second file is named '002_Null_allele_detection' and contains the data generated with FreeNA (Chapuis and Estoup, 2007). The third file is named '003_Bottleneck_result' and contains the results obtained with Bottleneck v.1.2.2 (Piry et al., 1999).

Danksagung

Die Danksagung wurde aus der elektronischen Version der Arbeit entfernt.

The acknowledgement was removed from the electronic version of the paper.

Eidesstattliche Erklärung

Erklärung gemäß der Promotionsordnung des Fachbereichs 09 vom 07. Juli 2004 § 17 (2)

„Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe.

Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten."

Gießen, den 08.12.2021

_____

Sarah Petermann