DOCTOR OF PHILOSOPHY

Population genetics of the farmland sawfly Dolerus aeneus (Hymenoptera, Symphyta)

Cook, Nicola

Award date:
2011

DOCTOR OF PHILOSOPHY

Population genetics of the farmland sawfly Dolerus aeneus (Hymenoptera, Symphyta)

Nicola Cook

2011

University of Dundee

Population genetics of the farmland sawfly *Dolerus aeneus* (Hymenoptera, Symphyta)

Nicola Cook

Presented for the degree of Doctor of Philosophy at the University of Dundee

September 2011

List of Contents

Publications

Cook N, Aziz N, Hedley PE, Morris J, Milne L, Karley AJ, Hubbard SF and Russell JR (2011) Transcriptome sequencing of an ecologically important graminivorous sawfly: a resource for marker development. *Conservation Genetics Resources* **3**, 789-795.

List of Tables

List of Figures

List of Plates

Acknowledgements

Dr. Joanne Russell, Dr. Alison Karley, Dr. Dave Parish and Prof. Stephen Hubbard who provided excellent supervision throughout the project. Particularly to Joanne who put up with an endless stream of questions and was always supportive.

Gaynor McKenzie and Allan Booth who taught me the tricks of the molecular trade and provided the best lab banter known to science!

Edward Baxter and David Soutar who gave permission to collect sawflies on their land.

Dr. Alan Barclay for a huge amount of help with the field work throughout the project and for the provision of much-needed hamburgers!

Dr. Nicola Randall, Dr. Steve Moreby and colleagues of the Game and Wildlife Conservation Trust and Lynne Karley who provided help collecting sawflies from field sites south of the border.

Dr. Naveed Aziz and colleagues at the University of York who contributed to the project in the form of Next-generation sequencing.

Dr. Rosemary Clarke and Arlene Wigham for their help with the flow cytometry.

Dr. Linda Milne who performed the bioinformatics on the NGS data.

Dr. Pete Hedley and Jenny Morris for their help with the RNA work.

Clare Booth and Louise Donnelly who provided sequencing and genotyping services.

Dr. Ian Dawson and Dr. Katrin Mackenzie who provided advice on population genetic and statistical aspects of the project.

Philip Smith for thoroughly proofreading this document.

Emily Clark for being a great friend, for helping out with my field work (for reasons best left unstated) and for answering daft questions all the time in the early part of the project!

My friends (in no particular order) Gaynor McKenzie, Allan Booth, Rachael Adams, Anna Macrae, Natalie Walker, Hannah Clarke, Fiona Fraser, Nicky Bonar, Kelly Houston, Hazel Bull, Lea Wiesel and Helen Downie who made SCRI a fun place to be and listened to my whining when I was writing up!

My partner, Grant Reid who has put up with so much and has been my rock, thank you.

And last but definitely not least my mum, Lorraine Cook, for all her help and support, I couldn't have done it without you!

Declaration

I hereby declare that the following thesis is based on the results of investigation carried out by myself, and that this thesis is my own composition. This thesis has not in whole, or any part, been previously presented for a higher degree. Work other than my own is clearly stated in the text with reference to the relevant researchers or their publications.

Nicola Cook

Abstract

Over the last 50 years populations of certain farmland birds have undergone severe declines over the same timescale that agriculture has intensified. The larvae of grassland sawflies (Hymenoptera, Symphyta) are a key component of the fledgling diet and it is thought that agri-intensification has reduced the numbers of these insects to such an extent that the populations of bird species dependent on them are limited. Sawfly populations may be more than usually susceptible to disturbance because firstly, their adult stages are poor dispersers and secondly, some species possess Complementary Sex Determination (CSD), a mechanism which can produce sterile males under inbreeding conditions.

This study has produced a sawfly transcriptomic library through the use of 454 pyrosequencing, the first genetic resource for any farmland sawfly. From this library, a set of 13 polymorphic microsatellite markers were isolated for use in the common farmland sawfly *Dolerus aeneus*. Using these markers, three Scottish populations of *D. aeneus*, a species common to all six UK sites sampled in this study, exhibited similar levels of genetic diversity and low levels of genetic differentiation. However, evidence of inbreeding was detected in each of the three populations. In addition, potential diploid males were detected in *D. aeneus* using microsatellite markers, a primary indication that CSD may be operating in this species.

The population genetic analysis in the current study suggests that fragmentation of suitable sawfly habitat as a result of agricultural intensification has not yet acted to isolate *D. aeneus* populations, although some genetic effects (inbreeding and low diversity compared with non-threatened hymenopterans) are apparent. In addition, diploid males have been detected which may have compromised fertility. This study will be of interest to research groups working on the genetics of the Hymenoptera and on the conservation and management of sawflies and the bird species dependent on them.

1. General Introduction

1.1 Agricultural intensification and the declining populations of farmland birds

1.1.1 Overview

Over recent decades, agriculture has intensified throughout the UK as farmers combine new technologies with more intensive practices in an attempt to meet the food and energy demands of an ever-increasing human population. This is generally referred to as "agricultural intensification". In Britain, intensification has involved changes in a number of practices, including the increased use of agro-chemicals, both pesticides and fertilisers, widespread switching on lowland farms to crops such as oilseed rape, while spring-sown cereals, root crops, hay crops and fallow periods (periods where the land is uncultivated) have all declined, and there is a general trend towards crop monoculture (Chamberlain *et al.* 2000). The trend towards homogenization of the landscape has been exacerbated by the removal of hedgerows and other uncultivated land to produce larger fields for crop production. As a consequence the amount of semi-natural habitat present in the agricultural environment has declined with the overall result of greatly reduced farmland biodiversity (Benton *et al.* 2003; Newton 2004).

Many types of farmland wildlife have been adversely affected in terms of their population size by the multivariate process of agricultural intensification including mammals, arthropods and flowering plants (Benton *et al.* 2003). However, farmland birds have long been the major focus of research into the ecological effects of land-use changes. Their conspicuous nature makes them easy to research and their patterns of behaviour, distribution and seasonal life cycle track closely the temporal and spatial changes in the agricultural landscape (Ormerod and Watkinson 2000).

The effects of agricultural intensification can be difficult to quantify as the process consists of a number of different components each of which affect different species in different ways. Also, the fact that the component processes of intensification have occurred concurrently and are co-dependent makes it difficult to separate the effects of one change from that of another (Newton 2004). The main mechanisms proposed to affect farmland birds are a lack of nutritious invertebrate food sources, a lack of suitable nesting

habitat or direct mortality due to farming operations (Newton 2004; Chamberlain *et al.* 2000; Brickle *et al.* 2000; Chamberlain and Crick 1999; Crick *et al.* 1994; Potts 1986).

The Grey Partridge (*Perdix perdix*) in particular has been studied extensively with reference to agricultural intensification. A long-term monitoring study in the Sussex Downs in southern England started by Dick Potts in 1968 was designed to continuously monitor changes in the abundance and diversity of weeds, insects and birds on farmland with the intention of explaining long-term changes in the arable environment (Potts 1986; Potts *et al.* 2010). Prior to World War I, gamekeepers in the region recorded an average of 50 breeding pairs per km^2 each spring. In the first year of the study, numbers of breeding pairs were shown to have declined to 20% of the pre-war density (Potts 1986). The main cause for the decline in the number of Grey Partridges was identified as a herbicide-induced reduction in the abundance of cereal-crop invertebrates leading to reduced chick survival (Potts 1986).

Research has confirmed that, like the Grey Partridge, other farmland bird species have experienced severe population declines due to changes in farmland management. Some species have experienced reductions of 80% or more in numbers and thus exhibit reduced geographical range over a period of less than 20 years (Fuller *et al.* 1995). Newton (2004) reviewed the declines of 30 such bird species closely associated with the farmland environment. He found that in the vast majority of these species, habitat reduction and/or a reduction in food supply were the main factors limiting population size.

The widespread reduction of invertebrates present in modern cereal crops, in terms of both abundance and diversity, is largely attributed to the increased use of agro-chemicals (Newton 2004; Potts 1986). Insecticides directly deplete arthropods that are taken by adult birds or fed to the young during the breeding season whereas herbicides indirectly deplete the food source by reducing the abundance of weeds that are host-plants for the arthropods (Boatman *et al.* 2004). The declining practice of undersowing temporary grassland within a mixed arable/ grass ley rotation (Sotherton 1998) and the reduction in the area of uncultivated field margins (Wilson *et al.* 1999) are also thought to be contributing factors in invertebrate decline.

1.1.2 The invertebrate food source of farmland birds including sawflies

In a review of farmland bird diets (Holland *et al.* 2006) invertebrates comprised 65 – 100% of the diet of 15 of the most common farmland birds in Europe. The list includes the Grey Partridge (*Perdix perdix)*, the Corn Bunting (*Miliaria calandra*), the Skylark (*Alauda arvensis*) and the Yellowhammer (*Emberiza citronella*) all of which inhabit British farmland and are of conservation concern. Wilson *et al.* (1999) noted that during the breeding season when invertebrates are a more highly sought-after food source, especially on behalf of dependent young, that the following groups were the most important: Orthoptera: Acrididae (grasshoppers); Hymenoptera: Symphyta (sawflies, more specifically the larvae): Arachnida: Araneae (spiders); Coleoptera (beetles); Lepidoptera (butterflies, more specifically the larvae); Hemiptera: Aphididae (aphids); Diptera; and Tipulidae (crane-flies and their larvae). It was found that grasshoppers, sawflies, leaf-beetles and spiders were all significantly more important in the diet of bird species in population decline compared to non-declining bird species. All four of these important invertebrate groups were found to be particularly sensitive to insecticide applications and are likely to be affected by declines in marginal habitats and increased tillage.

Sawflies (Hymenoptera: Symphyta) were reported to be of particular importance to the Grey Partridge as they are a highly nutritious food source for the chicks (Potts 1986; 1970). Preliminary investigations suggested that the average sawfly larva is the nutritional equivalent of 20 cereal aphids (Potts 1970). Since then, research has highlighted the significance of sawflies in the diet of other farmland birds, particularly in the nestlings, including the corn bunting (Aebischer and Ward 1997) and the skylark (Poulsen *et al.* 1998).

It is important to note that although some invertebrate groups have suffered as a result of the intensification of cereal farming others, such as cereal aphids, have thrived (Borg and Toft 1999; 2000). An experiment carried out by Borg and Toft (1999) illustrated that Grey Partridge chicks perform less well on a diet entirely composed of the cereal aphid *Rhopalosiphum padi* as opposed to a diet of mixed insect species containing those individuals of higher nutritional value. Chicks that fed on diets composed entirely of cereal aphids, or with a higher proportion of cereal aphids than would be preferentially

selected by the birds, exhibited poor growth rates and poor flight feather development. This experiment demonstrates that although there may be a larger biomass of alternative insect species present as a result of agricultural intensification, this is no substitute for a diet of insects such as sawflies that are of higher nutritional value.

1.2 Biology of sawflies

1.2.1 Taxonomy

The insect order Hymenoptera, comprising over 120,000 species worldwide (Chinery 1993), is split into two distinct sub-orders; the Apocrita containing the familiar wasps, bees and ants, and the Symphyta consisting of the sawflies alone. The Symphyta is by far the smaller sub-order (10,000 species worldwide) and its members are the most primitive of the Hymenoptera as a whole (Wright 1990). The Sawfly family Tenthredinidae is the predominant sawfly family in all parts of the world and over 380 of the 500 British species belong to this family. The remaining British species are distributed among eleven other families (Benson 1950).

The common name "sawfly" is derived from the shape of the female ovipositor, which takes the form of a pair of serrated appendages or "saws" held within a sheath, which is used to slice into plant material to create a space to lay eggs (Benson 1950) (Plate 1.1).

The Symphytans can be easily distinguished from their Apocritan counterparts by the lack of a marked constriction between the 1st and 2nd abdominal segments (Quinlan and Gauld 1981), the characteristic "wasp-waist". Also, all adult sawflies, with the exception of those in the family Cephidae, possess a pair of structures known as the cenchri located posterior to the scutellum. The cenchri come into contact with a rough area on the underside of the wings holding them in place when folded. It is thought that members of the Cephidae, who also display signs of a constricted waist, form a link between the primitive Symphytans and the more "advanced" Apocrita.

1.2.2 Life history

Sawfly adults emerge during the spring and summer period, with the flight of individuals spread over weeks or months depending on the species. In some species individuals appear at staggered intervals throughout the season (Benson, 1950). The peak season for adult sawflies in the UK is May (England) or June (Scotland), although

Plate 1.1: Female sawfly* (Dolerus nitens) *using her ovipositor to slice into a leaf. Image © Bruce Marlin. http://www.cirrusimage.com

individuals of the *Dolerus, Amauronematus* and *Aglaostigma* genera (common Tenthredinid sawflies) have been known to appear as early as March (Benson 1950).

When the adult initially emerges from the pupal skin it remains some days within its shelter allowing the sclerotinous covering to develop and harden. The length of time spent in this activity varies from 1–7 days between species. Upon emergence the adults will feed. Sawflies are very specific in their choice of foodplant, with many species feeding only on one tissue type of a single plant species. Shortly after emergence females will begin to oviposit, often without waiting for males to mate with them (Benson 1950).

Eggs are laid in the chosen food plant using the saw-like ovipositor to make a slit in the plant material. Eggs may be found entirely, or at least partly, embedded in the tissue with typically one egg per slit, although there may be many of these slits or pockets per leaf or stem. The oviposition habits of female sawflies are naturally closely related to the size and shape of their ovipositor, as are the size, shape and number of eggs laid. The length of incubation period varies within and between species according to environmental factors such as temperature and humidity. Hatching after a period of 9–14 days is considered normal. Sawfly eggs are generally very sensitive to excessive drought, heat, damp or cold and are easily killed (Benson, 1950).

Sawfly larvae can be found in abundance from June through to late autumn in the UK, first appearing typically two to three weeks after the flight of the adults. The majority of larvae feed on leaves, either in groups and conspicuous, or solitary and inconspicuous. Camouflage is achieved in two ways: the larvae can resemble the leaves of their food plant in colour and stay close to the leaf surface while feeding; or the larvae can feed only on the underside of leaves. Some sawfly larvae only come out to feed after dark or in wet weather, although the likely cause of this behaviour is probably driven by their humidity requirements. It is common for the solitary larvae to be more cryptic in appearance than the gregarious larvae. The gregarious individuals tend to be more vividly marked and rear up when provoked, using “scare tactics” as a defence mechanism (Benson, 1950; Wright, 1990).

The larvae feed over a few weeks, although in some species, for example *Tenthredopsis* species, larvae can be found late in the year feeding over the course of two or three months. When the larvae are fully grown, they pass into the pre-pupal resting state, which is how the vast majority of sawfly species over-winter. During this stage, the

pre-pupae either burrow into soil or leaf-litter, into galls (sometimes the galls of other insects), under soft bark or attach themselves to the stems or leaves of a living plant. The choice of location is directly linked to the humidity and desiccation levels preferable to the species. The pre-pupa is much like the larva in form except that the colour patterns, spines and hairs are much reduced (Benson 1950; Wright 1990).

In the spring, a pupa is formed from the pre-pupa, the change probably stimulated by shifts in temperature and humidity, and the adult soon sheds the pupal case and takes flight. Although a single brood per year is common in most sawfly species, there are some species that produce up to 3 generations per year in which case the life cycle is greatly accelerated. In the UK, the majority of sawflies are univoltine (Benson 1950; Wright 1990).

1.2.3 Sex determination in sawflies

1.2.3.1 Haplodiploidy and Complementary Sex Determination (CSD)

The dominant mode of sex determination in the Hymenoptera (Symphyta) is arrhenotokous haplodiploidy (arrhenotoky). This is a form of haplodiploidy by which males develop from unfertilized eggs (by parthenogenesis) and are haploid and females develop from fertilized eggs and are diploid (Heimpel and de Boer 2008). Mated hymenopteran females have control over fertilization by choosing whether or not to release stored sperm from the spermatheca at the point of oviposition, meaning that females can actively adjust the sex ratio of their offspring (van Wilgenburg *et al.* 2006). The other main form of haplodiploid sex determination known in insects is paternal genome elimination (PGE) whereby males develop from fertilised eggs and are diploid in the first instance but lose their paternal chromosome set early in development (Heimpel and de Boer 2008).

The two methods by which arrhenotoky can be achieved that have received empirical support are genomic imprinting, under which activation of the female developmental pathway requires paternally derived genes, and Complementary Sex Determination (CSD) (Heimpel and de Boer 2008). CSD is the system that is best understood and has been confirmed, since its discovery by Whiting (1943), using a variety

of methods, in more than 60 species of Hymenoptera including three Symphytans: *Athalia rosae*, *Neodiprion nigroscutum* and *Neodiprion pinetum* (Heimpel and de Boer 2008; van Wilgenburg *et al.* 2006; Stahlhut and Cowan 2004).

Under CSD, the sex of an individual not only depends upon ploidy, but also upon the allelic composition at a single sex-determining locus. Heterozygotes at this locus develop as females, whereas homozygous individuals develop as males. Haploid individuals develop normally into males due to being hemizygous at this locus (Heimpel and de Boer 2008; van Wilgenburg *et al.* 2006). When only one locus is involved, the process is referred to as single-locus Complementary Sex Determination (sl-CSD), but incidences of multi-locus CSD (ml-CSD) have also been reported. Under the ml-CSD model, males are produced only if diploid individuals are homozygous at all of the sex determining loci; a female will be produced if one or more of the multiple sex-determining loci is heterozygous (Heimpel and de Boer 2008).

The sex determination locus (SDL) has recently been sequenced in the honeybee *Apis mellifera* (Hasselmann *et al.* 2008; Beye *et al.* 2003). The SDL contains five known genes, two of which have sex determination function: the complementary sex determiner gene (*csd*) and the newly discovered *feminizer* (*fem*) gene (Gempe *et al.* 2009; Hasselmann *et al.* 2008) (Figure 1.1). Heterozygosity at *csd* leads to a female-specific splice-variant at *fem* which encodes a functional protein allowing female development to progress. Homo- or hemizygosity at *csd* leads to the production of the default male-specific splice variant at *fem* which contains a premature stop codon and yields no functional protein; therefore male development ensues (Hasselmann *et al.* 2008).

However, little is known about the regulation linking sex determination to sexual differentiation. RNAi knockdown experiments indicate that the other mRNA-producing genes present in SDL (GB11211, GB13727 and GB30480) (Figure 1.1) do not have sexual differentiation function and that the paralogous gene pair *fem* and *csd* is required for the control of female differentiation in both the somatic and the germ cells (Gempe *et al.* 2009). On the basis of the protein amino acid sequence, *csd* appears to be homologous to the *tra* protein in *Drosophila melanogaster* (Beye *et al.* 2003) which plays a role in sex determination by regulating the downstream sex-specific splicing of the doublesex (*dsx*) gene at the end of the sex determining pathway.

Figure 1.1: Position of known genes within the sex determining locus (SDL) of* Apis mellifera. *Genes are orientated 5' to 3' according to the direction of the arrows. The names of the functionally characterised genes are underlined.

Dsx controls the activity of the final target genes necessary for both male and female somatic differentiation. It has been confirmed that *dsx* in honeybees (*Am-dsx*) is sex-specifically spliced and is a potential target for the signal initiated by *csd* (Cho *et al.* 2007). In addition, it has been shown that heterozygosity at *csd* is only required to initiate female development in early embryogenesis whereas *fem* maintains this course of action throughout development. Expression of the female-specific *fem* protein directs further female-specific splicing of *fem* in a positive feedback splicing loop maintaining the female signal (Gempe *et al.* 2009).

1.2.3.2 Evolution of arrhenotoky and Complementary Sex Determination (CSD)

Arrhenotoky is not exclusive to the Hymenoptera; at least 12 independent origins have been proposed. The most likely precursor to haplodiploidy in the Hymenoptera is thought to be standard diplodiploidy (both males and females are diploid) and a number of evolutionary routes to arrhenotoky from a diplodiploid ancestor have been suggested. These routes have been broadly grouped into those that invoke inbreeding (breeding amongst closely related individuals) as a precursor and those that invoke outbreeding (Reviewed by Heimpel and de Boer 2008).

Early models suggested that inbreeding may have been involved in the evolution of arrhenotoky due to the fact that persistent inbreeding creates a situation in which males compete for access to their sisters as mates thus selecting for a female-biased sex ratio. Arrhenotoky provides a mechanism by which these sex ratios can be achieved (Hamilton 1967). Also, inbreeding in diplodiploid ancestors and the purging of deleterious alleles associated with inbreeding was seen as a possible preadaptation to arrhenotoky (Borgia 1980; Brown 1964). However, in the Hymenoptera at least two lines of reasoning suggest that this was not the case. Firstly, inbreeding does not seem prevalent in the ancestral Hymenoptera (the Symphyta) (Coppel and Benjamin 1965; Walter *et al.* 1994) and secondly inbreeding depression as a result of the deleterious effects of Complementary Sex Determination (CSD) is widespread in ancestral hymenopterans (discussed in Section 1.2.3.3) (Heimpel and de Boer 2008).

There are also models of arrhenotoky evolution that do not assume inbreeding as a precursor and are based on the fact that haploid sons are more closely related to their

mothers under arrhenotoky than diploid sons to their mothers under diplodiploidy. This asymmetry in relatedness could cause a rare arrhenotoky allele to spread in an ancestral diplodiploid population and with continued outbreeding this allele could become fixed. Arrhenotoky would only spread under these conditions if the fitness of the haploid males was sufficient (Borgia 1980; Smith 2000; Bull 1979).

Normark (2003) reviewed the evolution of alternative genetic systems in insects and examined the evolution of haplodiploidy from an ecological perspective. He noted that haplodiploid lineages were very few and ancient and appeared to arise in lineages that used woody plants and stems as a food source (either the wood itself or the sap within the plant). Woody plants are a nutritionally poor food source and the insects that rely on them usually also rely on maternally inherited bacterial endosymbionts. Some of the basal lineages of the Hymenoptera feed on dead wood and harbour intracellular bacteria but the correlation between the bacteria and haplodiploidy is unclear. Hamilton (1993) speculated that haplodiploidy may have been the outcome of a history of conflict over sex determination between the intracellular bacteria and their hosts. The endosymbiont seeks to feminize the host and in response the host multiplies and/or moves the sex-determining elements around the genome. This in turn presents more targets for the endosymbiont until finally all surviving X chromosomes behave as autosomes and sex determination is based on chromosome dosage alone. More recent models also lend support to the "endosymbiont-induced haplodiploidy" hypothesis but the conditions for this to occur are very specific and/or the endosymbiont must confer some additional direct benefit to the host to allow it to persist in the population (Kuijper and Pen 2009).

To summarise, a wide range of models for the evolution of arrhenotoky have been put forward and because of the fact that arrhenotoky has evolved independently on several occasions there is no requirement for a "one-size-fits-all" hypothesis, although, it is unlikely that inbreeding was involved in the evolution of arrhenotoky in the Hymenoptera due to the lack of inbreeding in the ancestral hymenoptera and the inbreeding depression brought about as a result of CSD. The presence of CSD in a representative of every major hymenopteran group has led to the assumption that CSD is the ancestral mode of sex determination in the order. However, knowledge of the phylogenetic distribution of CSD is incomplete and interestingly species with and without CSD occur within the same genus (*Cotesia*; de Boer *et al.* 2007; Zhou *et al.* 2007; 2006; Niyibigira *et al.* 2004; Gu

and Dorn 2003). An expansion of the search for CSD into the more primitive members of the Hymenoptera, including the Symphyta, will help to determine whether CSD is in fact the ancestral mode of sex determination (van Wilgenburg *et al.* 2006; Cook and Crozier 1995).

1.2.3.3 Complementary Sex Determination (CSD) and the consequences of diploid male production

Diploid males produced as a result of CSD have been shown to be inviable or sterile in the vast majority of hymenopteran species studied (Elias *et al.* 2009; Heimpel and de Boer 2008; van Wilgenburg *et al.* 2006). For example, diploid males of the sawfly *Neodiprion nigroscotum* are incapable of mating properly (Smith and Wallace 1971) whereas diploid males of the sawfly *Athalia rosae ruficornis* mate with females without any difficulty, although the resulting offspring are triploid and sterile (Naito and Suzuki 1991). It had become a general assumption that diploid males perform poorly across a range of traits associated with reproductive fitness and that if they did succeed in fathering surviving offspring, these offspring in turn were sterile, but this generalization was called into question when new research highlighted functionally reproductive males in the solitary wasp *Euodynerus foraminatus* (Cowan and Stahlhut 2004) and more recently in the parasitoid *Cotesia glomerata* (Elias *et al.* 2009).

It is generally accepted that haplodiploids are less susceptible to the effects of inbreeding due to the fact that deleterious alleles are more effectively expelled from the population via the haploid males (Zhou *et al.* 2007; Butcher *et al.* 2000). However, those haplodiploid species possessing the Complementary Sex Determination (CSD) system and producing reproductively compromised diploid males will be prone to inbreeding depression (Zayed and Packer 2005). This is because inbreeding increases the chance of homozygosity at the sex determining locus (loci) and therefore increases diploid male production (DMP) (Cook and Crozier 1995; van Wilgenburg *et al*. 2006). Increased DMP initially reduces population growth rate and effective population size (the number of individuals in a theoretically ideal population exhibiting the same level of genetic drift as the actual population (Hartl and Clark 1997)). In small closed populations, genetic drift combined with demographic and environmental stochasticity leads to a reduction in the

number of sex alleles and therefore higher DMP. The cycle continues in a process termed the "diploid male vortex" which ultimately leads to extinction (Zayed and Packer 2005).

Selection against inbreeding is to be expected in species that are CSD positive (Zhou *et al.* 2007) with a range of strategies being applied to combat DMP. Indeed, the mating systems of many species with CSD appear to be characterized by outbreeding with both temporal and spatial segregation of offspring to reduce the occurrence of sib-mating (Heimpel and de Boer 2008). For example, females of the sawfly *Athalia rosae* lay fertilised eggs early in life and unfertilised eggs later on resulting in a temporal segregation of opposite-sex kin (Lee *et al.* 1998). In the gregarious wasp *Bracon hebetor* females and males are unreceptive to mating for the first two hours after emergence, and thus disperse from their natal area before mating (Ode *et al.* 1995)

Many species with CSD have mechanisms, which may or may not have arisen in response to CSD, that counteract DMP as described above. However, DMP levels are necessarily dependent on the sex allele diversity within a population. Estimates of the number of sex alleles (for a single sex-determining locus) in hymenopteran populations lie generally between nine and 20 corresponding to levels of DMP production of 5–11% (Cook and Crozier 1995), although, up to 86 alleles at the sex locus have been detected in natural populations of the fire ant *Solenopsis invicta* (Ross *et al.* 1993). Large haplodiploid populations can maintain many sex alleles and therefore have lower levels of DMP. However, in small, isolated populations, genetic drift reduces sex allele richness and increases DMP (Zayed and Packer 2005).

1.3 Susceptibility of sawflies to agri-intensification

Studies carried out as part of a long-term monitoring project of farmland flora and fauna in the Sussex Downs, run by the Game and Wildlife Conservation Trust, have shown that sawfly numbers declined steadily over the period 1970–1991. Over the same period of time, the use of fertilizers and pesticides has increased, hedgerows and marginal areas have declined, field sizes have increased and traditional ley farming and crop rotations have been largely abandoned (Aebischer 1991). A combination of the sawfly life cycle and the potential presence of CSD in these species could mean that sawflies are more susceptible to the effects of agricultural intensification than other invertebrates occupying the same habitat.

As discussed in Section 1.2.2 the vast majority of sawfly species overwinter as pupae in the soil. Mechanical disturbance of the overwintering pupae due to increased levels of winter cropping could increase sawfly mortality by 50% (Barker *et al.* 1999). In addition, sawflies have suffered as a result of increasing pesticide application. As a general rule sawflies are slow-reproducing insects (many species are univoltine) and research has shown that populations can take up to four years to recover from a single summer application of a broad-spectrum insecticide (Sotherton 1990; Aebischer 1990). The increased use of herbicides and a decline in the practice of undersowing (the use of cereals as a nurse crop for grass) is thought to have eliminated a number of sawfly host plants from the agricultural landscape (Sotherton 1998, 1990). This may be a particular problem as sawflies are very specific in their choice of foodplant (Section 1.2.2).

The processes described above could be acting in combination to fragment suitable sawfly habitat. Given that sawflies are thought to disperse poorly (Benson 1950), and could therefore struggle to colonise new habitat, it is possible that sawfly populations could become isolated and subject to inbreeding depression. The potential presence of CSD, and possibly of sterile diploid males, could compound the effects of inbreeding (Cook and Crozier 1995) and ultimately lead to local extinction of sawfly populations (Zayed and Packer 2005).

1.4 Conservation genetics

1.4.1 The importance of genetic variation

Small, endangered populations differ from larger populations in two main respects. The level of inbreeding is increased and the importance of genetic drift (stochastic loss of alleles) in relation to the population genetic structure is increased. Both processes lead to a loss of genetic diversity which means reduced evolutionary potential and reproductive fitness and a heightened extinction risk (Spielman *et al.* 2004). Small, potentially threatened populations should exhibit lower genetic diversity as compared to taxonomically related non-threatened populations (Höglund 2009). In an extensive analysis of the genetic diversity of 170 endangered species and taxonomically related non-threatened species, 77% of threatened species exhibited lower heterozygosity, and heterozygosity was 35% lower, on average, in threatened taxa compared with non-threatened taxa (Spielman *et al.* 2004).

In addition to estimating genetic diversity levels within populations it is also important to determine the extent of population fragmentation to identify threatened populations. Loss or fragmentation of habitat induces higher population substructuring in endangered species through reduced migration between remaining habitat fragments. Isolation of groups of individuals in this fashion is a major cause of inbreeding and can potentially lead to local extinction (Höglund 2009), which may be of particular consequence for species that are thought to disperse poorly, such as sawflies (Benson 1950).

Therefore, estimates of genetic diversity and levels of population substructure are commonly determined in population genetics and provide important information for conservation genetics studies (Väli *et al.* 2008). Principally, it is the genetic diversity at loci with functional importance that will affect the ability of a population to respond to selection. Thus, genetic variability measured from within these regions is useful for conservation (Höglund 2009; Väli *et al.* 2004). For example, information relating to allelic diversity at the *complementary sex determiner* (csd) locus, recently characterised in the honeybee *Apis mellifera*, would be particularly useful in assessing the reproductive fitness of sawfly populations in light of the consequences of CSD discussed in Section

1.2.3.3. However, in the absence of genome-wide information in a species of interest, molecular markers can be used as useful indicators of overall variability in the genome (Höglund 2009; Väli *et al.* 2004; Beebee and Rowe 2004) to facilitate estimates of population extinction risk.

1.4.2 Molecular markers

1.4.2.1 Available molecular markers

In early studies of genetic variation, allozyme (or isozyme) loci, assayed at the protein level via starch electrophoresis, were the marker of choice. Allozyme studies required large amounts of tissue and often the destructive sampling of individuals of the target species for conservation. In addition, it was suspected that a large amount of variation was not detected due to redundancy in the genetic code; therefore the use of this marker type has been superseded by DNA-based markers (Höglund 2009; Beebee and Rowe 2004).

RFLP (Restriction Fragment Length Polymorphism) (Botstein *et al.* 1980) was the first of the genomic DNA-based markers to be developed and RFLPs are considered to have marked the beginning of a new era in this field (Liu and Cordes 2004). The principle of RFLP is simple: genomic DNA is digested with restriction endonucleases and results in fragments whose number and size can vary between individuals, populations and species due to mutations in the restriction site. Traditionally, fragments are separated and analysed using Southern blots (Southern 1975) whereby digested genomic DNA is subjected to agarose gel electrophoresis, transferred to a membrane and visualised by hybridisation to specific probes. More recent uses of RFLP have replaced time-consuming Southern blot methods with PCR-based analyses, although the latter method requires sequence information for the target species (Liu and Cordes 2004).

One of the first methods that used the PCR-based technique was RAPD (Randomly Amplified Polymorphic DNA). With this method, short universal primers are used that randomly anneal to the target DNA and amplify the DNA between any two random primer pairs. If the primers anneal to the template DNA and the region between two primer pairs is short enough, an amplification product will be produced known as a RAPD profile for each individual sample. Genetic variation between individuals and

groups of individuals or populations is then assessed by the presence or absence of each product in a RAPD profile (Höglund 2009; Beebee and Rowe 2004; Liu and Cordes 2004). The RAPD method does not require any prior knowledge of genetic sequence of the target species. However, due to the low annealing temperature required to allow the universal primers to amplify, the reproducibility of this method is questionable (Liu and Cordes 2004).

By the mid-1990s (Vos *et al.* 1995), a more reliable descendant of the RAPD and RFLP methods was developed (Höglund 2009; Beebee and Rowe 2004) known as AFLP (Amplified Fragment Length Polymorphism). With AFLP, restriction enzymes are used to digest genomic DNA in the first instance. Complementary double-stranded adaptors are then ligated to the ends of the restriction fragments and PCR is performed using primers complementary to the adaptor sequences (Höglund 2009). The result is a series of PCR products forming an AFLP profile for each individual that can be analysed in a similar way to a RAPD profile.

In recent years microsatellites or Simple Sequence Repeats (SSRs) have become the marker of choice for population genetic studies (Squirrel *et al.* 2003; Zane *et al.* 2002). Microsatellites are tandemly repeated motifs of 1-6 bp sequences found in coding and non-coding regions of all prokaryotic and eukaryotic genomes analysed to date (Zane *et al.* 2002; Tautz 1989). These tandem repeat sequences are subject to a high level of single-motif insertion or deletion mutations resulting in a high level of polymorphism in the length of the repeat sequence (Metzgar *et al.* 2000). The length of a microsatellite repeat sequence can therefore vary between individuals, and populations, and it is these length variants (alleles) that are used to analyse genetic diversity (Figure 1.2).

The characteristically high level of polymorphism, along with the Mendelian co-dominant mode of inheritance, makes microsatellite markers a powerful tool for population genetic studies (Zane *et al.* 2002). However, the use of microsatellites requires a high amount of initial investment as each locus has to be identified and sequenced to facilitate the design of PCR primers. Only then can the alleles at each locus be identified following the determination of the size of the PCR products (Beebee and Rowe 2004; Liu and Cordes 2004). Microsatellites have the highest level of polymorphism of any of the

Allele 1 – 16 repeat units (48 bp)

TGACTGCTCACACACACACACACACACACACACACACACAGTTTCGGA
ACTGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCAAAGCCT

Allele 2 – 9 repeat units (34 bp)

TGACTGCTCACACACACACACACACACAGTTTCGGA
ACTGACGAGTGTGTGTGTGTGTGTGTGTCAAAGCCT

Allele 3 – 12 repeat units (40 bp)

TGACTGCTCACACACACACACACACACACACACAGTTTCGGA
ACTGACGAGTGTGTGTGTGTGTGTGTGTGTGTCAAAGCCT

Figure 1.2: Example of three alleles from a single microsatellite locus. Microsatellite repeat regions are shown in green and the microsatellite flanking regions are shown in red. Alleles are normally described by their length (repeat + flanking region) in base pairs (bp).

molecular markers due to the large number of alleles that can occur at each locus (Liu and Cordes 2004).

Although microsatellites are the common tool for population genetic analysis, Single Nucleotide Polymorphisms (SNPs) have been developed more recently as an alternative type of marker for population genetic studies (Coates *et al.* 2009). SNPs are single-base substitutions found at a single genomic locus. Theoretically within a locus a SNP can produce up to four alleles but typically SNPs are considered bi-allelic and are restricted to either the two pyrimidines (C/T) or the two purines (A/G) (Liu and Cordes 2004) due to the low probability of two independent base changes occurring at a single position (Vignal *et al.* 2002). In a genetic diversity study the lack of polymorphism at SNPs has to be compensated for by using a larger number of these markers, which in turn can be complicated to isolate (Höglund 2009).

1.4.2.2 Choosing a molecular marker

There is a range of molecular markers available with varying characteristics, and no single marker is suitable for all applications (Beebee and Rowe 2004). For population genetic studies it is important that the chosen marker is selectively neutral to ensure that the genetic structure observed within and between populations is solely due to the frequency of inbreeding and gene flow (Höglund 2009).

As a general rule, the marker chosen for a population genetic study is (1) polymorphic and evenly distributed throughout the genome, (2) provides adequate resolution of genetic differences, (3) has multiple, independent and reliable markers, (4) is simple, quick and inexpensive to use, (5) requires small amounts of tissue/DNA and (6) requires no prior information about the genome of the species in question (Agarwal *et al*. 2008). The choice of any DNA-based marker involves a trade-off between precision and convenience with expense being a significant factor (Agarwal *et al.* 2008; Sunnocks 2000).

The pros and cons of each popular marker type introduced in Section 1.4.2.1 (RFLP, RAPD, AFLP, SSRs and SNPs) are listed in Table 1.1. The drawback of all three of the random amplification techniques is that the subsequent analysis cannot distinguish between heterozygotes and homozygotes. So-called "dominant" markers are inherently less informative than "co-dominant" markers for population genetic studies as the vast majority of genetic diversity analyses rely on comparisons of heterozygosity levels within and between populations (Beebee and Rowe 2004). Therefore, the decision not to use RAPD, RFLP or AFLP analysis in the current study was made in spite of the fact that the use of these markers requires no information relating to the target DNA (Table 1.1). Furthermore, it was decided to use microsatellite markers as opposed to SNPs (both are technically difficult to isolate) due to the fact that microsatellite markers are characterised by a high degree of polymorphism and that fewer markers should be required in comparison to SNPs (Table 1.1). In addition, the utility of microsatellite markers in non-model species such as that in the current study is well-documented (Coates *et al.* 2009; Zane *et al.* 2002) whereas SNP discovery and genotyping is still a challenging endeavour (Höglund 2008; Liu and Cordes 2004).

Table 1.1: The benefits and disadvantages of five popular marker types in terms of their utility for population genetic studies.

Marker type	Dominant/Co-dominant	Pros	Cons
RAPD	Dominant	No knowledge of target DNA required. Relatively inexpensive to use. Straightforward system.	Sensitive to laboratory conditions. High quality template necessary. Poor reproducibility.
RFLP	Dominant	No knowledge of target DNA required. Considered more reliable and reproducible than RAPD. Straightforward system. Restriction sites abundant throughout the genome.	Can be labour-intensive. High concentrations of high quality DNA required. Relatively low polymorphism. To use the PCR-based method sequence information is required.
AFLP	Dominant	Considered more reliable and reproducible than RAPD. No knowledge of target DNA required. Amount and quality of DNA required less than that for RFLP.	Slightly more technically difficult than RAPD and RFLP. Can be labour-intensive.
Microsatellites (SSRs)	Co-dominant	Highly polymorphic. Reliable and reproducible. Relatively low number of markers required compared to RAPD/AFLP/RFLP/SNPs. Abundant within the genome.	Long development time. Labour-intensive. Can be expensive to establish. Specific primers required. Complicated mutation process. Questionable neutrality
SNPs	Co-dominant	Abundant within the genome. Simple mutation process as compared to microsatellites. Reliable and reproducible.	Low polymorphism. Questionable neutrality. High no. of markers required. Can be expensive to establish. Complex isolation procedure. High volume of sequence information required relating to target DNA.

1.4.3 Microsatellite markers

1.4.3.1 Properties of microsatellite markers

Microsatellite sequences can range in length up to a few hundred base pairs (Beebee and Rowe 2004) and are present in both coding and non-coding regions of the genome (Chistiakov *et al.* 2006). However, as compared to a random distribution pattern, microsatellites have been shown to be more abundant in the non-coding regions (Metzgar *et al.* 2000). In addition, the abundance and distribution of certain repeat types varies between genomic region and between taxa; in arthropods the most common repeat motif is $(CA)_n$ (Toth *et al.* 2000).

The high degree of polymorphism associated with microsatellite sequences arises due to insertion or deletion mutations resulting in the addition or removal of repeat units (Section 1.4.2.1). The origin of these mutations is still under debate (Zane *et al.* 2002). However, slipped-strand mis-pairing during DNA replication, which results in the replicated strand possessing a different number of repeats from the template strand, is thought to be the predominant mutational mechanism (Schlötterer and Tautz 1992). During DNA replication, longer sequences of repeated units will pose more of a problem to DNA polymerases than shorter sequences making longer sequences more prone to slipped-strand mis-pairing (Chistiakov *et al.* 2006) generating the higher levels of allelic diversity associated with longer repeat sequences (Primmer *et al.* 1996).

Mutation rates also vary between repeat type, with dinucleotide repeats exhibiting higher mutation rates and therefore being the most polymorphic of the repeat types (Lee *et al.* 1999; Chakraborty *et al.* 1997). However, variability in mutation rate is also present between repeat motifs. In a study comparing the relative mutation rates of dinucleotide repeat motifs in *Drosophila melanogaster* CA repeats appeared to have the highest mutation rate and AT repeats the lowest (Bachtrog *et al.* 2000). The authors suggest that this may be due to preferences in the DNA mismatch repair system and could also account for differences in the proportions of dinucleotide motifs represented in the genome between organisms. For example, in an AT-rich genome an adaptation of the mismatch repair system to AT mismatches would be selectively advantageous. If AT

mismatches are then repaired more efficiently the mutation rate in AT repeats will be reduced (Bachtrog *et al.* 2000).

Importantly, mutation via slipped-strand mis-pairing allows the same microsatellite allele to arise multiple times, resulting in size homoplasy (similarity of genes or traits for reasons other than co-ancestry). This can violate the basic assumptions of analysing genetic markers in that alleles are assumed to be derived from a common ancestor. Several mutational models that account for this homoplasy have been proposed. Determining which model is most appropriate is important because the estimation of genetic distances between populations based on microsatellite data relies on the underlying assumptions of the chosen model (Chistiakov *et al.* 2006).

1.4.3.2 Selective neutrality of microsatellite markers

As previously stated (Section 1.4.2.2) molecular markers for use in population genetics should ideally be selectively neutral to allow the researcher to determine that observations of genetic structure and diversity are due to levels of inbreeding and gene flow and are not biased by selection. Microsatellites are typically considered to be neutral markers, in which case no deviation from the expectations under a neutral model is expected by selection acting on a microsatellite itself (Chistiakov *et al.* 2006; Schlötterer *et al.* 1997). However, some research groups have linked microsatellites to functional properties (Chistiakov *et al.* 2006) and there have been reports of associations between allelic length variants and gene expression (reviewed in Kashi and King 2006), and therefore the potential for selection on microsatellites cannot be ignored (Väli *et al.* 2008).

However, as more researchers are beginning to use EST-SSRs (microsatellites derived from transcriptomic regions of the genome), due to the increasing amount of publicly available sequence data (Ellis and Burke 2007), the effects of selection on microsatellite analysis of population genetics have been considered in more detail. A study by Woodhead *et al.* (2005) showed a significant correlation in the rank order of population diversities in ferns as determined by EST-SSRs and genomic (anonymous) SSRs. Another study by Väli et al. (2008) showed that genetic diversity estimated by anonymous microsatellite markers and by sequence variation in non-coding regions of the genome were correlated at the population level. The results of these studies may indicate

that the effects of selection on microsatellites are negligible in some cases. Ellis and Burke (2007) suggest in their review that the effects of selection can best be minimized by increasing the number of markers utilised in population genetic studies and by using a common set of markers across taxa when working in a comparative manner.

1.5 Project Objectives

The aim of this study was to quantify levels of genetic diversity in populations of common farmland sawfly species and to assess the population substructure. Conservation research in graminivorous sawflies has been census-based to date and this study represents the first step into molecular research in the taxon. The four key objectives of the study were as follows:

1. Sample sawflies on a UK-wide scale in order to identify a suitable study species (Chapter 3).
2. Develop a set of microsatellite markers for use in the chosen study species (Chapter 4).
3. Determine the presence or absence of Complementary Sex Determination in the same species (Chapter 5).
4. Estimate the levels of genetic diversity present in populations of the chosen study species and gain an understanding of how these populations are structured (Chapter 6).

2. General Methods

2.1 Sawfly collection and processing

2.1.1 Collection of adult sawflies

Adult sawflies were collected using a Malaise trap (Plate 2.1). The Malaise trap, originally developed by Rene Malaise (1937) and later modified in 1972, is a form of neutral flight-interception trap (Townes 1972). In appearance the trap is similar to a ridge tent, but with open walls and a central barrier. Flying insects are directed upwards by the shape of the structure into a collecting vessel located at the highest point of the trap; the collection vessel contains 70% ethanol, which kills and preserves the trapped insects. The trap is designed to catch insects with an innate tendency to fly upwards upon encountering a barrier making it particularly useful for trapping the Hymenoptera or Diptera. Malaise traps generally have white roof panels and navy or black side panels; these neutral colours are chosen so that the trap does not specifically attract insects (Southwood 1978). This type of trap can be left unattended in the field for one to two weeks and is therefore both convenient and cost-effective to use.

Malaise traps are designed to be positioned across insect flight paths (Southwood 1978). All Malaise traps used in the following studies were positioned with the open sides of the trap facing into the grass margin and with the tallest side of the trap (containing the collection bottle) facing into the cropped area of the field. Collection bottles containing approximately 200 ml of 70% ethanol were replaced weekly and captured insects were taken back to the lab for storage and identification.

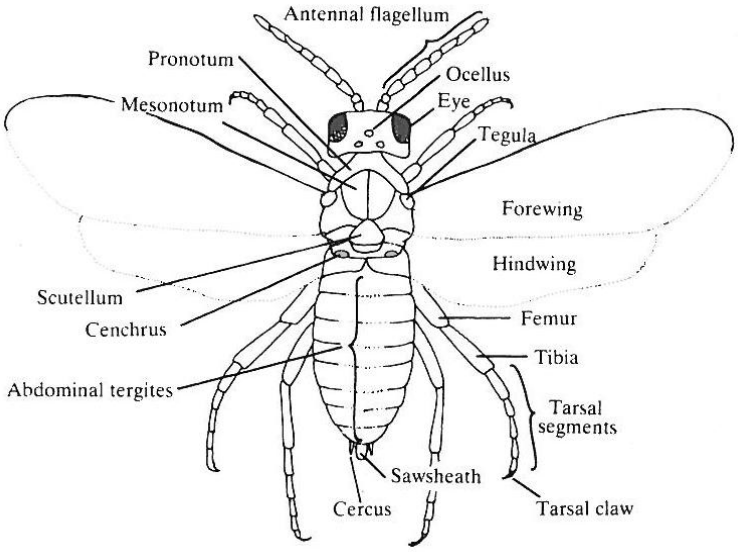
Plate 2.1: Malaise trap in side-view identical to those used in the study.

2.1.2 Identification of adult sawflies

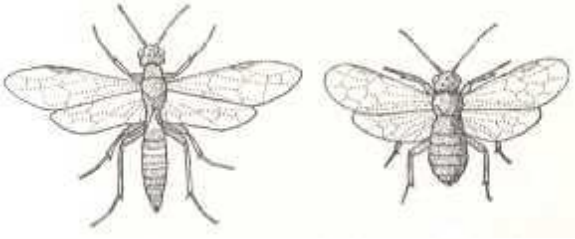
As the Malaise trap is not insect-specific, it was first necessary to separate sawflies from the total catch. The contents of the collection bottle (ethanol + insects) were emptied into a glass Petri dish for easier viewing. Hymenopteran insects were identified by the presence of two membranous pairs of wings with the hindwings usually smaller in size than the forewings. In flight, the forewings are linked to the hindwings by a series of interlocking hooks known as hamuli (Plate 2.2). When a dissection microscope is used to magnify the specimen (approx. ×10), this row of minute hooks is clearly visible on the front edge of the hindwings in all Hymenoptera and is the defining characteristic of the order (Chinery 1993).

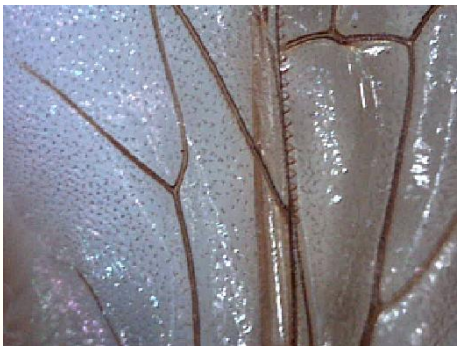
The Hymenoptera comprises two distinct sub-orders: the Apocrita, which includes the wasps, bees, ants, ichneumons and other groups; and the Symphyta, which contains only the sawflies. The Symphyta were segregated from all other hymenoptera due to the lack of a marked constriction between the thorax and the abdomen, otherwise known as a "wasp waist" (Plate 2.2). Also, with the exception of those in the family Cephidae, all adult sawflies possess a pair of structures known as cenchri (singular "cenchrus") (Plate 2.2) situated posterior to the scutellum. These structures come into contact with a scaly area on the underside of the forewings holding them in position when the insect is at rest (Wright 1990) and were easily recognisable with the naked eye.

Individuals were viewed under a dissection microscope and microdissection forceps and fine-tipped paintbrushes were used to manipulate the specimen to visualise characteristics more clearly. Sawflies were identified to genus level using the AIDGAP guide (Wright 1990). Subsequent identification to species was carried out using the relevant sections of "Handbooks for the Identification of British Insects" (Benson 1952; Benson 1958; Quinlan and Gauld 1981). Once identified, sawflies were preserved individually in vials of 70% ethanol and stored at 4 ºC.



A





B C

Plate 2.2: (A) Diagram of an adult female sawfly, note the position of the cenchri (Wright 1990). (B) left: an example of a hymenopteran with the characteristic "wasp-waist", right: a sawfly with no "wasp-waist" (www.bumblebee.org). (C) Row of hamuli on the front edge of the hind wing of a hymenopteran. (http://allencentre.wikispaces.com/Junior+Insects+in+Close-up).

2.1.3 Sexing of adult sawflies

Sexing of adult sawflies is relatively simple; in females the last segment of the abdomen is longitudinally divided on the ventral surface by the sawsheath containing the saws (Plate 2.3). The corresponding last segment of male sawflies has a smooth surface as the male insects lack saws and therefore have no sawsheath division (Wright 1990).

2.1.4 Collection of sawfly larvae

Sawfly larvae resemble the larvae of butterflies and moths (Lepidoptera) but sawfly larvae can be distinguished by the presence of at least six pairs of abdominal prolegs (Chinery 1993) (Plate 2.4).

Sawfly larvae were collected by sweepnetting the vegetation in the grass and wildflower margin of a cereal field at New Gilston Farm, Fife, Scotland (Grid Reference: NO 443 067). A sweep net made from thick cotton mesh with a round mouth of diameter c. 0.5 m was used to sweep the vegetation in the field margin. Any invertebrates present in the top half of the vegetation were knocked into the net as it was swept across the plant material. Sawfly larvae were collected throughout May and June in 2009, with sampling beginning a few weeks after the first observed emergence of adult sawflies.

Captured sawfly larvae were transferred to plastic 5ml vials and live insects were frozen at -80 ºC upon return to the laboratory.

Plate 2.3: The terminal ventral segments of an* Aglaostigma fulvipes *female (left) and male (right) showing the sawsheath of the female longitudinally dividing the terminal segments of the abdomen and the smooth undercarriage of the male.

Plate 2.4: Sawfly larva (*Dolerus *species) (GWCT 2010).

2.2 Preparation of sawfly DNA

2.2.1 Extraction of genomic DNA from sawfly tissue

Genomic DNA was extracted from the heads of adult sawflies using the DNeasy© Blood and Tissue Kit (QIAGEN Ltd.; Crawley, UK) following the manufacturers instructions for extracting from chitinous tissue. DNA was extracted only from the sawfly heads to minimise the presence of bacterial DNA in the sample, particularly bacteria from the gut.

Heads were excised from the bodies of individuals using a scalpel previously sterilised with 100% ethanol. The head(s) (including antennae) were placed into a 1.5 ml microcentrifuge tube containing 180µl Buffer ATL and the sample was homogenised using a micropestle. Following addition of Proteinase K (20 µl of 20 mg/ml), the sample was mixed by vortexing and incubated for 4 hours at 56 °C to ensure adequate lysis of the tissue. RNase A (4 µl of 100 mg/ml) was added to remove RNA and the sample was mixed by vortexing. Following incubation at room temperature for 2 minutes, the sample was vortexed, and Buffer AL (200 µl) was added and mixed again. An aliquot of 100% ethanol (200 µl) was added and the sample was thoroughly mixed to yield a homogeneous solution.

The sample was centrifuged at 5,900 rcf for 2 minutes to pellet the exoskeleton debris, which can interfere with the function of the spin columns in subsequent steps. The supernatant was transferred to a DNeasy© Mini spin column placed in a 2 ml collection tube and centrifuged for 1 minute at 5,900 rcf. The DNeasy© Mini spin column containing bound DNA was placed in a clean 2 ml collection tube and the DNA sample was washed by adding Wash Buffer AW1 (500 µl) to the column and centrifuging for 1 minute at 5,900 rcf. The flow-through was discarded and the wash step was repeated by the addition of Buffer AW2 (500 µl) to the column followed by centrifugation at 15,700 rcf for 3 minutes.

The DNeasy© Mini spin column was transferred to a clean 1.5 ml microcentrifuge tube for DNA elution. Buffer AE (150 µl) was added directly to the column membrane and the sample was incubated at room temperature for 1 minute prior to centrifugation at 5,900 rcf for 1 minute. A second elution step was performed with Buffer AE (50 µl) to increase yield. DNA samples were stored at -20 ºC until required.

2.2.2 Estimation of DNA concentration and quality

DNA fragments were separated by gel electrophoresis in 1% (w/v) agarose gel ("Hi-Pure" Low EEO agarose, BioGene Ltd.; Kimbolton, UK) suspended in 1 X TBE buffer (89 mM Tris-HCl pH 8.3, 89 mM Boric Acid, 2 mM EDTA). Gels were stained with SYBR® Safe DNA gel stain (Invitrogen Ltd.; Paisley, UK) at 1X concentration.

Concentration of DNA samples was determined by absorbance at 260 nm using a Nanodrop Microphotometer ND-1000 (Fisher Scientific UK Ltd.; Loughborough, UK).

2.2.3 PCR conditions

All Polymerase Chain Reactions (PCR) were carried out in a GeneAmp® PCR System 9700 thermocycler (Applied Biosystems; Warrington, UK). Taq DNA polymerase used in all reactions was sourced from Roche Applied Science (Burgess Hill, UK) supplied at 5U/µl and used according to the manufacturer's instructions. 10X PCR buffer used in all reactions was supplied with the Taq and consisted of 100 mM Tris-HCl, 15 mM $MgCl_2$ and 500 mM KCl, pH 8.3. Deoxyribonucleotides (dNTPs; Roche Applied Science; Burgess Hill, UK) were prepared as 2 mM stocks and final concentration in the reaction for each dNTP (dATP, dCTP, dGTP, dTTP) was 200 µM.

2.2.4 Purification of the PCR product

Purification of PCR products was performed using the MinElute® PCR Purification Kit (QIAGEN Ltd.; Crawley, UK) following the microcentrifuge protocol provided by the manufacturers.

The MinElute® Gel Extraction Kit (QIAGEN Ltd.; Crawley, UK) was used to isolate DNA fragments after electrophoresis through agarose gel. Fragments of gel containing PCR product of the required size were excised using a clean, sharp scalpel whilst viewing the gel under UV light. The products were isolated and purified following the manufacturer's instructions.

For both of these procedures, the product was routinely eluted in a final volume of 15µl, as opposed to the recommended 10µl, which did not adversely affect the samples and provided a larger sample volume for subsequent procedures.

2.2.5 Whole genome amplification

The illustra™ Genomiphi™ V2 DNA Amplification Kit (GE Healthcare; Buckinghamshire, UK) was used in some circumstances to amplify genomic DNA post extraction.

Due to the nature of the sawfly life cycle and the relative rarity of these insects, sample collection in the spring and summer months frequently yielded only small numbers of each species, which limited the availability of sawfly tissue for the development of molecular markers. Using the Genomiphi™ Kit to amplify genomic DNA allowed the molecular techniques to be optimised with only a small number of tissue samples.

Approximately 15 ng of sawfly DNA extracted using the DNeasy© Kit was used in one Genomiphi™ reaction (20 µl) following the manufacturer's instructions. Aliquots of the reaction could be used in downstream applications without the need for further purification.

3. Species composition of farmland sawfly (Hymenoptera, Symphyta) populations

3.1 Introduction

3.1.1 Introduction

The primary aim of the current study was to develop molecular markers to facilitate population genetic analyses in a common British farmland sawfly. In the first instance this required identifying a study species common to the arable landscape in the UK and is therefore most likely to be a food source for farmland bird species. Secondly, it was necessary to obtain enough individuals of the study species to optimise the stages involved in molecular marker development (Chapter 4).

In the agricultural landscape sawflies are most likely to be found in the grass and wildflower margins surrounding the cropping areas where suitable host plants are located. Therefore, these habitats were the focus of the field studies. Furthermore, as discussed in Section 1.2.2, the majority of British sawflies are only present in the adult form for one to two months of the year during the spring to summer period (Benson 1950). Sawfly larvae are present for a few weeks after the emergence of the adults (Benson 1950) but are notoriously difficult to identify due to the absence of clear taxonomic keys (Barker 1998). Therefore, to be sure of the identity of the individuals sampled, attention was focused on sampling adult sawflies.

Sampling was initially implemented in sites where sawflies were known to be present. Data from studentships funded by the Game and Wildlife Conservation Trust (GWCT) to study sawflies has highlighted such areas including Mains of Glamis Farm (north of Dundee) and New Gilston Farm in Fife. These sites, along with a well-established wildflower strip at the James Hutton Institute, provided an ideal starting point to concentrate sampling effort.

A flight interception trap is a simple but effective method of sampling flying adult insects. The Malaise trap was deemed suitable for this study as it should not attract one species more than any other and should therefore allow unbiased determination of the

most common species present at the sampling sites. The Malaise trap can be left unattended in the field for periods of 7–10 days, depending on weather conditions, allowing insect catches to be accumulated and then collected at regular intervals.

As discussed in Section 1.2.3 it is possible that farmland sawfly species sampled during the current study could exhibit Complementary Sex Determination (CSD). One of the primary indications that CSD is present in a species is an exceptionally high male-biased sex ratio (van Wilgenburg *et al.* 2006; Johns and Whitehouse 2004). For this reason, the sex ratio was documented for the most common species sampled as part of this study.

The sampling programme described in this chapter was designed, in the first instance, to identify a single study species and latterly to facilitate population genetic analysis of that species on a range of spatial scales. However, during the course of the sampling a wealth of data was accumulated regarding species composition at each sampling site. To make use of these data preliminary comparisons were drawn between sampling locations in terms of species composition and diversity.

3.1.2 Chapter summary

Sawflies were collected from three main localities in Scotland over three consecutive years (2008 to 2010). In the final year, sampling was extended to include three additional locations in England. A range of sawfly species was captured from all sites and successfully identified facilitating the selection of a single study species for molecular marker development. The sex ratios observed in the study species are discussed. In addition, tentative comparisons are drawn between sampling locations in terms of species composition and species diversity using commonly-used diversity indices.

3.2 Materials and Methods

3.2.1 Location of sampling sites

Sawflies were sampled during the spring/summer periods of 2008, 2009 and 2010 at a number of field sites. Three sites in Scotland were visited in each year of the study: Mains of Glamis Farm at the "Behind the Houses" field (NO 396478); New Gilston Farm at the "Harewheel" (NO 445068) and "Exercise" (NO 450070) fields, either singly or in combination; and The James Hutton Institute at the "mini-rotation" field (NO 337298) (Plate 3.1). In 2008 only, an additional Scottish site was used, which was a private grassland site near Old Meldrum (NJ 823279) (Plate 3.1).

During 2010 sampling was extended to include three further sites located in England: "Field B" at Down Farm near Dorset (SU 000147); a field on farmland belonging to Harper Adams University College (SJ 708208); and a field at Willoughby Farms in Claxby St Andrew (TF 452707) (Plate 3.1).

Plate 3.1: Map of Britain showing the positions of the Old Meldrum (A), Mains of Glamis (B), James Hutton (C), New Gilston (D) Claxby St Andrew (E), Harper Adams (F) and Down Farm (G) sampling sites.

3.2.2 Sampling strategy

3.2.2.1 Spring/summer 2008

One Malaise trap (Section 2.1.1) was positioned in the north grass/wildflower margin of the "Behind the Houses" field at Mains of Glamis Farm (Plate 3.2), one in the grass/wildflower margin adjacent to the "mini-rotation" field on farmland belonging to The James Hutton Institute (Plate 3.3) and one in the grass/wildflower margin surrounding the "Exercise" field at New Gilston Farm (Plate 3.4). One Malaise trap was positioned in roughly the centre of private grassland at the Old Meldrum site. All Malaise traps were erected 5th May 2008 and remained in place until they were dismantled on 14th July 2008.

3.2.2.2 Spring/summer 2009

In 2009 a sampling programme was again implemented at the three main sites in Scotland. In this sampling year, Malaise traps were repositioned on a weekly basis to a new random position within their respective sampling sites. Random positions were chosen by superimposing 10 m × 10 m grid-squares onto scaled drawings of the field margin area or beetle bank where sampling was to take place. Grid squares were numbered and a random grid square was chosen each week with the use of a random number generator. The Malaise trap was then moved to the flattest area within the designated 10 m × 10 m square. After a grid square had been used once it was not permitted to be used again following a sampling without replacement strategy (Southwood 1978).

Two Malaise traps were erected at Mains of Glamis, one in the grass margin on the northern edge of the "Behind the Houses" field and one in the margin on the southern edge of the field (Plate 3.2). One Malaise trap was erected at the James Hutton Institute in the grass margin adjacent to the "mini-rotation" field (Plate 3.3). One Malaise trap was erected at New Gilston Farm in the beetle bank crossing the "Harewheel field" (Plate 3.4) and two additional Malaise traps were erected in the margin surrounding the "Exercise" field (Plate 3.4). The "Exercise" field was split into two halves and one of the Malaise

traps was used to sample the margin surrounding the northern half of the field and the second to sample the margin surrounding the southern half of the field. All Malaise traps were erected on 21st April 2009, repositioned on a weekly basis as described, and taken down on 4th September 2009.

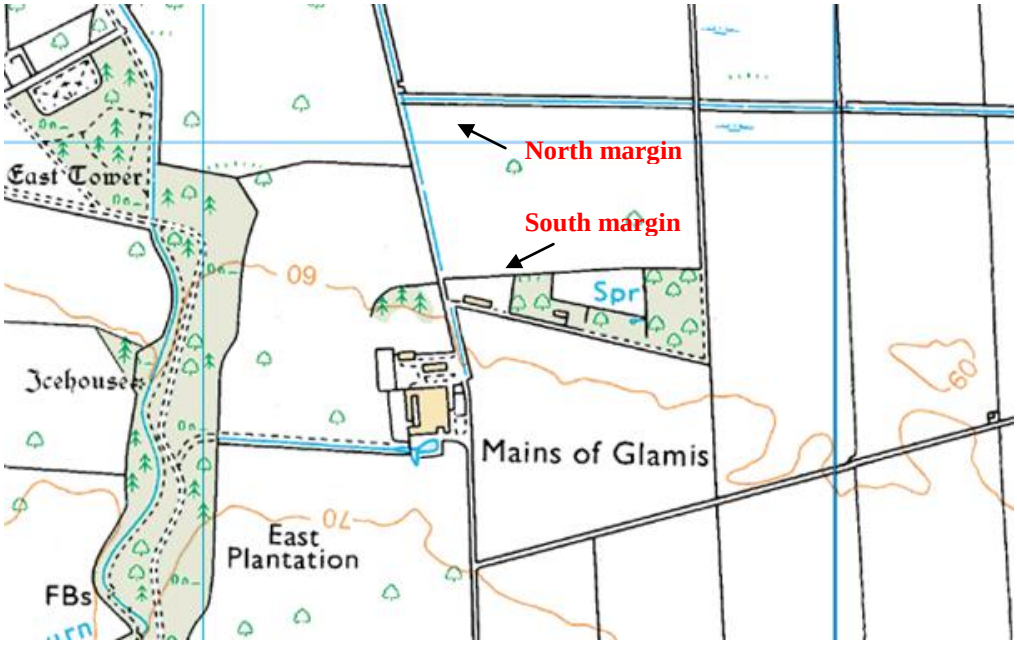


Plate 3.2: Section of Ordnance Survey© Map showing the "Behind the Houses" field at Mains of Glamis Farm and indicating the position of the north and south field margins.

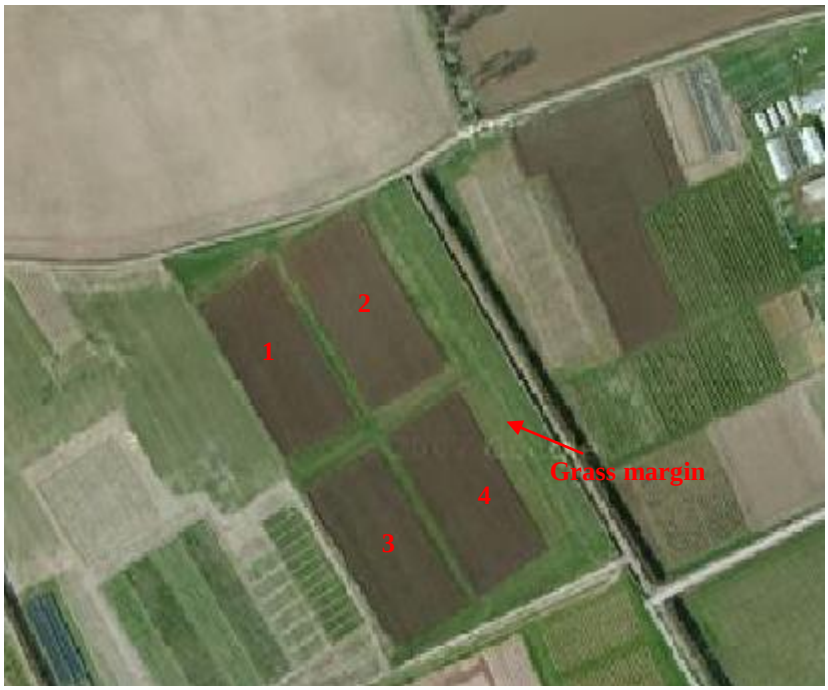


Plate 3.3: Aerial image of the "mini-rotation" field on The James Hutton Institute farmland indicating the positions of the four cropping quadrants and the grass margin where trapping took place.

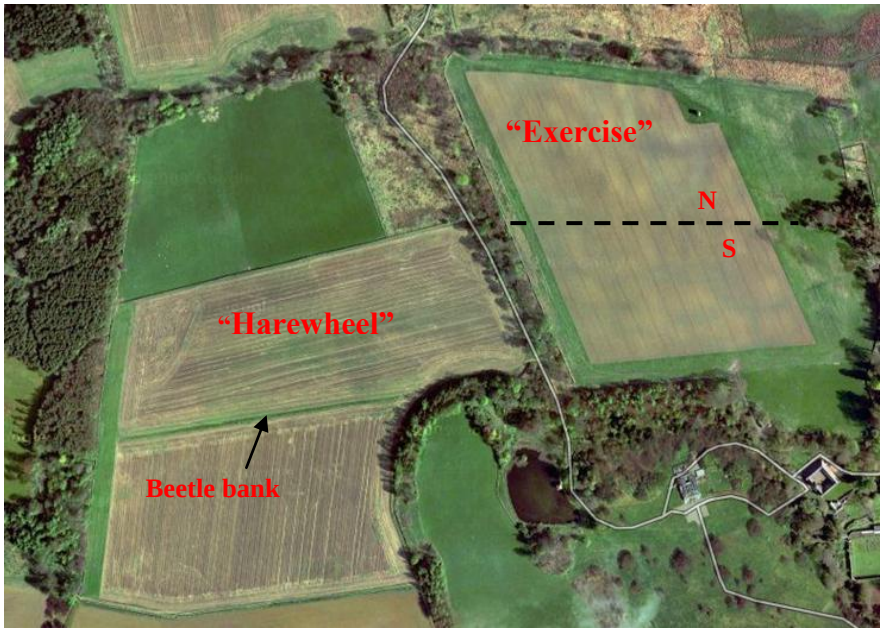


Plate 3.4: Aerial image showing the "Exercise" and "Harewheel" fields at New Gilston Farm. The dashed line indicates where the "Exercise" field was split into north and south sampling areas labelled N and S respectively. The Beetle bank used for sampling the "Harewheel" field in 2010 is shown.

3.2.2.3 Spring/summer 2010

In the third field season, the three main sites in Scotland were sampled again and sampling was extended to include three sites in England. During this season all Malaise traps remained stationary, without weekly repositioning.

Two Malaise traps were erected at the Mains of Glamis site, both positioned randomly in the northern margin of the "Behind the Houses" field (Plate 3.2). Two Malaise traps were deployed at the James Hutton Institute, both positioned randomly in the margin adjacent to the "mini-rotation" field (Plate 3.3). Two Malaise traps were erected at the New Gilston site; one was positioned in the northern half of the margin surrounding the "Exercise" field and the second in the southern half of the margin (Plate 3.4). Malaise traps were erected at these sites week beginning the 12th April 2010 and taken down on the 17th August 2010.

The three additional sites in England were as follows. One Malaise trap was positioned in an area of grassland adjacent to a cereal field at Willoughby Farms by Claxby St Andrew (Plate 3.5). One Malaise trap was erected in a grass margin adjacent to a set-aside field on land belonging to Harper Adams University College (Plate 3.6). Finally, a third Malaise trap was deployed in a grass field at Down Farm, Dorset (Plate 3.7). The Malaise traps at Claxby St Andrew, Harper Adams and Down Farm were in operation from 12th April to 17th August, 10th May to 1st August and 8th April to 11th June respectively.

Plate 3.5: Aerial view of the grassland where the Malaise trap was positioned at Willoughby Farm by Claxby St. Andrew.

Plate 3.6: Aerial image of the field at Harper Adams University College indicating the position of the margin where the Malaise trap was sited.

Plate 3.7: Aerial image showing the position of the field at Down Farm, Dorset where the Malaise trap was positioned.

3.2.3 Collection, identification and sexing of sawflies

Insects captured by the Malaise traps were collected on a weekly basis and returned to the laboratory for storage as described in Section 2.1.1. Sawflies were separated from the total catch and identified to genus and, where possible, to species as described in Section 2.1.2. The total number of individuals of each species collected by each Malaise trap was recorded. Individuals were sexed as described in Section 2.1.3 and deviation from a 50:50 male to female sex ratio, at each sampling site and in each sampling year, was analysed using a Chi-squared test. A Chi-squared test was performed only when the "expected value" for each sex was five individuals or more.

3.2.4 Characterisation of sampling sites

The Mains of Glamis and New Gilston sampling sites are representative of land under Integrated Farm Management practices whereby crop production is segregated from biodiversity by reducing weed abundance in the cropped areas and managing for biodiversity in the field margins (Hawes *et al.* 2010). The field used for sampling at The James Hutton Institute, although part of a long-term research project, is still representative of integrated farming approaches; the spraying of pesticides and other agro-chemicals is confined to the cropped area and the marginal areas are managed to maximise vegetation biodiversity.

Information relating to the cropping and general management practice at the sampling sites was only available for the sites in Scotland. Cropping records for the period 2005–2011 were made available for the "Behind the Houses" field at Mains of Glamis (Table 3.1), 2001–2010 for the James Hutton "mini-rotation" field (Table 3.2) and 2001–2010 for the "Exercise" and "Harewheel" fields at New Gilston (Table 3.3). The cropping history of the fields at Mains of Glamis and New Gilston was dominated by cereals (particularly winter barley and wheat), which typifies many cropping systems in the lowland arable areas of Eastern Scotland (Hawes *et al.* 2010). The James Hutton site exhibited a more diverse history of crop types, including brassicas, potatoes and grass/fallow, due to the nature of the experimental field (i.e. a mini-rotation). Set-aside was only present in the cropping history of the New Gilston field sites.

The “Behind the Houses” field at Mains of Glamis has a total area of 12.98 Ha with an 11.85 Ha average working area (91.3%) per year. The “mini-rotation” field at The James Hutton Institute measures 2.8 Ha in total with 2 Ha per year working area (0.5 Ha per quadrant) (71%). At New Gilston the “Exercise” field measures 15.2 Ha and the “Harewheel” field 16.7 Ha with an average of 9.4 Ha (61.8%) and 13.0 Ha (77.8%) working area per year respectively.

Table 3.1: Cropping records for the "Behind the Houses" field at Mains of Glamis Farm for the period 2005-2011 inclusive.

Mains of Glamis - "Behind the Houses"		
Year	**Crop**	**Working area (Ha)**
2005	Winter Wheat	12.67
2006	Broccoli	12.67
2007	Spring Oats	11.10
2008	Winter Wheat	11.10
2009	Potatoes	Not available
2010	Winter Wheat	11.70
2011	Winter Barley	Not available

Table 3.2: Cropping records for the "mini-rotation" field at The James Hutton Institute for the period 2001-2011 inclusive.

The James Hutton Institute - "mini-rotation" field				
	Quadrant (0.5 Ha each)			
Year	**1**	**2**	**3**	**4**
2001	Spring Barley	Potatoes	Grass	Oilseed Rape
2002	Oilseed Rape	Winter Barley	Potatoes	Grass
2003	Grass	Oilseed Rape	Winter Barley	Potatoes
2004	Potatoes	Grass	Rape and Swedes	Spring Barley
2005	Spring Barley	Potatoes	Grass	Brassica
2006	Potatoes	Spring Barley	Swedes	Fallow
2007	Fallow	Potatoes	Spring Barley	Swedes
2008	Swedes	Fallow	Potatoes	Spring Barley
2009	Spring Barley	Swedes	Fallow	Potatoes
2010	Potatoes	Spring Barley	Swedes	Fallow

Table 3.3: Cropping records for the "Exercise" and "Harewheel" fields at New Gilston Farm for the period 2001-2010 inclusive.

New Gilston			
Field	**Year**	**Crop**	**Working area (Ha)**
Exercise	2001	Spring Barley	9.97
	2002-03	Winter Barley	9.97
	2004	Spring Barley	8.70
	2005	Spring Barley	8.70
	2006	Spring Barley	8.70
	2006-07	Winter Barley	9.70
	2007-08	Set-aside	9.70
	2008-09	Winter Wheat	9.70
	2010	Spring Oats	9.40
Harewheel	2001	Set-aside	Not available
	2001-02	Winter Barley	15.46
	2003	Spring Barley	7.07
	2003-04	Winter Barley	15.31
	2004-05	Set-aside	13.47
	2005-06	Winter Barley	14.59
	2006-07	Winter Barley	14.59
	2007-08	Spring Barley	14.59
	2008-09	Winter Oats	8.40
	2009-10	Winter Wheat	13.46

3.2.5 Quantifying sampling effort

For each sampling site and year, the total sampling effort was calculated. Sampling effort was quantified in terms of sampling units, where one sampling unit was equal to one full day of one Malaise trap being in continuous operation within a field. Due to the susceptibility of Malaise traps to high winds and the fact that some traps were maintained by others on our behalf, sampling effort was quantified post-hoc.

In 2008, the Malaise traps at all four sampling sites (Mains of Glamis – North Margin, The James Hutton Institute, New Gilston "Exercise" field and Old Meldrum) were in continuous operation for the period 5th May to 14th July, corresponding to a cumulative total of 70 sampling units per field.

In 2009, all six Malaise traps were in operation for the period 21st April to 4th September potentially corresponding to a cumulative total of 136 sampling units per trap. However, bad weather caused some of the traps to collapse at points during the sampling season resulting in loss of sampling units. Data for Malaise traps positioned in the same field or "locality" were combined, namely those at Mains of Glamis ("Behind the Houses" field; North and South Margins) and at New Gilston ("Exercise" field; North and South trapping areas). In total, 265 sampling units were accumulated for Mains of Glamis (North and South margins), 134 for The James Hutton Institute, 265 for New Gilston "Exercise" field (North and South sampling areas) and 129 for New Gilston "Harewheel" field (Table 3.4).

In 2010, six Malaise traps were erected at three localities in Scotland on varying days of the week beginning 12th April (Mains of Glamis, The James Hutton Institute and New Gilston "Exercise") and taken down 17th August. However, on the 7th July 2010 high winds caused the collapse of and damage to one of the Mains of Glamis traps and to both of the James Hutton traps. To compensate for this, the trap originally positioned in the north sampling area of the New Gilston "Exercise" field was relocated to the James Hutton site to allow sampling to continue with one Malaise trap at each of the three localities. In total, 205 sampling units were accumulated at Mains of Glamis, 208 at the James Hutton and 203 at New Gilston "Exercise" (Table 3.4). No sampling units were lost from any of the sites in England resulting in a total of 126 sampling units accumulated at Claxby St. Andrew, 82 at Harper Adams and 63 at Down Farm (Table 3.4).

Table 3.4: Total sampling effort per trap and locality for sampling years 2009 and 2010.

Year	Site	Field and trap position	Sampling effort per trap	Total sampling effort per locality
2009	Mains of Glamis	"Behind the Houses" (North margin)	129	265
	Mains of Glamis	"Behind the Houses" (South margin)	136	
	The James Hutton	"Mini-rotation"	134	134
	New Gilston	"Exercise" (North trapping area)	129	265
	New Gilston	"Exercise" (South trapping area)	136	
	New Gilston	"Harewheel" (Beetle bank)	129	129
2010	Mains of Glamis	"Behind the Houses" (North margin)	122	205
	Mains of Glamis	"Behind the Houses" (North margin)	83	
	The James Hutton	"Mini-rotation"	124	208
	The James Hutton	"Mini-rotation"	84	
	New Gilston	"Exercise" (North trapping area)	81	203
	New Gilston	"Exercise" (South trapping area)	122	
	Claxby St. Andrew	grassland adjacent to cereal field	126	126
	Harper Adams	margin adjacent to set-aside field	82	82
	Down Farm, Dorset	grass field	63	63

3.2.6 Species diversity measures

Every biological community has an attribute known as "species diversity" and there have been many different suggestions for measuring this accurately (Southwood 1978; Krebs 1989). Methods range from looking at the "species richness" of a community (simply the number of species present) to the use of different species diversity indices which take into account both the number of species present and the abundance of these species (Speight *et al.* 2008). There is much discussion over which of the diversity indices is the best and caution is urged in the use of all diversity indices (Magurran 2004; Krebs 1989). Therefore, for the data presented here two of the most widely-used indices of diversity (Shannon-Wiener and Simpson's) were calculated in addition to species richness.

3.2.6.1 Species richness

In the current study species richness is defined as the number of sawfly species present in a locality. A rarefaction method was employed to compare species richness between localities, within each sampling year, to take into account differences in the number of individuals sampled. The following rarefaction algorithm, originally proposed by Sanders (1968) and later modified by Hurlbert (1971) and Simberloff (1972), was used (Krebs 1989) and computed using the online rarefaction calculator provided by the University of Alberta (http://www.biology.ualberta.ca/jbrzusto/rarefact.php):

$$E(S_n) = \sum \left[1 - \frac{\binom{N - Ni}{n}}{\binom{N}{n}} \right]$$

(Eq. 1)

where $E(S_n)$ = Expected number of species in a random sample of n individuals

S = Total number of species in the entire collection

N_i = Number of individuals of species i

N = Total number of individuals in collection = ΣN_i

n = Value of sample size chosen for standardisation ($n \leq N$)

$\binom{N}{n}$ = number of possible combinations of n individuals from a set of N = $N!/n!(N-n)!$

3.2.6.2 Shannon-Wiener diversity index

The commonly-used Shannon-Wiener function was used to compare species diversity between localities. This function is sensitive to the number of rare species in the community (Krebs 1989). Due to the differences in sampling strategy between years, the data for different sampling years were analysed separately for each locality. The Shannon-Wiener function is as follows (Krebs 1989):

$$H = -\sum (p_i)(\log_2(p_i)) \qquad \text{(Eq.2)}$$

where H = the information content of the sample, i.e. the index of species diversity

p_i = proportion of total sample belonging to the ith species

3.2.6.3 Simpson's diversity index

Simpson's diversity index was also used to estimate species diversity between localities. In contrast to the Shannon-Wiener function, this index is more sensitive to the abundance of the common species in a community (Krebs 1989). Again, to compensate for the differences in sampling strategy between years, the data for different sampling years were analysed separately for each locality. To estimate the complement of Simpson's original measure the following equation was used (Krebs 1989):

$$1\text{-}D = 1 - \Sigma (p_i)^2 \qquad \text{(Eq. 3)}$$

Where 1-*D* = index of diversity

p_i = proportion of individuals belonging to species *i* in the community

3.3 Results

3.3.1 Rank abundance

Tenthredopsis excisa, *T. nassata* and *Dolerus aeneus* were consistently present amongst the three or four most abundant sawfly species in Malaise trap catches at the Scottish sites in 2008, 2009 and 2010 (Figures 3.1, 3.2 and 3.3). In addition, relatively high numbers of *Pachyprostasis rapae* were collected at Mains of Glamis in 2009 and 2010 and of *Ametastegia equiseti* at New Gilston "Exercise" in 2010. A similar dominance by these species (*T. excisa*, *P. rapae* and *D. aeneus*) was observed at the Harper Adams site in 2010, but these species were far less dominant at the other two English sites (Figure 3.4). Interestingly, at Down Farm, six species of a single genus (*Dolerus*) were collected, with *Dolerus picipes* being the most abundant (Figure 3.4). Dominance by a single genus was not observed at any of the other localities in any sampling year. Furthermore, at Claxby St Andrew neither *Tenthredopsis* nor *Dolerus* were represented in the three most common species.

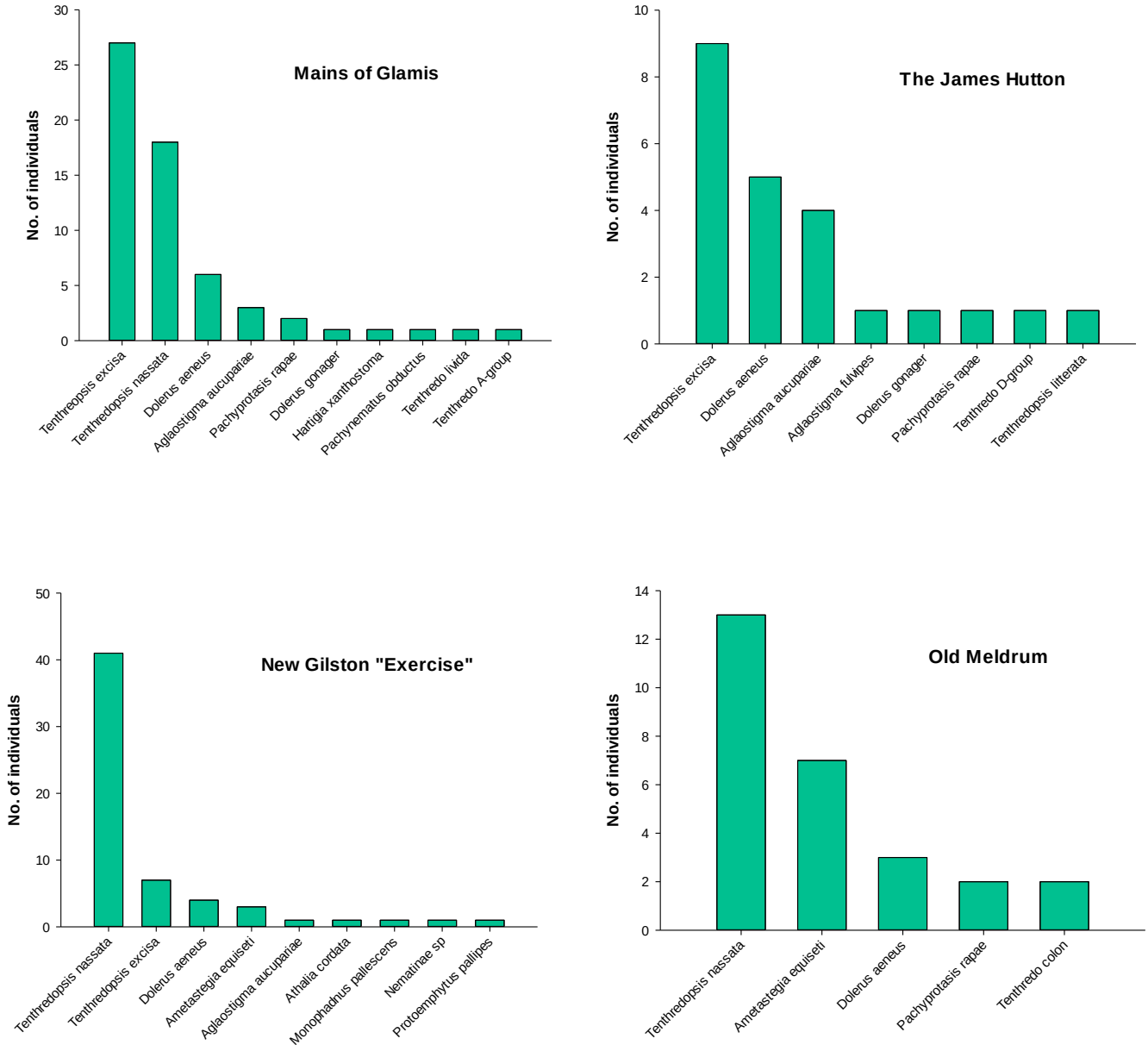


Figure 3.1: Rank abundance of sawfly species sampled at four localities in 2008.

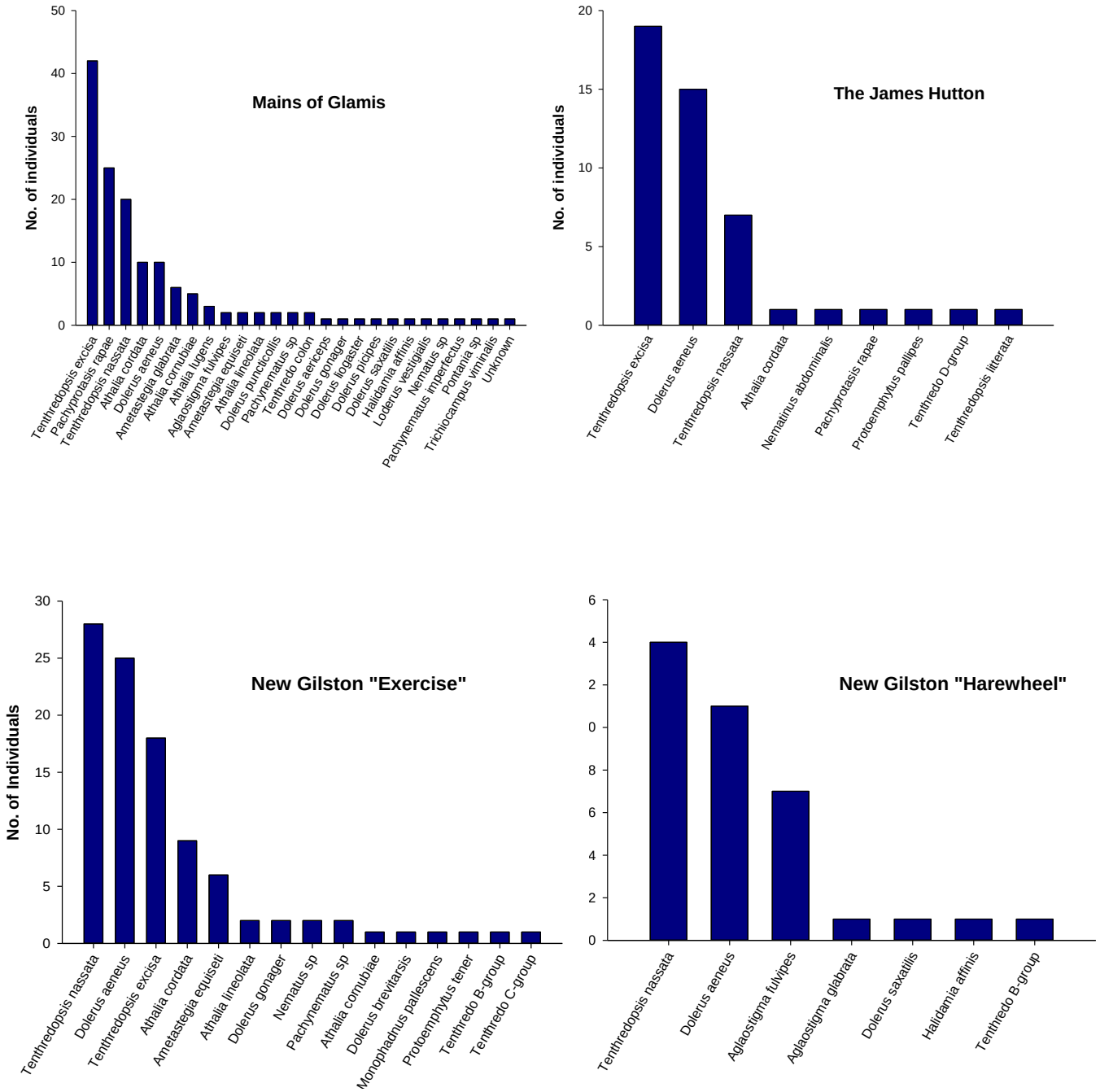


Figure 3.2: Rank abundance of sawfly species sampled at four localities in Scotland in 2009.

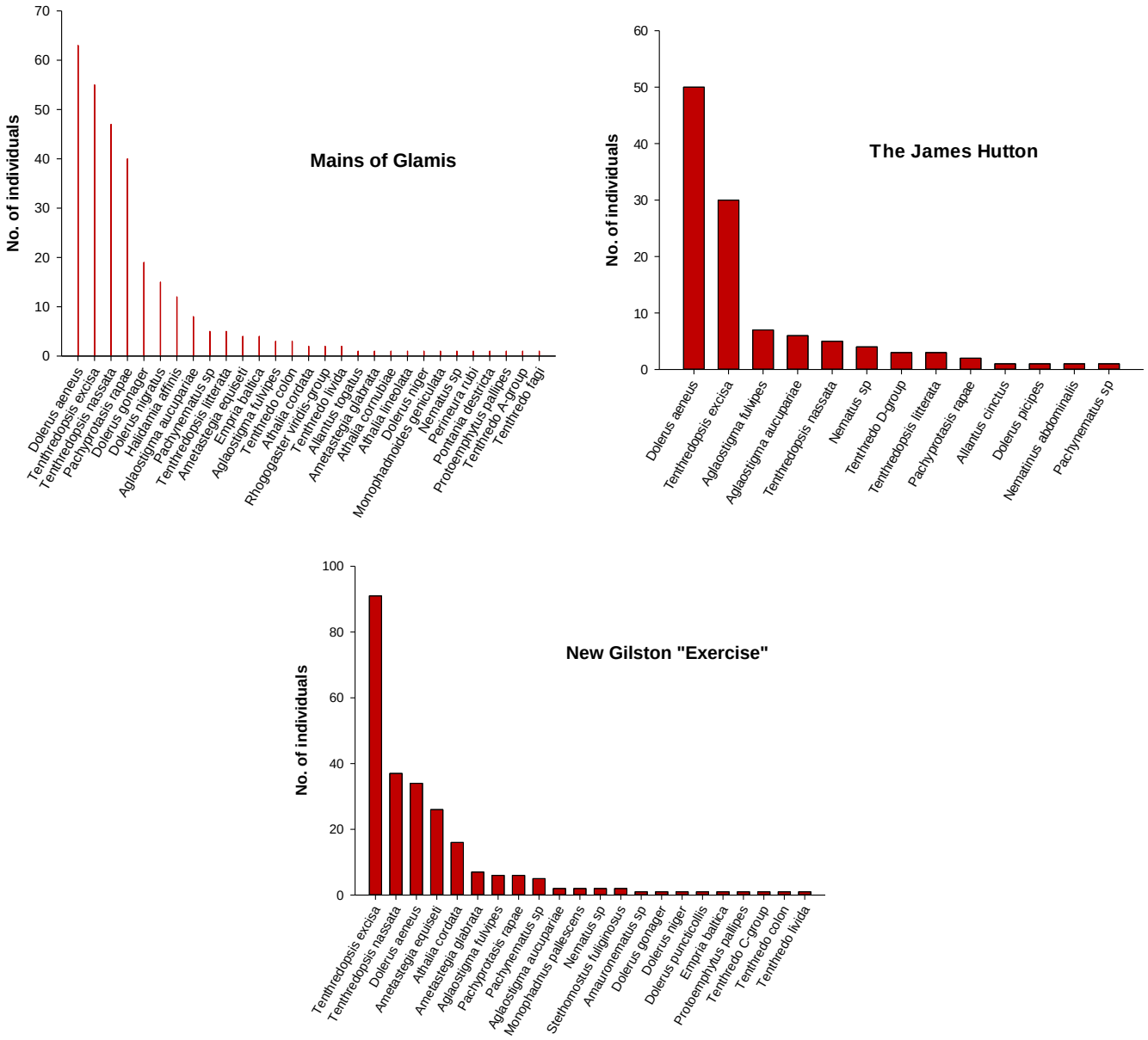


Figure 3.3: Rank abundance of sawfly species sampled at three localities in Scotland in 2010.

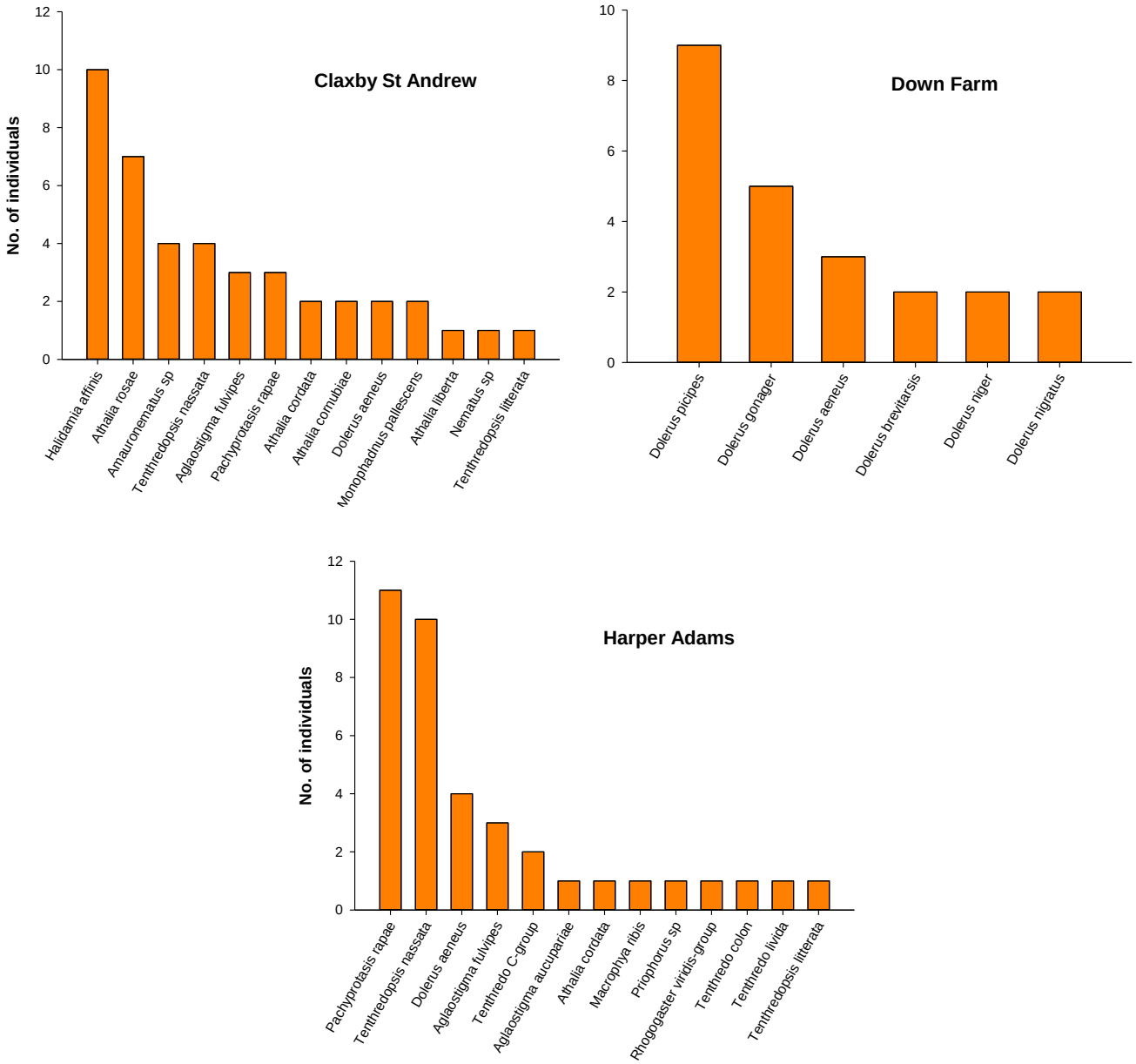


Figure 3.4: Rank abundance of sawfly species sampled at each of three localities in England in 2010.

3.3.2 Sex ratios of common species

The most common species sampled across all localities and sampling years were *Tenthredopsis excisa*, *Tenthredopsis nassata* and *Dolerus aeneus* (Section 3.3.1). For each of these species, in all cases where a Chi Square test was permissible, a significant deviation from a 50:50 male:female sex ratio was detected (Figures 3.5, 3.6 and 3.7).

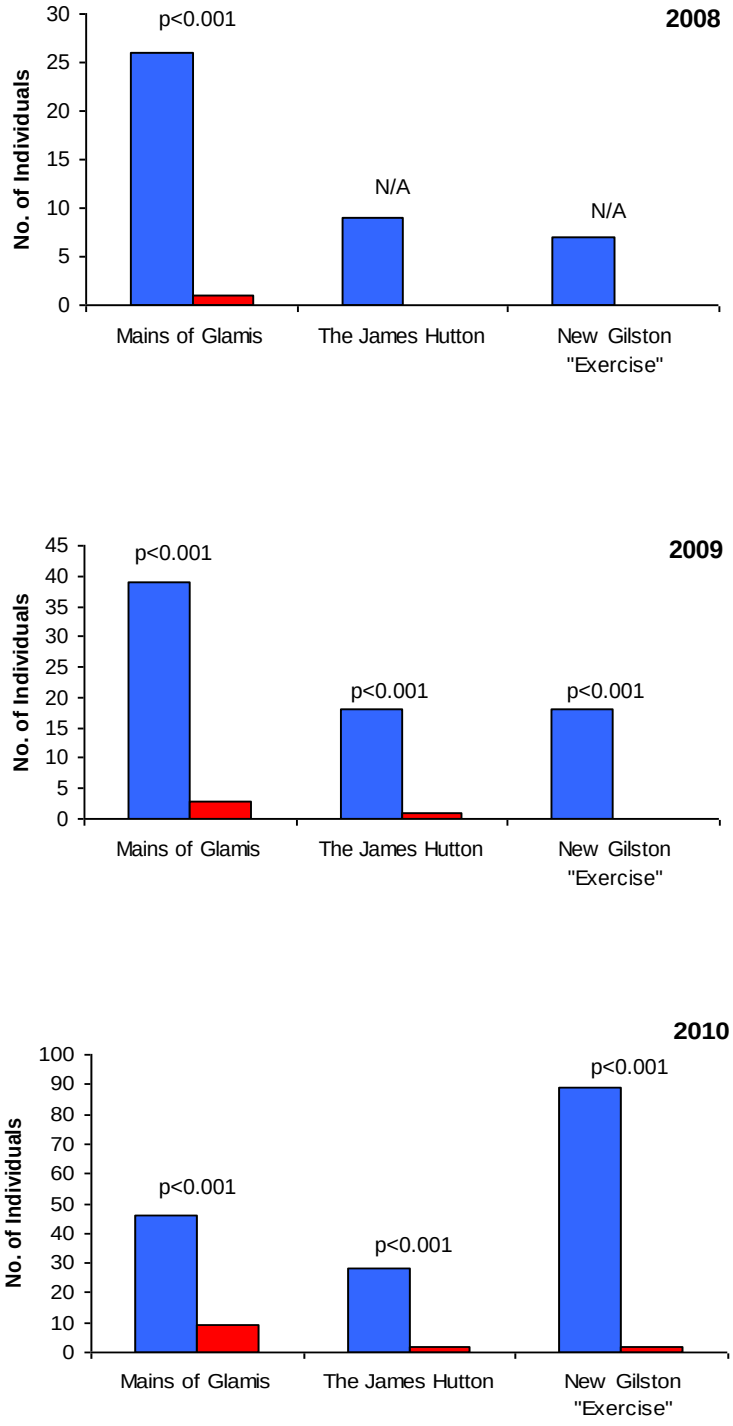


Figure 3.5: Numbers of male (blue bars) and female (red bars) individuals of* Tenthredopsis excisa *collected at each of the Scottish sites in 2008–2010. Missing data for some localities indicates that no individuals of this species were sampled. Outcome of Chi Square test shown where applicable, NS = not significant.

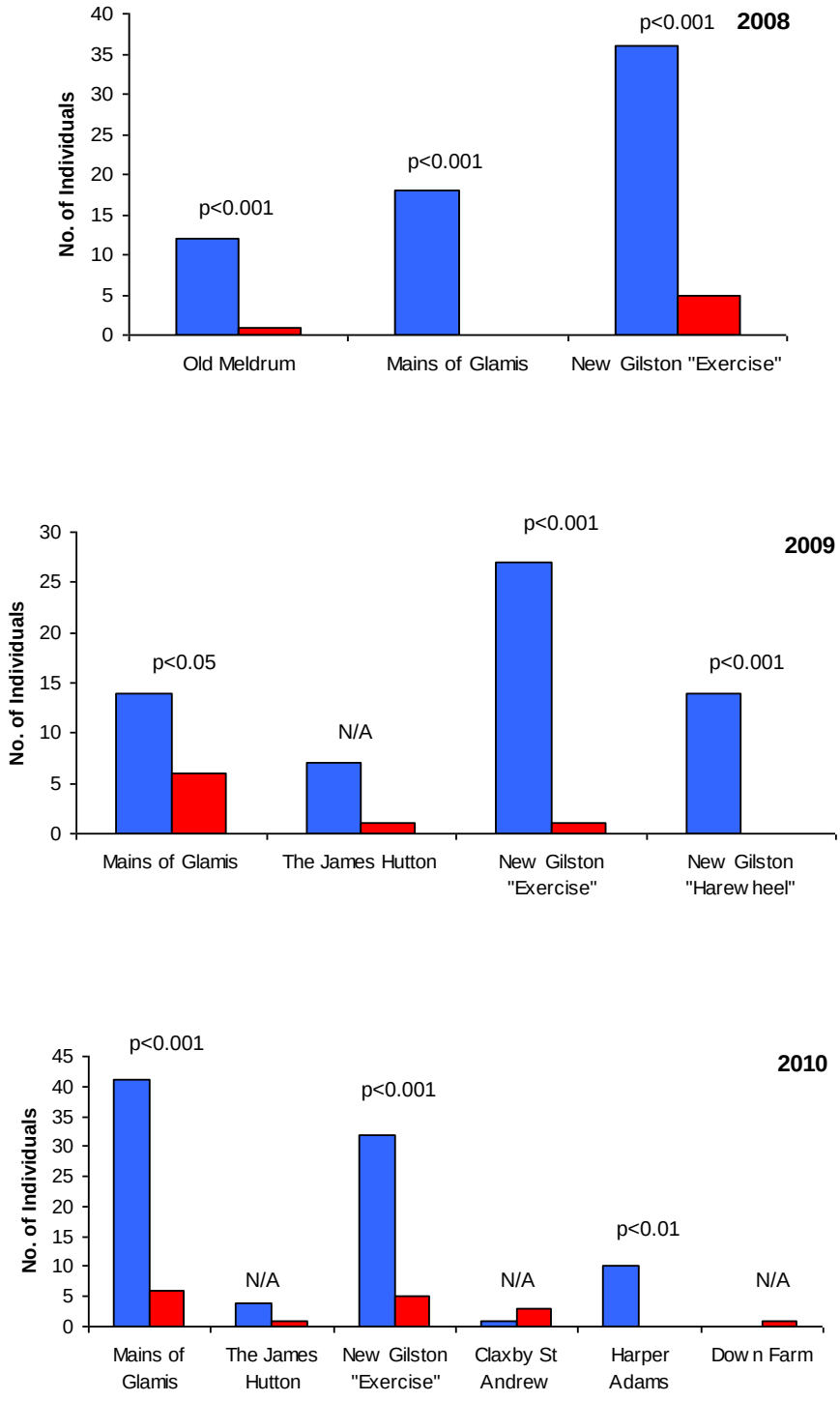


Figure 3.6: Numbers of male (blue bars) and female (red bars) individuals of* Tenthredopsis nassata *collected at each site in 2008–2010. Missing data for some localities indicates that no individuals of this species were sampled. Outcome of Chi Square test shown where applicable, NS = not significant.

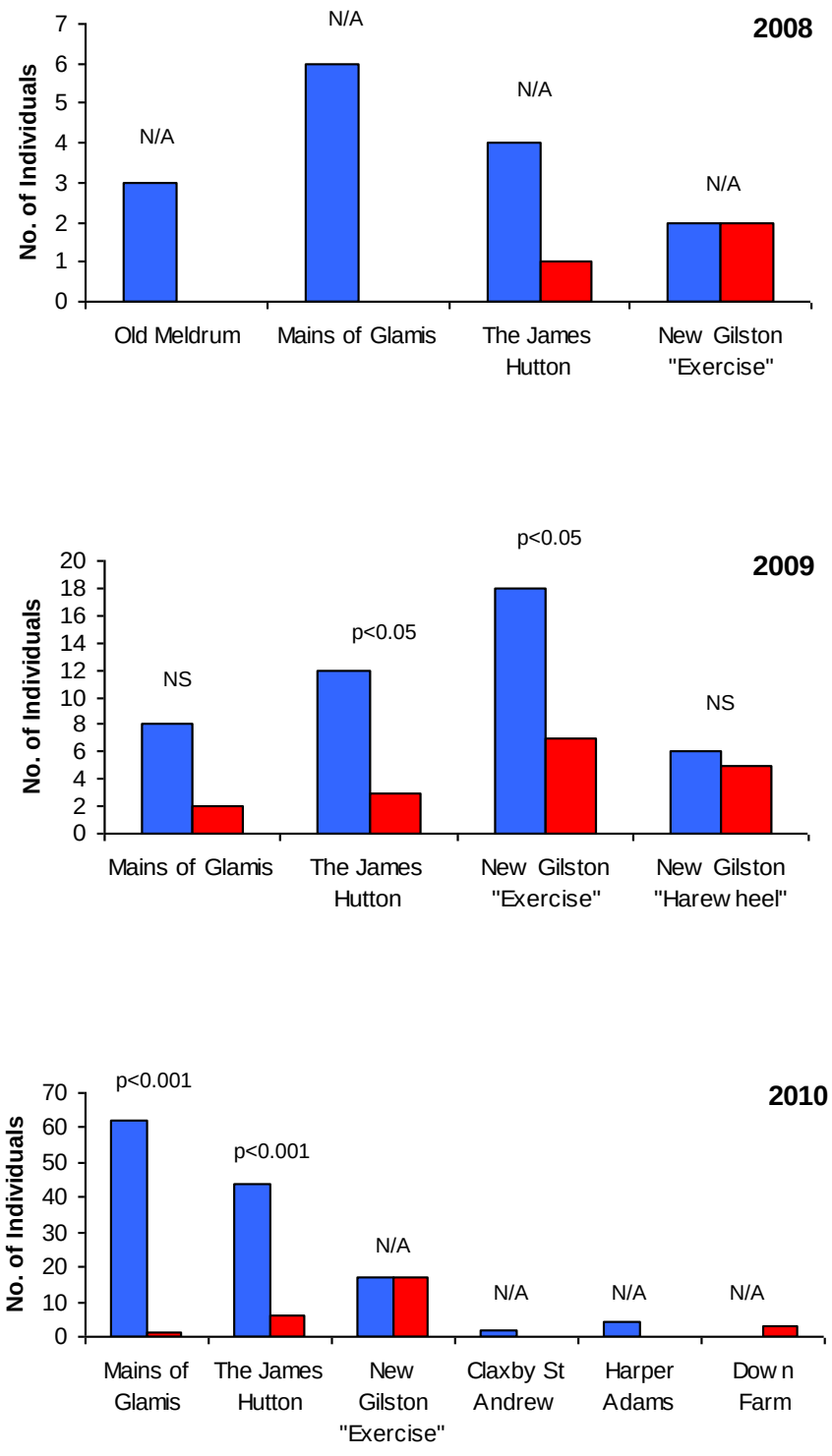


Figure 3.7: Numbers of male (blue bars) and female (red bars) individuals of* Dolerus aeneus *collected at each site in 2008–2010. Missing data for some localities indicates that no individuals of this species were sampled. Outcome of Chi Square test shown where applicable, NS = not significant.

3.3.3 Species diversity

3.3.3.1 Species richness

Across all three sampling years species richness did not appear to differ markedly between localities (Figure 3.8). In 2008, species richness was highest at The James Hutton Institute and lowest at New Gilston. In 2009, Mains of Glamis exhibited the highest level of species richness and New Gilston "Harewheel" the lowest. In 2010, species richness was highest at Claxby St. Andrew and lowest at Down Farm.

3.3.3.2 Heterogeneity measures, Shannon-Wiener and Simpson's

The Shannon-Wiener (*H*) and Simpson's (*1-D*) indices of diversity were calculated as described in Section 3.2.6 for all localities and sampling years. The diversity values produced from the indices were closely correlated with each other (Pearson product-moment correlation coefficient; $r = 0.914$, $p<0.01$) and with the values for species richness (Shannon-Wiener, $r = 0.897$, $p<0.001$; Simpson's, $r= 0.833$, $p<0.001$ respectively) lending weight to any observed differences between sampling localities.

In 2008, The James Hutton Institute was ranked as the most diverse by both indices whereas in 2009 and 2010 the New Gilston "Exercise" locality was ranked as most diverse (Table 3.5). Of the three localities sampled in England, Claxby St Andrew was ranked as the most diverse by both indices (Table 3.5). In all but one case the Shannon-Wiener function ranked the localities in the same order as the Simpson's index; in 2009 The James Hutton Institute and New Gilston "Harewheel" localities were ranked the opposite way round by each index (Table 3.5).

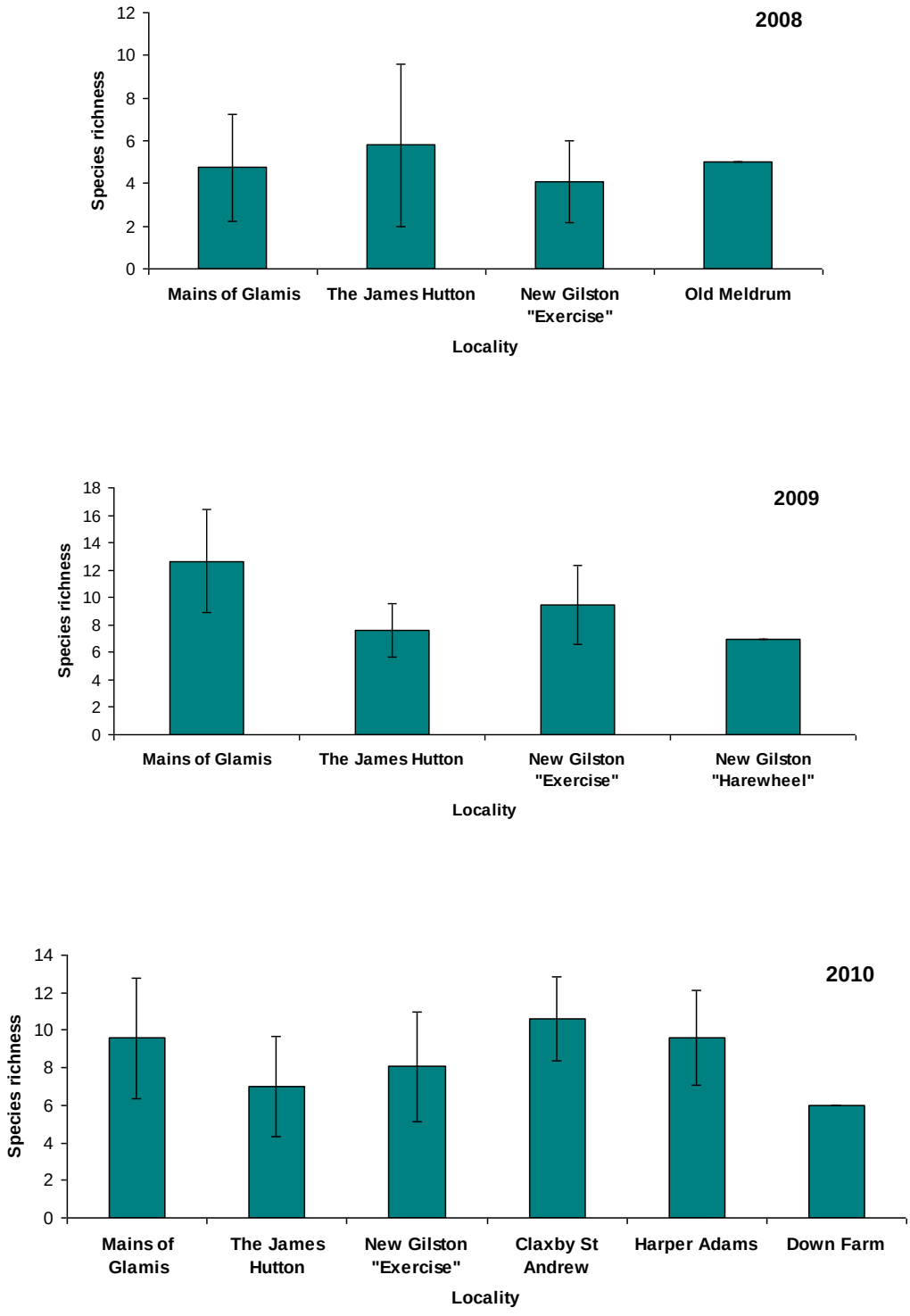


Figure 3.8: Species richness at each site in 2008–2010. 95% confidence intervals are given for species richness values that were calculated by rarefaction.

Table 3.5: Shannon-Wiener (H) and Simpson's (1-D) indices of diversity. For each year of the study, the localities are ranked from 1-4 (1 being the most diverse) according to the H index. *Note that Simpson's 1-D ranked these two localities the other way round.

Year	Site	H	1-D	Rank
2008	The James Hutton	2.431	0.760	1
	Mains of Glamis	2.231	0.702	2
	Old Meldrum	1.921	0.678	3
	New Gilston Exercise	1.706	0.511	4
2009	Mains of Glamis	3.460	0.853	1
	New Gilston "Exercise"	2.866	0.813	2
	The James Hutton*	2.172	0.710	3
	New Gilston "Harewheel"*	2.086	0.715	4
2010 (Scotland)	Mains of Glamis	3.505	0.871	1
	New Gilston Exercise	2.980	0.802	2
	The James Hutton	2.485	0.727	3
2010 (England)	Claxby St Andrew	3.336	0.876	1
	Harper Adams	2.984	0.821	2
	Down Farm	2.311	0.760	3

3.4 Discussion

3.4.1 Tenthredopsis excisa *identified as a study species*

Tenthredopsis excisa was the dominant species at all of the main localities sampled in Scotland in 2008 at the beginning of the study, suggesting that this species could contribute significantly to the sawfly diet of farmland birds in the region. Thus, attention was initially focused on *T. excisa* for molecular marker development (Chapter 4).

3.4.2 Male-biased sex ratios observed in common farmland sawfly species

Extreme male-biased sex ratios were observed at most or all of the localities for the three most common species detected in the current study (*Tenthredopsis excisa*, *Tenthredopsis nassata* and *Dolerus aeneus* respectively) across all sampling years. There are several explanations for this. It is possible that the male-biased sex ratios observed here are simply consequences of male preference for the position of the Malaise traps and therefore males were caught more frequently. However, in 2009 Malaise traps were moved around the sampling sites in order to compensate for any bias in the preference of a particular species for a particular area of the site. By proxy, this should minimise any bias induced by male preference for a particular area. In addition, it is possible that the male-biased sex ratios observed in the current study could be attributed to the fact that males of most insect species exhibit higher activity levels than that of females (Speight 2008) and were therefore represented in higher numbers in the Malaise trap catches regardless of sampling regime. The difference in activity levels between males and females could be accounted for by sampling using a method such as sweep netting which captures stationary insects from vegetation. However, this method of sampling is time-consuming and labour-intensive and was therefore not used in the current study due to time constraints.

It is also possible that variation in resource quality, insect population structure, and differential mortality of the sexes can contribute to biased sex ratios (Godfray 1994). This is a consequence of the fact that Hymenopteran females typically have control over fertilisation of their eggs, facultatively adjusting the sex ratio of their progeny (Heimpel

and de Boer 2008; van Wilgenburg *et al.* 2006). For example, females of *Athalia rosae* (the turnip sawfly) allocate the sex of their progeny in response to growth of the host plant, with male-biased sex ratios prevailing in populations where plant growth is slow and female-biased ratios where plant growth is rapid (Craig *et al.* 1992). At the present time there are no studies which document such behaviour in *T. excisa*, *T. nassata* or *D. aeneus.*

In addition, genetic factors could also explain male-biased sex ratios. The presence of Complementary Sex Determination (CSD; Section 1.2.3), combined with conditions that promote inbreeding, can lead to the production of diploid males. Additional (diploid) males in a population could cause the sex ratio to become male-biased. The operation of CSD in the study species is investigated in Chapter 5.

3.4.3 Species diversity at three localities in Scotland ranked in the same order across two sampling years

Two commonly-used diversity indices, Shannon-Wiener and Simpon's, ranked the three main localities in Scotland in the same order in 2009 and 2010 (Section 3.3.3.2); Mains of Glamis was ranked the most diverse in terms of sawfly species diversity followed by New Gilston "Exercise" and then The James Hutton Institute. It is difficult to explain why this is so with the limited amount of information available relating to the sampling localities.

Sawfly survival is known to be affected by the level of winter-cropping due to disturbance of the over-wintering pupae (Barker and Reynolds 2004). It is possible that this would have a faster, more significant effect on the rarer species within a community which would be detectable in species diversity indices. However, species diversity at the sampled localities did not appear to be related to incidence of winter-cropping; Mains of Glamis ranked higher than the James Hutton Institute in terms of species diversity (Section 3.3.3) yet the cropping area was more often cultivated during winter at Mains of Glamis than at The James Hutton Institute (Section 3.2.4). In addition, species diversity did not appear to correlate with the proportion of the field put into set-aside (in this case grass margin area; a suitable habitat for sawflies). Mains of Glamis, which ranked highest in terms of species diversity, had the highest working area to set-aside ratio in the sampled field (Section 3.2.4).

The information regarding the cropping history at the collection sites was limited to recent years and only available for the fields where sawflies were collected making it difficult to explain any differences in diversity between sites. Furthermore, the sampling regime was not specifically designed to analyse differences in species composition and diversity resulting in unequal sample duration across sites. Therefore, the conclusions that can be drawn explaining the differences in species composition and diversity between localities based on the available information are minimal. Importantly, there are a number of other factors relating to management practices which could be affecting sawfly species diversity including the frequency and intensity of agrochemical use, the range of host-plants available in set-aside land, the cropping history of neighbouring fields and the presence or absence of landscape features such as woodland or bodies of water.

Thus, sawfly species diversity in each of the sampled localities, indeed any locality in the arable landscape, will be regulated by a large number of factors. As a result, it is likely that species diversity estimates for each locality will change on a yearly basis even with consistency of sampling effort and methodology. To gain a true estimate of how sawfly populations are affected by their environment, it would likely be more productive to take a rigorous approach with a single study species, for example using molecular markers to examine in detail the genetic variation and population structure of these insects.

3.5 Conclusion

This chapter reports the sawfly species composition of three main localities in Scotland facilitating the initial selection of *Tenthredopsis excisa* as a study species. Preliminary comparisons were drawn between localities in terms of their species diversity and this did not appear to be correlated with the incidence of winter-cropping in the fields used for sampling or to the proportion of the field put into set-aside. However, this conclusion should be treated with caution as the sampling regime was not specifically designed to assess species diversity.

Extremely male-biased sex ratios were observed for *T. excisa* and for another two common species, *Tenthredopsis nassata* and *Dolerus aeneus*, at most or all localities sampled. Although this could be a consequence of the sampling method or sex allocation strategies by female sawflies, it is also possible that this result could be due to the presence of Complementary Sex Determination (CSD) in combination with inbreeding in sawfly populations. The presence of CSD in the study species is examined in Chapter 5.

4. Isolation and characterisation of microsatellite loci in a common farmland sawfly

4.1 Introduction

4.1.1 Methods for the isolation of microsatellite markers

Microsatellites have become the marker of choice for population genetic studies due to their multi-allelic nature, reproducibility, co-dominant mode of inheritance and abundance within the genome (Schlötterer 2004). Characterised by a high level of polymorphism, they are powerful genetic markers. However, their widespread use has been hindered due to the time-consuming and expensive methods required to isolate them (Squirrel *et al.* 2003; Zane *et al.* 2002).

The main drawback to using microsatellites is that, traditionally, they need to be isolated *de novo* from most species, involving the complex, laborious process of constructing an enriched genomic library (Zane *et al.* 2002). Briefly, DNA fragments are digested with restriction enzymes and then preferentially size-selected and cloned into a plasmid vector. Enrichment for microsatellite-containing sequences takes place before or after the cloning step using repeat-containing probes (Zane *et al.* 2002). In spite of the difficulties, microsatellites have been successfully isolated using this method for many insect species including the bumblebee *Bombus terrestris* (Stolle *et al.* 2009) and the harvester ant *Messor structor* (Arthofer *et al.* 2005). To date microsatellites have been isolated for only one sawfly species, *Cephus cinctus* or wheat stem sawfly (Hartel *et al.* 2003).

More recently, the public availability of huge volumes of sequence data for numerous eukaryotic genomes has accelerated microsatellite-based research (Sharma *et al.* 2007). Mining microsatellites from existing sequence databases significantly reduces the time and costs involved in isolating these markers in the target species. However, existing genetic resources for sawflies are scarce, with less than 1,500 nucleotide sequences available in the public database (NCBI 2011).

Within the last decade, new high through-put sequencing technologies referred to as “Next Generation Sequencing” (NGS) have revolutionized the field of molecular

biology. NGS enables the generation of enormous volumes of data relatively cheaply and on a far shorter timescale than was previously possible (Morozova and Marra 2008; Wall *et al.* 2009; Metzker 2010). In some cases, in excess of one billion short sequence reads can be generated per instrument run (Metzker 2010). Technologies range from the pyrosequencing method commercialized by Roche 454 to sequencing by ligation, a platform available from Life/APG (Metzker 2010).

Each NGS technology or "platform" has a distinct profile in terms of the average read length generated and the run time of the instrument. Additionally, one technology may be more suited for a specific application than another (Metzker 2010). The field of NGS development is fast-moving and, as new technologies emerge and existing platforms improve, it is necessary for researchers to match their study to the appropriate NGS platform.

The applications of NGS data seem limitless, ranging from comparative genomics to RNA expression profiling (Morozova and Marra 2008). However, the use of genome or transcriptome NGS to identify molecular markers, including microsatellites, is relatively novel with only a few studies reporting this approach in detail, especially for non-model insect species. Nonetheless, authors using this approach for marker development have published encouraging results. For the water strider *Gerris incognitus* (Insecta: Hemiptera) a one-quarter 454 pyrosequencing run yielded 30,820 (16.8 %) sequence reads from the total sequence set containing microsatellite repeats with 3 weeks of development time (Perry and Rowe 2011). NGS as a method of isolating microsatellite markers, when compared with the traditional genomic library-based approaches, is fast, efficient and less technically demanding.

4.1.2 Utility of microsatellite markers in related species

Transfer of primer sequences between closely related species can help to minimise the effort and/or cost involved in isolating microsatellites. The success of this 'cross-amplification' of microsatellite loci into other related species depends on the level of conservation of the flanking primer regions between species and is therefore inversely related to evolutionary distance between the species (Rico *et al.* 1996; Moore *et al.* 1991; Schlötterer *et al.* 1991). Cross-species amplification of microsatellite loci is well

documented in insect species including within the Hymenoptera. Examples include loci isolated for use in the bumblebee *Bombus terrestris* of which 25 were polymorphic in at least 1 of 9 other *Bombus* species (Stolle *et al.* 2009). The microsatellite loci isolated for the sawfly *Cephus cinctus* cross-amplified into 2 closely related *Cephus* species and the more distantly related *Trachelus tabidus* (Hartel *et al.* 2003).

4.1.3 Chapter summary

This chapter reports on the isolation and characterisation of microsatellites in a common farmland sawfly species via a traditional enriched genomic library method and latterly via an NGS transcriptome sequencing approach. The relative success of these two methods is analysed and the isolated microsatellites are considered in terms of their utility for population genetic studies and their ability to cross-amplify into other closely-related species.

4.2 Materials and Methods

4.2.1 The development of a genomic library enriched for microsatellite sequences in the species Tenthredopsis excisa

4.2.1.1 Preparation of genomic DNA for enriched-library construction

Genomic DNA was isolated from the heads of *Tenthredopsis excisa* individuals (Section 2.2.1) and in some cases amplified using the Genomiphi™ kit (Section 2.2.5) to increase the concentration of total genomic DNA. Two approaches were then used to prepare genomic DNA for the enrichment step. Either a standard single enzyme restriction digest was performed or a digest using two enzymes, one rare and one frequent cutter, in combination.

Briefly, genomic DNA was digested (37 ºC overnight) with either *Sau*3AI (1 unit/1 µg DNA) (Promega UK Ltd.; Southampton, UK) in a 50 µl reaction or with the frequent cutter *Mse*I (1 unit/500 ng DNA) and the rare cutter *Eco*RI (5 units/500 ng DNA) (New England Biolabs (UK) Ltd.; Hitchin, UK) simultaneously in a 250 µl reaction. *Sau*3AI digest reactions were purified using the MinElute® Kit (Section 2.2.4) into a 15 µl volume prior to running on a 1% (w/v) agarose gel (Section 2.2.2). A gel slice containing fragments in the size range 300-1000 bp was excised and purified using the MinElute® Kit (Section 2.2.4) into a 15 µl volume. *Eco*RI/*Mse*I digest reactions were neither size-selected nor purified prior to downstream applications as indicated in the FIASCO protocol (Fast Isolation by AFLP of Sequences Containing repeats) (Zane *et al*. 2002).

Subsequently, size-fractionated DNA was ligated to double stranded oligonucleotide adaptor sequences depending on the restriction enzyme used (Table 4.1). The adaptor sequences were prepared prior to use as follows. In the case of *Sau*3AI, 10 µl each adaptor (50 µM) was added to 79.2 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and 0.8 µl 5M NaCl. The mixture was then placed in a thermocycler and incubated at 95 °C for 3 minutes, 65 °C for 2 minutes, 45 °C for 2 minutes and 25 °C for 1 minute followed by a 4 °C hold. *Eco*RI adaptors were prepared by combining 3 µl *Eco*-F adaptor (100 µM), 3 µl *Eco*-R adaptor (100 µM) and 54 µl SDW, placing the mixture in a 65 °C water bath and allowing the mixture to cool slowly to room

temperature. *Mse*I adaptors were prepared by combining 30 µl *Mse*-F adaptor (100 µM) and 30 µl *Mse*-R adaptor (100 µM) and heat treating to 65 ºC as described above. After heat-treatment, the adaptors should have formed double-stranded adaptor sequences via complementary base-pairing.

Ligation reactions were performed in a 50 µl volume at 12 ºC for 12 hours. In the case of *Sau*3AI-digested fragments, size-fractionated DNA (10 µl) was ligated to 1.25 nmol *Sau*3AI adaptor (12.5 µl adaptor mix) using 2 units T4 DNA ligase and 5X ligase buffer (10 µl) (Promega Ltd.; Southampton, UK). In the case of fragments double-digested with *Eco*RI and *Mse*I, size-fractionated DNA (30 µl) was ligated to 5 pmol *Eco*RI adaptor (1 µl adaptor mix) and 50 pmol *Mse*I adaptor (1 µl adaptor mix) using 1 unit T4 DNA ligase, 5X ligase buffer (10 µl) (Promega Ltd.; Southampton, UK) and 10 mM ATP (1 µl).

Ligated DNA was then amplified by PCR. In the case of *Sau*3AI ligated DNA, 10 µl ligation reaction mixture was used per PCR (5 reactions) with 20 pmol *Sau*3AIA as the primer (Table 4.1) in each reaction. In the case of *Mse*I/*Eco*RI ligated DNA, 2.5 µl ligation reaction mixture was used per PCR (5 reactions) with 300 ng universal *Eco*RI primer E00 (5'GACTGCGTACCAATTC3') and 300 ng universal *Mse*I primer M00 (5'GATGAGTCCTGAGTAA 3') in each reaction. PCR was performed according to the conditions described in Section 2.2.3 with the following programme: denaturation at 94 ºC for 5 minutes, 35 cycles of 94 ºC for 30 seconds, 55 ºC for 30 seconds, 72 ºC for 1 minute, then a 7 minute final extension at 72 ºC and 4 ºC hold. Replicate PCRs were combined and purified using the MinElute® Kit as described in Section 2.2.4 eluting into a 15 µl final volume.

Table 4.1: Recognition sites and adaptor sequences for restriction digest enzymes.

Enzyme	**Recognition site**	**Adaptor sequences**	
*Sau*3AI	GATC	*Sau*3AIA	5'GCGGTACCCGGGAAGCTTGG3'
		*Sau*3AIB	5'CGCCATGGGCCCTTCGAACCCTAG3'
*Mse*I	TTAA	*Mse*-F	5'GACGATGAGTCCTGAG3'
		Mse-R	5'TACTCAGGACTCAT3'
*Eco*RI	GAATTC	*Eco*-F	5'CTCGTAGACTGCGTACC3'
		Eco-R	5'AATTGGTACGCAGTCTAC3'

4.2.1.2 Enrichment for microsatellite sequences using magnetic particle separation

A total of 200 pmol of biotinylated oligonucleotide probe (either $(CA)_{13}$ alone or $(CA)_{13}$ and $(GA)_{13}$ in equal amounts) (Sigma-Aldrich Co. Ltd.; Poole, UK) was hybridised to the 15µl of combined PCR product from the previous step in 6X SSC (0.9 M NaCl, 90 mM trisodium citrate, pH 7.0) and 1% SDS (1% w/v sodium dodecylsulphate). Hybridisation was carried out in a microcentrifuge tube placed at 42 ºC with rotation overnight. Isolation of microsatellite-containing fragments from the PCR product was then carried out using Dynabeads® M-280 Streptavidin magnetic beads (Invitrogen Dynal; Paisley, UK) in conjunction with a Dynal® magnetic stand.

Dynabeads were prepared according to the manufacturer's instructions. Briefly, 1 mg of beads were washed prior to use in 2X Binding and Washing (B&W) Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 mM NaCl) and resuspended in 200 µl 1X B&W Buffer. The hybridisation reaction mixture was then made up to a volume of 200 µl with SDW before adding to the 200 µl washed beads in suspension. The mixture was incubated at room temperature for 15 minutes with rotation to allow the biotinylated molecules to bind to the streptavidin beads.

The magnet was applied for 2 minutes and then the beads were washed twice in 1X B&W Buffer to remove any unbound fragments prior to resuspension of the beads in 100 µl of 10 µM EDTA, pH 8.2. Subsequently, the beads were incubated at 90 ºC for 2 minutes to release the annealed fragments, which were separated from the beads by re-applying the magnet; the released fragments were then purified using the MinElute® Kit as described in Section 2.2.4 and eluted into 10 µl of Buffer EB.

4.2.1.3 Sequencing of a fragment library enriched for microsatellite repeats

4.2.1.3.1 Cloning of enriched fragments using pGEM®-T Easy Vector

Fragments from the enriched library were ligated with pGEM®-T Easy Vector (Promega UK, Southampton, UK) for 12 hours at 4 ºC at a molar ratio of 1:1 (insert DNA:vector) in a reaction volume of 10 µl containing 3 units T4 DNA ligase and 1X Ligase Buffer (Promega Ltd.; Southampton, UK). Subsequently, 2 µl of ligation reaction was

transformed using ElectroMAX™ DH10-B™ electrocompetent cells (Invitrogen, Paisley, UK). Briefly, 2 µl of ligation reaction was mixed on ice with 18 µl of competent cells for 1 minute. This mixture was transferred to a 0.2 cm electroporation cuvette (Invitrogen, Paisley, UK) and pulsed at 1.6 volts using an *E. coli* pulser (Bio-Rad Laboratories Inc.; Hemel Hempstead, UK). The cuvette was then removed from the pulser and 1 ml of SOC (peptone 20g/l, yeast 5g/l, NaCl 0.584g/l, KCl 0.186g/l, $MgCl_2$ 2.03 g/l, $MgSO_4$ 2.46 g/l, glucose f/s 3.60 g/l) was added to the cuvette. The resulting mixture was transferred to a 15 ml conical tube and incubated with shaking at 37 ºC for 1 hour.

The cells were spread onto 1.5 % "LB AIX" agar plates containing peptone (10 g/l), yeast (5g/l) and NaCl (5g/l) supplemented with 0.5 ml Ampicillin (100 µg/ml), 1.6 ml IPTG (32 µg/ml) and 1.6 ml X-gal (32 µg/ml). The plates were incubated at 37 ºC overnight to allow blue/white colony selection. Individual white colonies were picked from the plates by hand using a pipette tip and placed into individual wells of a 96-well deep block containing 1 ml LB (peptone 10 g/l, yeast 5g/l, NaCl 5g/l, glucose 1g/l) supplemented with Ampicillin (100 mg/l) and grown overnight at 37 ºC with shaking.

4.2.1.3.2 Preparation of plasmids for large-scale sequencing

Plasmids from individual colonies were prepared for sequencing using an adaptation of the Millipore Multiscreen Plasmid Preparation Protocol (Millipore UK Ltd., Watford, UK) for large scale sequencing.

Incubated transformed cells were pelleted by centrifuging the 96-well blocks at 1,811 rcf for 5 minutes to and the supernatant was discarded. Pellets were resuspended in 80 µl of Solution I (30 mM glucose, 15 mM Tris-HCl pH 8.0, 30 mM EDTA pH 8.0, 60 µg/ml RNase A) using a vortex. Solution II (80 µl) was added (0.2 M NaOH, 1% SDS) and mixed thoroughly by vortexing before incubation at room temperature for 2 minutes. Solution III (80 µl) was added (3.6 M potassium acetate, 14% (v/v) acetic acid) before vortexing for 1 minute to yield a homogeneous bacterial lysate.

A Multiscreen MAFBNOB plate was positioned in the base of a vacuum manifold (both Millipore UK Ltd; Watford UK) and 160 µl binding solution (8 M guanidine hydrochloride) added to each well of the plate. The upper part of the manifold was replaced and a Multiscreen MANANLY clearing plate (Millipore UK Ltd.; Watford UK)

positioned on top. The bacterial lysate (130 µl) was transferred to each well of the clearing plate. Vacuum (10 ″Hg) was applied to the manifold for 3 minutes to draw the lysate through to the binding plate. The lysate and the binding buffer were thoroughly mixed and the binding plate transferred to the upper part of the manifold. Vacuum was applied for 1 minute (30 ″Hg) to bind plasmids to the plate and remove waste solution.

A 200 µl volume of ethanol (70%) was added to each well and vacuum (30 ″Hg) was applied for 1 minute followed by an identical wash with vacuum (30 ″Hg) applied for 3 minutes to remove ethanol. The membranes of the binding plate were dried by spinning at 1,811 rcf for 10 minutes and the plate was incubated at room temperature for 10 minutes to allow the ethanol to evaporate. The binding plate was then transferred to a microtitre storage plate and SDW (100 µl) applied to each well. Plasmids were eluted by spinning the plate at 1,811 rcf for 5 minutes.

Plasmid preparations (5 µl) were double-checked for inserts by digesting with *Eco*RI (New England Biolabs (UK) Ltd.) in a total reaction volume of 10 µl with 1X digest buffer (supplied with enzyme) and 0.1 µg BSA at 37 ºC for 1 hour. Digest reactions were then run on a 1% (w/v) agarose gel (Section 2.2.2) to determine the presence of inserts. Inserts are present if two bands can be visualised on the gel; one very large band around 3,000 bp in size representing the plasmid and another smaller band (approximately 100-800 bp in size) representing the insert.

4.2.1.3.3 Large-scale sequencing of plasmids

Preparations of plasmids known to contain inserts were sequenced. Plasmid template (5 µl) was placed in a sequencing reaction containing 0.5 µl Big Dye v3.1, 1.75 µl 5X dilution buffer (both Applied Biosystems Inc.; Warrington UK) and 1 µl universal M13F primer (10 µM) (Eurofins MWG Operon; London, UK). Reactions were then subjected to the following thermocycling programme: 96 ºC for 1 minute, followed by 25 cycles of 96 ºC for 10 seconds, 50 ºC for 5 seconds and 60 ºC for 4 minutes, followed by a 4 ºC hold.

Sequencing reactions were purified by EDTA/Ethanol precipitation. Briefly, 125 mM EDTA (2.5 µl) was added to each reaction followed by 95 % ethanol (30 µl). The reactions were incubated at room temperature for 15 minutes followed by centrifugation

at 1,811 rcf for 45 minutes at 4 ºC. Plates were then inverted and centrifuged at 100 rcf for 10 seconds to remove the supernatant. The plasmids were then washed twice by the addition of 70 % ethanol (150 µl) and centrifugation at 1,811 rcf for 10 minutes at 4 °C followed by inversion of the plate onto a tissue to remove the supernatant. Finally, the plate was incubated at room temperature for 1 hour.

Sequencing of the purified plasmids was performed using an ABI PRISM® 3730 Genetic Analyzer (48 capillary).

4.2.1.3.4 Discovery of microsatellite loci

The vector sequence was trimmed from the individual sequence data files using Sequencher 4.9 sequence analysis software (Genecodes, Ann Arbor, MI, USA). Microsatellite loci were found by manually screening the sequences. The initial oligonucleotide probe used to enrich the fragments was either $(CA)_{13}$ or $(CA)_{13}$ and $(GA)_{13}$ together. Therefore, the sequences were searched for $(CA)_n$ or the reverse complement $(GT)_n$ and $(GA)_n$ or the reverse complement $(CT)_n$. Primers were designed to amplify the microsatellite loci using Primer3 (v. 0.4.0) software (Rozen and Skaletsky 2000) employing the default settings.

4.2.1.4 Amplification of microsatellites and testing for polymorphism

DNA was extracted from the whole bodies of 8 *Tenthredopsis excisa* individuals as described in Section 2.2.1 and used to test the potential microsatellite loci. Individuals of both sexes were represented in this test population along with those from different sampling years and sites (Table 4.2). PCR was carried out in a total reaction volume of 10 µl containing approximately 15 ng total genomic DNA from each individual with 1 µM forward and 1 µM reverse primer under conditions described in Section 2.2.3. PCR reactions were subjected to an initial hold of 94 ºC for 5 minutes followed by 35 cycles of 94 ºC for 30 seconds, 55 ºC for 30 seconds and 72 ºC for 1 minute, followed by a 72 ºC final extension and a 4 ºC hold. PCR products were visualised on a 1% (w/v) agarose gel (Section 2.2.2) to determine whether amplification was successful across the test population.

Table 4.2: **Tenthredopsis excisa** ***test population used in the initial amplification of microsatellite loci.***

Test population: *Tenthredopsis excisa*		
Sex	Collection site	Year collected
M	SCRI	2008
M	SCRI	2008
M	GLAMIS	2008
M	GLAMIS	2008
F	SCRI	2009
F	GLAMIS	2009
F	GLAMIS	2009
F	GLAMIS	2009

If the locus was amplified in all or most of the test population then the forward primer of the pair was 5' fluorescently-labelled with 6-FAM (6-Carboxyfluorescein). PCR was performed again, as described above, using the same test population, the original reverse primer and the new fluorescently-labelled forward primer. Fluorescent PCR products were then diluted 1 in 10 and combined with 8.84 µl Hi-Di™ Formamide and 0.16 µl GeneScan™ 500 Rox™ internal size standard (both Applied Biosystems Inc., Warrington, UK) and visualized using the ABI PRISM® 3730 Genetic Analyzer (48 capillary). Fragments were analysed using GeneMapper® software (Applied Biosystems Inc., Warrington, UK). Microsatellites were considered to be polymorphic if 2 or more alleles were present in the test population. Any polymorphic loci were assayed on a larger test population of 79 *T. excisa* individuals collected from a range of field sites in 2009.

4.2.2 The development of a sawfly transcriptomic library utilising next-generation sequencing technologies

4.2.2.1 Extraction of total RNA from sawfly larval tissue

Total RNA was extracted from 5 pooled sawfly larvae of mixed species (310 mg tissue) using the "RNeasy© Mini Kit" (QIAGEN Ltd., Crawley, UK) with some modifications. Briefly, larvae were ground in liquid nitrogen using a mortar and pestle, Buffer RLT (6.18 ml) was added (600 µl per 30 mg of tissue as recommended) then the mixture was aliquoted into QIAshredder columns placed in 2 ml collection tubes. This was centrifuged at 15,700 rcf for 2 minutes; the flow-through was centrifuged for a further 3 minutes at 15,700 rcf and the supernatants transferred to clean microcentrifuge tubes.

To each, 1 volume of 70% ethanol was added and mixed thoroughly by pipetting. A 700 µl volume of each sample, including any precipitate that had formed, was then transferred to an RNeasy spin column placed in a 2 ml collection tube. The samples were centrifuged at 9,300 rcf for 15 seconds and the flow-through was discarded. Wash Buffer RW1 (700 µl) was added to each of the RNeasy spin columns and centrifuged for a further 15 seconds at 9,300 rcf. Flow-through was discarded and the columns washed with Buffer RPE (500 µl) followed by centrifugation at 9,300 rcf for 15 seconds. Wash Buffer

RPE (500 µl) was added a second time and the sample was centrifuged at 9,300 rcf for 2 minutes.

The spin columns were then transferred to new 1.5 ml collection tubes and RNase-free water (40 µl) was pipetted directly onto each spin column membrane. The columns were centrifuged for 1 minute at 9,300 rcf to elute the RNA. RNA concentration and quality was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies UK Ltd.; Cheshire, UK).

4.2.2.2 mRNA enrichment from total RNA

Total RNA was enriched for mRNA using the Oligotex© mRNA Mini Kit (QIAGEN Ltd., Crawley, UK). The volume of total RNA was adjusted to 250 µl with RNase-free water as directed by the manufacturer and Buffer OBB (250 µl), pre-heated to 37 ºC, was added with 15 µl Oligotex suspension also pre-heated to 37 ºC. This was mixed thoroughly by pipetting and then incubated at 70 ºC for 3 minutes in a water bath to disrupt the secondary structure of the RNA. The sample was then incubated at room temperature for 10 minutes. The Oligotex:mRNA complexes, formed during incubation, were pelleted by centrifugation at 15,700 rcf for 2 minutes. The supernatant was then carefully removed.

To reduce rRNA contamination the pellet was resuspended in RNase-free water (125 µl) and Buffer OBB (125 µl) and subjected to a second round of incubation at 70 ºC for 3 minutes followed by 10 minutes at room temperature. The sample was centrifuged again for 2 minutes at 15,700 rcf to pellet the Oligotex:mRNA complexes and the supernatant carefully removed.

The Oligotex:mRNA pellet was then suspended in Buffer OW2 (400 µl) to wash. The suspension was transferred to a small spin column (provided with the kit) and placed in a 1.5 ml RNase-free collection tube. The spin column set-up was then centrifuged for 1 minute at 15,700 rcf and the flow-through was discarded. This step was repeated to wash the complexes thoroughly

The spin-column was transferred to a new 1.5 ml RNase-free collection tube and 40 µl of Buffer OEB (pre-heated to 70 ºC) was pipetted directly onto the column membrane and pipetted up and down 3–4 times to maximise contact between the Oligotex:mRNA complexes and the elution buffer. This was then centrifuged for 1 minute

at 15,700 rcf. To ensure maximal yield a second 40 µl volume of pre-heated Buffer OEB was applied and the procedure repeated. The 80 µl elution was then stored at -80 ºC.

4.2.2.3 cDNA synthesis

Complementary DNA (cDNA) was synthesised from poly A^+ RNA (isolated as described in Section 4.2.2.2) using the SMARTer™ PCR cDNA Synthesis Kit (CLONTECH UK Ltd, Hampshire, UK) following the manufacturer's instructions with some modifications.

In the first instance, 3.5 µl of poly A^+ RNA (approximately 168 ng) was placed in a 0.5ml microcentrifuge tube with 1 µl of 3' SMART CDS Primer II A (12 µM). The contents of the tube were mixed by pipetting gently prior to spinning the tube briefly in a microcentrifuge. The tube was then incubated at 72 ºC for 3 minutes and then 42 ºC for 2 minutes in a thermal cycler.

For one cDNA synthesis reaction the following Master Mix was prepared: 2 µl of 5X First-Strand Buffer, 0.25 µl of DTT (100 mM), 1 µl of dNTP Mix (10 mM), 1 µl SMARTer™ II A Oligonucleotide (12 µM), 0.25 µl RNase Inhibitor, 1 µl of SMARTscribe™ Reverse Transcriptase (100 U) for a total volume of 5.5 µl. The Master Mix was prepared at room temperature and the reverse transcriptase added immediately prior to use of the Master Mix. This was added to the reaction tube giving a total volume of 10 µl. The tube was then incubated at 42 ºC for 90 minutes and the reaction terminated by incubation at 70 ºC for 10 minutes. The first-strand reaction product was diluted by adding 90 µl of TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA).

A 3 µl volume of the diluted single-stranded cDNA was aliquoted into a 0.5 ml microcentrifuge tube and made up to a volume of 10 µl using deionised water. The 3 µl volume is chosen relative to the starting amount of poly A^+ RNA used in the initial stages of cDNA synthesis.

A thermal cycler was pre-heated to 95ºC whilst the PCR Master Mix was prepared. One reaction required: 74 µl of deionised water, 10 µl of 10X Advantage 2 PCR Buffer, 2 µl 50X dNTP Mix (10 mM), 2 µl 5' PCR Primer II A (12 µM) and 2 µl of 50X Advantage 2 Polymerase Mix for a total volume of 90 µl per reaction. The Master Mix was vortexed thoroughly and then spun briefly in a microcentrifuge to collect the contents

at the bottom of the tube. At this point, 90 µl of Master Mix was added to the 10 µl diluted single-stranded cDNA for a total reaction volume of 100 µl.

The reaction mixture was then subjected to the following thermal cycling programme: 95 ºC for 1 minute followed by 27 cycles of 95 ºC for 15 seconds, 65 ºC for 30 seconds and 68 ºC for 6 minutes. The full reaction mixture was subjected to 15 complete cycles at which point the program was paused and 70 µl of the reaction mixture was stored at 4 ºC while the remaining 30 µl of the reaction mixture was subjected to further PCR cycling for a total of 27 cycles with 5 µl of reaction mixture removed after every third cycle. These 5 µl aliquots were analysed by gel electrophoresis to determine the optimum number of PCR cycles for good quality double-stranded cDNA. The optimum number of PCR cycles was 15, after which there was no improvement or increase in product with further cycling. The 70 µl of reaction mixture initially stored after 15 cycles was purified using the MinElute® Kit (Section 2.2.4) and eluted in Buffer EB (20 µl).

4.2.2.4 Next-generation sequencing and sequence assembly

Double-stranded cDNA (4 µg) was submitted to the Department of Biology Technology Facility (University of York; UK) for standard transcriptome sequencing on a 454 Genome Sequencer FLX Instrument, Titanium series (Roche Applied Science; Burgess Hill, UK). Fragmentation and library preparation were carried out as recommended by the manufacturer prior to the sequencing run whereby a half-plate was utilised. The resulting reads were screened for the presence of adaptor sequences originating from both the cDNA preparation and the 454 experimental procedures. Raw SFF files were generated using GS FLX on-board software. Reads lacking the adaptor sequences are removed by the sequencer software.

Prior to assembly into contiguous sequences (contigs) the sequence reads were screened a second time for the presence of adaptor sequences originating from both the cDNA preparation and the 454 experimental procedures. Adaptor contamination was masked using CROSS_MATCH (http://www.phrap.org/phredphrapconsed.html), and then trimmed from the reads using custom perl scripts. The matching quality scores for the reads were also removed. Any reads that had adaptor contamination in the middle were

discarded as possible chimeric sequences. After adaptor trimming, the remaining sequences were assembled using the TGICL suite (http://compbio.dfci.harvard.edu/tgi/software) running on a single CentOS Linux machine with four processors. The assembly CAP3 parameters used were –p 75 –d 200 –f 250 –h 90.

4.2.2.5 Homologies to known sequence

Contigs and singletons were annotated with descriptors of their closest homologues using BLAST (Altshul *et al*. 1990) (with an e-value cut off of 1.00 e-10) to search them against the non-redundant nucleotide sequences located within the NCBI database (accessed 17/08/11).

All sequences were then compared using BLAST (Altshul et al. 1990) (with an e-value cut off of 1.00 e-10) with the *Drosophila melanogaster* proteins listed in the "Annotated Proteins" database within FlyBase: A Database of *Drosophila* Genes and Genomes (Tweedie *et al.* 2009). Each hit was then compared to the generic GO (gene ontology) data on the Gene Ontology Consortium website (The Gene Ontology Consortium 2000) (accessed 02/09/11). The GO database has developed three structured vocabularies or "ontologies" which enable researchers to describe gene products in terms of their associated (1) biological processes, (2) cellular components and (3) molecular functions in a species independent manner. Rather than using a full GO (gene ontology) analysis, a generic "GO Slim" classification was used to give a broad overview of the ontology content within these three classifications.

4.2.2.6 Microsatellite discovery and primer design

Microsatellites of repeat length 2–5 were discovered in the final assembly by searching with Sputnik software (Abajian 1994). Primers were designed using Primer3 software v. 0.4.0 (Rozen and Skaletsky 2000) employing the default settings. Primer pairs were grouped into those that amplified di-, tri-, tetra- and pentanucleotide repeat sequences respectively. Primer pairs designed to amplify 48 and 24 of the longest di- and

trinucleotide repeat regions respectively were synthesised (Sigma Aldrich; Gillingham, UK).

4.2.2.7 Amplification of microsatellite loci in Tenthredopsis excisa *and* Dolerus aeneus

Each of the 72 primer pairs (Section 4.2.2.6) were run against the same test population of eight *Tenthredopsis excisa* individuals used previously (Table 4.2) and on a test population of 15 *Dolerus aeneus* individuals, comprising 14 males and one female, collected during the 2008 field season (Table 4.3). Different PCR conditions were employed to allow amplification of the microsatellite loci. Firstly, identical PCR conditions to those used previously (Section 4.2.1.4). Secondly, an identical thermocycling programme as the first but with an increased annealing temperature of 58 °C. Finally a touchdown PCR programme beginning with a 94 °C incubation for 5 minutes followed by 7 cycles of 94 °C for 30 seconds, 65 °C for 30 seconds (reducing by 1 °C/cycle) and 72 °C for 1 minute. This was followed by 25 cycles of 94 °C for 30 seconds, 58 °C for 30 seconds and 72 °C for 1 minute and finished with 72 °C incubation for 7 minutes and an 8 °C hold.

If a single product was amplified in all or most of the test population then the forward primer of the pair was 5' fluorescently-labelled with 6-FAM (6-Carboxyfluorescein). PCR was then performed again using the same test population, the appropriate thermocycling programme, the original reverse primer and the new fluorescently-labelled forward primer. Genotyping was performed as described in Section 4.2.1.4.

Microsatellite loci were considered to be polymorphic if two or more alleles were present at a locus. Polymorphism Information Content (PIC) values were calculated for microsatellite loci according to Botstein *et al.* (1980) using the Microsatellite Toolkit add-in (Park 2001) for Microsoft Excel.

Table 4.3: Test population of* Dolerus aeneus *individuals (sampled 2008) used in the initial amplification of microsatellite loci.

Test population: *Dolerus aeneus*	
Sex	Collection Site
M	SCRI
F	New Gilston
M	New Gilston
M	SCRI
M	Glamis
M	Glamis
M	Glamis
M	Glamis
M	New Gilston
M	SCRI
M	SCRI
M	Glamis
M	Aberdeen
M	Aberdeen
M	Aberdeen

4.2.2.8 Scoring of microsatellite loci

Scoring of microsatellite alleles using the GeneMapper® software (Applied Biosystems Inc., Warrington, UK) can result in the presence of wrongly-scored alleles. To screen for the presence of insertion or deletion mutations accounting for the presence of erroneous alleles (i.e. alleles scored one base pair apart for a dinucleotide repeat sequence) a selection of alleles were sequenced for each microsatellite locus where suspect alleles arose.

Microsatellite loci were amplified as described in Section 4.2.2.7 and resulting PCR products purified by combining 5 µl PCR product with 2 µl ExoSAP-IT (USB Corporation; Cleveland, USA). The resulting mixture was incubated at 37 °C for 15 minutes followed by 80 ºC for 15 minutes according to the manufacturer's instructions. Approximately 5 ng of purified PCR product was placed in a 10 µl sequencing reaction containing 0.5 µl Big Dye v3.1, 1.75 µl 5X dilution buffer (both Applied Biosystems Inc.; Warrington UK) and 1 µl of the correct forward primer (10 µM) (Eurofins MWG Operon; London, UK). Reactions were then subjected to the following thermocycling programme: 96 ºC for 1 minute, followed by 25 cycles of 96 ºC for 10 seconds, 50 ºC for 5 seconds and 60 ºC for 4 minutes, followed by a 4 ºC hold.

Sequencing reactions were purified by EDTA/Ethanol precipitation. Briefly, 125 mM EDTA (2.5 µl) was added to each reaction followed by 95 % ethanol (30 µl). The reactions were incubated at room temperature for 15 minutes followed by centrifugation at 1,811 rcf for 45 minutes at 4 ºC. Plates were then inverted and centrifuged at 100 rcf for 10 seconds to remove the supernatant. The reaction products were then washed twice by the addition of 70 % ethanol (150 µl) and centrifugation at 1,811 rcf for 10 minutes at 4 °C followed by inversion of the plate onto a tissue to remove the supernatant. Finally, the plate was incubated at room temperature for 1 hour. Sequencing of the purified reaction products was performed using an ABI PRISM® 3730 Genetic Analyzer (48 capillary) and the resulting sequences were screened using Sequencher 4.9 sequence analysis software (Genecodes, Ann Arbor, MI, USA).

4.2.2.9 Cross-species amplification of microsatellite loci

Polymorphic dinucleotide microsatellite primer pairs (three alleles or more) were tested on eight randomly chosen individuals of the following 11 species: *Dolerus puncticollis, Dolerus nigratus, Dolerus picipes, Dolerus gonager, Aglaostigma aucupariae, Aglaostigma fulvipes, Halidamia affinis, Ametastegia glabrata, Pachyprotasis rapae, Athalia cordata* and *Tenthredopsis nassata.*

PCR was performed using the conditions described in Section 4.2.1.4 with an increased annealing temperature of 58 ºC. PCR products were visualised on a 1% (w/v) agarose gel (Section 2.2.2) to determine whether or not the loci amplified successfully.

4.3 Results

4.3.1 The development of a genomic library enriched for microsatellite sequences

4.3.1.1 Overview of enriched genomic library construction

The enriched genomic library construction procedure is summarised graphically in Plate 4.1. Total genomic DNA was extracted (Step 1) and subsequently digested with restriction enzymes; both *Mse*I and *Eco*RI were used simultaneously or *Sau*3AI alone (Step 2). The fragments were size-selected (300–1000 bp) and compatible, double-stranded adaptors were ligated to the ends of the fragments (Step 3). PCR was then performed to increase the concentration of the DNA using primers homologous to the known adaptor sequences (Step 4). Fragments containing microsatellite repeats were then separated from the rest using a biotinylated oligonucleotide probe in conjunction with magnetic, streptavidin-coated beads (Step 5). A second PCR was then performed to increase the number of microsatellite-containing fragments (Step 6). The enriched fragments were then cloned and sequenced (Step 7). Sequences were searched manually for microsatellite loci and primers designed using suitable software (Step 8).

4.3.1.2 Isolation and characterisation of microsatellite loci

A total of 768 clones were sequenced during the construction of the enriched library, 336 of which were derived from enriched genomic libraries arising from the *Sau*3AI digestion method and 432 of which were derived from the *Mse*I/*Eco*RI digestion method. Twenty-seven microsatellite loci were discovered within these sequences, of which it was possible to design primers for 23 (Table 4.4). Twelve of the primer pairs amplified a single product in all or most of the test population of eight *Tenthredopsis excisa* individuals), one of which yielded a polymorphic microsatellite locus with two alleles (Table 4.4). When this locus (Texcisa11) was assayed on the larger 2009 *T. excisa* population of 79 individuals there was no increase on two alleles. Due to the lack of polymorphism in the discovered loci no diversity statistics were performed.

In addition to the expected (CA/GT/GA/CT)n repeats, several other dinucleotide repeat motifs were discovered, along with some tri-, tetra- and pentanucleotide repeats, in the sequenced clones. Seventeen of the repeat sequences found were dinucleotide repeats, seven were trinucleotide repeats and the rest were tetra- or pentanucleotide repeats (Table 4.4). Thirteen of the sequences were composed of perfect tandem repeats with the remaining 14 sequences composed of imperfect tandem repeats.

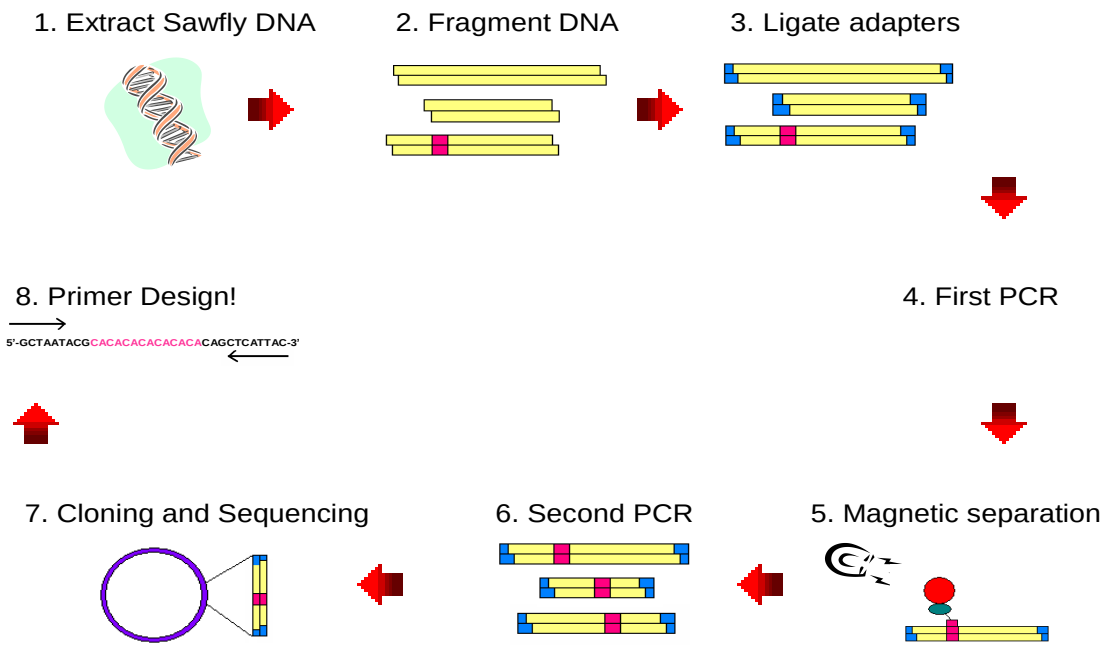


Plate 4.1: Outline of the process of enriched genomic library construction. In Step 2 the microsatellite region of the DNA is marked in red. In Step 3 the blue sections represent the adaptor sequences. In Step 5 the oligonucleotide probe is shown in red adjacent to the microsatellite region. The probe is biotinylated represented by the green ellipse, and hybridised to a streptavidin bead represented by the red circle.

Table 4.4: Twenty-seven microsatellite loci isolated for* Tenthredopsis excisa *using an enriched genomic library method.

Locus name	Repeat Sequence	Motif type	Repeat type	Expected product size (bp)	Primers designed	Single products amplified Y/N	Polymorphic Y/N
SawCAGA1	$(CT)_{13}$	Dinucleotide	Perfect	140	F 5'TGACTGCGTACCAATTCACC3'		
					R 5'CGTACTACGGAGGGACTCCA3'	N	N/A
SawCAGA1a	$(CTT)_3$	Trinucleotide	Perfect	N/A	No - SSR too short.	N/A	N/A
SawCAGA1b	$(CA)_4$…$(AC)_3A$	Dinucleotide	Perfect/Compound	N/A	No - SSR too short.	N/A	N/A
Saw CAGA1c	$(CT)_4$	Dinucleotide	Perfect	N/A	No - SSR too short.	N/A	N/A
SawCAGA2	$G(GT)_5(GA)_5GG(GA)$	Dinucleotide	Imperfect	173	F 5'CGCACACGTTACACACAACC3'		
					R 5'TCCTGAGTAACCCCGACAAC3'	N	N/A
SawCAGA3	$GAAAA(GAAGA)_3GAAAA$	Pentanucleotide	Imperfect	184	F 5'CTCGAAAGCAAGGGAACAAG3'		
					R 5'TTCCTGAGACTTTGCGCTTT3'	N	N/A
SawCAGA3a	$(CCG)_3$	Trinucleotide	Perfect	N/A	No - SSR too short.	N/A	N/A
SawCAGA4	$(GA)_{10}$	Dinucleotide	Perfect	169	F 5'TGATGAGTCCTGAGTTAACGAGA3'		
					R 5'TCCATATAACGGGGAAATCG3'	N	N/A
SawCAGA5	$(GA)_{11}$	Dinucleotide	Perfect	132	F 5'TGACTGCGTACCAATTCACC3'		
					R 5'ATGCTCAGGTTCGGTGTCTC3'	N	N/A
SawCAGA6	$(GA)_{10}$	Dinucleotide	Perfect	165	F 5'TCCTGAGTAACGAGAGAGAGAGAGA3'		
					R 5'ACTGCGTGCCAATTCCATA3'	N	N/A
SawCAGA7	$(AAAG)_3(GAAG)(AAAG)$	Tetranucleotide	Imperfect	124	F 5'CCTCCAAAAGCCAACGAATA3'		
					R 5'TGACTGCGTACCAATTCCAA3'	N	N/A
Texcisa1	$GG(GT)_6(GG)_2$	Dinucleotide	Imperfect	207	F 5'TGACTGCGTACCAATTCG3'		
					R 5'GGAAGCTGTCCAGCCTACTG3'	N	N/A
Texcisa2	$(GT)_4$	Dinucleotide	Perfect	205	F 5'GTAATGATCCTTCCGCAGGT3'		
					R 5'AAGCTCGCGTTGATTACGTC3'	Y	N
Texcisa3	$(TG)_4T$	Dinucleotide	Perfect	168	F 5'AGGGGGAACTGGGTGAATAA3'		
					R 5'GCCGTATCCTACCACGAGAC3'	Y	N
Texcisa4	$(GA)_4GC(GA)_4T(AG)_5$	Dinucleotide	Imperfect	155	F 5'TTCCGGACCGAAATAAATGT3'		
					R 5'CGGTCGTCGAAGAGGAATAA3'	Y	N
Texcisa5	$(GT)_2$ $TT(GT)_2(GTTTGT)_2$	Dinucleotide	Imperfect	211	F 'AATCGCCGCACTATAACCAC3'		
					R 5'ACCTCATTTGACCCAGAACG3'	Y	N
Texcisa6	$(TA)_3$	Dinucleotide	Perfect	94	F 5'GGCTTTTAGCGCGTATTCAG3'		
					R 5'CCCGTGGATTACTCAGGA3'	N	N/A

Texcisa7	$(CA)_5$	Dinucleotide	Perfect	163	F 5'TGACTGCGTACCAATTCTGG3' R 5'GTAATCGGGGTAAGCTGCTG3'	N	N/A
Texcisa8	$(TAA)_2CAATAA$	Trinucleotide	Imperfect	166	F 5'TACGCCGGTGAAGGGTATAG3' R 5'GGTTACGCGACGAATCAACT3'	Y	N
Texcisa9	$(CAC)_6$	Trinucleotide	Perfect	162	F 5'ACGATCGCGAATCTTACACC3' R 5'ACCAATTCCCGACATAAACG3'	Y	N
Texcisa10	$(GTT)(GT)(GTT)_2$	Trinucleotide	Imperfect	170	F 5'TGACTGCGTACCAATTCCAC3' R 5'TTAGGCCAGAAAAACGGAAA3'	N	N/A
Texcisa11	$(CA)_6$…$TAGA(TGGA)_7TGTA$	Tetranucleotide	Compound/Imperfect	184	F 5'TTTTTCGCTGCAGTTACACG3' R 5'AGGGGAAGACAATCCCTGTC3'	Y	Y
Texcisa12	$(CTT)_3CTC(CTT)_6TTT(CTT)_2$	Trinucleotide	Imperfect	206	F 5'CCAAGCTTCGGATCTGTACG3' R 5'CCCGTGTTGAAAACAATCCT3'	Y	N
Texcisa13	$(CG)_8$….$(CA)_5CG(CA)_4(CG)_3(CA)_2CG(CA)_{11}$	Dinucleotide	Imperfect	151	F 5'CCCAAGCTTCGAGTCGAGTA3' R 5'AGACGCGGGGAGGTTAAGT3'	Y	N
Texcisa14	$(ACG)_3ACA(ACG)_3C(ACG)$	Trinucleotide	Imperfect	235	F 5'ACGGGGAGAGTGTGTTTGTC3' R 5'CTTCGGATCCATCTGGGTTA3'	Y	N
Texcisa15	$(GT)_2(AT)_2(GT)_{11}AT(GT)_3$	Dinucleotide	Imperfect	207	F 5'ATCTGTTCGAGCGCGTAGTT3' R 5'TTTATCCCAGTTCGCAGTCC3'	Y	N
Texcisa16	$(GC)_2(GT)_8GA$	Dinucleotide	Imperfect	154	F 5'TTTGCAACAGCTCCTGATGT3' R 5'CCAACAGCGTCTCGAATACA3'	Y	N

4.3.2 The development of a sawfly transcriptomic library using next-generation sequencing technologies

4.3.2.1 Transcriptome sequencing and contig assembly

Using 454 pyrosequencing 150,725 high quality sequence reads were assembled *de novo* into 18,539 contigs and 260 singletons representing 7 Mbp of transcriptome (Figure 4.1). The mean contig length was 372 nucleotides. The longest contig in the assembly was 3916 bp in length and was composed of 124 individual reads. The contig with the largest number of component reads was 2344 bp in length and contained 1597 individual reads.

4.3.2.2 Homologies to known sequences

Of the 18,799 total sequences (contigs + singletons), 8,809 had significant matches in the public sequence databases (NCBI) at the 1.00 e-10 e-value cut off level. These blast matches correspond to 6,714 unique accession numbers, 4,273 (63.64 %) of which were insect sequences. Of the insect sequences, 3,574 (83.64 %) were hymenopteran sequences (Figure 4.2).

When all sequences were annotated by comparison with the Annotated Proteins present in Flybase (Tweedie *et al.* 2009) and then compared with the Gene Ontology database (The Gene Ontology Consortium 2000) 19,998 GO hits to the sequences were grouped under “biological processes” and 11,396 hits were grouped under “cellular components” totalling 31,394 hits (note that one contig can match a single protein in FlyBase but that protein can have more than one GO term attached to it). Within the “biological processes” group the two main sub-groups of gene products were “biological processes” (21 %) and “anatomical structure development” (10 %) (Figure 4.3). Within the “cellular components” group the two main sub-groups of gene products were “protein complex” (22 %) and “cytoplasm” (14 %) (Figure 4.4).

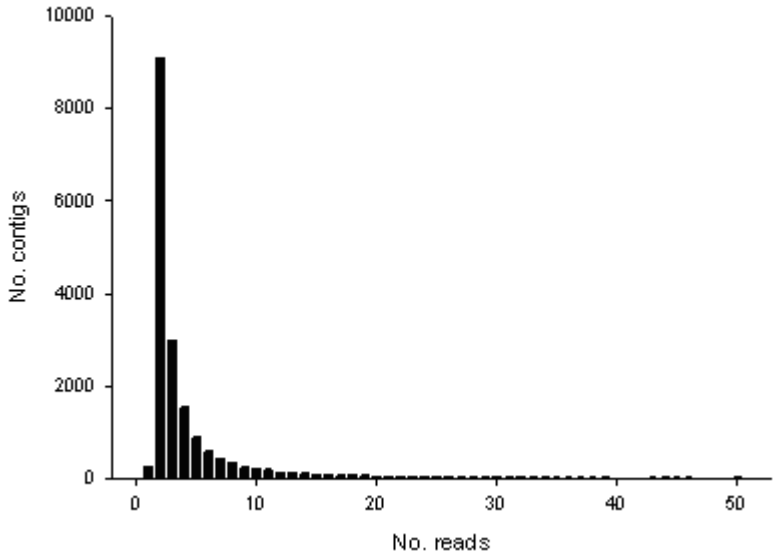


Figure 4.1: Number of reads per contig within a transcriptome sequence assembly constructed from cDNA isolated from five sawfly larvae of unknown species. Data are shown for contigs containing 1-50 reads, representing approximately 97.9% of all sequences (contigs + singletons).

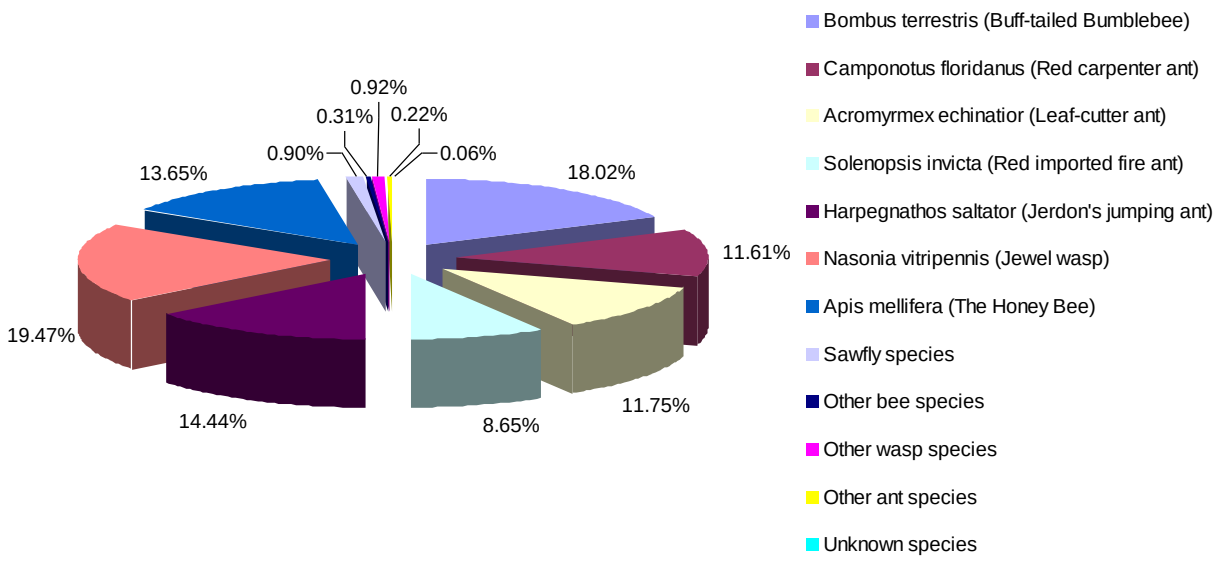


Figure 4.2: Characterisation of BLAST hits to hymenopteran sequences in the NCBI non-redundant protein sequence database.

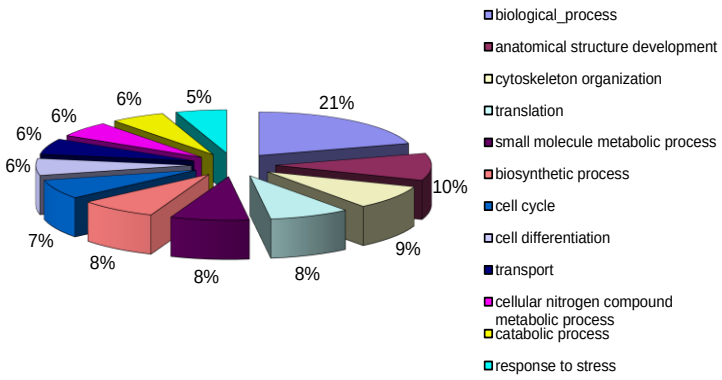


Figure 4.3: Sub-groups of GO terms with more than 500 hits to our sequences under the heading "biological processes" accounting for 53.42 % of the total hits under this heading.

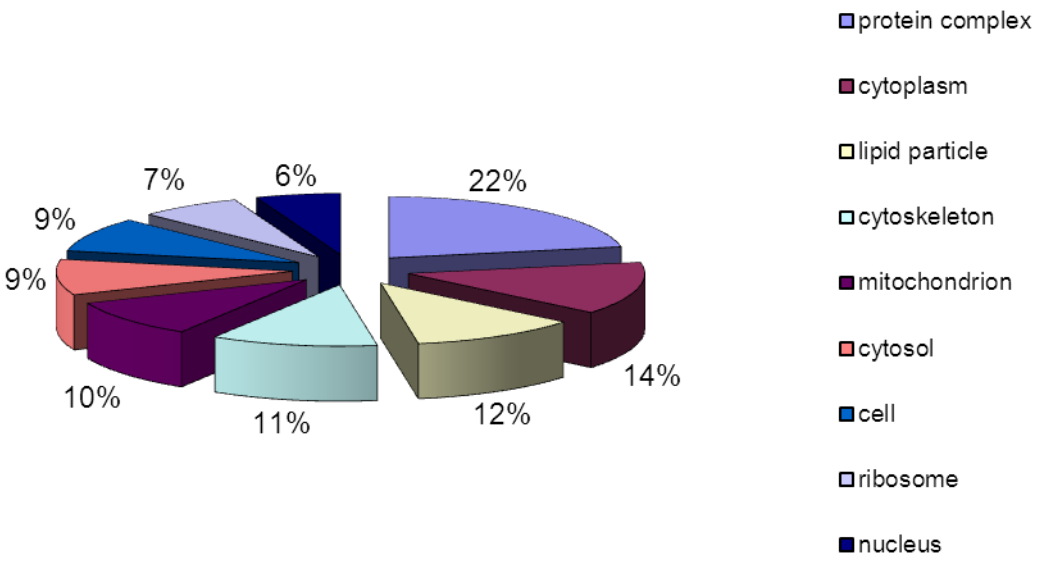


Figure 4.4: Sub-groups of GO terms with more than 500 hits to our sequences under the heading "cellular components" accounting for 78.41 % of the total hits under this heading.

4.3.2.3 Isolation and characterisation of microsatellite loci

After the final assembly, 1,030 of the 18,539 contigs and 9 of the 260 singletons were found to contain one or more SSRs. This gave a total of 1,284 isolated SSRs, of which 17% were dinucleotide repeats, 52% were trinucleotide repeats and the rest were tetra- and pentanucleotide repeats (Table 4.5). Primer pairs were designed for 847 of the 1,284 loci discovered. As previously stated, 72 of these primer pairs were synthesised (Table 4.5).

4.3.2.3.1 Amplification of microsatellite loci in *Tenthredopsis excisa* and *Dolerus aeneus*

Each of the 72 primer pairs was run on test populations of 8 *Tenthredopsis excisa* (Table 4.2) individuals and 15 *Dolerus aeneus* individuals (Table 4.3) as stated in Section 4.2.2.7.

In the *T. excisa* test population, 18 of the 72 primer pairs amplified a single product in all or most of the test population. The remaining 54 primer pairs did not amplify a product at all or produced multiple products or smears. Of the 18 primer pairs that gave single PCR products, nine were fluorescently-labelled. Four of these fluorescent primer pairs produced polymorphic microsatellite loci with two alleles. When assayed on the larger test population of 79 *T. excisa* individuals, there was no increase on two alleles for any of the four loci.

In the *D. aeneus* test population, 43 of the 72 primer pairs amplified a single product in all or the majority of the test population. Nineteen of these primer pairs were fluorescently-labelled and subsequently 14 of these loci were found to be polymorphic exhibiting two or more alleles. One of the 14 polymorphic loci failed to amplify a fragment in four individuals over multiple attempts suggesting the presence of null alleles and this locus was removed from further analysis.

For the 13 remaining polymorphic loci, the number of alleles per locus ranged between two and eight and diversity (PIC) values ranged from 0.117 to 0.850 (mean 0.527) (Table 4.6). Six of the thirteen microsatellite loci had matches in NCBI with known insect sequences, five of which were hymenopteran sequences (Table 4.6).

Table 4.5: Summary of microsatellite sequences isolated via sawfly transcriptome NGS (Reproduced in part from Cook* et al. *2011).

SSR type	Total no. each SSR type	Repeat number range	Primers designed[a]	Primers tested[b]
Dinucleotide	212	5-16	137	48
Trinucleotide	662	4-54	460	24
Tetranucleotide	210	3-15	128	0
Pentanucleotide	200	2-8	122	0
Total	1284	N/A	847	72

[a] The number of possible primer pairs designed using the default settings in Primer3 (Rozen and Skaletsky 2000). [b] The number of primer pairs synthesised and tested.

Table 4.6: Primer sequences of 13 microsatellite loci for use in **Dolerus aeneus** *(Reproduced from Cook* **et al.** *2011).*

Locus name[a]	Repeat motif[b]	Primer sequence 5'-3'[c]	Allele size range (bp)	No. of alleles	PIC values	Homologies
SAW454_1 (JF304726)	$(AG)_5(AT)(AG)_6(GG)(AG)_3$	F GTAGCTGAATGGGATTAAGCGAAGC R AGTGATTTTCAGTGAATGTTCCATCC	197-229	8	0.809	None
SAW454_2 (JF304727)	$(AC)_{14}$	F TCGCTTATAGTACGCAGATACCCGT R CAAGTGGCTCTCTGCTCCTATGCT	203-230	4	0.425	Cyclic AMP-dependent transcription factor ATF-4 [Camponotus floridanus]. GenBank Accession number: EFN64595.1
SAW454_3 (JF304728)	$(AC)_9(GC)(AC)_3$	F AAACGTCACAATCATCACCGA R AGGGTATGTATAGCGACACAGAATAAG	210-220	4	0.524	None
SAW454_4 (JF304729)	$(AG)_5$……$(AG)_3$	F TTTTTGGTTGAATTATTATCACGGG R GGATAGAGACGATATAAGCAATCTCCA	235-254	8	0.850	None
SAW454_7 (JF304730)	$(GA)_6(GC)(GA)_4$	F GGAGTCCGTAAGCGGTCCCT R TCCGTTTCCCTTGGCACACT	209-228	4	0.335	hypothetical protein EAG_13038 [Camponotus floridanus] GenBank Accession number: EFN60563.1
SAW454_11 (JF304731)	$(AT)_{11}$	F CGGAGGGAATCGAGTCGAAC R CACCCGAGTACCTCTCCCGA	234-242	7	0.786	None
SAW454_14 (JF304732)	$(AT)_{10}$	F TTGCTGCACAGCTTTTGATCC R GTTCGCCGGCAAGTTCTTTG	198-200	2	0.117	ACYPI008213 [Acyrthosiphon pisum] Accession number: BAH72837.1
SAW454_16 (JF304733)	$(GT)_{10}$	F CCGAAAAGGGGGAATTACGG R TGGCGGTGAAGAAAAACCCA	188-204	5	0.629	None
SAW454_19 (JF304734)	$(AT)_9$	F CAAGCCGGAGTTGCACAAGA R TGACCGATTGATAGACACATCATAGG	214-220	3	0.370	Vacuolar proton pump subunit G [Camponotus floridanus] GenBank Accession number: EFN71124.1
SAW454_23 (JF304735)	$(TA)_8$	F CGCTACGCCTTAAGAAGTTCAATCC R AAATCTCACGTGCGAGTACCGA	197-214	4	0.550	None
SAW454_24 (JF304736)	$(TG)_8$	F GAAGTGCGGATGATACTGCGA R AAAGATGAGGCGGAAAAGAAAAA	232-245	7	0.783	Cleft lip and palate transmembrane protein 1-like protein [Harpegnathos saltator] GenBank Accession number: EFN88050.1
SAW454_31 (JF304737)	$(TAT)_2(TAA)(TAT)_{15}$	F TCAGTCGATCCTTGCCGTCTC R CGAGGAAACACAATGCCTAATGC	224-234	3	0.370	None
SAW454_41 (JF304738)	$(ATA)_6$……$(ATA)_4$	F CACCGCCGTCCTCTCTCTTT R CGCGATTGTATCGATGTCTTCTTG	197-198	2	0.309	PREDICTED: similar to CG17927-PF [Nasonia vitripennis] GenBank Accession number: XP_001607303.1

[a] GenBank Accession numbers are given in parentheses. [b] one dot indicates one base. [c] all forward primers are 5' labelled with 6-FAM.

4.3.2.3.2 Scoring of microsatellite loci

Sequencing of microsatellite repeat regions within loci showed that erroneous alleles were attributable to insertion or deletion mutations in differing regions outwith the microsatellite repeat sequence. Therefore, scoring of microsatellite alleles was not altered in any way from that observed in the GeneMapper® software (Applied Biosystems Inc., Warrington, UK). Sequencing of several individuals at locus Saw454_1 is shown as an example (Plate 4.2).

4.3.2.3.3 Cross-species amplification of microsatellite loci

Nine of the most polymorphic microsatellite loci (three alleles or more) (Table 4.6) were tested on a range of other sawfly species and genera as listed in Section 4.2.2.9. The nine loci were amplified successfully in other closely related *Dolerus* species with minimal presence of multiple products or smearing (Table 4.7). However, none of the loci produced single products in the more distantly related genera of the Tenthredinidae.

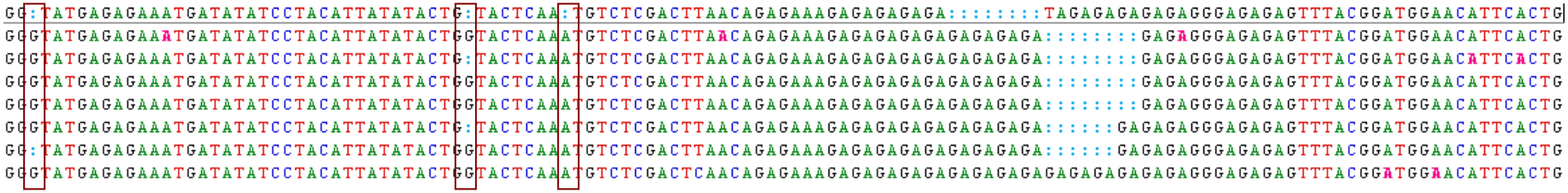


Plate 4.2: Screenshot taken from Sequencher 4.9 sequence analysis software. Part of the reference sequence (200 bp total (119 shown); containing a GA repeat region) for locus Saw454_1 is represented by the first row. The two rows directly below the reference sequence represent two individuals that were scored as 200 bp followed by two individuals scored as 199bp, followed by two individuals scored as 201 bp and finally one individual scored as 206 bp. Insertion/deletion mutations are highlighted. Variation in the length of the microsatellite sequences can be attributed to regions other than the repeat sequence thus accounting for the presence of one bp differences between alleles.

Table 4.7: Within-genus cross-species amplification of nine* Dolerus aeneus *microsatellite loci (Reproduced from Cook* et al. *2011).

Locus	Saw4541	Saw4542	Saw4543	Saw4544	Saw45411	Saw45416	Saw45419	Saw45423	Saw45424
D. puncticollis	+	+	+	+	+	+	+	MP	+
D. nigratus	-	+	+	+	+	+	+	-	+
D. picipes	+	+	+	+	+	+	+	+	+
D. gonager	+	+	+	+	+	MP	-	-	+

+ indicates a clear PCR product of the expected size in at least 7 out of the 8 individuals tested for that species, - indicates no product amplified in any individual, MP indicates multiple products or smears.

4.4 Discussion

4.4.1 A genomic library derived from sawfly DNA and enriched for microsatellite sequences produced low marker yield

In spite of the difficulties involved in using enriched genomic libraries to isolate microsatellite markers (Section 4.1.1) variations on this method have been used successfully to produce microsatellites for numerous species including those belonging to the Hymenoptera (Stolle *et al*. 2009; Arthofer *et al.* 2005; Hartel *et al.* 2003). However, the protocol described in this chapter for the isolation of microsatellites from sawfly DNA produced markers with very low yields.

The DNA used to construct the libraries was extracted solely from the heads of the sawflies in order to minimise the presence of contaminant DNA in the sample i.e. DNA from bacterial populations in the gut. Extracted sawfly DNA was of good quality but present at low concentrations and therefore the Genomiphi kit was trialled to amplify the genomic DNA. However, due to the tendency of the Genomiphi enzyme to amplify smaller fragments preferentially, its use resulted in amplification of the small fraction of degraded, fragmented DNA present in the initial extraction and the use of Genomiphi was therefore discontinued.

In spite of low DNA concentrations, restriction digests were successful and progression to ligating known adaptor sequences to the digested DNA fragments was possible. However, subsequent PCR amplification of adaptor-ligated DNA fragments was inconsistent suggesting that the adaptor-ligation step of the protocol was not optimal. Fragments of DNA digested simultaneously with *Mse*I and *Eco*RI and then ligated to the appropriate adaptor sequences were found to amplify slightly more consistently than those digested with *Sau*3AI.

Across all of the libraries generated here, 768 clones were sequenced of which 27 contained microsatellite sequences (3.5 %) (Section 4.3.1.2). This low frequency suggests that CA and GA repeats are uncommon in *Tenthredopsis excisa* or that the enrichment step of the protocol described in this chapter was unsuccessful. The first explanation is unlikely given that these repeat types are known to be common in other Hymenoptera (Estoup et al. 1993; Thorén et al. 1995). It is, however, highly plausible that the

enrichment protocol was not functioning optimally. Arthofer *et al.* (2005) reported 6.9 % of clones containing microsatellite repeats when using an enrichment protocol to isolate CA and GA repeat sequences from the ant *Messor structor*, which is nearly double the value obtained in the present study.

Overall, the main problem with the genomic library protocol was the quantity of DNA obtained in the initial extractions, which had a knock-on effect on the success of subsequent steps. To optimise the protocol using larger quantities of DNA would have necessitated destructive sampling of a large proportion of the sawfly material collected from field sites. Collection of additional sawfly samples at this stage of the project was constrained by the short season when sawfly adults are on the wing and available for collection (Section 1.2.2). Therefore, effort was focussed on the alternative approach of constructing an RNA library, which requires much smaller quantities of starting material (Hedley *pers. comm.*)

4.4.2 Next-generation sequencing (NGS) of a sawfly transcriptomic library yields large numbers of microsatellite markers

The enriched library method gave low returns in terms of polymorphic microsatellite markers isolated (1 from 768 sequenced clones) for the expense (approximately £5,000) and level of effort involved (18 months work) (Section 4.4.1). The increasing availability of new sequencing methods was explored as an alternative approach. The next-generation pyrosequencing technology, commercialised by Roche/454 (Margulies *et al.* 2005), generates large volumes of sequence data very quickly. Most importantly, sequence data generated by 454 pyrosequencing can be assembled *de novo* i.e. the short sequence reads generated can be assembled into contigs without relying on an existing sequenced genome for alignment, making it ideal for insects, such as sawflies, for which there is little existing genetic information.

In eukaryotes, microsatellites are known to be more abundant in the non-coding regions of the genome compared to the coding regions (Hancock 1995). However, any markers isolated from the transcriptome should more readily cross-amplify as these regions are likely to be more highly conserved between species. This provided another reason to pursue an RNA library.

As reported in Section 4.2.2, cDNA synthesised from sawfly larval mRNA was submitted for 454 pyrosequencing as opposed to that from sawfly adult tissue. The sawfly adult samples available were unsuitable for RNA extraction as they were automatically ethanol-fixed by the Malaise trap sampling method (Section 2.1.1). Therefore, sawfly larvae, captured by sweepnetting and stored live at -80 ºC (Section 2.1.4) were used. Unfortunately, sawfly larvae are difficult to identify due to a lack of full published descriptions and illustrated keys (Barker 1998). In addition, storage at -80 ºC affected the morphological characteristics of the larvae therefore the actual species used were not identified.

After the final assembly, 1,039 of the resulting sequences (contigs + singletons) were found to contain one or more microsatellite sequences (6.9 SSRs identified for 100 ESTs), which was double the yield achieved with the enriched genomic library method and on a much shorter timescale (12 weeks compared to 78 weeks, respectively). In the first instance, the isolated markers (72 microsatellite loci) were tested on the same *Tenthredopsis excisa* population used for the genomic library method. *T. excisa* had been selected as the study species as it was the most abundant species present at all sampling sites in the 2008 field season (Section 3.3.1) and it was likely that the larval samples used to isolate SSRs belonged to the same species. When difficulties arose in amplifying the markers in *T. excisa*, a second test population of *Dolerus aeneus* individuals was used, which was the second most abundant species present at all of the 2008 sampling sites, and the markers amplified successfully.

Cross-amplification of the isolated markers into other sawfly species was tested and was successful in the four *Dolerus* species but not in the more distantly related genera of the Tenthredinidae (Section 4.3.2.3.3). Lack of cross-amplification outwith the genus *Dolerus* in the present study does not rule out cross-amplification in other more closely-related sawfly genera that were not tested.

Due to the small numbers of markers obtained from the enriched genomic library method, i.e. only one detectable polymorphic microsatellite locus, these were not tested for cross-amplification in closely-related species. Therefore, it is not possible to compare the utility of markers isolated using the genomic and transcriptomic approaches directly in terms of cross amplification. However, when the clonal sequences (those containing microsatellites) isolated using the genomic library method were BLAST searched against

the public sequence database (NCBI), no hits to insect sequences were obtained. In contrast, 89.8 % of the microsatellite-containing contigs from the 454 pyrosequencing had hits to insect sequences suggesting that these sequences are conserved among the class Insecta and therefore more likely to yield cross-amplifying markers.

4.4.3 A polymorphic set of microsatellite markers for the farmland sawfly Dolerus aeneus

The 13 polymorphic microsatellite loci reported in this chapter generated diversity (PIC) values (min: 0.117, max: 0.850) indicating that the majority of these loci are sufficient for population genetic analysis. PIC values of greater than 0.5 indicate highly informative loci, 0.5>PIC>0.25 indicates reasonably informative and PIC values less than 0.25 indicate slightly informative (Botstein *et al*. 1980). Microsatellite loci isolated from the mosquito *Aedes aegypti* showed a similar range of PIC values to those reported here (min: 0.171, max: 0.867) and were useful in revealing genetic differentiation between populations of the species (Lovin *et al*. 2009).

The commonly used diversity statistics, observed and expected heterozygosity, were not calculated using this test set of individuals as this statistic can only be generated when there is a significant proportion of diploid individuals in the test population. However, only one diploid female individual was present in this test population. Due to the presence of haplodiploidy in *Dolerus aeneus*, the other sampled individuals were most likely to have been haploid males. If Complementary Sex Determination (CSD) (Section 1.2.3) also operates in this sawfly species (i.e. in addition to haplodiploidy), it is possible that some of these males were in fact diploid. Unfortunately, the GeneMapper® software used to type the microsatellite loci is not capable of distinguishing between a haploid individual and a diploid homozygote and, based on the 13 SSR loci, did not detect any heterozygous males (which would have confirmed the existence of diploid males). For the purposes of this study, the male individuals were therefore assumed to be haploid.

However, the diploid female displayed heterozygosity at six of the 13 loci tested indicating that greater sampling effort to collect more diploid individuals could have allowed calculation of observed and expected heterozygosity, as has been reported for the sawfly *Cephus cinctus* (Hartel *et al*. 2003).

4.5 Conclusion

This chapter compares the efficiency of the traditional enriched genomic library method and the relatively new Next-generation Sequencing method for the isolation of polymorphic microsatellite markers. This study has demonstrated that 454 pyrosequencing is a fast, cost-effective and low labour method of microsatellite discovery in a non-model species. The transcriptome assembly presented here, in addition to its primary function as a marker resource, serves as the first large-scale sequencing project in any sawfly species and a stepping-stone to further molecular research in sawflies, for example in gene expression studies and comparative genomics.

The 13 polymorphic microsatellite loci described in this chapter will now be used to determine the presence of diploid males in *Dolerus aeneus* (Chapter 5) and to determine the levels of genetic variation and the dynamics of gene flow in *Dolerus aeneus* populations (Chapter 6).

5. Detecting Complementary Sex Determination (CSD) in *Dolerus aeneus*

5.1 Introduction

5.1.1 Complementary Sex Determination

Insects of the order Hymenoptera are characterised by haplodiploid sex determination where males are haploid and females are diploid (Section 1.2.3). Complementary Sex Determination (CSD), as described in Section 1.2.3.1, is one such mechanism by which haplodiploidy can be achieved and, under conditions that promote inbreeding, can lead to the production of diploid males (Heimpel and de Boer 2008). Diploid males can be effectively sterile or exhibit reduced reproductive success (Armitage *et al.* 2010; Cook and Crozier 1995; Whiting 1943). The presence of diploid males can, therefore, have far-reaching consequences at the population or species level including an increased extinction risk (Zayed and Packer 2005). Detecting the presence or confirming the absence of CSD in *Dolerus aeneus* is likely to be significant for predicting any threat to the survival of this species.

5.1.2 Detecting diploid males

If arrhenotokous haplodiploidy was the sex determination mechanism for a given species (within the Hymenoptera) then all individuals present should be either diploid females or haploid males (Heimpel and de Boer 2008). The presence of diploid males indicates that Complementary Sex Determination (CSD) operates within the species.

A number of methods have been employed to detect diploid males in the Hymenoptera to determine whether CSD is in operation. In a search of the recent literature, two methods were found to be common practice for the detection of diploid males; flow cytometry and microsatellite analysis. Both of these methods can be applied to detect diploid males in sawfly populations.

5.1.2.1 Flow cytometry

The term flow cytometry is derived from the measurement (metre) of single cells (cyto) as they flow within a liquid stream through a focused light source. Cells to be analysed are stained with different fluorescent dyes depending on the target cell structure.

When fluorescently stained cells are passed through the light source, termed "events", they produce signals in the form of light scatter and emitted fluorescent light. These light signals (photons) are detected and converted to a voltage pulse which is proportional to the original number of photons. The voltages are amplified on a linear scale prior to passing through an analogue-to-digital converter which automatically assigns each signal a specific channel number (between 0 and 1023). This channel number is proportional to the original amount of fluorescence or light scattered. In the present study, nuclear DNA is the cellular component of interest and the amount of fluorescence emitted by a cell stained with a DNA-binding fluorescent dye is directly related to the nuclear DNA content.

Dolerus aeneus females are diploid, irrespective of the presence or absence of CSD. Therefore, the nuclear DNA content of a *D. aeneus* female determined by flow cytometry can be used as a reference value for determining the relative nuclear DNA content of *D. aeneus* males. It should then be possible for the males to be designated as haploid or diploid.

5.1.2.2 Microsatellite analysis

Microsatellites, currently the most popular source of genetic marker (Sharma *et al.* 2007; Zane *et al.* 2002), are traditionally used for genetic diversity studies. However, microsatellite loci have also been used to detect diploid males in the Hymenoptera including the leaf-cutting ant *Atta sexdens* (Armitage *et al.* 2010) and the solitary wasp *Euodynerus foraminatus* (Stahlhut and Cowan 2004). Male individuals are genotyped at a defined number of microsatellite loci (Chapter 4) and are classified as diploid if two alleles are detected at one or more microsatellite loci.

5.1.3 Chapter summary

In this chapter two methods are tested to determine the level of ploidy in *Dolerus aeneus* males. Firstly, a direct measure of ploidy level is attempted using flow-cytometric analysis of nuclear DNA content. Secondly, diploidy is inferred from observations of heterozygosity at a number of microsatellite loci. The relative success of each method is evaluated and the likelihood of Complementary Sex Determination operating in *D. aeneus* is discussed.

5.2 Materials and Methods

5.2.1 Flow-cytometric analysis

5.2.1.1 Dissection of sawfly brain tissue

Sawfly heads were excised from the bodies with a sharp sterile scalpel and placed into a Petri dish half-filled with 1X PBS (137 mM NaCl, 1.5 mM KH_2PO_4, 8.1 mM $Na_2HPO_4.12H_2O$, 2.7 mM KCl). The Petri dish was then placed under a dissection microscope set to approximately X40 magnification. The head of the sawfly was held in place with a dissection needle, and the exoskeleton removed from the ventral surface of the head using microdissection forceps. Exposed brain tissue was recognised with the aid of a 3-dimensional image of the honeybee brain (Plate 5.1) (Haddad *et al.* 2004). Brain structures were easily distinguished due to their smooth texture and white colouration.

Brain tissue from each individual was excised with microdissection forceps, placed into a separate sterile 1.5 ml microcentrifuge tube and covered with a drop of 1X PBS to prevent desiccation.

5.2.1.2 Preparation of a single-cell suspension

Sawfly neuronal cells were prepared for flow-cytometric analysis working with two samples at a time (i.e. two microcentrifuge tubes each containing brain tissue from a different individual). This helped to keep to the strict timing required in this process to ensure that cells were not ruptured due to overexposure to enzymatic activity.

The brain tissue was suspended in Drosophila Ringer's Solution (DRS) (100 µl) (130 mM NaCl, 5 mM KCl, 1.5 mM $CaCl_2$, 2 mM Na_2HPO_4 supplemented with 0.25 mg/ml Trypsin, 0.25 mg/ml Chymotrypsin and 0.1 mg/ml BSA) (Butcher 1998). Each sample was drawn up and down a 25 gauge needle (i.e. a needle with an inner diameter of 0.26 mm) several times, occasionally using the tip of the needle to disperse any larger tissue clumps. At this stage and throughout the protocol a different needle was used for each sample.

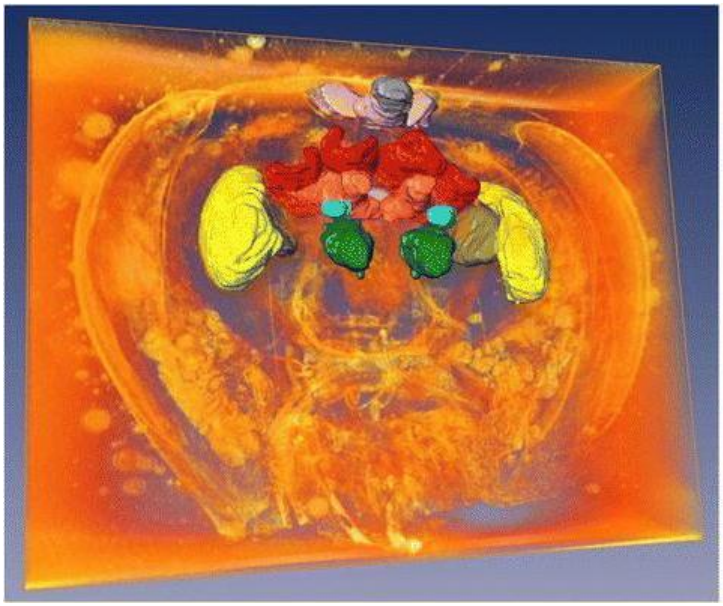


Plate 5.1: Volume rendering of a 3-dimensional dataset from an NMR (nuclear magnetic resonance) image of the honeybee brain (Haddad* et al*. 2004). Surface reconstructions have been overlain onto the semi-transparent image to show the relative orientation of the brain structures in relation to the bee's head capsule.

Ice-cold 70% ethanol (500 µl) was added to each sample and the mixture was immediately drawn up and down the 25 gauge needle to prevent re-aggregation of the cells. The addition of the ethanol halts enzyme function, fixes the cells and permeabilises them. Cell suspensions could be stored at this point for several hours to facilitate the preparation of multiple samples in the following steps.

Each mixture was then pipetted into a 12×75 mm polypropylene tube (BD Pharmingen, Oxford, UK) and wash buffer (1X PBS + 1% BSA) (500 µl) was added. The tubes were then centrifuged at 400 rcf for 5 minutes and the supernatants discarded. This wash step was repeated a second time and the supernatants again discarded. Stain solution (1X PBS + 1% BSA supplemented with 50µg/ml Propidium Iodide and 50 µg/ml RNase A) (250 µl) was then added to each tube and very slowly pipetted up and down to resuspend the cells. The cell suspensions were then incubated at room temperature for 20 minutes in darkness to allow the RNase to act and the Propidium Iodide to bind to the DNA.

5.2.1.3 Flow-cytometric analysis

Single-cell suspensions were prepared, as described in Section 5.2.1.2, from the brain tissue of 150 *Dolerus aeneus* individuals sampled in 2010 (Section 3.2.2.3) and were analysed using a FACSCalibur flow cytometer (Becton, Dickinson and Company, Oxford, UK). Excitation of propidium iodide bound to DNA was achieved using a 488 nm laser light source and fluorescence was detected in the FL2 parameter (585±40 nm). CellQuest Pro software was used to acquire the data (Becton, Dickinson and Company, Oxford, UK) and FlowJo software (Tree Star, Inc., Oregon, USA) was used to interpret the data.

Due to time constraints it was not possible to prepare and run all samples on the same day therefore it was necessary to prepare control samples to run at the beginning of each batch to calibrate the instrument. In this case the control samples were single-cell suspensions from the brain tissue of *Dolerus aeneus* females, which are known to be diploid. As is common practice when using flow cytometry to analyse DNA content, the 2N peak (in diploid female controls) was set at channel 400 in the FL2-Area parameter to allow good resolution of 1N (haploid) peaks and doublet peaks.

For each sample, events representing connective tissue cells or "debris" (i.e. material from lysed cells) were excluded from the final analysis by plotting forward scatter (FSC) against side scatter (SSC) for each event and gating to exclude these events (Figure 5.1). Forward scatter is light scattered in the direction of the laser beam and gives an indication of cell size. Side scatter is light scattered perpendicular to the laser beam and gives an indication of shape/texture. These parameters are commonly used in flow cytometry for the identification of cells and the exclusion of debris.

"Doublet" events (cells joined together) were also excluded from the final analysis as they can contribute to inaccurate readings of DNA content i.e. a doublet of two 1N cells will have the same nuclear DNA content as a single 2N cell. Doublet events have a greater pulse width than a single cell event as doublets take longer to pass through the laser beam and can thus be detected and excluded from the analysis. The area (FL2-A) of the emitted fluorescent light pulse was plotted against the width (FL2-W) of the light pulse for each event and doublets were removed by gating to exclude them (Figure 5.2).

After debris and doublet events had been excluded, the DNA content of the remaining events was plotted on a graph for each sample. The position of the 2N peak for the control sample (i.e. a diploid female) was recorded and all other samples processed in the batch were characterised by the position of their DNA peaks relative to that of the control (Figure 5.3). Haploid males will have a DNA content approximately half that of the control and a diploid male should have a DNA content more or less identical to the control. Samples were deemed to have 'failed' when either no events were detected in the sample at all, or there was no discernable DNA peak after gating.

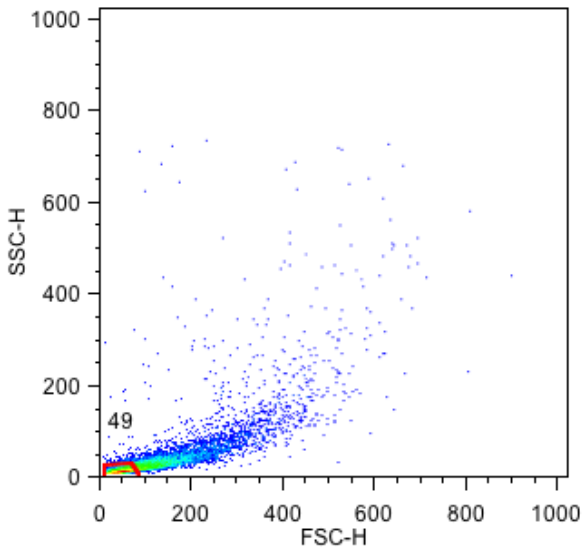


Figure 5.1: Example of a forward scatter (FSC) versus side scatter (SSC) plot for the exclusion of debris and connective tissue. The gate was set to ensure that the majority of DNA-containing material was located in the gated region. In this example 49% of events have been gated by the red line and are selected for further analysis.

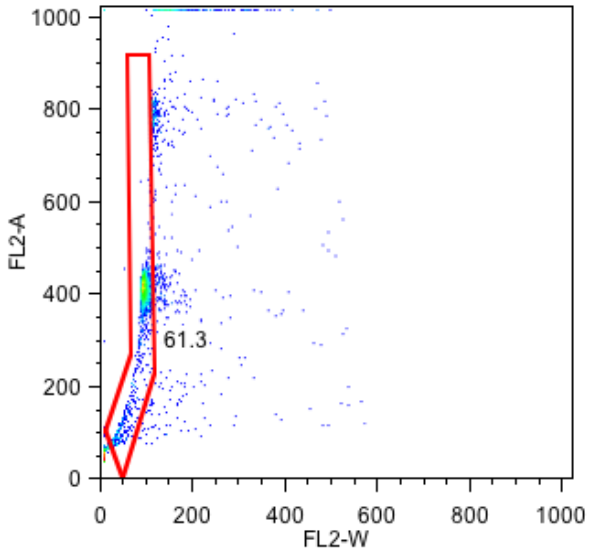


Figure 5.2: Example of a plot used to identify single cells showing the area of the emitted light pulse (FL2-A) versus the width of the emitted light pulse (FL2-W). FL2-A represents a measure of the fluorescence intensity (i.e. DNA content) whereas FL2-W indicates the time taken for a particle to traverse the laser beam. Doublets and clumps of cells will have higher FL2-W than single cells and can therefore be excluded from further analysis. In this example 61.3% of events were selected within the red line for analysis of DNA content.

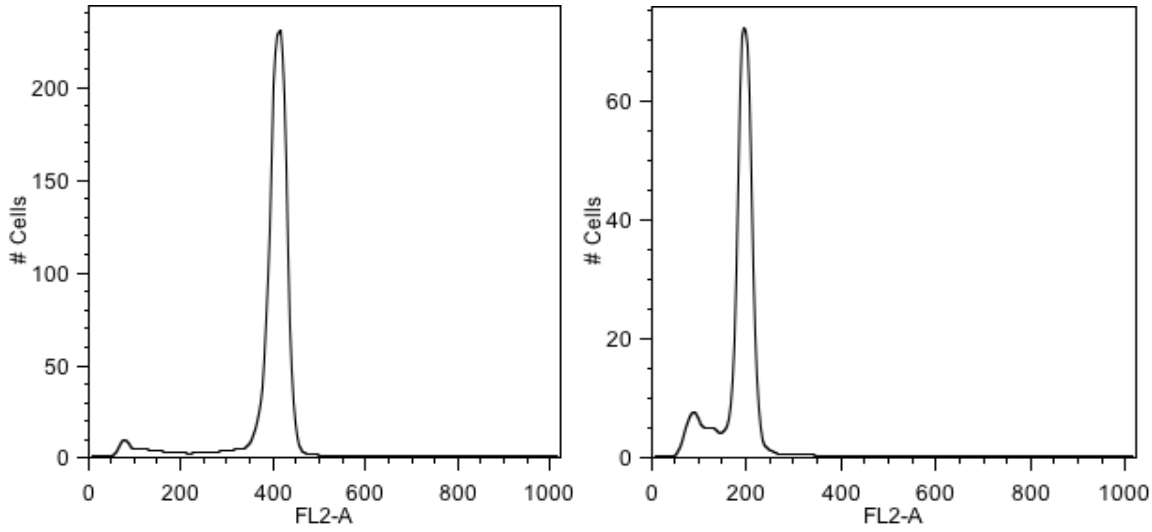


Figure 5.3: Left: relative number of cells plotted against DNA content (FL2-A) for a sample taken from a* Dolerus aeneus *female individual (diploid control). Right: the same data for a sample taken from a male* Dolerus aeneus *individual. The DNA content is approximately half that of the control sample and therefore the individual is haploid.

5.2.2 Microsatellite analysis

5.2.2.1 Microsatellite genotyping

Thirteen polymorphic microsatellite loci (Chapter 4; Table 4.6) isolated for use in *Dolerus aeneus* as described in Chapter 4 were assayed on 44 male *Dolerus aeneus* individuals sampled in 2009 and 130 *D. aeneus* males sampled in 2010.

PCR was carried out in a total reaction volume of 10 µl containing approximately 15 ng total genomic DNA from each individual with 1 µM fluorescently-labelled forward primer and 1 µM reverse primer under conditions described in Section 2.2.3. PCR reaction mixtures were subjected to an initial hold of 94 ºC for 5 minutes followed by 35 cycles of 94 ºC for 30 seconds, 58 ºC for 30 seconds and 72 ºC for 1 minute, followed by a 72 ºC final extension and a 8 ºC hold. Fluorescent PCR products were then diluted 1 in 10 and combined with 8.84 µl Hi-Di™ Formamide and 0.16 µl GeneScan™ 500 Rox™ internal size standard (both Applied Biosystems Inc., Warrington, UK) and visualized using the ABI PRISM® 3730 Genetic Analyzer (48 capillary). Fragments were analysed using GeneMapper® software (Applied Biosystems Inc., Warrington, UK).

Individuals were designated diploid if two alleles were detected at one or more of the thirteen loci. Any detected diploid individuals were re-amplified and alleles scored using the same protocol along with a subset of haploid individuals to determine the genotyping error rate.

5.2.2.2 Calculating the power of microsatellite markers for diploid male detection

Non-detection of diploid males can occur if the subset of microsatellite markers used exhibit low allelic diversity (low heterozygosity). To calculate the power of the subset of microsatellite markers, namely the probability that a diploid individual will be heterozygous at one or more of the markers (P_{het}), the following equation was used (Souza *et al* 2010):

$$1-\prod_{j=1}^{L}\sum_{i=1}^{N}\left(x_i^{2}\right)$$

where L = the number of loci,

N = the number of alleles at a locus and,

x = the frequency of allele *i.*

P_{het} was calculated using both a diploid female dataset (all *D. aeneus* females sampled in 2010) and a haploid male dataset (all *D. aeneus* haploid males sampled in 2010) independently. Null alleles were detected by MicroChecker within the diploid female dataset (Section 6.3.2) at loci (Saw454_4, Saw454_7, Saw454_11, Saw454_24, Saw454_31 and Saw454_41). Therefore when using the diploid female dataset the following equation, number four from Brookfield (1996), was used to calculate the null allele frequency (r) at each of these loci:

$$\frac{H_e - H_o}{1 + H_e}$$

where H_e is the expected heterozygosity and,

H_o is the observed heterozygosity.

5.3 Results

5.3.1 Flow cytometric analysis

A total of 150 single-cell suspensions prepared from *Dolerus aeneus* neuronal tissue were analysed by flow cytometry, comprising 130 males and 20 female controls. The mean number of events acquired per sample was 3,470 (range: 300–10,000 events). Of the 130 male samples processed, 76 (58.5 %) were "successfully analysed"; cell populations of a statistically robust size were not achieved. However, the data for these individuals was examined in order to give a preliminary indication of ploidy. For these 76, the mean number of events analysed after gating was 507 (range: 69-1530 events). Seventy-four of the males were designated haploid and two were designated "hyperhaploid" (DNA content higher than that of a haploid individual but lower than that of a diploid individual) (Table 5.1). No diploid males were detected.

Notably, the control sample for batch 3 failed (Table 5.1) therefore it was not possible to determine the ploidy of the male samples in that batch by comparison with a diploid control in the normal fashion. Two strategies were employed to overcome this issue. Firstly, flow cytometer instrument settings were recalled from the previous batch of analysis and these were used to analyse the samples. Although this is not 'proper practice', due to slight variations in instrument sensitivity on a day-by-day basis, it did allow us to ensure similar signal intensity from the samples analysed in this batch compared to others. Secondly, it was possible to compare the 14 successfully analysed male samples from batch 3 with each other and it was clear that 13 of these samples were very similar to each other in terms of DNA content and one had higher DNA content. Given that the frequency of diploid males is likely to be very low (Heimpel and de Boer 2008) it is more likely that these samples are all haploid rather than all diploid. Therefore, it was concluded that these 13 samples were haploid individuals and one was "hyperhaploid".

The two "hyperhaploid" individuals exhibited nuclear DNA content higher than that of a haploid individual but lower than that of a diploid individual. Hyperhaploid individuals were detected in batches 1 and 3. Elevated DNA content of the hyperhaploid

individual in batch 1 was identified by comparison with diploid and haploid individuals (Figure 5.4) in a plot of the DNA content against relative number of cells.

As the diploid control for batch 3 failed, hyperhaploid individuals were detected by calculating the percentage increase in DNA content of each individual from the haploid level, rather than the decrease in DNA content from the diploid level. Haploid samples with more than 100 cells in the final DNA analysis were selected and the modal DNA content for each sample determined. (Samples with fewer than 100 cells were omitted in order to limit increased variance in the mode due to low numbers.) The mean of these modal values was then calculated (Table 5.2). The percentage increase in DNA content was determined by comparing the modal DNA content of the hyperhaploid sample with this calculated (haploid) mean. For the hyperhaploids from batch 1 and 3, the percentage increase in DNA content was 25.1 % and 23.0% respectively (Table 5.2).

Table 5.1: Summary of male samples of **D. aeneus** *analysed by flow cytometry.*

Batch No.	No. male samples tested	No. successfully processed samples	No. haploids	No. hyperhaploids	Valid control
1	39	34 (87.2 %)	33	1	Y
2	18	9 (50.0 %)	9	0	Y
3	40	14 (35.0 %)	13	1	N
4	33	19 (57.6 %)	19	0	Y

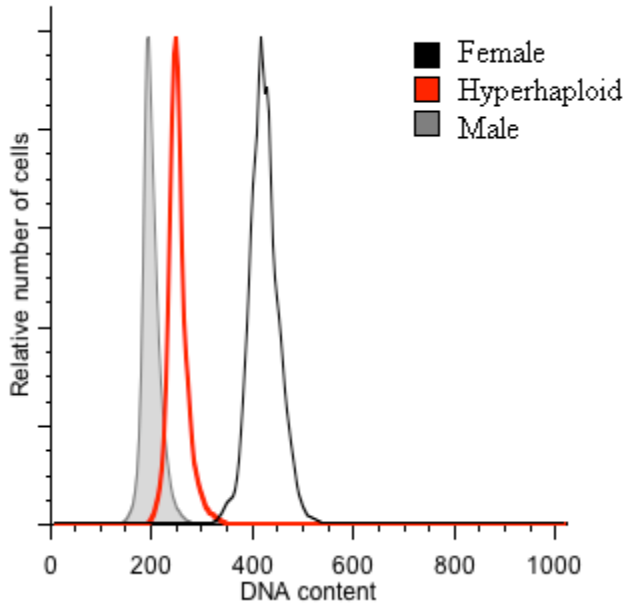


Figure 5.4: Nuclear DNA content of the hyperhaploid from batch 1 in comparison with that of a female control (diploid) and a haploid male from the same batch. The hyperhaploid male individual has DNA content approximately 25% higher than that of a haploid individual.

Table 5.2: The modal DNA contents of haploid male samples from batches 1 and 3 and the mean DNA content values compared to hyperhaploid male individuals.

Batch	Sample Index Number*	Modal DNA content	Mean of Modal DNA contents	Modal DNA content of hyperhaploid**
1	9	185	199	249 (25.1 %)
	12	197		
	21	196		
	22	195		
	24	179		
	39	196		
	40	205		
	43	167		
	44	200		
	45	189		
	52	214		
	53	200		
	55	189		
	60	195		
	61	192		
	64	208		
	69	208		
	70	215		
	71	204		
	73	198		
	74	207		
	79	207		
	86	211		
	87	199		
	102	205		
	107	213		
	120	196		
	121	208		
	122	200		
	124	200		
3	170	276	300	369 (23.0 %)
	173	278		
	180	321		
	190	323		
	193	328		
	201	303		
	219	278		
	220	314		
	226	307		
	227	285		
	234	308		
	241	338		
	243	236		

**** Only successfully processed haploid samples with more than 100 events in the final analysis are included. **Figures in parenthesis give the percentage increase in DNA content of the hyperhaploids relative to mean haploid DNA content.***

5.3.2 Microsatellite analysis

5.3.2.1 Microsatellite genotyping

Of the 44 *Dolerus aeneus* males sampled in 2009 and assayed across 13 polymorphic microsatellite loci, five displayed heterozygosity at either locus 2 or 4 and were designated diploid (Table 5.3). Subsequently, of the 130 *D. aeneus* males, sampled in 2010, five males displayed heterozygosity at either locus 1, 4 or 23 and were designated diploid (Table 5.3). The results from locus Saw454_2 were disregarded for the 2010 individuals as it was not possible to score alleles unambiguously on this dataset.

Allele scoring across duplicate analyses of the same individual-locus combinations was identical therefore extremely low genotyping error rates were estimated.

5.3.2.2 Calculating the power of microsatellite markers for diploid male detection

The resolving power of the microsatellite markers (P_{het}) was calculated as 0.999, to three decimal places, when using either the haploid male dataset or the diploid female dataset (with corrections for null alleles). This translates to a 99.9% chance of detecting a diploid male by heterozygosity at one or more of the loci used for analysis.

5.3.3 Combined results of flow cytometric and microsatellite analyses

At the time of running the flow-cytometric analysis, the *Dolerus aeneus* individuals collected in 2009 were no longer available to be tested. Therefore, it was only possible to compare the two methods of ploidy analysis for the *D. aeneus* males collected in 2010. Five individuals were designated diploid by microsatellite analysis, of which two were designated haploid by the flow-cytometric analysis (Table 5.4). For the remaining three individuals, flow-cytometric analysis failed to give a result.

Table 5.3: Detection of diploid males at each microsatellite locus for 44* Dolerus aeneus *males collected in 2009 and 130* D. aeneus *males collected in 2010.

	No. diploid males detected	
Locus	**2009**	**2010**
Saw454_1	-	2
Saw454_2	4	N/A
Saw454_3	-	-
Saw454_4	1	2
Saw454_7	-	-
Saw454_11	-	-
Saw454_14	-	-
Saw454_16	-	-
Saw454_19	-	-
Saw454_23	-	1
Saw454_24	-	-
Saw454_31	-	-
Saw454_41	-	-
Total:	5	5

Table 5.4: Comparison of flow cytometry and microsatellite analysis in five* Dolerus aeneus *males, sampled in 2010, classified as diploid by microsatellite analysis.

Sample Index No.	Microsatellite Result	Flow Cytometry (FC) Locus	Flow Cytometry (FC) Result	FC Batch No.
79	Diploid	Locus 4	Haploid	1
86	Diploid	Locus 4	Failed Sample	1
140	Diploid	Locus 1	Failed Sample	2
153	Diploid	Locus 1	Haploid	2
164	Diploid	Locus 23	Failed Sample	3

5.4 Discussion

5.4.1 Flow cytometric analysis proved difficult with the **Dolerus aeneus** *samples available*

Of the 130 single-cell suspensions prepared from *Dolerus aeneus* males, only 76 (58.5%) samples produced detectable DNA peaks (Section 5.3.1). The high failure rate was due to a combination of reasons including insufficient material, the presence of high levels of debris, and few intact cells/nuclei remaining in the sample. The success of the cell isolation procedure from brain tissue was probably compromised by prior fixation of samples in ethanol, necessitated by the Malaise trap sampling method (Section 2.1.1). Disaggregation of cells from fixed tissue requires greater mechanical force compared to fresh tissue, resulting in increased cellular damage. In addition, ethanol-fixed tissue will not necessarily show the same susceptibility to enzymatic cleavage as unfixed tissue, and it is possible that the digestion buffer, which was adapted from a study using flow cytometric analysis to determine the ploidy level of *Diadegma chrysostictos* (Hymenoptera: Ichneumonoidea) individuals (Butcher *et al.* 2000), was unsuitable for use on fixed tissue.

Furthermore, the tissue samples in the present study were very small and the number of cells present in the suspensions was correspondingly low. Brain tissue, although present in small amounts, was considered appropriate for this study as the likelihood of contaminating bacterial cells in neuronal tissue is low (Hubbard *pers. comm.*). Also, the vast majority of nerve cells should be permanently arrested in the G0 phase of the cell cycle giving a more accurate estimation of ploidy level than a tissue sample with a higher proportion of mitotic cells. The protocol for flow-cytometric analysis used here was optimised initially using a different species of sawfly, *Tenthredopsis nassata*, which is marginally larger than *Dolerus aeneus* with a correspondingly larger brain. Sample failure rate with the test species was much lower (21.4 %) compared to with the failure rate for the *D. aeneus* samples analysed in the present study (41.5 %; (Section 5.3.1).

Flow-cytometric analyses would typically aim to acquire a minimum of 10,000 events per sample (considered a statistically robust population; Clarke *pers. comm.*). The

mean number of events acquired per sample in the present study was 3,470 (Section 5.3.1), and the number of events included in the final analysis for the successfully processed samples was even smaller (mean of 507). Therefore, the flow cytometry results presented in this chapter should be viewed with caution. Inevitably cells are lost during preparation and some data points are subsequently gated to exclude polyploid cells and debris. Higher densities of cells in the cell suspension might have been achieved with larger tissue samples from each individual and could have decreased the failure rate.

The flow cytometry approach detected two hyperhaploid individuals with DNA contents 25.1 % and 23.0 % higher than that of a haploid individual (Section 5.3.1). Several explanations are possible for the presence of these hyperhaploid individuals. Firstly, the ratio of DNA dye to cellular DNA might have been higher in these samples, which would result in an increased labelling density of the DNA. This is highly unlikely as saturating levels of propidium iodide are used in all preparations and therefore the dye:DNA ratio should be consistent over a wide range of cell densities. Secondly, the hyperhaploids may have an increased chromosome number relative to their counterparts due to some genetic condition. This possibility cannot be confirmed or discounted as information relating to the genome size of *Dolerus aeneus* and genetic disorders in the species is not available. Thirdly, these individuals may have been misidentified and belong to a different species with a larger genome size. This, unfortunately, cannot be checked because dissection of the brain tissue destroys head capsule morphology required for species identification.

5.4.2 Microsatellite analysis highlighted potential diploid males in Dolerus aeneus

Forty-four *Dolerus aeneus* males sampled in 2009 and 130 *D. aeneus* males sampled in 2010 were successfully genotyped across 13 microsatellite loci (Section 5.3.2). Five of the 2009 males were designated as potential diploids along with five of the 2010 males.

However, none of the ten male individuals designated diploid displayed heterozygosity at more than one locus. Whereas, all the female individuals (2009 and 2010 samples combined) sampled were heterozygous at an average of 4.51 loci (min = 1, max = 8) (Appendix 1) which is a possible indication that the heterozygosity in these

males may be attributable to error. However, the genotyping error rates were estimated to be extremely low (Section 5.3.2.1).

It is possible that these males are heterozygous at only one locus purely by chance but it is also possible that the microsatellite markers which detected heterozygosity in the male samples (Saw454_1, Saw4542_, Saw454_4 and Saw454_23) are part of a gene family. Therefore, in these male individuals one allele could be PCR-amplified from one locus and the second allele from another locus, meaning that the individual is in fact haploid though appearing diploid (heterozygous at one microsatellite locus) in the GeneMapper® output. In order to determine if this is the case, it would be necessary to clone and sequence the PCR products from the individual-locus combinations in question along with a subset of controls. This check, although valuable given the low number of heterozygous loci in the potential diploid males, would present an extra expense. In addition, if the allele scoring at these loci is questionable in the "diploid" males for the reasons stated, this would mean questioning these loci across all samples genotyped and indeed, all studies in the literature using microsatellite markers.

5.4.3 A comparison of microsatellite and flow cytometric analyses as methods of determining the ploidy of male sawflies

Flow-cytometric analysis of ploidy in *Dolerus aeneus* males was unreliable with the samples available in the present study. If flow cytometry could be performed on fresh samples (i.e. not ethanol-fixed prior to preparation) and a variety of different enzymes tested for breaking up the tissue, it is likely that the quality of the resulting single cell suspensions could be improved and that flow cytometry would be more successful.

Good quality genotyping data was obtained from microsatellite analysis quickly and with relative ease in comparison with the flow cytometry approach. However, the ploidy of the detected "diploid" males is questionable (Section 5.4.2). Without further investigation it is not possible to confirm whether or not these individuals are in fact diploid males.

5.4.4 The presence of Complementary Sex Determination (CSD) in **Dolerus aeneus**

Inferring the operation of Complementary Sex Determination (CSD) in a haplodiploid species by confirming the presence of diploid males has been commonplace in the literature (van Wilgenburg *et al.* 2006) The potential presence of diploid males in *D. aeneus* indicates that CSD could be operative in this species although further research to confirm the results of the microsatellite analysis would be required to confirm this.

However, if these diploid males were confirmed as a true result, i.e. by cloning and sequencing of suspect PCR products, it should be noted that diploid males have been found in species where CSD has been shown to be absent and are as a result of mutation or the hybridisation of two species (van Wilgenburg *et al.* 2006). For example, in the parasitoid *Nasonia vitripennis* a mutant strain exists where females are triploid. Unmated triploid females can lay both haploid and diploid eggs that typically develop into viable, fertile males with the diploid males producing diploid sperm (Beukeboom and Kamping 2006). Also, Molbo *et al.* (2004) recorded diploid males produced as a result of matings between two subspecies of the fig wasp *Pegoscapus hoffmeyeri*. Therefore, caution is required when directly inferring the operation of CSD from the presence of diploid males and these examples illustrate the need to confirm CSD via other means such as breeding experiments and/or molecular techniques (van Wilgenburg *et al.* 2006).

5.4.5 Diploid male production rate in **Dolerus aeneus**

Once diploid males have been confirmed in a species (and CSD confirmed by additional means), the next logical step is to determine the frequency of diploid male production in native populations in order to help predict how CSD will affect the genetic diversity and population dynamics in the future.

Assuming for illustrative purposes that the “diploid” males detected in the present study are true diploids; the diploid male production rate (DMR: the percentage of diploid individuals that are male) in the Scottish *Dolerus aeneus* population was 22.72 % (2009) and 15.15 % (2010) respectively. These figures are considerably higher than that recorded by Fujiwara *et al.* (2004) in their study of the turnip sawfly *Athalia rosae* (0.08 %) where 1306 diploid individuals were sampled from the field. However, without knowledge of a

range of other factors affecting the DMR in either species (i.e. effective population size, dispersal rate, sex allele diversity, viability of diploid males etc.) it is difficult to say whether a relatively high DMR observed in *D. aeneus* could be due to increased incidence of inbreeding relative to that in *A. rosae*.

In future studies, it would be beneficial to use an optimised flow cytometry approach in combination with microsatellite analysis to detect diploid males. Thereby eliminating the question of whether males that are heterozygous at fewer loci than females are in fact diploid. A combined analysis with an optimised flow cytometry componentt, in addition to more information relating to the characteristics of *D. aeneus* populations, would permit data such as that gathered in this study to confirm the presence of CSD and to estimate the DMR.

5.5 Conclusion

This chapter compares the success of flow cytometric analysis and microsatellite analysis for the detection of diploid males in *Dolerus aeneus* and therefore the potential presence of Complementary Sex Determination (CSD). This study has highlighted the potential presence of diploid males in *D. aeneus* and the requirement for a highly-optimised protocol that combines both microsatellite analysis and flow cytometry.

Flow cytometric analysis failed to detect diploid males in this study, but the method could be improved by using fresh (non-ethanol-fixed) samples for analysis and by optimising the reagents involved in the preparation of single cell suspensions. An optimised flow cytometry method, used in combination with SSRs in future studies, would eliminate any questionability surrounding the ploidy of individuals which exhibit heterozygosity at a single microsatellite locus. This would enable accurate data to be collected to confirm or deny the presence of CSD and on the frequency of diploid male production to parameterise models of sawfly population dynamics.

6. Population structure and genetic diversity in the farmland sawfly *Dolerus aeneus* (Hymenoptera; Symphyta)

6.1 Introduction

6.1.1 Sawfly population decline

A long-term monitoring study initiated in 1970 by the Game and Wildlife Conservation Trust highlighted that populations of farmland sawflies are in decline. Specifically, over the period 1970 to 1990 sawfly numbers decreased at a rate of 4.4 % per year (Aebischer 1991). These declines correlate significantly with agricultural intensification procedures such as a decline in the practice of undersowing (temporary grassland within a mixed arable/ grass ley rotation) and the increased use of agrochemicals in the arable landscape (Aebischer 1991).

Although it is known that populations of British farmland sawflies are decreasing in size, and reasons to explain this have been put forward, there have been no previous studies that examine the genetic factors involved in sawfly population decline. In general, populations of any organism are rarely driven to extinction before genetic factors have time to take effect (Frankham 2005). Therefore, if the genetic factors involved in sawfly populations are ignored it is possible that their extinction risk could be underestimated and inappropriate recovery strategies applied.

6.1.2 The genetic effects on sawfly populations

The processes contributing to agricultural intensification are likely to lead to fragmentation of suitable sawfly habitat. Given that sawflies are thought to disperse poorly (Benson 1950), it is possible that their populations have become isolated as their habitat fragments. Small, isolated populations can be subject to increased levels of inbreeding and genetic drift (stochastic loss of alleles). Both of these processes lead to a loss of genetic diversity which results in reduced evolutionary potential and a heightened extinction risk (Spielman *et al.* 2004).

It is important to determine the extent to which sawfly populations have become fragmented in order to identify threatened populations. Fragmentation of habitat can induce higher levels of population substructuring in threatened species via reduced migration and gene flow between habitat fragments, a major cause of inbreeding and a potential pathway to extinction for an isolated population (Höglund 2009).

6.1.3 Population genetic analyses and conservation

Population genetic analyses such as estimates of allelic richness and heterozygosity levels (a reduction in heterozygosity is an indicator of the occurrence of inbreeding) at certain genetic loci can help to identify threatened populations (Höglund 2009). Principally it is the allelic diversity at loci with functional importance that will affect the ability of a population to respond to selection. However, the information obtained from molecular markers such as microsatellites (Chapter 4) can give an estimation of the overall levels of diversity and gene flow currently present within and between populations (Höglund 2009). Researchers have used microsatellite markers with success to determine the extinction risk of other hymenopteran insects such as the solitary bees *Colletes floralis* (Davis *et al.* 2010) and *Andrena fuscipes* (Exeler *et al.* 2010).

6.1.4 Chapter summary

This chapter represents the first genetic analysis of farmland sawflies in Britain, with the intention of using the data to promote their conservation. The levels of genetic diversity within and between populations of the graminivorous sawfly *Dolerus aeneus* are quantified and levels of genetic differentiation between populations are assessed. These analyses are carried out using a set of 13 polymorphic microsatellite markers developed specifically for use in *Dolerus aeneus* (Chapter 4). These studies will help to determine the extinction risk of *Dolerus aeneus* populations.

6.2 Materials and Methods

6.2.1 Sampling of Dolerus aeneus

Sawflies were sampled using Malaise traps (Section 2.1.1) in two consecutive years following the sampling programmes described in Section 3.2.2.2 (2009) and Section 3.2.2.3 (2010). *Dolerus aeneus* individuals were identified from the weekly Malaise trap collections as described in Section 2.1.2. The sex of each individual was determined as described in Section 2.1.3 and the ploidy of the male individuals was examined via flow cytometry and/or microsatellite analysis (Chapter 5).

Individuals sampled with different traps but from the same locality (farm) were considered as representing a single population. However, individuals sampled in 2009 were not combined with those sampled in 2010. In 2009, three Scottish populations were sampled: Mains of Glamis, The James Hutton Institute (TJHI) and New Gilston. In 2010, the same three Scottish populations were sampled and, in addition, three English populations: Claxby St Andrew (Lincolnshire), Harper Adams (Shropshire) and Down Farm (Sussex).

The total number of *Dolerus aeneus* individuals sampled from each population, the date of collection (the day the collection bottle was taken from the Malaise trap) and the sex and ploidy of each individual was recorded.

6.2.2 Microsatellite genotyping

Thirteen polymorphic microsatellite loci isolated for use in *Dolerus aeneus* (Chapter 4) were assayed on all individuals sampled in 2009 and 2010. PCR was carried out in a total reaction volume of 10 µl containing approximately 15 ng total genomic DNA from each individual with 1 µM fluorescently-labelled forward primer and 1 µM reverse primer under conditions described in Section 2.2.3. PCR reaction mixtures were subjected to an initial hold of 94 ºC for 5 minutes followed by 35 cycles of 94 ºC for 30 seconds, 58 ºC for 30 seconds and 72 ºC for 1 minute, followed by a 72 ºC final extension for 5 minutes and a 8 ºC hold. Fluorescent PCR products were then diluted 1 in 10 and combined with 8.84 µl Hi-Di™ Formamide and 0.16 µl GeneScan™ 500 Rox™ internal

size standard (both Applied Biosystems Inc., Warrington, UK) and visualized using the ABI PRISM® 3730 Genetic Analyzer (48 capillary). Fragments were analysed using GeneMapper® software (Applied Biosystems Inc., Warrington, UK).

The presence or absence of null alleles (alleles that fail to amplify by PCR) was investigated using the software MicroChecker 2.2.3 (Van Oosterhout et al. 2004). MicroChecker accepts datasets composed of diploid individuals and therefore a dataset comprising the allelic composition of all of the diploid females sampled in 2010 was used.

6.2.3 Genetic diversity analysis

6.2.3.1 Hardy-Weinberg Equilibrium and genotypic linkage disequilibrium

Tests for genotypic linkage disequilibrium (non-random association of alleles) between pairs of loci were carried out in FSTAT v. 2.9.3.2 (Goudet 1995). This was accomplished using a dataset detailing the allelic composition at 12 microsatellite loci of all haploid males sampled in Scotland in 2010 (the largest sample of individuals with the same ploidy level). Departure from Hardy-Weinberg Equilibrium (HWE) at 12 microsatellite loci within each of the Scottish populations was tested for in both FSTAT v. 2.9.3.2 (Goudet 1995) and GENALEX v. 6.4.1 (Peakall and Smouse 2006). The former software package uses a randomisation technique whereas the latter uses a Chi square method. Analysis was performed using a dataset composed of all the known diploid individuals (females + potential diploid males) sampled in 2010 (the largest sample of diploid individuals).

6.2.3.2 Allelic richness

Allelic richness (the number of alleles) at each of 12 microsatellite loci was calculated for the three Scottish populations in both sampling years using the haploid male individuals (the largest sample of individuals with the same ploidy level). Allele counts were made using Microsatellite Toolkit (Park 2001) then allelic richness was quantified by rarefaction using the online rarefaction calculator provided by the University of Alberta (http://www.biology.ualberta.ca/jbrzusto/rarefact.php). Rarefaction on the 2009 and 2010 datasets was based on a sample size of five and 17 respectively. The sample size for the

2010 dataset is derived simply from the sample size from the smallest population in that year (New Gilston). The sample size for the 2009 dataset is the sample size for the smallest population minus one (only five individuals could be scored at all microsatellite loci).

Allelic richness was not quantified using any of the diploid datasets (females and females + potential diploid males) due to the small sample sizes of these individuals in both sampling years. The use of rarefaction on the diploid datasets would have meant working with the sample size of the smallest population which would be only one or two individuals.

To test for significant difference in mean allelic richness over all loci between populations SigmaStat software (v 3.1.0) (Systat Software Inc., London, UK) was used to run an Analysis of Variance (ANOVA).

6.2.3.3 Comparisons of allele frequencies within and between populations

For the three most polymorphic loci (Saw454_1, Saw454_4 and Saw454_24) allele frequencies were plotted using the haploid male dataset (the largest sample of individuals with the same ploidy level) for the three Scottish populations. Data from different sampling years was analysed separately. In addition, the allele frequencies of the male haploid individuals and the known diploid individuals (females + potential diploid males) were compared within each of the Scottish populations at the same three loci. Again, data from different sampling years was analysed separately.

6.2.3.4 Observed and expected heterozygosity

Levels of observed and expected heterozygosity within sampled populations and for each of twelve microsatellite loci were calculated using Microsatellite Toolkit (Park 2011). Observed heterozygosity (H_O) was calculated by dividing the number of observed heterozygotes at each locus by the number of individuals in the population. The mean observed heterozygosity over all loci was then calculated. The mean expected heterozygosity (H_E) or gene diversity (the probability that any two alleles drawn from a

population are different to each other for a given locus) for each locus was calculated according to Nei's (1987) unbiased gene diversity:

$$H_E = \frac{\sum_{1}^{L} \left(1 - \sum p_i^2\right)}{L}$$

where L = the number of loci

p_i = the frequency of the ith allele at a given locus.

The mean expected heterozygosity over all loci was then calculated.

Expected heterozygosity was calculated for each of the three Scottish populations for both sampling years using datasets composed of all of the haploid individuals sampled. Subsequently, using a diploid dataset (females + potential diploid males), both expected and observed heterozygosity were calculated for each of the three Scottish populations for sampling year 2010. However, when calculating both observed and expected heterozygosity from the diploid individuals (females + potential diploid males) sampled in 2009, only the New Gilston population was used. This was due to the small sample size of diploids at Mains of Glamis (four individuals) and The James Hutton Institute (three individuals) in this sampling year.

6.2.3.5 Level of inbreeding

The inbreeding coefficient (F_{IS}; the reduction in heterozygosity of the individuals relative to the subpopulation to which they belong, or, in biological terms: the fractional reduction in heterozygosity of a sub-population relative to a random-mating subpopulation with the same allele frequencies) was calculated using FSTAT v 2.9.3.2 (Goudet 1995). F_{IS} was quantified using two datasets, one consisting of only the female individuals and a second consisting of females and potential diploid males, respectively (diploid individuals are necessary for computation of heterozygosity levels) for the New Gilston population in each sampling year for each of the 12 microsatellite loci. F_{IS} was computed for only the New Gilston population as the number of diploid individuals sampled from other populations was very low.

The significance of the F_{IS} values was determined using a randomisation-based test in FSTAT v 2.9.3.2 (Goudet 1995).

6.2.4 Analysis of population structure

6.2.4.1 Principal Coordinate Analysis

Principal Coordinate Analysis (PCA) was used for a preliminary assessment of genetic relatedness among populations of *Dolerus aeneus*. Briefly, PCA is a method of reducing the dimensionality of a dataset consisting of a large number of variables whilst retaining as much of the variation present in the dataset as possible. This is achieved by transforming the dataset to a smaller number of variables or "co-ordinates". As consecutive coordinates (axes of variation) are extracted they account for less and less of the variability in the dataset such that most of the variation present in all of the original variables is represented by the first few coordinates. With the dataset simplified in this way, plotting the first two principal coordinates will give the best possible 2D graphical representation of the similarities and differences between data points as possible (Peakall and Smouse 2010).

In this case the datasets are composed of the genotypes of individual *Dolerus aeneus* sawflies at 12 microsatellite loci. PCA was used to look for groups of genetically related individuals using three datasets (composed of haploid male individuals, diploid female individuals and females + potential diploid male individuals) respectively and to determine whether any groups of related individuals were sampled from the same population. Samples collected from Mains of Glamis, The James Hutton Institute and New Gilston in 2009 and 2010 were analysed.

Genetic relatedness matrices were generated and PCA performed on these, whereby 3 coordinates were extracted, using GENALEX v. 6.4.1 (Peakall and Smouse 2006). The extraction of 3 components was deemed sufficient after a pilot study revealed that the majority of the variation in the dataset was accounted for at this level.

6.2.4.2 STRUCTURE Analysis

To assess more thoroughly the patterns of differentiation between the *Dolerus aeneus* populations sampled in Scotland (Mains of Glamis, The James Hutton Institute and New Gilston) in 2009 and 2010 respectively, a Bayesian model-based cluster analysis was carried out using the STRUCTURE v 2.3.1 software package (Pritchard *et al.* 2000). STRUCTURE assumes that all of the genetic material from the sampled individuals comes from one or more (unobserved) user-defined populations (K) each of which is characterised by a set of allele frequencies. These allele frequencies and the population of origin of each allele copy of each individual are assumed to be unknown and are estimated from the dataset. To do this, the software uses a Markov Chain Monte Carlo (MCMC) approach.

The MCMC begins with an arbitrary configuration of parameter values (allele frequencies in each of K populations and the population of origin of each allele copy of each individual) and iteratively updates the configuration in steps conditional on the dataset and the current configuration of parameters until the most likely configuration is determined (Falush *et al.* 2007). For each MCMC step the likelihood of the current configuration of assignment of individuals to populations (given the dataset) is quantified. The number of MCMC steps used must be sufficient to allow the likelihood values to stabilise. For each MCMC run, for a given value of K, STRUCTURE calculates the posterior probability of the data (P(D)) and displays ln P(D).

The end result of the STRUCTURE analysis is that each individual is given a population membership coefficient profile (Q) which can be visualised graphically and also the "true" value of K is chosen. Traditionally, the approach for selecting the true K for the dataset in question is to adopt the value of K for which ln P(D) is maximal. More commonly, K is selected by choosing the lowest value of K for which the ln P(D) values calculated have begun to "more or less" plateau (Pritchard *et al.* 2010; Falush *et al.* 2007). However, not all datasets behave in such a manner that detecting K in this fashion will be straightforward (MacKenzie *pers. comm.*) and the authors of the programme (Pritchard *et al.* 2010) suggest that the more formal criteria for estimating the true K described by Evanno *et al*. (2005) can be used.

Evanno *et al.* (2005) suggest that, to select the true K, it may be more accurate to use their *ad hoc* statistic (ΔK). The rationale behind using this statistic is that the true K may not be obvious in some datasets simply by looking at the distribution of ln P(D) values (denoted L(K) by Evanno *et al.* 2005) whereas in calculating ΔK the true K becomes more apparent. In the first instance, the second order rate of change (L″(K)) of L(K) with respect to K is quantified for all values of K greater than 1. The mean (over the number of MCMC runs per K) of the absolute values of L″(K) is divided by the standard deviation of L(K) for all values of K greater than 1 giving the ΔK statistics. The modal value of the distribution of ΔK is the true K as defined by Evanno *et al.* (2005).

STRUCTURE Analysis was carried out using 100,000 "burn-in" steps and a further 150,000 MCMC steps (sufficient to ensure the conversion of likelihood values to equilibrium). No *a priori* assignment of individuals to populations was given. Analysis was carried out on only the haploid individuals sampled from Mains of Glamis, The James Hutton Institute and New Gilston in 2010 therefore ploidy was set to 1. The number of assumed populations (K) was 1 to 10 and 10 runs of the MCMC algorithm were carried out for each value of K. The "admixture" model was used to run the analysis as opposed to the "no admixture" model. The admixture model allows individuals to have mixed ancestry, a more flexible model for dealing with natural populations (Pritchard *et al.* 2000b). All other user-definable parameters were kept at the default settings. ΔK according to the Evanno *et al.* (2005) method was quantified using the online software Structure Harvester (Earl 2011).

6.2.4.3 Pairwise F_{ST} and *isolation by distance*

The fixation index F_{ST} (Wright 1921), which is based on the Infinite Alleles Model (IAM) of microsatellite mutation, was calculated using FSTAT v 2.9.3.2 (Goudet 1995) for all possible pairs of the three Scottish populations using all diploid individuals (females + potential diploid males) sampled in 2010 (the largest diploid dataset). This analysis allows levels of genetic divergence between pairs of populations to be quantified. F_{ST} has a theoretical minimum of 0 (indicating no genetic divergence) and a theoretical maximum of 1 (indicating fixation of alternative alleles in different populations) (Hartl and Clark 1997).

Levels of divergence among populations as measured by F_{ST} were compared with the geographic distance between populations to determine whether genetic divergence reflected isolation by distance. Geographic distance in km was calculated using the grid references for the sampling localities (Chapter 3). The appropriate statistical test to examine the relationship between genetic divergence and geographic distance is a Mantel test. However, a pilot study using both FSTAT (Goudet 2005) and GenStat 14th Edition (VSN International; Hemel Hempstead, UK) revealed that the number of populations used in the current study was too low to permit a Mantel test. Therefore, to give a preliminary indication of any isolation by distance effect, the Pearson product-moment correlation coefficient was used.

6.3 Results

6.3.1 Sampling of Dolerus aeneus

In 2009, a total of 61 *Dolerus aeneus* individuals were sampled across three Scottish populations (Table 6.1(A)). In 2010, a total of 158 *D. aeneus* individuals were sampled across 6 populations with 147 of those from the three Scottish populations and the remaining 11 individuals from the English populations (Table 6.1(B)). In each sampling year each population was dominated by haploid males apart from at New Gilston in 2010 where equal numbers of haploid males and (diploid) females were caught and at Down Farm in 2010 where higher numbers of females than males were caught (Table 6.1(B)).

In 2009, collections of both males and females occurred over approximately the same time period at Mains of Glamis and The James Hutton Institute (between 26th May and 30th June) whereas at the New Gilston site the season extended from 28th April to 11th August (Figure 6.1). Although sampling began earlier in 2010, the extension of the season was observed again at the New Gilston site with samples collected from 21st April to 28th July (Figure 6.2).

At Claxby St Andrew in 2010, one haploid male was caught in mid-May and a second at the end of May. No females or potential diploid males were caught. At Harper Adams in the same year, a total of four haploid male individuals were caught on the 21st May. At Down Farm female individuals were caught from the beginning of April through to the beginning of May and one haploid male was sampled in mid-May.

Table 6.1: The number of* Dolerus aeneus *individuals sampled per population in 2009 (A) and 2010 (B). The portion of the total no. of individuals sampled per population composed of females, haploid males and potential diploid males respectively is shown.

(A)

	Mains of Glamis	**TJHI**	**New Gilston**	**Totals**
No. females	2	3	12	**17**
No. haploid males	6	12	21	**39**
No. potential diploid males	2	0	3	**5**
Total no. individuals	**10**	**15**	**36**	**61**

(B)

	Mains of Glamis	**TJHI**	**New Gilston**	**Claxby**	**Harper Adams**	**Down Farm**	**Totals**
No. females	1	6	17	0	0	4	**28**
No. haploid males	58	43	17	4	2	1	**125**
No. potential diploid males	4	1	0	0	0	0	**5**
Total no. individuals	**63**	**50**	**34**	**4**	**2**	**5**	**158**

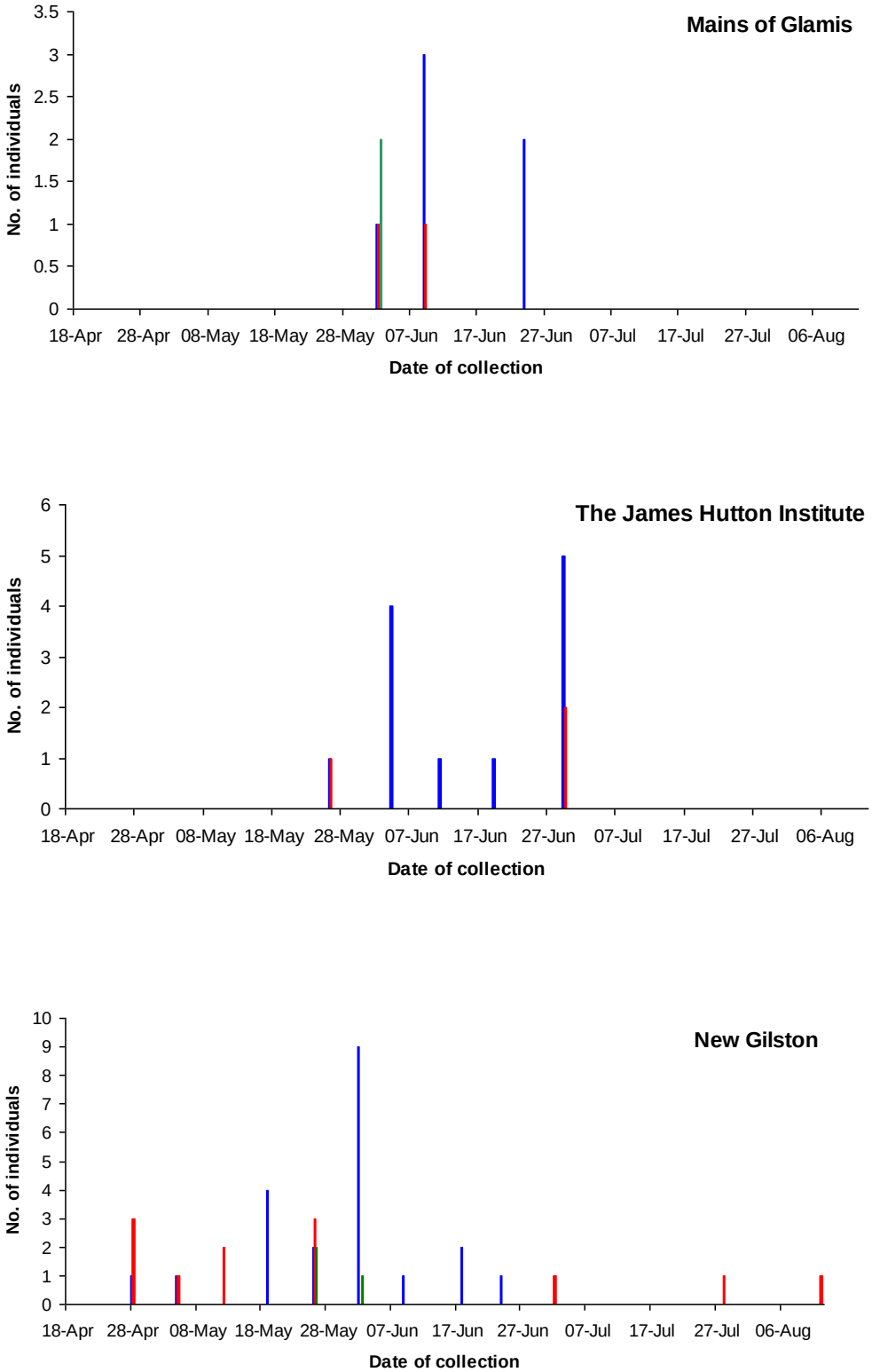


Figure 6.1: The collection dates of* D. aeneus *males (blue bars), females (red bars) and potential diploid males (green bars) sampled from Mains of Glamis, The James Hutton Institute and New Gilston in 2009.

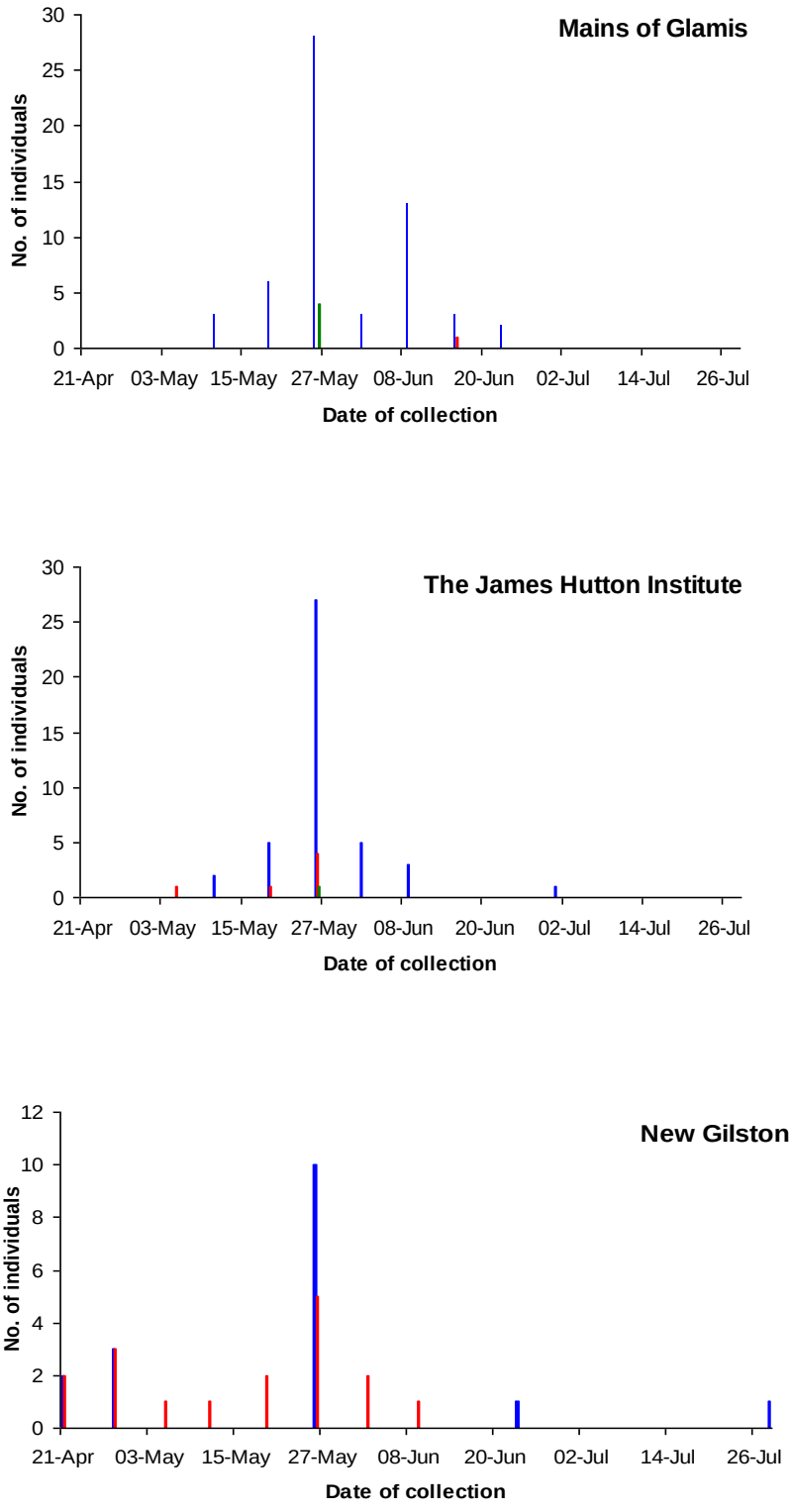


Figure 6.2: The collection dates of* D. aeneus *males (blue bars), females (red bars) and potential diploid males (green bars) sampled from Mains of Glamis, The James Hutton Institute and New Gilston in 2010.

6.3.2 Microsatellite genotyping

In total, 61 *Dolerus aeneus* individuals sampled in 2009 and 158 *D. aeneus* individuals sampled in 2010 were genotyped at 13 microsatellite loci. Of the 61 individuals sampled in 2009, 16 (19.75%) could not be genotyped at every locus after multiple attempts. For these 16 individuals data was missing for between 1 and 4 loci (mean: 1.44). Similarly, of the 158 individuals sampled in 2010, two (1.27%) could not be genotyped at all loci after multiple attempts with data missing for only one locus per individual (Appendix 1). The results from locus Saw454_2 were not possible to score unambiguously for individuals sampled in 2010. Therefore, the data from this locus was not used in any subsequent population genetic analysis.

Missing data suggests the presence of null alleles and although MicroChecker detected the presence of null alleles at six loci (Saw454_4, Saw454_7, Saw454_11, Saw454_24, Saw454_31 and Saw454_41), there were no missing values (i.e. an individual with no detectable microsatellite alleles) in the dataset that was used (all *D. aeneus* females sampled in 2010; Appendix 1). In addition, only two haploid individuals from a total of 125 sampled in 2010 could not be genotyped at all loci and data was missing for only one locus per individual. Based on this, and the fact that it is unlikely that null alleles have major impacts on estimations of genetic differentiation and the outcome of assignment testing (e.g. cluster analyses) (Carlsson 2008), all 12 tested loci were included in subsequent analyses without any modification to allele frequencies.

The number of alleles per locus ranged from one to nine for the 2009 dataset and from three to 17 for the 2010 dataset. Also, the polymorphism information content (PIC) of each locus, as a general rule, increased between 2009 and 2010 with the exception of locus Saw454_11 where the PIC value decreased from 0.755 in 2009 to 0.738 in 2010 (Table 6.2).

Notably, the Down Farm samples exhibited population-specific alleles at seven out of the 12 microsatellite loci scored for the 2010 dataset. The percentage of population specific alleles ranged from 40 to 100 (mean = 71.43%) per locus. No population-specific alleles were recorded for either the Harper Adams or Claxby St Andrew populations. Also of interest, three of the four female individuals (index no.'s: 725, 727 and 755) sampled from Down Farm had identical allelic composition across all 12 microsatellite loci. The

remaining female individual sampled from the same location (index no.: 723) had identical allelic composition to the three aforementioned individuals at all loci apart from locus Saw454_14 (Appendix 1).

It should be noted that, as discussed in Chapter 5, male individuals which are heterozygous at only one locus may or may not in fact be diploid males. Given that this can neither be proven nor disproven without further research, these individuals are considered “potential diploid males” and where they are used in population genetic analysis throughout this chapter is clearly indicated.

Table 6.2: No. of alleles per locus and allele size ranges for 13 microsatellite loci assayed on **Dolerus aeneus** ***individuals sampled in 2009 (n = 61) and 2010 (n = 158) respectively.***

Locus	No. of alleles (2009)	Allele size range in bp (2009)	PIC values	No. of alleles (2010)	Allele size range in bp (2010)	PIC values
Saw454_1	13	195-222	0.714	17	195-229	0.807
Saw454_2	5	203-228	0.501	N/A	N/A	N/A
Saw454_3	3	209-217	0.191	10	209-231	0.507
Saw454_4	14	231-256	0.825	15	231-251	0.863
Saw454_7	2	209-226	0.033	4	209-228	0.061
Saw454_11	8	234-242	0.755	9	234-242	0.738
Saw454_14	1	198	0.000	3	196-228	0.043
Saw454_16	6	188-202	0.466	9	188-205	0.533
Saw454_19	6	213-221	0.505	6	213-221	0.525
Saw454_23	4	209-215	0.506	6	197-215	0.546
Saw454_24	8	210-235	0.746	13	211-242	0.796
Saw454_31	4	222-246	0.385	6	216-246	0.489
Saw454_41	4	185-199	0.452	5	185-199	0.489

6.3.3 Genetic Diversity

6.3.3.1 Hardy-Weinberg Equilibrium and genotypic linkage disequilibrium

No evidence of linkage disequilibrium was detected within the 2010 haploid male dataset between any pair of microsatellite loci, therefore all loci could be considered independent, a prerequisite for population genetic analyses.

Departure from Hardy-Weinberg expectations (HWE) as tested by a randomisation method in FSTAT (using a dataset composed of all female and all potential diploid males sampled in 2010) was detected only at locus Saw454_11 in the New Gilston population.

6.3.3.2 Allelic richness

An Analysis of Variance revealed no significant difference between the mean allelic richness of each of the Scottish populations, in either sampling year, as tested using the haploid male datasets (Table 6.3).

Table 6.3: Comparison of allelic richness between three* Dolerus aeneus *populations in Scotland over 2 sampling years using datasets composed of all the haploid individuals sampled in the respective sampling year.

Year	Population	No. of Individuals	Mean Allelic Richness	Significance (p-value)	F	Degrees of Freedom *
	Mains of Glamis	6	2.444			
	The James Hutton Inst.	12	2.161			
2009	New Gilston	21	2.478	0.722	0.329	2 (35)
	Mains of Glamis	58	4.118			
	The James Hutton Inst.	43	4.196			
2010	New Gilston	17	4.250	0.988	0.012	2 (35)

***** First value indicates degrees of freedom between groups. Values in parentheses indicate total degrees of freedom.

6.3.3.3 Allele frequencies

Allele frequencies at the three most polymorphic loci (Saw454_1, Saw454_4 and Saw454_24) were not observed to vary dramatically between populations in sampling year 2009 or 2010 (Appendix 2; Figures A1 and A2). In 2009, the allele present at the highest frequency was the same for each of the three Scottish populations at loci Saw454_1 and Saw454_4 (alleles "199" and "237" respectively), with a number of the rare alleles shared between populations (Appendix 2; Figure A1). At locus Saw454_24 the most common allele at The James Hutton Institute and New Gilston was "233" whereas at Mains of Glamis the most common allele was "232" (Appendix 2; Figure A1). In 2010, the allele present at the highest frequency was the same for each of the three Scottish populations at loci Saw454_1 and Saw454_24 (alleles "199" and "234" respectively) with a number of the rare alleles shared between populations (Appendix 2; Figure A2). At locus Saw454_4 the most common allele at The James Hutton Institute and New Gilston was "240" whereas at Mains of Glamis alleles "235" and "236" were the two most common alleles (Appendix 2; Figure A2).

The allele frequencies of the haploid male individuals and the diploid individuals (females + potential diploid males) within each of the three Scottish populations appeared to follow a similar distribution at each of the three most polymorphic loci (Appendix 2; Figures A3-A8). To summarise the information shown in Appendix 2: Figures A3-A8, Table 6.4 shows the allele present at the highest frequency for both the haploid and diploid components of each of the Scottish populations in both sampling years. For individuals sampled in 2009, the haploid and diploid components of each of the Scottish populations had the same highest frequency allele at two out of the three most polymorphic loci. Also, at the remaining locus, the highest frequency alleles were either one base or one repeat motif apart in size (Table 6.4). For individuals sampled in 2010, the haploid and diploid components of each population had the same highest frequency allele ("199") at locus Saw454_1 (Table 6.4). In addition, the highest frequency alleles for the Mains of Glamis haploids and diploids differed by only a single base or single motif at loci Saw454_4 and Saw454_24 (Table 6.4). Similarly, at The James Hutton Institute the highest frequency allele differed by a single base at loci Saw454_24 (Table 6.4).

Table 6.4: The most frequent alleles present in both the haploid and diploid components of each of three Scottish* Dolerus aeneus *populations sampled in 2009 and 2010.

Sampling Year	Population	Locus	Most frequent allele(s) (haploid individuals)	Most frequent allele(s) (diploid individuals)
2009	Mains of Glamis	Saw454_1	199	199
		Saw454_4	237	237 and 244
		Saw454_24	232	233
	The James Hutton Institute	Saw454_1	199	199
		Saw454_4	237	235
		Saw454_24	233	233
	New Gilston	Saw454_1	199	199
		Saw454_4	237 and 240	237
		Saw454_24	233	234
2010	Mains of Glamis	Saw454_1	199	199
		Saw454_4	235 and 236	237
		Saw454_24	234	233 and 235
	The James Hutton Institute	Saw454_1	199	199
		Saw454_4	206	199
		Saw454_24	234	235
	New Gilston	Saw454_1	199	199
		Saw454_4	240	236
		Saw454_24	234	231

6.3.3.4 Observed and expected heterozygosity

Overall levels of diversity (H_E: expected heterozygosity) varied subtly between the three Scottish populations when using the haploid individuals sampled in 2010 (min H_E = 0.538, max H_E = 0.550). When using haploid individuals sampled in 2009, gene diversity was noticeably higher at New Gilston (H_E = 0.520) than at either Mains of Glamis or The James Hutton Institute, which showed relatively similar gene diversities (0.446 and 0.411 respectively) (Table 6.5).

Observed heterozygosity was calculated in addition to expected heterozygosity for the diploid individuals (females + potential diploid males). When using individuals sampled in 2010, expected heterozygosity was higher than that observed for all three populations and the difference between observed and expected heterozygosity was largest at Mains of Glamis and lowest at New Gilston (Table 6.5). When using diploid individuals (females + potential diploid males) sampled in 2009, observed and expected heterozygosity was only calculated for the New Gilston population. In this case, expected heterozygosity also exceeded observed heterozygosity (Table 6.5).

At Mains of Glamis in 2009, four diploid individuals were sampled, two of which were potential diploid males. The two potential diploid male individuals were homozygous at 12 out of 13 loci whereas the two diploid female individuals were homozygous at six and eight of 12 loci (only 12 loci due to missing data) respectively (Appendix 1). At The James Hutton Institute in 2009, three diploid individuals were sampled, all of which were female. These three individuals were homozygous at six out of 12 loci (missing data point) and nine and 11 out of 13 loci respectively (Appendix 1).

6.3.3.5 Level of Inbreeding

Positive inbreeding coefficients were noted for the New Gilston population in both sampling years regardless of whether the female dataset (F_{IS} = 0.223 (2009) and 0.300 (2010)) or the all diploids (females + potential diploid males) dataset (F_{IS} = 0.338 (2009) and 0.301 (2010)) was used (Table 6.6). In each case the randomisation test revealed that the inbreeding coefficients were significant.

Table 6.5: Levels of expected heterozygosity (gene diversity) and observed heterozygosity in three Scottish populations of* Dolerus aeneus. *Data not shown for the Mains of Glamis and The James Hutton Institute populations as calculated using the diploid individuals sampled in 2009 due to very low sample size.

		Mains of Glamis		**The James Hutton Institute**		**New Gilston**	
Dataset	**Sampling year**	$\mathbf{H_E}$	$\mathbf{H_O}$	$\mathbf{H_E}$	$\mathbf{H_O}$	$\mathbf{H_E}$	$\mathbf{H_O}$
Haploid individuals	2009	0.446	N/A	0.411	N/A	0.520	N/A
	2010	0.550	N/A	0.538	N/A	0.541	N/A
Diploid individuals	2009	-	-	-	-	0.493	0.320
	2010	0.446	0.15	0.505	0.274	0.528	0.373

Table 6.6: F_{IS} values (averaged over 12 microsatellite loci) for the New Gilston* Dolerus aeneus *population sampled in 2009 and 2010 and using both the female dataset and the diploids (females + potential diploid males) dataset.

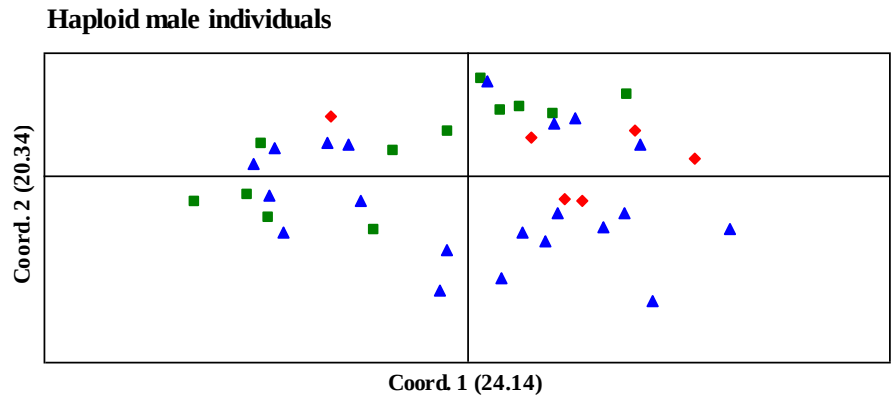
Sampling year	Dataset	New Gilston	Sample size
2009	Females	0.223	12
	Diploids	0.338	15
2010	Females	0.300	17
	Diploids	0.300	17

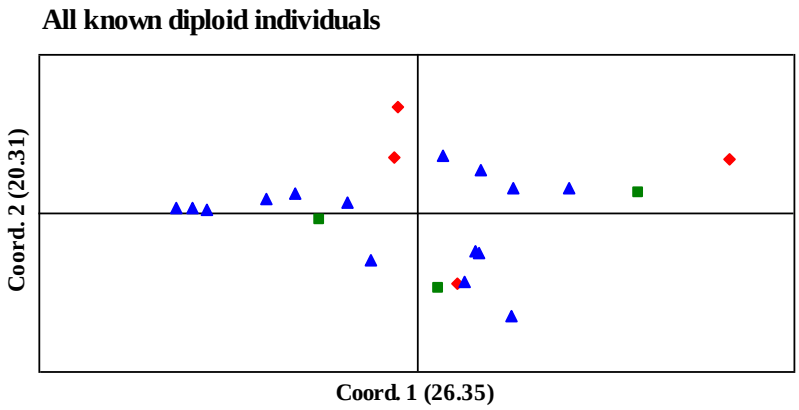
6.3.4 Population structure

6.3.4.1 Principal Coordinates Analysis

Principal Coordinates Analysis (PCA) did not reveal any distinct "clusters" of genetically related individuals regardless of the dataset used (haploid males, diploid females or all known diploids (females + potential diploid males)) or the year that samples were taken (Figures 6.3 and 6.4).

In each case three coordinates were extracted from the dataset with between 55.83 and 66.58 % of the variation present in the original dataset explained by these coordinates (mean = 62.01 %, standard deviation = 3.643).





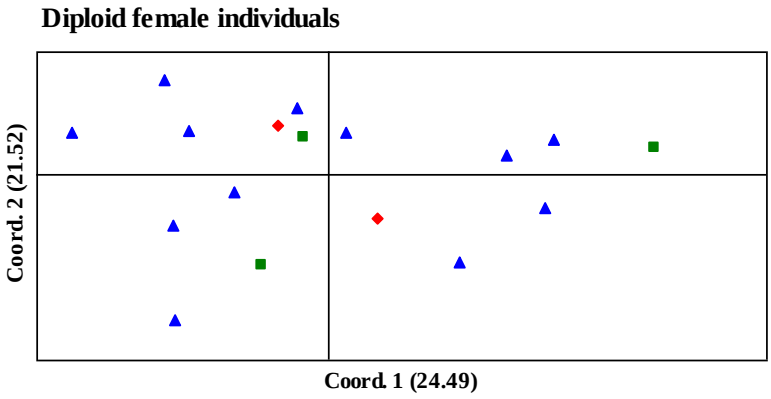


Figure 6.3: Principal Coordinate Analysis (PCA) of three populations of* Dolerus aeneus*: Mains of Glamis (red diamonds), The James Hutton Institute (green squares) and New Gilston (blue triangles). Graphical representations of PCA for datasets consisting of haploid individuals, all known diploid individuals and all diploid female individuals sampled in 2009 are shown. Numbers in parenthesis indicate the percentage of variation explained by the coordinates.

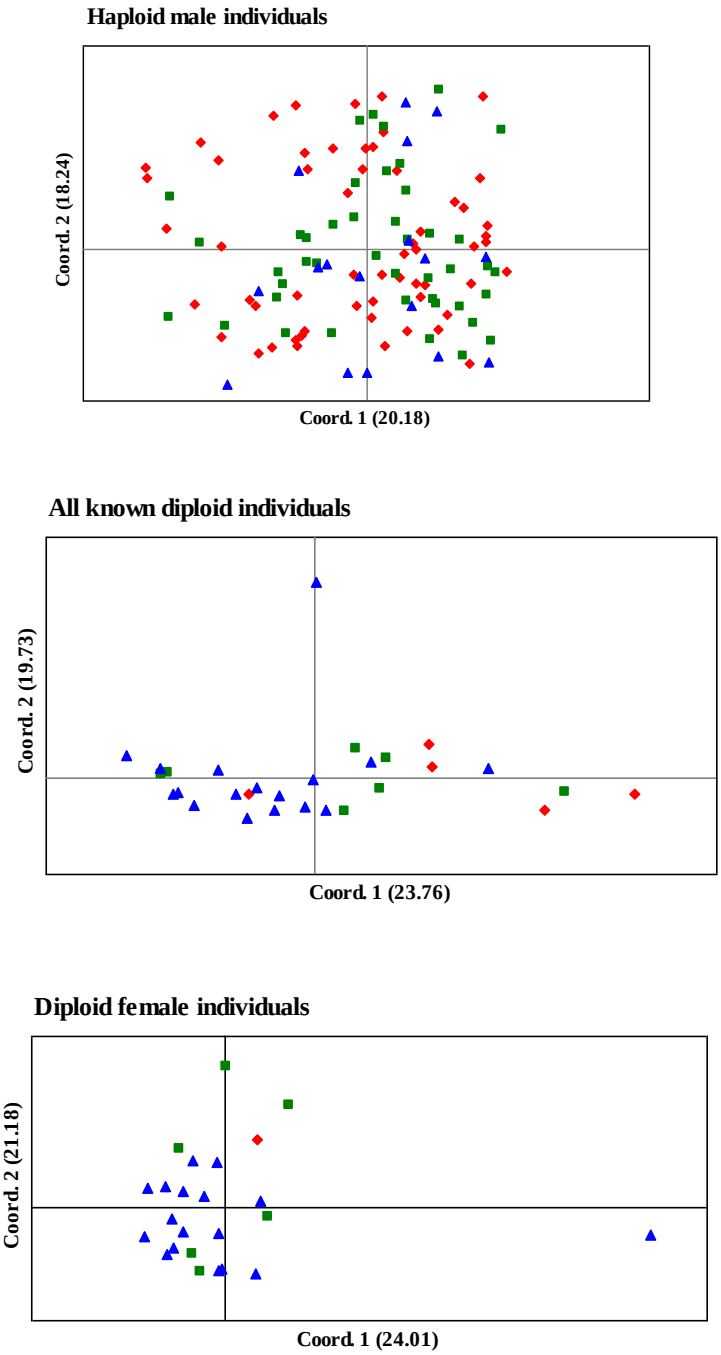


Figure 6.4: Principal Coordinate Analysis (PCA) of three populations of* Dolerus aeneus*: Mains of Glamis (red diamonds), The James Hutton Institute (green squares) and New Gilston (blue triangles). Graphical representations of PCA for datasets consisting of haploid individuals, all known diploid individuals and all diploid female individuals sampled in 2010 are shown. Numbers in parenthesis indicate the percentage of variation explained by the coordinates.

6.3.4.2 STRUCTURE analysis

The STRUCTURE analysis showed that the most likely number of populations was one for the 2009 dataset (ln P(D) = -503.72) and three for the 2010 dataset (ln P(D) = -1727.35) as these were the values for which ln P(D) was maximal (Table 6.7). However no clear plateau was visible when mean ln P(D) was plotted against K (Figure 6.5). ΔK was calculated for both datasets according to Evanno *et al.* (2005). The true K was found to be two for the 2009 dataset and three for the 2010 dataset (Figure 6.6).

Both the traditional STRUCTURE method and the Evanno method determined that three was the most likely number of populations for the 2010 dataset. However, the STRUCTURE analysis did not assign genetically similar individuals to distinct geographical locations. Genetically similar individuals were distributed across collection sites (Figure 6.7).

Table 6.7: ln P(D) values (mean over 10 MCMC runs) for each value of K. By traditional interpretation (Pritchard* et al. *2010) the true values of K for the 2009 and 2010 datasets respectively are one and three; the values for which ln P(D) is maximal.

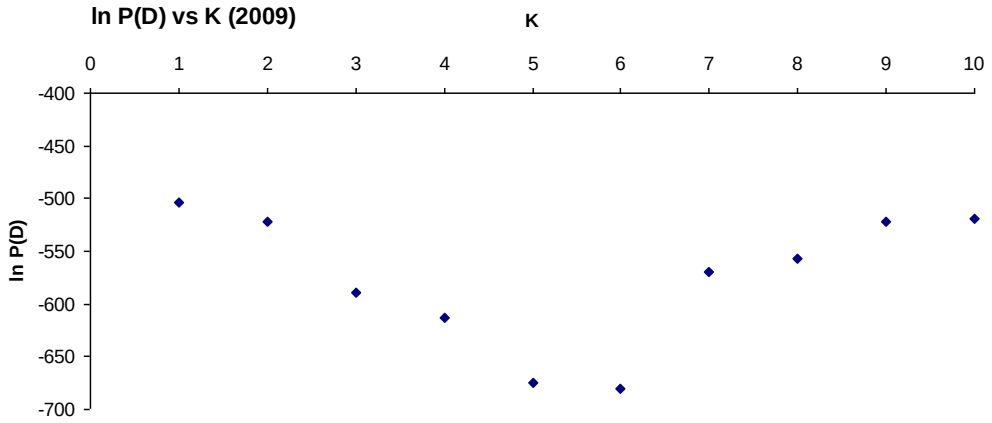
(A) 2009

K	Reps	ln P(D)
1	**10**	**-503.72**
2	10	-521.56
3	10	-588.71
4	10	-612.98
5	10	-674.37
6	10	-680.08
7	10	-569.86
8	10	-557.22
9	10	-521.67
10	10	-519.38

(B) 2010

K	Reps	ln P(D)
1	10	-1727.72
2	10	-1867.82
3	**10**	**-1727.35**
4	10	-1757.48
5	10	-1839.22
6	10	-1967.78
7	10	-2155.57
8	10	-2276.98
9	10	-2312.89
10	10	-2477.06

(A)



(B)

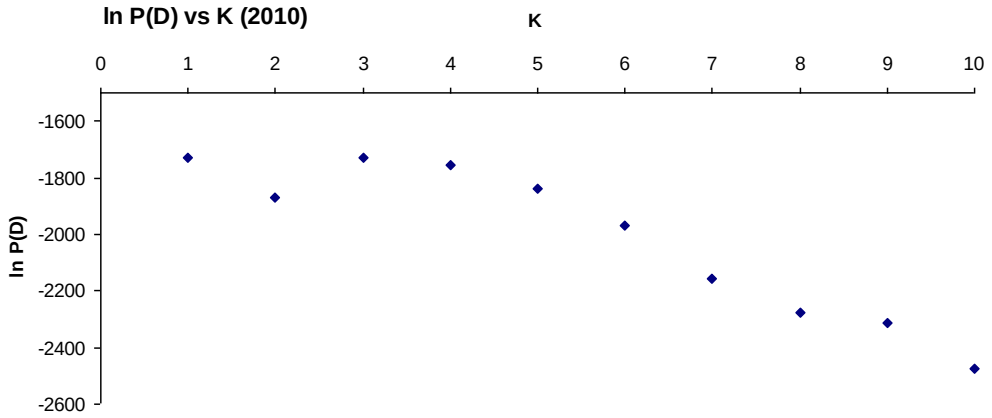
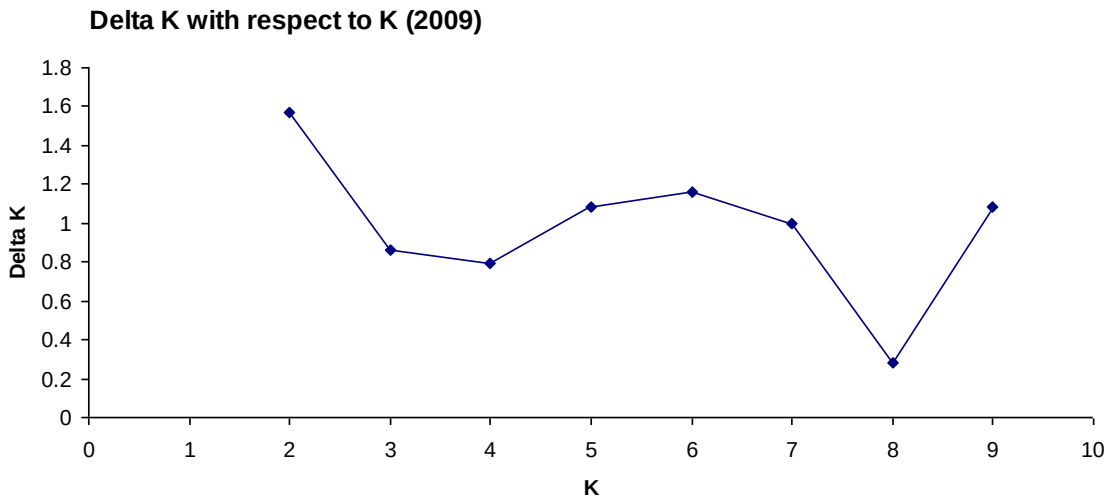


Figure 6.5: Mean ln P(D) values over 10 MCMC runs for each value of K for the 2009 (A) and 2010 (B) datasets respectively. There is no discernible plateau in the ln P(D) for either dataset.

(A)



(B)

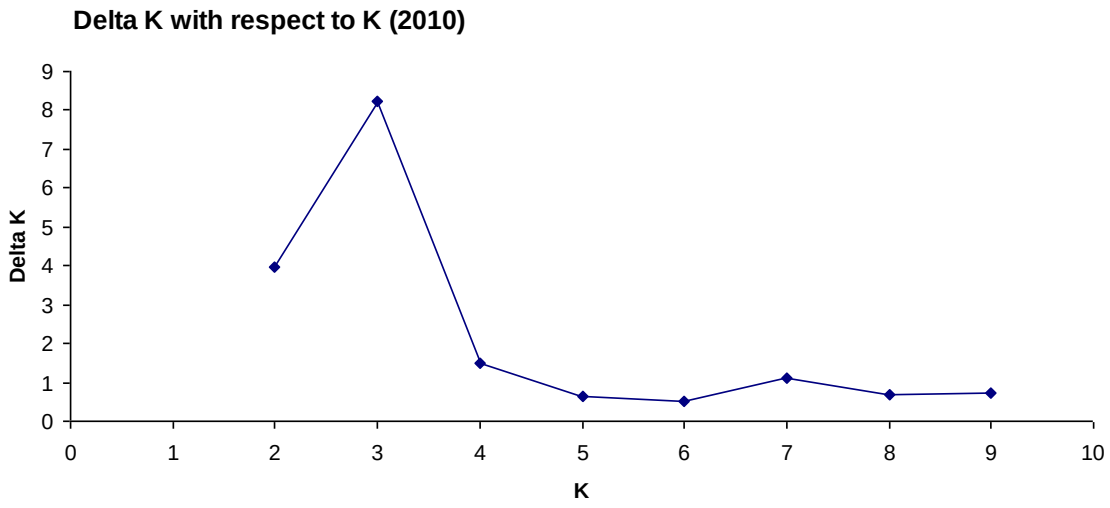


Figure 6.6: ΔK with respect to K for the 2009 and 2010 datasets respectively. ΔK is maximal for K=2 in 2009 and K=3 in 2010.

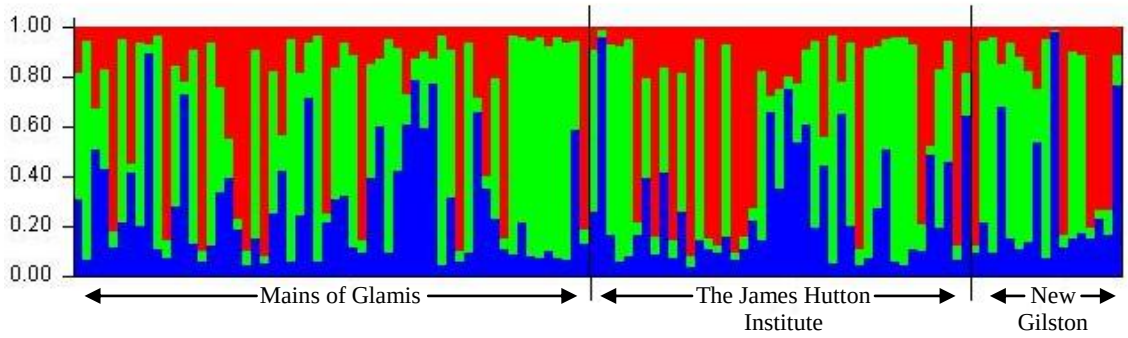


Figure 6.7: Membership coefficient profile (Q) for each individual included in the analysis. One bar represents one individual and the three colours represent each of the K populations. The proportion of the bar taken up by one colour indicates the probability of that individual belonging to that (K) population. Genetically similar individuals are distributed across collection sites (Mains of Glamis, The James Hutton Institute and New Gilston).

6.3.4.3 Pairwise F_{ST} and isolation by distance

Pairwise F_{ST} and geographic distance appeared to be positively correlated for the three Scottish populations (2010 samples; all diploid individuals used (females + potential diploid males)), indicating increased genetic divergence with distance between localities, although there was no significant relationship between the two variables (Pearson product-moment correlation coefficient r = 0.980, p = 0.129) (Table 6.8).

Table 6.8: Pairwise FST values below the diagonal and geographic distance in km above the diagonal for three Scottish populations of* Dolerus aeneus *sampled in 2010.

	Mains of Glamis	**The James Hutton Institute**	**New Gilston**
Mains of Glamis	*	18.94	41.16
The James Hutton Institute	0.0096	*	25.45
New Gilston	0.0807	0.0164	*

6.4 Discussion

6.4.1 Relatively high numbers of Dolerus aeneus *sampled in Scottish localities and samples dominated by haploid males*

In sampling year 2010, higher numbers of *Dolerus aeneus* were sampled in Scotland than in 2009 (Section 6.3.1) which is most likely due to the more intensive sampling programme used in this year (Chapter 3). In addition, the vast majority (93.04%) of *D. aeneus* individuals sampled in 2010 were captured in the Scottish localities as opposed to those in England (Section 6.3.1). It could be that *D. aeneus* populations are declining at the English localities, relative to the Scottish ones. More simple explanations are that the sampling sites in England were poorly chosen and not representative of suitable habitat for the species or that the timing of sampling at the English sites occurred too late for collection of *Dolerus aeneus* (Chapter 3). The small sample sizes at all of the English sampling locations resulted in the absence of these samples in any detailed population genetic analysis.

Populations were dominated by haploid males in both sampling years apart from at New Gilston in 2010 where haploid males and diploid females were present in equal numbers (Section 6.3.1). There are several explanations for the over-representation of haploid males at some sampling locations. It could be that there is differential mortality between males and females in the field under certain conditions, leading to higher probabilities of males surviving. In addition, female hymenopterans can facultatively adjust the sex ratio of their progeny (van Wilgenburg *et al.* 2006) and it is possible that female *D. aeneus* individuals have produced male-biased broods in response to prevailing environmental factors; females of the sawfly *Euura lasiolepis* are known to produce male-biased broods in response to poor host plant quality or slow host plant growth (Craig *et al.* 1992). A more simple explanation is that male *D. aeneus* individuals may disperse more widely than females. As a general rule female insects must invest far more resources in reproduction than males which frequently means that they are larger and heavier than the males of their species. Correspondingly, the female wing-load is higher than that of the male resulting in lower levels of dispersal (Speight *et al.* 2008).

Dolerus aeneus individuals were captured over a longer time period at New Gilston in both sampling years than at Mains of Glamis or The James Hutton Institute (Section 6.3.1). The emergence of *D. aeneus* is triggered by warm sunny days at the beginning of spring (Benson 1950) but sawflies also have very high humidity requirements (Benson 1950). It is possible that moisture levels in the soil at New Gilston were comparatively higher than at the other sampling locations and led to earlier emergence of individuals at this locality with the onset of spring. The low geographic distance between the sampling localities makes it unlikely that localised weather conditions, i.e. earlier onset of spring, are responsible for the earlier emergence times at New Gilston.

6.4.2 Three populations of **Dolerus aeneus** *exhibit similar levels of genetic diversity*

No significant difference was detected in allelic richness between the three Scottish populations of *Dolerus aeneus* for either sampling year (Section 6.3.3.2). Data obtained from a population genetic study of the sweat bee *Halictus rubicundus* (Soro *et al.* 2010), a species not known to be endangered (IUCN Red List of threatened species accessed 04/09/11), showed that three British populations of *H. rubicundus* (solitary phenotype) also exhibited similar levels of allelic richness (ANOVA; $p = 0.186$). This suggests that genetic diversity levels in three Scottish *D. aeneus* populations are relatively equal and may not be characteristic of a species under threat.

A scan of the allele frequencies in each of the three Scottish populations, using three highly polymorphic loci, did not reveal any obvious differences between populations (Section 6.3.3.3 and Appendix 2). For the most part the common alleles in each population were the same and other alleles were shared between populations, including relatively rare alleles. Examining the data in this manner gave a basic understanding of the relationship between these populations prior to applying more advanced statistical analysis and lends weight to the theory that these populations are of relatively equal genetic diversity.

The allele frequencies of the haploid male component of the populations appeared to follow a similar distribution to that of the diploid component (Section 6.3.3.3). Owen (1986) suggested after an in-depth modelling experiment that the gene frequencies of both

sexes at haplodiploid loci should be equal unless certain assumptions are violated: no dominance of alleles in females, equal dispersal in both sexes and equal effect of alleles in both sexes. The similar allele frequencies observed between the sexes in this study indicate to some extent that these populations are interbreeding normally. Also, in analyses where only the haploid male component of the population could be used, it was possible to assume that variation was not being severely under- or overestimated in any or all populations.

Levels of gene diversity (expected heterozygosity) varied only subtly between populations (haploid individuals 2010; H_E Mains of Glamis = 0.550, The James Hutton Institute = 0.538 and New Gilston = 0.541), another indication that diversity in each of the three Scottish populations is relatively equal. However, these values are far lower than that of three British populations of (the non-threatened) *H. rubicundus* (H_E = 0.775, 0.807 and 0.829) (Soro *et al.* 2010) but comparable with that of three fragmented Scottish populations of the solitary bee *Colletes floralis* (H_E = 0.51, 0.57 and 0.58) (Davis *et al.* 2010). Therefore, the diversity levels observed in the *Dolerus aeneus* populations in this study could be characteristic of that of a fragmented solitary hymenopteran population.

In addition, regardless of the dataset used (haploid, diploid female or all diploids) observed heterozygosity was lower than expected if these populations were mating completely at random (expected heterozygosity) (Section 6.3.3.4). This is a primary indication that inbreeding may be occurring to some extent in these populations. Davis *et al.* (2010) sampled diploid females from eight populations in their UK-wide study of the threatened solitary bee *Colletes floralis*. Of these eight populations, four exhibited an observed heterozygosity level lower than expected. For these four populations the mean difference between observed and expected heterozygosity was 0.05 (over all loci and populations). However, in a population genetic study of the non-threatened solitary bee *Andrena vaga* (IUCN Red List of threatened species accessed 04/09/11) four populations sampled from north-west Germany exhibited a mean decrease of 0.22 in observed heterozygosity relative to the expected (over all loci and all populations) (Exeler *et al.* 2008). The mean difference in observed and expected heterozygosity recorded in this study (2010 samples; diploid individuals; all 3 populations) was 0.227. The reduction in observed heterozygosity relative to expected heterozygosity observed in the current study is comparable to that of the non-threatened hymenopteran *Andrena vaga* suggesting that

the levels of inbreeding occurring in the three Scottish populations are not detrimental to the survival of *D. aeneus* populations. However, this result should be viewed with caution for a number of reasons. Firstly, the diploid sample sizes available for analysis in the current study were small. Secondly, three of the four female individuals (index no.'s: 725, 727 and 755) sampled from Down Farm had identical allelic composition across all 12 microsatellite loci possibly suggesting the occurrence of inbreeding at this site (Appendix 1). Finally, the reduction in observed relative to expected heterozygosity recorded in the threatened *Colletes floralis* was much smaller than recorded in this study suggesting that a comparatively high proportion of homozygotes may not be sufficient to diagnose a threatened population.

Furthermore, positive inbreeding coefficients were detected for the New Gilston population (the population with largest sample of diploid individuals) in both sampling years (Section 6.3.3.5; diploid female dataset; F_{IS} = 0.223 and 0.300 respectively). It is very unlikely in nature that any population will exhibit an inbreeding coefficient of 0 (the theoretical minimum; Hartl and Clark 1997) and it would be prudent to compare these values to that of a "healthy" sawfly population; information that is not available. However, the non-threatened (IUCN Red List of threatened species accessed 04/09/11) solitary bee *Andrena fuscipes* exhibited a mean F_{IS} of 0.195 over 12 sampled populations (min n = 8, max n = 26) and eight microsatellite loci (Exeler *et al.* 2010). The comparatively higher values observed in this study could therefore indicate that inbreeding is occurring excessively in *D. aeneus* populations. To determine more accurately the extent of inbreeding in *D. aeneus* populations it would be necessary to sample from a larger number of more widely-separated locations and attempt to obtain more samples of diploid individuals from each.

6.4.3 Three Scottish populations of **Dolerus aeneus** *do not appear to be genetically distinct*

A preliminary investigation into population substructure of *Dolerus aeneus* using principal coordinates analysis (PCA) did not highlight any "clusters" of related individuals in either sampling year regardless of the dataset used (Section 6.3.4.1). This

result suggests that the three Scottish populations of *Dolerus aeneus* could be part of one larger population.

STRUCTURE analysis detected that the most likely number of populations from the range of K values tested was three (2010 samples). However, genetically similar individuals were distributed across collection sites (Section 6.3.4.2) indicating that individuals sampled from these three localities (Mains of Glamis, The James Hutton Institute and New Gilston) are all part of one larger population. When the 2009 samples were used to perform STRUCTURE analysis, the algorithm appeared to "struggle" to settle on a likely configuration of parameters (likelihood values did not reach equilibrium with increasing number of MCMC reps). This is likely due to a combination of small sample size and the high number of missing values in the 2009 dataset. In addition, for the 2009 dataset the traditional STRUCTURE method and the Evanno *et al.* (2005) method differed in their estimation of K (Section 6.3.4.2). If the "true K" is indeed one, as the traditional STRUCTURE method detected, then by default ΔK (based on rates of change) cannot detect this (Evanno *et al.* 2005).

Pairwise F_{ST} values calculated for the three Scottish populations of *D. aeneus* ranged from 0.0096 to 0.0807 for geographic distances ranging from 18.94 km to 41.16 km (Section 6.3.4.3). F_{ST} values in this range indicate little to moderate genetic differentiation (little = 0 to 0.05, moderate = 0.05 to 0.15) between populations as originally defined by Wright in 1978 (Hartl and Clark 1997).

Pairwise F_{ST} and geographic distance appeared to be correlated although the relationship was not significant (Section 6.3.4.3). By means of comparison, the Mains of Glamis and The James Hutton Institute populations of *D. aeneus* sampled in this study were situated 18.94 km apart with an F_{ST} of 0.0096 whereas one pair of *Colletes floralis* populations sampled by Davis *et al.* 2010 were situated 18.6 km apart with an F_{ST} of 0.07. Also, The James Hutton Institute population and the New Gilston population were situated 25.45 km apart with an F_{ST} of 0.0164 whereas a second pair of populations from the Davis *et al.* (2010) study was situated 24.8 km apart with an F_{ST} of 0.062. The examples from the Davis *et al.* (2010) study listed above highlight comparable F_{ST} values on a similar geographic scale. These examples were part of a larger study of 12 populations on a UK-wide scale over which an isolation by distance effect was found to be significant. The lack of significance in this study is most likely attributable to small

sample size (no. populations sampled). If it was possible to sample more populations of *D. aeneus* on a wider geographic scale, it is likely that an isolation by distance effect would be significant. This assumption is corroborated by the presence of population-specific alleles in the individuals sampled from Down Farm (Section 6.3.2).

6.5 Conclusion

This chapter investigates the levels of genetic diversity and population substructure in Scottish populations of *Dolerus aeneus*. In the three populations of *D. aeneus* examined (Mains of Glamis, The James Hutton Institute and New Gilston), levels of genetic diversity are relatively equal and these populations do not appear to be genetically differentiated from one another, although the presence of inbreeding within each of the three populations was noted.

The small sample sizes obtained at the English sampling sites meant that it was not possible to analyse the population genetics of *D. aeneus* on a wider geographic scale. However, this preliminary investigation has shown that populations of *D. aeneus* may not be as fragmented as has been suggested previously and that a more extensive study with an optimised sampling programme would provide the means to examine the population genetics of this species on a UK-wide scale.

7. Principal findings and future perspectives

7.1 Summary of principal findings

A detailed knowledge of the underlying genetic diversity and population structure in a species is essential for its efficient conservation (Höglund 2009). As such, the main purpose of the study outlined in this thesis was to develop a set of molecular markers for use in a common farmland sawfly, an important food source for declining populations of farmland birds, in order to examine the levels and structure of genetic variation within sawfly populations. An additional goal was to determine the presence or absence of Complementary Sex Determination (CSD), a sex determination mechanism present in some hymenopterans which has deleterious consequences under inbreeding conditions (Section 1.2.3.3).

To summarise the content of this thesis, *Dolerus aeneus* was selected as a study species (Chapter 3 and Chapter 4) and a set of polymorphic microsatellite markers were developed for use in this species (Chapter 4; Cook *et al.* 2011). Using these markers (Chapter 5), potential diploid males were detected in *Dolerus aeneus*, a primary indication that CSD may be operative in this species, although further study is required to confirm this. Finally, using the microsatellite markers developed in the present study (Chapter 4), three Scottish populations *D. aeneus* were found to exhibit relatively equal genetic diversity and low levels of genetic differentiation. However, evidence of inbreeding was detected in each of the three populations (Chapter 6).

7.2 Implications of this research

This study represents the first molecular genetics research on any farmland sawfly species directed towards promoting the conservation of sawflies and of birds such as the Grey Partridge and the Yellowhammer that depend on them. It was proposed in Chapter 1 that agricultural intensification may have fragmented suitable sawfly habitat to such an extent that sawfly populations had become isolated and their genetic diversity reduced. The sawfly life cycle and the potential presence of CSD could mean that sawflies will be more

susceptible to the effects of agricultural intensification than other invertebrates occupying the same habitat (Section 1.3). Previous research has shown that sufficient underlying genetic diversity in a population or species is essential for its survival (Frankham 2005; Spielman *et al.* 2004). Therefore, it is possible that genetic factors could be, in part, responsible for the decline in sawfly numbers recorded since the 1970s (Aebischer 1991).

This study showed that levels of genetic diversity in each of the three Scottish *D. aeneus* populations examined were relatively similar but were comparable with that of a threatened solitary hymenopteran. In addition, evidence of inbreeding in each of the three populations was detected (Section 6.3.3.5). However, tests for genetic differentiation between populations revealed that these three populations do not appear to be genetically distinct (Section 6.3.4). This analysis suggests that fragmentation of suitable sawfly habitat as a result of agricultural intensification has not yet acted to isolate *D. aeneus* populations although some genetic effects (inbreeding and low diversity in comparison with non-threatened hymenopterans) are apparent. It would have been beneficial to compare populations separated by greater geographic distance to lend weight to these conclusions.

In addition, potential diploid males were detected in *D. aeneus* (Chapter 5) which, if confirmed by further research, could indicate that CSD may be operative in this species. If CSD is indeed present in *D. aeneus*, further inbreeding in populations of this species, promoted by intensive agricultural management practices, could increase the number of potentially sterile diploid males being produced. This increased production of diploid males could initially reduce population growth rate and effective population size. In the smaller populations which result, genetic drift could lead to a reduction in the number of sex alleles and therefore higher diploid male production. Research has shown that under certain conditions this cycle continues in a process termed the "diploid male vortex" which can ultimately lead to extinction (Hein *et al.* 2009; Zayed and Packer 2005).

To summarise, this research shows that, at the present time, sawflies do not appear to be have been isolated by agricultural intensification. However, the fact that evidence of some inbreeding and the potential presence of CSD were detected within *D. aeneus* populations merits further research to confirm or deny the presence of CSD and to enable the prediction of how sawfly populations are likely to respond to changing agricultural management practices.

7.3 Future perspectives

7.3.1 Sampling of Dolerus aeneus *on a UK-wide scale*

To accurately predict how *D. aeneus* populations will respond to changing agricultural management practices in years to come it would be beneficial to develop a model that could estimate corresponding changes in their genetic diversity and differentiation levels. The development of such a model would depend on a population genetic study such as that outlined in Chapter 6 of this thesis to assess the current levels of diversity and differentiation in sawfly populations on a UK-wide scale. Although population genetic analysis of *D. aeneus* populations south of the border was attempted in the current study, the low number of samples collected at these localities prevented the inclusion of these populations in detailed analysis. Sampling of *D. aeneus*, and indeed of any other sawfly to be studied, on such a large scale would need to be considered carefully to ensure that enough individuals were sampled at each locality.

Sampling of sawflies for use in this study was carried out using Malaise traps (Section 2.1.1), a form of flight-interception trap known to be effective for sampling flying insects particularly Hymenoptera and Diptera (Southwood 1978). *D. aeneus* individuals were successfully captured at all sampling locations but in lower numbers at those in England (Section 6.3.1). However, the low numbers cannot solely be explained by the difference in sampling effort between sites (Section 3.2.5). It is possible that the sampling sites selected south of the border did not contain a high enough proportion of suitable host plants. In addition, sawflies have been known to emerge as early as March in southern England (Benson 1950) and it is possible that the timing of the sampling in this study (no sites in England were sampled before mid-April) occurred too late. The location and quality of new sampling sites and the timing of sampling would need to be considered very carefully in future studies.

Relatively high numbers of the selected study species *Dolerus aeneus* were sampled by the Malaise traps at all of the Scottish sampling locations. In particular, high numbers of haploid males were caught. However, for some population genetic analyses diploid individuals are necessary (i.e. comparison between observed and expected heterozygosity, calculation of the inbreeding coefficient F_{IS} and assessment of

differentiation using F_{ST}) but diploids were sampled in relatively low numbers in the current study. It is possible that (diploid) females do not spend as much time in flight as males of the species (Speight *et al.* 2008) and to increase the number of females collected at each locality it may be necessary to alter the sampling programme.

Malaise traps were used in the current study due to their neutral status, neither attracting nor deterring particular insect species. This meant that a study species could be chosen based on the proportions of different species sampled by the traps. However, now that a study species has been selected it will be possible to alter the trapping method used in future studies. Sampling by sweep-netting the vegetation at chosen sampling localities could allow the collection of male individuals and potentially higher numbers of the more sedentary females of the species. However, sweep-netting is a labour-intensive task; to sample by this method on a UK-wide scale within the small time period that *D. aeneus* is on the wing would require the help of many people and this may not be possible. The use of attractively-coloured water traps could be a more efficient alternative. *Dolerus* sawflies have been demonstrated to show a slight preference for black (or dark coloured) water traps although no sex related differences in colour selectivity were noted (Barker *et al.* 1997). Therefore, the use of this type of trap may serve to increase the number of *D. aeneus* individuals sampled overall including diploid individuals.

It is important to bear in mind that the intention of a large-scale population genetic study on *D. aeneus* is conservation-based. The goal of any sampling programme should be to sample as non-destructively as possible.

7.3.2 Confirmation of the presence of Complementary Sex Determination in Dolerus aeneus

In the current study, potential diploid males were detected in *Dolerus aeneus* (Chapter 5) which could be a primary indication that Complementary Sex Determination (CSD) is in operation in this species (Heimpel and de Boer 2008). However, the true ploidy status of these male individuals is questionable based on the results of the current study. Notably, the preliminary results obtained from the flow cytometry analysis (no diploid males) were in contradiction with that of the microsatellite analysis (ten diploid males). This may largely be attributable to the fact that flow cytometry is difficult on samples that are

preserved in ethanol such as those in the current study which were necessarily stored in this way (Section 5.4.1).

Also, it was highlighted that all males detected as "diploid" via the microsatellite analysis were heterozygous at only locus whilst all (diploid) female individuals genotyped were heterozygous at a mean of 4.51 loci. A plausible explanation for this, as stated in Section 5.4.2, is that these males are not diploid at all but that the two alleles observed at the individual-locus combinations in question are in fact amplified from two separate gene loci belonging to the same gene family. It is not possible to say, without further investigation, whether these individuals are in fact haploid or are truly diploid males. Therefore, the only conclusion that can be drawn at the present time is that these are "potential diploid males".

Further research comprising the use of a combined flow cytometry/microsatellite analysis could confirm the presence of diploid males and thus the potential presence of CSD as is common in the literature. However, diploid males have been detected in a number of hymenopteran species that have not arisen as a result of CSD but are a consequence of mutation or hybridization (Section 5.4.4). Therefore it would be beneficial to confirm the operation of CSD in *D. aeneus* not only on the basis of diploid male production but through the use of inbreeding experiments and/or molecular techniques (van Wilgenburg *et al.* 2006).

Inbreeding experiments combined with cytological analysis to confirm ploidy have been used by numerous researchers to infer the presence of CSD including the sawfly *Athalia rosae ruficornis* (Naito and Suzuki 1991). For example, in a species with CSD, a matched mating (i.e. a female A_iA_j mates with a haploid male that carries a sex allele identical to one of her own, either A_i or A_j) results in 50 % of the diploid offspring developing as homozygous diploid males (Cook and Crozier 1995). The ploidy of these males is then confirmed by cytological analysis.

The sex determination locus (SDL) has recently been sequenced in the honeybee *Apis mellifera* (Hasselmann *et al.* 2008; Beye *et al.* 2003) and in a related study Cho *et al.* (2006) successfully sequenced the *csd* gene in *Apis mellifera* and two additional related species *A. cerana* and *A. dorsata* in order to assess intra-specific polymorphism at the *csd* locus. The success of these studies suggests that it may be possible to sequence the *csd* gene in *D. aeneus* or any other sawfly in order to confirm the presence of CSD.

If CSD is in operation in *Dolerus aeneus* this will have to be incorporated into any model used to predict how populations of the species will respond to changes to their habitat. The next logical step would be to determine the number of sex alleles present in any *D. aeneus* populations examined. In small, isolated populations, sex allele diversity will be relatively low and the probability of matched matings increases. Higher numbers of matched matings would lead to production of larger numbers of potentially sterile diploid males with deleterious effects at the population level.

The all important question remains of whether or not any diploid males produced in *D. aeneus* are indeed sterile. Until recently, the sterility of diploid males produced as a result of CSD in the Hymenoptera was considered a general phenomenon. However, a recent study by Elias *et al.* (2009) demonstrated that diploid males of *Cotesia glomerata*, a species known to exhibit CSD, are just as competitive as haploid males in obtaining matings and father fully fertile diploid daughters. In addition, Cowan and Stahlhut (2004) uncovered fertile diploid males in the solitary wasp *Euodynerus foraminatus*. These studies highlight that the sterility of diploid males cannot be assumed. The level of their fertility must be fully assessed and incorporated into any theoretical model attempting to explain changes in hymenopteran population dynamics in the face of environmental change.

7.3.3 Future genetic research in Dolerus aeneus

The microsatellite markers isolated in the current study were developed from a transcriptomic library (Chapter 4). Although the main purpose of this library was the development of molecular markers for population genetic analysis, this genetic resource is a first for any farmland sawfly and could pave the way for future genetic research in this taxon.

Of the 1,284 microsatellites discovered within the transcriptome assembly, only 72 were tested and narrowed down to a functional polymorphic set of 13 markers for use in *D. aeneus* (Chapter 4). Nine of these polymorphic markers were found to cross-amplify successfully into other *Dolerus* species suggesting that with a small amount of further research this genetic resource could be mined successfully for microsatellite markers suitable for use in other *Dolerus* species and additional closely related species.

Transcriptome libraries can also be utilised for the discovery of other types of marker such as the increasingly common SNPs (Single Nucleotide Polymorphisms; Section 1.4.2.1). In addition to population genetic analyses, SNPs can potentially be linked to candidate genes of known function (Renault *et al*. 2010). Studies using 454 pyrosequencing in other species such as the Glanville fritillary butterfly *Melitaea cinxia* (Vera *et al.* 2008) and the flesh fly *Sarcophaga crassipalpis* (Hahn et al. 2009) have successfully isolated thousands of SNPs from one sequencing run.

It may also be desirable at some point to extend beyond the population genetic studies described in this thesis and proposed in Section 7.3.1 in order to link genetic diversity of *Dolerus aeneus* to physiological performance and ecological factors. For example, sawflies are known to exhibit low resistance to insecticides (Aebischer 1990) and it may be beneficial to examine the genetic basis for this observation. A starting point could be to screen the *D. aeneus* transcriptomic library for contigs annotated (by comparison with e.g. the Gene Ontology Consortium database) as having functions related to pesticide resistance in other Hymenoptera. This type of research is complex and would require extensive planning and resources, but the transcriptome library produced during the current study could be an invaluable tool for such research.

7.4 Conclusion

The research outlined in this thesis has provided the molecular means to examine the population genetics of *D. aeneus*, and potentially that of other sawfly species on a UK-wide scale. The next-generation sequencing method used to develop microsatellite markers has yielded the first genetic resource for any farmland sawfly and could be used in future studies to develop additional markers or to isolate candidate genes. The potential presence of CSD has been detected and merits further research to confirm the operation of CSD and to assess the fertility of *D. aeneus* diploid males. Finally, a population genetic analysis of *D. aeneus* was carried out encompassing three Scottish populations. This analysis showed that the effects of agricultural intensification on *D. aeneus* are not yet extensive in this geographic area but evidence of inbreeding within these populations was detected. This analysis merits further research on a UK-wide scale to obtain a comprehensive understanding of the population genetics of this species and to permit

theoretical modeling of how the species will respond to environmental change. The studies outlined in this thesis will be beneficial to any research group interested in the genetics of the Hymenoptera and/or their conservation and that of the bird populations dependent on them.

References

Abajian C (1994) Sputnik. University of Washington, Department of Molecular Biotechnology.

Aebischer NJ (1990) Assessing pesticide effects on non-target invertebrates using long-term monitoring and time series modelling. *Functional Ecology* **4**, 369-373.

Aebischer NJ (1991) Twenty years of monitoring invertebrates and weeds in cereal fields in Sussex. 305-331 In: Firbank LG, Carter N, Darbyshire JF and Potts GR (eds) *The Ecology of Temperate Cereal Fields.* Blackwell Scientific Publications, Oxford, UK.

Aebischer NJ and Ward RS (1997) The distribution of corn buntings *Miliaria calandra* in Sussex in relation to crop type and invertebrate abundance. 124–138 In: Donald PF and Aebischer NJ (eds) *The ecology and conservation of corn buntings* Miliaria calandra. U.K. Nature Conservation, No. 13. Joint Nature Conservation Committee, Peterborough.

Agarwal M, Shrivastava N and Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* **27**, 617-631.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Armitage S, Boomsma J and Baer B (2010) Diploid male production in a leaf-cutting ant. *Ecological Entomology* **35**, 175-182.

Arthofer W, Schlick-Steiner BC, Steiner FM, Konrad H, Espadaler X and Stauffer C (2005) Isolation of polymorphic microsatellite loci for the study of habitat fragmentation in the harvester ant *Messor structor*. *Conservation Genetics* **6**, 859-861.

Bachtrog D, Agis M, Imhof M and Schlötterer C (2000) Microsatellite Variability Differs Between Dinucleotide Repeat Motifs—Evidence from *Drosophila melanogaster*. *Molecular Biology and Evolution* **17**, 1277-1285.

Barker AM, Sanbrooke KJ and Aebischer NJ (1997) The water trap colour preferences of farmland sawflies. *Entomologia Experimentalis et Applicata* **85**, 83-86.

Barker AM (1998) The identification of larvae of eight graminivorous species of the sawfly genus *Dolerus* Panzer 1801 (Hymenoptera:Tenthredinidae) regularly found in grass and cereal fields in southern England. *Journal of Natural History* **32**, 1181-1215.

Barker AM, Brown NJ and Reynolds CJM (1999) Do host-plant requirements and mortality from soil cultivation determine the distribution of graminivorous sawflies on farmland? *Journal of Applied Ecology* **36**, 271-282.

Barker A and Reynolds CJM (2004) Do Host-Plant Interactions and Susceptibility to Soil Cultivation Determine the Abundance of Graminivorous Sawflies on British Farmland? *Journal of Agricultural and Urban Entomology* **21**, 257-269.

Beebee TJC and Rowe G (2004) An introduction to Molecular Ecology. Oxford University Press, New York, USA.

Benson RB (1950) An Introduction to the Natural History of British Sawflies. *Transactions of the Society for British Entomology* **10**, 45-142.

Benson RB (1952) Hymenoptera 2. Symphyta. Section (b). *Handbooks for the Identification of British Insects* **6.2(b)**, 51-137.

Benson RB (1958) Hymenoptera 2. Symphyta. Section (c). *Handbooks for the Identification of British Insects* **6.2(c)**, 139-252.Benton TG, Vickery JA and Wilson JD (2003) Farmland biodiversity: is habitat heterogeneity the key? *Trends in Ecology and Evolution* **18**, 182-188.

Benton TG, Vickery JA and Wilson JD (2003) Farmland biodiversity: is habitat heterogeneity the key? *Trends in Ecology and Evolution* **18**, 182-188.

Beye M, Hasselmann M, Fondyrk KM, Page REJ and Omholt SW (2003) The Gene *csd* Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. *Cell* **114**, 419-429.

Beukeboom LW and Kamping A (2006) No Patrigenes Required for Femaleness in the Haplodiploid Wasp *Nasonia vitripennis*. *Genetics* **172**, 981-989.

Boatman ND, Brickle NW, Hart JD, Milsom TP, Morris AJ, Murray AWA, Murray KA and Robertson PA (2004) Evidence for the indirect effects of pesticides on farmland birds. *Ibis* **146**, 131-143.

Borg C and Toft S (1999) Value of the aphid (*Rhopalosiphum padi*) as food for grey partridge chicks. *Wildlife Biology* **5**, 55-59.

Borg C and Toft S (2000) Importance of insect prey quality for grey partridge chicks *Perdix perdix*: a self-selection experiment. *Journal of Applied Ecology* **37**, 557-563.

Borgia G (1980) Evolution of haplodiploidy: models for inbred and outbred systems. *Theoretical Population Biology* **17**, 103-128.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* **32**, 314-331.

Brickle NW, Harper DGC, Aebishcher NJ and Cockayne SH (2000) Effects of agricultural intensification on the breeding success of corn buntings *Miliaria calandra*. *Journal of Applied Ecology* **37**, 742-755.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5**, 453-455.

Brown SW (1964) Automatic frequency response in the evolution of male haploidy and other coccid chromosome systems. *Genetics* **49**, 797-817.

Bull JJ (1979) An advantage for the evolution of male haploidy and systems with similar genetic transmission. *Heredity* **43**, 361-381.

Butcher RDJ (1998) Complementary Sex Determination and mating patterns in the Campoplegini tribe (Hymenoptera: Ichneumonoidea: Ichneumonidae: Campopleginae). [Ph.D. Thesis] University of Dundee.

Butcher RDJ, Whitfield WGF and Hubbard SF (2000) Single-Locus Complementary Sex Determination in *Diadegma chrysostictos* (Gmelin) (Hymenoptera: Ichneumonidae). *The American Genetic Association* **91**, 104-111.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ and Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences* **94**, 1041-1046.

Chamberlain DE and Crick HQP (1999) Population declines and reproductive performance of Skylarks *Alauda arvensis* in different regions and habitats of the United Kingdom. *Ibis* **141**, 38-51.

Chamberlain DE, Fuller RJ, Bunce RGH, Duckworth JC and Shrubb M (2000) Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *Journal of Applied Ecology* **37**, 771-788.

Chinery M (1993) Insects of Britain and Northern Europe. Harper Collins, London.

Chistiakov DA, Hellemans B and Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture Research* **255**, 1-29.

Cho S, Huang ZY, Green DR, Smith DR and Zhang J (2006) Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research* **16**, 1366-1375.

Cho S, Huang ZY and Zhang J (2007) Sex-Specific Splicing of the Honeybee *doublesex* Gene Reveals 300 Million Years of Evolution at the Bottom of the Insect Sex-Determination Pathway. *Genetics* **177**, 1733-1741.

Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD and Lewis LC (2009) Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* **100**, 556-564.

Cook JM and Crozier RH (1995) Sex determination and population biology in the Hymenoptera. *Trends in Ecology and Evolution* **10**, 281-286.

Cook N, Aziz N, Hedley PE, Morris J, Milne L, Karley AJ, Hubbard SF and Russell JR (2011) Transcriptome sequencing of an ecologically important graminivorous sawfly: a resource for marker development. *Conservation Genetics Resources* **3**, 789-795.

Coppel HC and Benjamin DM (1965) Bionomics of the nearctic pine-feeding diprionids. *Annual Review of Entomology* **10**, 69-96.

Cornuet JM and Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Craig TP, Price PW and Itami JK (1992) Facultative sex ratio shifts by herbivorous insect in response to variation in host plant quality. *Oecologia* **92**, 153-161.

Crick HQP, Dudley C, Evans AD and Smith KW (1994) Causes of nest failure among buntings in the UK. *Ibis* **41**, 88-94.

Davis ES, Murray TE, Fitzpatrick U, Brown MJF and Paxton RJ (2010) Landscape effects on extremely fragmented populations of a rare solitary bee, *Colletes floralis*. *Molecular Ecology* **19**, 4922-4935.

de Boer JG, Ode PJ, Vet LEM, Whitfield JB and Heimpel GE (2007) Diploid males sire triploid daughters and sons in the parasitoid wasp *Cotesia vestalis*. *Heredity* **99**, 288-294.

Earl DA (2011) Structure harvester v 0.6.1. Available at http://taylor0.biology.ucla.edu/struct_harvest/

Elias J, Mazzi D and Dorn S (2009) No Need to Discriminate? Reproductive Diploid Males in a Parasitoid with Complementary Sex Determination. *PLOS One* **4**, Article no. e6024.

Ellis JR and Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* **99**, 125-132.

Estoup A, Solignac M, Harry M and Cornuet J-M (1993) Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Research* **21**, 1427-1431.

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

Exeler N, Kratochwil A and Hochkirch A (2010) Does recent habitat fragmentation affect the population genetics of a heathland specialist, *Andrena fuscipes* (Hymenoptera:Andrenidae)? *Conservation Genetics* **11**, 1679-1687.

Exeler N, Kratochwil A and Hochkirch A (2008) Strong genetic exchange among populations of a specialist bee, *Andrena vaga* (Hymenoptera: Andrenidae). *Conservation Genetics* **9**, 1233-1241.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.

Frankham R (2005) Genetics and extinction. *Biological Conservation* **126**, 131-140.

Fujiwara Y, Akita K, Okumura W, Kodaka T, Tomioka K and Naito T (2004) Estimation of Allele Numbers at the Sex-Determining Locus in a Field Population of the Turnip Sawfly (*Athalia rosae*). *Journal of Heredity* **95**, 81-84.

Fuller RJ, Gregory RD, Gibbons DW, Marchant JH, Wilson JD, Baillie SR and Carter N (1995) Population declines and range contractions among lowland farmland birds in Britain. *Conservation Biology* **9**, 1425-1441.

Gempe T, Hasselmann M, Schiott M, Hause G, Otte M and Beye M (2009) Sex Determination in Honeybees: Two Separate Mechanisms Induce and Maintain the Female Pathway. *PloS Biology* **7**, Article no. e1000222.

Godfray HCJ (1994) Parasitoids, Behavioural and Evolutionary Ecology. Princeton University Press, West Sussex, UK.

Goudet J (2005) Fstat (ver. 2.9.4), a program to estimate and test population genetics parameters. Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Gu H, Dorn S (2003) Mating system and sex allocation in the gregarious parasitoid *Cotesia glomerata*. *Animal Behaviour* **66**, 259-264.

Haddad D, Schaupp F, Brandt R, Manz G, Menzel R and Haase A (2004) NMR Imaging of the honeybee brain. *Journal of Insect Science* **4**, Article no. 7.

Hamilton WD (1967) Extraordinary Sex Ratios. *Science* **156**, 477-488.

Hamilton WD (1993) Inbreeding in Egypt and in this book: a childish perspective. In *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives*, ed. NW Thornhill, pp. 429–50. University of Chicago Press, Chicago.

Hahn DA, Ragland GJ, Shoemaker DD and Denlinger DL (2009) Gene discovery using massively parallel pyrosequencing to develop ESTs for the fleshfly *Sarcophaga crassipalpis*. *BMC Genomics* **10**, article no: 234.

Hancock JM (1995) The contribution of slippage-like processes to genome evolution. *Journal of Molecular Evolution* **41**, 1038-1047.

Hartel KD, Frederick BA and Shanower TG (2003) Isolation and characterization of microsatellite loci in wheat stem sawfly *Cephus cinctus* and cross-species amplification in related species. *Molecular Ecology Notes* **3**, 85-87.

Hartl DL and Clark AG (1997) Principles of population genetics, Third ed. Sinauer Associates, Inc., Sunderland, MA, USA.

Hasselmann M, Gempe T, Schiott M, Nunes-Silva CG, Otte M and Beye M (2008) Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* **454**, 519-523.

Hawes C, Squire GR, Hallett PD, Watson CA and Young M (2010) Arable plant communities as indicators of farming practice. *Agriculture, Ecosystems and Environment* **138**, 17-26.

Heimpel GE and de Boer JG (2008) Sex Determination in the Hymenoptera. *Annual Review of Entomology* **53**, 209-230.

Hein S, Poethke H and Dorn S (2009) What stops the 'diploid male vortex' ? – A simulation study for species with single locus complementary sex determination. *Ecological modeling* **220**, 1663-1669.

Höglund J (2009) Evolutionary Conservation Genetics. Oxford University Press, Oxford, UK.

Holland J, Hutchison MAS, Smith B and Aebischer NJ (2006) A review of invertebrates and seed-bearing plants as food for farmland birds in Europe. *Annals of Applied Biology* **148**, 49-71.

Hurlbert SH (1971) The non-concept of species diversity: a critique and alternative parameters. *Ecology* **52**, 577-586.

Johns CV and Whitehouse MEA (2004) Mass rearing of two larval parasitoids of *Helicoverpa* spp. (Lepidoptera: Noctuidae): *Netelia producta* (Brullé) and *Heteropelma scaposum* (Morley) (Hymenoptera: Ichneumonidae) for field release. *Australian Journal of Entomology* **43**, 83-87.

Kashi Y and King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics* **22**, 253-259.

Krebs CJ (1989) Ecological Methodology. Harper and Row Publishers. New York, USA.

Kuijper B and Pen I (2009) The evolution of haplodiploidy by male-killing endosymbionts: importance of population structure and endosymbiont mutualisms. *Journal of Evolutionary Biology* **23**, 40-52.

Lee JS, Hanford MG, Genova JL and Farber RA (1999) Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Human Molecular Genetics* **8**, 2567–2572.

Lee JM, Hashino Y, Hatakeyama M, Oishi K and Naito T (1998) Egg Deposition Behavior in the Haplodiploid Sawfly *Athalia rosae ruficornis* Jakovlev (Hymenoptera: Symphyta: Tenthredinidae). *Journal of Insect Behaviour* **11**, 419-428.

Liu ZJ and Cordes JF (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

Lovin DD, Washington KO, deBruyn B, Hemme RR, Mori A, Epstein SR, Harker BW, Streit TG and Severson DW (2009) Genome-based polymorphic microsatellite development and validation in the mosquito *Aedes aegypti* and application to population genetics in Haiti. *BMC Genomics* **10**, Article no. 590.

Magurran AE (2004) Measuring biological diversity. Blackwell Publishing. Oxford, UK.

Malaise R (1937) A new insect-trap. *Entomologisk Tidskrift, Stockholm.* **58**, 148-160.

Margulies M *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Metzgar D, Bytof J and Wills C (2000) Selection Against Frameshift Mutations Limits Microsatellite Expansion in Coding DNA. *Genome Research* **10**, 72-80.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.

Molbo D, Machados CA, Herre EA and Keller L (2004) Inbreeding and population structure in two pairs of cryptic fig wasp species. *Molecular Ecology* **13**, 1613-1623.

Moore SS, Sargeant LL, King TJ, Mattick, JS, Georges, M and Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**, 654-660.

Morozova O and Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-264.

Naito T and Suzuki H (1991) Sex Determination in the Sawfly, *Athalia rosae ruficornis* (Hymenoptera): Occurrence of Triploid Males. *Journal of Heredity* **82**, 101-104.

Nei M (1987) Molecular Evolutionary Genetics, Columbia University Press, New York, USA.

Newton I (2004) The recent declines of farmland bird populations in Britain: an appraisal of causal factors and conservation actions. *Ibis* **146**, 579-600.

Niyibigira EI, Overholt WA and Stouthamer R (2004) *Cotesia flavipes* (Cameron) and *Cotesia sesamiae* (Cameron) (Hymenoptera: Braconidae) do not exhibit complementary sex determination: evidence from field populations. *Applied Entomology and Zoology* **39**, 705–715.

Normark BB (2003) The evolution of alternative genetic systems in insects. *Annual Review of Entomology* **48**, 397-423.

Ode PJ, Antolin MF and Strand MR (1995) Brood-mate avoidance in the parasitic wasp *Bracon hebetor* Say. *Animal Behaviour* **49**, 1239–1248.

Ormerod SJ and Watkinson AR (2000) Editor's Introduction: Birds and Agriculture. *Journal of Applied Ecology* **37**, 699-705.

Owen RE (1986) Gene frequency clines at X-Iinked or haplodiploid loci. *Heredity* **57**, 209-219.

Park SDE (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [Ph.D. Thesis] University of Dublin.

Peakall R and Smouse PE (2010) Read me GenAlEx 6.4. Available at http://appliedpopulationgenetics2010.wikispaces.com/file/view/GENALEX6.4_documntation.pdf

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

Perry JC and Rowe L (2011) Rapid Microsatellite Development for Water Striders by Next-Generation Sequencing. *Journal of Heredity* **102**, 125-129.

Piry S, Luikart G and Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502–503.

Potts GR (1986) The Partridge. Pesticides, predation and conservation. Collins, London, UK.

Potts GR (1970) Recent changes in the Farmland Fauna with Special Reference to the Decline of the Grey Partridge. *Bird Study* **17**, 145-166.

Potts GR, Ewald JA and Aebischer NJ (2010) Long-term changes in the flora of the cereal ecosystem on the Sussex Downs, England, focusing on the years 1968–2005. *Journal of Applied Ecology* **47**, 215-226.

Poulsen JG, Sotherton NW and Aebishcher NJ (1998) Comparative nesting and feeding ecology of skylarks *Alauda arvensis* on arable farmland in southern England with special reference to set-aside. *Journal of Applied Ecology* **35**, 131-147.

Primmer CR, Saino N, Moller AP and Ellegren H (1996) Directional evolution in germline microsatellite mutations. *Nature Genetics* **13**, 391-393.

Pritchard JK, Xiaoquan W and Falush D (2010) Documentation for structure software: Version 2.3. Available at http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf

Pritchard JK, Stephens M and Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.

Quinlan J and Gauld ID (1981) Hymenoptera 2. Symphyta. New edition section (a). Handbooks for the Identification of British Insects **6.2(a)**, 1-67.

Renaut S, Nolte AW and Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology* **19**, 115-131.

Rico C, Rico I and Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. *Proceedings of the Royal Society of London B* **263**, 549-557.

Ross KG, Vargo EL, Keller L and Trager JC (1993) Effect of a Founder Event on Variation in the Genetic Sex-Determining System of the Fire Ant *Solenopsis invicta*. *Genetics* **135**, 843-854.

Rozen S and Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. 365-386. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa NJ, USA.

Sanders HL (1968) Marine benthic diversity: a comparative study. *The American Naturalist* **102**, 243-282.

Schlötterer C, Amos B and Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**, 63-65.

Schlötterer C and Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**, 211-215.

Schlötterer C, Vogl C and Tautz D (1997) Polymorphism and Locus-Specific Effects on Polymorphism at Microsatellite Loci in Natural *Drosophila melanogaster* Populations. *Genetics* **146**, 309-320.

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

Sharma PC, Grover A and Kahl G (2007) Mining microsatellites in eukaryotic genomes. *Trends in Biotechnology* **25**, 490-498.

Simberloff DS (1972) Properties of the rarefaction diversity measurement. *The American Naturalist* **106**, 414-418.

Smith NGC (2000) The evolution of haplodiploidy under inbreeding. *Heredity* **84**, 186-192.

Smith SG and Wallace DR (1971) Allelic sex determination in a lower hymenopteran, *Neodiprion nigroscutum* Midd. *Canadian Journal of Genetics and Cytology* **13**, 617–621.

Soro A, Field SJ, Bridge C, Cardinal SC and Paxton RJ (2010) Genetic differentiation across the social transition in a socially polymorphic sweat bee, *Halictus rubicundus*. *Molecular Ecology* **19**, 3351-3363.

Sotherton NW (1990) The effects of six insecticides used in UK cereal fields on sawfly larvae (Hymenoptera: Tenthredinidae). *Proceedings of the Brighton Crop Protection Conference, Pests and Diseases* **3**, 999-1004.

Sotherton NW (1998) Land use changes and the decline of farmland wildlife: an appraisal of the set-aside approach. *Biological Conservation* **83**, 259-268.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Southwood TRE (1978) Ecological methods with particular reference to the study of insect populations (2nd ed.). Chapman and Hall, Cambridge, UK.

Souza RO, Del Lama MA, Cervini M, Mortari N, Eltz T, Zimmermann Y, Bach C, Brosi BJ, Suni S, Quezada-Euan JJG and Paxton RJ (2010) Conservation genetics of neotropical pollinators revisited: microsatellite analysis suggests that diploid males are rare in orchid bees. *Evolution* **64**, 3318-3326.

Speight MR, Hunter MD and Watt AD (2008) Ecology of insects, concepts and applications. Blackwell Publishing. West Sussex, UK.

Spielman D, Brook BW and Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences* **101**, 15261-15264.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M and Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* **12**, 1339-1348.

Stahlhut JK and Cowan DP (2004) Single-locus complementary sex determination in the inbreeding wasp *Euodynerus foraminatus* Saussure (Hymenoptera: Vespidae). *Heredity* **92**, 189-196.

Stolle E, Rohde M, Vautrin D, Solignac M, Schmid-Hempel P and Moritz, RFA (2009) Novel microsatellite DNA loci for *Bombus terrestris* (Linnaeus, 1758). *Molecular Ecology Resources* **9**, 1345-1352.

Sunnocks P (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **15**, 199-203.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**, 6463-6471.

The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29.

Thorén PA, Paxton RJ and Estoup A (1995) Unusually high frequency of $(CT)_n$ and $(GT)_n$ microsatellite loci in a yellowjacket wasp, *Vespula rufa* (L.) (Hymenoptera: Vespidae). *Insect Molecular Biology* **4**, 141-148.

Toth G, Gaspari Z and Jurka J (2000) Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. *Genome Research* **10**, 967-981.

Townes H (1972) A light-weight malaise trap. *Entomological News* **83**, 239-247.

Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R, Zhang H and The FlyBase Consortium (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research* **37**, D555-D559.

Väli U, Einarsson A, Waits L and Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* **17**, 3808-3817.

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

van Wilgenburg E, Driessen G and Beukeboom LW (2006) Single locus complementary sex determination in Hymenoptera: an "unintelligent" design? *Frontiers in Zoology* **3**, **DOI:** 10.1186/1742-9994-3-1

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

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-4414.

Wall PK, Leebens-Mack J, Chanderbali AS, Barakat A, Wolcott E, Liang H, Landherr L, Tomsho, LP, Hu Y, Carlson JE, Schuster SC, Soltis DE, Soltis PS, Altman N and dePamphilis CW (2009) Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**, Article no. 347.

Walter GH, Ruohomaki K, Haukioja E and Vainio E (1994) Reproductive behavior of mated and virgin females of a solitary sawfly *Dineura virididorsata*. *Entomologia Experimentalis et Applicata* **70**, 83–90.

Whiting PW (1943) Multiple Alleles in Complementary Sex Determination of *Habrobracon*. *Genetics* **28**, 365-382.

Wilson JD, Morris AJ, Arroyo BE, Clark SC and Bradbury RB (1999) A review of the abundance and diversity of invertebrate and plant foods of granivorous birds in northern Europe in relation to agricultural change. *Agriculture, Ecosystems and Environment* **75**, 13-30.

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M and Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Molecular Ecology* **14**, 1681-1695.

Wright A (1990) British Sawflies (Hymenoptera: Symphyta) A key to adults of the genera occurring in Britain. *Field Studies Council* **7,** 531-593.

Wright S (1921) Systems of mating. *Genetics* **6**, 111-178.

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

Zayed A and Packer L (2005) Complementary sex determination substantially increases extinction proneness of haplodiploid populations. *Proceedings of the National Academy of Sciences* **102**, 10742-10746.

Zhou Y, Gu H and Dorn S (2006) Single-locus sex determination in the parasitoid wasp *Cotesia glomerata* (Hymenoptera : Braconidae). *Heredity* **96**, 487-492.

Zhou Y, Gu H and Dorn S (2007) Effects of inbreeding on fitness components of *Cotesia glomerata*, a parasitoid with single-locus complementary sex determination (sl-CSD). *Biological Control* **40**, 273–279.

Appendix 1: Raw microsatellite genotyping data

Microsatellite scoring at 13 microsatellite loci for all* Dolerus aeneus *individuals sampled in 2009. Red cells indicate missing data, blue cells indicate a heterozygote locus for a male individual and peach cells indicate a heterozygote locus for a female individual.

Individual Index no.	Sex	Pop.	Saw454_1		Saw454_2		Saw454_3		Saw454_4		Saw454_7		Saw454_11		Saw454_14		Saw454_16		Saw454_19		Saw454_23		Saw454_24		Saw454_31		Saw454_41	
41	M	Glamis	208	208	203	228	217	217	244	244	209	209	236	236	198	198	196	196	213	213	209	209	232	232	231	231	199	199
44	M	Glamis	199	199	203	203	215	215	237	237	209	209	238	238	198	198	198	198			215	215	232	232	231	231	197	197
47	M	Glamis	199	199	203	203	217	217	233	237	209	209	241	241	198	198	188	188	215	215	213	213	233	233	231	231	198	198
55	F	Glamis	199	205	203	205	215	215	237	240	209	209	235	241			198	198	213	213	211	211	233	233	231	231	198	198
100	M	Glamis	206	206	210	210	215	215	242	242	209	209	234	234	198	198	198	198	213	213	213	213	210	210	228	228	197	197
138	M	Glamis	205	205	203	203	215	215	240	240	209	209	239	239	198	198	198	198	213	213	213	213	234	234	246	246	198	198
161	M	Glamis	199	199	203	203	215	215	237	237	209	209	238	238	198	198	198	198	213	213	209	209	233	233	231	231	198	198
175	F	Glamis	199	201	203	203	215	217			209	209	234	238	198	198	196	198	213	213	209	211	229	229	228	231	198	198
215	M	Glamis	197	197	210	210	215	215	233	233	209	209	238	238	198	198	198	198	218	218	209	209	230	230	231	231	198	198
216	M	Glamis	199	199			215	215	235	235	209	209	238	238	198	198	198	198	219	219	213	213	232	232	231	231	197	197
31	M	JHI	199	199	210	210	215	215	235	235			241	241	198	198	196	196	213	213	209	209	233	233	231	231	198	198
32	F	JHI	199	199	210	212	215	215	237	237	209	209	236	236	198	198	198	198	213	213	209	211	231	233	228	231	198	198
105	M	JHI	199	199	210	210	215	215	237	237	209	209	239	239	198	198	196	196	213	213	209	209	233	233	231	231	199	199
116	M	JHI	201	201	203	203	215	215	237	237	209	209	239	239	198	198	198	198	218	218	211	211	235	235	231	231	198	198
117	M	JHI	199	199	210	210	215	215	237	237	209	209	238	238	198	198	196	196	213	213	209	209	233	233	228	228	198	198
125	M	JHI	199	199	203	203	215	215	238	238	209	209	241	241	198	198	188	188	213	213	209	209	233	233	228	228	197	197
164	M	JHI	199	199	203	203	215	215	235	235	209	209	238	238			188	188	213	213	209	209	233	233	231	231		
209	M	JHI	206	206	203	203	215	215	242	242	209	209	236	236	198	198	198	198	213	213	213	213	231	231	228	228	198	198
225	F	JHI	200	205	205	210	215	215	235	240	209	209			198	198	198	202	213	213	209	213	234	234	228	231	198	198
226	M	JHI	205	205	203	203	215	215	240	240	209	209	238	238	198	198	198	198	213	213	209	209	233	233	231	231	198	198
229	F	JHI	199	199	205	205	215	215	235	235	209	209	241	241	198	198	196	202	215	215	209	209	233	233	228	231	198	198
230	M	JHI	220	220	210	210	215	215	256	256	209	209	238	238	198	198	198	198	217	217	209	209	233	233	228	228	198	198
231	M	JHI	201	201	203	203	215	215	237	237	209	209	238	238	198	198	198	198	213	213	213	213	233	233	231	231	198	198

232	M	JHI	207	207	203	203	215	215	242	242	209	209	236	236	198	198	198	198	217	217	209	209	233	233	228	228	198	198
233	M	JHI	199	199	203	203	215	215	235	235	209	209	238	238	198	198	198	198	213	213	209	209	234	234	231	231	198	198
1	F	Gilston	199	199	203	210	215	215	237	237	209	209	236	236	198	198	198	198	217	219	211	211	230	233	228	231	198	198
2	M	Gilston	203	203	210	210	215	215	241	241	209	209	241	241	198	198	198	198	213	213	209	209	232	232	231	231	197	197
3	F	Gilston	199	201	205	210	215	215	237	237	209	209	235	238	198	198	198	198	213	219	209	209	232	234	231	246		
4	F	Gilston	199	199	203	210	215	215	235	236	209	209	238	241	198	198	196	197	213	217	211	213	229	231	228	231	198	198
6	M	Gilston	197	197	203	203	215	215	233	233	209	209	238	238	198	198	198	198	217	217	209	209	231	231	228	228	198	198
8	F	Gilston	199	205	203	203	215	215	237	240	209	209	236	242	198	198	196	198	213	213	209	213	230	234	231	246	198	198
11	F	Gilston	199	199	203	203	215	217	235	237	209	209	238	242	198	198	188	196	213	219	209	211	234	234	228	228	198	198
12	F	Gilston	203	203	203	203	215	215	238	238	209	209	239	242	198	198	198	198	213	213	211	213	234	234	228	228	198	198
14	M	Gilston	203	203	210	210	217	217	239	239	209	209	241	241	198	198	196	196	215	215	209	209	234	234	231	231	198	198
16	M	Gilston	205	205	203	203	215	215	240	240	209	209	242	242	198	198	198	198	213	213	209	209	232	232	228	228	197	197
17	M	Gilston	222	222	228	228	209	209	245	245	226	226					189	189					232	232	222	222	185	185
18	M	Gilston	199	199	210	210	215	215			209	209			198	198	198	198	221	221	209	209	234	234	228	228	197	197
21	F	Gilston	199	199	210	210	215	215	237	237	209	209	239	239	198	198	198	198	213	213	209	213	233	233	231	231	198	198
24	M	Gilston	199	199	203	228	215	215	235	235	209	209	239	239	198	198	198	198	213	213	209	209	233	233	231	231	199	199
25	M	Gilston	203	203	203	228	215	215	238	238	209	209	236	236	198	198	198	198	217	217	209	209	234	234	231	231	198	198
26	M	Gilston	199	199	203	203	215	215	237	237	209	209	241	241	198	198	196	196	213	213	209	209	230	230	231	231	198	198
27	F	Gilston	199	206	203	203	215	215	235	242	209	209	238	241	198	198	196	198	218	219	209	213	231	231	231	231	198	198
28	M	Gilston	199	199	210	210	215	215	235	235	209	209	238	238	198	198	196	196	213	213	209	209	233	233	228	228	198	198
29	F	Gilston	199	199	203	210	215	217	235	235	209	209	234	238	198	198	196	198	213	215	209	215	231	233	231	231	198	198
58	M	Gilston	205	205			215	215	240	240			238	238	198	198	198	198	215	215	213	213	234	234	231	231	198	198
59	M	Gilston	199	199	210	210	215	215	236	236			237	237	198	198	198	198	213	213	209	209	233	233	231	231	198	198
60	M	Gilston	199	199	203	203	215	215	236	236	209	209	242	242	198	198	198	198	213	213	209	209	233	233	231	231	198	198
61	M	Gilston	203	203	210	210	215	215	238	238	209	209	238	238	198	198	198	198	213	213	209	209	234	234	231	231	197	197
63	M	Gilston	210	210	203	203	217	217	246	246	209	209			198	198	198	198	221	221	213	213	230	230	231	231	197	197
64	M	Gilston	199	199	203	203	215	215	235	235	209	209	237	237	198	198	196	196	213	213	209	209	230	230	231	231	199	199
72	M	Gilston	203	203	203	205	215	215	238	238	209	209	238	238	198	198	198	198	213	213	209	209			231	231	197	197
83	M	Gilston	205	205	210	210	215	215	240	240	209	209	238	238	198	198	196	196	213	213	211	211	234	234	231	231	197	197
88	M	Gilston	199	199	203	203	215	215	237	237	209	209	236	236	198	198	196	196	213	213	211	211	233	233	231	231	198	198
96	M	Gilston	201	201	203	203	215	215	237	237	209	209	238	238	198	198	198	198	213	213	213	213	230	230	231	231	199	199
163	M	Gilston	199	199			215	215	235	235	209	209	238	238			196	196	213	213	213	213	232	232	231	231	198	198

180	M	Gilston	205	205	203	203	215	215	240	240	209	209	241	241			196	196	221	221	209	209	233	233	231	231	197	197
205	M	Gilston	201	201	203	203	215	215	237	237	209	209	236	236	198	198	198	198	213	213	213	213	233	233	231	231	198	198
222	M	Gilston	203	203	210	210	215	215	238	238	209	209	242	242	198	198	196	196	215	215	209	209	231	231	231	231	197	197
247	F	Gilston	195	195	203	203	215	215	231	231	209	209	238	241	198	198	198	198	213	213	209	211	230	234	228	231	198	198
319	F	Gilston	199	210	210	210	215	217	237	246	209	209	237	237	198	198	198	198	213	215	209	213	230	235	231	231	197	197
335	F	Gilston	195	195	203	205	215	215	231	240	209	209	236	242	198	198	196	198	213	213	209	209	230	235	231	231	197	197

Microsatellite scoring at 12 microsatellite loci for all* Dolerus aeneus *individuals sampled in 2010. Red cells indicate missing data, blue cells indicate a heterozygote locus for a male individual and peach cells indicate a heterozygote locus for a female individual.

Individual Index no.	Sex	Pop.	Saw454_1		Saw454_3		Saw454_4		Saw454_7		Saw454_11		Saw454_14		Saw454_16		Saw454_19		Saw454_23		Saw454_24		Saw454_31		Saw454_41	
43	M	Glamis	199	199	214	214	237	237	209	209	240	240	198	198	198	198	215	215	209	209	233	233	231	231	198	198
44	M	Glamis	199	199	215	215	237	237	209	209	238	238	198	198	196	196	213	213	209	209	234	234	228	228	198	198
45	M	Glamis	197	197	215	215	233	233	209	209	241	241	198	198	196	196	213	213	209	209	234	234	246	246	197	197
52	M	Glamis	200	200	215	215	235	235	209	209	238	238	198	198	202	202	213	213	211	211	236	236	231	231	198	198
53	M	Glamis	206	206	215	215	242	242	209	209	238	238	198	198	198	198	213	213	213	213	233	233	228	228	197	197
55	M	Glamis	199	199	215	215	236	236	209	209	234	234	198	198	198	198	215	215	209	209	233	233	231	231	197	197
60	M	Glamis	204	204	215	215	240	240	209	209	238	238	198	198	196	196	213	213	215	215	232	232	231	231	198	198
61	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	198	198	215	215	209	209	232	232	228	228	197	197
64	M	Glamis	220	220	215	215	246	246	209	209	242	242	198	198	196	196	219	219	213	213	232	232	231	231	198	198
79	M	Glamis	199	199	215	215	235	237	209	209	241	241	198	198	196	196	213	213	213	213	233	233	231	231	198	198
86	M	Glamis	199	199	215	215	235	237	209	209	241	241	198	198	196	196	218	218	213	213	235	235	231	231	199	199
87	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	213	213	235	235	231	231	197	197
102	M	Glamis	206	206	216	216	242	242	209	209	238	238	198	198	198	198	213	213	209	209	231	231	231	231	197	197
107	M	Glamis	199	199	215	215	237	237	209	209	240	240	198	198	202	202	213	213	211	211	234	234	228	228	198	198
120	M	Glamis	203	203	218	218	238	238	209	209	238	238	198	198	198	198	213	213	209	209	232	232	246	246	197	197
121	M	Glamis	199	199	215	215	236	236	209	209	238	238	198	198	196	196	213	213	209	209	234	234	246	246	197	197
122	M	Glamis	205	205	215	215	240	240	209	209	240	240	198	198	198	198	213	213	209	209	234	234	231	231	199	199
124	M	Glamis	201	201	216	216	237	237	209	209	234	234	198	198	198	198	213	213	209	209	231	231	231	231	198	198
125	M	Glamis	199	199	216	216	237	237	209	209	236	236	198	198	200	200	213	213	209	209	234	234	246	246	198	198
127	M	Glamis	197	197	215	215	233	233	209	209	236	236	198	198	198	198	213	213	209	209	231	231	228	228	197	197
130	M	Glamis	205	205	216	216	240	240	209	209	237	237	198	198	196	196	215	215	211	211	236	236	231	231	198	198
131	M	Glamis	205	205	214	214	240	240	209	209	238	238	198	198	196	196	213	213	209	209	231	231	231	231	198	198
132	M	Glamis	199	199	214	214	235	235	209	209	238	238	198	198	198	198	215	215	213	213	231	231	231	231	198	198
133	M	Glamis	205	205	214	214	240	240	209	209	238	238	198	198	196	196	213	213	211	211	232	232	228	228	198	198
136	M	Glamis	201	201	215	215	237	237	209	209	240	240	198	198	198	198	213	213	213	213	232	232	231	231	198	198

137	M	Glamis	203	203	214	214	238	238	209	209	238	238	198	198	198	198	213	213	209	209	234	234	231	231	197	197
138	M	Glamis	199	199	215	215	236	236	209	209	238	238	198	198	196	196	213	213	209	209	234	234	231	231	198	198
139	M	Glamis	201	201	214	214	236	236	209	209	236	236	198	198	196	196	219	219	209	209	234	234	228	228	198	198
140	M	Glamis	199	203	215	215	238	238	209	209	235	235	198	198	198	198	213	213	209	209	235	235	231	231	198	198
141	M	Glamis	199	199	218	218	235	235	209	209	236	236	198	198	205	205	213	213	209	209	233	233	228	228	197	197
142	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	196	196	215	215	209	209	234	234	228	228	198	198
143	M	Glamis	206	206	214	214	242	242	209	209	238	238	198	198	196	196	213	213	209	209	225	225	231	231	199	199
145	M	Glamis	201	201	214	214	236	236	209	209	238	238	198	198	198	198	219	219	209	209	235	235	231	231	197	197
146	M	Glamis	199	199	213	213	235	235	209	209	236	236	198	198	196	196	213	213	213	213	234	234	231	231	198	198
147	M	Glamis	199	199	215	215	237	237	209	209	236	236	198	198	196	196	213	213	209	209	231	231	231	231	199	199
148	M	Glamis	206	206	214	214	242	242	209	209	240	240	198	198	198	198	213	213	209	209	231	231	246	246	198	198
150	M	Glamis	201	201	215	215	236	236	209	209	238	238	198	198	198	198	213	213	215	215	235	235	246	246	197	197
151	M	Glamis	210	210	215	215	246	246	209	209	238	238	198	198	196	196	217	217	209	209	234	234	228	228	197	197
152	M	Glamis	201	201	215	215	236	236	209	209	238	238	198	198	196	196	215	215	209	209	231	231	228	228	197	197
153	M	Glamis	199	203	214	214	239	239	209	209	241	241	198	198	198	198	215	215	209	209	233	233	231	231	197	197
154	M	Glamis	201	201	214	214	236	236	209	209	236	236	198	198	198	198	215	215	213	213	231	231	216	216	198	198
257	M	Glamis	203	203	215	215	238	238	209	209	236	236	198	198	198	198	217	217	209	209	233	233	231	231	197	197
261	M	Glamis	208	208	217	217	244	244	209	209	236	236	198	198	198	198	213	213	209	209	235	235	231	231	198	198
265	M	Glamis	210	210	215	215	246	246	209	209	241	241	198	198	196	196	217	217	209	209	234	234	228	228	197	197
345	M	Glamis	208	208	215	215	244	244	209	209	236	236	198	198	198	198	215	215	213	213	233	233	228	228	198	198
350	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	213	213	231	231	231	231	198	198
351	M	Glamis	199	199	215	215	235	235	209	209	236	236	198	198	198	198	218	218	213	213	235	235	228	228	199	199
355	M	Glamis	205	205	216	216	240	240	209	209	236	236	198	198	198	198	213	213	213	213	233	233	228	228	198	198
357	M	Glamis	201	201	216	216	236	236	209	209	241	241	198	198	198	198	213	213	209	209	235	235	231	231	198	198
360	M	Glamis	197	197	216	216	233	233	209	209	238	238	198	198	198	198	215	215	215	215	235	235	228	228	198	198
367	M	Glamis	207	207	215	215	242	242	209	209	237	237	198	198	196	196	213	213	209	209	233	233	231	231	198	198
369	M	Glamis	200	200	216	216	234	234	209	209	241	241	198	198	198	198	213	213	209	209	234	234	231	231	198	198
374	M	Glamis	206	206	215	215	242	242	209	209	237	237	198	198	198	198	213	213	213	213	234	234	231	231	198	198
375	M	Glamis	201	201	215	215	236	236	209	209	235	235	198	198	196	196	213	213	213	213	234	234	228	228	198	198
377	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	213	213	230	230	231	231	198	198
379	M	Glamis	201	201	215	215	237	237	209	209	241	241	198	198	198	198	213	213	213	213	233	233	228	228	198	198
382	M	Glamis	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	209	209	235	235	231	231	198	198

467	M	Glamis	199	199	215	215	237	237	209	209	238	238	198	198	196	196	213	213	209	209	233	233	228	228	198	198
485	F	Glamis	199	206	214	215	237	242	209	209	241	241	198	198	198	198	213	213	209	213	231	234	231	231	199	199
488	M	Glamis	199	199	215	215	236	236	209	209	241	241	198	198	198	198	218	218	209	209	231	231	231	231	198	198
507	M	Glamis	199	199	215	215	235	235	209	209	236	236	198	198	196	196	213	213	209	209	234	234	231	231	198	198
550	M	Glamis	199	199	215	215	236	236	209	209	241	241	198	198	198	198	217	217	211	211	211	211	246	246	197	197
573	M	Glamis	206	206	215	215	242	242	209	209	236	236	198	198	196	196	219	219	209	209	234	234	231	231	198	198
38	F	JHI	199	199	215	230	237	237	209	209	236	236	198	198	198	202	213	213	209	209	234	234	231	231	185	185
39	M	JHI	199	199	216	216	237	237	209	209	238	238	198	198	196	196	215	215	213	213	236	236	228	228	198	198
40	M	JHI	199	199	209	209	251	251	228	228	234	234	198	198	190	190	215	215	209	209	238	238	222	222	198	198
69	M	JHI	199	199	215	215	237	237	209	209	238	238	198	198	198	198	217	217	213	213	233	233	231	231	198	198
70	M	JHI	201	201	215	215	236	236	209	209	238	238	198	198	198	198	213	213	209	209	234	234	228	228	198	198
71	M	JHI	201	201	215	215	237	237	209	209	241	241	198	198	196	196	218	218	209	209	234	234	231	231	198	198
73	M	JHI	205	205	215	215	240	240	209	209	238	238	198	198	198	198	219	219	213	213	236	236	246	246	198	198
74	M	JHI	200	200	215	215	234	234	209	209	241	241	198	198	188	188	213	213	209	209	233	233	231	231	198	198
75	F	JHI	199	199	215	216	235	237	209	209	237	237	198	198	198	198	213	213	209	213	235	235	231	231	198	198
156	M	JHI	205	205	216	216	240	240	209	209	238	238	198	198	198	198	217	217	209	209	234	234	228	228	198	198
157	M	JHI	199	199	214	214	235	235	209	209	238	238	198	198	198	198	219	219	213	213	236	236	231	231	198	198
158	F	JHI	199	199	214	214	250	250	209	209	234	236	196	198	198	198	213	213	213	213	234	234	231	231	198	198
159	M	JHI	205	205	215	215	240	240	209	209	238	238	198	198	198	198	215	215	209	209	234	234	231	231	197	197
160	M	JHI	201	201	215	215	236	236	209	209	238	238	198	198	188	188	213	213	213	213	234	234	231	231	199	199
164	M	JHI	199	199	215	215	236	236	209	209	241	241	198	198	196	196	213	213	209	213	235	235	231	231	197	197
168	M	JHI	205	205	216	216	240	240	209	209	238	238	198	198	198	198	213	213	209	209	234	234	231	231	198	198
170	M	JHI	199	199	216	216	235	235	209	209	235	235	198	198	196	196	213	213	213	213	234	234	231	231	197	197
171	M	JHI	205	205	215	215	240	240	209	209	236	236	198	198	200	200	213	213	209	209	233	233	231	231	198	198
173	M	JHI	206	206	214	214	242	242	209	209	236	236	198	198	188	188	213	213	211	211	234	234	231	231	197	197
174	F	JHI	201	210	215	217	236	246	209	209	238	238	198	198	198	198	213	217	209	209	231	235	231	246	198	199
175	F	JHI	199	201	215	215	236	236	209	209	241	241	198	198	198	198	213	213	209	209	231	234	228	228	197	199
176	M	JHI	199	199	215	215	234	234	209	209	235	235	198	198	198	198	213	213	209	209	231	231	231	231	197	197
177	M	JHI	206	206	214	214	242	242	209	209	238	238	198	198	198	198	213	213	211	211	234	234	228	228	197	197
178	M	JHI	205	205	215	215	240	240	209	209	239	239	198	198	198	198	213	213	211	211	234	234	246	246	198	198
179	M	JHI	205	205	215	215	240	240	209	209	242	242	198	198	188	188	213	213	213	213	234	234	228	228	198	198
180	M	JHI	200	200	215	215	235	235	209	209	236	236	198	198	198	198	213	213	209	209	234	234	231	231	198	198

181	M	JHI	207	207	215	215	242	242	209	209	238	238	198	198	202	202	217	217	213	213	232	232	228	228	198	198
182	M	JHI	200	200	215	215	234	234	209	209	239	239	198	198	198	198	213	213	211	211	234	234	231	231	198	198
185	M	JHI	197	197	217	217	233	233	209	209	242	242	198	198	198	198	215	215	213	213	234	234	231	231	198	198
186	M	JHI	203	203	215	215	238	238	209	209	241	241	198	198	198	198	213	213	209	209	235	235	231	231	198	198
187	M	JHI	210	210	215	215	246	246	209	209	241	241	198	198	196	196	213	213	213	213	235	235	231	231	197	197
188	M	JHI	199	199	215	215	236	236	209	209	239	239	198	198	198	198	217	217	209	209	231	231	228	228	197	197
189	M	JHI	197	197	215	215	233	233	209	209	240	240	198	198	196	196	213	213	209	209	234	234	231	231	198	198
190	M	JHI	199	199	215	215	236	236	209	209	241	241	198	198	196	196	213	213	213	213	234	234	231	231	198	198
191	M	JHI	210	210	215	215	246	246	209	209	238	238	198	198	188	188	213	213	209	209	232	232	228	228	198	198
192	F	JHI	201	206	215	215	236	242	209	209	236	238	198	198	198	198	215	219	209	209	235	235	228	246	198	198
193	M	JHI	199	199	215	215	237	237	209	209	238	238	198	198	202	202	215	215	209	209	234	234	228	228	197	197
195	M	JHI	205	205	215	215	240	240	209	209	236	236	198	198	198	198	213	213	209	209	231	231	231	231	198	198
197	M	JHI	201	201	215	215	236	236	209	209	236	236	198	198	198	198	213	213	209	209	234	234	228	228	198	198
198	M	JHI	199	199	215	215	237	237	209	209	236	236	198	198	198	198	217	217	213	213	236	236	228	228	198	198
200	M	JHI	199	199	217	217	235	235	209	209	239	239	198	198	196	196	213	213	213	213	230	230	231	231	198	198
240	M	JHI	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	209	209	234	234	228	228	197	197
241	M	JHI	199	199	215	215	235	235	209	209	241	241	198	198	198	198	213	213	209	209	231	231	231	231	197	197
243	M	JHI	201	201	215	215	236	236	209	209	238	238	198	198	198	198	213	213	213	213	235	235	231	231	198	198
246	M	JHI	205	205	215	215	240	240	209	209	241	241	198	198	196	196	213	213	209	209	233	233	231	231	197	197
251	M	JHI	197	197	215	215	233	233	209	209	236	236	198	198	198	198	219	219	211	211	232	232	231	231	198	198
320	M	JHI	200	200	215	215	235	235	209	209	238	238	198	198	198	198	213	213	209	209	233	233	228	228	197	197
322	M	JHI	199	199	216	216	235	235	211	211	238	238	198	198	196	196	213	213	213	213	235	235	231	231	197	197
327	M	JHI	205	205	215	215	240	240	209	209	238	238	198	198	198	198	213	213	211	211	235	235	231	231	198	198
635	M	JHI	203	203	215	215	238	238	209	209	238	238	198	198	198	198	215	215	213	213	235	235	231	231	198	198
9	M	Gilston	205	205	215	215	240	240	209	209	236	236	198	198	198	198	213	213	215	215	232	232	231	231	199	199
10	F	Gilston	197	201	215	215	233	236	209	209	238	238	198	198	196	198	213	213	209	213	235	235	231	231	198	198
11	F	Gilston	197	199	215	215	233	236	209	209	238	238	198	198	196	198	213	213	211	213	232	234	231	231	198	198
12	M	Gilston	199	199	217	217	237	237	209	209	241	241	198	198	198	198	213	213	209	209	235	235	231	231	198	198
21	M	Gilston	199	199	215	215	237	237	209	209	234	234	198	198	198	198	213	213	209	209	231	231	228	228	198	198
22	M	Gilston	203	203	215	215	239	239	209	209	241	241	198	198	198	198	213	213	213	213	233	233	228	228	198	198
23	F	Gilston	205	205	216	217	240	242	209	209	238	238	198	198	197	198	213	213	209	209	231	231	228	231	197	198
24	M	Gilston	199	199	215	215	235	235	209	209	238	238	198	198	196	196	217	217	209	209	231	231	246	246	198	198

25	F	Gilston	197	201	215	217	233	237	209	209	238	242	198	198	196	198	213	213	209	209	231	235	228	231	198	198
26	F	Gilston	199	201	215	216	236	236	209	209	241	241	198	198	196	202	213	213	209	215	232	235	231	231	197	197
34	F	Gilston	197	197	214	216	236	240	209	209	238	238	198	198	196	199	213	217	209	209	231	235	231	231	198	198
50	F	Gilston	199	203	214	215	235	239	209	209	238	238	198	198	188	196	213	213	209	211	231	235	231	246	198	198
66	F	Gilston	201	201	215	215	237	237	209	209	234	241	198	198	196	196	213	213	209	209	231	235	231	231	198	198
67	F	Gilston	203	205	215	215	238	240	209	209	238	238	198	198	198	198	213	213	211	213	230	232	231	231	198	199
201	M	Gilston	201	201	215	215	237	237	209	209	238	238	198	198	196	196	213	213	211	211	234	234	231	231	198	198
205	F	Gilston	229	229	209	209	251	251	226	226	235	235	198	198	190	190	213	219	197	197	238	242	222	222	185	185
206	F	Gilston	199	201	215	217	237	237	209	209	236	238	198	198	196	198	213	213	209	209	233	235	231	231	198	198
218	M	Gilston	200	200	215	215	235	235	209	209	236	236	198	198	198	198	213	213	209	209	234	234	231	231	198	198
219	M	Gilston	203	203	215	215	238	238	209	209	241	241	198	198	196	196	213	213	213	213	234	234	231	231	199	199
220	M	Gilston	199	199	215	215	235	235	209	209	238	238	198	198	198	198	217	217	209	209	231	231	231	231	198	198
221	F	Gilston	195	197	215	215	233	233	209	209	241	241	198	198	198	198	213	213	209	209	231	234	228	246	198	198
222	F	Gilston	205	207	215	215	240	242	209	209	239	239	198	198	196	198	213	213	209	209	231	231	230	230	198	198
226	M	Gilston	229	229	209	209	251	251	226	226	234	234	198	198	190	190	217	217	197	197	242	242	222	222	185	185
227	M	Gilston	205	205	215	215	240	240	209	209	236	236	198	198	198	198	218	218	209	209	232	232	228	228	198	198
228	F	Gilston	199	203	215	217	237	239	209	209	236	236	198	198	196	198	213	213	209	209	230	230	231	231	198	198
229	M	Gilston	199	199	215	215	237	237	209	209	238	238	198	198	196	196	215	215	211	211	234	234	231	231	198	198
232	M	Gilston	199	199	215	215	235	235	209	209	238	238	198	198	198	198	213	213	213	213	232	232	231	231	198	198
233	M	Gilston	205	205	217	217	240	240	209	209	238	238	198	198	198	198	213	213	213	213	233	233	246	246	198	198
234	M	Gilston	205	205	215	215	240	240	209	209	238	238	198	198	198	198	221	221	209	209	235	235	231	231	199	199
304	F	Gilston	199	207	215	215	235	242	209	209	238	238	198	198	196	197	215	218	209	213	231	235	228	231	197	197
308	F	Gilston	199	201	214	215	236	236	209	209	234	236	198	198	198	198	213	213	209	213	231	231	231	231	198	199
434	F	Gilston	199	203	214	214	235	238	209	209	238	238	198	198	198	198	213	213	209	209	235	235	231	231	198	198
595	M	Gilston	205	205	215	215	240	240	209	209	241	241	198	198	198	198	218	218	209	209	232	232	228	228	197	197
713	M	Gilston	195	195	215	215	231	231	209	209	238	238	198	198	198	198	215	215	209	209	234	234	231	231	197	197
316	M	Claxby	222	222	214	214			209	209	239	239	198	198	198	198	213	213	213	215	235	235	231	231	198	198
532	M	Claxby	205	205	215	215	240	240	209	209	236	236	198	198	198	198	213	213	209	209	231	231	231	231	198	198
660	M	Harper Adams	200	200	215	215	235	235	209	209	241	241	198	198	196	196	213	213	209	209	233	233	231	231	199	199
662	M	Harper Adams	199	199	215	215	236	236	209	209	241	241	198	198	198	198	217	217	209	209	233	233	231	231	198	198
663	M	Harper Adams	200	200	215	215	235	235	209	209	241	241	198	198	198	198	219	219	209	209	234	234	231	231	198	198
665	M	Harper Adams	203	203	215	215	238	238	209	209	236	236	198	198	198	198	218	218	209	209	230	230	231	231	198	198

723	F	Down Farm	213	213	224	231	244	248	209	209	238	240	198	198	190	190	221	221	210	210	233	237	228	228	193	193
725	F	Down Farm	213	213	224	231	244	248	209	209	238	240	228	228	190	190	221	221	210	210	233	237	228	228	193	193
727	F	Down Farm	213	213	224	231	244	248	209	209	238	240	228	228	190	190	221	221	210	210	233	237	228	228	193	193
755	F	Down Farm	213	213	224	231	244	248	209	209	238	240	228	228	190	190	221	221	210	210	233	237	228	228	193	193
756	M	Down Farm	223	223	209	209	250	250	226	226	234	234	198	198	190	190			197	197	240	240	222	222	185	185

Appendix 2: A comparison of the allele frequencies within and between three Scottish *Dolerus aeneus* populations

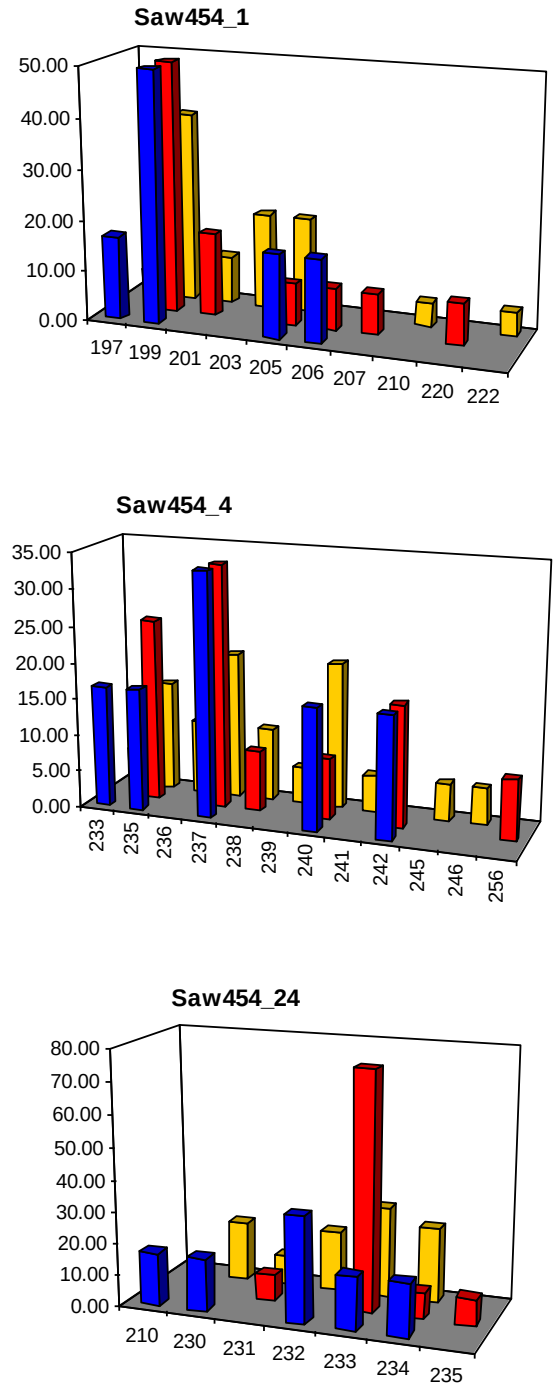


Figure A1: A comparison of the allele frequencies within 3 Scottish* Dolerus aeneus *populations [Mains of Glamis (blue bars), The James Hutton Institute (red bars) and New Gilston (yellow bars)] sampled in 2009 at three polymorphic microsatellite loci.

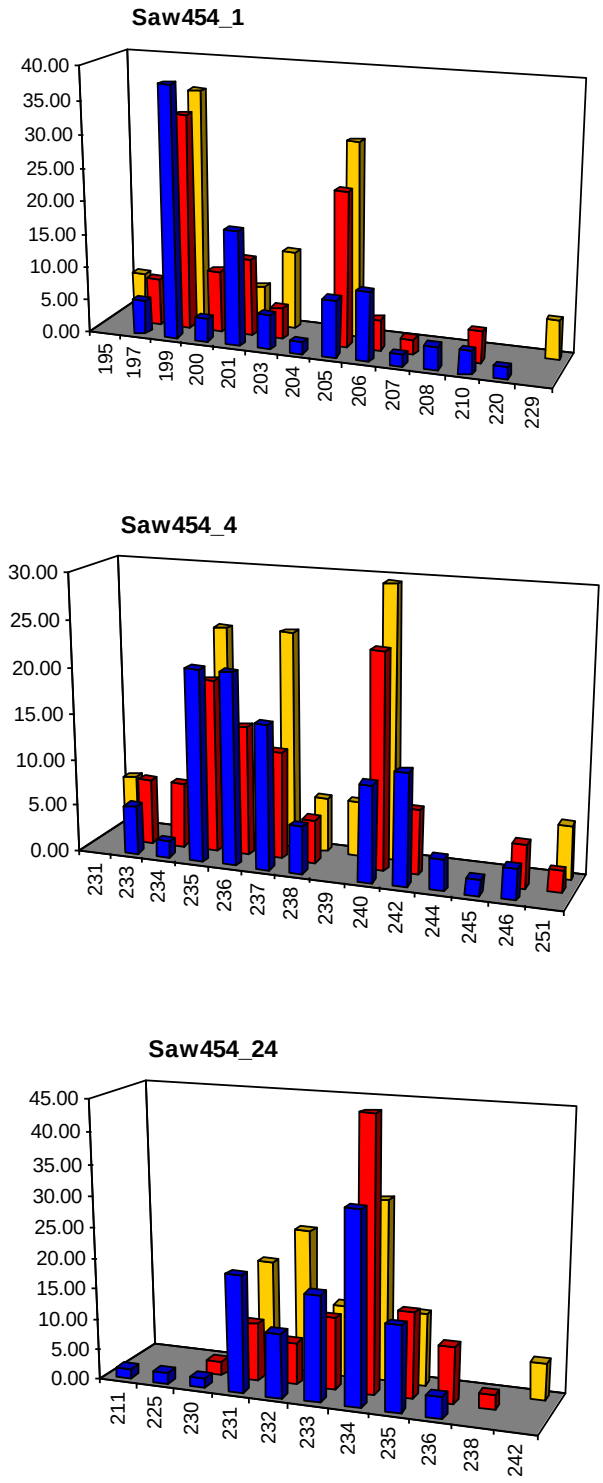


Figure A2: A comparison of the allele frequencies within 3 Scottish* Dolerus aeneus *populations [Mains of Glamis (blue bars), The James Hutton Institute (red bars) and New Gilston (yellow bars)] sampled in 2010 at three polymorphic microsatellite loci.

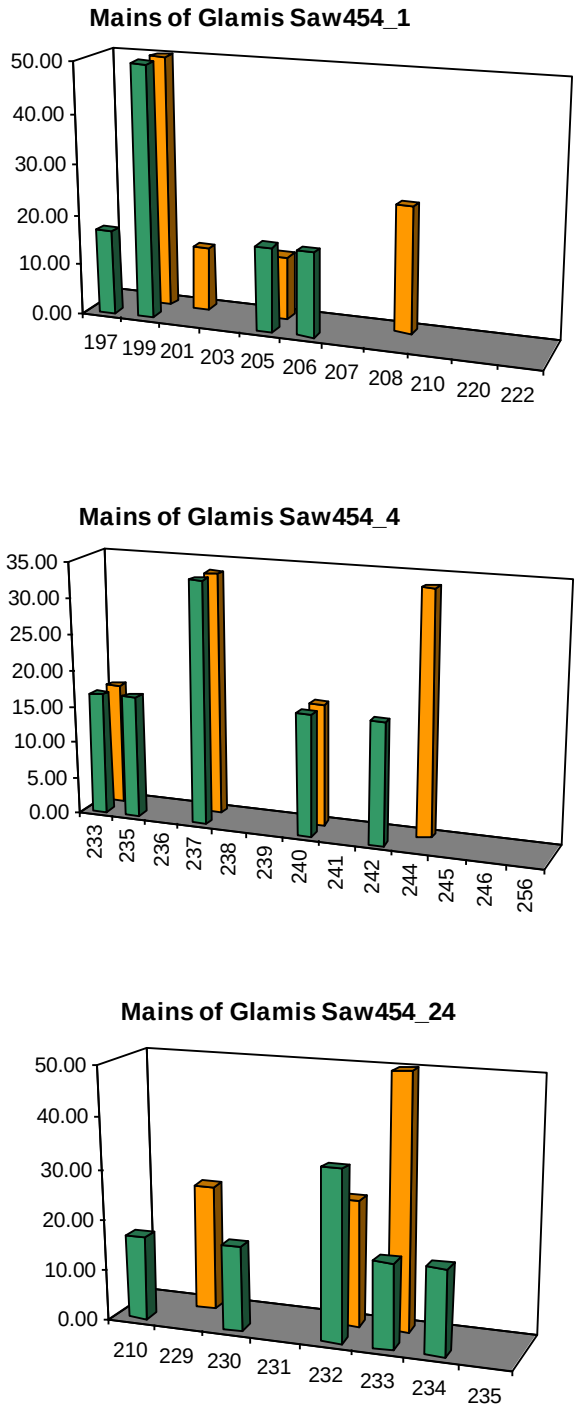
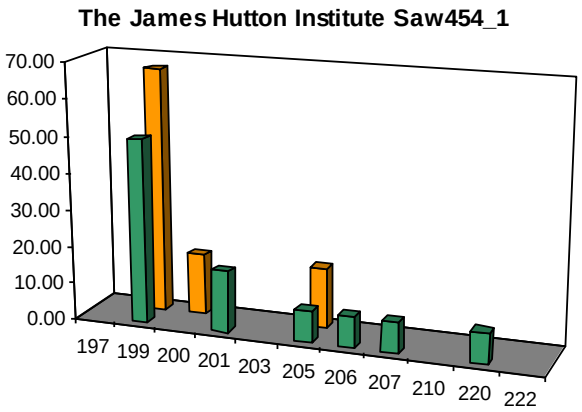
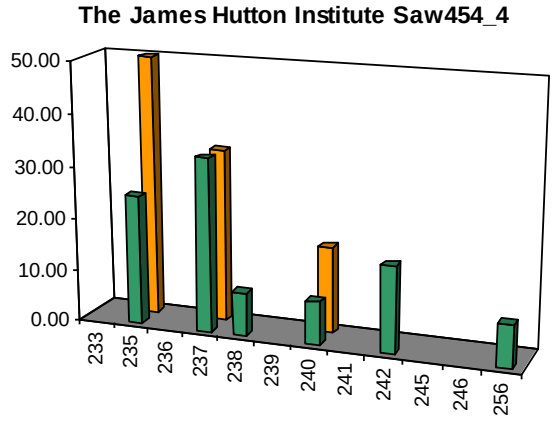


Figure A3: A comparison of the allele frequencies between the haploid male individuals (green bars) and all of the known diploid individuals (orange bars) sampled at Mains of Glamis in 2009 at three microsatellite loci.





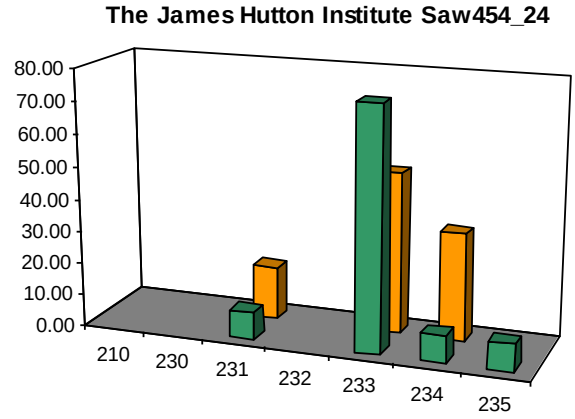
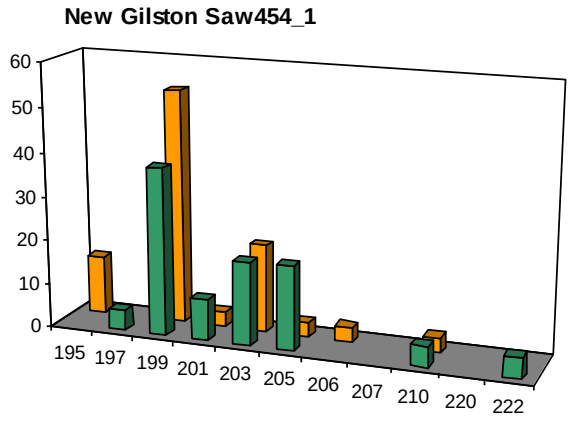
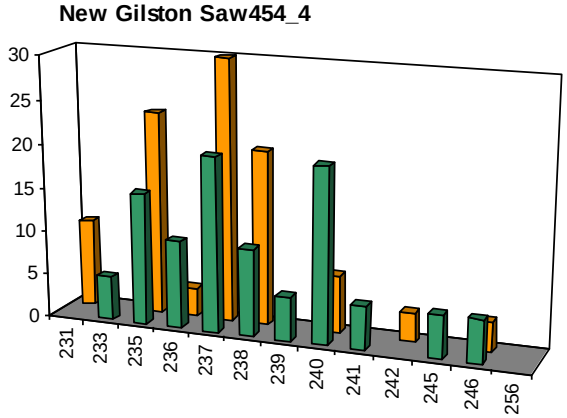


Figure A4: A comparison of the allele frequencies between the haploid male individuals (green bars) and all of the known diploid individuals (orange bars) sampled at The James Hutton Institute in 2009 at three microsatellite loci.





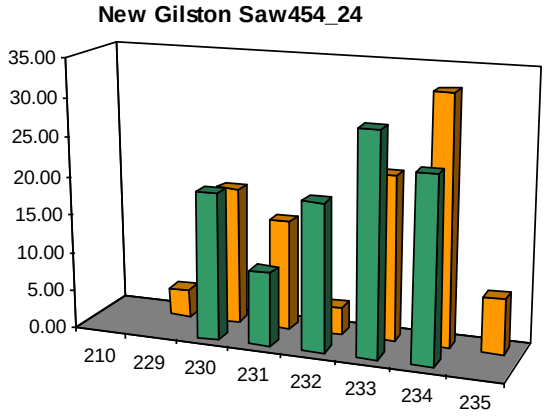


Figure A5: A comparison of the allele frequencies between the haploid male individuals (green bars) and all of the known diploid individuals (orange bars) sampled at New Gilston in 2009 at three microsatellite loci.

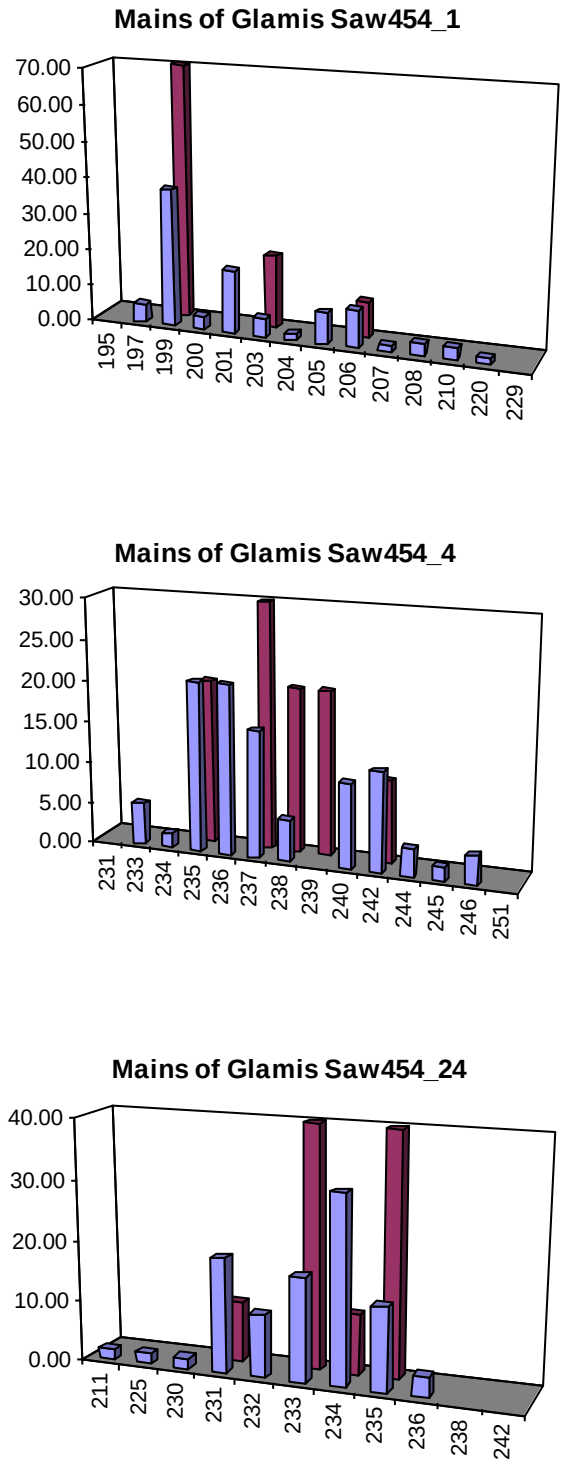


Figure A6: A comparison of the allele frequencies between the haploid male individuals (blue bars) and all of the known diploid individuals (red bars) sampled at Mains of Glamis in 2010 at three microsatellite loci.

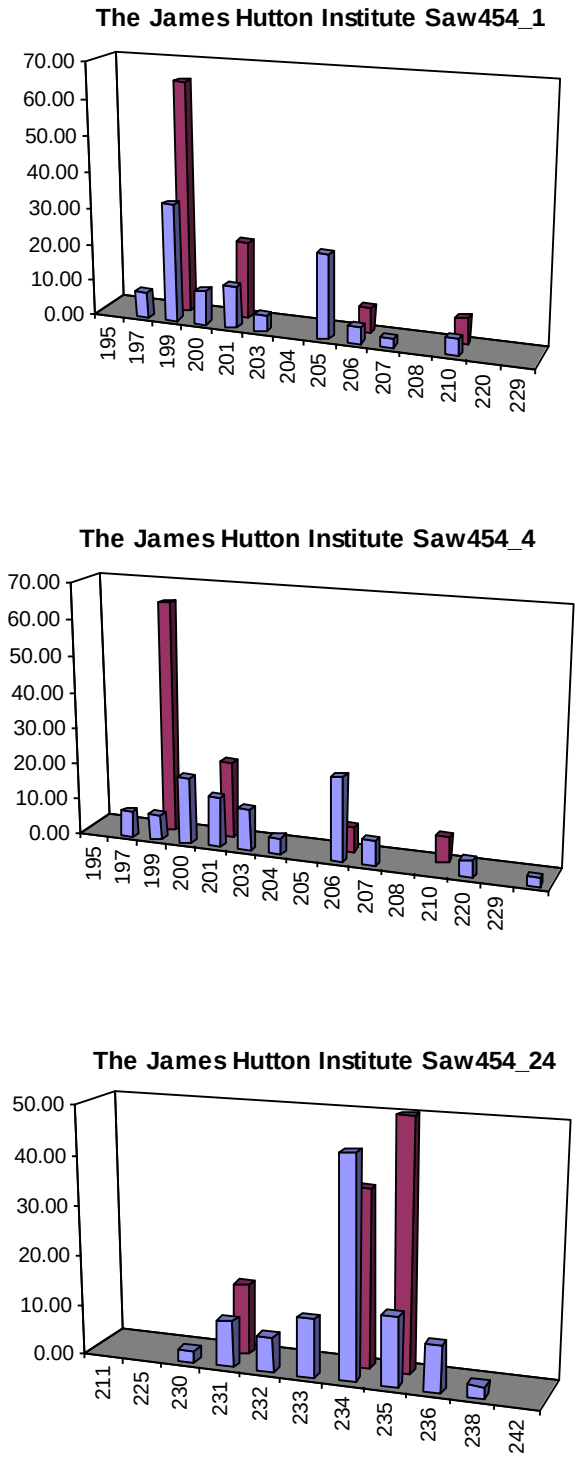


Figure A7: A comparison of the allele frequencies between the haploid male individuals (blue bars) and all of the known diploid individuals (red bars) sampled at The James Hutton Institute in 2010 at three microsatellite loci.

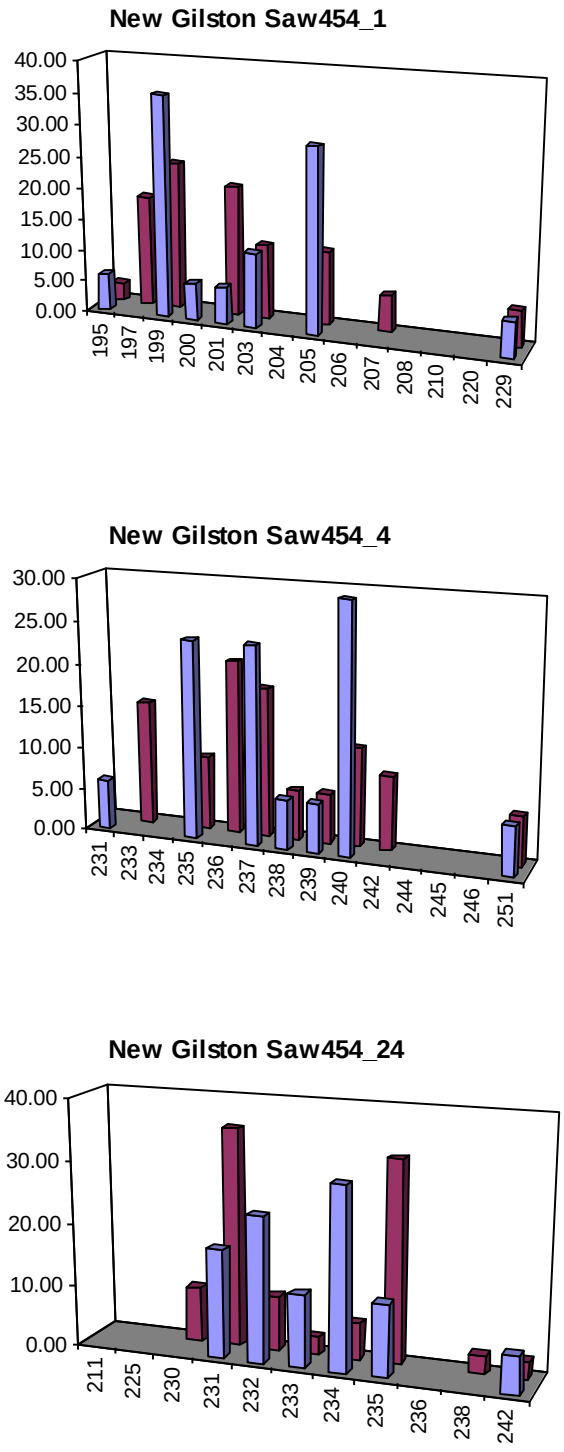


Figure A8: A comparison of the allele frequencies between the haploid male individuals (blue bars) and all of the known diploid individuals (red bars) sampled at New Gilston in 2010 at three microsatellite loci.