INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior Agrária

Development of a cost-effective SNP tool to detect genetic pollution in honey bee spermatheca content

Chiraz Soltani

Dissertation submited to Escola Superior Agrária de Bragança to obtain the Degree of Master in Biotechnological Engineering

Supervised by

Dora Sofia Martins Henriques

Maria Alice da Silva Pinto

This dissertation includes the comments and suggestions mentioned by the Jury

Bragança
2023

This work was supported by the German Federal Ministry for Food and Agriculture, (project no. 2818BM040). FCT (Fundação para Ciência e Tecnologia) provided financial support by national funds (FCT/MCTES) to CIMO (UIDB/00690/2020 and UIDP/00690/2020) and SusTEC (LA/P/0007/2021).

INDEX

List of abreviations

AIMs: Ancestry-Informative Markers

***A.m.*:** *Apis mellifera*

bp: Base pairs

CLUMPAK: Clustering Markov Packager Across K

DNA: Deoxyribonucleic Acid

FST: Fixation index of genetic differentiation

g: Gram

HiSeq: HiSeq Sequencing Systems

HS: High Sensitivity

K: Number of cluster

Lineage A: *African lineage*

Lineage C: Eastern Europe lineage

Lineage M: Western and Northern European lineage

Lineage O: Near East and Asia lineage

LLS: Log-Likelihood Scores

MALDI-TOF: Matrix-assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometry

MAF: Global minor allele frequency

Mbp: Maltose-binding protein

Ml: *millilitre*

NGS: Next-Generation Sequencing

°C: Celsius degree

PCR: Polymerase Chain Reaction

Q: Matrix of membership proportions

µ: Average

rpm: rotations per minute

RFLP: Restriction Fragment Length

RNA: Ribonucleic Acid

SNPs: Single Nucleotide Polymorphisms

UMI: UMI Adapter

Vcf: variant call format

WG: Whole Genome

WGS: Whole Genome Sequencing

3′ UTR: 3′ Untranslated Region

5′ UTR: 5′ Untranslated Region

µl: microlitre

List of figures

List of tables

Acknowledgments

I would like to show my sincere appreciation to those who have contributed to this dissertation and supported me in one way or the other during this amazing journey for without the generous support of them, this work would not have been possible.

First of all, I would like to express my pride to be supervised by widely known and highly experienced scientific researchers.

Professora Dora Sofia Martins Henriques and Professora Maria Alice da Silva Pinto, I am very thankful to have such a wonderful, professional and perfectionist persons as my thesis supervisors. I am deeply thankful for your continuous support and patience in this learning path, for your persistent help, and for the constant motivation to do more and better. You offered continuous advices and encouragement and I benefitted from your knowledge and scientific experience. I appreciate your effort and I want to thank you for your wise guidance, kindness and continuous availability.

I also want to thank all researchers of the Mountain Research Centre (CIMO) for their support and generosity. Thank you all for always being so helpful and friendly, as well as for all your efforts and support during the realization of this work. I am deeply grateful to all members of the jury who kindly honored me by their presence to participate in the defense of this thesis.

Keeping the best for last, I would like to thank my parents Hedia Mihoub, Naceur Soltani, my brothers Achref, Mahdi, Ali, Ilyes and my sisters Ines, Aida, Najoua, Dina, Sonia whose help made this journey possible from the beginning they believe in me and supported my studies abroad both morally and financially.

My big appreciation to my special and wonderful person in my life Mehdi, who helped me in every possible way to achieve my dreams and successfully complete this research work. I will never forget all your support and effort, thank you for your love, your help, your patience and especially your moral support.

They have been my inspiration and my unfailing source of passion and energy. Thank you for your irreplaceable support, for all your love, patience and kindness, for always encouraging me to pursuit my education, for believing in me and for everything that you have done to make me the person I am now. Without you none of this would have been possible. I love you!!

ABSTRACT

The loss of genetic complexes adapted to local conditions through genetic introgression is one of the many threats affecting the honey bee (*Apis mellifera*). Large-scale displacement of honey bees has altered their native distribution. One extreme example is *A. m. mellifera*, a European subspecies that was once widespread but is now threatened with extinction in numerous countries due to introgression and replacement by the C lineage. As a result, conservation measures may be needed to preserve the genetic diversity of these subspecies. Acknowledging the significance of native genetic diversity, several conservation and breeding programs have been established. Their effectiveness hinges on the availability of accurate and cost-effective molecular tools for assessing subspecies introgression. Whole-genome data has offered valuable insights into honey bee evolution, yet its practical application is hampered by the need for specialized bioinformatics expertise and computational resources, often unavailable in conservation and breeding centers. To bridge this gap, a novel SNP (Single Nucleotide Polymorphism) tool, based on the NEBNext Direct Genotyping Solution, has been developed. This tool was designed from 228 whole-genome sequence data generated from 148 M-lineage drones and 80 C-lineage drones. From 5,007 highly differentiated SNPs, we selected 130 SNPs. After eliminating problematic SNPs, we retained 82 SNPs that demonstrated exceptional accuracy in estimating the degree of genetic introgression in known samples. This innovative tool represents a significant advancement in the genetic analysis of honey bee colonies, with applications spanning breeding and conservation efforts for *A. m. mellifera*, *A. m. iberiensis*, *A. m. carnica* and *A. m. ligustica*.

Keywords: Single Nucleotide Polymorphism, introgression, conservation, *Apis mellifera*, whole genome sequencing, NEBNext Direct Genotyping Solution.

RESUMO

Uma das muitas ameaças para as abelhas melífera (Apis mellifera) é a perda de complexos genéticos adaptados localmente devido à introgressão genética. Os movimentos em larga escala das abelhas melíferas têm vindo a alterar a sua distribuição natural. Um exemplo é o da A. m. mellifera, uma subespécie europeia anteriormente amplamente distribuída que agora está ameaçada em muitos países devido à introgressão e substituição pela linhagem C.

O reconhecimento da importância da diversidade genética nativa levou à criação de vários programas de conservação e reprodução. A sua eficácia depende da disponibilidade de ferramentas moleculares precisas e económicas para avaliar a introgressão das subespécies. Os dados de genomas completos têm sido muito importantes para a compreenção da evolução das abelhas melífera, mas a sua aplicação prática é dificultada pela necessidade de conhecimentos especializados em bioinformática e pela necessidade de recursos computacionais, muitas vezes indisponíveis nos centros de conservação. Para colmatar esta lacuna, foi desenvolvida uma nova ferramenta de SNPs (Polimorfismo de Nucleotído Único), baseada na solução NEBNext Direct Genotyping Solution. Esta ferramenta foi feita a partir de 228 genomas completos de 148 zangões da linhagem M e 80 zangões da linhagem C. A partir de 5.007 SNPs altamente diferenciados, foram selecionados 130 SNPs. Após a eliminação de SNPs problemáticos, 82 SNPs demonstraram uma elevada precisão na estimativa do grau de introgressão genética em amostras conhecidas. Esta ferramenta inovadora representa um avanço significativo na análise genética de colónias de abelhas melíferas, com aplicações na conservação de *A. m. mellifera, A. m. iberiensis, A. m. carnica* e *A. m. ligustica*.

Palavras chave: Polimorfismo de nucleótido único, introgressão, conservação, *Apis mellifera*, sequenciação do genoma complete, Solução NEBNext Direct Genotyping

I. INTRODUCTION

1. Framework

The honey bee, *Apis mellifera*, is the most important crop pollinator. Yet, it is under pressure globally due to several factors, such as parasites, pathogens, pesticides and also the large-scale trading of commercial queens and/or colonies, threatening the genetic integrity of native populations (Henriques, Browne, et al., 2018). This is particularly worrisome for honey bees due to their mating system. Drones from the surrounding colonies join in a congregation area that the virgin queens visit to mate with tens of drones (Baudry et al., 1998). The sperm is stored in the spermatheca to fertilize eggs during the queen's life (Yániz et al., 2020). This process increases intracolonial diversity, which is important to colony fitness. However, when foreign and local honey bees are bringing together, this mating system facilitates introgressive hybridization and eventually displacement of native subspecies. This happened, for instance, with *A. m. mellifera*, a formerly widely distributed European subspecies that is now threatened with extinction in many countries due to introgression and replacement by C-lineage (Jensen et al., 2005; Pinto, Henriques, Chávez-Galarza, Kryger, Garnery, van der Zee, et al., 2014; Soland-Reckeweg et al., 2009).

Whole-genome sequencing has provided important insights into honey bee evolution and can identify management units important for conservation and breeding programs. However, adds little value to the real conservation world because their use requires bioinformatics expertise and computational power unavailable in conservation and breeding centres. Nevertheless, whole-genome data can be used to create low-cost tools. Several panels containing a reduced number (< 200) of SNPs have been designed from to address different goals such as (i) identifying Africanized honey bees (Chapman et al., 2015), (ii) estimating C-lineage introgression into *A. m. mellifera* and *A. m. iberiensis* subspecies (Henriques, Browne, et al., 2018; Henriques, Parejo, et al., 2018; Muñoz et al., 2017) or (iii) monitoring diversity in immune genes (Henriques et al., 2021). All these reduced SNP panels have been tailored for genotyping in the MassARRAY MALDI-TOF platform and they do not allow to calculate the exact allele frequency of each loci. The goal of this work was to create a new SNP tool based on the NEBNext Direct Genotyping Solution, a hybridization-based target enrichment approach that allows the estimation of allele frequencies and therefore allow the genotyping single individuals but also pools of individuals and/or spermatheca content.

Finally, the tool will be validated and the results analyzed. This tool will represent a major breakthrough in honey bee genetic analysis and could be applied in different breeding and conservation programs throughout Europe.

2. Objectives

The objective of this work was to use whole-genome sequencing data to design cost-effective and easy-to-use tools to accurately detect genetic pollution in *Apis mellifera* populations of M- and C- lineages. This tool will represent a major breakthrough in honey bee genetic analysis because will allow the use of pools of individuals and semen stored in the queen's spermatheca. To achieve this goal, we followed the following steps:

(i) Design a tool that includes the most informative SNPs to calculate genetic pollution between M- and C-lineage honey bees.

(ii) Genotype the samples using NEBNext Direct Genotyping System.

(iii) Validate the tool using the results obtained with pools of DNA with known allele frequencies and DNA extracted from semen stored in the spermatheca of artificially inseminated queens originating from a breeding program.

II. LITERATURE REVIEW

1. The genus *Apis*

1.1 An overview

Bees comprise approximately 20,000 known species categorized into seven distinct families (Ascher et al., 2014). Among this species, 7 to 12 honey bee species belonging to the *Apis* genus (Arias & Sheppard, 2005; Michael S. Engel, 1999). This genus is a small and morphologically and behaviourall cohesive group (Michener, 2000) and is part of the Apidae family, characterized by the presence of a pollen basket. The *Apidae* family is classified under the phylum *Arthropoda*, class *Insecta*, and order *Hymenoptera* (Table 1).

Bees evolved in response to a shift in food sources, from insect prey to pollen and nectar obtained from angiosperm flowers (Chahbar et al., 2013). Bees in general and the honey bee, *Apis mellifera*, are crucial to the ecosystem functioning and human food production due to the pollination service they provide. This service has been valued at >$200 billion annually worldwide (Gallai et al., 2009b; Le Conte & Navajas, 2008).

Table 1: The taxonomy of honey bees

Taxonomy	
Kingdom	*Animalia*
Phylum	*Arthropoda*
Class	*Insecta*
Order	*Hymenoptera*
Family	*Apidae*
Genus	*Apis*

The species that belong to *Apis* genus are divided in three clades (Figure 1), the cavity-nesting bees (*Apis mellifera*, *Apis cerana*, *Apis koschevnikovi*, *Apis nuluensis*, *Apis breviligula*), the giant bees (*Apis dorsata*, *Apis laboriosa*, *Apis binghami*, *Apis nigrocincta*, *Apis indica*) and dwarf bees (*Apis florea*, *Apis andreniformis* (Arias & Sheppard, 2005; Raffiudin & Crozier, 2007). *Apis mellifera* natural distribution encompasses Europe, Africa and Western Asia, whereas the other *Apis* species are found exclusively in Asia (Han et al., 2012).

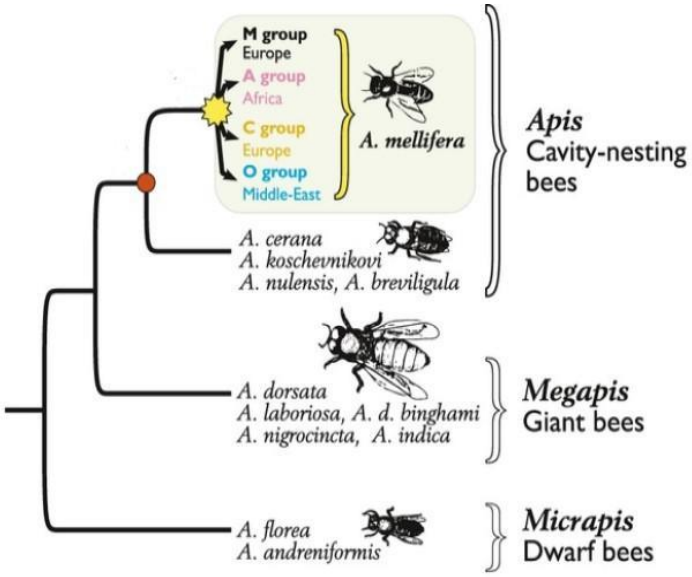

Figure 1: Phylogeny of the three clades of *Apis* (Han et al., 2012)

2. The honey bee biology

Honey bees are eusocial insects and the colonies are formed by three casts; the queen that is the single individual that lay eggs, up to 2,000 per day, the workers that carried out all the other tasks such as foraging, brood care, hive maintenance and defense and the drones that are responsible to mate with queens from other colonies (James et al., 2008).

Honey bees go through a complete metamorphosis and have four life stages: egg, larva, nymph and adult (Figure 2). The passage through the immature stages takes 16 days for queens, 21 days (on average) for workers, and 24 days for drones.

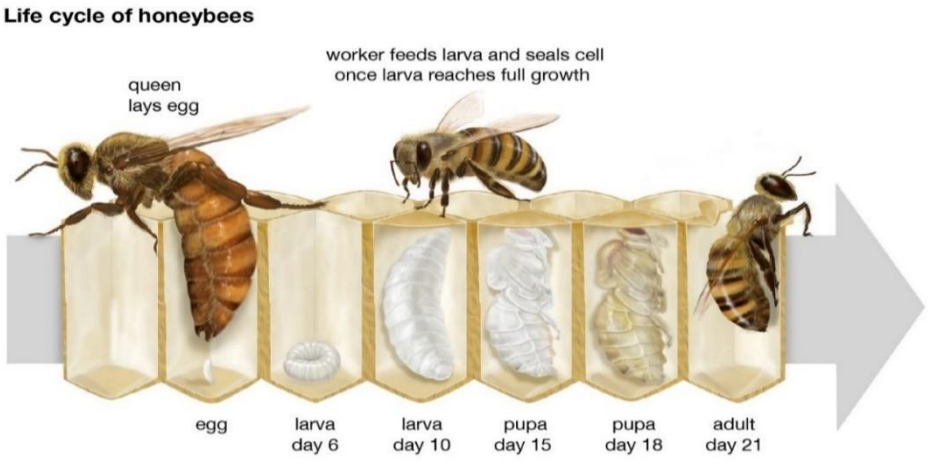

Figure 2: Life cycle of honey bees (https://honeyportal.keystone-foundation.org/morphology/)

Honey bees have an haplodiploid reproduction, meaning that males are haploid and females are diploids (Blackmon et al., 2015). Queens lay two kinds of eggs: fertilized and unfertilized. Unfertilized eggs develop into males (usually called drones). Fertilized eggs develop into females. This can be explained by the fact that sex in honey bees is determined by a single gene known as complementary sex determined (Beye et al., 2003). When only one copy of this gene is present or there is homozygosity, a male is produced; when two different alleles are present (heterozygosity) a female is produced.

Females can be sterile workers or a fertile queen, depending on their diet. The larvae that were laid in queen cells are fed with royal jelly, while those in worker cells are fed with worker jelly (Shi et al., 2011).

3. Mating system

The honey bee queens are polyandrous, meaning that they mate with several drones. A virgin queen can mate with 10 to 15 drones. This number, however, varies between species and subspecies (Adams et al., 1977; Hernández-García et al., 2009; Kerr et al., 1962). At the age of 15-23 days, drones and virgin queens' mate during a mating fight (Yániz et al., 2020). Drones fly from surrounding colonies to a congregation area to await the arrival of virgin queens (Baudry et al., 1998; Yániz et al., 2020). When the queen approaches the congregation area, drones pursue her, and several of them copulate with her before dying (Baudry et al., 1998; Yániz et al., 2020). During mating, 6-12 million spermatozoa are transferred from the drone's seminal vesicles into the queen's genital orifice. Approximately 10% of these spermatozoa are transferred to the queen's oviducts. Finally, only 3-5 % of the spermatozoa (2–7 million) are stored in the queen spermatheca and are used to produce eggs (Yániz et al., 2020).

Since the congregation area joins bees from the surroundings the number and the genetic background of the colonies influence the relatedness between the honey bees and the genetic diversity within the colony (Baudry et al., 1998).

4. *Apis mellifera*: The evolutionary history

The western honey bee, *A. mellifera*, is a generalist species foraging for nectar and pollen on an extensive range of flowering plants and was historically native to Africa, Europe, western Asia and the Middle East (Leclercq et al., 2018; Techer et al., 2017; Tihelka et al., 2020). This species has separated from its close relative, *A. cerana*, approximately 6-25 million years ago when it spread westward to colonize parts of Asia, Europe, and Africa (Ramírez et al., 2010; Sheppard & Meixner, 2003). As European settlers colonized various parts of the

globe, they transported and established several lineages of *A. mellifera*, along with elaborate beekeeping practices (Sheppard, 1989).

A. mellifera had to adapt to a wide and diverse range of habitats with different climates and his adaptation occurred by natural selection and resulted in about 33 subspecies (Chen et al., 2016; Michael S. Engel, 1999; Smith & Glenn, 1995; vanEngelsdorp & Meixner, 2010), showing a particular geographic distribution (Figure 3). Phylogeographical studies using a variety of genetic markers (Cornuet et al., 1991; Weinstock et al., 2006; Whitfield et al., 1988) classified this diversity into four major lineages: Africa (A), the Middle East (O), Southeastern Europe (C), and Western and Northern Europe (M). Other sublineages and lineages have been proposed such as lineage Y and sublineage Z. (Alburaki et al., 2011, 2013; Garnery et al., 1992; Kandemir & Kence, 1995; Smith, 2019; Smith et al., 1997). The distribution of each subspecies and the corresponding evolutionary lineage is shown in (Table 2).

Figure 3: The spatial distribution of 33 subspecies of honey bee *A. mellifera*

Table 2: Honey bee subspecies and *Apis mellifera* (Ilyasov et al., 2020)

Subspecies Name	Common name	Lineage	Distribution
Africa (11 subspecies)			
Apis mellifera lamarckii	The Egyptian honey bee	**O**	Egypt, Sudan
Apis mellifera litorea	The East African coastal honey bee	**A**	Kenya
Apis mellifera adansonii	The West African honey bee	**A**	Nigeria, Burkina Faso, Uganda, Tanzania, Zambia, Senegal, Sudan
Apis mellifera scutellata	The African honey bee	**A**	Kenya, Tanzania, Uganda, Republic of South Africa, Somalia
Apis mellifera monticola	The East African Mountain honey bee	**A**	Mountains of Kenya, Tanzania
Apis mellifera capensis	The Cape honey bee	**A**	Cape region in the Republic of South Africa
Apis mellifera unicolor	The Madagascar honey bee	**A**	Madagascar
Apis mellifera simensis	The Ethiopian honey bee	**A**	Ethiopia
Apis mellifera sahariensis	The Saharan honey bee	**A**	Morocco, Algeria, Tunisia, Libya, Mauritania Western Sahar
Apis mellifera intermissa (Synonym: *A. m. major*	The Tellian honey bee	**A**	Morocco, Libya, Tunisia
Apis mellifera jemenitica (Synonyms: *Apis mellifera yemenitica, Apis mellifera nubica, Apis mellifera sudanensis, Apis mellifera bandasii, Apis mellifera woyi-gambell*)	The Arabian honey bee	**Y**	Arabian Peninsula, Chad, Saudi Arabia, Somalia, Sudan, Uganda, Yemen
Western Asia and the Middle East (9 subspecies)			
Apis mellifera ruttneri	The Maltese honey bee	**A**	Malta
Apis mellifera syriaca	The Syrian honey bee	**Z**	Syria, Israel, Lebanon, Palestine, Jordan
Apis mellifera mellifera Linnaeus (Synonyms*: Apis mellifica germanica, Apis mellifica nigrita, Apis mellifica mellifica*	The European dark honey bee	**M**	France, United Kingdom, Switzerland, European, part of Russia, Poland, Denmark, Norway, Sweden, Ireland

Apis mellifera pomonella	The Tian Shan honey bee	**O**	Tian Shan mountains of Kazakhstan, Kyrgyzstan
Apis mellifera sinisxinyuan	The Xinyuan honey bee	**M**	Uygur Autonomous Region of China
Apis mellifera meda	The Persian honey bee	**Z**	Iran, Iraq, Syria, Turkey
Apis mellifera caucasia	The Caucasian honey bee	**C**	South Russia, Turkey, Georgia
Apis mellifera remipes Synonym *: Apis mellifera armeniaca*	The Armenian honey bee	**O**	South Russia, Armenia, Iran, Georgia
Apis mellifera anatoliaca	The Anatolian honey bee	**Z**	Iran, Armenia, Syria, Turkey
Europe (13 subspecies)			
Apis mellifera iberiensis New name for preoccupied: *Apis mellifera iberica,*	The Spanish honey bee	**M**	Spain, Portugal
Apis mellifera macedonica	The Macedonian honey bee	**C**	Bulgaria, Greece, Macedonia, Ukraine
Apis mellifera ligustica	The Italian honey bee	**C**	Italia
Apis mellifera carnica Synonyms : *Apis mellifica hymettea* *Apis mellifera carnica, Apis mellifera carniolica, Apis mellifica banatica, Apis mellifera banata*	The Carniolan honey bee	**C**	Slovenia, Bulgaria, Poland, Austria, Croatia, Bosnia and Herzegovina, Serbia, Hungary, Romania
Apis mellifera carpathica	The Carpathian honey bee	**C**	Ukraine, Bulgaria, Romania, Moldova
Apis mellifera rodopica	The Bulgarian honey bee	**C**	Bulgaria
Apis mellifera cecropia	The Greek honey bee	**C**	Greece
Apis mellifera siciliana Synonym : *Apis mellifera sicula*	The Sicilian honey bee	**C**	Sicily
Apis mellifera adami	The Cretan honey bee	**C**	Crete
Apis mellifera cypria	The Cyprus honey bee	**C**	Cyprus
Apis mellifera artemisia New name for preoccupied*: Apis millefera acervorum*	The Russian steppe honey bee	**C or O**	South Russia, Ukraine
Apis mellifera sossimai	The Ukrainian honey bee	**C or O**	Crimea, South Russia, Ukraine

New name for preoccupied: *Apis mellifera cerifera*			
Apis mellifera taurica	The Crimean honey bee	**C or O**	Crimea, South Russia, Ukraine

Ten of the 33 subspecies are endemic to Europe and belong to the evolutionary lineages M and C. European M lineage includes only two subspecies: *A. m. mellifera* or the Dark honey bee and *A. m. iberiensis* or the Iberian honey. *A. m. iberiensis* occupies the Iberian Peninsula, while *A. m. mellifera* ranges from France in the south to Scandinavia in the north, and from Ireland and the United Kingdom in the west to the Ural Mountains in the east (Ruttner, 1988a). Lineage C is restricted to the Apennine and Balkan peninsulas and includes the two most widely kept honey bee subspecies: The Italian *A. m. ligustica* and the and the Carniolan *A. m. carnica.*

A. m. iberiensis and *A. m. mellifera* are genetically not very divergent. Post-glacial recolonization from an Iberian refuge explains the genetic patterns of *A. m. mellifera*. The interpretation of *A. m. iberiensis* variation, on the other hand, has proven more difficult, with different studies yielding contradictory results. Morphology (Ruttner, 1988b) and allozymes (Smith & Glenn, 1995) revealed a gradual cline from Africa to northern Europe, with Iberian honey bees exhibiting intermediate phenotypes. This pattern is consistent with the primary intergradation hypothesis for M-lineage origin (Ruttner, 1988b; Smith & Glenn, 1995). Microsatellites revealed and homogenous Iberian population and no differentiation between *A. m. iberiensis* and *A. m. mellifera*, and a sharp disruption between M and A-lineages (Franck et al., 1998). A study using simultaneously mitochondrial and SNPs (Single Nucleotide Polymorphism) markers revealed two major clusters forming a well-defined cline that bisects *Iberia* along a northeastern–southwestern axis, supporting the hypothesis of secondary contact between divergent populations previously isolated in glacial refugia (Chávez-Galarza et al., 2015; Pinto et al., 2013).

In terms of the genetic integrity of the European M-lineages, while *A. m. iberiensis* exhibits no signals of introgression in its native range, *A. m. mellifera* has been severely compromised by introgressive hybridization to the point where it has been driven to extinction in some places (Chávez-Galarza et al., 2015; Pinto, Henriques, Chávez-Galarza, Kryger, Garnery, van der Zee, et al., 2014).

5. The economic value of honey bees

Honey bees play a crucial role in both ecological and economic systems, they are considered a major pollinator due to their efficiency and high accessibility. Pollination had a global economic value of about €153 billion in 2005 and honey bees (*Apis mellifera*) are the primary pollinators of many agricultural crops (Hung et al., 2018). They are vital pollinators for a wide range of crops, including fruits, vegetables, nuts, and oilseeds. Their

pollination services contribute to the successful reproduction and yield of these crops. It is estimated that honey bee pollination directly contributes to the production of agricultural commodities worth billions of dollars globally each year (Gallai et al., 2009a). In Agricultural Productivity, honey bee pollination increases crop yields and improves the quality of fruits and seeds. The economic value of honey bee pollination can be measured by the increased crop yield and the enhanced market value of the produce.

Hone bees also produce honey, which is not only consumed as a sweetener but also used in various food products, beverages, and medicinal applications. Beekeepers harvest honey from managed honey bee colonies, and the honey industry contributes to the agricultural economy (Morse & Calderone, 2000).

Apart from honey, they produce other valuable products such as beeswax, propolis (a resinous substance), royal jelly, and pollen are all hive products that have commercial value. These products are used in various industries, including cosmetics, pharmaceuticals, food processing, and crafts (Weis et al., 2022).

6. Honey bee threats

Honey bees are under threat from a variety of factors, the majority of which are caused by humans, they are extremely sensitive to pesticides, most likely because they have fewer genes encoding detoxification enzymes than other insects (Claudianos et al., 2006). This explains the massive losses caused by organochlorine, carbamate, organophosphorus, and pyrethroid pesticide exposure between 1966 and 1979 (Atkins, 1992). Fortunately, there were efforts to restrict pesticides applications. However, the residual activity of some pesticides was never addressed effectively (Johansen & Mayer, 1990; Johnson et al., 2010).

Climate change has a variety of effects on honey bees (Le Conte & Navajas, 2008) for example, rising CO2 levels in the atmosphere reduce pollen quality, which is critical for honey bee larval development (Brodschneider & Crailsheim, 2010). Climate change may also cause a shift in plant distribution, affecting the availability of resources important to bees (Le Conte & Navajas, 2008). Beekeepers attempt to use honey bees that were either more productive or easier to manage, for instance, selecting and using bees with low defensive behavior and high honey production. To achieve these goals, different subspecies are crossed, promoting gene flow and admixture or introgression between populations with different genetic backgrounds (Hernández-García et al., 2009; Pinto, Henriques, Chávez-Galarza, Kryger, Garnery, van der Zee, et al., 2014; vanEngelsdorp & Meixner, 2010). The natural admixture may be an important evolutionary force in speciation and genetic

diversity maintenance (Dowling & Secor, 1997; Nolte & Tautz, 2010). Human-induced admixture, on the other hand, may contribute to the irreversible loss of genotype combinations across the entire genome (Allendorf & Luikart, 2009). This is especially worrisome for honey bees due to their pool mating system *A. m. ligustica*, *A. m. carnica*, and *A. m. caucasica*, all from the C-lineage, have been widely used due to their docility and high productivity. Buckfast is a popular artificial strain that has been selected for its superior honey production and low defensive behavior, and it is also mostly of C-derived ancestry (Minozzi et al., 2021). The large-scale honey bee movement has altered the natural distribution of the honey bee to the point where the formerly widely distributed European subspecies, *A. m. mellifera*, is threatened with extinction due to introgression and replacement by C-lineage strains (De La Rúa et al., 2009; Jensen et al., 2005; Pinto, Henriques, Chávez-Galarza, Kryger, Garnery, van der Zee, et al., 2014; vanEngelsdorp & Meixner, 2010). A similar panorama is found in the *A. m. iberiensis* populations of the Canary Islands and the Azores (Ferreira et al., 2020; Muñoz et al., 2013, 2014).

The movement of individuals is accompanied by the movement of parasites and pathogens. *A. mellifera* has been particularly harmed by the mite *Varroa destructor* and the microsporidian *Nosema ceranae* (Antúnez et al., 2009; Dussaubat et al., 2012; Martín-Hernández et al., 2018; Traynor et al., 2020). Both are native to Asia and have spread rapidly worldwide following a host shift from *Apis cerana* to *Apis mellifera* (Martín-Hernández et al., 2018; Traynor et al., 2020) *V. destructor* suppresses bee immunity and serves as a reservoir, incubator, and transmission route for several viruses, including one of the most important honey bee pathogens, the Deformed Wing Virus (Yang & Cox-Foster, 2005). *N. ceranae* is an intracellular parasite that reduces colony longevity by inducing oxidative stress and altering metabolism and immune response (Antúnez et al., 2009; Botías et al., 2013; Dussaubat et al., 2012).

7. Tools and importance of preserving genetic diversity

To evolve in response to environmental change, populations and species need genetic diversity. In honey bees, large-scale movement of colonies and queens, artificial selection, and the widespread use of commercial stocks reduce adequate population size and genetic diversity (Estoup et al., 1995). Reciprocal translocation experiments highlight the importance of maintaining subspecies adapted to local conditions to promote the long-term survival of honey bees and have shown that local bees have higher survival rates and lower pathogen loads than introduced subspecies (Büchler et al., 2014).

The recognition that native genetic diversity is essential for bees facing multiple threats has led to a number of conservation and breeding programs and initiatives. Examples of them are some conservation programs created to protect the endangered *A. m. mellifera* in Denmark (Læsø), Scotland (the islands of Colonsay and Oronsay), the United Kingdom, France, Netherlands, Norway, Switzerland, Ireland, and Belgium (see the website "http://www.sicamm.org"). Other conservation and breeding programs were created to protect *A. m. carnica*, for instance in Slovenia, Austria, and Germany (De La Rúa et al., 2009). It may seem strange to create conservation centres for a subspecies that is widely used, but only a small portion of its diversity has been exploited.

The success of these initiatives relies on the capacity to accurately detect genetic pollution in a time- and cost-effective manner. In several programs, this has been done through wing morphometry. However, based on data from Africanized honey bees (Sheppard & Smith, 2000), wing morphometry is likely unable to detect low levels of genetic pollution. Another widely used marker is the mitochondrial (mtDNA) intergenic $tRNA^{leu}$-cox2 region. The $tRNA^{leu}$-cox2 is highly informative because it combines size and sequence variation. However, the mtDNA has a maternal inheritance, and thus important information may be missed (Bertrand et al., 2015). The microsatellites overcome the limitations of morphometry and mtDNA markers (Jensen et al., 2005; Soland-Reckeweg et al., 2009) because it is a neutral biparentally inherited marker. However, SNPs (Single Nucleotide Polymorphism) have several advantages over microsatellites. They are more abundant and widespread in the genome (Weinstock et al., 2006), have lower genotyping error and higher data quality, and are easily transferable between laboratories (Vignal et al., 2002). Furthermore, some studies have found that a small number of high-graded SNPs (Muñoz et al., 2015) outperform the multiallelic marker in estimating introgression (Muñoz et al., 2017; Parejo et al., 2018). These characteristics make SNPs an effective tool for testing breeding stocks. SNP data can be easily incorporated into shared genetic databases, facilitating the implementation of a conservation strategy at the European scale.

8. Single nucleotide polymorphism (SNP)

SNPs refer to single-base changes in a DNA sequence with an allele frequency of at least 1% (Vignal et al., 2002). At each position, there are four possible nucleotides, but due to the low mutation rate (approximately 10^{-8} to 10^{-9} changes per nucleotide per generation), SNPs are typically characterized by two alleles bi-allelic. This bi-allelic nature can also be attubuted to the clear bias towards the transition substitutions (purine to purine

or a pyrimidine-to-pyrimidine substitution) (Vignal et al., 2002). SNPs are the most recent genetic marker used in honey bee research and they a number of advantages over other molecular markers, they are biparentally inherited, evolve according to the infinite allele model which is more efficient than the stepwise mutation model (associated with STR) in the study of the origin of individuals. At a technical level, SNPs display low genotyping errors, provide high-quality data, are more amenable to automation, data interpretation can be standardized, therefore allowing for experiments to be easily replicated between laboratories (Vignal et al., 2002).

There are several methods for SNP genotyping. The choice of the genotyping technique is determined by both number of SNPs and the sample size (Sobrino et al., 2005). When a study requires a few SNP markers to be genotyped in a very large sample size, methods that share the costs of the SNPs by many samples should be chosen. TaqMan nuclease assay, the mass spectrometry assay MassARRAY MALDI-TOF platform of Agena BioScience™ and NEBNext Direct Genotyping Solution are Technologies that fit well in this scenario.(Rybicka et al., 2021)

In honey bees there are several reduced SNP panels (Chapman et al., 2015; Henriques et al., 2021; Henriques, Browne, et al., 2018; Henriques, Wallberg, et al., 2018) designed to be genotyped by the MassARRAY MALDI-TOF platform. This technology uses PCR to amplify the regions of the genome containing each targer SNP (Table 3).

During my thesis, a new tool based on NEBNext Direct Genotyping Solution, a hybridization-based target enrichment approach that can target 100 to 5000 markers from up to 96 samples within a single hybridization was developed (Textor et al.,2018).

Table 3: The strengths and weaknesses of SNP

Strengths	***Weaknesses***
More abundant and widespread in the genome (Reich et al., 2003).	Costly. https://www.police-scientifique.com/adn/evolutions
Provide genome-wide coverage and higher quality data (Vignal et al., 2002)**.**	Sometimes require advanced technologies to be implemented (Gaudet et al., 2009).
Easily transferable between laboratories (Vignal et al., 2002)**.**	Popular platforms for SNP typing use real-time PCR instruments and are usually not available in laboratories with limited resources, because

	of higher initial costs and the need for ongoing, highly technical instrument maintenance (Birdsell et al., 2012).
A small number of high-grade SNPs is more efficient than the multi-allelic marker to estimate introgression (Muñoz et al., 2017).	Sometimes require advanced technologies to be implemented (Gaudet et al., 2009).
More amenable to automation and facilitate the standardization of genetic material identification on a global scale. (Henriques, Browne, et al., 2018).	
Suitable for high-throughput automated technologies that allow genotyping of hundreds or thousands of loci in many different individuals (Vignal et al., 2002).	

9. From whole-genome data to reduced SNP assays

With next-generation sequencing (NGS) platforms, thousands to millions of SNPs can be identified in model and non-model organisms (Baird et al., 2008). For the honey bee, the development of a population-scale whole-genome-sequencing (WGS) dataset is facilitated due to several reasons. (i) A reference genome has been available since 2006 (Weinstock et al., 2006) and has been frequently updated (Elsik et al., 2014), making mapping or genome assemblage much easier. (ii) Because the honey bee genome is only 236 Mbp long is possible to have quality SNPs (with sufficient sequence coverage) at low cost. (iii) The honey bee haplodiploid system allows the sequencing of haploid males, which provides accurate calls of SNPs and the elimination of uncertainties in reconstructing haplotypes.

NGS have shift the scale of analysis from few genomic regions and loci to whole-genomes. The millions of SNPs identified by WGS have been important to gain insights into fundamental apicultural questions e.g. tolerance/resistance to *V. destructor* and introgression (Harpur et al., 2019, 2020; Parejo et al., 2016; Saelao et al., 2020) and into the processes (neutral and adaptative) shaping diversity patterns (Chen et al., 2016; Cridland et al., 2017; Fuller et al., 2015; Harpur et al., 2014; Henriques, Wallberg, et al., 2018; Nelson et al., 2017; Parejo et al., 2017; Wallberg et al., 2014). However, to develop

and analyse WGS data sophisticated laboratorial and computational tools and advanced bioinformatics skills are required. These resources are frequently unavailable in conservation and breeding centers, meaning that WGS data by itself adds little value to routine activities. Fortunately, WG can be used to develop reduced SNP-based tools that can be employed easily by the bee community.

Panels containing a reduced number (< 200 SNPs) of highly informative markers have been designed to address a variety of objectives, including (i) identifying Africanized honey bees (Chapman et al., 2015), (ii) estimating C-lineage introgression into the M-lineage subspecies (Henriques, Browne, et al., 2018; Muñoz et al., 2017; Parejo et al., 2018) and (iii) monitoring diversity in immune genes (Henriques et al., 2021). All these SNP panels have been designed for genotyping using the MassARRAY MALDI-TOF platform. Panels with thousands of SNPs were created using other technologies, such as Affymetrix or Illumina Infinium. These panels have different objectives such as performing genome-wide association screening for hygienic behaviour (~44,000 SNPs) (Spötter et al., 2016). Identification of honey bee subspecies (~4,000 SNPs) (Momeni et al., 2021) or genomic selection (~100,000) (Jones et al., 2020).

III. MATERIAL AND METHODS

1. Whole genome re-sequencing datasets and sample selection

The aim of the molecular tool developed here was to accurately detect genetic introgression between M- and C-lineages subspecies across their wide distribution range, as well as, to ensure that can be used in the Hohen Neuendorf Bee Institute's mating station. SNPs were identified from WGS scans of two datasets obtained from previous studies (Henriques, Browne, et al., 2018; Jones et al., 2020; Parejo et al., 2016) and sequenced on an Illumina HiSeq 2500 with an aimed sequencing depth of 5X per individual with sequencing libraries generated using the Illumina TruSeq Sample Preparation Kit (Henriques, Parejo, et al., 2018; Jones et al., 2020; Parejo et al., 2016).

The dataset1 contains 299 whole-genome sequences (WGS) from drones sampled across a wide geographical range, 228 belong to the M-lineage (117 *A. m. iberiensis* and 111 *A. m. mellifera*), and 71 belong to C-lineage (37 *A. m. carnica* and 34 *A. m. ligustica*). The mapping and variant calling of this dataset is described in (Henriques et al., 2018). The second sequenced dataset (dataset2) represents the genetic diversity of the mating station of Hohen Neuendorf Bee Institute and contains 56 drones, 44 classified as *A. m. carnica* and 12 as *A. m. mellifera*. The mapping and variant calling of this dataset is described in (Jones et al., 2020).

2. Estimation of introgression

Introgression proportions (Q values) were estimated by ADMIXTURE v1.3.0 (Alexander et al., 2009) for two ancestral groups (K=2). The program was configured for 10,000 iterations in 20 independent runs for each K to confirm consistency between runs. Convergence between iterations was checked by comparing log-likelihood scores (LLS) using the default termination criterion set to stop when LLS increases by < 0.0001 between iterations CLUMPAK (Kopelman et al., 2015) was used to summarize and visualize the Q -plots.

3. SNP assay design

The SNP assay design was based on 204 pure drone WGS, of which 140 belonged to lineage M (117 *A. m. iberiensis* and 34 *A. m. mellifera*) and 64 to lineage C (30 *A. m. carnica* and 34 *A. m. ligustica*) of the dataset1. Dataset2 contained 24 pure drones, 16 of

which were classified as *A. m. carnica* and 8 as *A. m. mellifera* (Figure 4, Table S1). The most informative SNPs were identified based on the high F_{ST} values between the lineages (FST≥0.99 for dataset1 and F_{ST}≥0.80 for dataset2) calculated in PLINK 1.9 (Chang et al., 2015; Purcell et al., 2007) between the M and C lineage subspecies.

The putative functional role of the highly differentiated SNPs was detected by SNPeff 4.3 (Cingolani et al., 2012) and the NCBI honeybee version 102 annotation. To test which subset of SNPs contained redundant information, PLINK 1.9 was used to identify haplotype blocks with the block function "--blocks no-pheno-req no-small-max-span" with the parameter "—block with the parameter "--blocks-max-kb 5000" and to calculate the genomic blocks.

To develop the SNP assay we used NEBNext Direct Genotyping Solution (the protocol followed can be found at (https://international.neb.com/protocols/2019/07/30/protocol-for-use-with-nebnext-direct-genotyping-solution-neb-e9500-e9530) and we aimed for a total of 130 SNPs wich is the maximum number allowed by this technology at a low cost. To downsize the number of SNPs we preferentially retained the SNPs that followed one or several of the following criteria (i) located in different haplotype blocks in the genome and distributed along the 16 chromosomes (ii) functionally relevant (located in UTRs, exons and splice regions) (iii) a single match of the flanking region on the reference genome. The 301-bp flanking regions (150 bp on either side) of the130 retained SNPs were sent to New England Biolabs Inc and served as input for the bait design.

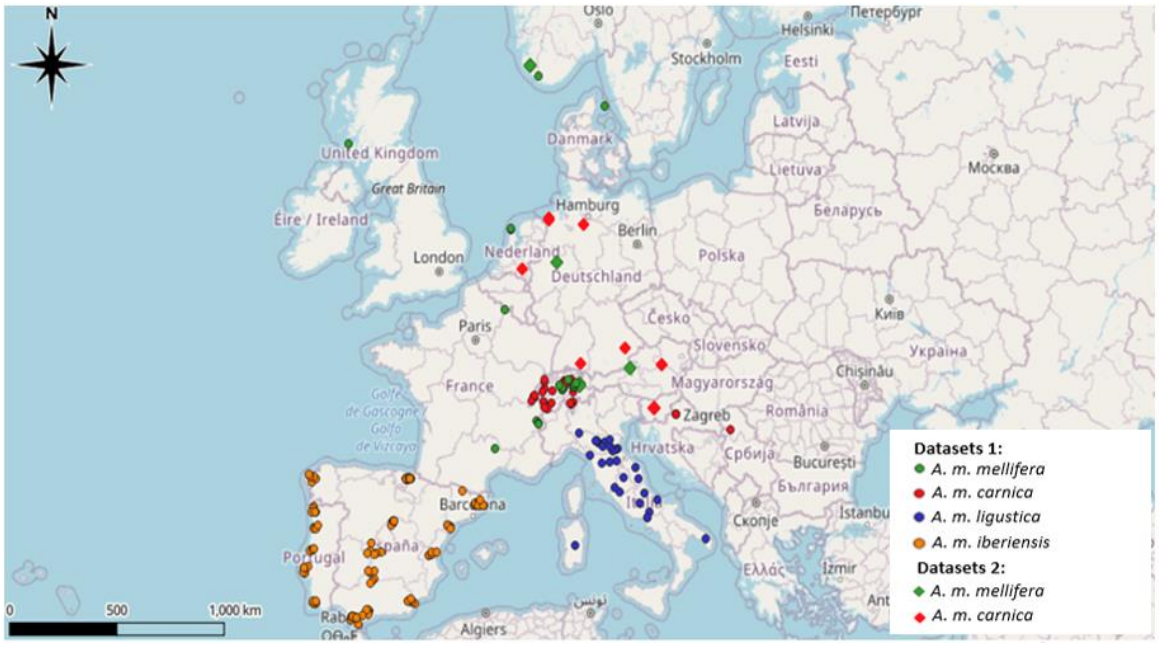

Figure 4: Map showing the spatial distribution of collected samples (datasets 1, datasets 2) (Map generated by QGIS 3.12.2- https://qgis .org/es/site/forusers/download.html)

4. Collection of the spermatheca content

The extraction of the spermatheca was peformed by researchers at the Hohen Neuendorf Bee Institute and send it to CIMO. Spermathecal fluid was collected by piercing and gently squeezing each spermatheca into a 20 µL droplet of K+ solution. The droplet, containing the fluid, was then transferred to a tube with 280 µL of K+ solution, briefly vortexed, and stored at -70°C. Later, frozen droplets were thawed and centrifuged at 5,000×g for 5 minutes. The supernatant was removed, and the resulting semen pellet was preserved in 400 µL of 98% ethanol.

5. Samples and DNA extraction

Genomic DNA was extracted from 168 samples. Different sources have been used such as legs (9), flight muscle (19) spermathecal content (117), semen (6), single individuals (75), and pools of individuals (9). The DNA from spermathecal content and semen were extracted using the tissue protocol of QIAmp DNA MicroKit (Qiagen®) following (Yadró et al., 2023) recommendations. The recommendations include to do three hours of lysis incubation, the addition of RNA-carrier and multiple re-elutions.

The DNA from the other samples were extracted using tissue protocol of Monarch Genomic DNA Purification kit (New England Biolabs® Inc. The minimum DNA

concentration needed to continue with the targeted genotyping is 1.7 ng/µL. All the samples were quantified for double-strand DNA (dsDNA) using Quantus ® Fluoremeter. The samples that did not have enough DNA concentration were evaporated. The DNA from six individuals were used to do DNA pools.

6. Targeted Genotyping

Using the NEBNext Direct ® Genotyping Solution sample preparation kit with minor modifications (New England BioLabs ® Inc.) was followed to genotype 130 SNPs in (Table S2). Three libraries were created, consisting of 24, 96, and 96 samples, respectively. In the initial step of this protocol, 10 ng of DNA was enzymatically fragmented for the first 216 samples and labeled at the 5' index with an Illumina®-compatible P5 adapter. This adapter incorporates two additional sequence essentials: (1) an inline 12 bp unique molecular identifier (UMI) that allows for error correction and improved sequencing accuracy and to mark each unique DNA fragment in each sample, followed by (2) an 8 bp inline sample index to barcode each sample prior to pooling. For barcoding, up to 89 samples are pooled into a single tube for processing by the NEBNext Direct Genotyping Solution target enrichment kit that hybridizes biotinylated baits to targets for 90 minutes. The pooled DNA was denatured and then hybridized to biotinylated oligonucleotide baits specifically designed for honey bees by New England Biolabs® and these baits define the 3 ends of each target sequence, then the target fragments will be bound to streptavidin beads and adjusted to remove the off-target flanking sequence enzymatically. In a second step, the 3' sequencing adapter is ligated and the captured fragments will be PCR amplified to generate automated libraries in a comprehensive manner. Due to the low DNA input and complexity in this specific application, the number of PCR cycles, was adjusted by adding 6 cycles to the recommended PCR cycles. The multiplexed libraries were sequenced on Illumina's MiSeq, which was done at CIBIO (University of Porto).

7. Miseq Sequencing

The libraries were evaluated using the TapeStation 2200 with the HS D1000 kit (Agilent Technologies) and also quantified by a SYBR green quantitative PCR assay using the KAPA Library Quantification kit (Kapa Biosystems). They were combined equimolarly into one single sequencing library. The sequencing library was diluted to 16 pM and the Illumina PhiX bacteriophage genome was spiked in at a concentration of 2%. The prepared

final library was sequenced on the Illumina MiSeq using the 2 × 150 cycles v3 chemistry, according to the manufacturer's instructions

8. MiSeq Sequencing and Data Preprocessing

After sequencing, the reads were demultiplexed with Illumina Bcl2fastq2 (v2.20). Trimmomatic-0.39 was used to cut the adapters and other Illumina-specific sequences from the reads. Reads less than 30 bp in length were also removed. Reads were aligned to the honeybee Amel_4.5 reference genome using BWA-MEM2 (Vasimuddin et al., 2019). The generated bam files were annotated with UMIs (Unique Molecular Indices), and mating information was set using the AnnotateBamWithUmis and SetMateInformation tools from fgbio-2.0.0 (http://fulcrumgenomics.github.io/fgbio/). Reads were then grouped by UMI and a consensus was, generated using the GroupReadsByUmi (strategy = adjacency) and Call MolecularConsensus Reads (min-input-base-quality=30 = 30, min-reads = 1) tools. After these steps, the reads were unmapped again. To map the reads again against the reference genome, the BAM reads have to be converted to FASTQ using SAMtools 1.15(Danecek et al., 2021) (tool fastq) and map consensus reads to the reference genome using BWA-MEM2. The UMI information of the unmapped consensus was transferred to the new mapped consensus BAM using Picard (http://broadinstitute.github.io/picard/) MergeBamAlignment. Finally, the consensus reads generated by CallMolecularConsensusReads were filtered using FilterConsensusReads (--min-reads=2 and --min-base-quality=10).

The baseline for the majority of the analyses was the synchronized (.sync) file from Popoolation2 v1.201(Kofler et al., 2011) that contains the allele frequencies after base quality filtering. To create this file, Popoolation2's mpileup2sync.jar script requires a mpileup file which was created by using SAMtools mpileup. A vcf (variant call format) file was also created using BCFtools 1.15 (Danecek et al., 2021), (Figure 5)

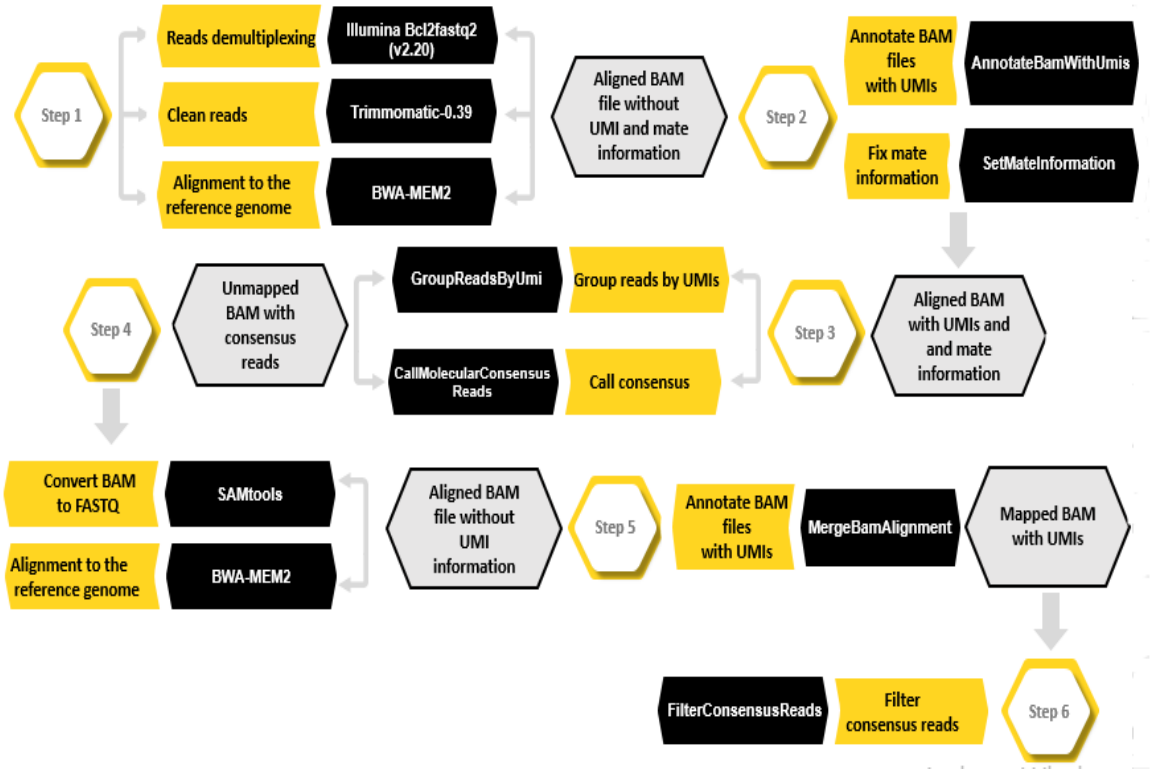

Figure 5: Data pre-processing flow chart

9. Quality control

A total of 216 samples underwent genotyping for the 130 SNPs. Genotypes were then subjected to quality control filters to remove SNP loci and samples with insufficient or inconsistent amplification. Specifically, samples with fewer than 20 reads in over 50% of loci and loci with fewer than 20 reads in over 25% of samples were excluded from the dataset. Loci with inconsistent SNP calls in at least two samples were excluded from further analysis.

i. Assessing the accuracy of the SNP tool to calculate allele frequency

To evaluate the sensitivity and accuracy of the SNP tool in determining allelic frequencies from pooled samples, DNAs from two haploid honey bees (one *A. m. carnica* and one *A. m. mellifera*) were pooled at different ratios (0:10, 1:9, 2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 9:2, 9:1, 10:0) at a concentration of 2 ng/µL.

In one experiment, pure samples of *A. m. mellifera* and *A. m. carnica* were sequenced using Illumina's HiSeq 2500 platform, in another experiment, the samples were obtained from the mating station of the Hohen Neuendorf Bee Institute and identified as pure using the geometric morphometric wing approach.

For each experiment, two different sets of *A. m. mellifera* and *A. m. carnica* samples were used, and two replicates were performed for each set in the WG sequencing. The frequency of each locus generated from the pooled DNAs was compared with the expected frequency. Loci with an error rate of ≥25% in more than 50% of the pools were excluded from the analysis.

After removing the problematic SNPs, the obtained M-allele frequencies of each of the six pools were compared with the expected frequencies and the absolute error calculated.

ii. Application of the SNP tool in spermatheca content

A total of 18 queens were artificially inseminated with five different semen mixtures, each containing known proportions of *A. m. carnica* and *A. m. mellifera* sperm (0%:100%, 20%:80%, 50%:50%, 80%:20%, 100%:0%). Each semen mixture was used for inseminating a group of three to four queens. The DNA was extracted from the contents of the spermathecae as well as the remaining semen mixtures, and genotyping was performed. The obtained M allele frequencies from the spermathecae contents were compared with the frequencies observed in the semen mixtures.

A total of 15 reciprocal crosses were conducted to examine whether the ancestry of the queen influences the tool accuracy. Three crosses involved *A. m. carnica* queens inseminated with *A. m. mellifera* semen, while three other crosses involved both the queen and the drone being of *A. m. mellifera* origin. Four crosses were performed using *A. m. carnica* queens and drones, and five crosses were conducted using *A. m. mellifera* queens and *A. m. carnica* drones. Genotyping was carried out for the queen's DNA, the semen used for insemination, and the spermatheca content, and the frequencies of the M allele were compared among these samples.

IV. RESULTS AND DISCUSSION

1. Genetic Structure

The development of SNP assay was based on pure drones, and to confirm the subspecies' ancestry and purity, we used ADMIXTURE. In dataset 1, 95 individuals were excluded due to introgression, with the majority being *A. m. mellifera*. Consequently, only 23 *A. m. mellifera* individuals out of the initial 111 were retained (Figure 6a). Additionally, seven C-lineage individuals were eliminated as they exhibited more than a 10% *A. m. mellifera* component (Figure 6a). All *A. m. iberiensis* individuals were pure. These findings align with previous studies, that show that *A. m. mellifera*, is threatened with extinction due to introgression and replacement by C-lineage strains (De la Rúa et al., 2009; Jensen et al., 2005; Pinto, Henriques, Chávez-Galarza, Kryger, Garnery, Van Der Zee, et al., 2014; vanEngelsdorp & Meixner, 2010). In contrast, *A. m. iberiensis* remains well preserved in its native range (Chávez-Galarza et al., 2017; Chávez-Galarza et al., 2015; Pinto et al., 2013). Therefore, it is very important to avoid the introduction of honey bees from other lineages in order to maintain its genetic integrity, which has proven to be highly diverse and complex, enabling it to live in a wide range and variety of environments (Chávez-Galarza et al., 2015, 2017; Chávez-Galarza et al., 2013; Henriques, Wallberg, et al., 2018; Pinto et al., 2013).

In the dataset2 24 out of 56 drones (Figure 6b) were not pure and therefore removed. Here, the majority of the removed individuals were from *A. m. carnica*, however four *A. m. mellifera* were also deleted.

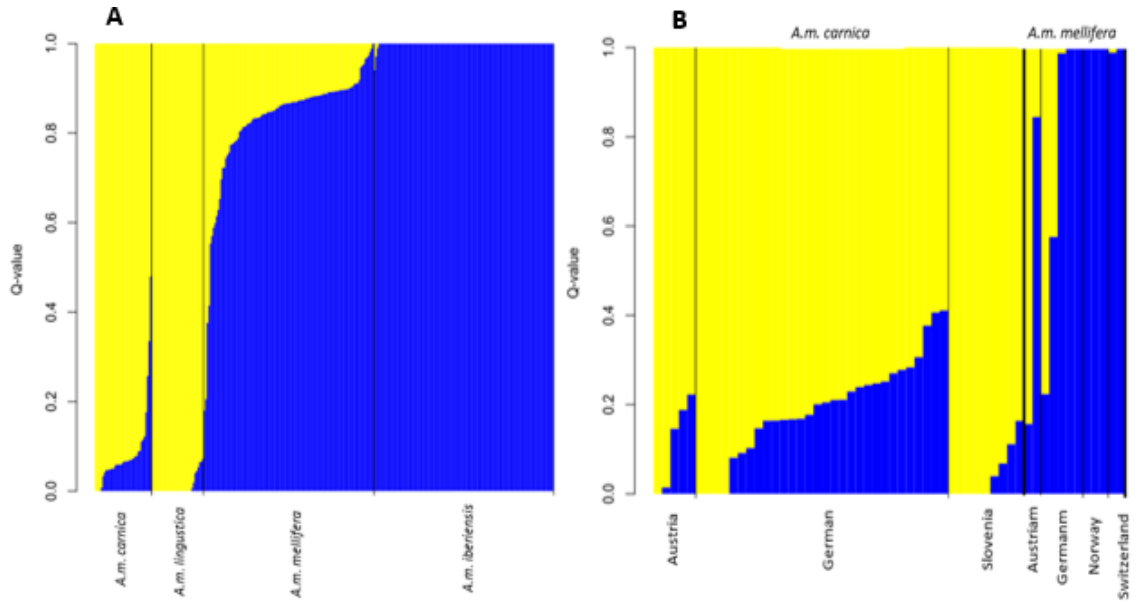

Figure 6: ADMIXTURE plot showing the genome partitioning into two clusters (K = 2) for each individual, represented by a vertical bar in A) dataset1 and B) dataset2. Blue represents the M-lineage cluster and yellow the C-lineage cluster. The black lines separate.

2. Selection and genomic information of highly informative SNPs

The development of SNP assay was based on pure drones. In dataset1, a total of 204 WGS were analyzed, with 140 belonging to the M-lineage (117 *A. m. iberiensis* and 23 *A. m. mellifera*) and 64 to C-lineage (30 *A. m. carnica* and 34 *A. m. ligustica*). The dataset2 consisted of 24 pure drones, 16 classified as *A. m. carnica* and eight as *A. m. mellifera.*

A dataset that accurately represents the original diversity of populations is essential for the development of a reliable molecular tool (Henriques, Parejo, et al., 2018). In this context, we have an example of pure individuals from various countries, expected to encompass a significant portion of the M and C-lineage diversity.

A total of 2,334,554 SNPs were detected with MAF > 0.05 in the pure individuals of dataset1 and 576,206 SNPs in the pure individuals of dataset2. In dataset1 4,930 SNPs were highly differentiated (F_{ST}≥0.99), 3,728 SNPs with F_{ST}=1 and 1,202 F_{ST}=0.99. In dataset2 the highest F_{ST} was 0.89 (4 SNPs) and we found we found 488 SNPs with a F_{ST}≥0.80. Of the 488 highly differentiated SNPs (F_{ST}≥0.80) found in the dataset2, 113 were also among the 2,334,554 SNPs of the dataset1 and only 77 had F_{ST}≥0.80 in the dataset1.

The high number of differentiated SNPs was expected because the C and M lineages are highly divergent (Wallberg et al., 2014) and a similar number of differentiated SNPs were reported by others (Henriques, Parejo, et al., 2018; Parejo et al., 2016).

In dataset1, chromosome 11 contained the highest proportion of informative SNPs (67.7%, 3338 SNPs), whereas chromosome 7 contained the least (0.2%, 8 SNPs). Regarding the dataset 2, chromosome 1 contained the highest proportion of informative SNPs (13.7%, 67 SNPs), and chromosome 16, the lowest proportion (2.9%, 14 SNPs) (Annexes, Table S3). For the next steps of the test design, we retained 5,007 fixed SNPs (77 highly differentiated SNPs in both datasets + 4,930 SNPs from dataset 1).

Most of the highly differentiated or fixed SNPs are located in introns (2,821 SNPs) and intergenic regions (1,782 SNPs); however, a number are located in regions of putative functional relevance, such as, non synonymous_variant or missense variants (18 SNPs) and splice acceptor and donor variants (3 SNPs).

NEBNext direct genotyping solution allows for a maximum of 130 SNPs at a reasonable expense, therefore, the number of SNPs was reduced by selecting preferentially putative functional SNPs from various genomic blocks distributed across the 16 chromosomes of honey bees.

The final 130 SNPs of the reduced SNP assay are distributed across 106 haplotype blocks, spanning all 16 chromosomes (Figure 7). Chromosome 1 has the highest number of SNPs, with 16, while chromosomes 7 and 16 have the lowest, each containing 4 SNPs (Figure 7; Annexes, Table S3 and S4). Out of the initially identified 18 informative SNPs, a total of 10 were retained in the assay (Annexes, Table S4). All the selected SNPs have an $F_{ST} \geq 0.88$ being 106 fixed ($F_{ST}=1$; Annexes, Table S4).

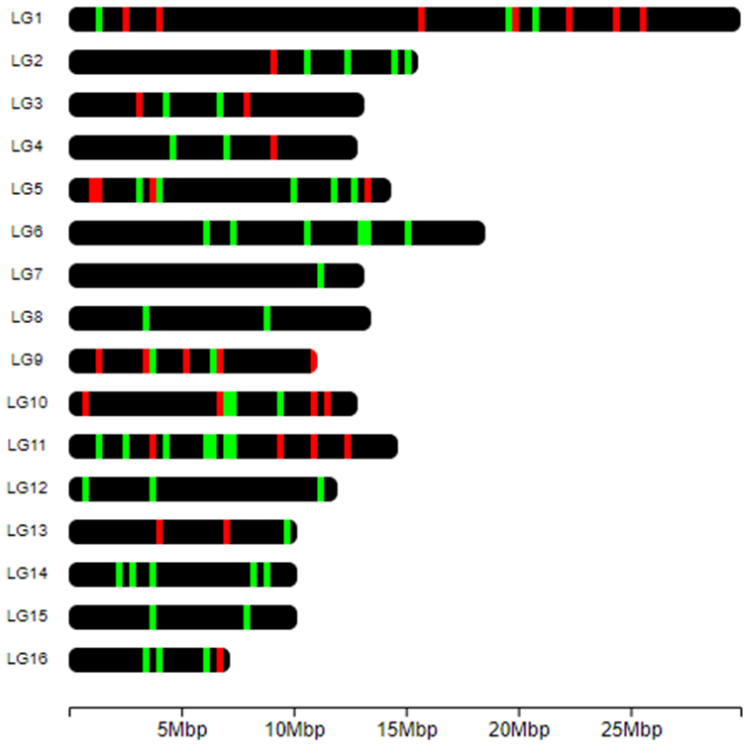

Figure 7: Distribution of the selected SNPs. Bars may represent multiple closely spaced SNPs. Deleted SNPs are highlighted in red. Image generated using chromoMap v0.4.1.

3. Quality control

On average we have 166.6 reads per sample and SNP. Out of the initial 216 samples, five were excluded from the dataset due to insufficient coverage; they had ≤20 reads in ≥50% of the loci. One sample (WGS2.4) exhibited no reads across all loci. After applying the coverage threshold of ≤20 reads in ≥25% of samples, 14 SNPs were removed from the dataset. A total of three SNPs (chr9:15736[Amel4,5] A, G; chr11:2983919[Amel4,5]C, T; chr11:7543229[Amel4,5]C, T) did not work in any individual. The comparison between Miseq and WG-sequenced samples revealed consistent calls among the loci without missing data.

4. Accuracy of SNP tool to determine allele frequency

A total of six DNA pools were generated by combining DNA from two pure haploid drones, one from *A. m. carnica* and one from *A. m. mellifera*, at varying ratios. (Table S5). Among the 130 SNPs initially analyzed, 47 exhibited an absolute error of ≥25% in ≥50% of the pools, leading to their exclusion from the dataset. Within these 47 SNPs, 13 had ≤20

reads in ≥25% of the samples. Subsequently, after removing SNPs with high missing data and those displaying an absolute error of ≥25% in ≥50% of the pools, 82 SNPs were retained for further analysis (Figure 8). Despite the reduction in the number of SNPs, several studies have demonstrated that approximately 50 SNPs suffice to differentiate M- from C-lineage (Henriques, Browne, et al., 2018; Henriques, Parejo, et al., 2018; Muñoz et al., 2015; Parejo et al., 2016) particularly when the majority of these SNPs are fixed (Pardo-Seco et al., 2014).

The novelty of this tool does not lie in its ability to detect genetic pollution between C and M lineages, as other tools have been developed for that purpose (Henriques, Browne, et al., 2018; Henriques, Parejo, et al., 2018; Muñoz et al., 2015; Parejo et al., 2016). Instead, it aims also to quantify the proportion of M and C lineage alleles. This is achieved by calculating the number of reads carrying the M and the C allele, allowing us to obtain the frequency of M and C lineage alleles. For instance, if the M allele is present at 10% of the reads, it means that the C allele represents 90% of the genetic composition. For simplicity, this discussion will focus solely on the M lineage proportion.

To assess the accuracy of the SNP tool in determining allele frequencies from pooled samples, we combined DNA from two haploid drones, one from *A. m. carnica* and one from *A. m. mellifera*, at various ratios (see Table S5). Two separate experiments were conducted. The first utilized pure *A. m. mellifera* and *A. m. carnica* samples, which had been previously subjected to whole-genome sequencing by Henriques et al. (2918) (Figure 8.1). The second experiment used samples obtained from the Hohen Neuendorf Bee Institute's mating station, where bees were identified as pure through wing geometric morphometric analysis (Figure 8.2). The M allele frequency per locus, generated from the pooled DNAs, was then compared to the expected frequency (Figure 8). In both experiments, the M allele frequency closely matched the expected values (Figure 8). However, it is worth noting that the representation of the M allele consistently exhibited a slight overestimation in both experiments and across nearly all mixtures. The most accurate estimation of allele frequencies for the whole-genome sequenced individuals was achieved when using DNA exclusively from *A. m. mellifera* (error=0.06) and *A. m. carnica* (error=0.27) (Figure 8.1). For the samples from the Hohen Neuendorf Bee Institute's mating station, the most accurate results were obtained when the DNA pools were composed of a 3:7 ratio of *A. m. mellifera* to *A. m. carnica* (error=0.5; Figure 8.2).

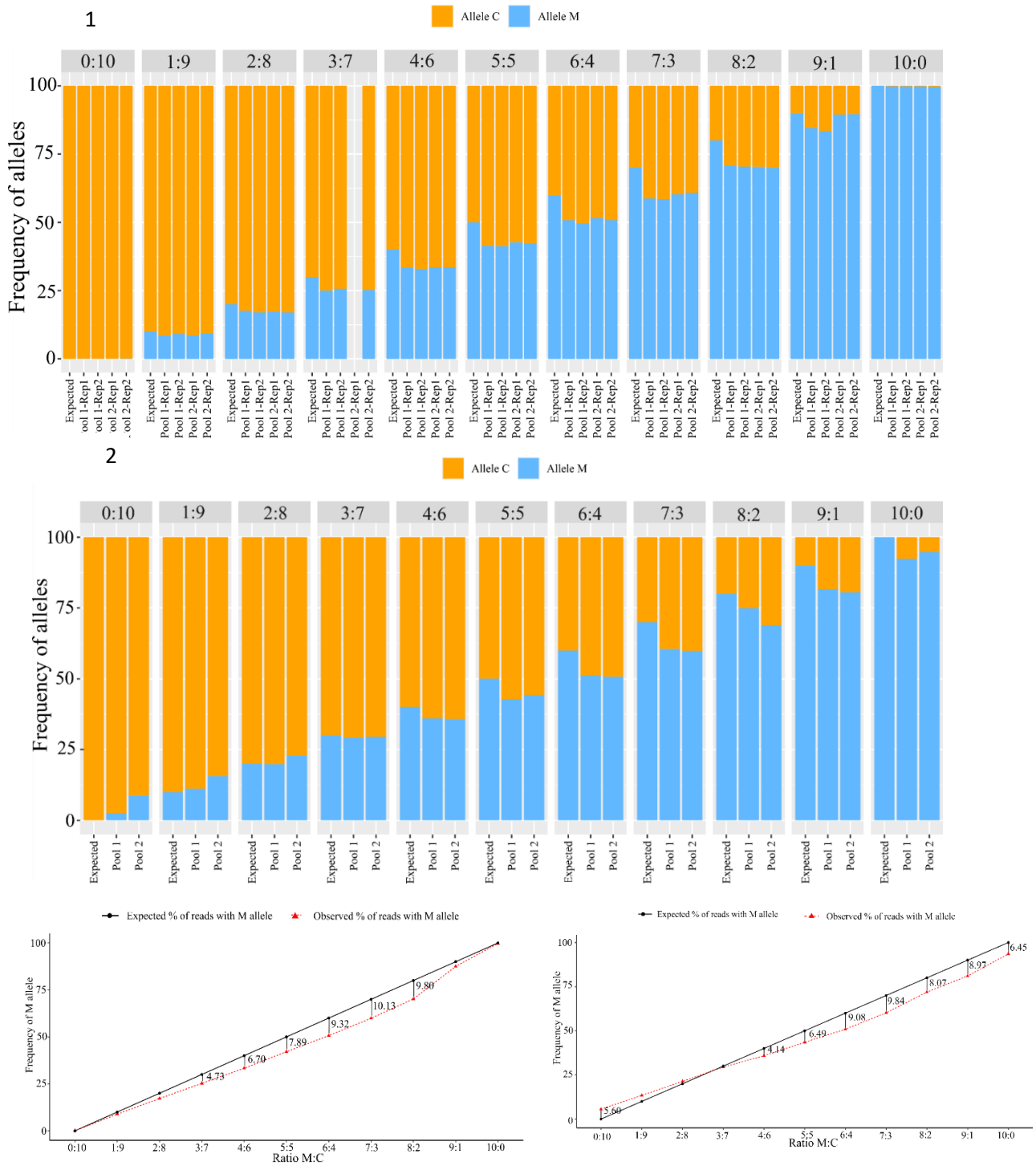

Figure 8: A) Histogram displaying expected M (in blue) and C (in orange) frequencies in the first bar, and M and C frequencies in DNA pools created from whole-genome sequenced pure individuals (1) and pure individuals from Hohen Neuendorf Bee Institute's mating s station (2). B) Expected and obtained M allele frequencies with corresponding errors.

5. Application of the SNP tool to estimate purity in spermatheca content

To verify if the SNP-based tool is suited to the analysis of DNA extracted from spermatheca, 18 queens were artificially inseminated using five semen mixtures with known proportions of *A. m. carnica* and *A. m. mellifera* semen. There is an excellent agreement between the M allele frequency found in the spermatheca content and the semen in all the mixtures. This shows that this tool can be used to estimate allele frequencies in the spermatheca content (figure 9).

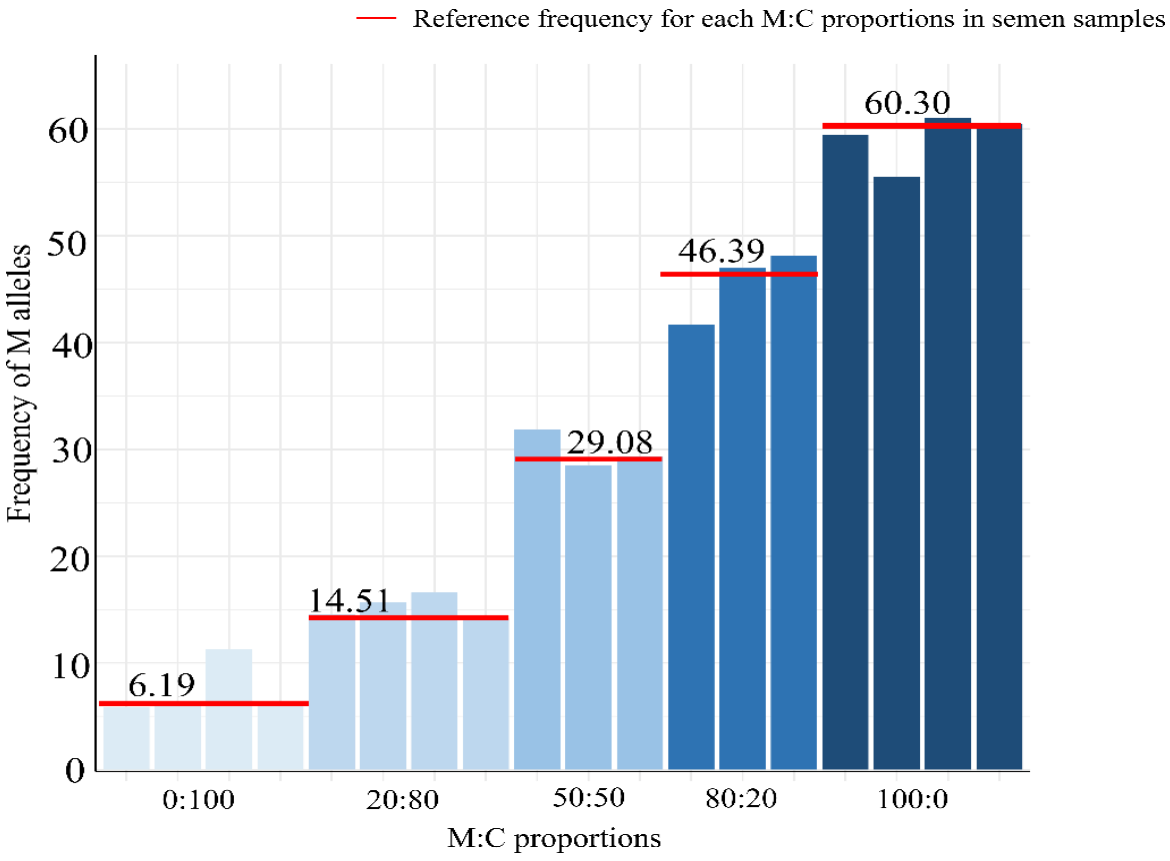

Figure 9: Frequency of the M allele in the spermatheca content of artificially inseminated bees using five semen mixtures, with the M allele frequency of the semen samples indicated by red lines.

6. Influence of queen ancestry on allele frequency estimation

We conducted reciprocal crosses to investigate whether queen ancestry influences allele frequency estimation (see Figure 10). When M-lineage semen (M allele = 90.33%) was used to inseminate a queen, the spermatheca consistently contained a similar percentage, approximately 90%, of M lineage alleles, irrespective of the queen's ancestry (see Figure 10a and 10b). The same pattern was observed when C-lineage semen (M allele

= 9.46%) was used to inseminate queens of different ancestries (see Figure 10c and 10d). These results indicate that queen ancestry does not have a significant influence on allele frequency estimation, suggesting the applicability of this tool in assessing spermatheca content.

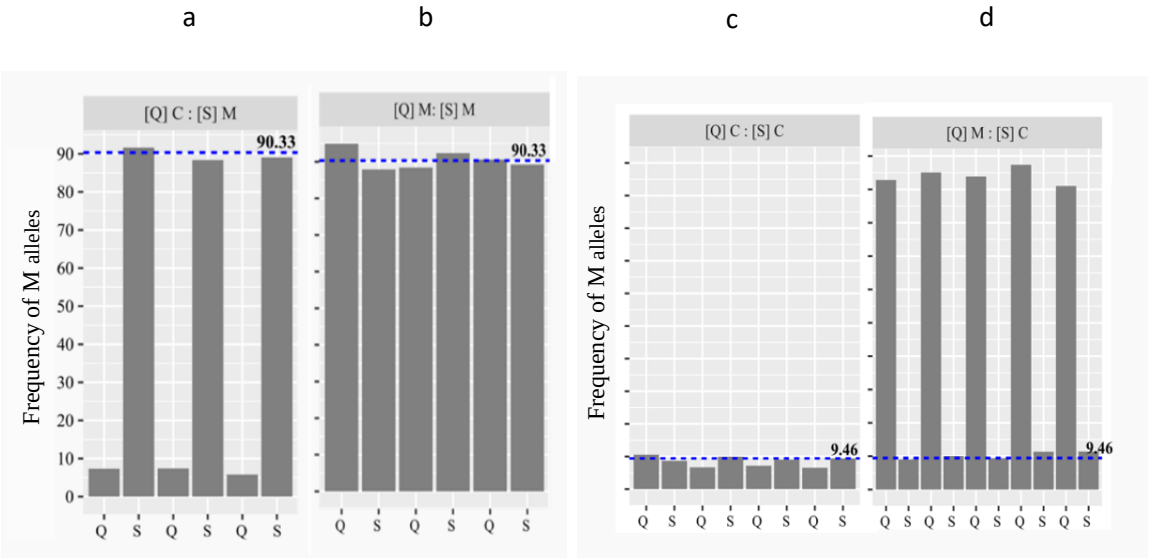

Figure 10: Results of reciprocal crosses. The frequency of M allele in the semen used to inseminate the queen it is represented in dashed blue lines, and the bar show the frequency of M allele in the queens (Q) and in the spermatheca contente (S).
A) the semen from individuals of M ancestry were used to inseminate queens of C ancestry.
B) the semen from individuals of M ancestry were used to inseminate queens of M ancestry.
C) the semen from individuals of C ancestry were used to inseminate queens of C ancestry.
D) the semen from individuals of C ancestry were used to inseminate queens of C ancestry.

The experiments of reciprocal crosses and the artificial insemination of the queen show that this tool can be applied to study the purity of the spermatheca content. This is very interesting because drones from surrounding colonies congregate in an area that virgin queens visit for mating Baudry et al. (1998). Therefore, while studying the spermatheca content involves sacrificing queens, it provides a reliable means of evaluating the genetic isolation of newly established mating stations.

V. Conclusions

Whole-genome data has offered valuable insights into the evolution of the honey bee, and has the potential to inform conservation and breeding programs.

However, a significant challenge arises as whole genome data often requires bioinformatics expertise and computational resources that are typically scarce in conservation and breeding centers. To address this challenge, we have developed a novel SNP tool based on the NEBNext direct genotyping solution. Analysis of 228 whole-genome sequences from 148 M-lineage drones and 80 C-lineage drones revealed 5,007 highly differentiated SNPs. From these, we selected 130 SNPs spanning the 16 honey bee chromosomes. We removed 48 problematic SNPs from the dataset.

Numerous experiments, involving DNA pools with known allele frequencies and semen extracted from the spermatheca of artificially inseminated queens within a breeding program, have showcased the versatility of this tool for both individual pools and spermatheca content.

This toll represents a crucial step forward in the genetic analysis of honey bees, offering the potential for widespread use in various breeding and conservation programs throughout Europe. In particular, when applied to spermatheca content, it provides a reliable means of evaluating the genetic isolation of newly established mating stations.

REFERENCES

Adams, J., Rothman, E. D., Kerr, W. E., & Paulino, Z. L. (1977). Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera*. *Genetics*, *86*(3), 583–596.

Alburaki, M., Bertrand, B., Legout, H., Moulin, S., Alburaki, A., Sheppard, W. S., & Garnery, L. (2013). A fifth major genetic group among honeybees revealed in Syria. *BMC Genetics*, *14*(1), 1–11.

Alburaki, M., Moulin, S., Legout, H., Alburaki, A., & Garnery, L. (2011). Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie*, *42*(5), 628–641.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Antúnez, K., Martín-Hernández, R., Prieto, L., Meana, A., Zunino, P., & Higes, M. (2009). Immune suppression in the honey bee (*Apis mellifera*) following infection by Nosema ceranae (Microsporidia). *Environmental Microbiology*, *11*(9), 2284–2290.

Arias, M. C., & Sheppard, W. S. (2005). Phylogenetic relationships of honey bees (*Hymenoptera:Apinae:Apini*) inferred from nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, *37*(1), 25–35. https://doi.org/10.1016/j.ympev.2005.02.017

Ascher, J. S., Kornbluth, S., & Goelet, R. G. (2014). Bees (*Hymenoptera: Apoidea: Anthophila)* of Gardiners Island, Suffolk County, New York. *Northeastern Naturalist*, *21*(1), 47–71.

Atkins, E. L. (1992). Injury to honey bees by poisoning. *The Hive and the Honey Bee*, 1153–1208.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, *3*(10), e3376.

Baudry, E., Solignac, M., Garnery, L., Gries, M., Cornuet, J., & Koeniger, N. (1998). Relatedness among honeybees (Apis mellifera) of a drone congregation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1409), 2009–2014.

Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E., & Omholt, S. W. (2003). The Gene

csd Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. In *Cell* (Vol. 114). http://genemachine.nhgri.nih.gov/

Birdsell, D. N., Pearson, T., Price, E. P., Hornstra, H. M., Nera, R. D., Stone, N., Gruendike, J., Kaufman, E. L., Pettus, A. H., & Hurbon, A. N. (2012). Melt analysis of mismatch amplification mutation assays (Melt-MAMA): a functional study of a cost-effective SNP genotyping assay in bacterial models. *PLoS One*, *7*(3), e32866.

Blackmon, H., Hardy, N. B., & Ross, L. (2015). The evolutionary dynamics of haplodiploidy: Genome architecture and haploid viability. *Evolution*, *69*(11), 2971–2978. https://doi.org/10.1111/evo.12792

Botías, C., Martín-Hernández, R., Barrios, L., Meana, A., & Higes, M. (2013). Nosema spp. infection and its negative effects on honey bees *(Apis mellifera iberiensis)* at the colony level. *Veterinary Research*, *44*(1), 1–15.

Brodschneider, R., & Crailsheim, K. (2010). Nutrition and health in honey bees. *Apidologie*, *41*(3), 278–294.

Büchler, R., Costa, C., Hatjina, F., Andonov, S., Meixner, M. D., Conte, Y. Le, Uzunov, A., Berg, S., Bienkowska, M., & Bouga, M. (2014). The influence of genetic origin and its interaction with environmental effects on the survival of *Apis mellifera L.* colonies in Europe. *Journal of Apicultural Research*, *53*(2), 205–214.

Chahbar, N., Muñoz, I., Dall'Olio, R., De la Rúa, P., Serrano, J., & Doumandji, S. (2013). Population structure of North African honey bees is influenced by both biological and anthropogenic factors. *Journal of Insect Conservation*, *17*(2), 385–392. https://doi.org/10.1007/s10841-012-9520-1

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-015.

Chapman, N. C., Harpur, B. A., Lim, J., Rinderer, T. E., Allsopp, M. H., Zayed, A., & Oldroyd, B. P. (2015). A SNP test to identify Africanized honeybees via proportion of 'African'ancestry. *Molecular Ecology Resources*, *15*(6), 1346–1355.

Chávez-Galarza, J., Garnery, L., Henriques, D., Neves, C. J., Loucif-Ayad, W., & Pinto, M. A. (2017). Mitochondrial DNA variation of *Apis mellifera iberiensis*: Further insights from a large-scale study using sequence data of the tRNA leu-cox2 intergenic region. *Apidologie*, *48*, 533–544.

Chávez-Galarza, J., Henriques, D., Johnston, J. S., Carneiro, M., Rufino, J., Patton, J. C., & Pinto, M. A. (2015). Revisiting the Iberian honey bee (*Apis mellifera iberiensis*)

contact zone: Maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Molecular Ecology*, *24*(12), 2973–2992. https://doi.org/10.1111/mec.13223

Chávez-Galarza, J., Henriques, D., Johnston, J. S., Azevedo, J. C., Patton, J. C., Muñoz, I., De la Rúa, P., & Pinto, M. A. (2013). Signatures of selection in the I berian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology*, *22*(23), 5890–5907.

Chávez-Galarza, J., Henriques, D., Johnston, J. S., Carneiro, M., Rufino, J., Patton, J. C., & Pinto, M. A. (2015). Revisiting the I berian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. *Molecular Ecology*, *24*(12), 2973–2992.

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., Liu, S., Lu, H., Tian, S., Li, R., & Shi, W. (2016). *Article for the Discoveries section*. http://mbe.oxfordjournals.org/

Claudianos, C., Ranson, H., Johnson, R. M., Biswas, S., Schuler, M. A., Berenbaum, M. R., Feyereisen, R., & Oakeshott, J. G. (2006). A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Molecular Biology*, *15*(5), 615–636.

Cornuet, J., Garnery, L., & Cornuet, J. M. (1991). *Mitochondrial DNA variability in honeybees and its phylogeographic implications*. https://hal.archives-ouvertes.fr/hal-00890963

Cridland, J. M., Tsutsui, N. D., & Ramírez, S. R. (2017). The complex demographic history and evolutionary origin of the western honey bee, *Apis mellifera*. *Genome Biology and Evolution*, *9*(2), 457–472.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008.

De la Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., & Serrano, J. (2009). Biodiversity, conservation and current threats to European honeybees. *Apidologie*, *40*(3), 263–284.

De La Rúa, P., Jaffé, R., Dall'Olio, R., Muñoz, I., & Serrano, J. (2009). Biodiversity, conservation and current threats to European honeybees. In *Apidologie* (Vol. 40, Issue 3, pp. 263–284). https://doi.org/10.1051/apido/2009027

Dowling, T. E., & Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, *28*(1), 593–619.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J. K., Lopez, J., Choi, J.-H., Martin-Hernandez, R., Botias, C., Cousin, M., & McDonnell, C. (2012). Gut pathology and responses to the microsporidium Nosema ceranae in the honey bee *Apis mellifera*. *PloS One*, *7*(5), e37017.

Elsik, C. G., Worley, K. C., Bennett, A. K., Beye, M., Camara, F., Childers, C. P., de Graaf, D. C., Debyser, G., Deng, J., & Devreese, B. (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, *15*(1), 1–29.

Estoup, A., Garnery, L., Solignac, M., & Cornuet, J.-M. (1995). Microsatellite variation in honey bee (*Apis mellifera L.*) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, *140*(2), 679–695.

Ferreira, H., Henriques, D., Neves, C. J., Machado, C. A. S., Azevedo, J. C., Francoy, T. M., & Pinto, M. A. (2020). Historical and contemporaneous human-mediated processes left a strong genetic signature on honey bee populations from the Macaronesian archipelago of the Azores. *Apidologie*, *51*(3), 316–328.

Franck, P., Garnery, L., & Solignac, M. (1998). The origin of west European subspecies of honeybees (*Apis mellifera*): New insights from microsatellite and mitochondrial data. *Evolution*, *52*(4), 1119–1134. https://doi.org/10.1111/j.1558-5646.1998.tb01839.x

Fuller, Z. L., Niño, E. L., Patch, H. M., Bedoya-Reina, O. C., Baumgarten, T., Muli, E., Mumoki, F., Ratan, A., McGraw, J., & Frazier, M. (2015). Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools. *BMC Genomics*, *16*(1), 1–18.

Gallai, N., Salles, J.-M., Settele, J., & Vaissière, B. E. (2009a). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, *68*(3), 810–821.

Gallai, N., Salles, J. M., Settele, J., & Vaissière, B. E. (2009b). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics*, *68*(3), 810–821. https://doi.org/10.1016/j.ecolecon.2008.06.014

Garnery, L., Cornuet, J., & Solignac, M. (1992). Evolutionary history of the honey bee Apis mellifera inferred from mitochondrial DNA analysis. *Molecular Ecology*, *1*(3), 145–154.

Gaudet, M., Fara, A.-G., Beritognolo, I., & Sabatti, M. (2009). Allele-specific PCR in SNP genotyping. In *Single nucleotide polymorphisms* (pp. 415–424). Springer.

Han, F., Wallberg, A., & Webster, M. T. (2012). From where did the western honeybee (*Apis mellifera*) originate? *Ecology and Evolution*, *2*(8), 1949–1957.

https://doi.org/10.1002/ece3.312

Harpur, B. A., Guarna, M. M., Huxter, E., Higo, H., Moon, K.-M., Hoover, S. E., Ibrahim, A., Melathopoulos, A. P., Desai, S., & Currie, R. W. (2019). Integrative genomics reveals the genetics and evolution of the honey bee's social immune system. *Genome Biology and Evolution*, *11*(3), 937–948.

Harpur, B. A., Kadri, S. M., Orsi, R. O., Whitfield, C. W., & Zayed, A. (2020). Defense response in Brazilian honey bees (*Apis mellifera scutellata*× *spp*.) is underpinned by complex patterns of admixture. *Genome Biology and Evolution*, *12*(8), 1367–1377.

Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J. M. D., Alqarni, A. S., Owayss, A. A., & Zayed, A. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences*, *111*(7), 2614–2619.

Henriques, D., Browne, K. A., Barnett, M. W., Parejo, M., Kryger, P., Freeman, T. C., Muñoz, I., Garnery, L., Highet, F., & Jonhston, J. S. (2018). High sample throughput genotyping for estimating C-lineage introgression in the dark honeybee: an accurate and cost-effective SNP-based tool. *Scientific Reports*, *8*(1), 1–14.

Henriques, D., Lopes, A. R., Chejanovsky, N., Dalmon, A., Higes, M., Jabal-Uriel, C., Le Conte, Y., Reyes-Carreño, M., Soroker, V., & Martín-Hernández, R. (2021). A SNP assay for assessing diversity in immune genes in the honey bee (*Apis mellifera L*.). *Scientific Reports*, *11*(1), 1–12.

Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M. T., & Pinto, M. A. (2018). Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evolutionary Applications*, *11*(8), 1270–1282.

Henriques, D., Wallberg, A., Chávez-Galarza, J., Johnston, J. S., Webster, M. T., & Pinto, M. A. (2018). Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula. *Scientific Reports*, *8*(1), 1–14.

Hernández-García, R., De La Rúa, P., & Serrano, J. (2009). Mating frequency in *Apis mellifera iberiensis* queens. *Journal of Apicultural Research*, *48*(2), 121–125. https://doi.org/10.3896/IBRA.1.48.2.06

Hung, K.-L. J., Kingston, J. M., Albrecht, M., Holway, D. A., & Kohn, J. R. (2018). The worldwide importance of honey bees as pollinators in natural habitats. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1870), 20172140.

Ilyasov, R. A., Lee, M. lyeol, Takahashi, J. ichi, Kwon, H. W., & Nikolenko, A. G. (2020).

A revision of subspecies structure of western honey bee *Apis mellifera*. In *Saudi Journal of Biological Sciences* (Vol. 27, Issue 12, pp. 3615–3621). Elsevier B.V. https://doi.org/10.1016/j.sjbs.2020.08.001

James, R., James, R. R., & Pitts-Singer, T. L. (2008). *Bee pollination in agricultural ecosystems*. Oxford University Press on Demand.

Jensen, A. B., Palmer, K. A., Boomsma, J. J., & Pedersen, B. V. (2005). Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, Apis mellifera mellifera, in northwest Europe. *Molecular Ecology*, *14*(1), 93–106.

Johansen, C. A., & Mayer, D. F. (1990). *Pollinator protection: a bee & pesticide handbook.* Wicwas Press.

Johnson, R. M., Ellis, M. D., Mullin, C. A., & Frazier, M. (2010). Pesticides and honey bee toxicity–USA. *Apidologie*, *41*(3), 312–331.

Jones, J. C., Du, Z. G., Bernstein, R., Meyer, M., Hoppe, A., Schilling, E., Ableitner, M., Juling, K., Dick, R., & Strauss, A. S. (2020). Tool for genomic selection and breeding to evolutionary adaptation: Development of a 100K single nucleotide polymorphism array for the honey bee. *Ecology and Evolution*, *10*(13), 6246–6256.

Kandemir, I., & Kence, A. (1995). Allozyme variability in a central Anatolian honeybee (*Apis mellifera* L) population. *Apidologie*, *26*(6), 503–510.

Kerr, W. E., Zucchi, R., Nakadaira, J. T., & Butolo, J. E. (1962). Reproduction in the Social Bees (Hymenoptera: *Apidae*). In *Source: Journal of the New York Entomological Society* (Vol. 70, Issue 4).

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). *Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. 15 (5), 1179-1191*.

Le Conte, Y., & Navajas, M. (2008). Climate change: impact on honey bee populations and diseases. *Revue Scientifique et Technique-Office International Des Epizooties*, *27*(2), 499–510.

Leclercq, G., Gengler, N., & Francis, F. (2018). How human reshaped diversity in honey bees (*Apis mellifera L.*): a review. *Entomologie Faunistique-Faunistic Entomology*.

Martín-Hernández, R., Bartolomé, C., Chejanovsky, N., Le Conte, Y., Dalmon, A.,

Dussaubat, C., García-Palencia, P., Meana, A., Pinto, M. A., & Soroker, V. (2018). Nosema ceranae in *Apis mellifera:* a 12 years postdetection perspective. *Environmental Microbiology*, *20*(4), 1302–1329.

Michael S. Engel. (1999). The taxonomy of recent and Fossil Honey Bees(*Hymenoptera: Apidae; Apis*). *Hymenoptera: Apidae; Apis J. Hym. Res.*, *8*, 165-196.

Michener, C. D. (2000). *The bees of the world* (Vol. 1). JHU press.

Minozzi, G., Lazzari, B., De Iorio, M. G., Costa, C., Carpana, E., Crepaldi, P., Rizzi, R., Facchini, E., Gandini, G., & Stella, A. (2021). Whole-Genome sequence analysis of Italian honeybees (*Apis mellifera*). *Animals*, *11*(5), 1311.

Momeni, J., Parejo, M., Nielsen, R. O., Langa, J., Montes, I., Papoutsis, L., Farajzadeh, L., Bendixen, C., Căuia, E., & Charrière, J.-D. (2021). Authoritative subspecies diagnosis tool for European honey bees based on ancestry informative SNPs. *BMC Genomics*, *22*(1), 1–12.

Morse, R. A., & Calderone, N. W. (2000). The value of honey bees as pollinators of US crops in 2000. *Bee Culture*, *128*(3), 1–15.

Muñoz, I., Henriques, D., Jara, L., Johnston, J. S., Chávez-Galarza, J., De La Rúa, P., & Pinto, M. A. (2017). SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Molecular Ecology Resources*, *17*(4), 783–795. https://doi.org/10.1111/1755-0998.12637

Muñoz, I., Henriques, D., Johnston, J. S., Chávez-Galarza, J., Kryger, P., & Pinto, M. A. (2015). Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). *PloS One*, *10*(4), e0124365.

Muñoz, I., Pinto, M. A., & De la Rúa, P. (2013). Temporal changes in mitochondrial diversity highlights contrasting population events in Macaronesian honey bees. *Apidologie*, *44*(3), 295–305.

Muñoz, I., Pinto, M. A., & De La Rúa, P. (2014). Effects of queen importation on the genetic diversity of Macaronesian island honey bee populations *(Apis mellifera Linneaus 1758). Journal of Apicultural Research*, *53*(2), 296–302. https://doi.org/10.3896/IBRA.1.53.2.11

Nelson, R. M., Wallberg, A., Simões, Z. L. P., Lawson, D. J., & Webster, M. T. (2017). Genomewide analysis of admixture and adaptation in the Africanized honeybee. *Molecular Ecology*, *26*(14), 3603–3617.

Nolte, A. W., & Tautz, D. (2010). Understanding the onset of hybrid speciation. *Trends in*

Genetics, *26*(2), 54–58.

Parejo, M., Henriques, D., Pinto, M. A., Soland-Reckeweg, G., & Neuditschko, M. (2018). Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*. *Journal of Apicultural Research*, *57*(4), 504–506.

Parejo, M., Wragg, D., Gauthier, L., Vignal, A., Neumann, P., & Neuditschko, M. (2016). Using whole-genome sequence information to foster conservation efforts for the European Dark Honey Bee, *Apis mellifera mellifera*. *Frontiers in Ecology and Evolution*, *4*, 140.

Parejo, M., Wragg, D., Henriques, D., Vignal, A., & Neuditschko, M. (2017). Genome-wide scans between two honeybee populations reveal putative signatures of human-mediated selection. *Animal Genetics*, *48*(6), 704–707.

Pinto, M. A., Henriques, D., Chávez-Galarza, J., Kryger, P., Garnery, L., van der Zee, R., Dahle, B., Soland-Reckeweg, G., De la Rúa, P., & Dall'Olio, R. (2014). Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, *53*(2), 269–278.

Pinto, M. A., Henriques, D., Chávez-Galarza, J., Kryger, P., Garnery, L., Van Der Zee, R., Dahle, B., Soland-Reckeweg, G., De La Rúa, P., Dall'Olio, R., Carreck, N. L., & Johnston, J. S. (2014). Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera)* from protected populations: A genome-wide assessment using SNPs and mtDNA sequence data. *Journal of Apicultural Research*, *53*(2), 269–278. https://doi.org/10.3896/IBRA.1.53.2.08

Pinto, M. A., Henriques, D., Neto, M., Guedes, H., Muñoz, I., Azevedo, J. C., & De La Rúa, P. (2013). Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honeybees. In *Apidologie* (Vol. 44, Issue 4, pp. 430–439). https://doi.org/10.1007/s13592-013-0192-y

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Raffiudin, R., & Crozier, R. H. (2007). Phylogenetic analysis of honey bee behavioral evolution. *Molecular Phylogenetics and Evolution*, *43*(2), 543–552. https://doi.org/10.1016/j.ympev.2006.10.013

Ramírez, S. R., Nieh, J. C., Quental, T. B., Roubik, D. W., Imperatriz-Fonseca, V. L., &

Pierce, N. E. (2010). A molecular phylogeny of the stingless bee genus Melipona (*Hymenoptera: Apidae*). *Molecular Phylogenetics and Evolution*, *56*(2), 519–525.

Reich, D. E., Gabriel, S. B., & Altshuler, D. (2003). Quality and completeness of SNP databases. *Nature Genetics*, *33*(4), 457–458.

Ruttner, F. (1988a). *Biogeography and taxonomy of honeybees*. Springer-Verlag.

Ruttner, F. (1988b). Morphometric analysis and classification. In *Biogeography and taxonomy of honeybees* (pp. 66–78). Springer.

Rybicka, M., Miłosz, E., & Bielawski, K. P. (2021). Superiority of MALDI-TOF mass spectrometry over real-time PCR for SARS-CoV-2 RNA detection. *Viruses*, *13*(5), 730.

Saelao, P., Simone-Finstrom, M., Avalos, A., Bilodeau, L., Danka, R., De Guzman, L., Rinkevich, F., & Tokarz, P. (2020). Genome-wide patterns of differentiation within and among US commercial honey bee stocks. *BMC Genomics*, *21*(1), 1–12.

Sheppard, W. S. (1989). *A history of the introduction of honey bee races into the United States. Part 1*.

Sheppard, W. S., & Meixner, M. D. (2003). *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie*, *34*(4), 367–375. https://doi.org/10.1051/apido:2003037

Sheppard, W. S., & Smith, D. R. (2000). Identification of African-derived bees in the Americas: a survey of methods. *Annals of the Entomological Society of America*, *93*(2), 159–176.

Shi, Y. Y., Huang, Z. Y., Zeng, Z. J., Wang, Z. L., Wu, X. B., & Yan, W. Y. (2011). Diet and cell size both affect queen-worker differentiation through DNA methylation in honey bees (*apis mellifera*, *apidae)*. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0018808

Smith, D. R. (2019). Mitochondrial DNA and honey bee biogeography. In *Diversity in the genus Apis* (pp. 131–176). CRC Press.

Smith, D. R., Crespi, B. J., & Bookstein, F. L. (1997). Fluctuating asymmetry in the honey bee, *Apis mellifera*: effects of ploidy and hybridization. *Journal of Evolutionary Biology*, *10*(4), 551–574.

Smith, D. R., & Glenn, T. C. (1995). Allozyme Polymorphisms in Spanish Honeybees (Apis mellifera iberica). In *Ann Arbor (Glenn). T. C. Glenn Is currently Journal of Heredity* (Vol. 86).

Sobrino, B., Brión, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on

SNP typing methodologies. *Forensic Science International*, *154*(2–3), 181–194.

Soland-Reckeweg, G., Heckel, G., Neumann, P., Fluri, P., & Excoffier, L. (2009). Gene flow in admixed populations and implications for the conservation of the Western honeybee, Apis mellifera. *Journal of Insect Conservation*, *13*(3), 317–328.

Spötter, A., Gupta, P., Mayer, M., Reinsch, N., & Bienefeld, K. (2016). Genome-wide association study of a Varroa-specific defense behavior in honeybees *(Apis mellifera).* *Journal of Heredity*, *107*(3), 220–227.

Techer, M. A., Clemencet, J., Simiand, C., Turpin, P., Garnery, L., Reynaud, B., & Delatte, H. (2017). Genetic diversity and differentiation among insular honey bee populations in the southwest Indian Ocean likely reflect old geographical isolation and modern introductions. *Plos One*, *12*(12), e0189234.

Textor, B., Emerman, A. B., Patel, K. M., Bowman, S. K., Adams, S. M., Desmond, B. S., Dunn, J. S., Barry, A., Corbett, S. E., & Elfe, C. D. (n.d.). High-Throughput Screening of 2300 Genetic Markers in S. lycopersicum using the NEBNext Direct Multiplexed Genotyping Approach. *Bioinformatics*, *14*, 178–192.

Tihelka, E., Cai, C., Pisani, D., & Donoghue, P. C. J. (2020). Mitochondrial genomes illuminate the evolutionary history of the Western honey bee *(Apis mellifera).* *Scientific Reports*, *10*(1), 1–10.

Traynor, K. S., Mondet, F., de Miranda, J. R., Techer, M., Kowallik, V., Oddie, M. A. Y., Chantawannakul, P., & McAfee, A. (2020). Varroa destructor: A complex parasite, crippling honey bees worldwide. *Trends in Parasitology*, *36*(7), 592–606.

vanEngelsdorp, D., & Meixner, M. D. (2010). A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology*, *103*(SUPPL. 1). https://doi.org/10.1016/j.jip.2009.06.011

Vasimuddin, M., Misra, S., Li, H., & Aluru, S. (2019). Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–324.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275–305.

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., Simões, Z. L. P., Allsopp, M. H., Kandemir, I., & De la Rúa, P. (2014). A worldwide survey of genome sequence variation provides insight into the evolutionary history of the

honeybee *Apis mellifera. Nature Genetics*, *46*(10), 1081–1088.

Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C., Evans, J. D., Maleszka, R., Robertson, H. M., Weaver, D. B., Beye, M., Bork, P., Elsik, C. G., Hartfelder, K., Hunt, G. J., Zdobnov, E. M., Amdam, G. V., Bitondi, M. M. G., Collins, A. M., Cristino, A. S., Lattorff, H. M. G., … Wright, R. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera. Nature*, *443*(7114), 931–949. https://doi.org/10.1038/nature05260

Weis, W. A., Ripari, N., Conte, F. L., da Silva Honorio, M., Sartori, A. A., Matucci, R. H., & Sforcin, J. M. (2022). An overview about apitherapy and its clinical applications. *Phytomedicine Plus*, *2*(2), 100239.

Whitfield, C. W., Behura, S. K., Berlocher, S. H., Clark, A. G., Johnston, J. S., Sheppard, W. S., Smith, D. R., Suarez, A. V, Weaver, D., & Tsutsui, N. D. (1988). *Thrice Out of Africa: Ancient and Recent Expansions of the Honey Bee, Apis mellifera*. www.sciencemag.org

Yadró, C. A., Lopes, A. R., Henriques, D., Musin, E., Wegener, J., & Pinto, M. A. (2023). Genomic DNA extraction from honey bee (*Apis mellifera*) queen spermathecal content. *Journal of Apicultural Research*, 1–4.

Yang, X., & Cox-Foster, D. L. (2005). Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. *Proceedings of the National Academy of Sciences*, *102*(21), 7470–7475.

Yániz, J. L., Silvestre, M. A., & Santolaria, P. (2020). Sperm quality assessment in honey bee drones. *Biology*, *9*(7), 174.

Appendix

Table S1. Samples used for the SNP assay design

Subspecies	Sample	Country	Dataset	Geographical coordinates (X)	Geographical coordinates (Y)	C-lineage Q-value
A. m. carnica	car_2722	Croatia	Dataset1	45.830	15.980	1.00
A. m. carnica	car_2724	Croatia	Dataset1	45.830	15.980	1.00
A. m. carnica	car_2737	Serbia	Dataset1	45.183	19.717	1.00
A. m. carnica	CAR01	Switzerland	Dataset1	46.895	8.366	0.93
A. m. carnica	CAR02	Switzerland	Dataset1	47.230	8.389	0.91
A. m. carnica	CAR03	Switzerland	Dataset1	46.519	6.370	0.94
A. m. carnica	CAR04	Switzerland	Dataset1	46.878	7.076	0.94
A. m. carnica	CAR06	Switzerland	Dataset1	47.106	8.395	0.93
A. m. carnica	CAR07	Switzerland	Dataset1	47.258	8.866	0.95
A. m. carnica	CAR10	Switzerland	Dataset1	46.560	6.301	0.93
A. m. carnica	CAR11	Switzerland	Dataset1	46.989	8.377	0.92
A. m. carnica	CAR12	Switzerland	Dataset1	46.989	8.377	0.95
A. m. carnica	CAR13	Switzerland	Dataset1	46.293	7.506	0.93
A. m. carnica	CAR14	Switzerland	Dataset1	47.149	7.035	0.94
A. m. carnica	CAR15	Switzerland	Dataset1	46.371	6.219	0.93
A. m. carnica	CAR17	Switzerland	Dataset1	46.338	6.976	0.94
A. m. carnica	CAR18	Switzerland	Dataset1	46.749	6.955	0.95
A. m. carnica	CAR19	Switzerland	Dataset1	47.253	7.058	1.00
A. m. carnica	CAR20	Switzerland	Dataset1	46.338	6.976	0.94
A. m. carnica	CAR21	Switzerland	Dataset1	47.160	8.535	0.91
A. m. carnica	CAR22	Switzerland	Dataset1	46.054	7.246	0.92
A. m. carnica	CAR23	Switzerland	Dataset1	46.097	7.037	0.95

A. m. carnica	CAR26	Switzerland	Dataset1	47.160	8.534	0.97
A. m. carnica	CAR28	Switzerland	Dataset1	46.766	9.008	0.95
A. m. carnica	CAR30	Switzerland	Dataset1	47.262	8.874	0.95
A. m. carnica	CAR31	Switzerland	Dataset1	46.816	7.623	0.93
A. m. carnica	CAR34	Switzerland	Dataset1	46.636	6.385	0.96
A. m. mellifera	ConGL11	Switzerland	Dataset1	46.937	9.016	0.10
A. m. mellifera	ConGL17	Switzerland	Dataset1	47.078	9.151	0.10
A. m. mellifera	ConGL2	Switzerland	Dataset1	47.028	8.983	0.10
A. m. mellifera	ConGL4	Switzerland	Dataset1	46.994	9.074	0.10
A. m. mellifera	ConGL7	Switzerland	Dataset1	46.985	9.096	0.09
A. m. mellifera	ConGL8	Switzerland	Dataset1	47.063	9.055	0.10
A. m. mellifera	ConME3	Switzerland	Dataset1	46.834	8.289	0.10
A. m. mellifera	ConME5	Switzerland	Dataset1	46.834	8.289	0.03
A. m. iberiensis	iber_2006	Portugal	Dataset1	41.775	-8.592	0.00
A. m. iberiensis	iber_2011	Portugal	Dataset1	41.711	-8.662	0.00
A. m. iberiensis	iber_2019	Portugal	Dataset1	41.532	-8.719	0.00
A. m. iberiensis	iber_2022	Portugal	Dataset1	41.555	-8.473	0.00
A. m. iberiensis	iber_2029	Portugal	Dataset1	41.567	-8.562	0.00
A. m. iberiensis	iber_2032	Portugal	Dataset1	40.823	-8.618	0.00
A. m. iberiensis	iber_2033	Portugal	Dataset1	40.764	-8.612	0.00
A. m. iberiensis	iber_2043	Portugal	Dataset1	40.765	-8.498	0.00
A. m. iberiensis	iber_2054	Portugal	Dataset1	40.925	-8.331	0.00
A. m. iberiensis	iber_2063	Portugal	Dataset1	39.799	-8.834	0.00
A. m. iberiensis	iber_2068	Portugal	Dataset1	39.793	-8.779	0.00
A. m. iberiensis	iber_2069	Portugal	Dataset1	39.777	-8.861	0.00
A. m. iberiensis	iber_2074	Portugal	Dataset1	39.835	-8.658	0.00
A. m. iberiensis	iber_2076	Portugal	Dataset1	39.844	-8.585	0.00

A. m. iberiensis	iber_2090	Spain	Dataset1	37.272	-1.749	0.00
A. m. iberiensis	iber_2095	Spain	Dataset1	37.366	-1.952	0.00
A. m. iberiensis	iber_2106	Spain	Dataset1	37.394	-2.106	0.00
A. m. iberiensis	iber_2110	Spain	Dataset1	37.388	-2.230	0.00
A. m. iberiensis	iber_2112	Spain	Dataset1	37.392	-2.248	0.00
A. m. iberiensis	iber_2118	Spain	Dataset1	37.486	-1.990	0.00
A. m. iberiensis	iber_2152	Spain	Dataset1	39.572	-0.808	0.00
A. m. iberiensis	iber_2154	Spain	Dataset1	39.570	-0.799	0.00
A. m. iberiensis	iber_2160	Spain	Dataset1	39.589	-0.632	0.00
A. m. iberiensis	iber_2163	Spain	Dataset1	39.588	-0.604	0.00
A. m. iberiensis	iber_2166	Spain	Dataset1	39.716	-0.633	0.00
A. m. iberiensis	iber_2168	Spain	Dataset1	39.688	-0.298	0.00
A. m. iberiensis	iber_2180	Spain	Dataset1	40.868	0.669	0.00
A. m. iberiensis	iber_2183	Spain	Dataset1	40.964	0.559	0.00
A. m. iberiensis	iber_2187	Spain	Dataset1	40.923	0.486	0.00
A. m. iberiensis	iber_2191	Spain	Dataset1	40.923	0.485	0.00
A. m. iberiensis	iber_2199	Spain	Dataset1	40.766	0.678	0.00
A. m. iberiensis	iber_2242	Spain	Dataset1	41.889	2.344	0.00
A. m. iberiensis	iber_2244	Spain	Dataset1	41.888	2.344	0.00
A. m. iberiensis	iber_2246	Spain	Dataset1	41.868	2.420	0.00
A. m. iberiensis	iber_2247	Spain	Dataset1	41.868	2.420	0.00
A. m. iberiensis	iber_2248	Spain	Dataset1	41.868	2.420	0.00
A. m. iberiensis	iber_2250	Spain	Dataset1	41.927	2.344	0.00
A. m. iberiensis	iber_2251	Spain	Dataset1	41.927	2.344	0.00
A. m. iberiensis	iber_2253	Spain	Dataset1	42.012	2.619	0.00
A. m. iberiensis	iber_2254	Spain	Dataset1	42.012	2.619	0.00
A. m. iberiensis	iber_2255	Spain	Dataset1	41.995	2.634	0.00

A. m. iberiensis	iber_2259	Spain	Dataset1	41.844	2.673	0.00
A. m. iberiensis	iber_2263	Spain	Dataset1	41.834	2.690	0.06
A. m. iberiensis	iber_2266	Spain	Dataset1	41.855	2.583	0.00
A. m. iberiensis	iber_2267	Spain	Dataset1	41.865	2.867	0.03
A. m. iberiensis	iber_2270	Portugal	Dataset1	37.293	-8.483	0.00
A. m. iberiensis	iber_2273	Portugal	Dataset1	37.301	-8.504	0.00
A. m. iberiensis	iber_2277	Portugal	Dataset1	37.288	-8.615	0.00
A. m. iberiensis	iber_2279	Portugal	Dataset1	37.281	-8.671	0.00
A. m. iberiensis	iber_2284	Portugal	Dataset1	37.392	-8.658	0.00
A. m. iberiensis	iber_2289	Portugal	Dataset1	37.389	-8.461	0.00
A. m. iberiensis	iber_2330	Portugal	Dataset1	38.701	-9.243	0.00
A. m. iberiensis	iber_2341	Portugal	Dataset1	39.013	-8.985	0.00
A. m. iberiensis	iber_2345	Portugal	Dataset1	38.759	-9.321	0.00
A. m. iberiensis	iber_2354	Portugal	Dataset1	38.962	-9.270	0.00
A. m. iberiensis	iber_2358	Portugal	Dataset1	38.973	-9.303	0.00
A. m. iberiensis	iber_2368	Spain	Dataset1	36.921	-4.874	0.00
A. m. iberiensis	iber_2371	Spain	Dataset1	36.991	-4.918	0.00
A. m. iberiensis	iber_2376	Spain	Dataset1	36.793	-4.968	0.00
A. m. iberiensis	iber_2384	Spain	Dataset1	36.730	-4.929	0.00
A. m. iberiensis	iber_2388	Spain	Dataset1	36.979	-5.025	0.00
A. m. iberiensis	iber_2427	Spain	Dataset1	38.455	-4.483	0.00
A. m. iberiensis	iber_2430	Spain	Dataset1	38.364	-4.631	0.00
A. m. iberiensis	iber_2434	Spain	Dataset1	38.231	-4.807	0.00
A. m. iberiensis	iber_2442	Spain	Dataset1	38.769	-4.826	0.00
A. m. iberiensis	iber_2447	Spain	Dataset1	38.799	-4.484	0.00
A. m. iberiensis	iber_2452	Spain	Dataset1	38.828	-4.908	0.00
A. m. iberiensis	iber_2485	Spain	Dataset1	42.915	-8.741	0.00

A. m. iberiensis	iber_2489	Spain	Dataset1	43.163	-8.761	0.00
A. m. iberiensis	iber_2498	Spain	Dataset1	43.039	-8.504	0.00
A. m. iberiensis	iber_2502	Spain	Dataset1	43.156	-8.602	0.00
A. m. iberiensis	iber_2507	Spain	Dataset1	43.155	-8.856	0.00
A. m. iberiensis	iber_2512	Spain	Dataset1	43.219	-8.954	0.00
A. m. iberiensis	iber_2544	Spain	Dataset1	36.280	-5.578	0.00
A. m. iberiensis	iber_2547	Spain	Dataset1	43.019	-2.355	0.00
A. m. iberiensis	iber_2548	Spain	Dataset1	43.019	-2.355	0.00
A. m. iberiensis	iber_2549	Spain	Dataset1	43.034	-2.347	0.00
A. m. iberiensis	iber_2550	Spain	Dataset1	43.034	-2.347	0.00
A. m. iberiensis	iber_2553	Spain	Dataset1	43.053	-2.052	0.00
A. m. iberiensis	iber_2554	Spain	Dataset1	43.053	-2.052	0.00
A. m. iberiensis	iber_2556	Spain	Dataset1	43.046	-2.155	0.00
A. m. iberiensis	iber_2557	Spain	Dataset1	43.046	-2.155	0.00
A. m. iberiensis	iber_2558	Spain	Dataset1	42.986	-2.312	0.01
A. m. iberiensis	iber_2560	Spain	Dataset1	42.986	-2.312	0.00
A. m. iberiensis	iber_2561	Spain	Dataset1	43.035	-2.343	0.00
A. m. iberiensis	iber_2565	Spain	Dataset1	43.092	-2.134	0.00
A. m. iberiensis	iber_2569	Spain	Dataset1	43.045	-2.124	0.00
A. m. iberiensis	iber_2570	Spain	Dataset1	43.044	-2.069	0.00
A. m. iberiensis	iber_2605	Spain	Dataset1	41.098	-3.307	0.00
A. m. iberiensis	iber_2607	Spain	Dataset1	41.066	-3.245	0.00
A. m. iberiensis	iber_2611	Spain	Dataset1	40.919	-3.358	0.00
A. m. iberiensis	iber_2620	Spain	Dataset1	41.134	-3.220	0.00
A. m. iberiensis	iber_2624	Spain	Dataset1	41.161	-3.160	0.00
A. m. iberiensis	iber_2629	Spain	Dataset1	41.149	-3.178	0.00
A. m. iberiensis	iber_2636	Spain	Dataset1	40.107	-4.738	0.00

A. m. iberiensis	iber_2647	Spain	Dataset1	39.770	-4.779	0.00
A. m. iberiensis	iber_2649	Spain	Dataset1	39.706	-4.066	0.00
A. m. iberiensis	iber_2652	Spain	Dataset1	39.659	-4.402	0.00
A. m. iberiensis	iber_2656	Spain	Dataset1	39.630	-4.514	0.00
A. m. iberiensis	iber_2659	Spain	Dataset1	39.665	-4.984	0.00
A. m. iberiensis	iber_2851	Spain	Dataset1	42.500	1.504	0.00
A. m. iberiensis	iber_2856	Spain	Dataset1	42.106	2.661	0.00
A. m. iberiensis	iber_2867	Spain	Dataset1	36.721	-5.343	0.00
A. m. iberiensis	iber_2869	Spain	Dataset1	36.645	-5.981	0.00
A. m. iberiensis	iber_2873	Spain	Dataset1	36.500	-6.149	0.00
A. m. iberiensis	iber_2874	Spain	Dataset1	36.500	-6.149	0.00
A. m. iberiensis	iber_2875	Spain	Dataset1	36.500	-6.149	0.00
A. m. iberiensis	iber_2877	Spain	Dataset1	36.659	-5.903	0.00
A. m. iberiensis	iber_2879	Spain	Dataset1	36.655	-6.070	0.00
A. m. iberiensis	iber_2880	Spain	Dataset1	36.655	-6.070	0.00
A. m. iberiensis	iber_2882	Spain	Dataset1	36.280	-5.578	0.00
A. m. iberiensis	iber_2884	Spain	Dataset1	36.280	-5.578	0.00
A. m. iberiensis	iber_2885	Spain	Dataset1	36.659	-6.008	0.00
A. m. iberiensis	iber_2886	Spain	Dataset1	36.659	-6.008	0.00
A. m. iberiensis	iber_2887	Spain	Dataset1	36.659	-6.008	0.00
A. m. iberiensis	iber_2889	Spain	Dataset1	36.564	-5.716	0.00
A. m. iberiensis	iber_2890	Spain	Dataset1	36.564	-5.716	0.00
A. m. ligustica	ITA10A	Italy	Dataset1	44.373	12.032	1.00
A. m. ligustica	ITA11A	Italy	Dataset1	41.542	14.223	1.00
A. m. ligustica	ITA12A	Italy	Dataset1	43.111	12.449	1.00
A. m. ligustica	ITA13A	Italy	Dataset1	40.024	9.171	0.96
A. m. ligustica	ITA14A	Italy	Dataset1	45.016	9.445	1.00

A. m. ligustica	ITA15A	Italy	Dataset1			0.93
A. m. ligustica	ITA16A	Italy	Dataset1	44.588	10.738	1.00
A. m. ligustica	ITA17A	Italy	Dataset1	43.527	13.246	1.00
A. m. ligustica	ITA18A	Italy	Dataset1	42.417	13.870	1.00
A. m. ligustica	ITA19A	Italy	Dataset1	43.839	11.966	1.00
A. m. ligustica	ITA1A	Italy	Dataset1	40.350	18.046	1.00
A. m. ligustica	ITA20A	Italy	Dataset1	43.718	10.948	1.00
A. m. ligustica	ITA21A	Italy	Dataset1	43.773	11.503	0.95
A. m. ligustica	ITA22A	Italy	Dataset1	41.960	13.537	0.98
A. m. ligustica	ITA23A	Italy	Dataset1			0.93
A. m. ligustica	ITA24A	Italy	Dataset1	42.105	14.706	1.00
A. m. ligustica	ITA25A	Italy	Dataset1	41.295	14.085	1.00
A. m. ligustica	ITA26A	Italy	Dataset1	43.057	13.423	1.00
A. m. ligustica	ITA27A	Italy	Dataset1	44.699	10.630	1.00
A. m. ligustica	ITA28A	Italy	Dataset1	44.290	11.877	0.95
A. m. ligustica	ITA29A	Italy	Dataset1	44.641	11.189	1.00
A. m. ligustica	ITA2A	Italy	Dataset1	44.359	11.626	1.00
A. m. ligustica	ITA30A	Italy	Dataset1	44.276	11.731	1.00
A. m. ligustica	ITA3A	Italy	Dataset1	44.438	11.031	1.00
A. m. ligustica	ITA4A	Italy	Dataset1	44.400	11.589	1.00
A. m. ligustica	ITA5A	Italy	Dataset1	44.058	10.183	1.00
A. m. ligustica	ITA6A	Italy	Dataset1	42.428	12.155	1.00
A. m. ligustica	ITA7A	Italy	Dataset1	41.318	14.048	1.00
A. m. ligustica	ITA8A	Italy	Dataset1			0.93
A. m. ligustica	ITA9A	Italy	Dataset1	42.645	11.857	1.00
A. m. ligustica	lig_2680	Italy	Dataset1	44.481	11.401	1.00
A. m. ligustica	lig_2685	Italy	Dataset1	44.730	11.527	0.99

A. m. ligustica	lig_2690	Italy	Dataset1	44.673	10.586	1.00
A. m. ligustica	lig_2691	Italy	Dataset1	44.673	10.586	1.00
A. m. mellifera	mel_2704	France	Dataset1	-0.685	43.453	0.03
A. m. mellifera	mel_2759	Denmark	Dataset1	11.180	57.317	0.05
A. m. mellifera	mel_2772	Netherlands	Dataset1	4.748	53.067	0.02
A. m. mellifera	mel_2780	Scotland	Dataset1	-6.231	56.079	0.04
A. m. mellifera	mel_2781	Norway	Dataset1	6.666	58.274	0.00
A. m. mellifera	mel_2815	Netherlands	Dataset1	4.781	53.104	0.01
A. m. mellifera	mel_2827	France	Dataset1	3.674	44.362	0.05
A. m. mellifera	mel_2830	Belgium	Dataset1	4.338	50.067	0.03
A. m. mellifera	MEL20	Switzerland	Dataset1	47.270	8.684	0.10
A. m. mellifera	MEL35	Switzerland	Dataset1	47.101	9.232	0.09
A. m. mellifera	MEL40	Switzerland	Dataset1	46.922	8.263	0.10
A. m. mellifera	MEL6	Switzerland	Dataset1	47.019	8.056	0.09
A. m. mellifera	SavA11	France	Dataset1	45.520	6.515	0.10
A. m. mellifera	SavA2	France	Dataset1	45.462	6.651	0.08
A. m. mellifera	SavA9	France	Dataset1	45.390	6.647	0.09
A. m. carnica	Ticino1	Switzerland	Dataset1	46.361	8.982	0.99
A. m. carnica	Ticino6	Switzerland	Dataset1	46.223	8.858	0.93
A. m. carnica	Ticino8	Switzerland	Dataset1	46.297	8.790	0.93
A. m. carnica	NL1	German	Dataset2	7.355	53.400	
A. m. carnica	S1	German	Dataset2	7.395	53.498	
A. m. carnica	S17	German	Dataset2	5.538	51.616	
A. m. mellifera	S2	German	Dataset2	12.896	47.724	
A. m. carnica	S21	German	Dataset2	9.716	53.268	
A. m. carnica	S30	Slovenia	Dataset2	14.573	46.078	
A. m. mellifera	S31	German	Dataset2	7.900	51.841	

A. m. carnica	S33	German	Dataset2	12.518	48.546	
A. m. carnica	S36	Slovenia	Dataset2	14.509	46.062	
A. m. carnica	S37	Slovenia	Dataset2	14.509	46.062	
A. m. mellifera	S42	Switzerland	Dataset2	9.477	47.081	
A. m. mellifera	S48	Norway	Dataset2	6.091	58.639	
A. m. mellifera	S49	Norway	Dataset2	6.091	58.639	
A. m. mellifera	S50	Norway	Dataset2	6.091	58.639	
A. m. mellifera	S53	Switzerland	Dataset2	9.450	47.030	
A. m. mellifera	S54	German	Dataset2	8.384	52.005	
A. m. carnica	S58	Austria	Dataset2	15.030	47.861	
A. m. carnica	S59	Austria	Dataset2	15.030	47.861	
A. m. carnica	S6	German	Dataset2	9.542	47.936	
A. m. carnica	S64	Slovenia	Dataset2	14.503	46.067	
A. m. carnica	S66	Slovenia	Dataset2	14.503	46.067	
A. m. carnica	S67	Slovenia	Dataset2	14.503	46.067	
A. m. carnica	S68	Slovenia	Dataset2	14.505	46.070	
A. m. carnica	S69	Slovenia	Dataset2	14.505	46.070	

Table S2. Genomic information, fla**nking region and statistics of the 130 highly-informative SNPs**

Linkage group	Position	SNP variants	SNP position in Amel_HAv3.1	Sequence Ontology term	Annotation impact	Gene / Genes	Haplotype blocks	FST	Mean coverage	Median coverage	Maximum coverage	Minimum coverage	% MD	Frequency error	Start-end position of flanking region	flanking sequence
1	164952	A,G	LG1:1220161	intergenic_region	MODIFIER	GB42162-GB42165	377	1.00	201.01	200.00	376.00	38.00	0.00	13.85	Group1:164802-165102	GTTATCTTTCGAAGCAAGCC TCGWCGGATTAAAAAAGAT ATAAACGTTTGGACGAGGTA TCTTCGTTTGAACTCGAGCA TGTATACTCTTCTTATTTGGA ATACCCGTATCCAACTCCAT CTCGGATTACAGAAGAAAA CCCAAGGGAAG[G/A]GGTCG GTTGATTTCACGTTGATRGC GGACCCTTTCATGGATTCCC GTGAAAAATAACTCTATTCC CAAGGGCTCGGGAAGAAGA CGTGCAAAGTATGAGCGAC

															GAGTATAATCCTCGCGCCAC TCTTTGAAATTCATCGCGCT CCAT
1	1449068	A,G	LG11:4405615	synonymous_variant	LOW	GB47659	2587	1.00	76.20	72.00	197.00	1.00	23.36	64.62	Group1:1448918-1449218	GCATCTAATAAAAAAGTAA ATAATGATAATCAAGAARTT TTATTTTTAATGAATAACAA AAATAATGATACAAATGATA TCATAGATGAAGAAGAAAA TATTTCTGCTTTAAAAAGCA ATAATAATGATAAAAATGA AGACACTACCCTA[A/G]TAT ATTATCCTTTTCCCTTGGATA TTG AAAATTCACAAGAAAATTCA AGTGAACAAGGTAAAAAGA TAAGTAATTTAAAATTAAAA TAAAAATTTAATTATTTAAT TATAGATAATAAATTAAATA ATTAATTAATAATAYGTTAT TATT
1	1452890	C,T	LG11:4401793	5_prime_UTR_premature_start_codon_gain_variant	LOW	GB47660	2587	1.00	127.27	128.00	220.00	25.00	0.00	1.54	Group1:1452740-1453040	AATATGTGAAATCAATATCA ATATATAATAATAATTTACA TAATAATTTACACGATCTTA ATATAAAAACAYGTGATATG TCTATGTCAGAAAGAGAATG AAAAAATTGAATGACGCRTT AGTAGTAAAATGTATCAGAA CGAATCAAATYGAAACGAT CAACTAAAATTTTAATCTCT TAAKTTTAAGCAAAA AAATTTTAGTAAAATTATTA GGCATTAATTAAAAAATAAA ATAATCAGATCTTTAATATA AAATAAAACAATTGGAAAA AGTTAATCCCACAGATTGAA ATGCTATT
1	2921459	A,T	LG1:2425405	intergenic_region	MODIFIER	GB50364-GB50341	4803	0.93	329.36	309.00	921.00	0.00	16.79	73.85	Group1:2921309-2921609	GATTTTGCTATTTTTAATTTC GATACTTCGAAATCTTTTTG TTTGGCATGATCGAATAATT ACGAGGATTTTGCTATTTTT AGTTTCGAAAGGTAATYTGT TGGATACAAAATTTGGAAGA GTAAGTGAAACGATGGATCT CAGATCTTA[T/A]ATTTTAAC TTGGGATATTTATTTCTATTC TAGAAATTAGCTAAAGATTT TGAAGAATAGTTTGAAATAY ACAAAATACGAGGGTTTTAC TATTTTTAGTTTCGAATTGTA ATCTGTTGAATATAAAATTT GGAAAAGTAAGAAACGGTG A
1	2921459	A,T	LG1:2425405	intergenic_region	MODIFIER	GB50364-GB50341	4803	0.93	329.36	309.00	921.00	0.00	16.79	73.85	Group1:2921309-2921609	GATTTTGCTATTTTTAATTTC GATACTTCGAAATCTTTTTG

																TTTGGCATGATCGAATAATT ACGAGGATTTTGCTATTTTT AGTTTCGAAAGGTAATYTGT TGGATACAAAATTTGGAAGA GTAAGTGAAACGATGGATCT CAGATCTTA[T/A]ATTTTAAC TTGGGATATTTATTTCTATTC TAGAAATTAGCTAAAGATTT TGAAGAATAGTTTGAAATAY ACAAAATACGAGGGTTTTAC TATTTTTAGTTTCGAATTGTA ATCTGTTGAATATAAAATTT GGAAAAGTAAGAAACGGTG A
1	4374646	A,G	LG1:3949704	intergenic_region	MODIFIER	GB52268-GB52267	7353	1.00	236.36	217.00	555.00	0.00	5.84	67.69	Group1:4374496-4374796	CGTTGGCAGGGGGCAGCAC GAACCCACTTGTGTTAACGA AGGGACTTATGCTTGGCTTC GGATAGAATCTCGATTTCTC TCGCCGTGTATTGATAAGGA GAGGGGGGGGAGTGAAACC GCGCGTTTCTTTCGAAAATT TRTATTTCTTCG[A/G]AATTT TGGAAATGTTCTTTTTTTTT TTTTTGCTTTTCGTGCAGCTC GATATTCGATTTCGAARTTA TGGATRGAARAAGGAGTTTC TTTTTCGGATTATGGATTGT ACAAGTTTCGGGGSAGGAAA ATTTYGGAGAATTCTTGAGG AAA
1	2921459	A,T	LG1:2425405	intergenic_region	MODIFIER	GB50364-GB50341	4803	0.93	329.36	309.00	921.00	0.00	16.79	73.85	Group1:2921309-2921609	GATTTTGCTATTTTTAATTTC GATACTTCGAAATCTTTTTG TTTGGCATGATCGAATAATT ACGAGGATTTTGCTATTTTT AGTTTCGAAAGGTAATYTGT TGGATACAAAATTTGGAAGA GTAAGTGAAACGATGGATCT CAGATCTTA[T/A]ATTTTAAC TTGGGATATTTATTTCTATTC TAGAAATTAGCTAAAGATTT TGAAGAATAGTTTGAAATAY ACAAAATACGAGGGTTTTAC TATTTTTAGTTTCGAATTGTA ATCTGTTGAATATAAAATTT GGAAAAGTAAGAAACGGTG A
1	4374646	A,G	LG1:3949704	intergenic_region	MODIFIER	GB52268-GB52267	7353	1.00	236.36	217.00	555.00	0.00	5.84	67.69	Group1:4374496-4374796	CGTTGGCAGGGGGCAGCAC GAACCCACTTGTGTTAACGA AGGGACTTATGCTTGGCTTC GGATAGAATCTCGATTTCTC TCGCCGTGTATTGATAAGGA GAGGGGGGGGAGTGAAACC GCGCGTTTCTTTCGAAAATT TRTATTTCTTCG[A/G]AATTT TGGAAATGTTCTTTTTTTTT TTTTTGCTTTTCGTGCAGCTC

																GATATTCGATTTCGAARTTA TGGATRGAARAAGGAGTTTC TTTTTCGGATTATGGATTGT ACAAGTTTCGGGGSAGGAAA ATTTYGGAGAATTCTTGAGG AAA
1	1.7E+07	A,G	LG1:15724409	synonymous_variant	LOW	GB47391	22963	1.00	89.45	89.00	179.00	16.00	1.46	1.54	Group1:16795087-16795387	TAATATATTTCAACAGGTTC GTTGCGGTCGTCTTCAAAAA GCACAGGCATTAGCTGAACA TTGTGGACAGCCTTGGAAAG CTGCATGTTTATTAGGATGG ATACCTCATCATGATCCTAA TTATCAAAATCCATTAATTG ATACTAAATTRCCCATTGAA GGAAATCCTAATAGAAGCTT ATGGAAATTATGTGCTTGGG AACTTTCACAAGATAAACGC GTAGGTAATAATTTTCTTAT TRTATTCATTATTAAAAATA ATTTATGAAAATATTTATAC TTAAATTTTTTTTATTTAGGT
1	1.7E+07	C,G	LG1:15722797	missense_variant	MODERATE	GB47391	22963	1.00	181.28	185.00	375.00	22.00	0.00	1.54	Group1:16796699-16796999	TTAATTTTATAGGCAGTTAA ACAATTATTATTCAATGTAT TATTATTTCCGGATGGTGGT TGGTTAGTTGATTCAAATAA TAATAATAATGATGAACCAT GTACACCAGAAGAAATATCT CGTGATCATCAAATGGAAAA ATTGCGTGAA[C/G]TTTGTAT TCCAAAAATTACACTTCTTT TACATTCTGTAATGTCTGAA ATGAATGAACACGCTGAATG TATTCAATTAGCAGATATTC TTGCTTCTGAACAATATCAA CTTTATAAGGTACATGTTAT ATTCATTAATATATTCATAA GAT
1	1.7E+07	A,G	LG1:15708044	intron_variant	MODIFIER	GB47499	22994	1.00	74.15	70.00	217.00	0.00	11.68	73.85	Group1:16811782-16812082	TATTCATCTTTTGGAAGAAA TTTTCTTTTGWYAATGAATG AAATATCATGTAATTAGGTA AAAATTAAATGARATTAATA ATTTCTAATGAAATATTATA TRTARTAGATTTCTTGTCAG GRAAAATTGTAATGATTGTA ATGATTGCTG[A/G]TTAAAA AATAATTTTTGGATGCAAYG TTTTTSTAATAATTCRAAGA AATTTTTCACGTGCCGAGAA ATATTTCAACCTCATAGATG AAATCTTCTCTTAAAAGAGA TWGGTGATTCTTAGATTAAT CTCTGATGGATACTCATTTA AGCA
1	2.1E+07	C,T	LG1:19475245	3_prime_UTR_variant	MODIFIER	GB42215	29705	1.00	354.39	355.00	774.00	52.00	0.00	29.23	Group1:21085728-21086028	CTTCACCGACCGCCTTTCGA GAACTTGATCGATATCGCRA ACCCTGTCTAGAACCTTCTY

																GTAATTGTTGTTATTCGTTTA CATTAATTACACCACGCGCG CCAACCAAACTTTTAGTTAA CGTTGGATCGAATTGTTTAT AGRTAAATA[T/C]AAATAAT TGCGAATCAATCCAGGGTTT TTGTTATTTTCTTTCTTTTAC CTATTCTTGTTTTATTTTCTC ATCTCAATATCTTTCCAAGC AGTAAAAGTCAAAAGTGAT GTATATATAAACCTTTTAAA AATTCYTTTTCTTTCTTTTTT T
1	2.1E+07	A,G	LG1:19739092	synonymous_variant	LOW	GB45511	29985	1.00	74.36	66.00	180.00	6.00	2.92	50.77	Group1:21434812-21435112	ATGCCACAGAAATCCGGAC GTGACGGTGGCCGCGATCAA AGGTCTGCAAATGGCGATAT CCGAGTGTCAGCATCAATTC ATGTGGCACAGGTGGAATTG TTCATCCTTGACGCCTAGYA GCAGGACGCAGCAGAGCAG CGTCCTCCTTCA[G/A]AGAG GTAAGATAGAGACAGTGGC GGATAGAACAGGTGRTGGG AGGGAGCGATGATAGAACG CGTGAACGATAGAACGAAT AGAAGTAGAAATTGAATTGT GTTGCGTACGACCTGAAAAA AAAAGAAAAAGAAAAAAAA AGAGAAAAAGA
1	2.3E+07	A,G	LG1:22400605	intergenic_region	MODIFIER	GB51609-GB51576	32139	0.91	280.23	262.00	771.00	0.00	16.79	73.85	Group1:22975402-22975702	CAAGAGAATTTCTTAAATTT GTTCAAAACGTTTCCCGTTT TCGTGGTTCGTGCTCTTTAC GCGGCGGTATAAACGCAATT TCCAACAAACGTTAACAAAT TAAGACTTGTCRATATTTAA GTGTAATTCGTGTCCTCMTT ATTTCRATTC[G/A]TTYTGCA TAAATAAAAATACTTGTTAG ATATTTTTTCCTGATATCAGT TTYTCTTATTRATTAAGAGA TTAAGAATCAATCTTTTATC CTCTTTTGTTTTGAAAACTCA TTCGAACAGGAGAATTCCTT AAATTTGTTCAAAACGTTTC Y
1	2.5E+07	C,T	LG1:20843251	intron_variant	MODIFIER	GB51630	35570	1.00	274.51	283.00	464.00	57.00	0.00	1.54	Group1:24534192-24534492	CTTCAAAGAACGTTTCYCCA AAGGAAGGGGAAAGACGCA CGACCCRTTAAGAGAGAGAT TCGTAATTTCTACGTYTGTTC TCGCGGCAAAAGTATCCACA AAGTTAATTATCGCGAACGA GTCCTCCTTTCAACCGTGCC ACGCGTGCCG[T/C]ACTCGAT GCGAGGAAGAAATTTCCGTT TTCAAAAGCCGTCTGTCTCT CCATTATCCGCTACCATCCT

																CTCGTACCACCATCTCTTTCT CGTGCCCCAAACCACCCCCC TTTCGATCCGTGTCTCTGTTT CGCRTCCCTGTGTGCGAGCG T
1	2.6E+07	C,T	LG1:24383575	intergenic_region	MODIFIER	GB55000-AME.7373	38764	1.00	295.15	273.00	812.00	1.00	17.52	64.62	Group1:26405976-26406276	CCAACGTCGCTCCGTCCGAA GRATGGAGAAAAGAGAGGA GAGTGCGCGACGTGCGTGG GGTAGGGAGAGAGAACGCG GAGGAGAAGGATTCTTCCGG AGGCTCTGAGCTGCGCTCAG CTGTGCTCGCTAGGCTGGGA TAGGGGRAGAGAG[T/C]GGC GGAGAAGAATTTTCTCCGGT GGAAATTTTAGCGAGCGCAA CGATAGGGGAGGGCGAGTA GGAGAGGCGGCACGCACAG YTCGGGGTCTACGTACTTTT GTACACGTACACGAGGGGG AGGGRGGGATTGGTTTGTGT CGGTAGARAT
1	2.6E+07	A,G	LG1:24397377	3_prime_UTR_variant	MODIFIER	GB54998	38788	1.00	185.69	187.00	363.00	0.00	4.38	64.62	Group1:26419841-26420141	ATTTCATTCTTTCTTTGTTAA AAGTAACGTGCCATTCAACR TATACAATTATTCGTTCATT AAAATAAATAGAACTTTAAT ACATATTGTTTTTAAATATG ATATTTGAATTCGAATRATA TCAGTTAAATAATTGCAGCA TTGATTTAGRATATGCTAAA TTGTTTTTTAAAACGTAATT ATTAAAAAAGAAAAACTGTT CATATTTAACTTTATACATTT YCAAATACGTATTTTATTTA TTTAAAGTTCAATTTTCTTCT TTGCAAATTATTTAAAGTTT TTTRTAATGTGGATAAGG
1	2.6E+07	A,G	LG1:24421199	intron_variant	MODIFIER	GB54998	38811	1.00	210.66	203.00	439.00	35.00	0.00	29.23	Group1:26443795-26444095	GCCGACGACGTGCTGTGATG AGAGGACGTGACCGGCGCTT GGGTCGAGACCGGTAAACT GCTGCTCGTCATRCTTGGCG TCGTCGAGCTGCTACTGCTC GCGGGAGTCTTGTGACTCTT GCTTGACTTAGCTGTTCCAC AACAGAGACAT[G/A]TGTTT TCAAACGTTTTTTCAGAAAT TCTAATCTCGTTCCAAAAAG ATAAATTAGATTCCTTTTTTC AGAAAAATTTTCACATTCGA TCGTATCTCGARTATAAGCG TTACGYATTACGTAATAAAT ATCTTATCTTTTTATAGAAA AGAG
1	2.8E+07	A,G	LG1:25670998	intron_variant	MODIFIER	GB53926	41281	1.00	121.96	115.00	414.00	0.00	21.90	73.85	Group1:27784310-27784610	AGATAACATAGAGATTGGA GAGAAACAAGAAAAATATA ATTTTAGATACAATTCGATT TTGCTTTGTATTTAAGTAAA

																GATTTTTGGAAAGAGAAATG ATTTTTAAATTCTGATGCAA AATTTATTTATTTCGCAAAC GTAYAGAATTTT[A/G]TTTCG TAYGTTTGTGAGAATTAATT TATCATAAAGATAACGTAGA GATTGAAGAGAAAAGAAAA ATGTAATTTTAAATACAATT CATCGAATTCTGATGASAAA TTTATTTATTTCTCAAACATA ATAAATTTATYATAACGATA TACAG
2	7861322	G,T	LG2:9036254	3_prime_UTR_variant	MODIFIER	GB46590	56913	1.00	33.55	34.00	71.00	1.00	25.55	63.08	Group2:7861172-7861472	CTTTCAATGTTGCCACTATC AAAAAATTATTAATTACTTT TTTTTCATTTATATAATGAA ATTCATCGTTTATAGGCACT TATTACACAACGATTTTAAA ATTCGTACATATAAATATTT TTATAGAAAYTGATTTATTA ACTAATTTAA[G/T]TCTACAT TTCTATTTAATAAAATATAA TTAARTCCTACGTAACTCTA TTAATGAATTAAAACCAAAA TCAACTTTCTTTTTCTAATTA TTTTCCTTTTTTTGTAATTTTT ATAATTGGAAGAAAAAATTT GCAATTTATTAAACTTTAAA
2	9095307	G,T	LG2:10596501	intron_variant	MODIFIER	GB52440	59401	1.00	110.23	106.00	244.00	20.00	0.00	41.54	Group2:9095157-9095457	CTTTGTTTCTCCATTTCTCGA GACGATAATACCACTTTTGA AAACTTTCTTTTATCTTTGAT CCCTTTTTTTCTTTTTCTTTA GAGAAATTGCTTGTAAAGAR CGACGAAATTCGATGGTTAG AAAACAAAGTGACAGATTTT TCATCCA[G/T]ACCTGGTATG AYATAATATATAATGTATAT ATATATAYGTTTCTCTGGAA CGTACATATACGCGCATGGA AGGAACCGTGTTATACATAG AGGAGGAGGGGGCGCGATT GGAGGGTTGGCCAAGATCTG TCTAAGGCATTCCTGGGCCT C
2	1.2E+07	C,T	LG2:12359463	intron_variant	MODIFIER	GB55349	63735	1.00	209.74	218.00	359.00	28.00	0.00	1.54	Group2:11757293-11757593	TGGCTCYCGGCCTCTCCCTC TTTATATATCCCTCCCCTATT AACCCGCCGTAATTACGATA CAGGCTTGGGCGGCCGTAAA CGAAGTAACAAGAGGCAAG GGTGCTCGTGTCGTCCGCGA TGCTGACAAATATCGGCGAT AAAACGCGCT[T/C]GGATAC GATGCCTATCTCTGGAGAAG GCGCAATTATCACCGACCGT CTCGTTRTCGTTTCATAAATT CGACGGAGGAGGGCGGAGA CGGGAYGCGCCCTCCTCCTC

																GTCCCCTCGTTATTTTCATCC GCCGACCTTCTTCTTCGATC CAC
2	1.4E+07	A,G	LG2:14587365	intron_variant	MODIFIER	INR-2	68011	0.97	276.58	274.00	591.00	37.00	0.00	38.46	Group2:14002694-14002994	TAAAAAATTCGAGCGTGTCG AGGGTCGAGGGTCGAGGGG TTCGTCGCCAAGTTTCGAGC TTTCGACCGGAGAGATTCGR CCGTCATCCTCRGAGAGTTA TATCTGTTACTCTTTCTGCTG CTCCRCATCGATTCCTTTTTT TTTTTTTTT[A/G]TGCGAATTT TGGAATTTTTTTTTTCCATCG CTGTTTCGATATCTTCGTTGG GGGAGTTTTTTTTTTTCTCTC TATATTTCAAATTTTCAACG AAAAATGCCATCCAACGTTG AAAAACGTTGGAACGGTTTT TCCTTTTTTTTTTTTTTT
2	1.4E+07	A,G	LG2:14994798	intron_variant	MODIFIER	GB55576	68704	1.00	138.29	141.00	303.00	20.00	0.00	16.92	Group2:14413511-14413811	CGTTGTCCGATTTCCTCTGTT TACTCGGCCAGGGTGTCGAT AATTGAACGCGGAAATCTTT ATCACCRTCCCCTCAAATTT TCCCCTTCCTTTTCTCTCTCT CTCTACACTAAAAATTTTAT ATATATTTTTCACGARTTTTT CGAAGAT[A/G]AGGATAATG GYGGCGTGTACGGAACAAG GAGAGGAGGAGATACGGTT CGAGATCAAACGGTTCCACC GTTTCGAGGTATTTCATCTC CGGCGGGATAGGTGGATGG GCRAAAGCAGCGGGGAGGC SGATACGTAAACACAAAGG GGGTTC
2	1.5E+07	A,C	LG2:15666464	synonymous_variant	LOW	GB55532	69425	1.00	314.79	316.00	684.00	36.00	0.00	50.77	Group2:15084112-15084412	TTCTCGATCCATTTACGCAT ATCTCGGCTAAAAGAAAAG ATCAAGATCAATTATAAAAA CRTATTAAGAAGAAAATAG WAATATTGYATGGRCAGAT ATTACTTACTCTTGACCACTT TCTTTGTCTTCTTCGTCTAAG TAAGAATATTC[C/A]CKCCA CGAGTCAAAATCATACCAAA ATGAATAAAATTTTTCTACT TTATCTCTARGTGTATCTGAT CCACCTAGACGCGGTACGGG TTTTTTTACAGACCATCTAG CATTTTCYTTAAATGCTTTAC CCATAATCGCGTAAAAATTA TTT
2	1.5E+07	C,G	LG2:15670378	synonymous_variant	LOW	GB55531	69425	1.00	436.35	431.00	867.00	98.00	0.00	1.54	Group2:15088038-15088338	AATTGTTTTTATTYAGCGAA ATATCATATTTAAATCTTTC AACACATTTTACCTTTTAAC ATTATTATTAACTTGTTTACC ATAAAGCTTTAAACCCCAYT CATCTTCTTTGATATTATCCA

																CCTTGGATAAGGAAATAGCA GAAGAGCTSACAGAGAATCT TGGGGGTTTATTTGGTACCG AATTCGTTGGCACTGATGGA AGTTGATGTGTTTCAAAATT ATTATTGGAGGAAGTAATTG ATGTCACCGAACCTACTCTG TTACTGTTACTGCCAACTAC TGGAGCATTTAGAGATGAT
2	1.5E+07	A,G	LG2:15760559	synonymous_variant	LOW	GB55483	69581	1.00	573.84	587.00	987.00	96.00	0.00	1.54	Group2:15175855-15176155	CAAGATTTACTACAGCGGGG ATCAACCGATGGACGTCACG TTGAAGAAAGATGGCCGAA AGGTGGTGGAGACCAGCCA CATCAAGTACACGGTGTTCG ACGAGTAYCTGATCATCTTC ATCAAGGATATCGAGAAGG ACGACGCGGGCGT[G/A]TAC GACGTTTCCATAACGAACGA CAGCGGGAGCGTGAGCGGC TCGTTCAACGTGTGCATCAC CGGYCTMCCSGGTCCGCCGA GCGAGCCGCTCGAGGTGAC YGACGTGAACAAGCACACG TGCACCGTCTCGTGGCGCCC TCCCAAGTTC
3	1701565	A,G	LG3:3307649	synonymous_variant	LOW	GB49089	72874	0.99	129.23	131.00	223.00	23.00	0.00	80.00	Group3:1701415-1701715	GATACAAGAAGGTTCATCAA CATCTGGCTCTGGTGTATCA AAGAATGGATGTCGTCCAAG AAGTTGAGGTACTAAAGGC AATGTTAAATATGGATGTGA TGCTCCCAATGCTGCTAAAC AATTTCTAATACTTTCTTTAT CTTGTGGATA[A/G]CGACTTA AATTATCAAGTAAACGATTT ACACACATATGAAGGCAATT TCTTGTAGCAAGCCGACTAG CCGCCARAGTAGCATGTAAA CCTTCTCGTACTTCACCTGA AAAATCTTCCAGTGCTCCTA AGATTGTTTCTAACTGATCT TCA
3	2596388	A,C	LG3:4271414	intergenic_region	MODIFIER	GB46930-GB46931	73682	0.97	675.91	662.00	1547.00	9.00	2.19	21.54	Group3:2596238-2596538	GTTGGAAAGGGAATATACG AATAGACGAGAATGGATCG AGAGACGAAAATGAATAAT AAAATTGATAATTTAAACTA TTTAGAATAAATAAATAAAT AAAAGATATAAAGAAACTA TGGAAAAAACTAATACACC GCTTCATCTATCGTC[A/C]AA AATTTTTCAATTCCACGAAA AATTATTCAATTCGATAAAA AAATACACGAGACGCTAAA AAATGATCGAAGGAACGCG TAATCGATCGAACGAAATTA AGAAAAGTCGTAACGCGAC

															CGCTCCTCGCGACTTTCGAT TGTAAACTGGC
3	4458140	A,G	LG3:6699602	missense_variant	MODERATE	GB55811	76237	0.99	140.89	132.00	288.00	33.00	0.00	20.00	Group3:4457990-4458290	GATCTAAAGGACCAGCTGTG TGGGGTTATAGAGCTCAAAT AGATCGTAAATTATATAAAA AACATCTACAGGAAGAACTT TTTAATACACCTGRTTTGCA AATATGTGAATCCTCTGTAG AAGATTTAATCATACATGGA GATTCTCCAA[A/G]ATGTTGT GGAATAATTCTTAGTATGTT TACTTTTGAATTGGATTATT ATTGTTCAATATTATTAATTT AAAAAAAAAAAAAAATAAA AAATTTATAAAAATCATATT ATAGAGAATGGAACAAAAA TTTATAGTGATGCAGTTGTT ATAA
3	4458749	C,T	LG3:6698993	missense_variant	MODERATE	GB55811	76237	0.99	104.27	102.00	203.00	19.00	0.73	13.85	Group3:4458599-4458899	TGGTTGCCTGTAGAAAAACA ATTGAACTGTTACTTAACAT ATACAACGGAAAAAGTAGA AAAAATTATAAAAGATAAC ATGCATTGTAATTTACATAT TACGGAAGAAATATGCGGTC CTAGATATTGTCCAAGCATA GAAAGCAAAATTYTCAARTT TAAGGGACATAAACATCCA ATTTGGTTAGAACCAGAAGG GCTAGATTCGCCTTTAATAT ATCCTGCTGGATTATCRTGT ACACTTCCAGCAGAGAAAC AAGAAGAACTTATTAAATKT ATTCCAGCATTAGAAAATGC AAAAA
3	4460272	C,T	LG3:6697470	synonymous_variant	LOW	GB55811	76237	0.99	155.77	150.00	330.00	19.00	0.73	49.23	Group3:4460122-4460422	AAGAARGAAATATTTTGGAT AGCATTGTTTCTTATTATCAT TGTAATATTATCTACTATTTT ATCTTCTTAGATTGCTGCAG CCACTCGAATTTCAGGAATT ACTCCATGTGCTATTTTACG ATTAATACATTATATAAGGA AAAATGAAYTTAATATAAA ATAAAATATGATAATRTATT AATTTAAATAAATAGTTTAA AAAATTTTATTTAAAACGAA ATGAAGATATTTATTAAAAT TTTGCCTTAAAATCGAAAAC AATTCAAATTTTTAAAAAAT TATTTAACCGTTTGAACGTA
3	7006583	C,T	LG3:7948185	intron_variant	MODIFIER	GB53701	80927	1.00	140.38	104.00	434.00	13.00	1.46	63.08	Group3:7006433-7006733	TTTTSTTTMGTGGTTTTTCCA AAGATTTTGCAATTTTTAAT AATATTCGGACATTGGATTC AGATCATGCCGCGAGAAAA GAATAAATCGCTTTTAGGAG ACGAAGAATTTGGACRAYC YTTCGCTTCGAATATATTTA

																GAAACGAAATT[C/T]GRTTTA KTTGAGAYYTGAGAYATTTT TCCCTGCTTTTGAGAAGAAT TATTAAGAAACACTTATGGA TCAYAGTTTCCTRGRAAACA GRGGATTTTGAGRRTGGCRT AGCTCYTYAGTTTCATTCCT TAAATTAGTGTTAACTCATG ATCC
3	7069573	C,T	LG3:8010641	intron_variant	MODIFIER	GB53701	81088	1.00	178.42	175.00	369.00	29.00	0.00	53.85	Group3:7069423-7069723	TTTATTTAGGTCGTGTGTATT TCTCGAATAAAATAACCCGC TCGATTCTTCTCTGTTCACAA ATCGTTGTTGATCGATTTCCT TGTGAATTATCTCCGATATA TACCTYGTATATTTTATAAG TCATCGTCATTTTTCATCTTC TTCYTTYRAATTATTATTTTC CTATTGGAAATATGCATAAA TATGTGAATTTGGATTTAAA TTGYTTCGTTAAARTAARTT TTCTTTARCAACAAATRATT AACTCGTTCTGGTTATTTATT ATATATAACAGGTGGAWTT GGAAACTAGGTTTGAR
3	7069664	A,G	LG3:8010732	intron_variant	MODIFIER	GB53701	81088	1.00	110.46	107.00	276.00	0.00	8.03	66.15	Group3:7069514-7069814	ATCTCCGATATATACCTYGT ATATTTTATAAGTCATCGTC ATTTTTCATCTTCTTCYTTYR AATTATTATTTTCCTATTGGA AATATGCATAAATATGTGAA TTTGGATTTAAATTGYTTCG TTAAARTAARTTTTCTTTARC AACAAAT[G/A]ATTAACTCG TTCTGGTTATTTATTATATAT AACAGGTGGAWTTGGAAAC TAGGTTTGARAARAAGAAAC GCTGTTAAATTTAATTTAAT ATTCTCCGGTTAGATTTAAC ACYAAAAGTGATCGTAATAT CGGGGAATTAAATTGACCAT T
4	4379048	A,G	LG4:4673399	downstream_gene_variant & intron_variant	MODIFIER	GB49569 & GB49566	98129	1.00	78.66	65.00	229.00	4.00	20.44	58.46	Group4:4378898-4379198	TACAATTCAATTATCTATAA CAAGTAATCAAATAYAYAT ATATATTATAGGTAGTTTTTT MTYAAAAACAATTCAGATA CARAGGTTTTTTATTAGATT ATYCTTCCACGTTTATCATTT CGCAARTAATATCARYTTTT CGAATGAAAT[A/G]TAACGT GCTTTTTAACCCTTGTCCTCG CGTTCATTCAAAAATATTCC TTAAAATCCACTTTTAAGGA TTAATTTTTCTTTTTCGACCA ACAAATTGTYGATATTCCCT GTTTGGTTAAATTGATAGAT TAACTTATTGTGAACAAGAG TA

4	4383542	C,T	LG4:4668909	upstream_gene_variant & intron_variant	MODIFIER	tRNA-Ala & GB49566	98138	1.00	83.46	75.00	234.00	12.00	4.38	35.38	Group4:4383392-4383692	AAGGGAAAGAAAAAGAAAA AAAAACCAACTTGGAGGCG CCGGGCATCGATCCCGGTAC CTCTCGCATGCTAAGCGAGC GCTCTACCATCTGAGCTACG CCCCCCGATAAGAAAGATTG GWATTCCGCGAAACTTCCG GTGYGATAACGCG[C/T]TCC CAATTTTTTYCCCCYTTTTTT YTTTTTTTAATTGATTTTTCA CTGATTCTCTGAGCTTTAATT GACAGCCGGTTAATTAAACG AACCGGTTGCAAYAGATTAC ACGTAAGAATCGARTGCTAT TATCCCTCGTTGAAATTCGT TTAT
4	7672595	A,G	LG4:6820258	intergenic_region	MODIFIER	GB50393-GB50392	104402	0.90	511.13	487.00	930.00	83.00	0.00	1.54	Group4:7672445-7672745	YATTTTAAATAATATCWCW ATTGCAMAAATATTAAATTT GAGYGARAATGTCACAGCA TCTTTCGTTCGAGTCGACGA ACATTCCAAAATTGAGATTC CAATTTTTTCCAATTTCGAG AATATATATATAAAAAAAA AAAGAAAAGAAAC[A/G]GTT CCATTAATTGTGAGAATGTC ACAGCATCTTTCGCTCGAGT CGACGAACATTCCAAAATTG AGATTCCAATTTTTTCCAATT TCCAGAATATATATAAAAAA AAAAAAGAAAAGAAARGAA ACGGTTCCATTAATTTTYCG GGAGGAA
4	7944946	A,G	LG4:9135796	intron_variant	MODIFIER	GB42990	104779	1.00	50.19	48.00	134.00	0.00	19.71	73.85	Group4:7944796-7945096	TGGTAATATTGATACTGTCT GCCCATATATCACAAACACG CATCTCGTGACTGCATAACG GATGCAGAACCGATGGTGTA ACCGGGGTATATACAACGTT GCAAATTGTGCCCCGYTTCA AATCGCTCGAGATTTACACA AACGTTCCGCRTATATATGT GTGTGTGAGTAAGCTGATCA GCCAGRATCGAGGATTTGCG ATTATTTTTGCATAAGCTTCT TAGATGTAYGGTGTGTGTTG TGGARAAGGAAATCAGATRT AAGTCGATGTAAATTGTAAG TTGAAAAAGATACGCAAGG T
4	7946420	A,G	LG4:9134324	intron_variant	MODIFIER	GB42990	104779	1.00	122.23	121.00	342.00	3.00	10.95	64.62	Group4:7946270-7946570	ATTTTGTAACATTGAATTCA ACAAGTGCATATTTGCATTT TTAATTTTAGAATTATAAAC ATATTATGAAATCTTGTAAC ATTGAATCCAACAAGTGTCG AGTGAAATAARTCCTATCAT ATAAAAATAAAATAACAAT AATAATATCCG[A/G]AGAAT

																ATTTTAATTTTAGAATTATA AACATATTGCGAAATCTTGT AACATTGAATCCAACAAGTG TCGRGTGAAATAAATCCTAT CATATAAAAATAACAATAAT ATCCGAGAARAATATTTTAA TTTTAGAATTKTWAATATAA AATGT
4	1.2E+07	A,G	LG4:12968747	intron_variant	MODIFIER	GB53016	111889	0.88	53.68	41.00	460.00	0.00	24.09	69.23	Group4:12216624-12216924	CCGAACGTTTGATTCAAGCG ACAACGATATTCGTGTGATC AATGCTTTAAGAAAAAGAAT GTATATAATAAAATATGTAA AAACATTAATGTAATTTATC AGAACAAATCCGGCAAATA AAAGAAAAGGATAATTAAT YATCGTATCCTC[G/A]CTATT ACGCTTCCATWTCGCTCGAA TCACGAATCGAAAAATTYCG ACCCCGATCGAATGGAAATT RGGAACGTTTTATTACGTAC AGGAYGTATATATCGAAGC GCGGTATCCTTTTGTCTCGG AGTGAAAAAGAAGACAACA CTTTTCG
5	316805	C,T	LG5:993002	intron_variant	MODIFIER	GB44402	112504	1.00	42.98	34.00	130.00	0.00	29.93	64.62	Group5:316655-316955	AATTGGGGAAAGTTGGTGA GCTTTCGTCTTTCTCGATGG GCTTTATTATCGAGGCGGTA ATAAAAACGCGCATATATCG TCCGGTTGCCCTCCACGTGC ATGTACGATCTGTGCACGCG CGTACAGACATGGCTAACGA TTTATTCTTAA[C/T]GTCGAT CGCCTCGCCATACCGCATAC TTTATGCTCCAGACGATTTTT TCAACCCACCTCTGCACCCG CCCCTCACCGTCTCCCGATA ATTCGAAATGTAAACCAGTT TTCTGAGCATACGCGTCGGA CCGTCGGACACACAATGAGC GAG
5	871727	C,G	LG5:1484240	intergenic_region	MODIFIER	GB52298-GB52303	113347	1.00	278.09	287.00	693.00	19.00	0.73	63.08	Group5:871577-871877	TGGAATATAGTTTTTCGTAT CCAAGARAGATAATAATAA ACTCGACAATCGTTCGAAAT ATCGAGTGGAATATTTCCAA CTAGRTAATTTTTAAATTGA ATGTAATGTTATCGAAAAGA TATCATAGATTGAAATAATA TTTTTTTTAAY[C/G]ATTCAT CTTTCCRGKATTTAATAAAT TGTAATCGAACAACCTTACG AAYATCGATTCTACAACTCT GTTTTAATCAYCGAAAACCA CCTAATTCTATCTATACAAA GTCCCTTAGTAATAGTTTGC AACGCTGTGCAAYACCAGA CAAAT

5	2361058	A,C	LG5:3177410	intergenic_region	MODIFIER	AME.21389-GB48879	116100	1.00	255.20	257.00	527.00	25.00	0.00	1.54	Group5:2360908-2361208	ACCGACCTTCGTGCGCGCAC ACGAGAAAAAACCACGCAG TCGAAAGGGGGTGGAAGCG AGACCGACCGCTTAAAAGA GAGACGTGACGCAGAAGCG GTGCGCGCGCGCGCGCGTGT GTGTGAGAAGGGCGAGAGA GGCGAAGAGAGAGAG[A/C]G CGCAAAAGCTTACTCTGCCG ACACACACGCGGGAACGGA CGAGCGCGAGTAGAACGTTC GTCTCGTGCACGTTCACGCG CGCGCGCGCGCGCGCACAC ACACTCGCATCCTAATTTTA GCAGGCATTCGGCGCAGCTG TGCCACCCACA
5	2364969	C,T	LG5:3181300	intergenic_region	MODIFIER	AME.21389-GB48879	116110	1.00	147.26	134.00	289.00	14.00	1.46	3.08	Group5:2364819-2365119	ATAATTAAACCTCTCAACGA TATCACGTAACAATTACCCA AAAAACCGATAGTTCGAAC AAYCTYCCCTAAATCATACA TCCTCCGACTAGGAAAGAAA GRAATAGAAAAAAGAAAAA GAAAGAAGGTGGACAAAAA TAGGATACATYTT[C/T]TCCC AAACGTTAAAATTCCGTAAT AAAAATCRTTTCACAACAAA CAAACACACTTAAACATCCG TGTATTAAAACGAATACAGR AAAGCTCGTGGAGTAGAYG TTTCATGKCCYCGTCTCGAG GATTCCTCAGGAGATCGAGA TTCTTGC
5	3181962	A,C	LG5:3607801	intergenic_region	MODIFIER	GB55036-GB55037	117141	1.00	146.04	134.00	421.00	20.00	0.00	55.38	Group5:3181812-3182112	TTTTCAARTTTTATTTATAAA TGTTAGAARCARAARTRATT AAAATTTSTTATTTGGAAAC AGTTACWCGATATTATACGT ATTAAAAAATAAATGCAAA GATCATAATTGTATTTTTTTK TTCTAGAAAAATTTAAGAAA TTTAGATCA[C/A]TAAATTTY AATTTTYACTTTCACGATAG ATACTTTCGAAATTKATTTT CTAAGAAATTTTAATCATCT TTGATTTGACATTCTATAAA TTCCCATGGAAGCAAAACTT CTCGATTACCATCCAGTAAA TATATTTCTTACATAGATTAT C
5	3507388	C,T	LG5:3827148	intron_variant	MODIFIER	GB51391	117354	1.00	359.85	363.00	883.00	10.00	1.46	50.77	Group5:3507238-3507538	TGCGTTACRATTCATGGARC TTCGTCTAATCTTTCCCGGAT GGGATGAAATATGTTCGAAG ATGTTCAGACTAAATGGTAC GAATATATTTTAGAATATTT TATTTCCATGGAAACGATGC RTAAATGTTTCAATTCTATW GGAATAAAA[C/T]ATTTCTTT

																CTAAAATATTCGTACATCTT TTTAGTATATTTGGAARCGA TGYTTTGAATAATCGTTTTA ATATCTGTAAATAATATGGA GGGAATTTCACGAATTTAAA ATAGTTCTTTTCATCTATGCG ATAGTTTAGAGGCAAGTCTT A
5	3510848	C,T	LG5:3830608	intron_variant	MODIFIER	GB51391	117348	1.00	150.24	130.00	519.00	0.00	18.98	73.85	Group5:3510698- 3510998	AATTACAYGTATTTATCATA AGTTTTMAAARAATAAGAA ATTATTAGAATAYCTTCTTTT AATTWAYCTTACAAAAGAA YATGTAAATTTAACCTCATA STTCTTYTTATAACATAYAT YTAATCACTGTTCACTATTA AACTTTRTTAY[T/C]GTAGAT TTTAAATTGAATTACGRATA TTTATTTCYCRTGAATYTAA AAGCATAAAAAATTGCCATT ATTTTAAGATCTTAATCGTG AAGATTCGTTAACTCGATTC GCGAGATTCTTGGAGGGACT CGACTCCAAGGATCAATGTC AACA
5	3805632	C,T	LG5:4122265	intergenic_region	MODIFIER	GB51387- GB51386	117817	1.00	83.77	84.00	147.00	18.00	0.73	1.54	Group5:3805482- 3805782	TCGATTAAAAATGTAATAAT AACACGCGTATTAATTAAAC AAAAATATTAATAATATACA AATTATACTACAATATCAAA TTCCATAAATTATATATCTA TTTATCTATATGAGATGACA AGTATTTTTCCTACGATTCCT GATAATAAA[T/C]TRATATTA TGTAATTAATTTTTCTTACCR TTAAATATTAGAAGCTTCCT GATTGACAAGAATCATCTAT CCTATTCAAATCGACTGCAC AAAAATTATATAATTTATTT CTAAATTTTTATAACTCCAA CAATGATCCASTGATTAAAA A
5	9447330	A,G	LG5:11796464	intergenic_region	MODIFIER	GB44551- GB44550	128010	1.00	213.74	219.00	399.00	28.00	0.00	1.54	Group5:9447180- 9447480	AGAGGAAGATTATCTCTCCT CCYMTCCCCYCCAACAATTY CTCCAACTTTSTTTTATTTTA AAATTATTTTTAAACCGTTTT TAATTCAACGTCAACGCCTC GCCTTTCATCTCTAATTTCGG TGATATTGATATCTTTCTCGT CGGAGA[G/A]ACGGCGAGGG ARAAAGAGGGAGGAGGAAG GGAGGCGTAATTAATCGCGT GGCCCGATGTTTAACTCGAT CACGAGGAAACCCTGTGCCT GATAATTTGCTCGCTGTTAA GTAATTTCACGGTCKACGAT ACCTTCTCCTTTTTCTCCTTT

5	1.1E+07	A,G	LG5:10018712	synonymous_variant	LOW	GB44641	131058	0.99	307.91	295.00	627.00	79.00	0.00	7.69	Group5:11210527-11210827	ATGTGTTTATTAGCATTTTG GTTTAAAATTTCAARTATAA AAATTATTAATTACTTATCT ATAATATGTTTACTTATTGG AAGTTATGTTACATATTTAG CAATATCTCCATCACCACCT TTACAAGCAACAAAACTTGG CATTTTTTTARTAATAATATC ATGGGCACTCTTAATGGGAT TTATTTCTTATTTAAAATTAA TTATTATATCAGTATTTAGA AATACAGTACTTCCCAATAT AYTCTTTAATGTAGGTGCAA TAATGCAAATAGCATCTACA AGTGGTGCAATATTTTCAT
5	1.3E+07	A,G	LG5:12717230	missense_variant	MODERATE	GB41366	133224	0.99	203.53	198.00	389.00	22.00	0.00	1.54	Group5:13034920-13035220	CTCGCTGCAGAGTCATCAAT GCACGGTGGATAAGTTAAGT TATTCTCTGTATAATAGTATT CGTTTCCAACTTCGTTTTCAG TCGAATATTGGAAACCTGCT GTGACTGTCTCCATATTGTC ACCGATGTGCATTTCCGAGT CTAACACGRTCGAGTCCCTA GATAGTCCGGACACGGACTG TACAATTGTTTCTATTCGTCT ACCTCTTTCGGTGCAGCTAT CTATTTTCATATTATCGTGTT CATCCTTCCTTTCCCTTGAAT TTTCACTGACTCTACCTCGTC CCACAGAAGGTGATC
5	1.4E+07	A,T	LG5:13222378	splice_region_variant&intron_variant	LOW	GB41325	133846	0.99	539.33	523.00	1316.00	7.00	10.95	64.62	Group5:13545922-13546222	GAAGGCAGACAAGGTTTCTA GAATAGCAGGATAATTTAAC TGTCGAGACCTGGATCTGGA ACAGAATTGAGATTACAATT TTGTAAAGCAATTTATCARA AAAGATTAAATAATATTTAC ATATAATAATARTTGATTGA TCTWGCATAAWACCTGTGT ATTCGATGCAYTGAATTGAT AAAATGTAATCATTCTTGAA CACTTTAACCTCAAATTTTG CTAACCTCTCTGCTSTGAAA CRTGAAGAGAAAGAAAAAT GAAGTTGAATCATTATTTTA AYAGTRAATTTTATCAAAAT TGT
6	5458628	A,G	LG6:6167306	intergenic_region	MODIFIER	GB46090-GB45995	142019	1.00	176.12	190.00	330.00	0.00	7.30	26.15	Group6:5458478-5458778	GGAACACTGCGTGAACGTGC AGGTTTATTCGCTGCTGTAA ATCTTGGCAAATCCTCGAAG AATATGGCCGAATTCYATCC GATTTCGGGGGTGAGGGKG ARGAAGGGAGRAGGGTAAC GTTTGACGTTTCTATTTTGAA AGGGAGGAAGG[A/G]TRGAG GAAATCGCAGAGGCGATTTT TTTTAATGGAGTTTCTGGAT

															TCGTGGAAATAAAATSGGGA ATTTTACGTATTTTTCGGGA CGAAGAATTTTAAAATGAAC GACGTRATTGTAGTTTTAAT GGAAAATGGACASGYTGGG AATTCT
6	5918385	C,T	LG6:13400166	synonymous_variant	LOW	GB46004	142908	1.00	92.33	78.00	255.00	7.00	2.92	41.54	Group6:5918235-5918535	AATTCACARTTGAGAGCAAA AGTATGTATTATTTTGAAGT GAAATTTTTARTAGATTTAA TATATTTTGTTYTTATTAGAT TTCAGTTTTAAATGTTGAAA AAGATAAGTTACTTTCAGAC AAGGAAAAATCGAAAACAG ATATAAAGGAYAAGGTAAA ATATATTAAATATATTAATA ATTTTTAYRTAAAAAAAAGA AAATAATTTATAGATTAAAT ATATGTTTAATGTTAGATGG AAACAAAATGTACYAAAAA ATRTGAAGAAAAGATTCAA GAAAATGAAGTTATGAAAA CTAAA
6	5921026	A,G	LG6:13397525	missense_variant	MODERATE	GB46077	142908	1.00	126.44	127.00	206.00	20.00	0.00	1.54	Group6:5920876-5921176	CTGCATCGTTCATTTCGTCTT GATTTTCRTATTTATCGTAA GGAGAATCTTCAAAATTTGT TTCGAATTGATTGAAATAYT TTTGACCGAATAAAACGTTT GATTTTGATAACGGAATAAT ATTGGGCTGGATAGATTTTG ATTGATCAA[A/G]TTTTAGGT CTCTTTGAATAGTTTCTGAA GGATTAAGATGACCGACACC ACTCACGCTGCGAACTTCAA AYGCAGAAGAATCGAGTCG CGGCCATGATGGATTCCAGT AATCGGTTGTGTATGGAAGA ATTTGCCATGCTTTTCCAATT CC
6	6003400	A,G	LG6:13315198	missense_variant	MODERATE	GB46070	142927	1.00	195.39	186.00	374.00	22.00	0.00	16.92	Group6:6003250-6003550	TTTAAATTTTGTTTTTTTAAT TGAGCAATATATGTTACTCT ATCTGATGATGTATTATTTTC ATCGAATGTGCTGTTAGAAT TAGTTATTTCTATATTCTCAG AATTAGTTATTAAATCTTTA TCATTTAATTTTTTTTGTTCA TCAGGC[A/G]TTTGTTCACCA TTTGGTACATGATCATTCAT TATCTCAGATGTTTCATTTTC TTTTTCCAAATCTATATGTCC ATTCATTGTCTTTGTTTCAGA ATTTTCACTATTTTCTTTATT TACTTGCACAGATGCACCTT CATCATCCGTAATAT
6	6024411	C,T	LG6:13293583	synonymous_variant	LOW	GRIP84	142927	1.00	53.01	52.00	99.00	11.00	4.38	15.38	Group6:6024261-6024561	TAAAATTAATTTTCTCAAAA ATATTATTTAAAAATTAAAT TGTCTTATAATGCTATATTTT

																ATGAAACATTCAGTATATTT TTTATATAAAACAAAAATAA AATTTTTAATATTATCATTA AATTTTATTTTTTCAATTAGG TTAGATTTYAATTTATTTTAT ACCGACATTTTAATAAAACG TGGAAGAGAAAAAACAATA CAAGAAGATATTTCTGGTTA ATTATTACTCATTAGAAGAA TATATATAATAAATTGTAAT TTTAGTTTGTGATTAATTATT GTAAATTAAAATTATAAT
6	6305901	A,T	LG6:13011712	intron_variant	MODIFIER	GB46038	142979	1.00	86.94	85.00	165.00	27.00	0.00	13.85	Group6:6305751-6306051	CTTGTAAGTTTAATACATAA TTCAATCATTTGTATCCAAT AAATATCTTTAATTTTTAGT AAAAACAAGTTCCTACGATT TCACTTRTAATTTTAAGATTC ATAAATGTATAAAGATGTTA CGATATCTAGTCAATGTATA CCAAGAGAT[A/T]TAATTTTT TATAATTCGGCAAACTTTCT ATGAAATTTTTTATTTATTAC ATGGAAGCGATTAAAATTAA TAAATATAATTAGAACACAT TTATCTATTGCTTAATATCAT TCAAATTTAAATTTTATTGT ATTATAAATTTTACGTATTA
6	6863177	C,G	LG6:7217420	intron_variant	MODIFIER	GB48608	143805	1.00	185.99	189.00	324.00	37.00	0.00	1.54	Group6:6863027-6863327	CGGAAAAGATWTCGAAGAG ATTTTCGCAAAGGTTTCTCG ATTCTAAAATTCTACGAGGA TTCTTTAATTAAAAATGTTC GWTAATTATCTTTATTCGGA AAAGATTTCGAAGARATTTT CGCAAAGGTTTCGATCGATT CTAAAATCTAS[G/C]AAGATT CTTTAATTTCAAAATTGTTTG ACAAGTGTTTTCCRGTAAGA GGATGATTCGTCGGATGATT GRACATACAGGAAATACRTA TGGGTTGARAAACGCTGGTT CASCGTTCGGAAACGATAAA CTTGAAATGGATGTACGGTA CGT
6	1.1E+07	C,T	LG6:10559799	intergenic_region	MODIFIER	GB52930-GB52950	150210	1.00	230.48	233.00	423.00	40.00	0.00	13.85	Group6:10923970-10924270	GTGCGGTTACRCCCGCCACT TTTTCATCCCCCGTGTACATT AAAAGAACGCGATGGCCAC GATGGTTTTTCCATCGTTCA AACRTTAACAAATCTCTAAT TACCGAATTCTTTCAACGTC TCTTCTTATTAGGAAATCGT MCTAACGCTT[T/C]TCGTGCT AATTAAACTCGCCGTGACCT GTCTCCYCATTCGAAAATCG ATCATTATTRCGTTTTCGTTT ACGTTTTCCAAGAATGAAAG ATAAAAAAGAACAGAGAGA

																GAGAGAGAGAGAGGAGATT AARAGAAATGGAGAGAAAC AGGAG
6	1.4E+07	C,T	LG6:12949004	synonymous_variant	LOW	GB42979	155072	1.00	106.17	89.00	261.00	14.00	2.92	44.62	Group6:13659894-13660194	ACATGAATATGCAGCTATAT ATTTAACCGCTGGAATCGAA AATCTGTTGGAAGAAATATT ATTGCAATGCATACCGAGTG ATTCTCATACGACTTTAACT GCAACTATGTTGGAACATGC TATTGCAAATAGYGGAGATT TATGGGGTCT[T/C]CTTCAAC CTTATGCACATCTTAATGCT GGTCGAACAGCATCAGGTA AATCTAGAAGCTAATTAATA TTTAATAATTATATATTTGTG ATTACTAATAACTGTTGAAA RATAATATGAATAGATTARA AAAATTTATTTAAAAAAAAT AAA
6	1.5E+07	C,T	LG6:14951382	intergenic_region	MODIFIER	GB53573-GB53574	158015	1.00	246.18	246.00	426.00	60.00	0.00	13.85	Group6:15332259-15332559	ATTTAACTCGCGTATTTACG CGTAAATCGTTATCGAGTTG TCATATMGATTCGTTGATAA TGATAATAAATAATTCGAAA GAAGAAACRCGGTGATTTCC ATCTTCTTGCATGCATCTGA CGTATCGATGGAGAAAAACT TGTAYGATGA[C/T]ATATAA AAAGACTTTATTTCATCTTG ATTACACAAGATACATCTTT TTACACATTTKTGATGCGAA GAAAGGAAATCCAACGAYG TACACGTATTTTTGTATTTTG ATTCACTTTACCTACGTTAA TTATAATTAAAAGATATTAA ACGC
7	8703231	C,T	LG7:11305036	intergenic_region	MODIFIER	GB42408-GB42409	173240	1.00	75.08	75.00	130.00	12.00	0.73	1.54	Group7:8703081-8703381	ACCGAACCAAGTTATATTTT CGTACACTTATTTCCGGGTT GGAAACCGAAACCTATGTG GAGAGATTTTCGTTAACGTT TGATTCGTAGCCGAGCAATT GATTATTTCGTTCTTTCTACT CYGCTCACGATTGCGTAATA TTTGCAATTT[C/T]TATTTCT AACGGAAATTTCGTTTYATT TCTTTYTTTTTTTTAATCTTTT TATTATATCGTAATTTTTAAT TGYAACTTATAACAAATATC GATTCTTTCTTTACTGATTTA TAAGTTGTTCTTGGAATTCG TGGTAATTCGATTGGAAGG
7	8703865	C,T	LG7:11304405	intergenic_region	MODIFIER	GB42408-GB42409	173240	1.00	164.81	175.00	279.00	20.00	0.00	18.46	Group7:8703715-8704015	AAATCTTCTTYAAATGCAGC TCGCATTCCCGATAAATCTA TTTACAACGAGCCGCCGATC TAAACGCGAGCGCATCGATG CCGAACGCATTCTAAATTCG ACGGAATTTCCCGTGGAAAG

																GGATATCGACGTTATCGATA CACCACCTCCYCTCGTACTC TAAAAAGTTACTCGATAYTC TCGATGATTCGTCGGAGCAC GTGAGTTCGTGACGAATCTC TCTTTCGATCTTGCRCCRATC GATCATTTCTTGCCGWGAGA AAGAAATCGTTGATGAAAC GACGAAATTTTATTCTAARR C
7	1.3E+07	C,T	LG7:13690902	intergenic_region	MODIFIER	GB48123-GB48160	180170	1.00	210.04	205.00	414.00	52.00	0.00	3.08	Group7:12723355-12723655	TATCTTAAATTTAATGTTGTT CGTTCGTTTCCTATCGTTTAG GAAGTAATAACAGTACAGA CGTGTGTGTGTGTGTGTATG TGTGTGTGTGTTTATGTGTRT GTCGATCTACTTATCTTTCGT CTGGATCAGTGTTGCCCGAT TGTTTCTYTACCTCGGTGAA AAYGAACSCGRGAAAARAG ATAAAATAAAAAACATATTC GTASTTCGACCTTCCTTTCTC CTYCTTCCAAATATCGATCG ATCTTGACAYGATATAATAA YATTTTTTTATTATAATAATA TTTTTCAAARCAGCGAG
7	1.3E+07	C,G	LG7:13690922	intergenic_region	MODIFIER	GB48123-GB48160	180170	1.00	126.96	121.00	400.00	0.00	21.90	73.85	Group7:12723375-12723675	TCGTTCGTTTCCTATCGTTTA GGAAGTAATAACAGTACAG ACGTGTGTGTGTGTGTGTAT GTGTGTGTGTGTTTATGTGT RTGTCGATCTACTTATCTTTC GTCTGGATCAGTGTTGCCCG ATTGTTTCTYTACCTCGGTG AAAAYGAAC[G/C]CGRGAAA ARAGATAAAATAAAAAACA TATTCGTASTTCGACCTTCCT TTCTCCTYCTTCCAAATATC GATCGATCTTGACAYGATAT AATAAYATTTTTTTATTATA ATAATATTTTTCAAARCAGC GAGCAAAAGGAATTTTTTTT TTT
8	3546391	C,G	LG8:3440882	synonymous_variant	LOW	GB40362	184420	0.99	126.32	121.00	247.00	28.00	0.00	7.69	Group8:3546241-3546541	ATTATAAATGATTCAAATTG AAATATTTTTGTAAACATTG ATATTCATATTTGAAAAAAA TTAAAAAGTTATAATAGTTT ACAATTTTTAAAGAATTATT TTGGTCTTCCTCGTGGTTTTC CTGTATTTTTTTTAGTCTTTT TAGTTGTT[C/G]CTTTTGTTT CTTCTATCTGTAAAAATATA ATATATATATATTAATTTAA GAGAAAAGAGAAAATAATA TAATATTAAAAATTGTTACT TATACTTACTTTACGTTTTGT AGGAGTTGAATTTGGTAATA

															CTTTGAAGAATGTATCCAATC
8	3611478	A,G	LG8:3504757	synonymous_variant	LOW	GB40357	184420	0.99	232.40	220.00	531.00	36.00	0.00	47.69	Group8:3611328-3611628	AATATATAATATATAAATCTTAATTATTATTAAAAAAAAAAATTTTTTWWAAAAATTTTTAAAAAAAAGTTTCTACCTTCCATTATATTTATTCCAATCATAAAGTTTAACAAGGAGGTTTTCACTTAATAAATATTCTAAACAAGGAGARTGTGTTGCTGTTGTAGATTGTCGATAATTATTGTAAAAGTTAAAATCACGTAATTCTAAAACTAAAAGTGTTGCCATCTGATCTACATGATTAACAACTCCAAGAACATCATCTTGATTACATATTCCTTTAGGTGGCTATGACAAAATA
8	9085611	A,G	LG8:8867739	upstream_gene_variant & intergenic_region	MODIFIER	GB52825 & GB52824-GB52825	192952	1.00	38.48	39.00	72.00	7.00	5.84	1.54	Group8:9085461-9085761	ACGTTTTGTTGAAATTTTAATAAATTTATTTATTTAAAAAAATATAAAATTAAATAATTATAAATAATAAAAAAGTATAAAATTATTTAAAGTTTGATCGAGAATATARATACTTAATAAAGGAAAAATGAAATGAAAGAAAAAATAAAARTGAATGAAAATAATTTCTGAGCAAATTTAATTTTCATTTAAATAGTTATAACTAAAAAAAGTTTYGTTCAAAATGACYTTACCATTGTCAAGAATAGATTATGCAACTGTGGGTGTCACATCCAGAAGAACCACCAAAATCTTTCCTGCA
8	9086908	A,G	LG8:8869036	splice_region_variant&intron_variant	LOW	GB52825	192952	1.00	91.57	86.00	200.00	17.00	1.46	12.31	Group8:9086758-9087058	TAAACATGAATGGATCAGAAATTACTTCTTTAGATACTTATGAATTACAAGATGGTCTAGATATACTTATTGGCAGACAGGATGGTACTGTTGAAGTATATACYTTTCCAGAAGAGGATACTTTAGCTTCACTTCGCTATCGTTATGTAA[A/G]AGCATTATATTTAAACATTTAAAGCTTTGATTTAATTTTTTAAATTAGCTTTTTAATATTATTCSCATAACTAAAATTTTTTTTWAAAAATTTATTTATAGAATGCCAGTGAAAGTATCAGTTCAGTGACAGGAGGAATAATTGGAGCAG
8	9087168	A,G	LG8:8869296	synonymous_variant	LOW	GB52825	192952	1.00	236.61	239.00	417.00	47.00	0.00	1.54	Group8:9087018-9087318	TGAAAGTATCAGTTCAGTGACAGGAGGAATAATTGGAGCAGCAGGATATCCTGAAGTRTTRGTTACTACATATTCAGGACGAATATTTGGCTTAACTACTAAACCACCAGGACTTCTGGAAGCCGATCAAAATGACGG

																TTTGACGAAATTRAAATTAG AAATACAACAATTACAAGA AAAGCTKAATGAAGAGAAA GAAGCAGCTATTTTTTCAAC AGATCCTTTAGCACCATTGA TTCTSTCTGTAAATTATAGGT ATAAAATTAATATAAATTAA ATAAATTGAARTCATATAGT ATTA
8	9115014	C,T	LG8:8897200	synonymous_variant	LOW	GB52848	192963	1.00	35.29	31.00	88.00	7.00	13.14	12.31	Group8:9114864-9115164	TTCAATTTTCAATTTTCAAA AATATCATTCAAGGATARTA CARTAGAATTTCTTTATTTTC TAYTTAGCATTAAAAATTAC TCACTTGTGATCATGTTTTTC ATTGTTGTGCCATAARTTRG ACAACGCTTTGTCCAGTAAA CCTCCTAA[C/T]GAAGAGTTC ATTTTAGAGCCCTTCAGCCC TGAAGATTCTAAATTCATTT TTCTATCGTTGTGCTCCTCTT GTGGRTCTCTAYTGTGCAAA GTATTGTAGTCTCCATCCGA CAGATTCGAACMAGCGATG CGTYCTTGATCTTGTAGACT A
9	15736	A,G	LG9:1304541	intron_variant	MODIFIER	GB45547	199195	1.00	0.26	0.00	2.00	0.00	100.00	73.85	Group9:15586-15886	AAATGAAATTTTGAAAAAAT ATAGAATATATAARTTACTT TTTTTAATTAAATTAATTAA AAATTAATAAAAAAATTTAT TAATTATTTAATTTACTAGTT GCTTAAATAATGTTACTTAA AGAAAAATAATATATTTATT TTAATATAA[A/G]TTATAAA AAAAATTTATATTTCTTAAA TTAATTTTTTCTTACGAAATA TTACTTAATGATTAAATTAA AATATTTAGTTTTTATATAAT ATAAATGTAAATACTAYTTT TATATTGTTGAATTGATTTAT CAATAATAAAAGATTTTATT
9	1971514	A,G	LG9:3522286	intron_variant	MODIFIER	GB43759	201244	1.00	160.96	161.00	379.00	19.00	0.73	52.31	Group9:1971364-1971664	ATACGGTGCATCCCACTTTT TTGATTTATGATGGAATTAC TCTCTGTCTCTGTTGTAACGT TCATACCGTCTCGCCTTGTA AATTATATCACATATGCGTA TTTCTCTCTCTCTCGAAAAAT TGTATATCTATATCTATACA TATATATG[A/G]ATTTCTATA AAATGTAGAAATYCATTTTG CCATRGACTAAACAAAAAC GATACCCGATACATYCTGTA TTTGAAAAAAAATCTATATT TAATTTTCATCAATCCAYGG AAAAATCAGATTTGAAAAAT AATAAAATATCGTCCAATAA TA

9	2115514	A,G	LG9:3666311	intron_variant	MODIFIER	GB43758	201595	1.00	174.99	171.00	325.00	27.00	0.00	13.85	Group9:2115364-2115664	AATCAATTCAACTTATCGAT CYGACGTCGATCTCCGAGCA GGAGAGAGARAAAAAAGAA AAAAATTCAATTTTCATCAA TCGAAACAAATAAARGGAA CGATTAAACGAGCCGGTTYC RAGTGGAGAGGGTGTAAAC GAGAGAATCCGCA[A/G]ACC CGTGTCCCGGCAACAACCGG CGACGACCGAGGCAAGAAG TGACCGCAACTTCAATCCCA ATCCGACTTCATTTCCATCC CACTTGAACCCGCGGTCTTC ATGCCTCCTCTTCACGGGCA CGCCCTTCCTGCTCCACCGG CACGACGT
9	3785070	A,G	LG9:5263665	intergenic_region	MODIFIER	GB44856-GB44844	204819	1.00	221.18	217.00	575.00	15.00	0.73	66.15	Group9:3784920-3785220	TGATCATACATTAYCACCGT TCTTTTAARCGGRGGAGGGG AGGAAAAARATATATCTCCC TTTATATCTCGCTCTTTTYCC TSTCGTTTCGCTSTTTCTTTTT CATTTGRAAATTGAATTTTTT AATCGTCGATATAAACRACG GCAAAA[A/G]ATAAATATTT GACGACAAAGGGATATTCTT TCGCGTCGTGAATTATAATG CAACAATTTGTTGGAAGAGG CGAAATAAACGAATTACAA CTTCAAAGCCRGTTGATCAT CGTTTACACGGATGATCGAT CAACTACGAAGAAACGTAR AA
9	5053380	A,G	LG9:6431269	intergenic_region	MODIFIER	GB42732-GB42731	207068	1.00	83.31	84.00	186.00	9.00	4.38	47.69	Group9:5053230-5053530	TATAACCTTGTTACTGTAGT ACTACATACGATCAATGTTT GACTTCAATGYGCCACATAT CTTACAATATGATCGCGCAT TCTTTACATTGATAATGCAG ATTAATAACTTCAATAAAAA CGCAATATTTTCTTAATCATT TTTAATTAT[A/G]AAAAAGA AAAATGATCTAATTCTTATG AGTTTTTCGTAAAAGAAAAA AGTAAAATRTAACATTACCA TTAAGATAAGAGAGTGATTT TATCGATCATATTGATCTCT ACTGTAGTTAATATCAATAA AAGTTTTGACAAATATTAGA ATA
9	5454899	A,G	LG9:6841250	intron_variant	MODIFIER	GB42708	207594	1.00	329.64	312.00	730.00	63.00	0.00	52.31	Group9:5454749-5455049	GAATAAATCCAAAGACGTCT GGCATGTTTAAAGAATTACG ACAACTGCAACACTGGAAG ATCGATATTTTTTTCTTTTTT TTTATCTTTTCTCTCTATTTA ATATCTAMRGATCACGAATT AAACGAACTTTCGACAAACG AGGAAACTCRAGGAAAACC

																ATTTGTAGACTCGTGATACA CAGATATCGGATGACAAAGT AGATCGCRCACGATTTAATC TTCCTCTGGCTGGAAAGAAA TGTATATATACTCGATGTAC GCGGCGCGACTTCATATCGA GCAATTTTCACTGGGTATTA C
9	5456771	A,G	LG9:6843119	synonymous_variant	LOW	GB42707	207594	1.00	176.93	167.00	363.00	23.00	0.00	18.46	Group9:5456621-5456921	ATGATAACATTCCTCCAAGT CTCACTGTAAAGATTAAATA AATTAATAATAATATTTCTT ATTATATATTAATATATTAT CATAAAATATAATTACAAAT AGTACCAAACAGCCTTAATA AATGAAATGCACCATATAGT TGACTTGGCA[G/A]TGTTTCA GGATTATCATTCATAATTTG AATAAATTGTGGACGTTCCC ATTTATATAATAATTGTAAA CCTAATGAAATATTGAAATA TTCGCGAATACCTTTAGTTA TTTCTAATGTACTTTCCCTAA AAATAGATAAAAATATTAGT TT
9	1.1E+07	C,T	LG9:10984145	splice_region_variant&intron_variant	LOW	GB53305	217286	1.00	42.33	43.00	82.00	5.00	4.38	3.08	Group9:10610437-10610737	CGATTCTATCGCTTTCTCTAA GCGAAATATTTCTTCCTGCA ATAAATGAAGGGTTCCAGTT TTACCACGATGACGCGCAAA CGCGGCCGCACAAGCCGAA TGTATACTGTTCACCCAGTT CTCGAGTTCCACTTGACACG GAGCCTAATG[C/T]AAAAGT GAAAAAAATTCCGTCACTGA ATTTGTAACTGATAATGAAA TAATTATTATCGCCGATAAA TAAATAAAAGTTTTCACTTC GATGTGATAAATGTATTTAW TTCATTATTTYTTTGATCGCG ATRAATTGAAGATAGAATRG TTG
9	1.1E+07	C,T	LG9:10948741	3_prime_UTR_variant	MODIFIER	GB53433	217322	1.00	229.10	221.00	389.00	65.00	0.00	13.85	Group9:10645901-10646201	TTTCGAAAAGATACGCACAA GCTGCCAAAAAAGGATTGG AATTAGTACGAAAGGAACG TGAACAAAGAGTTGGTAAAT TAACGGATGCTAATAGTGTA GAAAAAGTTTTAAATGTGGC TCGTTTAAACTGACGTAACG AGGGATCGGACT[C/T]TGAT CTCCACAATTTCTCGAATTT CAATTTTTCCTTAACAGAAT AATATTTAATATGTCTCGAA TTTCAATTTTTAATTAATAG AATAATATTTAATATATCCA TAAATTAACAATCTTTTTTA GATAATCTTTTTTAACGATA ATCTCT

9	1.1E+07	A,T	LG9:10890802	intergenic_region	MODIFIER	GB53297-GB53296	217341	1.00	42.22	38.00	150.00	0.00	32.12	73.85	Group9:10703717-10704017	ATTCGACGTAACAGTTATAC AATRTACATAATATTTGCTT ATGATGATCCATYTGTAATA CTGCACAYCCCTCAATATAT TTTTCCTCCGYTCCATCTGGT CTGGTCACTCTGGTCGCGAC ATCATCATAGTCTCTTGGAA ACTGATTTW[A/T]TTCCGTTC CAATCTCTCAATTATCATCG TATTATCTTACAATCTCACA AATCTGCTTTCGGCGTTACA TCGTTGAGAAGAATTTGTTC GTCAATCGTTTTTTGCGACA TAACACGCTTTTGGTCGATG ACATTTTATWGTCGTTCATA TT
10	164597	A,G	LG10:669552	intergenic_region	MODIFIER	GB43337-GB43334	218485	1.00	139.56	135.00	464.00	0.00	23.36	73.85	Group10:164447-164747	ACCTGTTAATGGTRTGCACA CCCGGGGGATCGATTTTCGT GAGTCTTCCATTAAAACTTG TGCCGGGATGCGCGTCTGGA TCTCGTAATCTACCACGTGT CTGGTCAAGAAAGGGGGTG RTTTCGTCTGATGGAATAGA GMCCAGRTTTCRARCCYAGG TRGAACGTTGTCGACGCGTG TCAACTATTCAGATTAGTCT GGMAYCCTTMGTTAATTASC AAAAGAAAACTGGCCAGAA GAAAAGAAGAAAAAGTART AYATAATTATTGAGAATYGA TATTTTCGTTTCACAATTTTG AAT
10	164648	A,C	LG10:669603	intergenic_region	MODIFIER	GB43337-GB43334	218485	1.00	145.03	139.00	454.00	0.00	22.63	73.85	Group10:164498-164798	TAAAACTTGTGCCGGGATGC GCGTCTGGATCTCGTAATCT ACCACGTGTCTGGTCAAGAA AGGGGGTGRTTTCGTCTGAT GGAATAGAGMCCAGRTTTC RARCCYAGGTRGAACGTTGT CGACGCGTGTCAACTATTCA GATTAGTCTGG[A/C]AYCCTT MGTTAATTASCAAAAGAAA ACTGGCCAGAAGAAAAGAA GAAAAAGTARTAYATAATTA TTGAGAATYGATATTTTCGT TTCACAATTTTGAATTTTATA TGTATATCGAAGAAATATGA GGATCGAATAATRTTTTAAA TTATT
10	6115851	C,T	LG10:6873071	intron_variant	MODIFIER	GB48386	226398	1.00	132.90	114.00	442.00	0.00	26.28	69.23	Group10:6115701-6116001	AAAGAGTTTCGTTCGAAAAG GTATTTATTCTTTTCCAGTAA TCGTATCGATCAGACYTTAC CTCAAAGATATTTTAATTAT CGCTCTAACGATATAAAATA ATTTTATTCGATAACAGTAA TACGAAATTATCCTCGCTCC YCTCTCCTC[T/C]CCCTCCTC

																CTCCTCTTTTTTCTCTTTTCC AAYTGGGAAATTAATAGCG GAACGAAATATAGCGAGAC GAAAATCACGTTGGGTCAAT CAAAGGGATATCGCCAGCGT TATTAATTACGAAATAACGA ACTCGTTTACACCCAGGCAG ACA
10	7933755	A,G	LG10:7424253	intron_variant	MODIFIER	GB54295	228810	1.00	442.28	435.00	962.00	58.00	0.00	26.15	Group10:7933605-7933905	GGAGAGAATTCGGATTCCAC CCGTGGAGGAAGATACATCC AAGGGTAATGGCTCGAGAG CCGCTCGCTACGAGTACTCA CTTTCGGACAACGTGACTTG GGTTCAGCCTCCTTGTTTCTT TTTTTTTYTTTTTTTTTTTGGA GGGAGGGG[A/G]AGGGTTTG ATGGCTGCTCGTTTCCAGCG TACCTCGAGTATTGTTGTTG CGAGTAGCTCGATAATGAGC ATCGATGAGATTTTTTYYCY TCCCCTCTTTCGGAGAAGGG GAATTCGTTCTTTCGAGACA AAAAAGGAAATTGGTCGGT GTC
10	7934803	A,G	LG10:7423214	intron_variant	MODIFIER	GB54295	228813	1.00	382.00	382.00	685.00	56.00	0.00	1.54	Group10:7934653-7934953	TCTTTGCGATGAAACGAAAA AAAAAGAAAGAAAGAAAAA AGRAAACGATATCCTTTCGC AAGTTCCGGTTAAAAATCGC GAAGAGGAAACGAATCAGA AACCGCTGATAAGATTCATT AATCGAAACATCGAGTGTAC GAGATGRGGGTG[A/G]GAGA AGAGGGAGGGGAAGAGAAA GATCGGGGTTAAGCGAGGG ATGAGGGAGGGTGAGAAGG TAATCGACCGGCATCTTGGA CGGATATCGAGGTGGCGTTG ATTCACTGGTGGCGAGRAAA TAAGAGAAAAATGAAAAGA ATTGAAATAA
10	8168036	C,T	LG10:7171291	intron_variant	MODIFIER	GB54295	229135	1.00	99.40	103.00	173.00	18.00	0.73	1.54	Group10:8167886-8168186	GAAAGTTTACCCCCTCCTYC CTCCCCCTACTACCACCGCC AAGCCTTTTTGGCCGGGATT ACGGGCYTTGGACGGACATC CATTCTCGTTTAACGCCTTTA TTTACCGTCCACTGCTTTTCC TTTCAACCCCTCCCGTCCCC ACCCCCCT[C/T]TCACCCACC CTTCCTGACCCCTTCTTGACT CCGCCAACTCCTTCGCCCCT CCTCTTCAGYTCTCCTCCATT CTCCTGCCGCTTCTTCYGTTC TTTCGTTCGTTGACACTCGA TTATCGCGGTGGTTCGAAGA TCCCAATCGCTCCATCCC

10	8884847	G,T	LG10:11035996	upstream_gene_variant & intergenic_region	MODIFIER	GB50874 & GB51176-GB50874	229901	1.00	11.13	11.00	23.00	1.00	93.43	27.69	Group10:8884697-8884997	TTCTTTCCATTCCCTCGACAC AATAACAAACCTACCRAACT ATACTATATATATATATATG TATATATATATATACACAAA CTATGCCTATTGGAGGATAG ATAGGGAGATGGAGAAACG CATCGAGCAATATAAATAAT AATAAGAATC[G/T]ACCCGT TGTTCCAGACGATGTATACT GACGGTCTCCTTACACTCCA CTATGGGAAGCATCCGGCGA CCTTGGCCGCGGAGGTCGGA GCCTGGTACACCGGGGACCG TCACGTGGACGTGACGTTGG CATGCGACGATGGGTCGGTA GTGA
10	1E+07	A,T	LG10:9422404	synonymous_variant	LOW	GB51072	231674	0.99	250.60	233.00	520.00	51.00	0.00	24.62	Group10:10499687-10499987	TTTTTTATAAAAATTTAAAA TTAAAAATAATAAATATAAT AATATATATAAAACAATAAA TAAAACAGTAAATAAAAAT AAAAATGCAAAATACCTTGT AGCAAAAGTATGCCCTTGTG AATCTTTAATATGACAAATA TCGAAAGATCC[T/A]GGATG ACGTTCGCGACTAACGACTG TACCTACACGTCCTAAATTT CTTCCACCAGTAATCATACA GAGATTACCTAAAAAAAAA TTATTTAAATGTAATTATTTA TATAATCATTTAAAAAAAAT TTTTATATTTATATTTAGCAT TATT
10	1.2E+07	A,G	LG10:11605217	intron_variant	MODIFIER	GB48760	234177	1.00	145.53	140.00	367.00	12.00	2.19	50.77	Group10:12157576-12157876	TCCGTCCTTCTCCTTTTTCCG TCCCCTCGTTTCGTCTCCTCT TCCTTTTTTTTTTCTTCACCG AGRAGCCTTCCCTCGGCTAA AARTAAAAACTTAGTAGCCC CTCCCTTACCTGAAATCATC CCCTSTGTGTATACCTGTGTG TCGCGC[A/G]GCGTTTTTCTC GAACCGATCCATGTCCTCTC CAAGTCTTCCAATTTTATTTC GTTCCCTTGGAATAATAATA TTGCGAGATCGGGCRTCTTT TTTTATTTTTTCGATCGGGCG AAATCGCGTGAGAGAGAGA GAGAGAGAAAGGGRGAYA
11	1468594	A,G	LG11:1293696	synonymous_variant	LOW	GB52239	237033	1.00	152.91	135.00	306.00	14.00	1.46	1.54	Group11:1468444-1468744	AGCTGGAAGTAGCCAGCCA GGCCYTGGCCGGCCTGGCGC ACTTCATCCGGCTGCAGCAA CGTTGTCGTGTCCGAGGTGA TCACGCTYACRGGTTGGTAA GAYGCTTCCTGGGTGATCAT CACGAGGCTTCTGTACTTGT TGTTTGCCGTA[G/A]CTGGCG AGCACTGATGAGCGGGGCC

																GTTGTYTCTGATGTTGAGCG AATCCTTGRTCGAAGCGGGG GATCCATTCTCCTTTCCAAT GGACGGGCGTCGCTCGACTG AGAGAAACCGTGTCGAAAC CCCTTCTATCCCCTCCWCCG CGCCGC
11	1778090	C,T	LG11:1483474	intergenic_region	MODIFIER	GB55189-GB55188	237378	1.00	26.20	20.00	77.00	0.00	47.45	63.08	Group11:1777940-1778240	TAYACACGCCTCGTACYGTC CCGTGCCGTTCACTTTAACC CTTCGCACCCTTGGGGTGAA AACAGTAACGTAATTRAACG GGGTGCAAGCACGCATTTAT MGATCGATCGAAATTGTAA AAAYAATAAATATAATATA AAAAAATATTTT[C/T]GMAT GTAAGAAAAATTAATATGTA AAGRTAYGATAATTATATTT AAAAAAAATTAAAATAAYA ATATTTTTAAAGATATTYAA TGACYAAAAGAATGTGTATY RTGAATCGTTCAATTTCTAA GAAATAAATTGACTCGCTAG ATTGTAA
11	2936665	C,T	LG11:2634270	missense_variant	MODERATE	GB55146	239096	1.00	91.26	85.00	211.00	11.00	2.19	23.08	Group11:2936515-2936815	TTTGAAAGATTTATCATGAA TCCTTCTTGTCGATGTTCTTC AATGCTTGCATTAATAGATT TTCCATTCTGTGATTCTAYAT ATTAAATAATAATAATAACA TTTTTTTCATAATAATTCTAT TTTTAAAGATCAAATATTAC CTTTCATYATACTTTGTGTTA TTTCATTTTCTAATAAAAATT CTGGAACTTCAGGTGGTGTA AATACTGGTGGTGCAGAGCT CGGTGGTGTTATATTCTCCT GATAAAAAAAAAATAAAGA TAAAAATGAGAATTATAGA ATTCAAACAAGCAAAGCA
11	2937477	C,T	LG11:2635082	missense_variant	MODERATE	GB55146	239096	1.00	228.98	228.00	409.00	48.00	0.00	1.54	Group11:2937327-2937627	ATTTACCTCCTATATGATTTT GTACGATAGTTTCYGATGTT GTCATTTGTAAATTGGCATC TGGATCTACTGCATCATGAT CATCGATAGTGTAATCTACT GAGGAAAAGGATATTTGTTG TACCAAACCTGGTCCTTGAG AAGTAAGAC[C/T]GGATGTT GTGTCTCTTTCAACATCAAG ATAATTTTGAGGAACAAAGC CTTCTTCACCACGATAATTT CGTGCTCGGAGCCATCCATC ACCATCACCTTCACCAACAA CTTCTAATTGCTCACTTTCGA CAATTGACAATTCATCAGGA TT
11	2983919	C,T	LG11:2681510	missense_variant	MODERATE	GB55142	239129	1.00	3.52	3.00	15.00	0.00	100.00	60.00	Group11:2983769-2984069	GGACGAAAATTATGAGACC ACTGYGAAATACAATGAGCT

																TCGTCTATGCATGCAAAAGC AATTGGTGGTAGCTCCCTTA ACAAYGCTCCAAATCCTGTT GATTTTTCTCCAGCTACTAC GGCTTCTGGAGAGATTAGCA AAATATTTATT[T/C]TYCCTT GTTTGATCATTTTCATAGTAT TATCTCTTACTTTTTGAGTTT GATTAGTATGCAAACAAGCT GCYGATAAAAAGGACGGAA CACCAGTTACTTGATCATCC ATTAAGGATACTAATGGTGA TATGACTAAAGTGATACAAG TGG
11	4155603	A,G	LG11:3713110	intergenic_region	MODIFIER	GB42097- GB42095	240451	1.00	279.47	266.00	670.00	0.00	3.65	56.92	Group11:4155453- 4155753	AAAAATAGAAAATTAAATTT TTTTTCATTCCTTTTYATTGC GAAAACCTTTATAAATTAAA ATAAAATTTATTTTAATTCG AATTATTTGCTATTTGTTGTT TTSATTTGCTAAAATGATAG ATTTATCAAGATATAAATGC YGTCTGAYRAAATRCTTAAA ACAAGAATTTTCARATGCAG TTTAATTATACGAATATAAT AAATTTATATTTAGTTTCAG TAAGTTATGTCAAATTAATT GTTATTTTTATTTGAATAAA ARTAGCTAAAATAAAAAGTT ATTTGAAAACAATTTATTT
11	4763028	C,T	LG11:6317800	splice_region_variant&synonymous_variant	LOW	GB55058	240480	1.00	196.28	199.00	336.00	42.00	0.00	1.54	Group11:4762878- 4763178	NNNNNNNNNNNNNCTCTT CCTATGCTGGGTAACGTGCT GCTCCTATGTTTCTTCGTCTT CTTTATATTTGGTATTGTCGG CGTGCAACTGTGGGAGGGC ATACTGCGTCAACGGTGCTT CCTGAAAGCTCTACCCAACG TCAAATATCC[T/C]GAGTAA GTGTCACATGAATCGATTCG TTTTYTGTCATGCTGTCATTT ATTGGTTCTTTAGATATATA GTTATATTTTTTAAGACGGT CAGAGTTCGTYTGATTCGTT TATTATTATGCTTTATTTCGC ATCGATTCCGCATTCTCTTTC T
11	4904280	A,G	LG11:6171384	splice_region_variant&synonymous_variant	LOW	GB55054	240480	1.00	156.67	150.00	336.00	27.00	0.00	27.69	Group11:4904130- 4904430	AGTGAGTCCCGCTTCTCTAT CACCTTCGAAACCCAAAAAG TGGATCACTTTTAAAAGAGA AGGTGGTAATAAACTGACAC ATAGCTGATAAATTCCATAA CCGAAACTAACTGCAGACAT YAATCGGGAGACCTCATTAG GTTCAACGAARCTGCAAAAT ATGTCACATCGATTTCCART TCTGTCGATTTTCAATAAGT AAATATATTTGTTATTAATA

																CAAATAACAATATAATATTT TAAAAAGAAATAATTATTTT AAATAATTAATACTTATGCA ATAAAATGTACATATATTAG C
11	5159160	A,G	LG11:7190739	synonymous_variant	LOW	GB54035	240480	1.00	85.15	84.00	163.00	21.00	0.00	1.54	Group11:5159010-5159310	TTTTTTATAATTATTAATCAT CATTCGGTGCTCTAAATTAA AGCCATAYATTGTAAAATAC TGAATATTATTTCATAATAA TCTATATAAGCAAACTCACC GAAGTTTCACGTAATCCGAT GAAGRTTGCATTTCGGTATA TACTGATTA[A/G]TATCCAAG TGCCAACCTTTTGTTAATGC ACAGGGATAATGTATAACA GCTTTACAACCACGTTTAAC GCAACCAATATTTGCTCCTG TTAAACCACAAGAGTCACAT ATCGATTTTGCTGCATCCCA CACAGCTTCTTGCAAACCTG TTA
11	5187989	A,G	LG11:7220232	synonymous_variant	LOW	GB54054	240480	1.00	239.80	227.00	522.00	36.00	0.00	40.00	Group11:5187839-5188139	AAATGAGAGAGCTGGTTTTC CACCTGATACGGAATTAGCT CTTTACGAAGAAATTAAACC AAATTTGGTAGAAAAAATA GATAATCTAACAGAACCATT AGAAAAAGTTCTTGAAGAAT TAATGGAYGGAGATATTATT GTTTTTCAAAARGAAGGAGA CAATCAAATGTATGAGCTTC CAACATGTAGAGAATATTTT AAGTAAGTTAGAATTRTTTA GAATTTTTTARAATTTGAAA TTTAATATTTTTCACAAAAA AAAATCTAATATGAAAATTT TATATTTTAGAGATCTATTTC A
11	7543229	C,T	LG11:9396017	intron_variant	MODIFIER	GB47257	243455	1.00	0.64	0.00	4.00	0.00	100.00	70.77	Group11:7543079-7543379	AAAATCTAATGGATGAAAAT TTATGAWTTCAATTATAATA TGATTCTAAGARGTYCAAAA TTTTATTAGATTCTGTATCAT TGCTGCATTTAAAGATAATW ATTATTATTATTATTACACAT TATATAATTTAAAAAATTAA TTAACRTA[C/T]TATAATTTA ATTTATATTACAATTCACAT ATAAATTCTAAAATATTCAA TCTTTTTTATAAAATTATTTT ATAAAATTTTAATTTTTCCA RATATATTTTTTTTTTTATTC TAATAAGCAAAATGATTGTA AAAAGTGATTTTTGCAAAA
11	8912535	C,T	LG11:11114556	intergenic_region	MODIFIER	GB47197-GB45090	245407	1.00	116.02	96.00	378.00	0.00	11.68	67.69	Group11:8912385-8912685	AGAAGATGAATAATCGACC GTTTGCCGAGGTCGTGGAAA AATATCCTTATCTATGTGCA CGATGATTTCGATTACGGGA

																TTTCTATATACAATATYATT GCCGATATAGTGATTGATAA ATTGAGACTAATTGTTCGAA GCATGGTTAGG[C/T]TAATTG AGCAATTTCTAAATTAATAA ATYCGAGAATATTTGAATA WATCTCGTTTYGCCGAGGA AAAAATTCTTTAACGAGAGA GAGAGAGAGAGAGAGGGAG AGGGAGGGAGAGGGAGAGA GAAGGCAACGCAAAAAACT TTCAATTCG
11	8917812	C,T	LG11:11119849	intergenic_region	MODIFIER	GB45090-GB45089	245420	0.99	131.07	125.00	360.00	6.00	8.76	64.62	Group11:8917662-8917962	AAAATTTATCAAAAAAAAA AAAATCCAATTTTCAGAAAT ATGATTTCCAACAAGATGCT ATTTGTTTATTAAATTCGTA AGAATTTTTTCTTTAACATTC AATCGATCAAAATTTCAAAA TTTAATCGTTTCTTTAAAAA ATTTACTTTT[T/C]TAYTAAA AAAATTCAATTCCAGAAAAG ACATGATTTCCAACGATGCT ATTTGTTTATTAAATTTGTAA AAATTTTTTCTTTAACATTCA ATCCATCAAAATTTCAAAAT TTAATCGTTTCTTTAAAAAA TTTACTTTTCTATTAAAAAA A
11	1.1E+07	A,G	LG11:12343894	intergenic_region	MODIFIER	GB45130-GB45131	248895	1.00	230.50	217.00	584.00	0.00	2.92	60.00	Group11:10588964-10589264	TCGATCGCGATTTTGACAGT TTTTCTTTTAATATAGATARA GTTTTATTATGYGATAATAT TATATAAACGATTCTTTCTTT TTGTTTTTCGATAATCCAATT CTTTGAATTGGAATGGAGTT TCGAAATTTCATGTGRATAG TTAAYTG[A/G]TCGCYTATGR AAAATTCCAACTAATCTCTC GYRTRGAAATARGAAACAY GCGCAAGTAAYAATGATAT AACTTTCTTTGACACGTTTCT CATCYTGAAYCGTTTATTTA TTACGGTGAATGATCCAGRC GATTCTTGATCRCCTCGATTT
11	1.5E+07	A,G	LG11:15471185	intron_variant	MODIFIER	GB43173	253884	1.00	164.72	164.00	272.00	40.00	0.00	1.54	Group11:14629568-14629868	CATTGATCAYTTTTCGTAGC ATACGGTCCAACCAAATTCT TCGGATCTTTAATAGCCGTC CAATTTTTATTTTTCACATTG TCGCATATCTATGAAAATGT AAACGTAAAAGGAAGAACA TTCAGTTTAAACTTCGAAYG TAAAAATAAC[A/G]TCGCAA TTTACATAAAARAATTTACG ATRGTAATTTGCATTAAATA CTTACTAAGTGTGTTATATA ATCATTTATGTATAGTACAT ACTATATATAAAGTGATTTA

																CTTATCAATCGATGAATTCA TACCTCGTAATAAGCAAGCA TGCC
12	1660591	A,G	LG12:822757	intron_variant	MODIFIER	GB40187	256375	0.93	164.88	165.00	403.00	20.00	0.00	41.54	Group12:1660441- 1660741	AAATGTGATAAAAATATATC RTTGAAATTTACAAATTTTA WWTTATTTTATTTTATTATA TTGAATTTATGAATCAGGAT TGTAATTCTGATATTATTGA CAATTGACTRTTTTAAATTA GCATCRAATATCARATACTA TTCTGTACAT[G/A]GACTAGG ATTACACCTACTTTAGTATA YGATTCCAACGACACAGGG AAAAATCCATTCGCGGTCAG ACAACAAACTCGTCTGGATC TCGTAATCACCATWATCAGA TCTTTTCGAAGCATTCTCCC GAAYTTGTCCATCGGCACGA TCGT
12	3713814	A,G	LG12:3751697	intergenic_region	MODIFIER	GB48920- GB53610	258012	1.00	39.67	41.00	74.00	5.00	6.57	1.54	Group12:3713664- 3713964	AATYTAAATTATTTTTTTTAG TTATRAGAAAAATTTATCTY GAATTTTTCAGATTTWSCRT ASAAAAGATATAMTGATGTT TTATATCGATTTTTTGATTTT TTAWATATATTTCTAAATGA GTTGTTTATCTTTTTARTTTT TCTTTTTRTTCTATATAATTY TATATAATATTTTTATTAGTT TTATATTATAAATTTRTTAAT TTTTTAAAATAAATATGTAA AAAAATATTTRYGTYATAAA TTAATATTAAAAATTTTTGC AAAATAATATTCTTYTTTAA CRTATTCYAAATCTA
12	3714188	A,G	LG12:3752071	intergenic_region	MODIFIER	GB48920- GB53610	258012	1.00	120.94	119.00	293.00	0.00	18.25	73.85	Group12:3714038- 3714338	ATAAAATYAATGAAAATAA TTTTTATATATARTTTATGAT ATTTTAAAAATTTGATTAAT TTTTTTTAATAATTTAAATAT CAATATGAAAAGAAACTTTY RTTCATTATTAAAAATYGTG AWAATAAAAATGAAKTTTG TTTAYTTTTC[G/A]YTTYATC TTAAWATATATTTAAARTAA ATTATCTAAGATTATAAATT ATTTTTGAAATTAATTATAT CRAAAAAAKTTTTACACAAA TTGAATAGTTTCGAAAMAA ARATAATTTATAAATGAAAT TTTACYAAATATAATCATAA YCGA
12	3714356	C,T	LG12:3752239	intergenic_region	MODIFIER	GB48920- GB53610	258012	1.00	119.20	121.00	291.00	0.00	18.98	73.85	Group12:3714206- 3714506	TTTAAARTAAATTATCTAAG ATTATAAATTATTTTTGAAA TTAATTATATCRAAAAAAKT TTTACACAAATTGAATAGTT TCGAAAMAAARATAATTTAT AAATGAAATTTTACYAAATA

																TAATCATAAYCGATTGCAAT TYATTAATTCYTTCAATGAA ATATTAATTATTTATATTAA AAAATCATTAATTCAGAAGA TATTTTAATTTGAATAAAAT TTGATTAYAAAAAACAAGG AACGTAAAATTGAYGAGTAT CGCTTTGTATATTTYTTATCT TTTATTTAATAAGTTTTATAA
12	9259013	C,T	LG12:11289860	intergenic_region	MODIFIER	GB52041-GB52042	267814	1.00	159.91	154.00	338.00	22.00	0.00	16.92	Group12:9258863-9259163	TTATGCACGAGACGGATCGA ACGAARTTGATAGACGCGA GTGAAAATTGGGAAATTGGT TTTTCTTTTCAAGCGTCTTTT AAGTAATAAAAAAAAGAAA AGAACGATTTTTGATTTCTG ATCGAAAGATATTTTATAAA TAAATGGGAAG[C/T]GATAA TTCGTAGCGTAAACGAAGCG GGAGATTAATTTGACGAACG GATTCAAGTCGAACGAAMC TTTGCGACTAATGCCGTTAA ATTGATCGGAGTGGATTCGA CTAAATCGAACGACGYTTTC GAATAAATGCGATTTCSATG CGATTT
13	2341902	G,T	LG13:3965179	splice_region_variant & intron_variant & downstream_gene_variant	LOW & MODIFIER	GB50671 & GB50679	276376	1.00	70.60	64.00	221.00	0.00	29.93	73.85	Group13:2341752-2342052	ATGGATTGCGTTGTTTGTAA ATGCATTCGAATTTACCACG AGCAGCTTTCTTATTTTTACG ACGATCTAAAACCACTTWA AAATAATTGTTTTKTTAAAA TAATTTTTCTCTATATTTTTA AGAAAATTCGARAAAAAAA TTTGATTKTA[T/G]CTACCTT TGATAGTATCTTCTATTATTT CTGGATAAAGTTTTTCGGTA ACGCTGCTACTTGGAGGATG CAATCTTGGATTTGTATTAA TTTCAATTAACCAGACTGAT AGATCATCCATTATTAAAAA ATCAGCACCGTATAATTGAA AG
13	6167082	C,T	LG13:7090606	intergenic_region	MODIFIER	GB53867-GB53868	283061	1.00	107.61	108.00	242.00	9.00	4.38	58.46	Group13:6166932-6167232	AGCGGCCARCCTCCTCGSCC CTTTGCAGACGAATCGAATC GATTCGATTCGATTCGTGGC GGGATGGCAACCTAGGATTA AATATCCAYCGCTTCCTCYC TGACGGATTCGTCGACGATA TGGATAAATYTATTCTAGCT TTCTGTGGAA[T/C]CGTACCA ATTTGATCAATCTYTTAATC GTCGATAATAACGATAATYC CTTCTCCTTTTTTTAAAAATA CGATTTTTGTGTAGARAATC GTGAAAAATTCCATYCGTAA ATCGTACGATATCTYGACTA

															GAGATGGAGARAAAMTGTG ATA
13	8773197	C,T	LG13:9904423	intergenic_region	MODIFIER	GB49915-GB49865	287712	1.00	170.31	178.00	401.00	18.00	0.73	33.85	Group13:8773047-8773347	AATTGCTTACGTTATTAGTT GGTCATAYGAAGTCATATTT YWTYCTATGACCACTTCAAA ATTGAAATTTATTACTATTT AATTTCCATTAATGTTTGCG GATAATCGAWGAAATYTTA AACTTCTTCTTTTTTGTTTTT ATTAAGCAAY[T/C]ACGTTA CAACTACGTTTRTAYRTRTA ATTTTTTTTAYTTTGCCCAAT TTTTAWTCGAGCAAAGTTTA TYTCAAARTTATCTCAAAAR TATCATTCAAATCGAAATAT TTTCTCAATATCCTATCATA AATYAAATCTTTAGARATTA ATT
13	1E+07	A,G	LG13:11223190	intron_variant	MODIFIER	GB41542 & GB41544	289465	1.00	244.55	220.00	784.00	26.00	0.00	47.69	Group13:10238366-10238666	TGAAATTTTAATCGATAAGA ATGAGATTCGAACAACTCGA AGTCTATCAATTTACTCGAT TCGATCAAAAGTTGGCGAAA CATGRCTTTAACGAGAACGA AATGTTATACTTAAACAAAG TACATAATTTATTTGGTTAG TTTAATATCGRTATATTGAC TTCGATATAGTTTTGCAATT ACCGTTTCTCGGAGAAAATT GGAAACACCGTTCACGTGTA AATRTATTCGAATACACTTA CCGGAAATTTGCTGCTATGA GTCTGCAAATTAGCYGGCTG TTGAGTTTGACTCGAGCTCT G
13	1E+07	C,T	LG13:11225910	intron_variant	MODIFIER	GB41542	289468	1.00	142.93	140.00	318.00	27.00	0.00	1.54	Group13:10241089-10241389	GGATCGGGATAAGAAATGG AGAAGGACAAACGTTCAGG GATAATTTATTCCGCGTCCA TACGCAGACGGGCTCCGCGT CCCACGAATATCAAGCTGCT CGTTATACCTTCCTTCTACTT ACTTCGATCATCTTTCGTTTC TACACATCTA[C/T]TTACGTT TATCTAATTTTTTTCTCTCGA TATTTTTCCCTTCTTTTTYTC CTTTTTCTTACGCGAGATTTT AACWCAACCCAACTTTCGA ATAATTTGATGCGTAYGAAC AATTGTTGATATATTGTTGA TYGATTCTAAAGAAGACGA AA
14	1794663	A,G	LG14:2397742	synonymous_variant	LOW	GB48483	292003	0.99	224.42	221.00	440.00	33.00	0.00	18.46	Group14:1794513-1794813	TCTCCCATTGATGAATCTAA ACGTTTTGGATTTATCATTTA ATTACATAGACTCGACACCA GAYAAACAATTTCAATATTT GAAAGATTTGAAGATTCTGT TGCTCGTGAACGATTCRCTG

																ACCTCTATGCCTAACGTCAA GTTGAATCT[A/G]TTGAGAG AACTTGACGTGTCAGGAAAT CCAATAGAGGTATTGACATA TCTATTTACGTGTTTTTGCAA TATTTTATTTAAATTATTAAT TACAAAGGACGAAAGMTAC ATCTAATTGAATTRAATTGA AAAATGTAGGTAYTGATGA AGG
14	2269579	C,G	LG14:2998714	synonymous_variant	LOW	GB43781	292355	0.99	186.29	185.00	373.00	31.00	0.00	12.31	Group14:2269429-2269729	ATTATAAATACCAAGTCTTA CAGTGGTGTACGTTGCCTGT CGTAAGAGACCAGCGCTAA GACCTGAATAAAACTTCAAA ATTGACTCTTCTCTGTATATG GAAGCAATTACATTAAGGAT CGAAGTTTTTTCCTTTTGTAC TTGGATCCG[C/G]GTTTTTAT YACATCCATGGGATGAACTA CACAGGTTGCTGCCATACTA TAATTTTTTATATGTATTGTA TTATACATATAATTATAGTA AAATCAAACAAATAGTAAA ATCTCTTTAAATGTAAGTAA TTARATAATTAAATTTACAT TT
14	3394800	A,G	LG14:3887186	intron_variant	MODIFIER	GB43598 & GB43596	294576	1.00	195.12	197.00	325.00	42.00	0.00	13.85	Group14:3394650-3394950	CGAATTTATAATTCAATCGC GAATCGTTTCATTCGCTAAA GAGATTGAAATTAYTTTCTT TTGCGACACGGAAATGAATT AARTGGAGYAAAATTGTAY AGGCAACGAGAACGAGGGA GCGCAACGCGGTRTCGAGG AGAGAGAGAAAGA[G/A]AG GCGCATCGAGTGGCAAAGA CAGACTTTACGCAATGCATA AAACAGAGCAYGGTTTTGCG ACACATAAAGGCAAAAGTG TCATAGTGTAGGCGGCCAAC GATGATTTATGATCCGCCTC GGAAGGRTGTATGCAGATTT TTGTGCATCG
14	7782149	C,T	LG14:8946691	synonymous_variant	LOW	GB52759	301991	1.00	187.58	194.00	351.00	23.00	0.00	1.54	Group14:7781999-7782299	GGTCCTMGGMGGGGAGGAC GATCTGCCCAAGATCTTCGA CAACATAGACGAGCTGTTGC GCGACTACGAGAGGAGCAA ACGTGGATTCAGGACGAATC CGTTCCTTCGAAGGGGCGAT GAGGACGACGAGGAGTGYC TCGAGGGGAATCG[C/T]CCA GGSGGCGACGAGACKTCCAC GTTCCACAGGAGGAGCAAC GTGTTCAAAGAYAGCGAGG CGCACAGGGGCGCGTGGGC GGAGATGTTGAGCACCGTGT TGGATTGGACGCTTGCCTTG

																TCRGTGAGTTRTATACTTCT YCTTCGTGCA
14	7786053	A,G	LG14:8942787	intergenic_region	MODIFIER	GB52759-GB52721	301997	1.00	111.47	107.00	309.00	0.00	18.25	73.85	Group14:7785903-7786203	GGAAGAGAGAAGGGAGAAA AAAAAAGGAAACTCGAAAC TCTGCGTTTAAACCGACGAT GACGAGACTGATGATTTCCA GGTAATTACAAGAGGAAAA TTGCTTTCGAATCCGTTTCGT CTGTAATTCTAATCGGGAAG AGGATTCTGTAGRYTCGTCG AGGAATAATTATTCTCGATC CGATAAATCGTTTYTAAAAA TCTTGATACGAAAAAAAACA ACGATTCGAGGTACAAAGAT TTTCGCTTTGTTTCGTAGGA AGATAAAAACGCGGGCGAC GAAGAGGGCGAAGCTGGCT GGGTC
14	7787263	C,G	LG14:8941577	synonymous_variant	LOW	GB52721	301999	1.00	178.21	189.00	301.00	33.00	0.00	1.54	Group14:7787113-7787413	CGGGTTGTTGTAGTAGCCCT GGTARTTGGGGTTGTAGTTG TGCGAGGCGACCGGGACRG GCGTGGCCGGGCCCGAGGTC CACGAGGGGCCGTAGTGAT AGCCGGGRCTCCCGCTSCCC GTCCCGCCCCCCGCGAAAAT GTCCGGGTCGCT[C/G]GACTT CGCGGGGACCAAGGGCAAC GAGGAGGTCGACGAGGCGC CGTGGYTCGGGTACATGGAC ACGCAGTTGGGCGTGTACGT GTCCTGTTGGCTCGTGTCCG ACGGAAGAACTGCAACAAT TTCGAAATCCATCTTTTTCCR AATCMAA
14	8458568	C,T	LG14:8270797	splice_region_variant&intron_variant	LOW	GB52697	303164	0.99	123.07	123.00	292.00	13.00	0.73	46.15	Group14:8458418-8458718	TGATTAGCAAATTTAATGTA ACGATTATAAATTGCTTCAA AATTGAATGGCTCTCCGTCA TAAATTTCTGTTTCATGTTTC ATTGATATTATCTAAAATAT ATAAATTGAATGTCAAAAAT ATATCCTATTAAAAAAAAGA TTATAAATA[C/T]ACTTACTA AACACATTTCTAATACAGAT AATCCTTCAAGCATTAATAC TTTGTCATCTTGGGAAAACA TTTTACTTGCTTCTATAAAAT CATTTATTTCTAGTCTTTGAT GTTTTTCTGATAATGAAGAA ACAGTTATTGCCAGAAAATT
15	3645893	A,G	LG15:3679113	3_prime_UTR_variant	MODIFIER	GB49484	308363	1.00	111.25	113.00	184.00	26.00	0.00	1.54	Group15:3645743-3646043	CCAGCAATTACATTGYTAAT TTGTGCAAATTTTTGTTGAC AAGTATTTCTCGTCTTTTCTC TGACCGAGTTGTGTATGTTA CGATCGCCTTGATTTCGCGA AGAAGAAAGGATTGAAAAT GGTTGAAGAACGAAAAACA

																AGAAAAATACGRGAGATAA AAAAAGAGTAAAATTCAAA GTATGTCAGAGTGTTTGGTA TCGAGAAAAGGATCACACA CACGCAATAAAGGTACACG ATTTAAAATTACAGCATGCA ACGTTTTTACAAGGCGAATT AACGTTTCTGCGTTTATTCC ATTGTT
15	3646331	C,T	LG15:3679551	3_prime_UTR_variant & downstream_gene_variant	MODIFIER	GB49484 & GB49460	308363	1.00	130.23	127.00	262.00	28.00	0.00	12.31	Group15:3646181-3646481	CACGTCAGGATCAAAGTCTG ACTATGTATTCGTTTTGTTTC ATCGACGTGCTTCGAATTTA GCTATATTATGTTTAACACT CGTTTGTTGTTTTTTCTTTTT CGTGTTGCTACTTTGTCGAA GTTGTCCGATTTAATATGTA CATAATTA[C/T]TTACGTAAT ATGCTATTATCGCTATTTTCA TTAACGTTCCARTTTTTATTA TTATTATTATTATTATTATAA TTATAATAATTATTATTAATT TTTATTATTATTATTATCATC ATTATCATTATAATTATTATT AATATTATCATCATC
15	3664432	A,G	LG15:3697527	synonymous_variant	LOW	GB49485	308367	1.00	60.27	59.00	123.00	14.00	1.46	3.08	Group15:3664282-3664582	TGAAATAAAACCAARATTTG GATTTTTAAGAAGTAGAGAA TTTATGATGAAAGATTTATA TACATTTGATATAAATTTAA ATAATGCACAAGAAACATAT GCTTTAGTATGCAATGCTTA TGAAAACATATTTAAAAAGA TAGGAATAACRTACATAAAA AGTAATAAAATGAATAATAT TAGTTATCTATTGACATTATT TATTTACATTGAGATTTACT ATGACATTATTTATAGGTAT TGGTGATCCAGGATTAATTG GAGATTCAATATCTCATGAA TATCATTATATAAGTGATAT
15	3720433	A,G	LG15:3753197	missense_variant	MODERATE	GB49452	308367	1.00	82.20	81.00	146.00	21.00	0.00	1.54	Group15:3720283-3720583	ATAACTGTCACATTTGTGTC TGTAACACGTCGAGAATGTC GTAGAACTAATGAAGTTAAG CTTAAAAATGGTAATTGGGC AAAAATTCCAGTTAAACCAA TAGCACCTTCAATTATTAGA CGACGAATATATGTATGACA TCCTCGTCTA[A/G]TTAAGGC ATTTAATGCAATTGTTGCAT TTTGCGGATATCTTATTTCTA CTTCTTTCCAAAGAGATGGA TGCCATGCTATTTCCCAAAG ACGTCTGCAAGTTTGAGCAA TGGAACATAGATCTCTGGTA CCCAACCAACTAAATATCTT CA

15	8266886	C,G	LG15:7806253	intron_variant	MODIFIER	GB49997	316103	1.00	199.13	207.00	456.00	32.00	0.00	47.69	Group15:8266736-8267036	AGGGCAAGGRCAAAAAATG GGGTGGGTAATCCCTYCCTA RCCTGGTAGATATTTTACAT ACTCTTACRTGTGATACGTC GAAAATATTAAAATCAGTTT TTTCTTCTTTTTTTTTTTTTC ATATTGAAATTGCAAAATTY AATCGTTAA[C/G]CGAAAGG TGAAGGTCGATTCTCTTTAC GATTCTGTACTTCTGTACTG ATTTATTATATGTTTATTAAG TAGATATACCGTTAATGCAC CCATTGGAAAAAAATAAAC ATACCGGATTGTCTAAATTC AKTCCGAATTTACGAAAAAA GAA
16	2817123	A,G	LG16:3919026	intron_variant	MODIFIER	GB51316 & GB51291	322558	0.99	271.05	266.00	647.00	30.00	0.00	49.23	Group16:2816973-2817273	ATGGCGCTTTGACTTTTAAC TGTTCCCCCTTTGTTCTCACT TTTGCACAAACTTGCAAATG GAAATTGATACTCTCATATT GCAGGTGAATTGTTTCAAGA AAGAAAGAAAGGAAGGAAA AACAAGGATGGATGTAGAT CTCTAGYAAAA[G/A]AGGAT TGTTTCAAGAAARAAAGAA AGGAAGGAAAAGRAAGGAT GGATRTAGATTTCTAGYAAA ARAAGATTGTTTCAARAAAG AAAGAAAGGAAGGAAAAGR AAGRATGGATGTARATYTCT AGRAAAAAAGRATTGGAAT GGAARGAGT
16	3262617	A,G	LG16:3470964	intergenic_region	MODIFIER	mir-iab-4-GB51320	323396	0.99	144.01	146.00	228.00	19.00	0.73	1.54	Group16:3262467-3262767	TCCGTACTTTATCGCGCGTG TGTGCCTTATACGTGTCACT TCGCTGCGTTACTTTGCCGA TATTAATGGCGACTGCCTGT CCAGGCCTACCAATCCTTCC CCYCTCCCTTCCCTTTTTATG CAGTTTATTTTATGGTAACA GCAATGCTC[A/G]AGAGAGT GTAACTCGTTCCTCGAATCG TCCCGCGAATGAAAATTCGT ATTTCGAAGAYGATTCGTTT ATTAGASATAAATTGAACTC GTCYCGTTAAATTATTTTATT ATATAATTAATCTTATCTCA TTTCAAAAATTCAAATGAAT AT
16	6479195	A,C	LG16:6678844	intron_variant	MODIFIER	GB45937	327675	1.00	167.72	171.00	409.00	0.00	18.25	73.85	Group16:6479045-6479345	TTCAAAGCTCCGTTAAAAAA TARAGATACGTGTTATCGRA TTGGAAAGAAAWAAAAATA CAGAGAAATATGTAATAATT TCTCTTGAGGTTTTTCTAATC GYTGAAGAAATTTCGAATTT TTTTWAYTTTCGTTTAGAAT CGAAATCTAT[A/C]TCTATAT

																CCATTGGAATATRTCTATCS AATTTTARTTTCAATTAAGA AAATACTAGGAACATCGATT AAGCRTCTYCTCTTCTGTTTG GACGGCGAAGATACCAAGT TAACCAACGAGTTGGTAACG AGGGTACAATAGCAATTTCG AAG
16	6927725	C,T	LG16:6230999	intergenic_region	MODIFIER	U1-GB45988	328448	0.99	55.53	54.00	106.00	8.00	2.92	1.54	Group16:6927575-6927875	CGGTMGAAAGAAAGGAAAT GATTTCACGATACATCCCCT TTACGACGTAATAAATCATC GAGGCAGTCGAGTCATKAA AGTTTGATCTCGTGACGCGT TTAACGTAATTCRATCTTGC GGATTTTCCACCGATAYCGY CCTATCGTAATT[C/T]ATCRT TAATTCCTCCRAGGAATTGA TTAATCTTTCATTCCCTTTGT GTCCAAACAGTTTTTCCAAT TAARCGAAATCGAAYGATCC CTCGARCGAGAAATTAGGCT TTYGTTTAGTAGGCTTCCAR TTTTATTTCCATGTTTGCAGT ACT

Table S3. Distribution of SNPs across the 16 chromossomes

Chromosome	# Informative SNPs dataset1	% Informative SNPs / total # informarive SNPs	# Informative SNPs dataset2	% Informative SNPs / total # informative SNPs	# SNPS in molecular tool
1	135	2.7	67	13.7	16
2	117	2.4	18	3.7	8
3	41	0.8	45	9.2	8
4	45	0.9	25	5.1	6
5	257	5.2	20	4.1	12
6	388	7.9	35	7.2	10
7	8	0.2	23	4.7	4
8	89	1.8	64	13.1	6
9	121	2.5	20	4.1	10
10	85	1.7	22	4.5	9
11	3338	67.7	29	5.9	15
12	49	1.0	33	6.8	5
13	64	1.3	21	4.3	5
14	69	1.4	22	4.5	7
15	114	2.3	30	6.1	5
16	10	0.2	14	2.9	4
Total	**4930**		**488**		**130**

Table S4. Information of the 130 highly-informative SNPs.

Linkage group	Position	Sequence ontology term	Annotation impact	Gene / Genes	Haplotype blocks	FST
1	164952	intergenic_region	MODIFIER	GB42162-GB42165	377	1,00
1	1449068	synonymous_variant	LOW	GB47659	2587	1,00
1	1452890	5_prime_UTR_premature_start_codon_gain_variant	LOW	GB47660	2587	1,00
1	2921459	intergenic_region	MODIFIER	GB50364-GB50341	4803	0,93
1	4374646	intergenic_region	MODIFIER	GB52268-GB52267	7353	1,00
1	16795237	synonymous_variant	LOW	GB47391	22963	1,00
1	16796849	missense_variant	MODERATE	GB47391	22963	1,00
1	16811932	intron_variant	MODIFIER	GB47499	22994	1,00
1	21085878	3_prime_UTR_variant	MODIFIER	GB42215	29705	1,00
1	21434962	synonymous_variant	LOW	GB45511	29985	1,00
1	22975552	intergenic_region	MODIFIER	GB51609-GB51576	32139	0,91
1	24534342	intron_variant	MODIFIER	GB51630	35570	1,00
1	26406126	intergenic_region	MODIFIER	GB55000-AME.7373	38764	1,00
1	26419991	3_prime_UTR_variant	MODIFIER	GB54998	38788	1,00
1	26443945	intron_variant	MODIFIER	GB54998	38811	1,00
1	27784460	intron_variant	MODIFIER	GB53926	41281	1,00
2	7861322	3_prime_UTR_variant	MODIFIER	GB46590	56913	1,00
2	9095307	intron_variant	MODIFIER	GB52440	59401	1,00
2	11757443	intron_variant	MODIFIER	GB55349	63735	1,00
2	14002844	intron_variant	MODIFIER	INR-2	68011	0,97
2	14413661	intron_variant	MODIFIER	GB55576	68704	1,00
2	15084262	synonymous_variant	LOW	GB55532	69425	1,00
2	15088188	synonymous_variant	LOW	GB55531	69425	1,00

2	15176005	synonymous_variant	LOW	GB55483	69581	1,00
3	1701565	synonymous_variant	LOW	GB49089	72874	0,99
3	2596388	intergenic_region	MODIFIER	GB46930-GB46931	73682	0,97
3	4458140	missense_variant	MODERATE	GB55811	76237	0,99
3	4458749	missense_variant	MODERATE	GB55811	76237	0,99
3	4460272	synonymous_variant	LOW	GB55811	76237	0,99
3	7006583	intron_variant	MODIFIER	GB53701	80927	1,00
3	7069573	intron_variant	MODIFIER	GB53701	81088	1,00
3	7069664	intron_variant	MODIFIER	GB53701	81088	1,00
4	4379048	downstream_gene_variant & intron_variant	MODIFIER	GB49569 & GB49566	98129	1,00
4	4383542	upstream_gene_variant & intron_variant	MODIFIER	tRNA-Ala & GB49566	98138	1,00
4	7672595	intergenic_region	MODIFIER	GB50393-GB50392	104402	0,90
4	7944946	intron_variant	MODIFIER	GB42990	104779	1,00
4	7946420	intron_variant	MODIFIER	GB42990	104779	1,00
4	12216774	intron_variant	MODIFIER	GB53016	111889	0,88
5	316805	intron_variant	MODIFIER	GB44402	112504	1,00
5	871727	intergenic_region	MODIFIER	GB52298-GB52303	113347	1,00
5	2361058	intergenic_region	MODIFIER	AME.21389-GB48879	116100	1,00
5	2364969	intergenic_region	MODIFIER	AME.21389-GB48879	116110	1,00
5	3181962	intergenic_region	MODIFIER	GB55036-GB55037	117141	1,00
5	3507388	intron_variant	MODIFIER	GB51391	117354	1,00
5	3510848	intron_variant	MODIFIER	GB51391	117348	1,00
5	3805632	intergenic_region	MODIFIER	GB51387-GB51386	117817	1,00
5	9447330	intergenic_region	MODIFIER	GB44551-GB44550	128010	1,00
5	11210677	synonymous_variant	LOW	GB44641	131058	0,99
5	13035070	missense_variant	MODERATE	GB41366	133224	0,99
5	13546072	splice_region_variant&intron_variant	LOW	GB41325	133846	0,99
6	5458628	intergenic_region	MODIFIER	GB46090-GB45995	142019	1,00
6	5918385	synonymous_variant	LOW	GB46004	142908	1,00
6	5921026	missense_variant	MODERATE	GB46077	142908	1,00
6	6003400	missense_variant	MODERATE	GB46070	142927	1,00
6	6024411	synonymous_variant	LOW	GRIP84	142927	1,00
6	6305901	intron_variant	MODIFIER	GB46038	142979	1,00
6	6863177	intron_variant	MODIFIER	GB48608	143805	1,00
6	10924120	intergenic_region	MODIFIER	GB52930-GB52950	150210	1,00
6	13660044	synonymous_variant	LOW	GB42979	155072	1,00
6	15332409	intergenic_region	MODIFIER	GB53573-GB53574	158015	1,00
7	8703231	intergenic_region	MODIFIER	GB42408-GB42409	173240	1,00
7	8703865	intergenic_region	MODIFIER	GB42408-GB42409	173240	1,00
7	12723505	intergenic_region	MODIFIER	GB48123-GB48160	180170	1,00
7	12723525	intergenic_region	MODIFIER	GB48123-GB48160	180170	1,00
8	3546391	synonymous_variant	LOW	GB40362	184420	0,99
8	3611478	synonymous_variant	LOW	GB40357	184420	0,99
8	9085611	upstream_gene_variant & intergenic_region	MODIFIER	GB52825 & GB52824-GB52825	192952	1,00
8	9086908	splice_region_variant&intron_variant	LOW	GB52825	192952	1,00
8	9087168	synonymous_variant	LOW	GB52825	192952	1,00
8	9115014	synonymous_variant	LOW	GB52848	192963	1,00
9	15736	intron_variant	MODIFIER	GB45547	199195	1,00
9	1971514	intron_variant	MODIFIER	GB43759	201244	1,00
9	2115514	intron_variant	MODIFIER	GB43758	201595	1,00
9	3785070	intergenic_region	MODIFIER	GB44856-GB44844	204819	1,00

9	5053380	intergenic_region	MODIFIER	GB42732-GB42731	207068	1,00
9	5454899	intron_variant	MODIFIER	GB42708	207594	1,00
9	5456771	synonymous_variant	LOW	GB42707	207594	1,00
9	10610587	splice_region_variant&intron_variant	LOW	GB53305	217286	1,00
9	10646051	3_prime_UTR_variant	MODIFIER	GB53433	217322	1,00
9	10703867	intergenic_region	MODIFIER	GB53297-GB53296	217341	1,00
10	164597	intergenic_region	MODIFIER	GB43337-GB43334	218485	1,00
10	164648	intergenic_region	MODIFIER	GB43337-GB43334	218485	1,00
10	6115851	intron_variant	MODIFIER	GB48386	226398	1,00
10	7933755	intron_variant	MODIFIER	GB54295	228810	1,00
10	7934803	intron_variant	MODIFIER	GB54295	228813	1,00
10	8168036	intron_variant	MODIFIER	GB54295	229135	1,00
10	8884847	upstream_gene_variant & intergenic_region	MODIFIER	GB50874 & GB51176-GB50874	229901	1,00
10	10499837	synonymous_variant	LOW	GB51072	231674	0,99
10	12157726	intron_variant	MODIFIER	GB48760	234177	1,00
11	1468594	synonymous_variant	LOW	GB52239	237033	1,00
11	1778090	intergenic_region	MODIFIER	GB55189-GB55188	237378	1,00
11	2936665	missense_variant	MODERATE	GB55146	239096	1,00
11	2937477	missense_variant	MODERATE	GB55146	239096	1,00
11	2983919	missense_variant	MODERATE	GB55142	239129	1,00
11	4155603	intergenic_region	MODIFIER	GB42097-GB42095	240451	1,00
11	4763028	splice_region_variant&synonymous_variant	LOW	GB55058	240480	1,00
11	4904280	splice_region_variant&synonymous_variant	LOW	GB55054	240480	1,00
11	5159160	synonymous_variant	LOW	GB54035	240480	1,00
11	5187989	synonymous_variant	LOW	GB54054	240480	1,00
11	7543229	intron_variant	MODIFIER	GB47257	243455	1,00
11	8912535	intergenic_region	MODIFIER	GB47197-GB45090	245407	1,00
11	8917812	intergenic_region	MODIFIER	GB45090-GB45089	245420	0,99
11	10589114	intergenic_region	MODIFIER	GB45130-GB45131	248895	1,00
11	14629718	intron_variant	MODIFIER	GB43173	253884	1,00
12	1660591	intron_variant	MODIFIER	GB40187	256375	0,93
12	3713814	intergenic_region	MODIFIER	GB48920-GB53610	258012	1,00
12	3714188	intergenic_region	MODIFIER	GB48920-GB53610	258012	1,00
12	3714356	intergenic_region	MODIFIER	GB48920-GB53610	258012	1,00
12	9259013	intergenic_region	MODIFIER	GB52041-GB52042	267814	1,00
13	2341902	splice_region_variant & intron_variant & downstream_gene_variant	LOW & MODIFIER	GB50671 & GB50679	276376	1,00
13	6167082	intergenic_region	MODIFIER	GB53867-GB53868	283061	1,00
13	8773197	intergenic_region	MODIFIER	GB49915-GB49865	287712	1,00
13	10238516	intron_variant	MODIFIER	GB41542 & GB41544	289465	1,00
13	10241239	intron_variant	MODIFIER	GB41542	289468	1,00
14	1794663	synonymous_variant	LOW	GB48483	292003	0,99
14	2269579	synonymous_variant	LOW	GB43781	292355	0,99
14	3394800	intron_variant	MODIFIER	GB43598 & GB43596	294576	1,00
14	7782149	synonymous_variant	LOW	GB52759	301991	1,00
14	7786053	intergenic_region	MODIFIER	GB52759-GB52721	301997	1,00
14	7787263	synonymous_variant	LOW	GB52721	301999	1,00
14	8458568	splice_region_variant&intron_variant	LOW	GB52697	303164	0,99
15	3645893	3_prime_UTR_variant	MODIFIER	GB49484	308363	1,00
15	3646331	3_prime_UTR_variant & downstream_gene_variant	MODIFIER	GB49484 & GB49460	308363	1,00
15	3664432	synonymous_variant	LOW	GB49485	308367	1,00

15	3720433	missense_variant	MODERATE	GB49452	308367	1,00
15	8266886	intron_variant	MODIFIER	GB49997	316103	1,00
16	2817123	intron_variant	MODIFIER	GB51316 & GB51291	322558	0,99
16	3262617	intergenic_region	MODIFIER	mir-iab-4-GB51320	323396	0,99
16	6479195	intron_variant	MODIFIER	GB45937	327675	1,00
16	6927725	intergenic_region	MODIFIER	U1-GB45988	328448	0,99

Table S5. Samples used to Assess the sensitivity and accuracy of SNP tool for determining allele frequencies

CIMO ID	Sample ID	M porportion (µL)	C porportion (µL)	Ratio
WGS1.1	**2759 (M)+2724 (C)**	0	10	0:10
WGS1.2		1	9	1:9
WGS1.3		2	8	2:8
WGS1.4		3	7	3:7
WGS1.5		4	6	4:6
WGS1.6		5	5	5:5
WGS1.7		6	4	6:4
WGS1.8		7	3	7:3
WGS1.9		8	2	8:2
WGS1.10		9	1	9:1
WGS1.11		10	0	10:0
WGS2.1	**2704 (M)+2722 (C)**	0	10	0:10
WGS2.2		1	9	1:9
WGS2.3		2	8	2:8
WGS2.4		3	7	3:7
WGS2.5		4	6	4:6
WGS2.6		5	5	5:5
WGS2.7		6	4	6:4
WGS2.8		7	3	7:3
WGS2.9		8	2	8:2
WGS2.10		9	1	9:1
WGS2.11		10	0	10:0
B1.1	**1D (c x4D (C)**	0	10	0:10
B1.2		1	9	1:9
B1.3		2	8	2:8
B1.4		3	7	3:7
B1.5		4	6	4:6
B1.6		5	5	5:5
B1.7		6	4	6:4
B1.8		7	3	7:3
B1.9		8	2	8:2
B1.10		9	1	9:1
B1.11		10	0	10:0
B2.1	2D (M)+3D (C)	0	10	0:10
B2.2		1	9	1:9

B2.3		2	8	2:8
B2.4		3	7	3:7
B2.5		4	6	4:6
B2.6		5	5	5:5
B2.7		6	4	6:4
B2.8		7	3	7:3
B2.9		8	2	8:2
B2.10		9	1	9:1
B2.11		10	0	10:0