Genetics of insect resistance to plant defence

Kim M. C. A. Vermeer

Thesis committee

Promotor
Prof. Dr M. Dicke
Professor of Entomology
Wageningen University

Co-promotor
Dr P.W. de Jong
Assistant professor, Laboratory of Entomology
Wageningen University

Other members
Prof. Dr B.J. Zwaan, Wageningen University
Prof. Dr M.E. Schranz, Wageningen University
Dr A. Biere, Netherlands Institute of Ecology (NIOO KNAW), Wageningen
Dr K. Vrieling, Leiden University

This research was conducted under the auspices of the C.T de Wit Graduate School for Production Ecology and Resource Conservation.

Genetics of insect resistance to plant defence

Kim M. C. A. Vermeer

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 18 February 2014
at 11 a.m. in the Aula.

Kim M. C. A. Vermeer
Genetics of insect resistance to plant defence
200 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)
With references, with summaries in Dutch and English

ISBN 978-94-6173-836-3

CONTENTS

1	General introduction	7
2	The potential of a population genomics approach to analyse geographic mosaics of plant-insect coevolution	25
3	Identification of a saponin degrading β-glucosidase from the flea beetle *Phyllotreta nemorum*	43
4	Changes in frequencies of genes that enable *Phyllotreta nemorum* to utilize its host plant, *Barbarea vulgaris,* vary in magnitude and direction as much within as between seasons	61
5	Natural selection against resistance of a flea beetle to host plant defences	75
6	Evaluation of a candidate resistance gene underlying flea beetle resistance to host plant defence with a population genomics approach	107
7	General discussion	133
	References	153
	Summary	171
	Samenvatting	177
	Acknowledgements	183
	Curicullum vitae	185
	Publications	187
	Supplementary	189

CHAPTER 1

General introduction

Kim M. C. A. Vermeer

This general introduction will provide the reader with background information of the studied system needed to understand the presented work. First, I will provide a general background of plant-insect interactions involving secondary metabolites. Subsequently, I will introduce the study species and give the rationale for using it as the study organism in this research. This is followed by an introduction of the most important concepts and models used in this thesis. Lastly, I will give the aims of the thesis and present a brief overview of the following chapters.

SCIENTIFIC BACKGROUND AND RELEVANCE

We live on a Green Planet. Without plants we would not refer to the earth as being green. Plants are abundant, despite being the primary food source of many other creatures that roam the same planet and despite being nearly completely immobile. Somehow plants manage to not become completely consumed, as the green color of the earth demonstrates. Different mechanisms of plant defences account for this. Plant defence mechanisms can be mechanical like thorns or wax layers. In addition, the plant can also be chemically defended by secondary metabolites (Fraenkel 1953). These chemical compounds defend the plant against potential herbivores by e.g. repellence, toxicity, or reduction of the digestibility by the attacker. A plant usually contains a variety of these secondary compounds (Schoonhoven et al. 2005) and they have different effects on different herbivores; some herbivores are attracted by the same compound that repels others. A well-known example of compounds that attract or repel herbivores are glucosinolates and their breakdown products in Cruciferae (Brassicaceae)(Verschaffelt 1910; Fraenkel 1959; Feeny 1977; Chew 1988). Glucosinolates reduce the performance of many generalist herbivores on the plant. For specialist insects that have adapted to this plant however, these glucosinolates can be a feeding and oviposition stimulant (Verschaffelt 1910; Thorsteinson 1953; Fraenkel 1959; Whittaker and Feeny 1971; Nielsen 1978; Renwick 2002). Where glucosinolates act as a first line of defence against generalists here, the plant can subsequently develop a second line of defence that wards off specialists (e.g. in Feeny 1977; Nielsen 1978; Sachdevgupta et al. 1993a; Sachdevgupta et al. 1993b; Shriver et al. 1993). For example, not the presence of glucosinolates but the presence of cardenolides explain rejection of the plant *Erysimum cheiranthoides* by the cabbage butterfly (Sachdevgupta et al. 1993b). The evolution of defences against herbivorous insects may occurs under selection pressure from these herbivores (Rausher 1996). This selection pressure may vary in time and space, as influenced by the composition of the community (Thompson 2005). Plant defences may also differ in constitutive and induced defences, meaning that the defence is present regardless of a herbivore being present or not, and that the defence will be initiated only when the herbivore attacks the plant, respectively. Even within a plant species, genotypes may differ in defence strategies against insects (e.g. Nielsen 1997b; Jander et al. 2001). Herbivores in their turn, can also evolve mechanisms to avoid, tolerate or resist the plant's second line of defence (Schoonhoven et al. 2005; Despres et al. 2007). The dynamic interplay between plant and herbivorous insects is also known as a (co)evolutionary 'arms race' be-

Glossary

Adaptation: the process that traits evolve by natural selection that influence the organism's fitness.

Bottleneck: In population genetics a bottleneck refers to a sudden and large decrease in population size.

Gene flow: the exchange of genetic information between (sub)populations (Endler 1977). Gene flow is not the same as migration as migration does not have to lead to genetic exchange.

Genetic drift: the change in allelic composition of populations caused by random sampling. The smaller the population, the bigger the effect of genetic drift is expected to be.

Genome: all the genetic information of an individual.

Glucosinolates: organic compounds that contain sulphur and nitrogen and glucose unit(s). We know them probably best from the sharp taste of mustard seeds.

Maladaptation: a trait evolved by natural selection that used to be advantageous but now negatively influences the organism's fitness.

secondary Metabolites: organic compounds that are not directly involved in the primary metabolism of an organism.

Metapopulation: a collection of (sub)populations that undergo extinction and recolonization events that affect genetic variation and evolution among this collection of subpopulations (Hanski 1998).

Neutral theory: Most polymorphisms are hardly or not at all under selection and are therefore mainly under influence of genetic drift and dispersal (Kimura 1983).

Panmictic population: a population where random mating occurs. Each individual has an equal chance of mating with another individual from that population.

tween plant and insect and is believed to have led to the variety of chemical compounds we find in plants (Ehrlich and Raven 1964).

Implications of the evolution of plant-insect interactions and resistance development concern agricultural and environmental issues. Highly invasive cruciferous weeds can, for example, be controlled with specialist insects (Jolivet et al. 1988). Moreover, plants that are high in oviposition stimulants but on which larvae cannot survive, can be planted near crops to serve as a "dead-end" trap for pest insects (Renwick 2002; Shelton and Nault 2004). *Barbarea vulgaris*, for example, can be used in integrated pest management against the diamond back moth (DBM), a notorious pest of various cultivated crucifers (Talekar and Shelton 1993; Idris and Grafius 1996; Shelton and Nault 2004; Badenes-Perez et al. 2010). Furthermore, deterrents, repellents, and feeding inhibitors of certain plants can be incorporated into crops to increase their resistance to pest insects. For the development of such applications and for them to remain successful, we need a better understanding of insect resistance mechanisms and how insect resistance evolves. This knowledge may also add to the prevention of resistance development by pest insects. In natural systems, the occurrence, intensity and direction of selection on the herbivore may vary among localities and through time. The speed of insect resistance development is

influenced by variation in plant defenses over time, although there is no real consensus as to whether it may increase or delay the process (Gardner et al. 1998; Gardner and Agrawal 2002; Renwick 2002). Seasonal variation in plant defences may offer a window of opportunity for insects to evolve resistance. On the other hand, temporal variation in the plant defence may reduce the intensity of the selection pressure on insect resistance. Geographical variation in plant defence and host plant range also influences the outcome of insect resistance development, as does the distributional range of the insect and the level of insect mobility. The spatial structure of a plant-insect interaction should therefore not be ignored when studying insect resistance in a natural system. Different components influencing population structure such as gene flow, genetic drift and strength of selection need to be taken into account for a better understanding of the evolution of local adaptation (as already acknowledged by Wright (1931)). Consequently, understanding insect resistance includes knowledge of seasonal, geographic and genetic variation in both plant defence and herbivore adaptation. A natural system of plant and insect which are both polymorphic with respect to a trait involved in resistance provides an excellent opportunity to study geographic variation in resistance.

STUDY SYSTEM

The main aim of my PhD project was to investigate the distribution of resistance of the crucifer specialist *Phyllotreta nemorum* to *Barbarea vulgaris* ssp. *arcuata* (Opiz.) Simkovics (Brassicaceae) defenses. The large striped flea beetle, *P. nemorum* L., is an herbivorous insect belonging to the family Chrysomelidae. Large is relative here, for an adult flea beetle is only 2,5 to 3 mm in size. Like other leaf beetles of this family, the flea beetle is a common pest of various cruciferous plants. They feed on Brassicaceae crops such as radish (*Raphanus sativus* L.) and turnip (*Brassica rapa* L.) (Alford 1999; Hedrick 1999). *Phyllotreta nemorum* larvae are leaf miners and just like the adult beetle they attack young leaves of the seedlings and can cause considerable damage to the plants due to their preference to attack the young plants (Neigel 1997; Alford 1999). Plant leaves are damaged by adults with a typical pattern called 'shot-holing'. In N-W Europe, this flea beetle has one generation per year in the field; the adults hibernate in winter. When adults appear in spring, they reproduce and lay eggs in the soil or on the plant. Larvae hatch after 2-4 weeks and the development from egg to adult takes 1-1½ month (see Figure 1 for the life cycle of the large striped flea beetle).

Barbarea vulgaris R. Br., a biannual plant, has two recognized subspecies in Denmark: *Barbarea vulgaris* ssp. *vulgaris* and *B. vulgaris* ssp. *arcuata*. Of the latter, two types are distinguished in Denmark; One has **p**ubescent leaves (hence called the P-type) and the other one has **g**labrous leaves (hence called G-type) (Nielsen 1997b). Beside this trait, they also differ in glucosinolates (Agerbirk et al. 2001; Agerbirk et al. 2003b), flavonoid profile (Dalby-Brown et al. 2011), saponin profile (Kuzina et al. 2009; Nielsen et al. 2010b) and cytogenetics ((Ørgaard and Linde-Laursen 2007), but see also (Ørgaard and Linde-Laursen 2008)). The *B. vulgaris* ssp. *arcuata* P-type and G-type also differ in their level of specialist insect de-

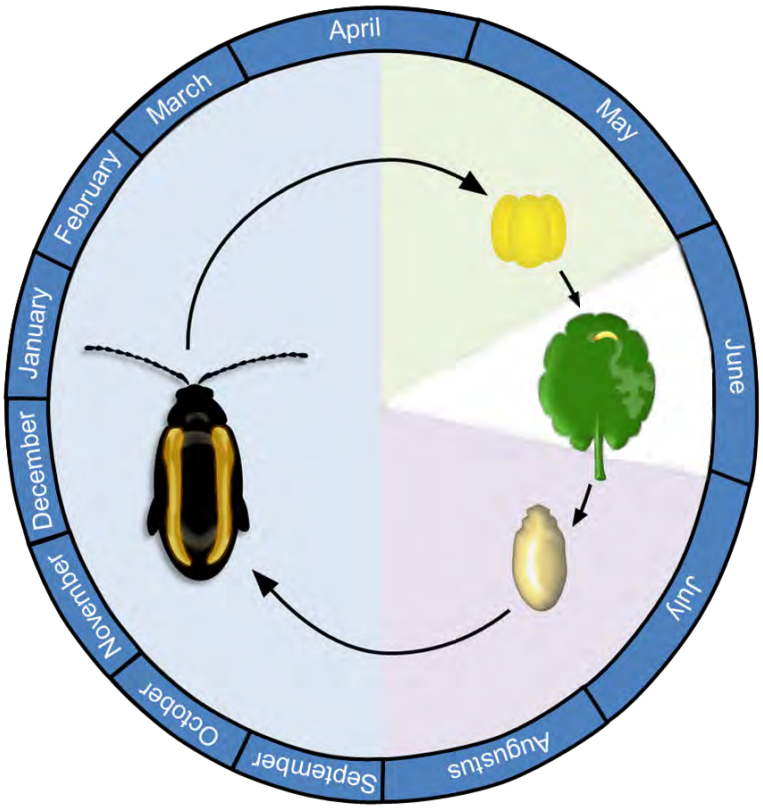


Figure 1 Schematic representation of the life cycle of the large striped flea beetle, *Phyllotreta nemorum*.

fence. The P-type is suitable for specialist insects such as the diamondback moth (DBM) (*Plutella xylostella* L.) and the flea beetle *P. nemorum* the whole year around. The G-type, on the other hand, is unsuitable for the DBM and *P. nemorum* during summer (Nielsen 1997b; Agerbirk et al. 2003a). Hybrids between both types are present in the field, but rare and there is a partial reproductive barrier between P-type and G-type (Agerbirk et al. 2001; Toneatto et al. 2010).

Interestingly, some flea beetle individuals can feed and develop on the G-type of *B. vulgaris* despite its repellence to most flea beetles (Nielsen 1996). Crosses between Danish flea beetles collected on G-type revealed that major resistance genes (R-genes) control both larval survival on, and adult acceptance of, defended G-type as a host plant (Nielsen 1996; Nielsen 1997a; Nielsen 1999; de Jong et al. 2000). Several modes of inheritance were found depending on the population of flea beetles studied; inheritance is mostly autosomal and in some populations there is sex-linked inheritance (Nielsen 1997a; de Jong and Nielsen 1999; Nielsen 1999; de Jong et al. 2000). In general, R-genes are rare in the Danish populations (de Jong and Nielsen 1999) and in some populations there is evidence of additional minor genes being involved in the resistance of the flea beetle (de Jong and Nielsen 1999). In the East of Denmark, near Kværkeby, inheritance is only autosomal and no effect of minor genes was found for these populations (de Jong, et al. 2000). Remarkably, on *B. vulgaris* G-type near Kværkeby nearly all beetles were found to be homozygous resistant, while on

nearby host plant patches the proportions of resistant beetles were surprisingly low (de Jong and Nielsen 1999; de Jong et al. 2000; Nielsen and de Jong 2005). One would expect this distribution of resistance with low insect mobility, but the low amount of genetic differentiation estimated with allozyme variation indicated that there is abundant gene flow among subpopulations (de Jong et al. 2001). Another explanation for this distribution of the R-gene is the existence of local differences in selection on resistant alleles (R-alleles). Logically, strong selection in favor of R-alleles is expected on the G-type, because beetles can only fully develop on the G-type when possessing at least one of these R-alleles. On other host plants, however, R-alleles may be under negative selection due to negative pleiotropic effects that are associated with these alleles. There are indications for negative fitness consequences of R-alleles. When an R-allele originating from a population on *B. vulgaris* G-type near Ejby was backcrossed in a near isogenic susceptible line, intraline crosses produced very few homozygous resistant beetles due to high pre-adult mortality (de Jong and Nielsen 2000) (Figure 2). Also for R-alleles from a population on *B. vulgaris* G-type near Kværkeby negative fitness effects were found, albeit again when backcrossed with a susceptible line

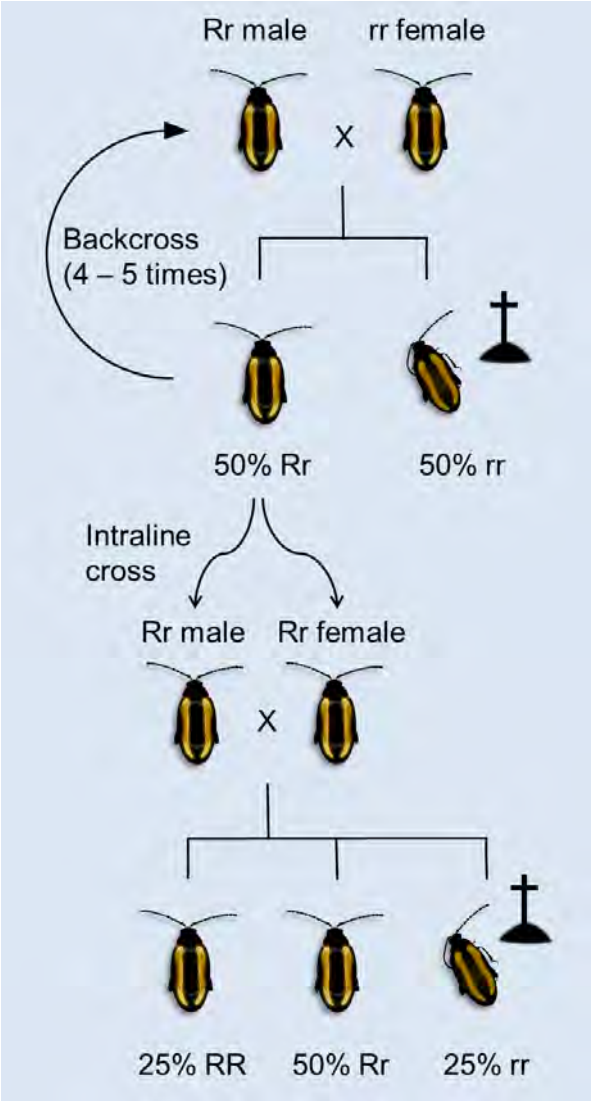


Figure 2 Crosses made between large striped flea beetles, *Phyllotreta nemorum*, to embed the R-gene in a susceptible genetic background. Heterozygous males with an autosomal R-allele originating from an Ejby population were initially crossed with susceptible females. Susceptibility is displayed as rr, heterozy-gous and homozygous resistance as Rr en RR, respectively. Percentages of expected genotypes are given for offspring survival on radish, under the assumption that all genotypes survive equally well on radish. These resistant beetles were backcrossed with susceptible beetles multiple times to produce heterozygous beetles with a susceptible genetic background. These heterozygous beetles were also crossed amongst each other in intraline crosses to produce offspring that would include homozygous resistant beetles (RR) with a susceptible genetic background. Only resistant beetles survive on *Barbarea vulgaris* G-type (hence the susceptible genotype displayed dead). Because the genotype cannot be distinguished from RR en Rr beetles when placed on G-type, analytical crosses were made with these beetles to determine their genotype.

(Breuker et al. 2007). If the R-allele on itself has negative consequences for the flea beetle fitness, we would not expect to find mostly homozygous resistant beetles on *B. vulgaris* G-type, especially not with the low survival that was found for homozygous resistant beetles in the lab (de Jong and Nielsen 2000). Another mechanism which could cause the observed distribution of resistant beetle frequencies near Kværkeby, is a preference for *B. vulgaris* G-type as a host plant. So far, however, no preference has been found for *B. vulgaris* G-type (Nielsen 1996)(P.W. de Jong & K.M.C.A. Vermeer, unpubl.).

We do not know if possession of R-alleles has negative consequences for the flea beetles in the field. There might have been strong selection in favour of modifier genes which counteract the negative effect of the R-alleles by epistasis (de Jong et al. 2000). The complex of resistance gene(s) and positive modifier gene(s) is called a coadapted gene complex. These complexes may play a role in the distribution of the R-alleles as found in the field (de Jong and Nielsen 2002). Crossbreeding between populations would cause these coadapted gene complexes to break up (Dobzhansky and Pavlovsky 1958; Lynch 1991) and without positive modifier genes, the R-gene would suffer from negative fitness effects. This way, even with moderate gene flow at the neutral level, resistance experiences a barrier resulting in the low frequency of R-alleles that we observe in populations near the *B. vulgaris* G-type.

Susceptible flea beetles respond the same to *B. vulgaris* as diamondback moth (DBM); *Barbarea vulgaris* P-type is fed on by both species and *B. vulgaris* G-type is only fed on when defence levels are low (Nielsen 1996; Nielsen 1997b; Agerbirk et al. 2003a). Glucosinolate levels are not related to the defence level of *B. vulgaris* G-type (Agerbirk et al. 2001), but certain saponin levels are (Agerbirk et al. 2003a; Kuzina et al. 2009). Saponins are a class of compounds occurring in many plant species with various biological effects, such as insec-

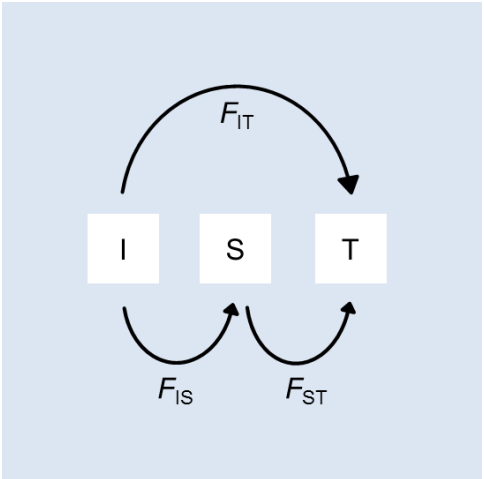


Figure 3 The classical view of population structure by Wright (1931). "*F*" is the fixation index that can be calculated for different levels of population structure (see text for how to calculate *F*). "I" stands for the individual level, "S" for the subpopulation level and "T" for the total population level.

ticidal properties (Sparg et al. 2004; Vincken et al. 2007). Hederagenin cellobioside was found as defence compound against DBM (Shinoda, et al. 2002) in *B. vulgaris* var. *variegata*. In *B. vulgaris* ssp. *arcuata*, oleanolic acid cellobioside was related to resistance against DBM (Agerbirk, et al. 2003a) and susceptible flea beeltes (Kuzina et al. 2009). The latter study also found that hederagenin cellobioside relates to resistance levels of *B. vulgaris* ssp. *arcuata* G-type and two other saponins that relate to the resistance levels, later identified as being gypsogenin cellobioside and 4-epihederagenin cellobioside (Nielsen et al. 2010b). Hederagenin cellobioside and oleanolic acid cellobioside were most abundant of these four compounds in defended plants. Bioassays with leaf discs treated with a solution containing hederagenin cellobioside demonstrated that this saponin inhibits flea beetle feeding (Nielsen et al. 2010a). Furthermore, food consumption was significantly less for susceptible flea beetles than resistant flea beetles. All of these findings show that saponins are responsible for the defence in the G-type. In the light of the previously mentioned evolutionary arm's race between plant and insect, the above findings suggest that saponins in the G-type act as a secondary line of defence against flea beetles, after the insects became adapted to glucosinolates.

Phyllotreta nemorum is a very suitable species for studying the mechanisms of insect adaptation to plant defenses in a natural system, because of the nature of its adaptation. The resistance is polymorphic, can be clearly assessed through a bio-assay, and inherited in Mendelian fashion. The species is abundant where it occurs and can be reared in a laboratory. Moreover, the species has already been studied to a certain extent (e.g. de Jong et al. 2000; Breuker et al. 2005; Nielsen and de Jong 2005; Breuker et al. 2007) as has its host plant *B. vulgaris* G-type (Huang et al. 1994; Agerbirk et al. 2001; Agerbirk et al. 2003b; Ørgaard and Linde-Laursen 2007; Kuzina et al. 2009; Nielsen et al. 2010a; Nielsen et al. 2010b; Kuzina et al. 2011). Consequently, scientific knowledge and molecular tools such as microsatellite loci, are available for this species (Verbaarschot et al. 2007; de Jong et al. 2009). Moreover, the type of compounds responsible for the defence in *Barbarea vulgaris* G -type are known. Despite of the knowledge on how resistance in flea beetles is inherited, it is not known yet how the flea beetle resistance mechanism operates.

The aim of this thesis was to answer the following questions: (1) which mechanism underlies the flea beetle adaptation to the defence of *B. vulgaris* G-type? (2) how is the resistance distributed through the flea beetle populations? and (3) what underlies this pattern?

In order to answer these questions, I investigated the flea beetle population structure in Denmark and used a population genomics approach.

A better understanding of the flea beetle adaptation to plant defences may also have considerable practical value. It might prove helpful in controlling damage on crops by *P. nemorum*, but also by closely related species that are also considered pests of *Brassica* species (Jolivet et al. 1988; Andersen et al. 2006).

METHODS

The study of population genetics, which forms a major part of this thesis, involves a rather large body of theory. Here, I introduce the most important concepts and models in population genetics that are used in this thesis.

Population genetics analysis

Population genetics is the study of the genetic composition (distribution and change) in a population. As most populations are structured to some extent, genetic drift, gene flow, difference in mutation levels and selection levels can all lead to changes in the genetic variation of subpopulations. These changes do not only occur over time but also between subpopulations.
The classical view of a structured population came from Wright (1931); a population is not a panmictic population, but is structured in subpopulations in which there is random mating and these subpopulations are connected with each other through gene flow. Wright also developed *F*-statistics (also known as fixation indexes) to estimate the amount of genetic variability of hierarchically subdivided populations. In his approach, a population can have a population structure at several levels (Figure 3) (Wright 1951); one from the individual (I) to the subpopulation (S), one from the subpopulation to the total population (T), and the third level then relates to the individual (I) to the total population (T).

Genetic variation is often measured as the amount of heterozygosity. In population genetics, the term heterozygosity does not refer to the genotype of a locus in a diploid individual, but refers to the fraction of individuals that is heterozygous for a particular locus at a certain population level. Nei (1973) extended this concept to multiple loci and a finite number of subpopulations. F_{IS} is then expressed as the level of heterozygosity in a population over loci, calculated as the difference in observed number of heterozygotes compared to the expected number of heterozygotes under the assumptions of the Hardy-Weinberg Principle (see Box 1). The formula for estimating F_{IS} is then

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

where H_S is the expected heterozygosity over loci within subpopulations and H_I is the observed heterozygosity over loci within subpopulations.
The formula for estimating F_{ST} (named G_{ST} by Nei) is

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T is the expected heterozygosity over loci in the total population and H_S is the expected heterozygosity over loci within subpopulations.

Box 1 Hardy-Weinberg Principle

According to the Hardy-Weinberg principle the frequencies of alleles will remain constant (in equilibrium) in the absence of selection, mutation, migration and genetic drift, in an infinitely large population with random mating (Griffiths et al. 2012). Although these requirements seem impossible to meet, the Hardy-Weinberg principle holds for surprisingly many populations. Subpopulations, for example, are expected to be in *Hardy-Weinberg equilibrium* (HWE) if there has been no recent bottleneck or a sudden increase in effective population size. Departure from HWE can be caused by many factors; genotyping errors, non-random mating, selection, genetic drift, population structure, selection, mutation, migration and genetic drift (Weir 1996). Testing for departure from HWE can be done with several statistical tests (Weir 1996). Another application of the Hardy-Weinberg principle is to derive genotypes from phenotypes, assuming that allele frequencies of that locus are in HWE.

If all three levels of population structuring are in Hardy-Weinberg Equilibrium (HWE), each *F*-statistic has a value of 0. With *F*-statistics one can calculate whether there is inbreeding in the investigated population and if bottlenecks have occurred. F_{ST} values are between 0 and 1, where F_{ST} = 0 means that the subpopulations are one panmictic population and F_{ST} = 1 means that subpopulations are completely different from each other. F_{IS} values are between -1 and 1, where values near 1 indicate a recent bottleneck and values below 0 may indicate balancing selection.

Most methods to measure genetic differentiation that are used nowadays are derived from Wright's *F*-statistics. The appropriate use of *F*-statistics and its derived measures to answer particular questions in the area of population genetics has been debated extensively (Weir and Cockerham 1984; Hedrick 1999; Neigel 2002; Jost 2008; Leng and Zhang 2011; Meirmans and Hedrick 2011). This has led to the formulation of other measures of differentiation (see Box 2), which are each taking into account some specific assumptions that are more realistic in particular cases and depending on what question one wants to answer. Each of these have been named differently by the authors, but are often referred to by others as F_{ST}. For example, software program fdist is said to estimate the expected F_{ST} but does this for bi-allelic as well as multi-allelic loci (Beaumont and Nichols 1996).

Population genetics work with the *neutral theory* (Kimura 1983). According to this theory many polymorphisms are hardly or not at all under selection and therefore mainly under influence of genetic drift, mutation rate and gene flow. As a consequence, an increase in population size results in less influence by genetic drift has and thus an increase in heterozygosity. Another prediction from this theory is that loci with a high mutation rate are expected

Box 2 Different estimators of population differentiation

Several methods are used to estimate the genetic differentiation of (sub) populations. The genetic variation of a sample has for a long time been measured as the level of heterozygosity. Several people questioned whether this is a good measurement for gene diversity and developed other measures like the effective number of alleles (e.g. Jost 2008). The methods used to estimate differentiation differ in underlying assumptions and models. They may depend on the type of mutation model used, subpopulation size, number of subpopulations, dominance of markers used, number of alleles per marker, mutation rate. Below, several methods are given for the measurement of genetic differentiation:

$\boldsymbol{F_{ST}}$ measures the genetic variation in a subdivided population (Wright 1951).

$$F_{ST} = \frac{s^2}{\bar{p}(1-\bar{p})}$$

where $\bar{p}$ is the average sample frequency of the allele over the samples and s^2 is the sample variance. Originally, the *F*-statistics were designed for bi-allelic loci. This method assumes an infinite number of subpopulations and the infinite allele mutation model.

$\boldsymbol{G_{ST}}$ is an analogue of F_{ST} developed to include loci with multiple alleles and a more realistic finite number of subpopulations (Nei 1973). By others often still referred to as F_{ST}.

$$G_{ST} = \frac{H_T - H_S}{H_T}$$

Where H_T is the expected genetic variation in the total population under Hardy-Weinberg assumptions and H_S is the expected level of heterozygosity among subpopulations under Hardy-Weinberg assumptions.

$\boldsymbol{G'_{ST}}$ is a standardised version of G_{ST} (Degen et al. 2004; Hedrick 2005). Several authors noticed some serious limitations of the *F*-statistics, especially for estimator G_{ST} (Weir and Cockerham 1984; Kappers et al. 2011). When the level of within subpopulation heterozygosity is high for loci, values of G_{ST} can be very small even if subpopulations only share a few alleles. Hedrick defined an estimator less biased towards highly variable loci:

$$G'_{ST} = \frac{G_{ST}\,(k - 1 + H_S)}{(k - 1)(1 - H_S)}$$

Where H_S is again the expected level of heterozygosity among subpopulations under Hardy-Weinberg assumptions and k is the number of subpopulations.

Continuation of Box 2 Different estimators of population differentiation

***R*ST** is analogous to F_{ST} but differs in underlying mutation model (Slatkin 1995). Slatkin developed R_{ST} specifically for microsatellite loci whose mutation process seems to correspond better to the stepwise mutation model (Slatkin 1995). According to this model alleles of similar sizes are more closely related than alleles of different sizes. R_{ST} is calculated as

$$R_{ST} = \frac{S_T - S_W}{S_T}$$

where S_T is the variance in allele size in the total population and S_W is the variance in allele size within subpopulations. Note that this equation is very similar to the one of G_{ST} and differs from G_{ST} in using the variance in allele size instead of the level of heterozygosity.

D does not use the level of heterozygosity to estimate gene diversity as well, instead the effective number of alleles is used (Nei 1977; Jost 2008). The estimator of D is calculated as

$$D_{est} = \frac{H_T - H_S}{1 - H_S} \times \frac{n}{n - 1}$$

Where H_T is the estimated genetic variation in the total population, H_S is the estimated genetic variation among subpopulations and n is the number of subpopulations. Though this is thought a better estimate of genetic differentiation between populations, it is less suitable than *F*-statistics for estimating population parameters like gene flow and migration (Jost 2009) because it is dependent on the mutation rate and level of heterozygosity of a locus (Degen et al. 2004; Whitlock 2011).

θ is an analogue to F_{ST} and G_{ST} that is often used as an estimator of genetic differentiation because it was developed especially to estimate genetic differentiation using the number of alleles per subpopulation / per locus / per individual as parameters (Weir and Cockerham 1984; Kappers et al. 2011). By regarding θ (F_{ST}) as a parameter rather than a statistic, their method incorporates a sampling error that F_{ST} and the like do not because they assume that all subpopulations are sampled.

to have a higher heterozygosity than loci with a low mutation rate. Indeed, for microsatellites – short tandem repeats that usually have a high mutation rate – the level of heterozygosity is often very high compared to other loci (Weber and Wong 1993).

Population genetic parameters

Parameters that influence the genetic composition of a population such as gene flow, genetic drift, mutation rate and selection pressure, are not known but can be estimated using *F*-statistics and the neutral theory. Because there is no direct way to estimate levels of gene flow in natural populations, this is done by an indirect approach (e.g. Slatkin 1981). Gene flow can be inferred from *F*-statistics, but is depending on underlying models. The original *F*-statistics as developed by Wright uses for example the infinite alleles mutation model (IAM) (Kimura and Crow 1964). According to this model, each time a mutation occurs it is a new unique allele. Another mutation model that has been developed because of the nature of some loci is the stepwise mutation model (SMM) (Ohta and Kimura 1973). In the SMM it is assumed that a mutation adds or deletes an integer in an allele so that the mutation process does not completely erase the prior allelic state, unlike in the IAM model. SMM seems especially suitable for microsatellites, because alleles of these loci consist of different numbers of short tandem repeats. Mutation is thought to add or delete a repeat of the microsatellite and hence the SMM seems a more accurate model than the IAM model (Henderson and Petes 1992; Weber and Wong 1993). Regardless, the SMM is not always preferable when microsatellites are used and it might be preferable to calculate estimators by using different mutation models (Takezaki and Nei 1996; Balloux and Lugon-Moulin 2002).

Lewontin and Krakauer (1973) realized that population genetic parameters affect the genome in different ways; processes like genetic drift, inbreeding and gene flow act in a similar way on nearly the whole genome, but natural selection acts specifically at a particular locus. Loci that are only affected by genome wide processes like genetic drift and gene flow are called neutral loci. Lewontin and Krakauer reasoned that neutrality of particular loci can be tested for by *F*-statistics. If the F_{ST} value of a particular locus deviates significantly from other F_{ST} values, then it is likely influenced by a locus-specific effect like selection.

Population genomics approach

The idea of Lewontin and Krakauer (1973) has been refined by Beaumont and Nichols (1996) to identify candidate loci that may be affected by natural selection. Their method is referred to as a population genomics approach. Population genomics is simply population genetics but with specific use of multiple loci across the genome, so it can distinguish between genome-wide effects and locus specific effects (Black et al. 2001; Luikart et al. 2003; Schlotterer 2003). Beaumont and Nichols (1996) advice to use a large number of loci across the genome, preferably with high levels of heterozygosity. The F_{ST} values or a derivate of F_{ST} is calculated for each locus in a pairwise comparison between subpopulations and should be weighted by the level of heterozygosity of that locus. A distribution of F_{ST} values can be drawn for each locus and outliers are candidate for locus-specific effects. Black et al.

(2001) mentioned the population genomics approach for the first time as an approach for plant-insect interactions to assess which parts of the genome only experience genome-wide effects and which parts also experience adaptive variation. The power of this approach lays in the ability to find candidate loci for being under selection without needing quantitative trait loci (QTL) or model species (Luikart et al. 2003). In CHAPTER 2 I will explain in further detail how the population genomics approach works and what kind of promising work can be done with it and in CHAPTERS 5 and 6 I apply a population genomics approach.

Candidate gene approach

In the population genomics approach many loci are sampled when looking for outliers. These loci should not be completely random, but include many neutral loci to set a baseline. In addition, loci can be included that one suspects to be a candidates for selection (Black et al. 2001). An example of such a candidate gene approach is the study of lake whitefish where the differentiation at growth-associated QTL was compared between normal and dwarf ecotypes and compared with genetic differentiation at the neutral level (Rogers and Bernatchez 2005). QTLs coming out as an outlier in the population genomics approach for pairwise comparisons between a normal and dwarf ecotype but not for comparisons between only normal or dwarf ecotypes, suggest that that particular QTL is under directional selection among these ecotypes.

In this thesis I took an integrated approach to examine geographical variation in resistance to the plant defense among wild insect populations. I have combined two complementary approaches: an empirical approach via the study of variation in resistance in flea beetle populations (CHAPTER 4), and a population genomics approach via molecular markers to gain insight in the genomic make-up of the population and its connection with the resistance trait (CHAPTERS 5 and 6).

OUTLINE OF THE THESIS

The principal goal of this study was to investigate the resistance of the flea beetle *P. nemorum* to *B. vulgaris* G-type 's defence, find the gene that is held responsible for the flea beetle's resistance and explain the distribution of this trait in natural populations.

In **CHAPTER 2,** I review the geographic mosaic theory of coevolution and the population genomics approach. I give an in-depth background of population genomics in plant-insect interactions and how to use it to work with predictions from the geographic mosaic theory of coevolution. Furthermore, I illustrate how useful the application of population genomics can be and how it can lead to more insights in a study system with a structured population and different modes of selection among the subpopulations.

In **CHAPTER 3**, I investigated the actual resistance gene in *P. nemorum*, with β-glucosidase as a candidate. β-glucosidases are enzymes that hydrolyzes glycosidic bonds, removing

monosaccharides such as glucose from the target molecule. Three β-glucosidase cDNA sequences were cloned from a resistant Danish flea beetle laboratory line. The enzyme of one of these sequences, named hereafter β-glucosidase C, was able to hydrolyse hederagenin cellobioside to hederagenin monoglucoside and glucose, when expressed in an insect cell line. These results suggest that expressed β-glucosidase C can deglycosylate antifeedant saponins and may play a role in the resistant flea beetle's ability to overcome the defence of *B. vulgaris*.

CHAPTER 4 focuses on variation of *P. nemorum* resistance to *B. vulgaris* G-type 's defence on other host plants. In order to make inferences about adaptive geographic variation, it is crucial to consider current and historical demographic processes that influence the distribution of genetic variation within and between populations. I investigated if the earlier found decrease in flea beetle resistance on other host plants forms a continuing trend. The frequency of resistant beetles did not decrease among the years, in contrast to what had been suggested. In addition, I addressed variation in resistance within the season. To this end, I sampled flea beetle populations in Denmark on *B. vulgaris* G-type and *S. arvensis* multiple times within the reproductive season of one year. I found that the frequency of resistant beetles in a population varied significantly within the season.

In **CHAPTERS 5 and 6** I used a population genomics approach to reveal geographic patterns of local adaptation in the flea beetle population near Kværkeby. In both chapters I sampled various microsatellites and the resistance phenotype of each beetle.

In order to look at differences within and between subpopulations, I first had to check if the populations sampled on different plant patches were at least partially genetically distinct, but also connected by gene flow or only recently subdivided from a common ancestral population. We used the program STRUCTURE in **CHAPTER 5** to look at the genetic diversity of the sampled subpopulations, using neutrally behaving microsatellites. The next step was to compare the genetic differentiation based on these microsatellites with genetic differentiation of the resistance trait between subpopulations on other host plants than *B. vulgaris* G-type. Pairwise comparisons between *B. vulgaris* G-type and *S. arvensis* would serve as additional evidence, because we know that the resistance trait is highly favoured on *B. vulgaris* G-type and therefore we expected to find for the resistance trait a significant deviation from genetic differentiation estimated with neutral loci. Indeed the resistance trait came out as an outlier in the comparison between the population on *B. vulgaris* G-type and the populations on *S. arvensis*. Even more interestingly was, that we also found significant deviation from genetic differentiation estimated with neutral loci for the resistance trait in a pairwise comparison between subpopulations on non-defended host plants at geographic and genetic different distances from populations on *B. vulgaris* G-type, suggesting that there is directional selection on the resistance trait on non-defended host plants as well.

In **CHAPTER 6** we repeated the outlier approach, this time with the gene coding for β-glucosidase found in chapter 3 as a candidate for resistance. Because β-glucosidase B and β-glucosidase C were so similar and originally thought to be the same sequence, and β-

glucosidase C was initially only cloned from resistant flea beetle lines, we hypothesized that both sequences were alleles of the major resistance gene, with β-glucosidase C being the dominant resistance allele. First, I investigated whether the resistance phenotype of the flea beetle is correlated with the frequencies of β-glucosidase B and β-glucosidase C found in CHAPTER 3. Additionally, I checked if the frequencies of β-glucosidase B and β-glucosidase C were according to Hardy Weinberg Expectations within the subpopulations. Subsequently, the genomic approach of CHAPTER 5 was implemented to find out if the candidate gene shows signs of being under selection.

Finally, **CHAPTER 7** summarizes and discusses the main results of the previous chapters and gives perspectives on future research in the light of findings of this thesis.

CHAPTER 2

The potential of a population genomics approach to analyse geographic mosaics of plant–insect coevolution

Kim M. C. A. Vermeer • Marcel Dicke • Peter W. de Jong

Modified version published in: Evolutionary Ecology (2011) 25: 977-992

Abstract

A central issue in the evolutionary ecology of species interactions is coevolution, which involves the reciprocal selection between individuals of interacting species. Understanding the importance of coevolution in shaping species interactions requires the consideration of spatial variation in their strength. This is exactly what the, recently developed, geographic mosaic theory of coevolution addresses. Another major development in the study of population ecology is the introduction of the population genomics approach in this field of research. This approach addresses spatial processes through molecular methods. It is of particular interest that population genomics is especially applicable to natural populations of non-model species. We describe how population genomics can be used in the context of the geographic mosaic of coevolution, specifically to identify coevolutionary hot spots, and to attribute genetic variation found at specific loci to processes of selection versus trait remixing. The proposed integration of the population genomics approach with the conceptual framework of the geographic mosaic of coevolution is illustrated with a few selected, particularly demonstrative, examples from the realm of insect-plant interactions.

INTRODUCTION

Organisms are under the constant threat of attack by their enemies and have evolved a range of defence strategies. How organisms evolve in the context of attack-defence relationships has been under intensive debate for several decades (e.g. Courtney 1988; Fox 1988; Rausher 1988; Thompson 1988). Since 1994 John Thompson has developed his influential theory "The Geographic Mosaic Theory of Coevolution" (Thompson 1994, Thompson 1999a, Thompson 2005). The primary premise of this conceptual framework to explain coevolutionary interactions and –patterns is that populations are generally genetically and ecologically structured and that, therefore, coevolution between interacting species must be studied at different scales, including that of local populations. Coevolution *sensu stricto* is thought to occur only in a limited range of the geographical distributions of the species involved, called "coevolutionary hot spots". Beyond such local coevolution, larger-scale coevolutionary dynamics are additionally influenced by geographic selection mosaics and trait remixing (here defined as the changing of the spatial distribution of alleles by processes including gene flow across landscapes, random genetic drift within populations, extinction and recolonization of local populations, and mutation; Thompson 2005; Gomulkiewicz et al. 2007). This view on coevolution has led to clear, testable predictions, and indeed these are taken up by an expanding range of scientists, who experimentally test these predictions in a variety of systems (e.g. Thompson 1999a, Thompson 1999b, Thompson 2009a; Burdon and Thrall 1999; Lively 1999; Brodie et al. 2002; Benkman et al. 2003; Neuhauser et al. 2003; Zangerl and Berenbaum 2003; Toju and Sota 2006). One of the central challenges is to explain the observed geographical distribution of alleles at adaptive loci in terms of selection (stabilizing, directional, disruptive or balancing) *versus* migration and genetic drift. In the context of the geographic mosaic of coevolution, this challenge translates into the identification of coevolutionary hot spots, and to distinguish their effects on genetic variation from those of trait remixing (Thompson 2005; Gomulkiewicz et al. 2007).

At roughly the same time as the development of these novel concepts in research on coevolution, a new approach was introduced in the field of evolutionary ecology and evolutionary genetics: population genomics (Black et al. 2001; Schlotterer 2002; Luikart et al. 2003; Dicke et al. 2004). This approach was inspired by two key factors: 1) the realization that, generally, populations are genetically structured, and consist of local sub-populations that are linked by migration, and 2) the progress in the development and analysis of molecular markers, that have become widely available for non-model organisms. Essentially, population genomics is an approach that attributes the distribution of genetic variation to locus-specific effects, such as selection, *versus* genome-wide processes such as migration, by genome-wide sampling of molecular markers. A recent development is the connection of population genomics to the field of phylogeography (Brito and Edwards 2009; Hickerson et al. 2010), which has traditionally provided a neutral template for the study of the evolution of coevolving traits, and aims to describe population history and estimate demographic parameters.

The approach of population genomics shows considerable congruence with the conceptual model of a geographic mosaic of coevolution. After introducing the two key topics of this paper in more detail, we will explore how the population genomics approach can lead to more insights in the dynamics of the geographic mosaics of coevolution.

THE EVOLUTION OF THE CONCEPT OF COEVOLUTION

Coevolution, i.e. reciprocal selection between individuals of interacting species, is an important process influencing adaptations. Classic examples of coevolution are found among the interactions between insects and plants (Ehrlich and Raven 1964; Berenbaum and Zangerl 1992; Funk et al. 2002; Cornell and Hawkins 2003; Pellmyr 2003; Nuismer and Thompson 2006; Futuyma 2009). This is not surprising, because insect herbivores represent more than 25% of all multi-cellular species (Strong et al. 1984) and our green world is abundantly covered by plants, making insect-plant interactions one of the most common interactions within ecosystems on this planet (Schoonhoven 2005).

Since the concept of coevolution was first conceived (Darwin 1859, Darwin 1862), there have been different views on how to define it. At one extreme, only reciprocal one-to-one interactions at the species level are included (Strong et al. 1984), also called pairwise coevolution by others such as Janzen (Janzen 1980). At the other extreme, coevolution is regarded as a process that is diffuse and multispecific (all species may interact with several other species; this is also known as diffuse coevolution) (Janzen 1980; Fox 1981; Futuyma and Slatkin 1983). As a result of using a very strict definition of coevolution, several authors such as Jermy (1984), Strong et al. (1984) and Schmitt et al. (1995) do not agree with the idea that coevolution is responsible for, or has at least a major impact on, insect-plant interactions.

The two extreme views on coevolution mentioned above, i.e. pairwise coevolution and diffuse coevolution, usually ignore the spatial structure of species interactions (Thompson 1999b). As a result, much of the coevolutionary process that occurs in between these two extremes is ignored. The spatial structure of species interactions is explicitly included in the more recently developed theory of the geographic mosaic of coevolution (Thompson 1994, Thompson 2005; Thompson et al. 1997). In this theory, much of the coevolutionary process occurs at levels in between those of local populations and species. This relates to the structure that many populations exhibit. Populations are generally not homogeneous, but geographically and genetically structured (Wright 1951, Wright 1968; Avise 2000; Thompson 2005), due to, for example, the patchy nature of their habitats and the patchy nature of biotic and abiotic environmental factors (Agrawal et al. 2001). One feature leading to structured populations is the lack of complete dispersal by individuals within a population so that mating within the population is not random. A metapopulation (Hanski 1998) often consists of subpopulations, that each have their own population dynamics and a certain degree of mating restriction between them.

Apart from the spatial structure of these metapopulations, that contributes to a geographic mosaic of coevolution, the fact that geographical ranges of interacting species usually do not completely overlap leads to spatial variation in species interactions and in the intensity of natural selection. For example, in plant-insect interactions, some patches of host plants may not be within the distributional range of a particular herbivore, while other patches might be under attack by this natural enemy, resulting in a difference in selection intensity on the host plant among patches. The distributional range of an insect species can be smaller than that of the host plant (Strong et al. 1984) or even exceed the host plant's range, for example when an insect uses multiple host plants. Both cases result in an unequal distribution of selective interactions across the distributional range. Also the geographical range of other herbivore species may have an effect on the selective interactions in a metapopulation of a species under study.

In most of north-western North America the moth *Greya politella* uses woodland stars (*Lithophragma parviflorum*) as its only host plant and acts as its pollinator when ovipositing in the flower (Thompson and Pellmyr 1992). In this case the costs of seed feeding by the moth's offspring are smaller than the benefit resulting from pollination. When other effective co-pollinators are involved, however, the pollination of *L. parviflorum* no longer depends on *G. politella*; the mutualistic effects of *G. politella* on *L. parviflorum* (Thompson and Pellmyr 1992) are swamped. Thus, other pollinators present in some subpopulations interfere with the selective interaction between *G. politella* and *L. parviflorum*.

Lodgepole pine (*Pinus contorta* ssp. *latifolia*) forms another example of a host plant experiencing differences in the interaction with different herbivores. The morphology of lodgepole pine cones in and near the Rocky Mountains has evolved differently because of different selection pressures imposed by red squirrels (*Tamiasciurus hudsonicus*) and red crossbills (*Loxia curvirostra*) (Benkman 1999; Benkman et al. 2001, Benkman et al. 2003). Both species eat the seeds from the partially closed pine cones. When no red squirrels are present, the red crossbill imposes a selection on the cone shape (cones are relatively narrow at their base) (Benkman 1999), but when red squirrels are present, the squirrels harvest most cones before crossbills have access to them. The presence of red squirrels imposes such a strong selection on the cones (cones are short and wide at the base when squirrels are present) that red crossbills no longer influence cone morphology (Benkman et al. 2001). Therefore, selection on cone shape is imposed by red crossbills only when red squirrels are absent. So, within the lodgepole pine species range, different reciprocal selective forces act due to differences in the distributional range of the herbivores.

Environmental variation (spatial and temporal) further contributes to this mosaic of coevolution. Individuals in subpopulations can evolve traits that make them best adapted to the local environmental conditions. However, because environmental conditions vary in space and time, an evolved trait may lose its adaptive value when local conditions change. Similarly, migration between patches can cause a locally adapted genotype to spread to localities where it does not perform as well as in the original patch. In an extreme case, the adaptation might even become a maladaptation in the new patch. So, a variable environment and mi-

gration between patches may cause local mismatches of traits and maladaptation (Thompson 1999b; Schoonhoven et al. 2005).

THE IMPORTANCE OF IMPLEMENTING A SPATIAL COMPONENT IN THE THEORY OF COEVOLUTION

Ignoring the geographical structure of populations might lead to an underestimation of the importance of coevolution. Coevolution between plants and insects is likely to be diffuse, because a herbivorous enemy of a plant species may alter the pattern of selection exerted by other natural enemies (Rausher 1996). Rausher also states that there can only be either pairwise or diffuse coevolution. Following Rausher's idea, if an interaction that evolves between a plant and its enemy is not influenced by the absence or presence of another species, this is called pairwise coevolution. All other evolutionary processes are diffuse (Iwao and Rausher 1997). By ignoring the geographical structure of populations, however, one ignores the possibility of the existence of both pairwise *and* diffuse coevolution. Where one subpopulation of the plant might be attacked by several enemies, plants in another patch can still be under selection exerted by only one enemy, thus being involved in pairwise coevolution at the local scale.

Strong et al. (1984) use geographical structure precisely as an argument why coevolution would not be common. Plant-insect interactions are variable and unpredictable in time and space and, therefore, Strong et al. (1984) argue that selection pressures differ locally, making it less likely that coevolution (which needs reciprocal and intense interactions) is actually working; in other words, they argue that the selection pressure is not stable and intense enough to drive coevolution. During some years a patch may consist of two interacting species, while in another year one of them can be absent or even more species may be present that interfere with this interaction. A reciprocal interaction would, thus, not be sustained and intense and, therefore, in the view of Strong et al. (1984) coevolution would not take place.

The geographical structure of plant-insect interactions, i.e. Strong et al.'s argument why coevolution is not common, is exactly what makes coevolution so flexible and such a major force in shaping ecologically relevant traits (Thompson 2009b). This is precisely what provides the raw material of the geographic mosaic of coevolution. As a certain local adaptation occurs in some places but not in others, a geographical structure arises of different selective interactions, the so-called selection mosaics. Given the overwhelming evidence that indeed populations of most, if not all, species are geographically structured even at the local scale (e.g. Bermingham and Avise 1986; Bohonak 1998; Althoff and Thompson 1999; Medrano and Herrera 2008; Nosil et al. 2008; Gomez et al. 2009a), Thompson's view of a dynamic mosaic of ever changing coevolutionary interactions, governed by (sometimes rapid) local adaptation and trait remixing through migration, in our view best describes the existing pattern of species interactions (Thompson 2005).

GEOGRAPHIC MOSAIC THEORY OF COEVOLUTION (GMTC)

Considering the process of coevolution as a hierarchical process operating at different spatial and temporal scales, the geographic mosaic theory of coevolution consists of three components: (a) geographic selection mosaics, (b) trait remixing and (c) coevolutionary cold and hot spots (Thompson 2005). "*Geographic selection mosaics*" refers to the occurrence of geographic differences in fitness of interacting species. Selection mosaics are not just variable selective forces on interactions, but relate to spatial differences in genotype-by-genotype-by-environment interactions amongst interacting species (Thompson 2005). "*Trait remixing*" refers to the changing of the distribution of alleles by processes such as gene flow and genetic drift. Gene flow between local subpopulations and genetic drift within the subpopulations may alter the distribution of traits by changing the genetic composition of the subpopulation, which can be further modified by local extinction of subpopulations and mutation. *Coevolutionary hot spots* are local communities where reciprocal selection through mutual interactions between individuals of different species occurs. Such reciprocal selection does not occur in *cold spots*. These three components are considered to make up the raw material of the GMTC, i.e. they constitute the driving force behind coevolution.

Based on these three components of the geographic mosaic of coevolution, different patterns and dynamics of interspecific interactions are expected than when considering a population that is not geographically structured. The patterns expected under the GMTC are trait mismatching among interacting species, few species-level coevolved traits and spatial variation in traits mediating interactions among species (Thompson 2005). Trait mismatching, for example, can be caused by one of the forces that contribute to trait remixing, e.g. gene flow between hot spots and cold spots. Because local communities differ, a certain trait can be an advantage in some hot spots whereas in others it is not, or may even cause a local maladaptation. Gene flow can prevent local coevolution of traits in some local communities, thereby causing local trait mismatching and maladaptation. Few species-wide coevolved traits are expected in a landscape of geographically and genetically structured subpopulations, because only in some cases have coevolved traits become fixed in the complete metapopulation (Thompson 2005).

Verifying these ecological predictions is valuable, but not enough to test the GMTC per se (Gomulkiewicz et al. 2007; Thompson 2005, Thompson 2009a). Spatially variable traits can, for example, also occur in antagonistic interactions, even when none of the three components of the GMTC are present. Local maladaptation can also take place without geographic selection mosaics, hot spots, cold spots and trait remixing. The same holds for the expected low number of coevolved traits found at the species level; they can also be observed when a geographic mosaic of interactions is not present. Even when gene flow is completely absent, populations can show patterns such as local maladaptation and trait mismatching (Nuismer et al. 2003). So, when the three patterns are all present, this does not necessarily mean that there is a geographical mosaic of coevolution.

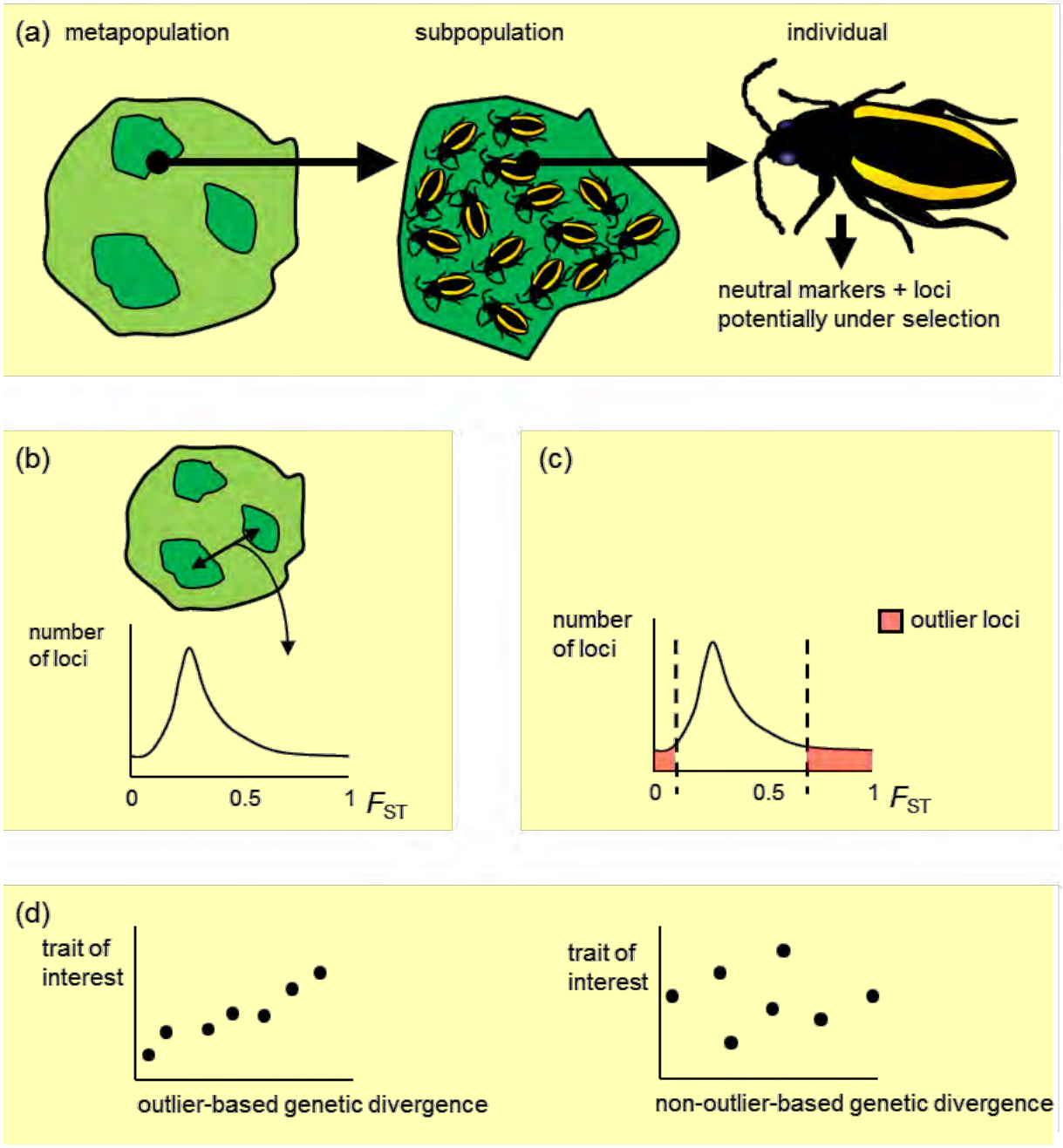


Figure 1 The population genomics approach consists of four steps: (a) sampling numerous (neutral) molecular markers – such as microsatellites – as well as testing loci that are suspected to be under selection in individuals of different tentative subpopulations; (b) A frequency distribution is drawn of a measure of the variation between different samples, in this example the F_{ST} value for each locus; (c) Screen the frequency distribution for statistical outliers; (d) Validation step to verify whether the outlier is caused by selection (or by hitch-hiking with/on a genome region that is under selection). In this example, non-outlier- and outlier-based genetic distances are correlated with differences in the ecological trait of interest.

It is important to realize that in addition to demonstrating the occurrence of the three *patterns*, i.e. local maladaptation, spatially variable traits and a paucity of coevolved traits at the species level, unraveling the components of the underlying *processes* that generate these patterns is vital to support the GMTC. The processes will reveal whether the patterns are indeed caused by the components of the GMTC. The components are the three components mentioned earlier: selection mosaics, cold and hot spots, and trait remixing. Since 2005, rapid progress has been made in formulating and testing of predictions from the GMTC (Thompson 2009a; Laine 2009). Approaches to test the underlying components are described by Gomulkiewicz et al. (2007).

The processes that generate the patterns predicted by the geographical mosaic theory of coevolution in the first place act on genetic variation. For a genotype to have effect on another genotype of an interacting species, i.e. the very basis of coevolution, there must be enough genetic variation in both species. There is a surprising paucity in empirical evidence of the genetics underlying natural adaptations, while the underlying genetic basis is a fundamental part of the understanding of natural adaptation (Orr 2005; Stinchcombe and Hoekstra 2008). Thompson emphasizes the need to study the *genetics* underlying traits that have evolved through coevolution and the processes influencing the spread and distribution of these genetic factors (Thompson 1999b, Thompson 2005). Hence, such knowledge is vital for understanding the evolutionary processes that shape the geographical mosaic of coevolution. To study genetic variation and how this influences the predictions of the GMTC, means that one cannot simply use a single disciplinary field, but must use a multidisciplinary approach, including ecology, genetics and molecular biology. A promising way to investigate the geographic mosaic theory of coevolution is to exploit a population genomics approach. Population genomics provides an interface between population genetics and molecular biology (Black et al. 2001), which links the molecular genetic basis of (coevolutionary) forces to their consequences at the population level.

POPULATION GENOMICS

Since Black et al. (2001) introduced the population genomics approach in the field of plant-insect interactions, it has proven to be useful in a range of studies (e.g. Rogers and Bernatchez 2005; Egan et al. 2008; Herrera and Bazaga 2008; Minder and Widmer 2008; Nosil et al. 2008; Schneider 2008; Butlin 2010; Ikeda and Setoguchi 2010). Briefly, the population genomics approach as developed by Black et al. (2001), involves the following. By sampling numerous markers throughout the genome in individuals from one or more populations, parts of the genome can be identified that differ in variation across populations from that of neutral markers, due to locus-specific effects such as selection (Black et al. 2001). Genome-wide effects, that affect all markers in the genome, include genetic drift (founder effects and population bottlenecks), migration and inbreeding (Black et al. 2001). What is done in the population genomics approach is that genome-wide effects and locus-specific effects are differentiated to gain more insight in evolutionary processes that influence variation within and across populations (Luikart et al. 2003; Stinchcombe and Hoekstra 2008).

In more detail, the procedure of a population genomics approach to detect selection is as follows (see also Figure 1): The first step is to formulate clear biological hypotheses with respect to contrasts between specific samples, and candidate loci to be included. A second step is the sampling of numerous loci (most of them thought to be neutral) in numerous individuals of the different groups of interest, e.g. putative subpopulations, by genotyping individuals for molecular markers such as microsatellites. Then a frequency distribution is produced of a measure of variation between the samples of interest, based on the range of loci that has been sampled. For neutral loci (e.g. most microsatellites), this is expected to be a continuous distribution with a certain maximum with stochastic variation around it. Subsequently, a screen for outlier loci that exhibit values of the measure of variation that deviate from values for the rest of the genome is carried out. Obviously, the correct identification of outlier loci is crucial to the success of this approach. Methods and software to do this are summarized in the excellent reviews of Luickart et al. (2003) and Butlin (2010). Assuming that the variation at the majority of the sampled loci is the result of genome-wide effects, the outlier loci are thought to mark adaptive variation, since they are the result of locus-specific effects such as selection. Examples of how different forms of selection (stabilizing, directional, disruptive and balancing) influence the measure of variation between samples of interest are given in Black et al. (2001). The selection of candidate loci in step 1 of the population genomics approach (see above), and the validation step that follows below, increase the rigor with which outliers can be adequately attributed to selection, rather than other locus-specific effects (e.g. mutation, assortative mating, and recombination). Other ways to confirm the presence of genuine outliers, and the cause of outlier behaviour, are reviewed in Luikart et al. (2003). Among others, repeatability of the detection of outlier loci in independent samples strengthens support for the correct identification of outliers, and, if the independent samples are from different localities with similar putative selection gradients, supports the interpretation that the outliers mark adaptive loci. Further validation of selection being indeed the cause of the deviation of the outlier locus should be obtained in a final step of the analysis. For the hawk moth-pollinated violet *Viola cazorlensis* this was done by comparing the phenotypic divergence of the floral trait of *V. cazorlensis* with the genetic divergence of the outlier loci (Herrera and Bazaga 2008). A strong relationship was found between the genetic divergence of the outlier loci and three floral traits, i.e. the length of the peduncle, spur and upper petal of the flower, while there was no relationship between the genetic divergence of the neutral loci and any of the investigated floral traits. In this validation, it is also important to consider alternative factors that may explain the presence of outlier loci. For example, Manel et al. (2009) correlated outliers in a phytophagous insect (the large pine weevil, *Hylobius abietis* (L.)) to host plant use vs. abiotic factors.

A specific advantage of the population genomics approach is that it enables the scanning of the genome for ecologically relevant (Thompson 2009b) genetic variation without having to know the phenotypes (Stinchcombe and Hoekstra 2008). Because the population genomics approach implies that the genome is scanned for outlier loci by using anonymous markers, no prior knowledge is needed regarding the phenotype of a trait of interest. Other advantages of the population genomics approach are that a possible breeding history is not needed and that this approach can be applied to naturally occurring non-model organisms.

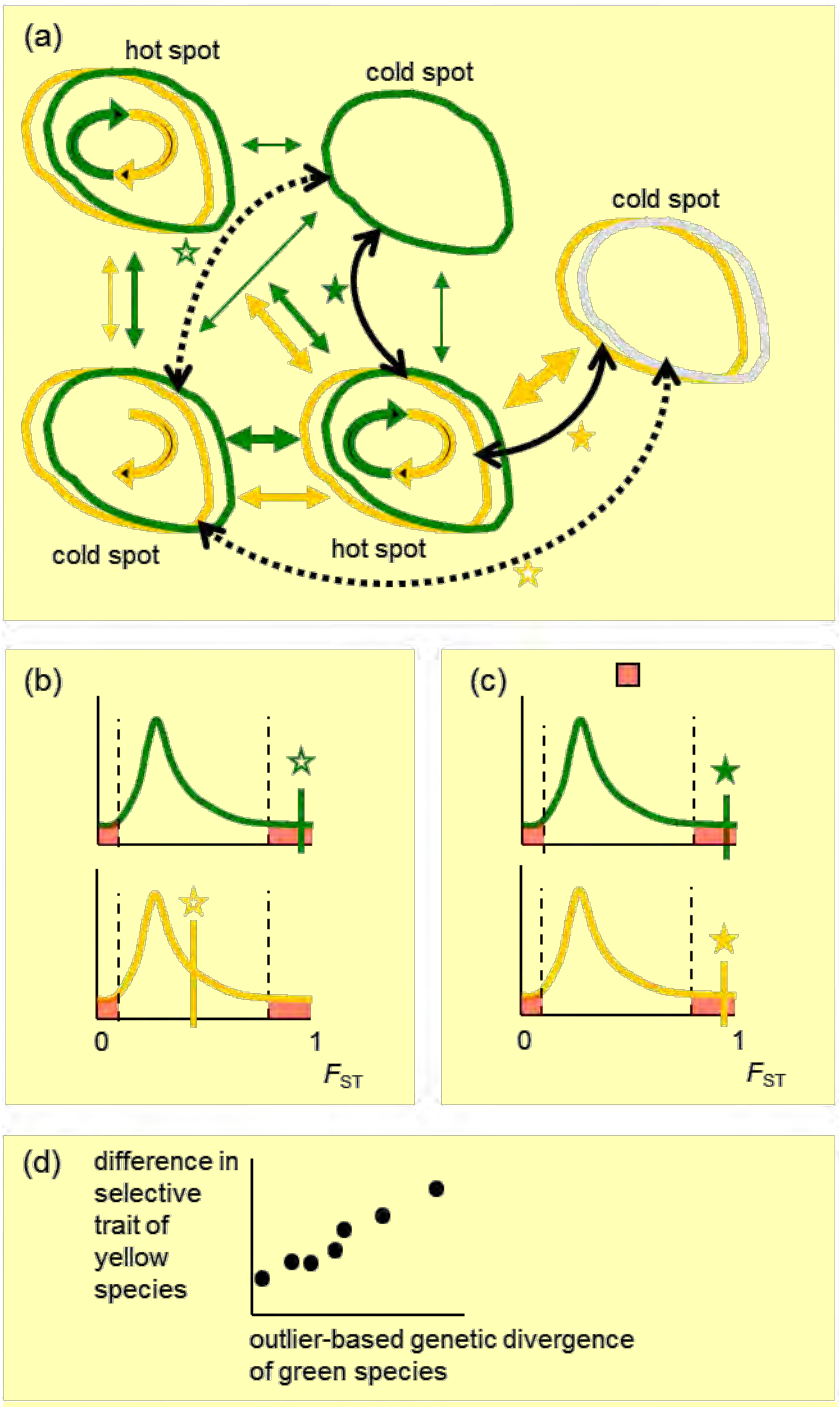


Figure 2 Using population genomics to detect selection and to identify hot spots and cold spots. (a): A hypothetical example has been drawn for a set of local populations of two interacting species (respectively green and yellow patches). Green and yellow arrows between patches represent dispersal of the respective species, with thicker arrows indicating more dispersal between the local populations. Curved arrows within patches show selective interactions within local populations; a green arrow shows selection from the green species on the yellow species, and a yellow arrow vice versa. The occurrence of both a yellow and a green arrow within one patch indicates reciprocal selection, so that such a patch becomes a coevolutionary hot spot. In some patches, selection only acts from one species upon the other, and in other patches, no selection occurs between the two species of interest, e.g. when one of them is not present in that local patch. This is, for example, the case in the right-hand patch, where the yellow species is interacting with a third species (grey), and the green species is absent. (b) and (c): examples of detection of selection within two local populations (the ones at the bottom of part (a) of the figure) by applying and F_{ST}-based population genomics approach, comparing each of the two populations with local patches where only one of the two species is present (and hence no selection is exerted between them). These comparisons are indicated with curved, black (respectively dashed and solid) double-headed arrows in part (a), accompanied by green and yellow open, and closed asterisks, respectively, for the two different pairwise comparisons. The vertical bars in (b) and (c) show the F_{ST} value for candidate loci, suspected to be under selection. In (b), the F_{ST} for the candidate locus for the green species is an outlier, but for the yellow species it is not. This shows that the yellow species exerts selection on the green one, but not vice versa, and the local community is thus a coevolutionary cold spot. In (c), the F_{ST} values for the candidate loci for both interacting species are outliers, showing that reciprocal selection may be occurring, and thus indicating a coevolutionary hot spot. (d) A validation involves the correlation of outlier F_{ST} values of one species with the hypothesized selective pressure imposed by the other species (the latter needs to be determined empirically).

Population genomics becomes particularly useful if segregating sites can be related to traits of interest, by identifying functional genetic polymorphisms (Storz 2005; Vasemagi and Primmer 2005). For the lake whitefish (*Coregonus clupeaformis)*, for example, a population genomics approach was combined with adaptive QTL (quantitative trait loci) mapping to examine growth rate differences between dwarf and normal ecotypes (Rogers and Bernatchez 2005). Significantly higher levels of divergence were found for several growth-associated QTLs.

INTEGRATION OF POPULATION GENOMICS APPROACH AND GMTC

In the context of the geographic mosaic of coevolution, the population genomics approach is an excellent tool to investigate the two basic processes leading to the three predictions of this theory (see also Figure 2):

a) *Detection of variable selection leading to the identification of hot spots and cold spots that cause spatial variation in traits mediating interactions among species.*

Cold spots are easy to identify if one considers a patch where only one of the two interacting species occurs (Gomulkiewicz et al. 2007). Population genomics is, therefore, not needed to identify cold spots. Hot spots, however, can only be identified by showing that reciprocal selection occurs. If a certain trait is expected to be under selection, the population genomics approach can be used to establish whether the locus of such a candidate trait is indeed under selection, like in the previous example of the growth rate differences between two ecotypes of the whitefish (Rogers and Bernatchez 2005). Traits that are expected to be under selection should be checked in both interacting species to identify a genuine hot spot, since there, by definition, selection should be reciprocal.

b) *Detecting trait remixing and more specifically, gene flow*

In a geographical mosaic of coevolution, trait remixing is thought to be the underlying cause of trait mismatching in interacting species. Gene flow, next to local extinction and random genetic drift, is one of the processes that influences trait remixing and can, therefore, influence geographic selection mosaics. If a population displays considerable genetic structure at certain loci, the influence of gene flow on homogenizing the variation at such loci in the metapopulation is limited. The population structure, however, needs not be the same at all loci (Thompson 2005); if selection at a certain locus is stronger than migration, then gene flow at that locus may be limited between subpopulations. So, only measuring average gene flow and showing population structure by using molecular markers, is not enough to prove the existence of a geographical mosaic of coevolution. Population genomics is useful to separate the processes of selection and migration/dispersal, which both influence gene flow. When a frequency distribution is produced of the variation of the sampled loci, statistical parameters – such as the F_{ST} statistic (Wright 1951), or any other appropriate measure of dif-

ferentiation (F_{ST} is not an appropriate measure under all circumstances, see Gregorius et al. 2007; Gillet and Gregorius 2008; other measures, each with their own (dis) advantages (see e.g. Meirmans and Hedrick 2010; Sefc et al. 2007) include: Q_{ST} (Spitze 1993), R_{ST} (Slatkin 1995), G_{ST} (Nei 1987), Θ (Weir and Cockerham 1984), Φ (Excoffier et al. 1992), *C* (Xu et al. 2009), G'_{ST} (Hedrick 2005), *D* (Jost 2008), δ (Gregorius et al 2007), D_m (Nei 1973)) – can be estimated for the loci that are considered to only undergo genome-wide effects. After removing outliers, the mean F_{ST} of these neutral loci, with a stochastic variation around it, should indicate to what extent subpopulations are differentiated with respect to genome-wide effects (Wright 1951), i.e. to what extent hot and cold spots are linked.

For *Viola cazorlensis* it was demonstrated that phenotypic traits such as the length of the flower petals are subject to selection by the pollinating hawk moth, because gene flow is high between the patches of *V. cazorlensis* and yet the phenotypic floral traits differed much between the patches (Herrera and Bazaga 2008). A strong average gene flow between the patches was found by sampling many neutral loci. Further details of the way population genomics can be employed to study the processes leading to a geographic mosaic of coevolutionary interactions are given in the examples below.

Examples of integrating the GMTC and population genomics

An interesting example of ecological adaptation is the interaction between the oligophagous flea beetle, *Phyllotreta nemorum* L., and one of its host plants, the crucifer *Barbarea vulgaris* ssp. *arcuata* (Opiz.) Simkovics. *Barbarea vulgaris* ssp. *arcuata* consists of two varieties (Nielsen 1997b): the P-type which can be used as food by all individuals of *P. nemorum* and the G-type which is unsuitable for most flea beetles during summer. These types can hybridize in the field, yielding hybrids with intermediate chemical defence. Flea beetle individuals can be susceptible or resistant to the defence of the *B. vulgaris* G-type.

The host plants, as well as the flea beetles, have a patchy geographical distribution (de Jong et al. 2001, de Jong et al. 2009). Resistance traits in both plant- and beetle varieties vary geographically and temporally (Nielsen and de Jong 2005; Toneatto et al. 2010). The defence of *B. vulgaris* varies in that some patches consist of the P-type and some of the G-type, thereby varying in suitability as well. In some patches nearly all *P. nemorum* are resistant to *B. vulgaris'* defence (G-type patches), whereas in other patches only some or none of the flea beetles are resistant (P-type patches). Most *B. vulgaris* patches (>80%) are free of *P. nemorum*, whereas *P. nemorum* is also found on other host plants than *B. vulgaris*, most notably *Sinapis arvensis* L. On the non-*B. vulgaris* patches, frequencies of *P. nemorum* that are resistant to *B. vulgaris* G-type are relatively low (de Jong and Nielsen 1999). This leads to a geographic mosaic of interactions between *B. vulgaris* and *P. nemorum* (Nielsen and de Jong 2005).

Crosses revealed that the flea beetle's resistance is caused by a dominant major resistance gene (R-gene). The exact genetic basis at the DNA level of the resistance trait of *P. nemorum* is not yet known, although a candidate gene has been identified (CHAPTER 3). For

B. vulgaris, the biochemical basis of the defence against phytophagous insects (including *P. nemorum*) has been unraveled (Kuzina et al. 2009; Nielsen et al. 2010a), and efforts have been made to elucidate the genetic basis of the production of this chemical defence (Kuzina et al. 2011). The presence of the different types of *B. vulgaris* on the one hand, and the different resistance genotypes of the flea beetles on the other hand, as well as their geographical distribution has prompted the question to what extent processes with genome-wide effect, such as dispersal/migration, and locus-specific effects, like selection, are responsible for these observed patterns. A population genomics approach is the obvious tool to examine this question. As prerequisites for application of this approach, microsatellite markers have been developed for both the flea beetles (Verbaarschot et al. 2007 and unpublished) and *B. vulgaris* (Toneatto et al. 2010). Furthermore, loci/markers are being identified for putative adaptive traits in both interacting species. One of the specific biological questions that can be addressed with population genomics is: can the difference in frequency of resistant flea beetles between other plants and *B. vulgaris* G-type be attributed to a limited dispersal of the beetles, or is selection on resistance in the flea beetles involved? By sampling flea beetles on *B. vulgaris* G-type and on other host plants and applying a population genomics approach using the microsatellite markers and the candidate gene for resistance, any involvement of selection can be detected. Samples of flea beetles on other host plants than *B. vulgaris* G-type, for example having different frequencies of resistant beetles (e.g. at different distances from a *B. vulgaris* G-type patch) can also be contrasted in a population genomics approach. In this way, hypothesized selection against resistance on these other plants by disruption of coadapted gene complexes (de Jong and Nielsen 2002) can be detected and distinguished from effects of dispersal. Analogously, the population genomics approach can be used to detect selection on chemical defence in *B. vulgaris*. For example, patches with and without presence of *P. nemorum* can be compared and the distribution of alleles involved in the level of chemical defence can be attributed to selection vs. dispersal. By combining the outcomes for the beetles and *B. vulgaris*, tentative coevolutionary hot spots, where selection for both interacting species is detected, can be identified (Figure 2). This way of applying population genomics would be a slightly different approach than that from Black's original perspective; whereas they *identified* candidate loci by using outlier loci, this investigation would study whether the a priori candidate gene is an outlier, in order to draw conclusions with respect to the involvement of selection in the observed geographical distribution of the different alleles of the trait of interest.

Another example of a population genomics approach in plant-insect interactions is the study of the leaf beetle *Neochlamisus bebbianae* and its host plants (Funk 1998; Egan et al. 2008; Funk and Nosil 2008). The populations of the leaf beetle are associated with either maple trees or willow trees. These leaf beetle populations are partially differentiated in host preference and performance traits and they exhibit premating reproductive barriers (Funk 1998; Egan and Funk 2006). By using population genomics, Egan and co-workers (Egan et al. 2008) aim to investigate the contribution of host-plant related divergent selection to genetic differentiation during ecological speciation.

An ecologically comparative genome scan of AFLPs for pairwise population comparisons of the beetle *N. bebbianae* was made. Then different host-population comparisons were contrasted with same host-population comparisons. Outlier AFLPs in the first set of comparisons, that were not outliers in the second set of comparisons, should represent regions experiencing host-specific selection. Certain regions were found that matched these criteria and those candidate gene regions are now further investigated to find specific genes (and their function) that contribute to ecological speciation (Nosil et al. 2008). Not only will these genes provide insight to the process of speciation, they are also nice examples of genes encoding for traits that vary across the populations, differing in cold and hot spots.

Local adaptation and maladaptation can be found even in a generalist plant-pollinator system when combining the strength of population genomics with the geographic mosaic theory of coevolution. Gomez and co-workers found that *Erysimum mediohispanicum* forms a selection mosaic of varying selective regimes mediated by different pollinators (Gomez et al 2009a, Gomez et al. 2009b). Although *E. mediohispanicum* and its interactors represent a generalist system, meaning that more than one-to-one interactions are present, it is composed of geographic mosaics of selection. This illustrates once again how important it is to also focus on the spatial scale of interactions, when investigating (co)evolution. Also generalist systems, involving many interacting organisms, can consist of different selective regimes caused by specific components of the community.

One of the predictions of the GMTC is the presence of a mosaic of local adaptation and maladaptation across the population. For *E. mediohispanicum* and its pollinators this was tested by comparing the plant's attractiveness to pollinators (Gomez et al 2009a). Plants originating from hot spots (i.e. spots where selection on the plant by pollinator assemblages is large) were more attractive than plants from cold spots. Randomly amplified polymorphic DNA (RAPD) was used to evaluate genetic differences between populations. Since most RAPDs are thought to be neutral, they can be used to calculate the genetic distance between populations. The genetic difference between populations was compared to differences in adaptation (being the attractiveness of the plants to pollinator assemblages, which can vary across different localities), to find out whether this trait was associated with the genetic differences between populations. Gene flow between the investigated populations was low, considering the large genetic differences found between populations with molecular markers. Knowing that plant fitness is highly influenced by pollinator visitation rate – plants in hot spots being more attractive than plants in cold spots – this indicates that *E. mediohispanicum* in hot spots is locally adapted, but maladapted in cold spots.

CONCLUSION AND FUTURE PERSPECTIVES

The concept of coevolution has been presented more than a century ago. So far, especially temporal aspects have been included in the study of coevolution, although temporal studies of coevolution in the true sense by monitoring reciprocal genetic change over time are relatively rare (J.N. Thompson, pers. comm.). To fully understand the evolutionary ecology of

insect-plant interactions, the inclusion of spatial variation in selection pressures and reciprocal interactions between individuals is essential. The geographical mosaic theory of coevolution has proven to be a valuable approach in this respect. Furthermore, by applying a population genomic approach when examining the geographical mosaic of coevolution, one can understand more of the genetic variation influencing adaptive traits under natural conditions and thereby more of coevolution. Given the important evolutionary questions that are open in the context of the effects of climate change on biodiversity, community structure, and species interactions, combining population genomics with a coevolutionary approach that includes geographical aspects is likely to yield valuable information to understand the past and predict future scenarios.

Acknowledgements

We are grateful to John Thompson for constructive comments on a preliminary draft of this paper, and sharing his thoughts about its topic with us. We also thank two anonymous reviewers for their helpful comments.

CHAPTER 3

Identification of a saponin degrading β-glucosidase from the flea beetle *Phyllotreta nemorum*

Kim M. C. A. Vermeer • Patrick Verbaarschot • Peter W. de Jong • Manabu Kamimura

Abstract

The crucifer *Barbarea vulgaris* is chemically defended by saponins against herbivores, including the flea beetle *Phyllotreta nemorum*. Some Danish and Swiss populations of this beetle, however, do naturally occur on defended *B. vulgaris*. We analyzed the molecular mechanisms that enable the beetles to feed on defended *B. vulgaris*, focusing on β-glucosidases as candidates for the enzymes involved in this resistance through degradation of saponins. We cloned three β-glucosidase cDNA sequences from a resistant Danish beetle laboratory line. Although all three cDNAs were cloned from another resistant population, one β-glucosidase – named *β-glucosidase C* – could not be cloned from a susceptible population, suggesting that this enzyme plays a role in the flea beetle resistance to the defence of *B. vulgaris*. A crossing experiment using a flea beetle family segregating for the resistance trait showed that mRNA expression of *β-glucosidase C* is correlated with the ability of the F_1 beetles to feed and survive on *B. vulgaris*. Recombinant β-glucosidase C protein expressed by an insect cell line hydrolyzed the most active and abundant *B. vulgaris* saponin, hederagenin cellobioside, to glucose and hederagenin monoglucoside. These results suggest that expressed β-glucosidase C can deglycosylate antifeeding saponins and may play a role in the resistant flea beetle's ability to overcome the defence of *B. vulgaris*.

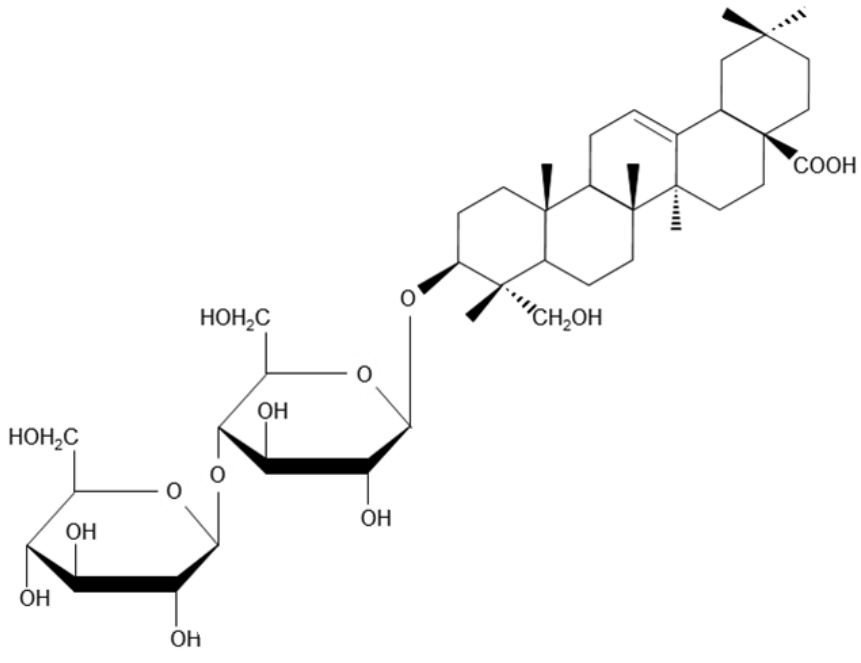


Figure 1 Structure of hederagenin cellobioside.

INTRODUCTION

The large striped flea beetle (*Phyllotreta nemorum* L., Coleoptera: Chrysomelidae) is a common pest of brassicaceous crops, such as radish and turnip (Alford 1999). Both larvae and adults feed from the preferably young plants, by mining (larvae) and pitting leaves, which can lead to considerable economic losses. Plants use secondary compounds to protect themselves against herbivory (chemical defense). Specialist insects, such as the large striped flea beetle, however, exploit these secondary compounds as feeding stimulant (Nielsen 1978; Lerin 1980; Fenwick et al. 1983; Nielsen 1988; Renwick 2002). *Phyllotreta nemorum* feeds on various plant species within the Cruciferae (Brassicaceae), all containing glucosinolates (Nielsen 1977). One of the crucifer species attacked by this insect is winter cress (*Barbarea vulgaris* ssp*. arcuata* R. Br.). In Denmark there are two types of this plant species. These types differ in glucosinolate profiles, saponin content, leaf pubescence and resistance towards insects like the diamondback moth and the flea beetle *P. nemorum* (Nielsen 1997b; Agerbirk et al. 2001; Shinoda et al. 2002; Agerbirk et al. 2003a; Agerbirk et al. 2003b). The so-called P-type (Agerbirk et al. 2001) has **p**ubescent leaves and can be eaten by all *P. nemorum* flea beetles while the other type has **g**labrous leaves (hence: G-type) and is defended against flea beetles during spring and summer (but not from autumn to spring) (Nielsen 1997b; Agerbirk et al. 2003b; Kuzina et al. 2009). Most flea beetles are susceptible to the G-type's defence, but some beetles are resistant, and can and do feed on it. The genetic basis of this resistance involves (a) major, dominant gene(s) (so called R-genes) (Nielsen 1997a; De Jong and Nielsen 1999). Only one R-allele already enables the flea beetle to feed and survive on the G-type. The major genes may be located at one or more loci, because previous studies revealed both autosomal and sex-linked inheritance (Nielsen 1997a; De Jong et al. 2000).

Saponin compounds have been demonstrated to be responsible for the defence in *B. vulgaris* G-type against insects such as *Plutella xylostella* and *P. nemorum* (Shinoda et al. 2002; Agerbirk et al. 2003a; Kuzina et al. 2009; Nielsen et al. 2010a). One of these, hederagenin cellobioside, appears to be a major defense compound. Shinoda and colleagues (2002) reported that the diamondback moth (DBM), *P. xylostella,* does not feed on *B. vulgaris* due to the feeding deterrent hederagenin cellobioside (see Figure 1). They suggested that this saponin might also act as a feeding deterrent to *P. nemorum*. Also oleanolic acid cellobioside may act as a feeding deterrent (Agerbirk et al. 2003a). Agerbirk and colleagues reported that the DBM can feed on the P-type throughout the year but that the G-type shows a seasonal variation in its defence to DBM, similar to the seasonal variation in defence against susceptible flea beetles. The variation in resistance in *B. vulgaris* G-type correlates with the amount of hederagenin cellobioside and the saponin is not found in the P-type of *B. vulgaris*.

Kuzina and colleagues confirmed that both hederagenin cellobioside and oleanolic acid cellobioside correlated significantly with plant resistance against the flea beetle, as well as two other saponins, identified as 3-O-cellobiosyl-4-epihederagenin and 3-O-cellobiosyl-gypsogeni (Kuzina et al. 2009; Nielsen et al. 2010b). More support for saponins being responsible for the defence of *B. vulgaris* G-type came from Nielsen and colleagues (Nielsen

et al. 2010a). The two saponins hederagenin cellobioside and oleanolic acid cellobioside isolated from *B. vulgaris* G-type had negative effects on the relative consumption rate of flea beetles when these compounds were offered on round leaf discs (diameter: 11 mm). Although relative consumption rate was negatively affected in resistant beetles by hederagenin cellobioside, the effect was much stronger for susceptible beetles. Hederagenin, the aglycone of hederagenin cellobioside, lacking the two glucose molecules at the C3 position of hederagenin cellobioside (Figure 2), had no effect on the consumption rate of both flea beetle genotypes. The effect of the saponin oleanolic acid cellobioside on the relative consumption rate differed only slightly between genotypes or was not significant at all. Hederagenin cellobioside is the most abundant of these tested saponins in the G-type of *B. vulgaris* (Kuzina et al. 2009). The removal of only one glucose molecule at the C3 position of heder-

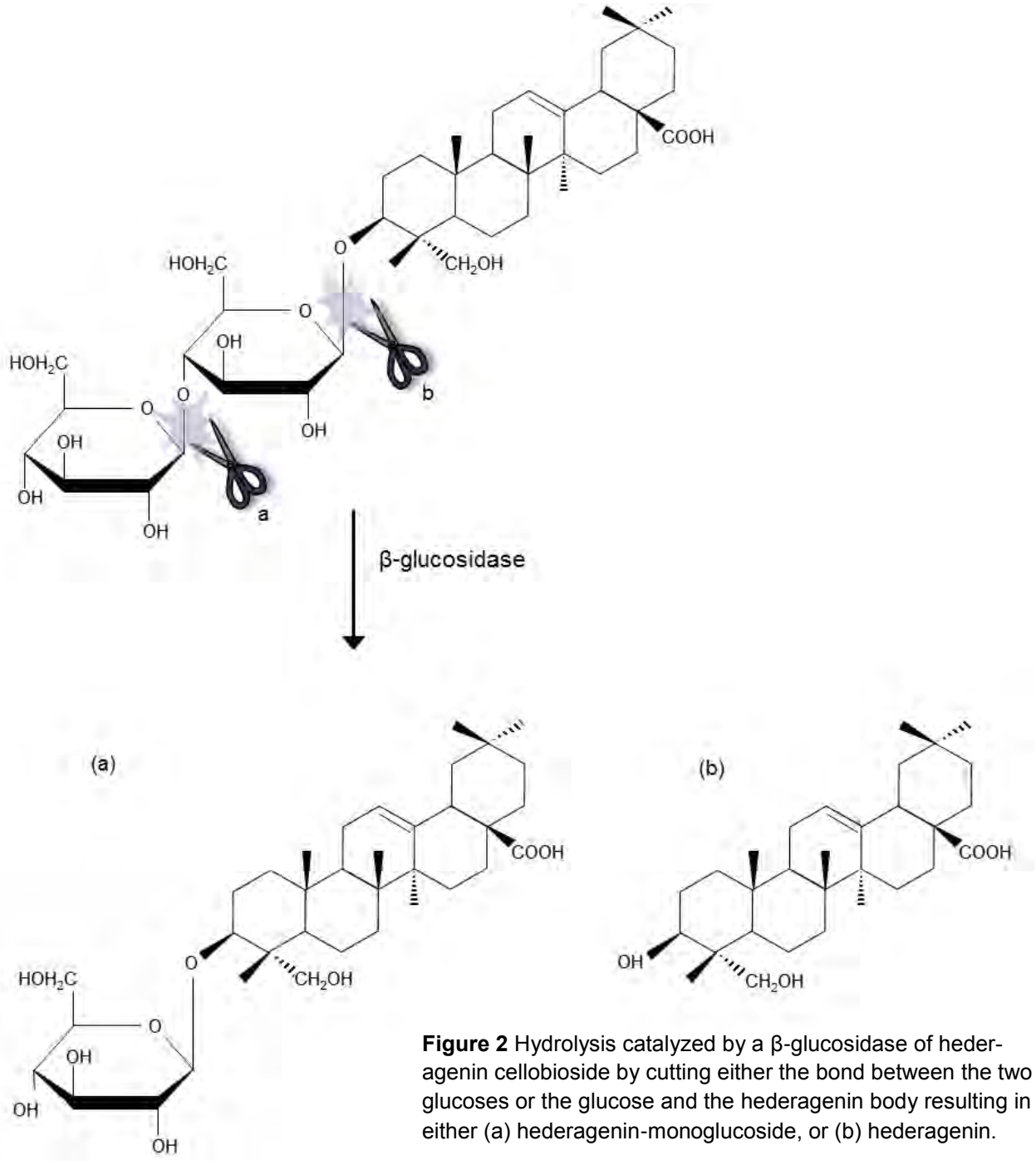


Figure 2 Hydrolysis catalyzed by a β-glucosidase of hederagenin cellobioside by cutting either the bond between the two glucoses or the glucose and the hederagenin body resulting in either (a) hederagenin-monoglucoside, or (b) hederagenin.

agenin cellobioside leads to hederagenin-monoglucoside and glucose. The monoglucoside has the same effect on the flea beetle as hederagenin cellobioside; it is less feeding deterrent for resistant beetles than for susceptible individuals in the same concentrations as hederagenin cellobioside (Augustin et al. 2012).

The observation that hederagenin cellobioside has a strong feeding deterrent effect on susceptible flea beetles compared to resistant beetles and that the aglycone does not show any feeding deterrent effect (Nielsen et al. 2010a), indicates that the R-alleles in the flea beetles might code for a saponin-degrading enzyme, more specifically for the degradation of hederagenin cellobioside.

Antifungal effects of saponins depend on their glucose units (Mshvildadze et al. 2000; Zhang et al. 2005). Some fungi, however, can detoxify host plant saponins by saponin degrading enzymes such as β-glucosidases (Bowyer et al. 1995; Osbourn et al. 1995; Morrissey et al. 2000; Faure 2002), but for insects such reports have not been published. β-Glucosidases (β-d-glucoside glucohydrolase, EC 3.2.1.21) are enzymes present in many living organisms such as animals, plants, fungi and bacteria (Esen 1993). The majority of these types of enzymes belongs to either glycoside hydrolases family 1 or family 3 (Henrissat 1991; Faure 2002). It catalyzes the hydrolysis of terminal β-glucose residues from oligo- or polysaccharides (see Figure 2). Because of its function we expect a β-glucosidase to be able to convert hederagenin cellobioside to hederagenin-monoglucoside or even hederagenin, by hydrolizing one respectively two glucose units.

In this chapter, we report on the analysis of the molecular mechanisms underlying the differences between susceptibility and resistance in the flea beetle *P. nemorum*, focusing on β-glucosidases as candidates of saponin-degrading enzymes. By means of cDNA cloning, mRNA expression analysis and enzymatic analysis using recombinant proteins, we found two β-glucosidases in the genome that were able to break down the saponin hederagenin cellobioside to its monoglycoside and glucose.

MATERIALS AND METHODS

Plants

Barbarea vulgaris ssp. *arcuata* (G-type) seeds were collected in 2004 in Herlev, Denmark (accession no. 3). Radish seeds (*Raphanus sativus* L. cv. Københavns Torve) were commercially obtained (Dæhnfeldt, Odense). Radish was grown at 20 ± 2 ˚C and a photoperiod of 18L:6D. *Barbarea vulgaris* plants were grown at 30 ± 6 ˚C and a photoperiod of 18L:6D (400W HPI/T-lamps, Philips, The Netherlands). *Barbarea vulgaris* plants were 3-8 weeks old and still in the vegetative stage when their leaves were used for bioassays.

Insects

Five near-isogenic flea beetle lines were used for initial cDNA cloning. Among them a susceptible line (rr genotype) originating from beetles collected on radish (*Raphanus sativus* L.) in Taastrup, Denmark (Nielsen 1999); this line is referred to as ST (Susceptible Taastrup) line. Beetles from this line are unable to survive on *Barbarea* and were maintained on radish in the laboratory. The other four near-isogenic lines (backcrossed with the ST line for at least 10 generations) were autosomal resistant lines originating from different Danish and Swiss populations: 1) the AE line originated from Ejby (A=autosomal, E=Ejby) (Denmark) (Nielsen 1997a), 2) the AK line originated from Kværkeby (Denmark) (De Jong et al. 2000), 3) the DE line originated from Delemont (Switzerland) (Nielsen et al. 2010a), and 4) the AT line originated from Try (Denmark). Beetles from the latter four lines have a single autosomal R-gene and can survive on the chemically defended G-type of *B. vulgaris* (Nielsen 1997a; de Jong et al. 2000). By backcrossing each of the resistant lines with the ST line for at least 10 generations, parts of the genome have been gradually replaced with that of the ST line except for the R-allele responsible for the resistance. Rearing conditions of the adults were as described in Nielsen (1997b), except that instead of 400 ml vials, vials of 158 ml volume were used. Mined leaves in glass boxes measuring 75 x 40 x 40 cm (w x d x h) were picked when the larvae were estimated to be ± 5 days old and transferred to plastic vials (V = 500 ml) with a layer of ± 5 cm moist peat mixed with vermiculite (in volume ratio 5:1). The leaves were removed from the vial, when larvae had left the leaf mines to pupate in the peat. Newly emerging adults from the peat were sexed and kept separate. They had no access to food until they were used for the rearing or used in bioassays.

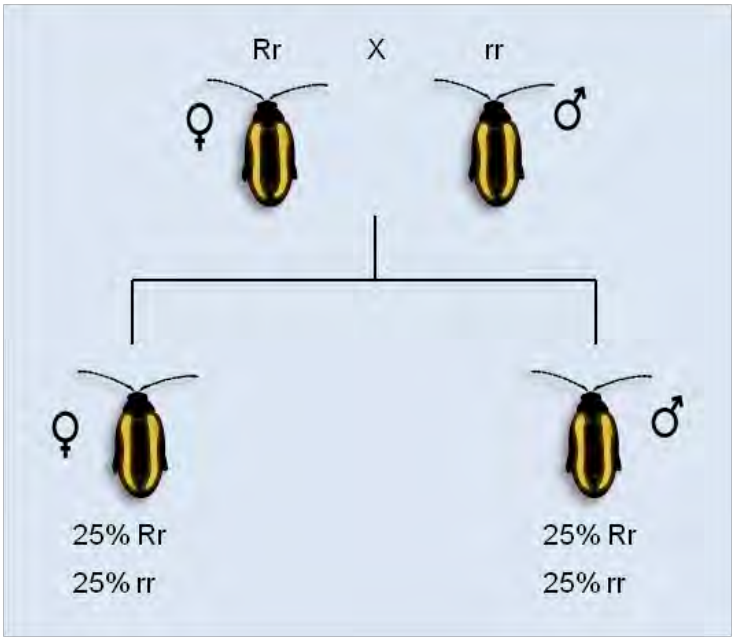


Figure 3 Schematic representation of the cross between an autosomal heterozygous resistant female (Rr) and a susceptible male (rr) to produce a segregating family.

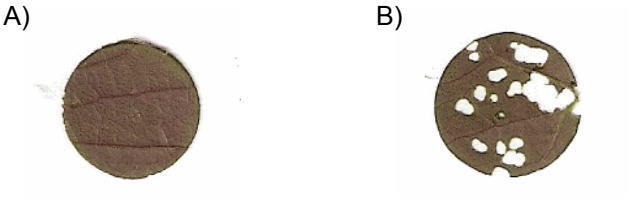


Figure 4 Examples of a leaf disk used in a bioassay when, A) nibbled on by a susceptible beetle, B) fed on by a resistant beetle.

A single segregating family was created by crossing a susceptible male with a heterozygous resistant virgin female (see Figure 3). Both male and female beetle were collected from a rearing which originated from field-collected larvae. We collected these larvae in the summer of 2007 on *B. vulgaris* G-type in Kværkeby, Denmark. The progeny of the single segregating family is expected to consist of resistant beetles and susceptible beetles with the same genetic background. The larval offspring of the cross were reared on radish in a climate room at the Laboratory of Entomology, Wageningen (Netherlands) and kept at 22±2°C and 60% RH in glass containers measuring 75 x 40 x 40 cm (w x d x h). Conditions of the rearing were as above. The F1 of the cross was sexed and kept separate when emerging from the peat as adults. The emerged F1 beetles had no access to food until they were used in bioassays.

Bioassays

No-choice test bioassays were performed by placing newly emerged adults from the segregating family cross in plastic vials (158 ml) with a ±1cm thick moist gypsum/charcoal bottom layer (Nielsen 1978). Two *B. vulgaris* leaf disks (round, diameter: 11 mm) were pinned in the bottom layer. Vials were kept for 72h at 24±2 °C and L18:D6. After three days we judged visually the amount of leaf tissue that the beetles had eaten, to distinguish between feeders and non-feeders. Non-feeders had only nibbled from the leaf disks and were scored as susceptible beetles, feeders had eaten a noticeable amount and were scored as resistant (see Figure 4). All offspring used for further analysis could clearly be categorized this way. Control bioassays were performed with susceptible beetles from the ST-line in order to verify that the *B. vulgaris* leaves were toxic. If these beetles did not eat from *B. vulgaris*, the plant was assumed to be fully defended against susceptible flea beetles.

Chemicals

Hederagenin cellobioside was isolated from *B. vulgaris* G-type as described previously (Shinoda et al. 2002). Hederagenin (aglycone) was purchased from Extrasynthese (France).

cDNA cloning of β-glucosidases

A blastP search was performed (Altschul et al. 1990; Altschul et al. 1997) of translated β-glucosidase sequences in the *Tribolium castaneum* genome to identify conserved amino acid regions using GLEAN (GLEAN 2006; Elsik et al. 2007).

Table 1 List of primers used for cDNA cloning and their sequences.

primer name	sequences (5'-3')	target	construction method	PCR template
gluc-F1	CAGTGGGAYYTNCCNCARGC	initial cDNA creation	PCR, ligation	5'-SMART RACE cDNA
gluc-R1	CCRTTYTCRGTDATRTADAT	initial cDNA creation	PCR, ligation	5'-SMART RACE cDNA
glA-52	GCCTTCACCGAGGACGAAAT	β-glucosidase A	3'-SMART RACE	3'-SMART RACE cDNA
glA-53	GTCCGACTGGACCCAAGGAA	β-glucosidase A	nested 3'-SMART RACE	3'-SMART RACE cDNA
glA-32	GTAGTCCAAAACGCCGTCGGAGAC	β-glucosidase A	5'-SMART RACE	5'-SMART RACE cDNA
glA-33	CGCCGAATACGTCCACTATG	β-glucosidase A	nested 5'-SMART RACE	5'-SMART RACE cDNA
glB-52	GAATGGCCGAGTTCTGCTTC	β-glucosidases B and C	3'-SMART RACE	3'-SMART RACE cDNA
glB-53	CCGTGGGGCATAAGAAAACT	β-glucosidases B and C	nested 3'-SMART RACE	3'-SMART RACE cDNA
glB-32	ACCTTATCTTGGCTATGTGGCTCACC	β-glucosidases B and C	5'-SMART RACE	5'-SMART RACE cDNA
glB-33	TCCACCGTGGCAGGTTTGTT	β-glucosidases B and C	nested 5'-SMART RACE	5'-SMART RACE cDNA
glA54	ACTTAGTTATCATCCAAAATGA	β-glucosidase A	PCR, ligation	cDNA
glA35	CGATTTATTTATTGAATTTATAC	β-glucosidase A	PCR, ligation	cDNA
glB54	CCATTATTATATCATAATTTCA	β-glucosidase B and C	PCR, ligation	cDNA
glB35	AGGCTTTATTTTCTCAATTTAT	β-glucosidase B and C	PCR, ligation	cDNA

Total RNA was extracted from beetles originating from the AT line by the acid guanidinium-phenol-chloroform extraction method (Chomczynski and Sacchi 1987) using RNAwiz™ Reagent (Ambion). cDNAs were synthesized from the total RNA with oligo (dT) primers using Ready-To-Go™ T-Prime First-Strand Beads (GE healthcare) and used as templates in a reverse transcription Polymerase Chain Reaction (RT-PCR) with degenerate primers. Forward primer gluc-F1 and reverse primer gluc-R1 (Table 1) were designed based on HWDLPQA and IYITENG amino acid residues, respectively, both of which are conserved in β-glucosidases belonging to the glycosyl hydrolase family 1. The full-length cDNAs were obtained by combining Rapid Amplification of 5'cDNA Ends (5'-RACE) and 3'-RACE reactions with primers designed based on the RT-PCR cDNA sequences (Table 1) using the SMART™ RACE cDNA Amplification Kit (Clontech) according to the manufacturer's instructions (SMART™ RACE cDNA Amplification Kit User Manual 2007). The RT-PCR was carried out with 35 cycles with an annealing temperature of 50°C. The PCR products were cloned into the pGEM-T vector (Promega) and sequenced by an automatic DNA sequencer (model 377, Applied Biosystems) using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Sequence data were collected and analysed by using Genetyx-Mac software (Software Development Co.).

Comparison of β-glucosidase cDNA sequences among five flea beetle lines

cDNAs were synthesized from total RNAs extracted from adult beetles of the susceptible ST line and the resistant AT, AK, AE and DE lines. A full-open reading frame (ORF) of β-glucosidase A was amplified by PCR using forward primer glA54 and reverse primer glA35 (Table 1) from each cDNA pool. Similarly, full ORFs β-glucosidase B and C were amplified using forward primer glB54 and reverse primer glB35 (Table 1). PCR products were cloned into the pGEM-T vector and sequenced.

Comparison of β-glucosidase cDNA sequences among resistant and susceptible beetles

Total RNA was isolated from each individual of the F1 from the segregating family cross using the TRI Reagent® kit (Applied Biosystems) following the TRI Reagent® Solution protocol and reverse-transcribed. A cDNA pool was synthesized by RT-PCR, carried out under the same conditions as above. The full coding region of the β-glucosidases A and B/C were amplified by PCR, cloned and sequenced as above.

Expression of β-glucosidase proteins

The full ORF of two subtypes of β-glucosidase C cDNAs, namely β-glucosidase C-11 and β-glucosidase C-21, were amplified by RT-PCR from the cDNA pools synthesized from AK beetles and cloned into the pIB/V5-His-DEST expression vector (Invitrogen). The resultant plasmids were called pl-gluC-11 and plB-gluC-21. Silk worm *Bombyx mori* NIAS-Bm-aff3 cells (Invitrogen) were transfected with pl-gluC-11 and plB-gluC-21 plasmids using a transfection reagent FuGENE HD (Roche Diagnostics) as described previously (Kanamori et al.

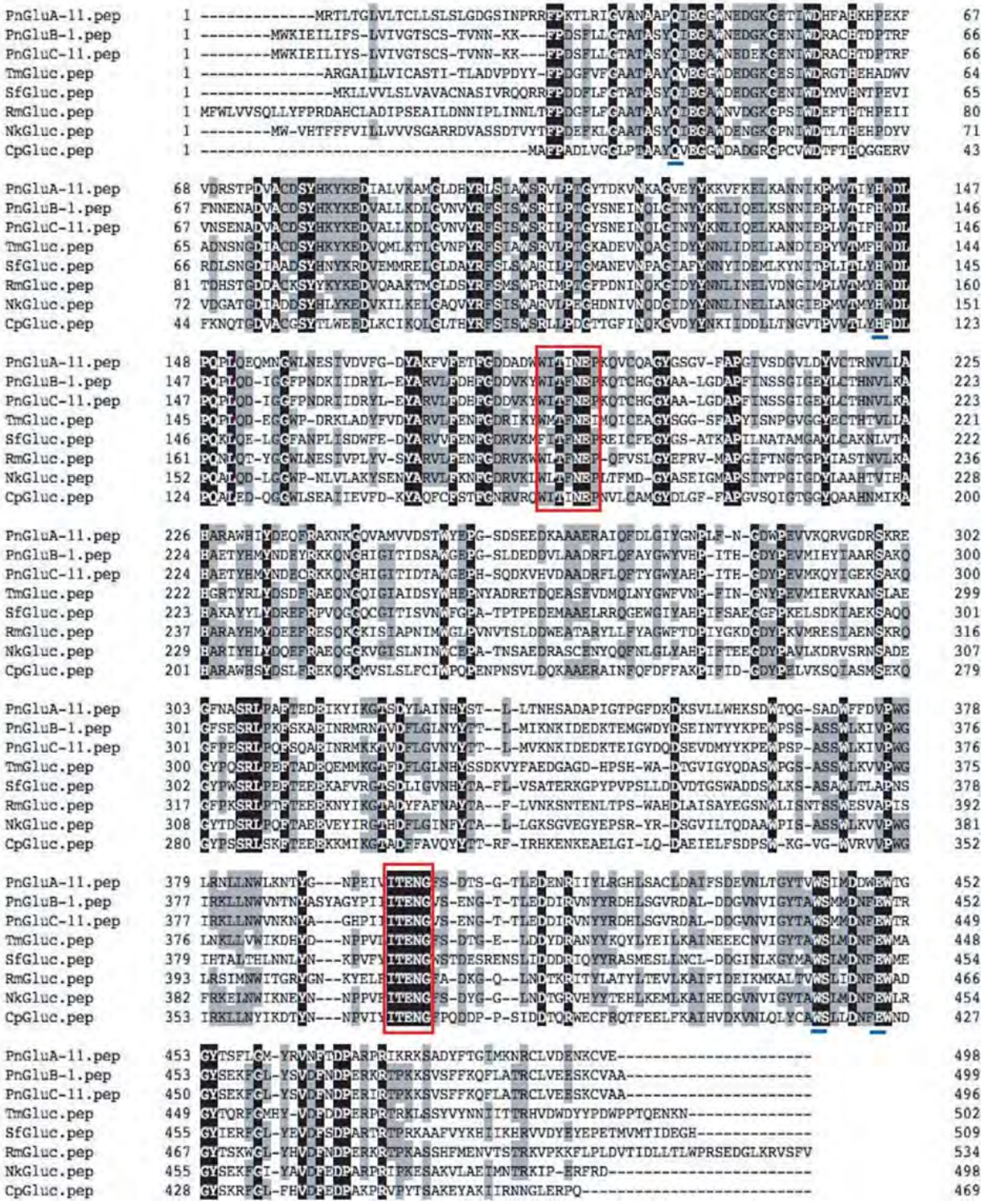


Figure 5 Pairwise amino acid sequence alignment of β-glucosidases of insects and a vertebrate. The red-boxed sequences contain glutamic acid residues (E) responsible for the catalysis (the upstream E is the proton donor and the downstream E is the nucleophile) (Marana et al. 2001). The blue-underlined sequences correspond to amino acid residues involved in substrate binding. Pn: *Phyllotreta nemorum* (large striped flea beetle), Tm: *Tenebrio molitor* (mealworm beetle), Sf: *Spodoptera frugiperda* (army worm), Rm: *Rhyparobia maderae* (madeira cockroach), Nk: *Nasutitermes takasagoensis* (nasute termite), Cp: *Cavia porcellus* (guinea pig).

2010). The conditioned medium into which expressed recombinant β-glucosidase C-11 and β-glucosidase C-21 proteins were secreted, was collected two weeks after the transfection and used for biochemical analysis.

Measurements of enzymatic activities of β-glucosidase C proteins

First, general β-glucosidase activities of β-glucosidase C-11 and β-glucosidase C-21, were tested using 4-methylumbelliferyl β-D-glucopyranoside (4-MUG: Sigma) as a fluorescent substrate. The two solutions were each diluted 25 times in a wide range buffer (Carmody 1961) and incubated for one hour at 25°C with 5mM of 4-MUG. The fluorescence of 4-MUG liberated in the reactions was measured with a 1420 ARVOmx multilabel counter (Perkin Elmer) with excitation at 355 nm and emission at 460 nm after adding 1 M glycine/NaOH buffer (pH 10.6).

Whether β-glucosidase C-11 and β-glucosidase C-21 proteins can hydrolyze the *Barbarea* saponins was tested using purified hederagenin cellobioside (Shinoda et al. 2002) as a substrate. The β-glucosidase C-11 and β-glucosidase C-21 solutions diluted in McIlvaine buffer (pH 5.0) were incubated with 0.5 mM of hederagenin cellobioside at 37°C for 24 h. Then an equal volume of saturated sodium chloride solution was added to each of the solutions and each solution was then twice extracted with 2-butanol. The 2-butanol extracts were mixed per solution, evaporated, and resolved again in methanol. This methanol extract was centrifuged at 3,000 *g* and the supernatant was run on thin layer chromatography (TLC) (solvent system, chloroform : methanol : water = 32 : 9 : 1). Hederagenin cellobioside and its derivative products were detected by spraying 10% sulfuric acid and heating. The molecular mass of the derived product was obtained by running the solution on an HP 1100 series high-performance liquid chromatography (HPLC: Agilent Technologies, Santa Clara, CA, USA) equipped with an HP1100MSD mass spectrometer. An aliquot of the samples was injected into the LC-MS and separated by a C18 reverse-phase column (TSK gel ODS-80Ts, 4,6 x 150 mm, TOSOH) eluted with 50% acetonitrile at 0.3 ml/min at 40°C. The eluent was ionized by negative electrospray ionization.

RESULTS

cDNA cloning of the candidate β-glucosidases

Initially two different sequences were amplified by RT-PCR using a pair of degenerate primers from a cDNA pool synthesized from the resistant AT line. Their full length cDNAs containing the entire ORF were obtained by a combination of 5'RACE and 3'RACE. In the course of the RACE analysis a third sequence was also cloned that was very similar to one of the two originally cloned sequences. As a result, we cloned three full length cDNAs. The deduced amino acid sequences of the three sequences were homologous to those of β-glucosidases belonging to the glycosyl hydrolase family 1 reported from other organisms (Figure 5). We therefore named them *P. nemorum* β-glucosidases A, B, and C (thereafter

abbreviated as β-gluA, β-gluB, and β-gluC). β-gluA cDNA was 1553 bp long excluding the poly A tail, encoding a 498 amino acid protein. β-gluB cDNA was 1563 bp long, encoding a 499 amino acid protein and β-gluC cDNA was 1552 bp long, encoding a 496 amino acid protein. β-gluC was highly homologous to β-gluB (94% identical on the nucleotide level and 93% identical on the amino acid level). In contrast, both β-gluB and β-gluC protein sequences showed only around 50% similarity to the β-gluA protein sequence. The N-terminal ends of the three proteins contained putative signal peptides, suggesting that they are secretory proteins. The two glutamic acid residues responsible for the catalysis in β-glucosidases were conserved in the three protein sequences (Figure 5).

Sequence comparison of the three β-glucosidases among five flea beetle lines

We set out to clone the entire coding regions of β-gluA, β-gluB, and β-gluC cDNAs by RT-PCR from the susceptible ST line and the resistant AK, AE and DE lines (next to the already cloned cDNAs from the resistant AT line) to explore the possibility that any of the β-glucosidases is specifically expressed in resistant beetles. In this experiment, the β-gluB, and β-gluC cDNAs were amplified using a common primer set (gluB54 and gluB35). We cloned in total 63 β-gluA, 42 β-gluB, and 55 β-gluC cDNAs from the five lines. We found four β-gluA variants, two β-gluB variants, and six β-gluC variants (Figure S1). The putative amino acid sequences of each β-glucosidase variant were 97.8 - 99.8% similar.

The β-gluA cDNAs were successfully cloned from all five lines (Table 2). The β-gluB cDNAs were successfully cloned from the ST, AK, AT, and AE lines, but not from the DE line. β-gluC cDNAs were successfully cloned from the AK, AT, AE and DE lines. Thus, β-gluC cDNAs were cloned from all four resistant lines but not from the susceptible line, suggesting that β-gluC is expressed specifically in resistant beetles.

Table 2 Number of β-glucosidase cDNA subtypes cloned from five flea beetle lines. The number right underneath each β-glucosidase is the number given to its subtype (= variant). The number in the table per flea beetle line is the number of clones made of this particular subtype.

flea beetle line	*β-gluA subtype*				*β-gluB subtype*		*β-gluC subtype*					
	1	2	3	4	1	2	1	2	3	4	5	6
ST	9	1	0	3	13	0	0	0	0	0	0	0
AK	1	12	4	1	4	21	3	0	22	13	0	0
AT	16	4	0	0	1	0	3	6	0	0	0	0
DE	1	4	0	0	0	0	0	0	0	0	5	0
AE	0	3	0	4	3	0	0	0	0	0	0	3

Table 3 Number of β-glucosidase cDNA subtypes cloned from individuals of a segregating cross. The number right underneath each β-glucosidase is the number given to its subtype (= variant). The number in the table per individual is the number of clones made of this particular subtype. Male individuals are coded started with 'M' and female individuals are coded with 'F'.

phenotype	individual	*β-gluB subtype*		*β-gluC subtype*					
		1	2	1	2	3	4	5	6
susceptible	M1	0	5	0	0	0	0	0	0
	M2	0	8	0	0	0	0	0	0
	M12	0	1	0	0	0	0	0	0
	M13	0	0	0	0	3	0	0	0
	M14	0	2	0	0	0	0	0	0
	M15	0	9	0	0	1	0	0	0
	M18	0	0	0	0	3	0	0	0
	F1	0	0	0	0	4	0	0	0
	F10	0	0	0	0	5	0	0	0
	F11	0	4	0	0	0	0	0	0
resistant	M10	0	0	0	0	5	0	0	0
	M11	0	0	0	0	6	0	0	0
	M101	0	4	0	0	3	0	0	0
	F1	0	1	0	0	6	0	0	0
	F8	0	10	0	0	0	11	0	0
	F19	0	6	0	0	2	2	0	0

Sequence comparison of the three β-glucosidases between beetles that are resistant and beetles that are susceptible to the defense of *B. vulgaris* G-type

Bioassays

The single segregating cross produced 22 adult males and 22 adult females whose phenotype could be distinguished by bioassays. From this F1, 13 males turned out to be resistant and 14 females turned out to be resistant as well. The other nine males and eight females were susceptible (Table 3). Half of the progeny was expected to be resistant and the other half to be susceptible. Pearson's chi-square test showed no significant deviation from this expectation (X^2 = 0.13, df = 1, P >> 0.05). The ratio male/female (22:22) was also as expected (22:22).

We tested if β-gluC mRNA was expressed only in the resistant offspring from the segregating cross by cloning the β-glucosidases from the resistant beetles and from the susceptible beetles separately. From six resistant individuals, the cDNAs of β-gluA, β-gluB and β-gluC

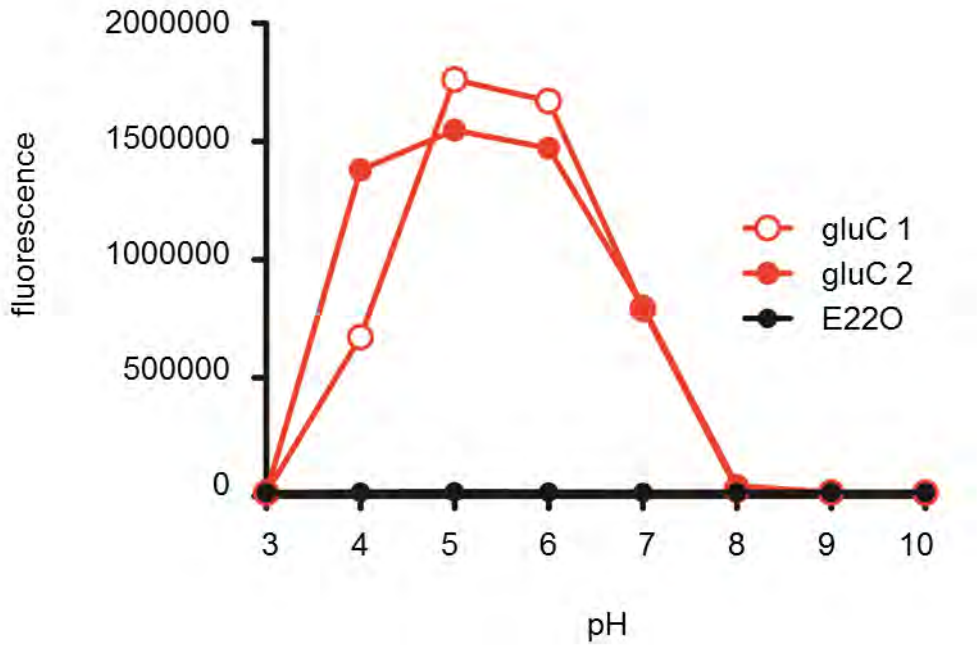


Figure 6 Enzymatic activities of *P. nemorum* flea beetle β-glucosidase C. Two subtypes of β-glucosidase C were expressed by aff3 cells; gluC-1 and gluC-2. E220 is another secretory enzyme but without β-glucosidase activity and was used as a negative control. The conditioned media of each cell line containing secreted β-glucosidases were incubated with a fluorescent substrate 4-methylumbelliferyl β-D-glucopyranoside under pH's varying from 3 to 10 for an hour. The Y axis shows the amount of fluorescence exhibited by the produced 4-methylumbelliferone.

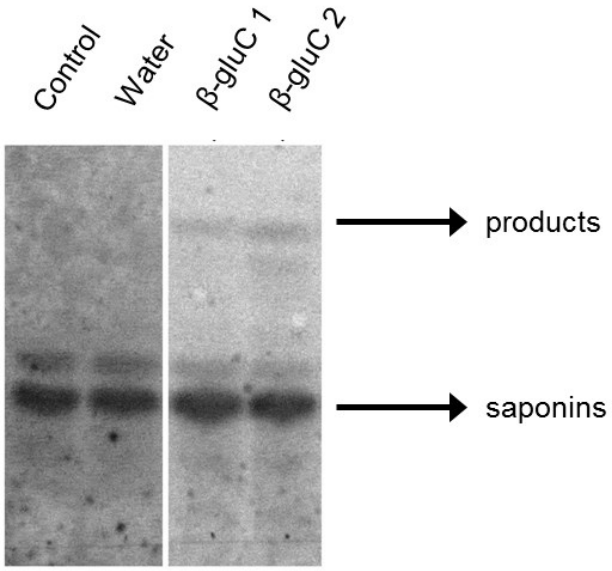


Figure 7 Western blot analysis of the saponin hederagenin cellobioside and possible breakdown products after hydrolysis by recombinant β-glucosidase proteins expressed in *Bombyx mori* NIAS-Bm-aff3 cell lines. The control medium contained no protein expressing cells. β-gluB-2 stands for the cell line expressing β-gluB-2 proteins, β-gluB-1 for the cell line expressing β-gluB-1 proteins, β-gluC-11 for the cell line expressing β-gluC-11 proteins, and β-gluC-21 for the cell line expressing β-gluC-21 proteins.

were cloned. Where we would have expected to only clone the β-gluA and β-gluB cDNAs from the susceptible individuals, β-gluC cDNA was also cloned from 5 of 10 susceptible beetles. We compared the ratios of β-gluC clones divided by the total of β-gluB and β-gluC clones (β-gluC/B+C ratio) between resistant and susceptible individuals. The β-gluC/B+C ratio of the 6 resistant beetles equalled 62.5% (35/56) whereas that of the 10 susceptible beetles equalled 35.6% (16/45). These ratios differed significantly from the expected ratios (Pearson's chi-square test, $X^2 = 7.25$, df = 1, $P = 0.0071$). The number of expressed β-gluC clones is therefore correlated with the ability of a flea beetle to feed and survive on *B. vulgaris* G-type, though expression of β-gluC is not restricted to resistant beetles.

Hydrolysis of a Barbarea vulgaris *saponin by β-glucosidase C proteins*

The recombinant protein expressed in each of the two aff3 cell lines β-gluC-1 and β-gluC-2, hydrolyzed 4-methylumbelliferyl β-D-glucopyranoside (4-MUG) (Figure 6). The recombinant proteins of β-gluC showed the highest activity to 4-MUG at pH 5.0 (Figure 6). These results indicate that the proteins of β-gluC have a β-glucosidase activity.

We then tested whether the recombinant β-gluC-1 and β-gluC-2 proteins can hydrolyze the *Barbarea* anti-feeding saponin hederagenin cellobioside. Incubation of hederagenin cellobioside with each of the two aff3 cell lines generated a substance with a lower mobility in the TLC analysis (Figure 7). LC-MS analysis showed that the molecular weight of this substance was 634.8, identical to that of hederagenin monoglucoside. These results indicate that both β-gluB and β-gluC can cut the β 1-4 bond within the cellobiose moiety of hederagenin cellobioside to generate hederagenin monoglucoside and glucose.

DISCUSSION

We cloned the entire coding regions of three β-glucosidase cDNAs, respectively β-gluA, β-gluB, and β-gluC by RT-PCR from susceptible as well as resistant *P. nemorum* beetles. We have shown that these genes were expressed both in beetles that are resistant and in beetles that are susceptible to the defence of *B. vulgaris* G-type, but that significantly fewer different β-gluC cDNAs were found for susceptible beetles than for resistant beetles. In addition, our results indicate that recombinant β-gluC proteins could hydrolyze the most active *B. vulgaris* G-type saponin hederagenin cellobioside into hederagenin monoglycoside and glucose. The results show that the β-gluC gene is not the so-called "R-gene" itself, because its mRNA is expressed in both susceptible and resistant beetles. While we have no experimental evidence concerning the functions of the genes that enable flea beetles to feed on defended *B. vulgaris*, these genes may have regulatory functions on the expression levels of these β-glucosidase genes and thereby influence the ability of the flea beetle to use *B. vulgaris* G-type as a food source. Therefore looking at the expression levels of subtypes of β-gluA, β-gluB, and β-gluC in both resistant and susceptible flea beetles before and after feeding on *B. vulgaris* G-type, could further elucidate the relation of these genes to flea beetle's resistance.

Nielsen et al. (2010a) showed that hederagenin cellobioside deterred feeding of susceptible *P. nemorum* beetles on radish, a host plant that is used by both susceptible and resistant flea beetles, while its aglycone did not prevent flea beetles from feeding. The range of hederagenin cellobioside concentrations used in the study was well within the range of the concentrations found in *B. vulgaris* by Shinoda et al. (2002), suggesting that hederagenin cellobioside could very well be the active compound deterring feeding of susceptible beetles on *B. vulgaris* G-type. The effect of hederagenin-monoglycoside on the consumption rate of both flea beetle genotypes corresponds with what was found for hederagenin cellobioside (Augustin et al. 2012). Accordingly, the ability of β-gluC to hydrolyze hederagenin cellobioside into hederagenin-monoglycoside does not make the saponin biologically inactive against the flea beetle. At present, we do not know if the enzyme that β-gluC is coding for, can remove only one sugar unit or additionally remove the other sugar unit as well. It might be one of more steps in degrading the saponin to the resistant compound. Therefore, future research on the capability of the enzymes of β-gluC – but also β-gluA and β-gluB – to hydrolyze hederagenin monoglycoside will be of much value. If β-gluC only targets the bond between the first and second glucose unit at the C3 position of the saponin, then an additional enzyme should be looked for.

Another future direction could involve primers developed for these β-glucosidase cDNAs. By testing more beetles from various sample sites, we could look for a consistent difference between susceptible and resistant beetles and subtypes of these β-glucosidases. We can also test if there is any selection pressure on these β-glucosidase genes in flea beetles found on *B. vulgaris* G-type by looking at the heterozygosity of the alleles compared to heterozygosity in neutral parts of the genome. If an excess of homozygosity is found for a β-glucosidase gene then this gene is likely to be under directional selection. If the β-glucosidase genes are involved in the resistance of the flea beetle, we would expect selection pressure on them in individuals found on *B. vulgaris* G-type.

All together, these results suggest that β-gluC might play a role in inactivating the saponins that *B. vulgaris* G-type uses as defence, by hydrolysing at least one of the sugars from a saponin. Saponins are secondary metabolites of plants and well known to act as defence chemicals against herbivores and pathogenic microorganisms (Yamane et al. 2010). Although it has been reported that some fungi detoxify host plant saponins by secreting special saponin-degrading enzymes (Osbourn et al. 1996; Morrissey et al. 2000; Morant et al. 2008), there has been no report for such enzymes in insects. As far as we know, the *P. nemorum* β-glucosidase C found in this study is the first insect enzyme that can deglycosylate saponins.

CHAPTER 4

Changes in frequencies of genes that enable *Phyllotreta nemorum* to utilize its host plant, *Barbarea vulgaris,* vary in magnitude and direction as much within as between seasons

Kim M. C. A. Vermeer • Patrick Verbaarschot • Peter W. de Jong

Modified version published in: Entomologia Experimentalis et Applicata (2012) 144: 37-44

Abstract

The interaction between the flea beetle, *Phyllotreta nemorum* L. (Coleoptera: Chrysomelidae), and its host plants is well suited to study the dynamics of a geographic mosaic of (co)evolution. The flea beetle can either be resistant or susceptible to the defense of one of its host plants, *Barbarea vulgaris* R.Br. G-type (Brassicaceae). Previous findings suggested that the frequency of resistant beetles on host plants other than the G-type of *B. vulgaris* had decreased over time within the period of 1999–2003. In 2008 and 2009 new sampling was performed to investigate whether or not this decrease in frequency of resistance of the flea beetles formed a continuing trend and whether or not the frequency of resistant beetles also varies within the year. The frequency of resistant beetles on different host plants was determined during the reproductive season of the flea beetles in both years. Overall, the frequencies of resistant beetles on *B. vulgaris* (G-type) remained close to 100%, as found before, but those on other host plants did not consistently decrease across the years, in contrast to what had been suggested. Furthermore, the repeated sampling revealed that the frequency of resistant beetles differed significantly within a season. The present data show that relative frequencies of different resistance phenotypes of *P. nemorum* on other host plants than *B. vulgaris* (G-type) are highly dynamic, both within and across years. Therefore, monitoring the changes in these resistance frequencies should involve season-wide sampling efforts. Although the monitoring in this study does not provide an explanation for the observed dynamics, we propose a testable scenario.

INTRODUCTION

Evolutionary processes have led to the present-day insect-plant interactions (Wheat et al. 2007). A possible evolutionary trajectory of a plant-insect interaction might look as follows: (1) phytophagous insects attack the plant, (2) the plant evolves a defence against the insects, and (3) an insect evolves resistance against the plant's defence. This type of interaction, a (co)evolutionary arms race between plants and insects, has been demonstrated in classical, as well as contemporary examples of interspecific evolutionary interactions (Ehrlich and Raven 1964; Wheat et al. 2007; Toju et al. 2011). If the insect species in step (3) of this trajectory was among those inducing the evolution of plant defence, one can speak of coevolution. Coevolution sensu stricto in a plant-insect interaction is reciprocal adaptation between plant and insect driven by natural selection (Thompson 2005). When additional species are involved in the interspecific interactions, then it is questionable if the natural selection exerted on each of the interacting species is reciprocal. Whereas earlier plant-insect coevolution was only invoked when the reciprocal evolutionary change between a plant and its enemy occurs at the whole species level (Strong et al. 1984), now we know that it is rare that geographical ranges of interacting species completely overlap, that environmental conditions differ within the range of a species, that populations are not panmictic, and that the interactions between an insect and its host plant can vary with time (CHAPTER 2). According to Thompson's influential theory about the geographic mosaic of coevolution, strict coevolution is only occurring in so-called 'coevolutionary hot spots'. The geographic mosaic theory of coevolution argues that beyond local coevolution there are broader dynamics (Thompson 2005; Nuismer et al. 2010; Thompson et al. 2010; CHAPTER 2). These broader dynamics have three components: geographical selection mosaics, coevolutionary hot spots, and trait remixing (Thompson 2005; Thompson et al. 2010; CHAPTER 2). Local adaptation is a crucial part of the geographic selection mosaics; within a metapopulation, local adaption leads to a shifting geographic mosaic of traits (Thompson 2005). Local adaptation and its influence on geographic selection mosaics needs further study.

A good example of a geographic mosaic of adaptation is that of the flea beetle *Phyllotreta nemorum* L. (Coleoptera: Chrysomelidae) and one of its host plants, *Barbarea vulgaris* ssp. *arcuata* (Opiz.) Simkovics (Brassicaceae) (Nielsen 1997b; de Jong et al. 2000; Nielsen and de Jong 2005). Both flea beetles and host plants have a patchy distribution in Denmark (de Jong et al. 2001, de Jong et al. 2009). In Denmark, the biannual *B. vulgaris* consists of two types (Nielsen 1997b; Toneatto et al. 2010): a pubescent and fully susceptible type (P-type) and a glabrous and partly resistant type (G-type). The P-type is the rarer one, and is suitable for all *P. nemorum* beetles throughout the whole year. The more common G-type develops chemical resistance to herbivory in the beginning of spring, when exposed to intense sunlight at ambient temperatures exceeding 15-20 °C (JK Nielsen pers. obs.), whereas in October the G-type gradually loses its resistance again (Nielsen 1997b; Agerbirk et al. 2001). Both *Barbarea* types have a patchy distribution and patches usually consist of a homogeneous group of one type of plants, although a limited amount of hybridization between the types has been observed (Toneatto et al. 2010). Most flea beetles are not able to survive on the defended G-type, but some beetles are resistant to its defence, enabling them to utilize

the plant (Nielsen 1997a, Nielsen 1997b, Nielsen 1999; de Jong and Nielsen 1999; de Jong et al. 2000). The resistance involves one or more major dominant gene(s), so-called R-genes. Beetles possessing at least one R-allele (Rr or RR) can survive and fully develop on the G-type. Beetles that lack the resistance alleles do not feed on the G-type (Nielsen 1996), and therefore cannot survive on this plant. Various modes of inheritance of resistance have been reported (Nielsen 1997a): sex-linked and autosomal. Because resistant beetles are relatively rarely found, including on *B. vulgaris* (there are many patches of *B. vulgaris* in East Denmark that have not been colonized by *P. nemorum* at all), the hypothesis that the use of *B. vulgaris* represents a host-range expansion has been proposed (Nielsen 1997b). Differences in the flea beetle's counter-defence and the occurrence of *P. nemorum* on some patches of *B. vulgaris* G-type likely leads to geographic differences in the interaction between the flea beetle and its host plant *B. vulgaris*, among other things depending on the local availability of alternative host plants. The annual plants *Sinapis arvensis* L. and *Raphanus raphanistrum* L. are most widely used as a host plant by *P. nemorum*. The flea beetle also uses *Lepidium draba* L. ssp. *draba* (formerly known as *Cardaria draba*), *Brassica nigra* L., *Brassica rapa* L. ssp. *Rapa*, and cultivated radish (*Raphanus sativus* L.) as host plants in Denmark (Nielsen 1977).

The precise details of the origin and evolution of the interactions between *P. nemorum* and *B. vulgaris* are currently under investigation. It is not clear whether the different resistance phenotypes of *B. vulgaris* and *P. nemorum* have evolved as a result of direct coevolution sensu stricto. Resistance of *B. vulgaris* may have evolved in response to other selection pressures than the use as host plant by *P. nemorum*, and indeed evidence exists that the polymorphism in chemical resistance of *B. vulgaris* is the result of allopatric processes during the ice ages, followed by secondary contact of the differently evolved types (Toneatto et al. 2010). Where glucosinolates have functioned as a general first line of defence in *B. vulgaris* (Renwick 2002; Kuzina et al. 2009), a second line of defence has evolved in *B. vulgaris* G-type in the form of saponins (Shinoda et al. 2002; Agerbirk et al. 2003a). These saponins have been proven to be responsible for the defence against susceptible flea beetles (Nielsen et al. 2010a). Even though it is not known what kind of selective pressure caused the evolution of defences by saponins in *B. vulgaris*, the flea beetle is likely to exert at least some selection on *B. vulgaris*. In Ejby (sample site in Nielsen and de Jong 2005), for example, the presence of resistant flea beetles seems to have affected the distribution and abundance of both the P- and G-types of *B. vulgaris* at this locality (Nielsen 1997b). Therefore, it is possible that currently there is a coevolutionary interaction (reciprocal evolutionary adaptation) between the two interacting species, with selection acting on the saponin defence in *B. vulgaris* and the resistance to this defence in *P. nemorum*, respectively.

Nielsen and de Jong (2005) investigated the geographical and temporal variation in frequency of resistance in the flea beetles. They found a significant difference in the frequency of resistant beetles among host plants. These beetles were high in all populations living on *B. vulgaris* whereas they were evidently lower in populations living on other host plants. In five out of six local patches of *S. arvensis*, they found a decrease of resistant beetles over years.

This decrease could be explained by selection against the resistance allele on other host plants than *B. vulgaris*, but as no decrease of resistant beetles was found on the other host plants besides *S. arvensis*, it could also be that there is no general trend of a decrease in R-genes found in populations on other host plants than *B. vulgaris*. The decrease found might also be a result of different sampling times within the reproductive season of the beetle, if the frequencies of resistant beetles change within a season at a sampling site.

The main objective of this study was to investigate whether the decrease in frequency of resistance of the flea beetles through the years formed a continuing trend and whether there is spatio-temporal variation that can account for the earlier results in Nielsen and de Jong (2005). The following questions were addressed: 1) is the pattern of variation in the frequencies of resistant beetles on different host plants found in Nielsen and de Jong (2005) still present? 2) is the decrease found in the frequency of resistant beetles on *S. arvensis* a continuing trend? and 3) is there temporal variation in the frequencies of resistant beetles in populations within one reproductive season?

For the present study, new sampling was done in 2008 and 2009. We collected the samples at the same locations in eastern Denmark as in Nielsen and de Jong (2005) and two additional locations in western Denmark, near Try (Figure 1). Apart from the trend over years, this time we also monitored within the reproductive season; if the frequency of resistant beetles varies significantly during the season, then timing of sampling may at least partly explain the decrease of resistant beetles across years that was found by Nielsen and de Jong (2005).

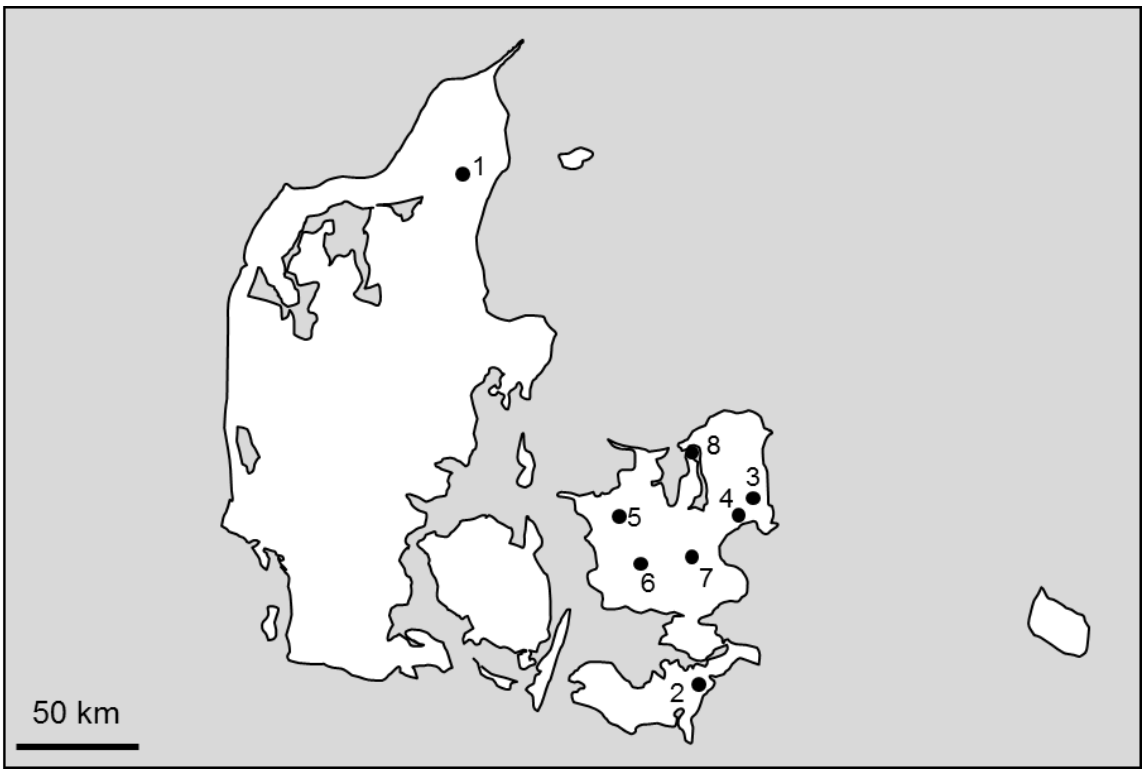


Figure 1 Map of *Phyllotreta nemorum* sample locations in Denmark: (1) Try, (2) Maglebrænde, (3) Taastrup, (4) Ejby, (5) Sveboelle, (6) Lynæs, (7) Suserup, and (8) Kværkeby.

Table 1 Temporal variation in the frequencies of R-genes in *Phyllotreta nemorum* populations living on different host plants. n, number of beetles; %, percentage of resistant beetles, as established via the feeding assay. *p*-values indicate differences in frequency of resistant beetles between sampling years. Sampling took place in June and early July in 1999, 2003, 2008, and 2009. Coordinates of the localities are given with WGS 84 (World Geodetic System 1984) as the reference coordinate system.

				1999		2003		2008		2009			
Locality	Site	Host plant	Coordinates	%	n	%	n	%	n	%	n	G	P
Kværkeby	1	G-type of *Barbarea vulgaris* ssp. *arcuata*	55°27'40.82", 11°55'0.08"	99.0	200	99.0	99	100.0	125	100.0	113	3.5	ns
Kværkeby	2	G-type of *B. vulgaris* ssp. *arcuata*	55°28'8.50", 11°53'23.09"	99.0	100	95.2	21	99.0	104	99.0	104	1.4	ns
Kværkeby	4	*Sinapis arvensis*	55°28'47.24", 11°53'2.90"	71.4	35	42.6	54	76.2	105	62.2	156	18.5	<0.0005
Kværkeby	5	*S. arvensis*	55°28'40.92", 11°51'28.16"	60.0	100	41.7	100	79.2	56	-	-	19.0	<0.0001
Ejby	1	G-type of *B. vulgaris* ssp. *arcuata*	55°42'27.48", 12°25'8.21"	98.5	199	100.0	51	100.0	119	-	-	3.7	ns
Ejby	1	*S. arvensis*	55°42'22.10", 12°24'59.57"	96.0	101	32.7	101	66.3	98	-	-	104.3	<0.0001
Sveboelle	-	G-type of *B. vulgaris* ssp. *arcuata*	55°39'35.83", 11°19'38.01"	98.9	87	100.0	7	98.2	56	-	-	0.3	ns
Suserup	-	*S. arvensis*	55°23'11.13", 11°32'27.17"	43.0	100	23.9	92	47.0	100	44.1	177	14.2	<0.005
Magle-brænde	-	*S. arvensis*	54°50'54.32", 12°1'55.89"	48.0	100	21.7	23	51.7	120	-	-	7.4	<0.05
Taastrup	-	*Raphanus sativus*	55°40'4.44", 12°18'22.24"	20.0	100	24.2	62	14.3	7	-	-	0.6	ns
Lynæs	-	*Cardaria draba*	55°56'31.43", 11°53'11.36"	15.5	97	21.4	28	19.3	31	-	-	0.6	ns
Try	1	G-type of *B. vulgaris* ssp. *arcuata*	57°6'46.41", 10°14'50.90"	-	-	-	-	99.1	109	99.0	102	0.0	ns
Try	2	P-type of *B. vulgaris* ssp. *arcuata*	57°7'26.05", 10°12'6.37"	-	-	-	-	74.7	79	87.9	140	6.0	<0.05

MATERIALS AND METHODS

Locations and time of sampling

Through June 2008 we collected leaves with leaf-mining flea beetle larvae from various locations (Figure 1 and Table 1) and various host plants (Table 1) in Denmark. All these sites were the same as those sampled in Nielsen and de Jong (2005) except for the sites sampled in Try. In 2009 we collected leaves with leaf-mining flea beetle larvae from some of the same locations as in 2008 (Table 1). All sampling in 2009 took place in June and early July. For comparison, the time of sampling in 1999 and 2003 in Nielsen and de Jong (2005; unpubl. data) was also in June and early July. The sampling of some locations in 2009 was done on different dates within the season, namely 11 and 26 June for Kværkeby on *S. arvensis*, and 4 June and 1 July for Suserup, also on *S. arvensis*.

Insects

We picked leaves containing flea beetle larvae from various plant individuals in each patch at each location to avoid bias from family structuring across plants. These leaves were transferred to 500 ml plastic vials that contained a layer of moist peat and vermiculite (in volume ratio 5:1). Most picked leaves contained 1-3 larvae, but some leaves (especially *S. arvensis*) occasionally contained more larvae. The vials were closed using a plastic lid with a central hole (diameter 1.5 cm), plugged with cotton wool. The vials were kept at 20 ± 2 °C. Larvae pupated in the vermiculite/peat. After 2.5 weeks the adult beetles started to emerge from the vermiculite and were captured within 24 h of emerging. They had no access to food. Beetles were sexed under a stereo microscope (25× magnification), looking at the ventral abdominal characters and antennal segments (see de Jong and Nielsen 1999). Sexes were kept separate and used immediately in bioassays (see below).

Bioassays

We used *B. vulgaris* G-type for the bioassays. The plants were maintained under 400-W HPI/T-lamps (Philips, Eindhoven, the Netherlands) supplying 160-200 µmol quanta m^{-2} s^{-1} on the level of the leaf surface, at 30 ± 6 °C and a L18:D6 photoperiod to ensure they were 100% defended against susceptible beetles (when kept under low light conditions, *B. vulgaris* can be used by all flea beetle genotypes; JK Nielsen, PW de Jong and KMCA Vermeer pers. obs.). *Barbarea vulgaris* plants were 3-8 weeks old and still in the vegetative stage when used in bioassays. To check whether the *B. vulgaris* G-type was fully defended against susceptible beetles, we exposed these plants to beetles of which the phenotype was already known to be susceptible in bioassays [ST-line, originating from beetles collected on radish in Taastrup (Nielsen 1999)]. If these beetles did not eat from the G-type, the plant was assumed to be fully defended against susceptible flea beetles.

Plastic vials (158 ml) with a moist charcoal-gypsum bottom layer (ca. 1 cm thick) were used for the bioassays. These vials were closed with a plastic lid containing a 15 mm hole closed

Figure 2 Examples of a *Barbarea vulgaris* leaf disk when A) nibbled on by a susceptible flea beetle, or B) fed on by a resistant flea beetle.

with cotton wool. The beetles were tested for their ability to feed on *B. vulgaris* by pinning two round *B. vulgaris* leaf discs (14 mm diameter) in the vial. The bio-assays were kept at 24 ± 2 °C and L18:D6 photoperiod for 72 h. The amount the beetles had eaten after 72 h was visually estimated, to distinguish between 'feeders' and 'non-feeders'. Non-feeders had only nibbled from the leaf discs, where feeders had eaten a noticeable amount of leaf surface (see Figure 2). More than 98% of the beetles could thus clearly be categorized as feeders or non-feeders. Beetles that we could not categorize this way were excluded from the study. Non-feeders were classified as susceptible beetles, feeding beetles were classified as resistant beetles (Nielsen and de Jong 2005).

Statistical analysis of differences in frequencies of resistant beetles collected in 1999, 2003, 2008, and 2009 was carried out with the likelihood-ratio test (G-test), which was also used by Nielsen and de Jong (2005). The sampling data of Kværkeby and Suserup for the various dates in 2009 were combined per patch for this analysis. The G-test was also used to examine whether the time of sampling within the season of 2009 had any effect on the frequency of resistant beetles for Kværkeby and Suserup.

RESULTS

The frequencies of resistant beetles found on different host plants differed substantially (Table 1) as had been found previously (Nielsen and de Jong 2005); in all sampled populations living on *B. vulgaris* G-type the frequencies of feeders were near 100%, whereas in populations on other host plants the frequencies were clearly lower.

When considering the data of all sampling years for each location, no consistent pattern could be discovered in percentage resistant beetles on alternative host plants or on *B. vulgaris* across years: not only apparent decreases with time, but also apparent increases were observed. We observed significant temporal differences in the frequency of feeders for all locations on *S. arvensis* over the sampled years (Table 1). Unlike in Nielsen and de Jong (2005), during this study the frequencies of feeders were not declining in populations on *S. arvensis* compared to the frequencies found earlier.

The frequencies of resistant beetles collected on various dates through the season in 2009 also differ significantly (Figure 3). We used the same test as before, the likelihood-ratio test (Kværkeby: $G = 14.64$, $P<0.0005$; Suserup: $G = 4.49$, $P<0.05$). For the two locations sampled on different dates in 2009 the frequency of feeders increased through the season.

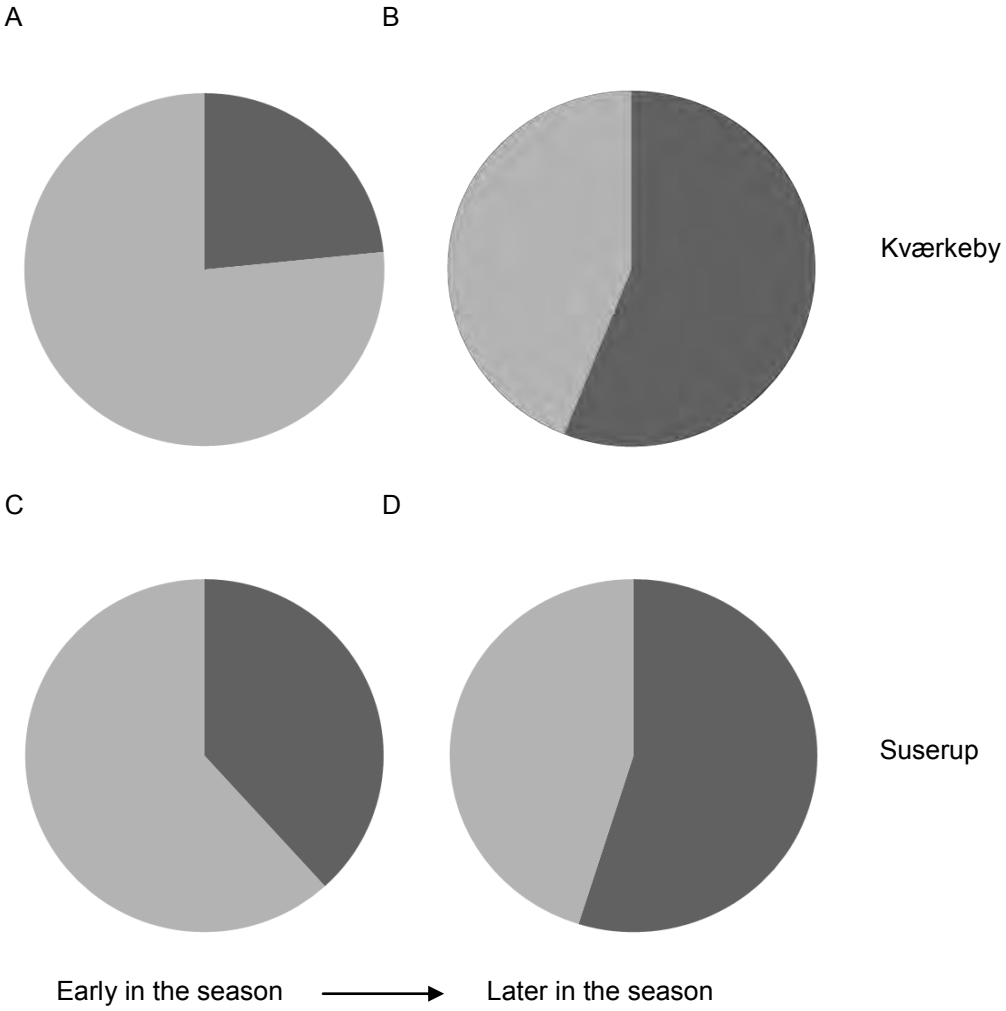


Figure 3 Relative frequencies of resistant and susceptible beetles in *Phyllotreta nemorum* populations for different sampling periods. Dark grey sections indicate feeders (resistant to *Barbarea vulgaris* G-type); light grey sections represent non-feeders. A) Kværkeby on 11 June 2009, B) Kværkeby on 26 June 2009, C) Suserup on 4 June 2009, and D) Suserup on 1 July 2009.

DISCUSSION

Our results clearly show that a) the percentage of resistant *P. nemorum* is consistently high on *B. vulgaris* at different locations and in different years; b) the percentage of resistant *P. nemorum* is generally lower on other plant species than on *B. vulgaris*, and varies between locations, years, and also within one season; c) there is no consistent pattern in the annual change in percentage of resistant *P. nemorum* on other plants than *B. vulgaris*, i.e., the decrease in the percentage resistance found at five out of six locations living on *S. arvensis* between 1999 and 2003 (Nielsen and de Jong 2005) was not part of a continuous trend; d) within the two locations where it was measured, the percentage of resistant flea beetles in a sample late in the season was significantly higher than early in the season. This spatial and temporal variation in allele frequencies at resistance loci in the flea beetles suggests a highly

dynamic geographic mosaic of interactions between flea beetles and their host plants. The within-seasonal variation in resistance allele frequencies shows that both within- and between seasonal sampling is required to study the dynamics of this mosaic, as apparent inter-annual trends in resistance allele frequencies are at least partly influenced by variation in phenology and/or timing of sampling between years. We cannot conclude that difference of sampling time within the season between the years 1999 and 2003 explains the earlier found decrease in frequencies of resistant beetles on *S. arvensis* (Nielsen and de Jong 2005). Despite knowing the exact sampling dates of each locality, it is not known whether sampling occurred relatively earlier or later in the reproductive season of the flea beetle compared to the previous sampling, because the start and length of the reproductive season of the flea beetle can differ per year depending on environmental conditions such as temperature.

Factors influencing (changes in) allele frequencies at adaptive loci include selection, migration, and genetic drift (Stephens 2010; Yeaman and Otto 2011). In the flea beetle system, current evidence suggests that all three of these play a role in determining the spatial and temporal patterns in the frequency of resistance phenotypes (de Jong et al. 2001; Breuker et al. 2005; Nielsen and de Jong 2005; this article). One factor that is likely to influence the temporal and spatial variation in frequency of resistance is migration. The dispersive ability of the flea beetle has been studied using molecular markers (de Jong et al. 2001; Breuker et al. 2005). Some genetic structure has been observed in flea beetle populations in East Denmark. Two factors have been found that correlate significantly with this structure: geographical distance and resistance genotype (de Jong et al. 2009). Thus, dispersive ability appears to be limited, at least partly (hence the correlation with geographical distance), and as yet unknown mechanisms have led to some genetic isolation between the different resistance phenotypes. The latter may, among other things, be caused by phenological effects as described below. In any case, because the genetic exchange between local populations and different resistance genotypes is limited, local frequencies of resistant flea beetles are also likely to be influenced by local predominance of host plants. The genetic structure of the flea beetle populations, combined with the observation that some patches of potential host plants are not attacked by flea beetles, suggests that the flea beetles have a metapopulation-like structure (Hanski 1998; Hanski et al. 2011), where founder effects (i.e., random drift) may also be responsible for local variation in the relative abundance of different resistance phenotypes. For example, just as recorded by Nielsen and de Jong (2005), the flea beetle population living on *B. vulgaris* G-type in Sveboelle was not well established, as in 2008 few larvae could be sampled again. The populations living on *B. vulgaris* G-type in Kværkeby, Ejby, and Try, on the other hand, were well established. At a distance of ca. 1 km from the *S. arvensis* site in Suserup, there was a *B. vulgaris* G-type patch but this patch was not colonized by flea beetles at all. Also in Sveboelle and Maglebrænde there was another *B. vulgaris* G-type patch closeby (ca. 250 m) but also here no *P. nemorum* were found at all in the G-type patch.

Sinapis arvensis is a fugitive plant that occurs in disturbed habitats, so that local availability of different host plant species may vary, local flea beetle populations may go extinct, recolo-

nization may take place, all leading to dynamic fluctuations and variation in local frequencies of resistance phenotypes of the flea beetles.

A third factor that undoubtedly influences local frequencies of flea beetle resistance phenotypes is selection. It has clearly been demonstrated that there is strong selection favouring resistance in the flea beetles that use *B. vulgaris* in late spring (when *B. vulgaris* has become fully toxic) (Nielsen 1997b). However, selection against resistance has also been suggested (de Jong and Nielsen 2002). This may explain the relatively low frequency of resistant flea beetles on other plants than *B. vulgaris*: when resistant migrants from populations of flea beetles on *B. vulgaris* G-type arrive on other host plants, where they have to compete with susceptible beetles, they may be at a selective disadvantage. A direct selective disadvantage has so far not been found (Nielsen 1999), but evidence has been found for a selective disadvantage of outcrossed, homozygous resistant flea beetles (de Jong and Nielsen 2000). This may slow down the geographical spread of the resistant flea beetle phenotype. Besides selective pressures that are associated with the genotype at the resistance locus, selection may also operate through trophic interactions. If, for example, the nutritional value of the *B. vulgaris* G-type is different from that of other host plants, this may influence the selective advantage of using this plant as food, and thereby the frequency of resistant phenotypes. So far, no such effects of host plant nutritional quality on the fitness of *P. nemorum* have been demonstrated. In contrast, there is evidence for the effect of another trophic interaction on the fitness of flea beetles: flea beetles using *B. vulgaris* G-type had a lower chance to be parasitized by parasitoid wasps than their conspecifics using *S. arvensis* as host (Kerstes and de Jong 2011). This advantage of using *B. vulgaris* G-type as host plant by avoiding parasitisation ('enemy-free space') appeared to involve a phenological component: flea beetles using this host plant, which occurs earlier in the season than alternative host plants, escaped the peak of parasitism that was found on the alternative plants (Kerstes and de Jong 2011).

Also phenological effects of both flea beetle development and that of its host plants may at least partially explain the observed intra-seasonal increase in percentage of resistant flea beetles on other host plants than *B. vulgaris* G-type. *Barbarea vulgaris* is a bi-annual plant that overwinters as a rosette and flowers in the subsequent spring. It is, therefore, already present very early in spring. Annual host plants, such as *S. arvensis*, of which the seeds first have to germinate (Nielsen 1997b), are present several weeks later in the year than *B. vulgaris* (e.g., Kerstes and de Jong 2011). *Phyllotreta nemorum* overwinters as adult beetles in the soil and in remaining stems of plants, and generally has one generation per year in Denmark (Nielsen 1997b). Somewhere during April or early May, when the weather conditions allow the flea beetles to terminate overwintering, the flea beetles emerge and start feeding on plants (Nielsen 1997b). At that time, *B. vulgaris* rosettes are present that can be used as food by the flea beetles when other plants such as *S. arvensis* are not present yet. The beetles start oviposition a few days after feeding (Nielsen 1989). Larval development lasts 20 days in the field (Nielsen 1977). The larvae pupate in the soil and new adults appear in July and August. Depending on the precise weather conditions, and thereby the development of chemical defence in the G-type of *B. vulgaris*, these plants may initially be attacked by sus-

ceptible, as well as resistant genotypes of the flea beetles. As the toxicity of these plants increases, they become less suitable, and eventually completely unsuitable for the susceptible flea beetles. During May the plants can already become completely unsuitable for susceptible beetles. If alternative, non-toxic host plants are nearby, the susceptible beetles may disperse onto these, whereas the resistant flea beetles may continue development on the *B. vulgaris* plants. Because of the earlier herbivory on the *B. vulgaris* plants, these plants deteriorate faster than the alternative host plants. In June when most flea beetle larvae are present, the *B. vulgaris* plants can already be completely deteriorated and unsuitable for both susceptible and resistant beetles (KMCA Vermeer and PW de Jong pers. obs.). Therefore, later in the season also the resistant flea beetles may disperse from *B. vulgaris* to other host plants. One must consider that the intra-seasonal increase in the percentage resistant beetles on other host plants than *B. vulgaris* G-type was only studied and found for two subpopulations. Therefore, further work is needed to investigate the generality of this finding of an increase of resistant phenotypes on other host plants than *B. vulgaris* G-type during the season.

The system of the phytophagous flea beetle *P. nemorum*, its host plants, and its natural enemies has all the characteristics of a geographic mosaic of local adaptations (Nielsen and de Jong 2005; Thompson 2005). There is local variation in the interaction between the flea beetle and its host plants, e.g., some patches of plants are not utilized by flea beetles at all. There seems to be a strong selection favouring the resistance loci on *B. vulgaris* G-type, whereas on other host plants the resistance locus does not seem to offer any advantage. Our data indicate that the geographic differences in adaptation are dynamic, leading to rapid temporal changes in allele frequencies at adaptive loci. Such rapid changes have been documented before (e.g., Foster et al. 2002), but not in any natural system involving resistance to plant defence. The polymorphism in the flea beetle's adaptation to the variable defence of one of its host plants and the rapid changes in the flea beetle's allele frequencies at the adaptive resistance loci, as well as the fact that we are at present close to unravelling the genetic basis of the flea beetle resistance trait at the DNA level, offer a unique opportunity to determine the causal factors involved in host plant use by this phytophagous insect.

Acknowledgements

We would like to thank Dr. Jens K. Nielsen for his expertise and helpful suggestions and for his hospitality and help when collecting the beetles. We thank Torben Sørensen, Dønnerup A/S (Svebølle), Ulrich B. Madsen, Eriksholm Skovdistrikt (Kværkeby), Torbjørn Nicolajsen, Municipality of Glostrup (Ejby), Anders Olesen and Bjørn Rasmussen, Falster Statsskovdistrikt (Suserup), Per Christensen, Hjallerup (Try G-plot) and Thomas Møller, Dronninglund (Try P-plot) for permission to work on their properties. We would also like to thank Bram Winkelman for technical support.

CHAPTER 5

Natural selection against resistance of a flea beetle to host plant defences

Kim M. C. A. Vermeer • Patrick Verbaarschot • Marcel Dicke • Peter W. de Jong

Abstract

The flea beetle *Phyllotreta nemorum* is polymorphic for resistance to the defence of the atypical host plant *Barbarea vulgaris* G-type. We investigated the factors influencing the geographical distribution of this resistance in a structured population in Denmark, where the proportion of flea beetles resistant to the defences of *B. vulgaris* is remarkably low for individuals collected on other host plants than *B. vulgaris*. We tested the hypothesis that the resistance trait is selected against when beetles live on non-defended host plants that are geographically separated from a patch of defended host plants. New primers were developed for 11 polymorphic microsatellite loci of *Phyllotreta nemorum* which, in addition to 10 previously developed microsatellite markers, were used to assess the genetic population structure nearby Kværkeby (Denmark). Of these 21 markers, 18 microsatellite loci were found to behave as neutral markers. The sampled distinct plant patches harboured genetically distinct flea beetle subpopulations. Outlier detection methods were applied to different measures of pairwise genetic differentiation between samples at microsatellite loci as well as the resistance trait and yielded the resistance trait as a candidate subject to selection. This also applied to the comparison between flea beetle subpopulations on non-defended host plants at different geographical distances from a patch of *B. vulgaris* G-type plants. These results suggest that the different frequencies of flea beetle resistance on patches of non-defended host plants cannot be explained by mere dispersal limitations, but that a locus-specific effect, such as directional selection, must also play a role. The possible mechanism of such selection, i.e. the breaking up of coadapted gene complexes, is discussed, as well as the comparison of different genetic differentiation measures and mutation models in population genomic studies.

INTRODUCTION

Coevolution is one of the key processes influencing and generating biodiversity (Sagan 1966; Thompson 1999b; Thompson and Cunningham 2002; Cornell and Hawkins 2003; Forde et al. 2004; Thompson 2005; Wheat et al. 2007; Hu et al. 2008; Laine 2009; Thompson 2009a). It involves the reciprocal evolutionary interaction between species. Coevolution seems more complex than formerly thought; reciprocal interactions are not ubiquitous throughout populations as previously depicted but populations are hierarchically structured (Wright 1938; Wright 1951; Wright 1968; Fox 1981; Agrawal et al. 2001; Avise 2004; Thompson 2005). Coevolution *sensu stricto* only occurs in some subpopulations within overlapping distributions of the interacting species. Moreover, the mode, direction and strength of selection is now thought to vary across geographical regions. Selection on traits that influence reciprocal species interactions may, for example, depend on the presence or absence of a third species (Benkman et al. 2001). In 2005, Thompson published his conceptual framework of the geographic mosaic of coevolution (Thompson 2005). This framework integrates multidisciplinary approaches to enable the study of the dynamics and structure of coevolving interactions. Local adaptation is a crucial part of the raw material for coevolution as a geographical mosaic (Thompson 2005); the basic template for the geographic mosaic of coevolution is created by local (co)evolutionary adaptation. Excellent examples of local adaptation have been found in the study of natural resistance of insect herbivores to plant defences (Ehrlich and Raven 1964; Strong et al. 1984; Mopper and Strauss 1998; Cornell and Hawkins 2003; Pellmyr 2003).

One of the challenges emerging from Thompson's conceptual framework is to explain the observed geographical distribution of alleles at adaptive loci in terms of (local) selection versus processes that affect the whole genome such as migration and genetic drift (CHAPTER 2). In Thompson's geographic mosaic theory of coevolution (Thompson 2005), the latter (genome-wide processes) alter the spatial distribution of traits, leading to one of the major attributes of the geographic mosaic of coevolution: "trait remixing". Trait remixing is predicted to influence the occurrence of traits amongst populations and only in some populations will traits of interacting species be well matched. As a result, reciprocal selection only occurs in so-called "coevolutionary hotspots" within the spatial distribution of an interaction, which are embedded in a broader network of "coldspots" where selection can occur but is not reciprocal. In his theoretical framework, Thompson presents specific evolutionary hypotheses and ecological predictions which make the theory amenable to empirical testing. Although suggestions for such tests have been published (Gomulkiewicz et al. 2007), empirical tests of the predictions of the geographic mosaic theory are still extremely rare. We recently proposed to employ a population genomics approach to attribute the distribution of alleles at adaptive loci to local selection or trait remixing, thus enabling the detection of coevolutionary hotspots (CHAPTER 2). In the present paper we apply this approach to an empirical study of a model system involving (co)evolutionary interactions between a herbivorous insect, *Phyllotreta nemorum* L. (Coleoptera: Chrysomelidae), and one of its host plants, *Barbarea vulgaris* R. Br. (Brassicales: Brassicaceae) (Nielsen 1997b; de Jong et al. 2000; Nielsen and de Jong 2005).

***Barbarea vulgaris* G-type – *Phyllotreta nemorum* system**

The flea beetle *P. nemorum* is an oligophagous herbivore, attacking a limited range of brassicaceous host plants. One of these, *B. vulgaris*, has two distinguished types in Denmark: a chemically defended type with **g**labrous leaves (hence called G-type) and a type with **p**ubescent leaves (P-type) that lacks chemical defence against *P. nemorum* and other herbivores, such as the diamondback moth *Plutella xylostella* (Shinoda et al. 2002; Agerbirk et al. 2003a). Flea beetles, in their turn, are polymorphic with respect to their resistance to the defence of the G-type of *B. vulgaris*: some show virulence (or resistance) towards its chemical defence, and some are susceptible to it. Crossing experiments demonstrated that major dominant resistance genes (R-genes) enable the flea beetles to overcome the plant defence (de Jong et al. 2000; Breuker et al. 2005; Nielsen and de Jong 2005). Different modes of inheritance of these genes (sex-linked and autosomal) present a puzzle with respect to the number of loci that may be responsible for the resistance of the flea beetles (Nielsen 1997a; Nielsen 1999; Nielsen 2012). However, it has been demonstrated that one dominant resistant allele at any of these putative loci is sufficient to enable the beetle to use *B. vulgaris* G-type as a host plant. Near Kværkeby the genetic basis of the resistance trait appears to be straightforward because a previous study showed that one dominant allele at a single major autosomal locus is sufficient to provide a beetle with resistance to the plant defence (de Jong et al. 2000).

The geographical distribution of resistance in the flea beetle is remarkable (De Jong and Nielsen 1999). Whereas virtually all flea beetles which are collected on *B. vulgaris* G-type are apparently homozygous resistant, the proportion of flea beetles using other host plants than this G-type that possess the resistance gene(s) is surprisingly low, and homozygous resistant individuals appear to be absent on these plants. This geographic distribution of adaptive genotypes may be explained in two different ways: 1) the flea beetle populations are highly genetically structured, and local selection favours resistance of the flea beetles on *B. vulgaris* G-type. Limited dispersal, a high frequency of local extinction and recolonization, and host fidelity may all contribute to a limited spread of flea beetle resistance to other host plants (Peterson and Denno 1998); 2) There is a cost associated with the resistance trait, so that selection acts against this specific genotype on other host plants than *B. vulgaris* G-type. This would also limit the spread of resistance, even when dispersal is not limited. Previous data show evidence that both of these mechanisms may operate in the field: allozyme analysis has shown a limited, though statistically significant, genetic differentiation between flea beetles collected on different plant patches, which was both influenced by the geographical distance between the collection sites, and the resistance phenotype of the beetles (de Jong et al. 2001; de Jong et al. 2009). Under certain conditions, resistant flea beetles appeared to have a lower fitness than susceptible ones through a high pre-adult mortality, both of homozygous and heterozygous resistant individuals, especially after outcrossing with a susceptible line (de Jong and Nielsen 2000; Breuker et al. 2007). This effect has only been found for laboratory lines into which the autosomal R-allele was crossed. Sex-linked R-genes do not seem to have any negative fitness consequences on other host plants than *B. vulgaris* G-type (Nielsen 1996; Agerbirk et al. 2003b). In earlier work, it was proposed that

the abundance of apparently homozygous resistant flea beetles on *B. vulgaris* G-type plants in the field, despite their relatively low fitness in the experiments mentioned above, may be explained by the presence of co-adapted gene complexes, consisting of the resistance gene and modifiers that counteract the negative pleiotropic effects (i.e. high mortality) of these R-genes (de Jong et al. 2000; de Jong and Nielsen 2002; Nielsen and de Jong 2005). Such co-adapted gene complexes are thought to break up after migrants from *B. vulgaris* outbreed with beetles living on other host plants, thus limiting the spread of resistance of the beetles (especially in genetically structured populations). At the same time, this mechanism would account for the relatively low levels of genetic differentiation that have been found between flea beetle populations, because the selection against resistance on other plants than *B. vulgaris* would take place only after outbreeding of resistant migrants with the populations on the non-*Barbarea* host plants, thus allowing for a large amount of genetic exchange.

The present study aims to explain the geographic distribution of resistance in natural flea beetle populations in terms of migration/dispersal versus selection. Based on our earlier results described above, we hypothesize that the observed geographical distribution of the resistance trait cannot solely be explained by limited dispersal, but also involves selection against resistant individuals on other host plants than *B. vulgaris* G-type. To test this hypothesis under natural conditions, it is necessary to employ a method to disentangle the roles of dispersal/migration and selection, i.e. a population genomics approach.

Flea beetle resistance is favoured by selection on the *B. vulgaris* G-type, as a logical result of only resistant beetles being able to survive on this plant. Our research interest, however, is to explain the relative rarity of the resistant flea beetles on other, non-defended host plants. The essential research question is therefore: is there evidence for selection *against* resistance in beetles that use other host plants than *B. vulgaris* G-type? The specific hypothesis is: *If coadapted gene complexes are indeed present in resistant flea beetles, and break up through recombination during outbreeding with other local flea beetle populations, then we expect that genetic differentiation between two flea beetle populations living on other host plants than the G-type, at different geographical distances from a G-type patch, is larger for the resistance trait than the expectation based on neutral markers.*

General approach

Most of the earlier work on the flea beetle-host plant interactions in Denmark has focused on populations around Kværkeby (de Jong et al. 2000; de Jong et al. 2001; Breuker et al. 2005; Nielsen and de Jong 2005). Even though allozyme studies have revealed limited, but significant, population structure for flea beetles using distinct plant patches around Kværkeby – even at a scale of ≈ 1 km between patches (De Jong et al. 2001), we tested for the sampled locations if they harbour genetically differentiated subpopulations. Ten microsatellite markers had already been developed for the flea beetle and we chose to use these highly variable microsatellite loci and a further eleven newly discovered microsatellite loci for our analyses. Microsatellite polymorphism of these new loci was assessed per locus using 192 up to 288 individuals of *P. nemorum* collected nearby Kværkeby from different locations. Genotyping

errors, such as stuttering and null alleles in microsatellites, can lead to apparent departure from the Hardy-Weinberg Equilibrium (HWE). Therefore, we first checked if the microsatellite loci were in HWE across subpopulations. We then tested if the subpopulations were each in HWE and used STRUCTURE (Pritchard et al. 2000), a software program to cluster genetic data to check for population structure. The basic algorithm was described by Pritchard et al.

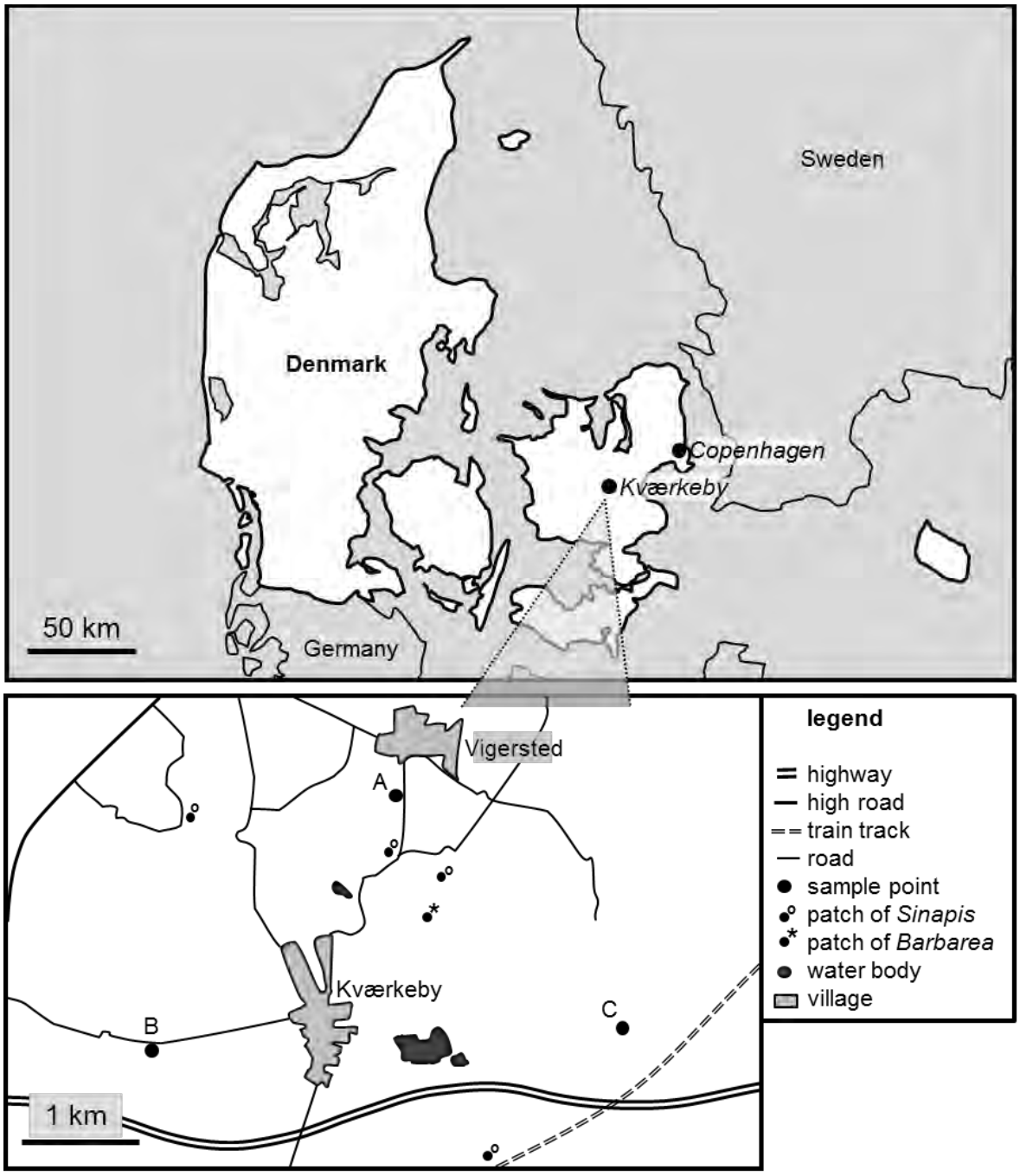


Figure 1 Map of the sites at which *Phyllotreta nemorum* was sampled, and other plant patches in the vicinity that were not studied in detail for the population genomics approach. (A) Kværkeby, *Sinapis arvensis*; (B) Kværkeby, *Sinapis arvensis*; (C) Kværkeby, *Barbarea vulgaris* G-type (for more details see Table 1). Only patches of *Sinapis arvensis* or *Barbarea vulgaris* G-type colonized by flea beetles are shown on the map.

(2000). Extensions to the method were published by Falush et al. (2003), (2007) and (2009). In addition, the differentiation between subpopulations was visualized by factorial correspondence analysis (FCA) using GENETIX version 4.03 (Belkhir et al. 1998) and principal coordinate analysis (PCA) using GENALEX version 6.13 (Peakall and Smouse 2006).

A population genomics approach was used to measure the variation in heterozygosity and allele frequencies within and between subpopulations. As in the classic population genomics approach (Black, et al. 2001; Luikart et al. 2003), numerous individuals were sampled from each subpopulation and various loci were genotyped across the genome. When a locus shows a significantly deviating level of differentiation between the subpopulations compared to the majority of the loci (i.e. if the value of the measurement of differentiation at this locus forms an outlier), then this locus is a candidate locus for being under selection. Our specific aim was to test whether flea beetle resistance is likely to be under selection in subpopulations on other host plants than *B. vulgaris* G-type. Because the genetic basis of resistance has not been resolved at the DNA level, we estimated genotype frequencies at the resistance locus using the phenotype frequencies, assuming that the resistance is governed by one resistance gene in the Kværkeby subpopulations (as supported by data from de Jong et al. (2000)), and that the subpopulations are in HWE at this locus. These estimates were then used to calculate differentiation at this locus between subpopulations, which was subsequently compared to differentiation at the neutral microsatellite loci. Different measures for genetic differentiation were used, because different underlying models lack consensus about which one of these is correct to use for microsatellites (Goldstein et al. 1995; Slatkin 1995; Takezaki and Nei 1996; Balloux and Lugon-Moulin 2002; Jost 2008; Leng and Zhang 2011). By using F_{ST} (estimator θ as in Weir and Cockerham (1984)) and R_{ST} (Slatkin 1995) both the infinite allele model of mutation (IAM) and the stepwise-mutation model (SMM) have been considered. We used the program LOSITAN for the frequentist method (fdist) (Beaumont and Nichols 1996; Antao et al. 2008) and ran it under the assumption of the IAM as well as the SMM. Moreover, we also used the program BAYESCAN for the bayesian method (Balding 2003; Beaumont and Balding 2004; Foll and Gaggiotti 2008). Further details of this approach are described in the next section.

Table 1 Sampling sites and their coordinates in Denmark, near Kværkeby. N refers to the number of sampled adult beetles that were subsequently used in bio-assays.

Location	host plant	Latitude	Longitude	N	distance (km) from A	distance (km) from B
A	*Sinapis arvensis*	+55° 28' 47.24"	+11° 53' 2.90"	162	0	2.5
B	*Sinapis arvensis*	+55° 27' 32.67"	+11° 51' 4.01"	129	2.5	0
C	*Barbarea vulgaris* G-type	+55° 27' 40.82"	+11° 55' 0.08"	113	2.3	3.4

Table 2 Summary of polymorphic microsatellite markers developed for *Phyllotreta nemorum*. Sequences are given for forward (F) and reverse (R) PCR primer.

Microsatellite locus (GenBank ID)	Primer sequence 5' - 3'	Repeat sequence motif†	Allelic size range	Annealing temperature	Number of alleles (Effective number)
PnA06	F: FAM-CCAGAAATGTCATCGTACA	(AC)8	190-228	60	20
(FJ217683)	R: GTTTGGTCTTTGTGATGGACAAGG				(5.02)
PnA10	F: NED-ACTCACGCCCGAATCGCTTC	(CGA)5	146-141	57	6
(FJ217684)	R: GTTTAGAATGGACATGGTCGGCGG				(2.57)
PnB11.311	F: FAM-CTTCGAAATAATCGTCTTC	(CGT)4N22(TCG)4	133-154	57	8
(FJ217687)	R: GTTTAATCTGGAGACGATGATGAC				(1.41)
PnC05.11	F: HEX-TCACCAAATTGTGACATGTACC	(TAAA)5	147-163	57	5
(FJ217688)	R: GTTTACCATAAACGCACTGTTGA				(2.27)
PnC10	F: NED-GGGCAACGAAACAAACGACGGTACC	(GAC)5	198-213	50	6
(FJ217689)	R: CGTTTGAGTACCGCTGCCGG				(1.71)
PnD01	F: FAM-AACTTTGCATCGAATTGTGC	(GCC)5N85(TGC)8	347-398	50	18
(FJ217690)	R: GTTTGTCTTGGAACGCTAATCGC				(7.4)
PnD02	F: HEX-ATCAGCAGGCCATGCAGCAG	(CAG)5	135-162	57	10
(FJ217691)	R: GTTTGCGTTGTAGCCGGATTTTG				(5.74)
PnD06	F: NED-ATCACGTTCGGCACCACCTG	(CTG)7	171-186	53	6
(DQ507814)	R: GTTTCTTCAGCAGCCTGATGGGC				(4.28)
PnE08.61	F: FAM-GCAGCAGGTCGAGGCGACTG	(CAG)4	133-160	50	10
(FJ217694)	R: GTTTATTCGCCACCGTACCGTTCG				(4.33)
PnE10	F: NED-TTTGAAAACTATTGCCCATC	(TGC)7N4(GCT)4	184-214	50	11
(FJ217695)	R: GTTTATGTCTTGGAACGCTAATCG				(5.14)
PnF03	F: FAM-AAATCCTTCAAAGGCTAAGCCAGC	(TGC)11	327-360	55	12
(FJ217696)	R: CAACGGTTCAGCAGCAACG				(1.62)
PnG06	F: HEX-CGTGGAATTTGTCAACACTTG	(GCC)8	192-228	50	13
(FJ217697)	R: GTTTGAGGACTTAGGTTTTGGTATAG				(4.66)

† The number after 'N' denotes the number of nucleotides in between repeat sequences.

MATERIALS AND METHODS

Sample collection

For the present study we sampled the same, and one new nearby, locations as in previous studies (Nielsen and de Jong 2005; CHAPTER 4) focussing on the flea beetle populations around Kværkeby (Denmark). Both *B. vulgaris* G-type and other host plants, such as *Sinapis arvensis, B. vulgaris* P-type and *Capsella bursa-pastoris* occur at this location. In 2009 on the 11th and 26th of June leaves with leaf mining flea beetle larvae were collected on *Sinapis arvensis* and *B. vulgaris* G-type nearby Kværkeby from various plant patches (see Figure 1 and Table 1 for exact locations). The area has been thoroughly searched for host plant- and flea beetle occurrence. Other patches with host plants were present, but not all were colonized by flea beetles. Patches of host plants that were colonized at least during the 2008 and 2009 sampling periods are shown in Figure 1. The host plants grow in distinct patches surrounded by areas where no suitable host plants for *P. nemorum* occur. On the 11th of June 75 leaves were collected from plant patch A (corresponding to site 4 in Nielsen and de Jong (2005)). On the 26th of June an additional ±100 leaves were collected from plant patch A and ±200 leaves from each of the two other plant patches, B and C. Location C (corresponding with site 1 in Nielsen and de Jong (2005)) consists of a G-type plant patch and is located respectively 2.3 and 3.4 km away from locations A and B where *S. arvensis* is the host plant. Mined leaves were picked from various plant individuals in each patch at each location to avoid potential bias from family structuring across plants. Picked leaves usually contained 1-3 larvae, though some leaves (especially *S. arvensis*) occasionally contained more larvae. These leaves were transferred to 500 ml plastic vials containing a layer of moist peat and vermiculite (5:1 volume ratio). Vials were closed using a plastic lid with a central hole of 15mm diameter, plugged with cotton wool, and kept at 20 ± 2 °C. Full-grown larvae pupated in the layer of soil, after which we removed plant material from the vials. After 2.5 weeks, adult beetles started to emerge from the soil. Individuals were sexed using a stereomicroscope (25 x magnification) (de Jong and Nielsen 1999). The sex ratio for each subpopulation was approximately evenly balanced between males and females (Chi square test, DF = 1, $P > 0.05$).

Bioassays

After sexing, each beetle was used in a no-choice bioassay within 24 h of emergence to determine whether they were resistant to defences of *B. vulgaris* G-type. Beetles had no access to food prior to the bioassay. The bio-assays were set up as in Chapter 4. Because 96-wells plates were used for the DNA analysis (see below) only 96 beetles per sample location were used for further analyses. Supernumerary beetles were also bio-assayed but excluded from further analysis. As in Chapter 4, non-feeders were classified as susceptible beetles and feeding beetles were classified as resistant beetles. Only < 1.5% of the bioassays were inconclusive when using this method.

The bioassays provide information on the flea beetle's phenotype in terms of the resistance to the G-type of *B. vulgaris*. To be able to use statistics on genotypic microsatellite data, the resistance genotype of each beetle was calculated assuming HWE within each sampled population. The frequency of the susceptible genotype is the same as for the susceptible phenotype (rr), because resistance is a dominant trait. The resistant phenotype can either be heterozygous (Rr) or homozygous (RR). We consider this assumption of HWE within samples justified, because a) we are interested in outliers of measures that indicate differentiation *between* subpopulations, and b) if directional selection at the resistance locus would lead to a deviation from HWE (with a deficiency of heterozygotes), our assumption of HWE actually represents a conservative approach with respect to the chance of finding the resistance to be an outlier relative to the neutral microsatellite loci.

DNA and microsatellite isolation and amplification

Whole genomic DNA was isolated from each beetle using the DNeasy® Blood & Tissue Kit (QIAGEN), following Purification of Total DNA from Animal Tissues (Spin-Column Protocol). Ten polymorphic microsatellite loci were already available for usage. The microsatellite primers for these loci have been characterized in Verbaarschot et al. (2007). New microsatellite markers were constructed by using di-, tri-, and tetranucleotide repeat-enriched libraries of *P. nemorum* genomic DNA constructed for previous microsatellite marker development using the same method as in Verbaarschot et al. (2007). The newly designed microsatellite primer pairs were evaluated via polymerase chain reaction (PCR) amplification using the isolated DNA of 96 beetles of each sampled plant patch. Together with the 10 microsatellites characterized by Verbaarschot et al. (2007) these newly designed microsatellite primer pairs were multiplexed (true multiplexing) for efficient amplification and scoring of alleles for the three locations near Kværkeby.

Amplification was performed in a 25 µl reaction volume using 15.875 µl nuclease-free water, 5 µl 5X colorless GoTaq® flexi buffer, 1.5 µl 25 mM $MgCl_2$, 0.5 µl 200 µM dNTPs, 0.5 µl 500 nM of each primer, 0.125 µl GoTaq® DNA polymerase, 1 µl of DNA (at least 4ng of DNA in water). Each forward primer was 5'-fluorescently labelled with either 6-FAM, NED and HEX (Applied Biosystems). Annealing temperatures used for the PCR amplification are shown in Table 2 and (Verbaarschot et al. 2007). PCR conditions consisted of an initial cycle at 94°C for 3 minutes, 30-35 repeat cycles at 94°C for 30 seconds, the annealing temperature for 30 seconds, 72°C for 1 minute, and a final extension at 72°C for 5 minutes. For all markers 35 repeat cycles were used except for markers PnD09, PnH09 and PnH12 (30 cycles) (see also Verbaarschot et al. (2007); coding of the markers was done as in Verbaarschot et al. (2007)).

The PCR products were then purified using Sephadex G-50 (Sigma) spin columns in a MultiScreen®-HV 96-well filter plate (Millipore) according to the Sequencing Reaction Purification Protocol (step 1 to 5) (Millipore 2000). The purified sequencing products were then run on an ABI PRISM 3700 sequencer (Applied Biosystems, Foster City, CA, at Greenomics,

Wageningen) using GeneScan-LIZ 500 as an internal marker (Applied Biosystems, Foster City, CA).

Data analysis

Genetic diversity. Results were compiled and analysed with GeneMarker® V1.5 software (Softgenetics®). Allele evaluation and the stutter peak filter failed to detect some alleles that were noticed after visual inspection, yet the program summed up many alleles that did not seem legitimate. Adjusting the parameters of the program did not satisfactorily improve this and thus fragment scoring was done mainly by eye using the electrophoregrams visualized by GeneMarker. We made sure that scoring occurred "blind", i.e. during scoring the sample location was not known to the person who did the scoring. Allelic size, total number of alleles (T_N), effective number of alleles (E_N), observed (H_o) and expected heterozygosity (H_e) were determined for each microsatellite locus. All these statistics were calculated using the GENALEX version 6.41 add-in (Peakall and Smouse 2006) in Excel and FSTAT version 2.9.3 (Goudet 1995; Goudet 2002). Parameters H_o and H_e were calculated using the estimate for genetic diversity as described by Nei, (1973; 1987) and E_N was calculated as

$$E_N = 1/ \Sigma p_i^2$$ (Hartl and Clark 1997).

Presence of null alleles and scoring errors due to stuttering were checked with MICRO-CHECKER version 2.2.3 (Van Oosterhout et al. 2004). As in Verbaarschot et al. (2007) locus PnBB08 showed an excess of homozygotes for all sampled locations. Also newly developed primer pair PnG06 showed possible presence of null alleles due to a heterozygote deficit when tested for all sample locations. Primer pair PnC10 only showed an excess in homozygotes for sample location A and B, but not for location C. Primer pairs PnF03 and PnH09 only showed a heterozygote deficit for sample location C and primer pair PnA06 only showed a heterozygote deficit for sample location A. There was no evidence for the presence of null alleles for all other loci. Scoring errors due to stutter bands and large allele drop-out were not detected either. FSTAT version 2.9.3.2 (Goudet 2002) was used to test deviation from HWE, inbreeding and linkage disequilibrium by calculating F_{IS} (Wright 1951) using the estimator f as in Weir and Cockerham (1984) and log-likelihood ratio G-statistics. Deviation from HWE was globally tested for each pair of loci across all three sample locations based on 1000 permutations. Consistent with MICRO-CHECKER, PnBB08, PnC10 and PnG06 were not in HWE (all P value < 0.001), with an observed heterozygosity significantly lower than expected (Table 3).

HWE deviation was also tested for all pairs of loci per sampled plant patch based on 6300 randomisations with a 5% nominal level. Again PnBB08, PnC10 and PnG06 were not in HWE, with an observed homozygosity significantly higher than expected for some sampled locations (each P value < 0.0001 for two sample locations, Table 3). Moreover, primer pair PnH12 showed an excess of heterozygotes for each sampled location (P value < 0.0001) and across all locations (P value < 0.001; data not shown). We ruled out a recent bottleneck in the population (assuming neutral theory) and a sudden significant increase in population size (Cornuet and Luikart 1996) by not finding a significant heterozygosity excess nor a sig-

Table 3 The Fixation Index (F_{IS}) is given for each sampled plant patch (i.e. A, B and C) and the mean for all samples together (All). Indicative adjusted nominal level (5%) per plant patch is: 0.00079, indicative adjusted nominal level (5%) for the whole population is: 0.00217. Asterisks indicate the proportion of randomisations for each microsatellite locus that gave a significantly larger or smaller F_{IS} than the observed based on 6300 randomisations for each sampled plant patch and 1000 randomisations for all.

Microsatellite locus	Fixation Index			
	A	B	C	All
PnA03	-0.008	-0.016	-0.137	-0.051
PnA04	0.185	-0.041	0.009	0.043
PnA06	0.090	-0.015	-0.112	-0.061
PnA10	0.133	0.098	-0.342[a]	-0.025
PnAB12	0.033	-0.109	0.098	0.016
PnBB08	0.154	0.447[a]	0.222[a]	0.288[b]
PnB11.311	-0.078	-0.034	0.087	-0.005
PnC05.11	0.019	0.183	-0.226	-0.087
PnC10	0.362[a]	0.346[a]	-0.143	0.232[b]
PnD01	0.037	-0.012	0.042	0.026
PnD02	-0.11	-0.076	-0.038	-0.069
PnD04	-0.161	-0.141	-0.18	-0.142
PnD06	-0.119	0.047	0.037	-0.018
PnD09	0.152	0.025	-0.04	0.013
PnE08.61	-0.191	-0.152	-0.091	-0.135
PnE10	-0.145	-0.061	-0.027	-0.073
PnE11	-0.064	0.015	-0.044	-0.02
PnF03	0.065	0.142	0.261	0.099
PnG06	0.162[a]	0.288[a]	0.27[a]	0.239[b]
PnH09	-0.085	-0.126	0.303	-0.102
PnH12	-0.348[a]	-0.254[a]	-0.252[a]	-0.268

[a] significance at $P < 0.0001$

[b] Significance at $P < 0.001$

nificant lack of heterozygosity as in Nei (1978) for all markers when compared to expected heterozygosity (Table 3). We found a heterozygosity excess for some markers, however. Heterozygosity deficiency of some markers may represent a presence of null alleles or genetic hitchhiking (Barton 2000). Because markers with a heterozygosity deficiency are not regarded as 'neutral', PnBB08, PnG06 and PnC10 were discarded in subsequent analyses. Linkage disequilibrium was tested between all pairs of loci based on a confidence level of 1%, i.e. 21000 permutations. Primer pairs PnE11, PnE10, PnD01 and PnD02 showed linkage amongst each other when considering a 1% nominal level after Bonferroni correction (P value < 0.00005). All other loci showed no significant deviation from linkage equilibrium (all P values > adjusted P value for a 5% nominal level) so we assumed that these loci were independent.

Analysis of population structure. A Bayesian clustering method (see Shoemaker et al. (1999) for a good review about the use of this method in genetics) for inference of population structure based on unlinked (or at least not tightly linked) markers has been implemented in the computer program STRUCTURE version 2.3.3 (Pritchard et al. 2000; Falush et al. 2003; Pritchard et al. 2007; Hubisz et al. 2009). This program assigns individuals to clusters, giving the number of clusters as *K*, estimating for each individual the proportion (*q*) of the genome that originated from a given genetic cluster. This estimation is done by the Maximum Likelihood method (see e.g. Beebee and Rowe (2004)). STRUCTURE assumes HWE within subpopulations and performs best with unlinked loci. Therefore, we excluded from the analyses those markers that are in linkage disequilibrium, namely PnE10, PnD01 and PnD02. Marker PnH12, which showed an excess of heterozygotes for all samples was excluded from the analyses as well. Individuals were first clustered in STRUCTURE with no prior population information. Afterwards the data sets were analysed again, this time with prior location information in the form of sample location A, B and C to assist clustering. STRUCTURE v2.3.3 can detect structure at lower levels of divergence than the original STRUCTURE models by using prior sampling location information, without bias towards detecting structure when none is present (Hubisz et al. 2009). Sampling information is ignored by this version of the program when there is evidence of population structure uncorrelated with sampling locations. For all analyses, we set most of the parameters to their default values as advised in the user's manual of STRUCTURE 2.3.3 (Pritchard et al. 2010). Exploring runs with varying λ, however, showed that the estimations of λ were often much lower than the λ = 1 expected under the independent allele frequencies model (Falush et al. 2003) and dependent on the number of assumed genetic clusters (*K*). We therefore let λ vary during each run. Data sets were analysed under the admixture model and dependent allelic frequencies (Falush et al. 2003). The algorithm was run 10 times for each number of genetic clusters, which was set from 1 to 5, for 100 000 burn-in steps and 100 000 replications, respectively. Additionally, the overall mean values of the *r* parameter in STRUCTURE were calculated for each genetic cluster when sample location information was used to assist clustering. The *r* parameter estimates if the location data were informative in defining the clusters. Values close to and/or below 1 indicate that the location data were informative, while values much larger than 1 indicate that the location data were not informative for the structure found (Hubisz et al. 2009; Pritchard et al. 2010). The results were compiled with STRUCTURE HARVESTER version 0.6.92 (Earl and Vonholdt 2012). Output files from STRUCTURE HARVESTER were run in CLUMPP version 1.1.2 (Jakobsson and Rosenberg 2007) to get summarized *q*-matrices of the repeated runs for each *K* and visualized in Distruct 1.1 (Rosenberg 2004). Differences between the data Maximum Likelihood of successive *K* values were tested using the non-parametric Wilcoxon two-sample test for dependent samples and SPSS 20 software. The most likely *K* value was estimated based on: (1) the standard method for estimating *K* as described in the manual of STRUCTURE (Pritchard et al. 2010) by calculating the posterior probability of *K* and (2) the largest *K* value with significantly higher likelihood than that from *K*-1 runs. Often the ΔK method of Evanno et al. (2005) is also used to estimate *K*, especially when the estimated probability of data for successive K values plateaus or continues to increase slightly after the presumed real K value is reached (Evanno et al. 2005; Campana et al. 2011). We chose not to include this method because the data Maximum Likelihood de-

creased after K = 3 and we only wanted to assure that our sampled plant patches could be considered genetically distinct subpopulations to at least some extent.

Finally, to corroborate clustering done by STRUCTURE, both a principal coordinate analysis (PCA) and a factorial corresponding analysis (FCA) were used to make a graphical representation of the repartition of the three individually sampled subpopulations. The PCA was done with data standardization after calculating the pair-wise genetic distance between samples using the GENALEX add-in (Peakall and Smouse 2006) in Excel. This PCA provides major axes explaining the variation in the dataset. The FCA was carried out with notice of sample information with GENETIX software (Belkhir, Borsa et al. 1998). This FCA plot identifies the major contributors to the overall genetic variation using a microsatellite individual-genotype matrix and puts higher priorities on finding relationships within a population than between populations.

Outlier detection. Both the frequentist method (fdist) and the bayesian method were used to identify outlier loci in the microsatellite dataset and test for outlier behaviour of the resistance trait and microsatellites (Beaumont and Nichols 1996; Beaumont and Balding 2004). Multiple analyses were done per program to get more robust results. LOSITAN (Antao et al. 2008), a user-friendly program in which simulations were done assuming either the infinite allele model (IAM) or the stepwise mutation model (SMM), was used for fdist. 50 000 simulations were run in LOSITAN with a confidence interval of 99% to test for patterns of possible selection by evaluating the relationship between the $F_{ST,}$ using the estimator θ as in Weir and Cockerham (1984) and expected H_e for each locus. A 'neutral' mean F_{ST} was computed for each pair of subpopulations and all three subpopulations together. To find the 'neutral' mean F_{ST}, LOSITAN removes potential outliers during the simulations when computing the F_{ST} value. We used the program BAYESCAN version 2.1 (Foll and Gaggiotti 2008) for the bayesian method (Balding 2003; Beaumont and Balding 2004). A multinomial-Dirichlet distribution is used as prior distribution of the allele frequencies in the subpopulation (Foll and Gaggiotti 2008). The program performs a logistic regression over two different models; one model that includes a locus-specific component (α) that is shared by all populations, and a population-specific component (β) that is shared by all loci, the other model has only the population-specific component β. BayeScan estimates a posterior probability of each one of these models using a reversible-jump MCMC algorithm. Because α is locus-specific, the posterior probability indicates for each locus if there is an effect of selection. 100 000 simulations were run in total, with 50 000 burn-in steps and 50 000 replications. Posterior probabilities were computed for each pair of subpopulations and all three subpopulations together. These posterior probabilities should not be seen as p-values. A posterior probability above 0.91 can already be considered strong support for a model including α. When for a locus the posterior probability exceeds 0.99, this provides decisive evidence for selection according to Jeffreys' scale of evidence for Bayes factors (Foll 2012).

Genetic differentiation between two subpopulations was further estimated by pairwise G_{ST} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005) and D (Jost 2008) for each microsatellite and the resistance trait, using SMOGD (Crawford 2010). Microsatellites that showed an excess of

homozygotes in a subpopulation according to MICROCHECKER and FSTAT were left out of the analysis, because we are only interested in neutral microsatellite loci. G_{ST} is equivalent to Wright's F_{ST} (Wright 1951; Nei 1973) but also accounts for multiple loci with multiple alleles. In 1983, Nei and Chesser improved G_{ST} by taking into account deviation from HWE within subpopulations and sample size (Nei and Chesser 1983). Hedrick developed the standardized measurement G'_{ST}, because the outcome of G_{ST} is dependent on the level of heterozygosity. The value of G_{ST} will always be very small with high levels of heterozygosity, even if not many alleles are shared by the subpopulations compared. Jost's *D* is calculated with the effective number of alleles rather than heterozygosity, so *D* is independent of average within-subpopulation heterozygosity. Furthermore, the estimator *D* is also dependent on the mutation rate of the marker. Leng and Zhang (2011) advise to calculate both G'_{ST} and Jost's *D* and the relative levels of heterozygosity when using microsatellites. The estimator of G_{ST}, G'_{ST} and Jost's *D* are given by SMOGD as well as their 95% confidence interval after 500 bootstrap replicates done at the individual level.

RESULTS

All newly found microsatellites were di-, tri- or tetra-nucleotide repeat motifs and polymorphic for the flea beetle population in Kværkeby; the number of alleles per marker ranged from 5 up to 20 (Table 2). The previously developed microsatellite markers were also polymorphic for this population (Table 4). For none of the sampled plant patches the markers were fixed. For each microsatellite the effective number of alleles (E_N) ranged between 1.42 and 6.42 per locus, with an average of 4.07 and expected heterozygosity (H_e) ranged from 0.29 to 0.89 (Table 4), with an average of 0.68 (average over all loci except PnBB08, PnC10, and PnG06 – see methods section). For each plant patch the average H_e was between 0.67 – 0.70. The average F_{IS} value per locus ranged between -0.27 and 0.29 (Table 3) with an overall F_{IS} of -0.012, indicating that inbreeding was not likely because F_{IS} values are expected to be close to zero when there is no inbreeding. Only for one locus a strong deviation from HWE was found. Hardy–Weinberg equilibrium for the whole population and over all loci was not rejected ($P > 0.99$).

Bioassays

All beetles from plant patch C (*B. vulgaris* G-type plants) were classified as being resistant. From plant patch A (*Sinapis arvensis* plants), nearly 70% of the beetles (66 of 95 individuals) were classified as resistant and from plant patch B (*Sinapis arvensis* plants) only 30% (28 of 93 individuals) were classified as resistant (Table 5). Of one beetle from plant patch A we could not with certainty determine the phenotype and for plant patch B this was the case for 3 beetles. All other beetles were classified as susceptible beetles. Table 5 displays both the resistance phenotype and the inferred genotype when HWE is assumed within subpopulations.

Table 4 Descriptive statistics of the microsatellite loci and their allelic size range, total number of alleles (T_N), effective number of alleles (E_N), observed (H_o) and expected (H_e) heterozygosity per locus based on HWE. H_o and H_e are given for each sampled plant patch (i.e. A, B and C) and all samples together (All).

				heterozygosity			
Microsatellite Locus	Allelic size range	T_N	E_N	A H_o / H_e	B H_o / H_e	C H_o / H_e	All H_o / H_e
PnA03	197-225	15	5.24	0.82 / 0.81	0.81 / 0.79	0.94 / 0.83	0.86 / 0.81
PnA04	171-183	5	1.87	0.33 / 0.40	0.51 / 0.49	0.50 / 0.50	0.45 / 0.47
PnA06	190-228	16	6.37	0.81 / 0.89	0.84 / 0.84	0.85 / 0.76	0.84 / 0.83
PnA10	146-141	6	2.55	0.55 / 0.63	0.44 / 0.48	0.90 / 0.67	0.63 / 0.61
PnAB12	185-224	14	6.42	0.83 / 0.86	0.97 / 0.87	0.72 / 0.79	0.84 / 0.85
PnBB08	209-257	16	6.02	0.73 / 0.86	0.44 / 0.78	0.67 / 0.85	0.61 / 0.84
PnB11.311	133-154	8	1.42	0.46 / 0.42	0.24 / 0.23	0.18 / 0.2	0.29 / 0.29
PnC05.11	147-163	5	2.24	0.56 / 0.57	0.47 / 0.55	0.66 / 0.54	0.16 / 0.49
PnC10	198-213	6	1.94	0.22 / 0.35	0.20 / 0.30	0.75 / 0.65	0.39 / 0.47
PnD01	347-398	17	7.18	0.85 / 0.87	0.88 / 0.86	0.81 / 0.84	0.85 / 0.89
PnD02	135-162	10	5.58	0.93 / 0.83	0.89 / 0.82	0.94 / 0.86	0.89 / 0.83
PnD04	136-166	10	4.31	0.83 / 0.71	0.85 / 0.74	0.98 / 0.82	0.89 / 0.78
PnD06	171-186	5	4.37	0.79 / 0.76	0.74 / 0.75	0.72 / 0.79	0.75 / 0.78
PnD09	164-197	10	4.56	0.60 / 0.68	0.80 / 0.82	0.84 / 0.81	0.75 / 0.79
PnE08.61	133-160	10	4.29	0.86 / 0.72	0.88 / 0.76	0.88 / 0.8	0.88 / 0.77
PnE10	184-214	11	5.08	0.93 / 0.81	0.86 / 0.80	0.82 / 0.79	0.87 / 0.81
PnE11	208-235	10	5.37	0.90 / 0.84	0.80 / 0.81	0.83 / 0.79	0.84 / 0.82
PnF03	327-360	12	1.51	0.42 / 0.45	0.26 / 0.31	0.15 / 0.21	0.28 / 0.32
PnG06	192-228	12	4.58	0.67 / 0.80	0.56 / 0.77	0.56 / 0.77	0.60 / 0.79
PnH09	159-165	3	1.90	0.58 / 0.52	0.39 / 0.34	0.66 / 0.52	0.54 / 0.46
PnH12	143-164	8	2.44	0.88 / 0.65	0.80 / 0.63	0.28 / 0.43	0.65 / 0.57

Genetic structure in the flea beetle population near Kværkeby

The genetic structure of all individuals was analysed in STRUCTURE with no prior location information using microsatellite markers that did not significantly deviate from HWE and did not show linkage disequilibrium. The data likelihood of successive *K* values (i.e. the number of genetic clusters assumed by the model) only decreased by increasing the *K* value and the posterior probability of *K* = 1 was the highest probability (Pr(*K*)=0.996) compared to the posterior probabilities found for the other *K* values, both suggesting *K* = 1 as the most likely *K* (Table 6). Additionally, the proportion of cluster membership (*q*) of each individual and each plant patch sample was fairly close to 1/*K* for each *K* value, meaning that each individual could be assigned to any genetic cluster. All of this suggests little to no population structure or that too few microsatellites/individuals were used for the program to pick up structure (Pritchard et al. 2010).

Table 5 Resistance phenotype of the flea beetle and inferred genotype when HWE is assumed within each sampled population. 'res' is a resistant phenotype and 'sus' a susceptible phenotype. R is the dominant allele for resistance, r the recessive allele.

Location	plant type	phenotype	#	genotype	#
A	*Sinapis arvensis*	res	66	RR	19
				Rr	47
		sus	29	rr	29
B	*Sinapis arvensis*	res	28	RR	3
				Rr	25
		sus	65	rr	65
C	*Barbarea vulgaris* G-type	res	96	RR	96
				Rr	0
		sus	0	rr	0

When individuals were classified into clusters with prior location information to assist clustering, the program was able to detect structure. The data likelihood of *K* values larger than 1 was higher than found when no prior location information was used (Table 6). Also the mean of the *r* parameter shows that for *K* values 2 to 5 the location data were helpful in defining the clusters, since the mean value of *r* was always close to 1 or smaller than 1. The data likelihood for each *K* value increased slightly when prior location information was incorporated into the clustering method until a *K*-value of 3, though the increase from $K = 2$ to $K = 3$ was not significant (non-parametric Wilcoxon two-sample test). Bigger *K*-values produced lower data likelihoods.

Table 6 Results of the population genetic analyses carried out with STRUCTURE and SPSS. *K* is the number of genetic clusters assumed by the model, Ln P(*K*) is the log likelihood of the *K* value and Pr(*K*) is the posterior probability of the *K* value. "*r*" is a parameter which estimates if the location data were informative in defining the clusters. Values below 1 indicate that the location data were informative and values much larger than 1 indicate that the location data were not informative for the structure found. "n.s." stands for *not significant* ($P > 0.05$), "NA" stands for *not available*.

K	prior location data	number of runs	Ln P(K)		Pr(K)	Wilcoxon significance	mean *r*
			mean	standard dev.			
1	no	10	-8041	0.08	1	NA	NA
2	no	10	-8065	8.31	0	0.005	NA
3	no	10	-8115	38.02	0	0.005	NA
4	no	10	-8141	33.99	0	n.s.	NA
5	no	10	-8170	47.15	0	n.s.	NA
1	yes	10	-8041	0.06	0	NA	8.88
2	yes	10	-7972	1.52	0.004	0.005	0.99
3	yes	10	-7967	7.66	0.996	n.s.	0.57
4	yes	10	-7994	36.59	0	n.s.	0.76
5	yes	10	-8004	32.06	0	n.s.	0.61

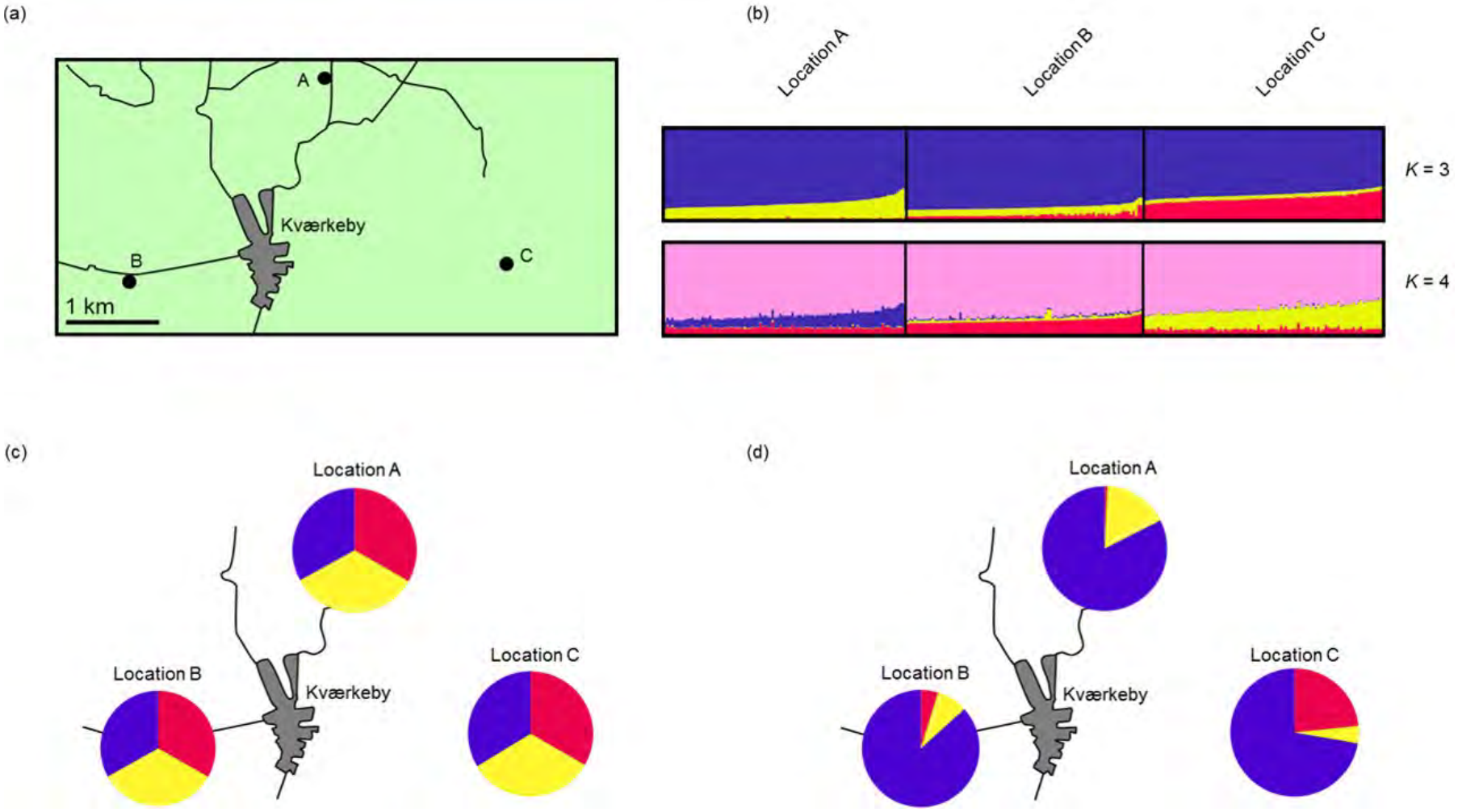


Figure 2 STRUCTURE analysis of *Phyllotreta nemorum* subpopulations near Kværkeby (Denmark). (a) Geographic representation of the sample area. (b) Graphical representations of ancestry membership coefficients inferred from STRUCTURE analysis with prior location knowledge. Each individual is shown as a vertical line divided into coloured segments representing the estimated membership proportions in the three and four genetic clusters (*K*) inferred with STRUCTURE. Individuals are arranged according to each sampled plant patch at location A, B and C. Magenta is the fraction of membership (*q*) of cluster 1, yellow of cluster 2, purple of cluster 3 and pink of cluster 4. (c) Graphical representation of each sampled plant patch and the percentage of each cluster as an average over 10 repeats at *K* = 3 *without* prior location knowledge. Each sample is shown as a pie divided into segments representing the overall membership proportions in the three genetic clusters inferred with STRUCTURE. As for (b) magenta stands for the fraction of membership (*q*) of cluster 1, yellow of cluster 2 and purple of cluster 3 (d) Graphical representations of each sampled plant patch and the estimated membership proportions of each cluster as an average over 10 repeats at *K* = 3 as in (c) but this time *with* prior location knowledge.

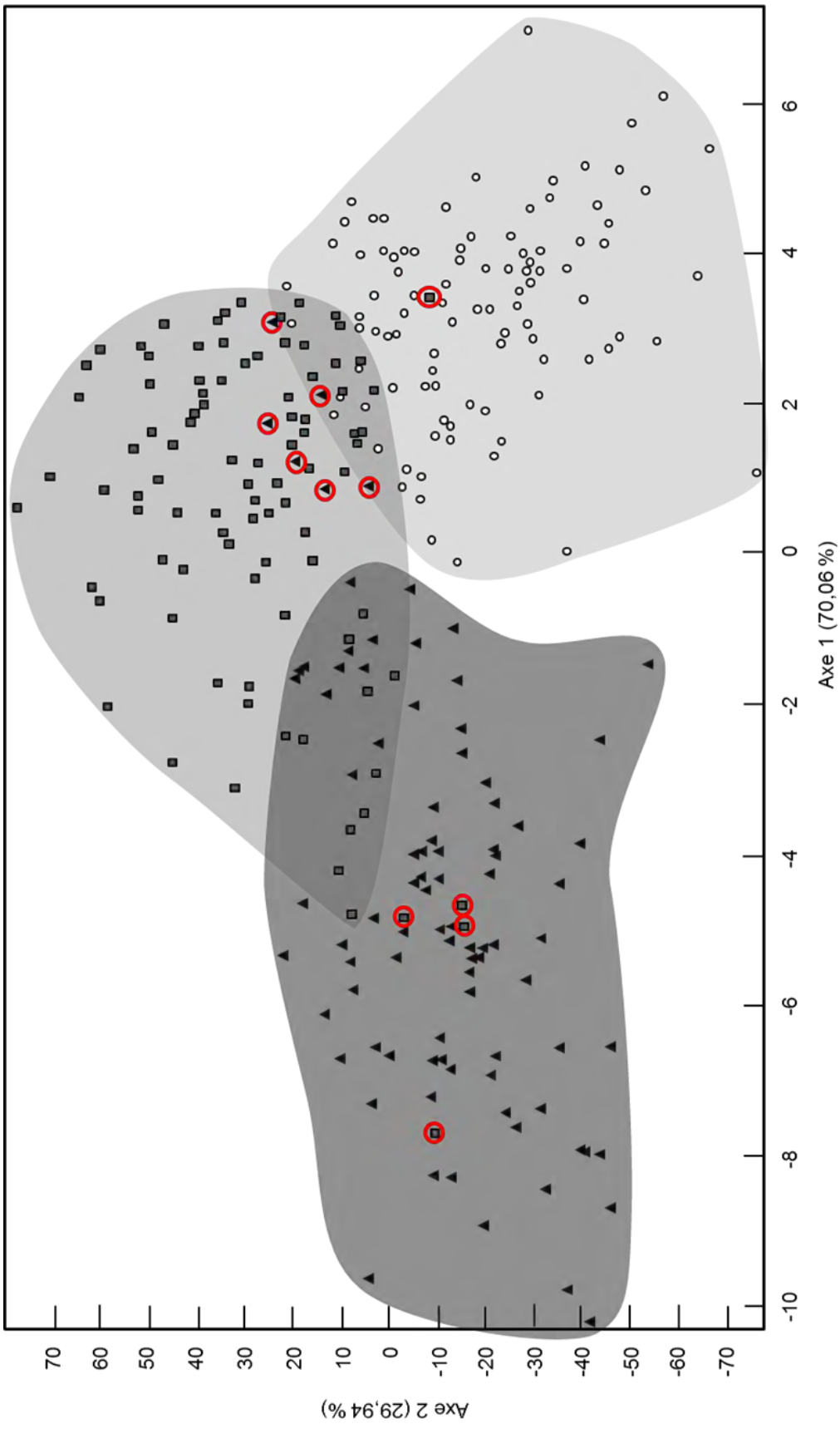


Figure 3 Graphical representation of the repartition of the three sampled subpopulations made with GENETIX software version 4.05.2 (Belkhir et al. 1998) using factorial correspondence analysis (FCA). Each grey square depicts an individual from location B, each black triangle from location A and each white circle from location C. Individuals that are situated in another cloud than the other individuals from that sampled plant patch are highlighted with a red circle.

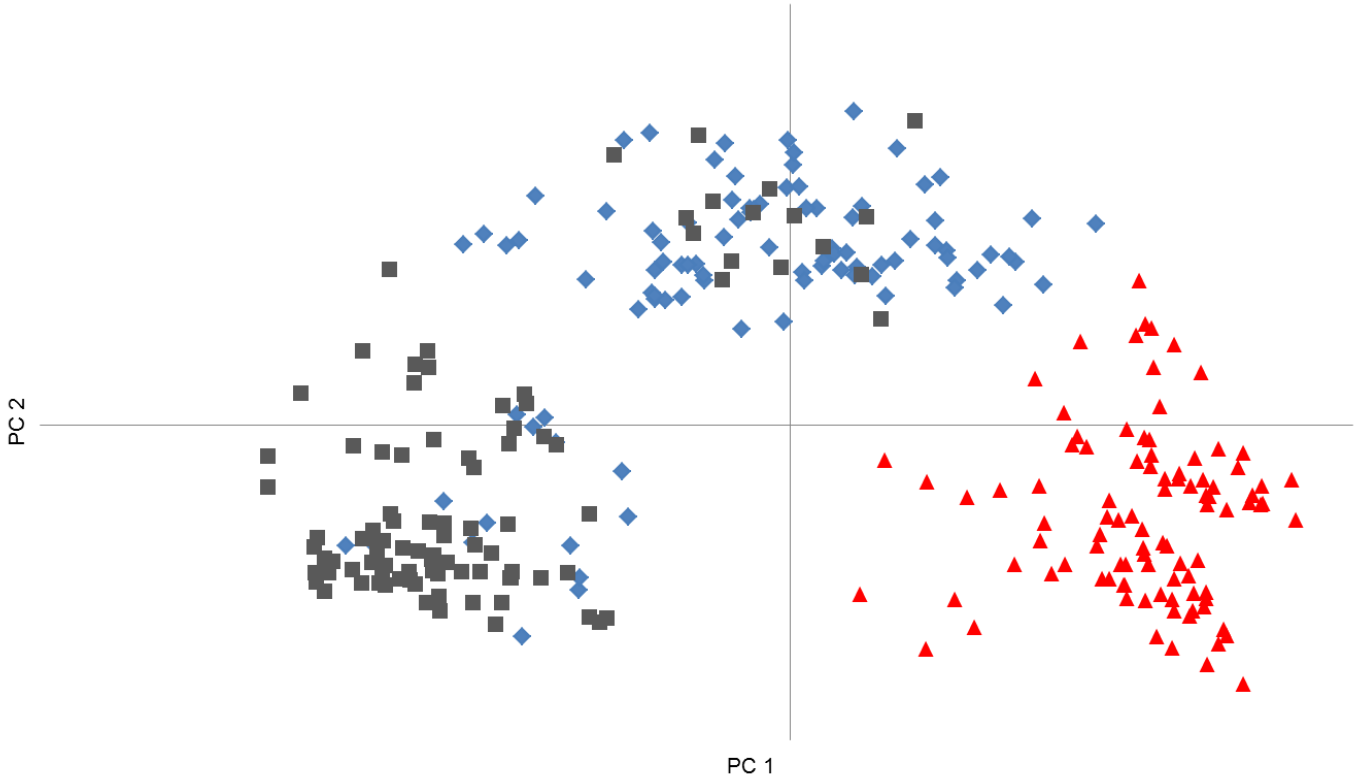


Figure 4 Graphical representation of the repartition of the three sampled subpopulations made with GENALEX software version 6.41 (Peakall and Smouse 2006) using principal component analysis (PCA). Each grey square depicts a sample/an individual from location A, each blue diamond an individual from location B and each red triangle an individual from location C.

Figure 2 shows for K = 3 for each plant patch sample the proportion of cluster membership (*q*) when prior location information is not used (Figure 2c), and when it is used to assist clustering (Figure 2d). Without prior location information the proportions of cluster membership are equally divided over each plant patch sample as mentioned earlier. When prior location information is included, the proportions of cluster membership are not equally divided over each plant patch sample, suggesting that there is genetic structure. All samples show an average estimated major membership proportion for the same genetic cluster suggesting that we sampled admixed subpopulations in which all individuals inherited most ancestry from the same ancestral population. Despite that, we can still see that the genetic profiles of each sample exhibit a different composition of cluster membership. Moreover, looking at the cluster membership of each individual from a certain plant patch (Figure 2b), we can see that the genetic profiles from individuals of one plant patch exhibit a different composition of subpopulation membership than genetic profiles from individuals of another plant patch, whether we assume three or four genetic clusters. Different compositions of genetic cluster membership of each plant patch sample at the subpopulation level and the individual level, both support the conclusion that the sampled plant patches harbour groups of flea beetles with distinct genetic divergence.

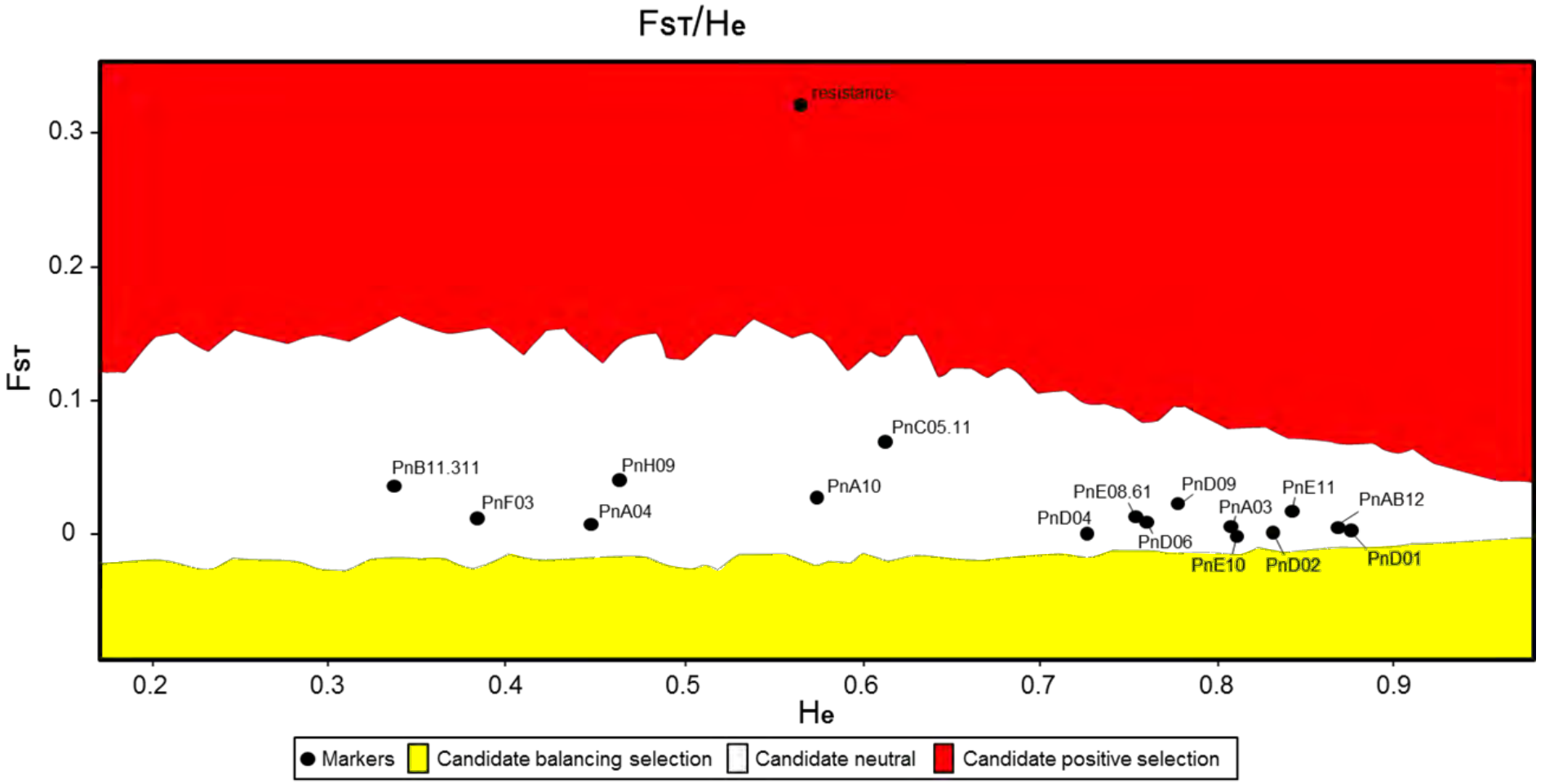


Figure 5 A comparison between the F_{ST} value and the expected heterozygosity (H_e) of each microsatellite locus to identify outliers using LOSITAN (Antao et al. 2008). In this case the F_{ST} values were simulated for the comparison between population A and population B (see text). Loci in the red area are candidates for directional selection and the ones in the white area are selectively neutral. None of the loci are positioned in the yellow area representing candidates for balancing selection.

Table 7 LOSITAN results of the simulated F_{ST} values per locus and the probability (*P*) that the simulated F_{ST} < sample F_{ST}, using 50 000 simulations, a 99% confidence interval and the infinite allele model. Loci with a probability higher than 0.99 (in bold) are highlighted with grey shading. "NA" stands for *not available*.

	Population A - B		Population A - C		Population B - C		All populations	
Locus	F_{ST}	*P*	F_{ST}	*P*	F_{ST}	*P*	F_{ST}	*P*
PnA03	0.0027	0.4765	0.0017	0.4039	0.0064	0.5302	0.0036	0.4106
PnA04	0.0041	0.6056	0.0068	0.6267	-0.0055	0.5278	0.0016	0.4847
PnA06	NA	NA	NA	NA	0.0117	0.633	NA	NA
PnA10	0.0261	0.7929	0.019	0.7134	0.0662	0.9498	0.037	0.9141
PnAB12	0.0005	0.3877	0.0171	0.7035	0.0361	0.9215	0.0179	0.8053
PnB11.311	0.0341	0.8476	0.0503	0.8845	-0.0002	0.5101	0.0313	0.8334
PnC05.11	0.0675	0.9557	0.0092	0.6189	0.023	0.7637	0.0342	0.8777
PnD01	0	0.3401	0.0131	0.633	0.0045	0.4728	0.0059	0.4639
PnD02	-0.0004	0.3975	0.0134	0.6432	0.0008	0.3984	0.0046	0.4438
PnD04	-0.0031	0.3867	0.0443	0.923	0.028	0.8471	0.0239	0.8448
PnD06	0.0050	0.5435	0.004	0.4748	0.0023	0.4811	0.0038	0.4421
PnD09	0.0191	0.7610	0.0185	0.721	-0.0008	0.3588	0.012	0.6402
PnE08.61	0.0099	0.6311	0.0202	0.7436	0.0043	0.4992	0.0114	0.6277
PnE10	-0.0038	0.2955	0.0032	0.4398	0.0009	0.4205	0.0001	0.2866
PnE11	0.0141	0.7039	0.0084	0.5527	0.0207	0.7706	0.0144	0.7128
PnF03	0.0088	0.6565	NA	NA	NA	NA	NA	NA
PnH09	0.0383	0.8676	NA	NA	NA	NA	NA	NA
Resistance	0.3218	**0.9998**	0.4021	**1**	0.8324	**1**	0.5760	**1**

The partitioning of individuals in the FCA and PCA (Figure 3 and 4) corresponded with the three samples as well. Factorial Correspondence axis (FC) 1 explained 70% of the variation in the FCA and FC 2 the remaining 30%. All three clouds of individuals from a location overlap partially, though not completely, so that three different groups can be distinguished that each roughly resemble a group of individuals from a certain plant patch. The first principal component (PC) in the PCA explained 33% of the variation and the second PC 26%. No individuals of locations A and B were partitioned in the cloud of individuals from location C. Neither did individuals from location C show up between individuals from locations A and B, while clouds of individuals from location A and B also contained individuals of the other sample location.

Thus, we have identified three distinct subpopulations which are roughly in agreement with the sampled plant patches. We therefore continued the analysis, while assuming that the three samples were genetically distinct subpopulations.

Table 8 BAYESCAN results of each locus F_{ST} and the posterior probability (P), showing that a model including selection for that locus fits the data better than a model without selection. Loci with a posterior probability higher than 0.91 (in bold) for at least one scenario are highlighted with grey shading. "NA" stands for *not available*.

	Population A - B		Population A - C		Population B - C		All populations	
Locus	F_{ST}	P	F_{ST}	P	F_{ST}	P	F_{ST}	P
PnA03	0.0230	0.1600	0.0222	0.4713	0.0246	0.4139	0.0132	0.6145
PnA04	0.0259	0.0888	0.0327	0.0964	0.0317	0.1622	0.0189	0.2146
PnA06	NA	NA	NA	NA	0.0347	0.0456	NA	NA
PnA10	0.0255	0.0730	0.0333	0.0642	0.0367	0.0754	0.0225	0.0524
PnAB12	0.0143	0.5947	0.0251	0.3525	0.0337	0.0626	0.0168	0.3617
PnB11.311	0.0258	0.0718	0.0538	0.2281	0.0336	0.1008	0.0227	0.0632
PnC05.11	0.0288	0.0966	0.0327	0.0886	0.0361	0.0782	0.0229	0.0760
PnD01	0.0240	0.1178	0.0321	0.0840	0.0316	0.1532	0.0160	0.4109
PnD02	0.0137	0.6107	0.0326	0.0646	0.0273	0.2975	0.0187	0.2204
PnD04	0.0260	0.0888	0.0608	0.4101	0.0351	0.0618	0.0232	0.0766
PnD06	0.0245	0.0924	0.0286	0.2088	0.0254	0.3577	0.0165	0.3403
PnD09	0.0263	0.0788	0.0317	0.1056	0.0285	0.2549	0.0169	0.3243
PnE08.61	0.0241	0.1064	0.0318	0.1030	0.0253	0.3823	0.0169	0.3331
PnE10	0.0166	0.4543	0.0325	0.0766	0.0338	0.0694	0.0173	0.3063
PnE11	0.0255	0.0592	0.0313	0.1216	0.0344	0.0582	0.0220	0.0526
PnF03	0.0241	0.1218	NA	NA	NA	NA	NA	NA
PnH09	0.0278	0.0900	NA	NA	NA	NA	NA	NA
resistance	0.0340	0.1714	0.2764	**1**	0.3808	**1**	0.3101	**1**

Genetic differentiation among subpopulations

Outlier detection. We used the program LOSITAN (Antao et al. 2008) for an F_{ST} outlier test, by comparing the 'neutral' mean F_{ST} with the heterozygosity in the loci. For both the infinite alleles model (IAM) and the stepwise mutation model (SMM) the resistance trait appeared as a potential outlier locus when using a 99% confidence interval with a probability of the actual F_{ST} being bigger than the simulated F_{ST} ($0.9998 < P < 1$)(see Table 7 and Figure 5 for the results under the IAM, the results under SMM were congruent). For each pairwise population comparison and also for all subpopulations together the resistance trait appeared as a potential outlier. In none of these comparisons the microsatellite loci used in the simulations were a candidate for selection.

The bayesian method was also used to test for outliers with the program BAYESCAN (Foll and Gaggiotti 2008). The posterior probability for each scenario did not exceed 0.91 for any microsatellite locus (see Table 8) which can be interpreted as no strong evidence for an effect of selection on those loci (Foll 2012). For the resistance trait the posterior probability

Table 9 Genetic differentiation estimators and 95% confidence interval (CI) per pair of subpopulations for each microsatellite and for the resistance phenotype.

pairwise comparison between locations A and B

microsatellite marker	PnA03	PnA04	PnA10	PnAB12	PnB11.311	PnC05.11	PnD01	PnD02	PnD04	PnD06	PnD09	PnE08.61	PnE10	PnE11	PnF03	PnH09	PnH12	resistance
G_{ST}	0.004	0.005	0.017	0.003	0.021	0.048	0.003	0.003	0.001	0.006	0.020	0.005	0.001	0.010	0.007	0.032	0.008	0.196
CI min	0	-0.002	0.002	0	0.005	0.008	-0.001	-0.001	-0.002	-0.001	-0.002	0.001	-0.002	0.002	-0.001	-0.008	0.001	0.114
CI max	0.011	0.021	0.038	0.009	0.041	0.108	0.009	0.008	0.005	0.014	0.056	0.018	0.005	0.018	0.045	0.099	0.017	0.289
G'_{ST}	0.037	0.014	0.058	0.038	0.040	0.167	0.041	0.028	0.005	0.040	0.131	0.034	0.008	0.100	0.016	0.086	0.037	0.430
CI min	0.003	-0.005	0.008	-0.007	0.010	0.029	-0.014	-0.010	-0.012	-0.005	-0.013	0.004	-0.019	0.026	-0.001	-0.015	0.005	0.266
CI max	0.091	0.054	0.123	0.110	0.081	0.347	0.126	0.081	0.031	0.095	0.321	0.112	0.050	0.186	0.045	0.271	0.079	0.588
D_{est}	0.033	0.008	0.043	0.035	0.020	0.127	0.038	0.026	0.004	0.034	0.114	0.029	0.007	0.092	0.009	0.057	0.030	0.295
CI min	0.003	-0.003	0.005	-0.006	0.005	0.021	-0.013	-0.010	-0.010	-0.005	-0.011	0.003	-0.017	0.024	-0.001	-0.008	0.004	0.174
CI max	0.081	0.033	0.091	0.102	0.043	0.272	0.117	0.074	0.026	0.081	0.282	0.097	0.045	0.171	0.025	0.201	0.063	0.423

pairwise comparison between locations A and C

microsatellite marker	PnA03	PnA04	PnA10	PnAB12	PnB11.311	PnC05.11	PnD01	PnD02	PnD04	PnD06	PnD09	PnE08.61	PnE10	PnE11	PnH12	resistance
G_{ST}	0.003	0.007	0.012	0.011	0.028	0.008	0.010	0.009	0.025	0.005	0.020	0.012	0.004	0.007	0.014	0.252
CI min	0	-0.002	0.004	0.003	0.010	-0.003	0.003	0.004	0.016	-0.001	-0.002	0.005	0	0	0.003	0.197
CI max	0.009	0.022	0.023	0.023	0.050	0.030	0.019	0.018	0.035	0.014	0.058	0.021	0.011	0.017	0.027	0.313
G'_{ST}	0.034	0.017	0.058	0.117	0.053	0.028	0.122	0.094	0.187	0.036	0.127	0.092	0.039	0.070	0.054	0.413
CI min	-0.001	-0.006	0.020	0.035	0.020	-0.009	0.039	0.036	0.123	-0.006	-0.015	0.035	-0.004	0.002	0.012	0.311
CI max	0.087	0.058	0.107	0.221	0.099	0.103	0.236	0.172	0.257	0.102	0.332	0.158	0.105	0.151	0.104	0.522
D_{est}	0.031	0.011	0.046	0.107	0.026	0.020	0.114	0.085	0.167	0.032	0.110	0.081	0.035	0.064	0.040	0.217
CI min	-0.001	-0.004	0.016	0.032	0.010	-0.006	0.036	0.032	0.109	-0.005	-0.013	0.030	-0.003	0.002	0.009	0.142
CI max	0.079	0.037	0.086	0.202	0.049	0.075	0.222	0.158	0.230	0.088	0.291	0.140	0.094	0.137	0.078	0.305

pairwise comparison between locations B and C

microsatellite marker

	PnA03	PnA04	PnA06	PnA10	PnAB12	PnB11.311	PnC05.11	PnD01	PnD02	PnD04	PnD06	PnD09	PnE08.61	PnE10	PnE11	PnH12	resistance
G_{ST}	0.006	0	0.010	0.037	0.021	0.002	0.023	0.005	0.003	0.016	0.004	0.003	0.005	0.003	0.013	0.007	0.717
CI min	0.001	-0.003	0.003	0.018	0.010	-0.002	-0.002	0	-0.001	0.009	-0.002	-0.002	0	-0.001	0.004	0	0.643
CI max	0.013	0.009	0.018	0.060	0.035	0.010	0.059	0.012	0.009	0.027	0.016	0.011	0.013	0.009	0.025	0.018	0.793
G'_{ST}	0.055	0.001	0.087	0.135	0.225	0.003	0.077	0.063	0.031	0.135	0.029	0.031	0.043	0.029	0.116	0.027	0.944
CI min	0.008	-0.008	0.027	0.073	0.120	-0.003	-0.007	0.001	-0.009	0.074	-0.014	-0.017	-0.004	-0.007	0.031	0.001	0.907
CI max	0.123	0.025	0.160	0.207	0.343	0.015	0.197	0.149	0.085	0.211	0.106	0.110	0.104	0.079	0.215	0.056	0.974
D_{est}	0.050	0.001	0.078	0.103	0.209	0.001	0.056	0.058	0.028	0.121	0.025	0.028	0.038	0.026	0.105	0.020	0.808
CI min	0.007	-0.006	0.024	0.054	0.110	-0.001	-0.005	0.001	-0.008	0.066	-0.012	-0.015	-0.003	-0.006	0.028	0.001	0.740
CI max	0.110	0.018	0.145	0.159	0.320	0.005	0.152	0.138	0.077	0.190	0.091	0.099	0.092	0.071	0.195	0.050	0.872

was 1 for the global analysis over all three populations and for the pairwise analysis of populations A and C and populations B and C. Only for the pairwise analysis of populations A and B the posterior probability was not high enough to infer evidence of selection. The inclusion of alpha in the model did not matter for any microsatellite locus nor for any pairwise comparison between subpopulations. For the resistance trait, however, alpha was estimated to be 3.05 for all populations together and 2.86 for populations B and C and 2.41 for populations A and C, all indicating that directional selection might be present.

Additionally, genetic differentiation between each pair of subpopulations was estimated by calculating G_{ST} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005) and D_{est} (Jost 2008) and their 95% Confidence Interval (CI) for each microsatellite and the resistance trait using the program SMOGD (Crawford 2010)(Table 9).

For the pairwise comparison between sampled subpopulations B and C (host plant *S. arvensis* versus *B. vulgaris* G-type, respectively) the resistance trait shows for all estimated parameters a substantially higher value (G_{ST} = 0.717, G'_{ST} = 0.944 and D_{est} = 0.808) than for each microsatellite marker (G_{ST} = 0 - 0.037, G'_{ST} = 0 - 0.225 and D_{est} = 0 - 0.209). In all cases the estimated parameter of the resistance trait even exceeds the upper level of the 95% confidence interval of the microsatellite estimated parameter.

For the pairwise comparison between sampled subpopulations A and C (host plant *S. arvensis* versus *B. vulgaris* G-type) the difference between estimators of the resistance trait and the microsatellites is not as pronounced as above. For G_{ST} the resistance trait shows an estimate substantially higher than the confidence interval of this estimator for each microsatellite marker. For G'_{ST} and D the resistance trait shows an estimate higher than all estimates for the microsatellites, but their confidence interval does not always come up with an upper limit lower than the estimator D of the resistance trait.

For the pairwise comparison between sampled subpopulations A and B (both patches with host plant *S. arvensis*) the resistance trait shows for all estimators a higher value than the upper limit of the confidence interval of this estimator for each microsatellite marker. Estimators of the resistance trait's genetic differentiation (G_{ST} = 0.196, G'_{ST} = 0.289 and D_{est} = 0.295) are substantially higher than the estimators of the microsatellites' genetic differentiation (G_{ST} = 0 - 0.048, G'_{ST} = 0 - 0.167 and D_{est} = 0 - 0.127).

DISCUSSION

The objective of this study was to provide support for an explanation for the geographical distribution of the resistant flea beetle phenotype. This distribution may theoretically be explained by either high population structure with little dispersal or by a cost to resistance on other host plants than *B. vulgaris* G-type. Because resistance is highly favoured on *B. vulgaris* G-type, we expect that for comparisons between populations on *B. vulgaris* G-type and another host plant genetic differentiation of the resistance trait is larger than for neutral

markers. If coadapted gene complexes are indeed present in resistant beetles and break up through recombination during outbreeding with other local flea beetle populations, then we expect a larger genetic differentiation for the resistance trait than expected for neutral markers, not only for comparisons between *B. vulgaris* G-type and another host plant, but also between other host plants at different geographical distances from a G-type patch.

The outlier detection methods showed in all comparisons between populations on G-type and on *S. arvensis* that resistance emerged as a candidate for a locus that is subject to selection. This was expected because there is strong selection favouring resistance on the *B. vulgaris* G-type. To also find an indication that resistance is an outlier in the comparison between two *P. nemorum* populations on *S. arvensis*, is more surprising, and indicates that also differences in resistance allele frequencies between populations on *S. arvensis* are influenced by local selection, rather than a restriction to dispersal. Whereas the genetic differentiation of microsatellites between the two *S. arvensis* patches is low (D_{est} < 0.1), nearly 70% of the flea beetles on plant patch A were resistant compared to only 30% on plant patch B. So there seems to be directional selection on the resistance trait even on other host plants than *B. vulgaris* G-type.

Earlier work with neutral allozymes had already shown a weak, though significant, genetic differentiation between the samples under the influence of geographical distance (de Jong et al. 2001). In addition, it was shown that the differentiation was not only influenced by geographical distance, but also by the resistance phenotype of the beetles (de Jong et al. 2009). Our results are consistent with the allozyme analysis by de Jong et al. (2009), for weak but significant differentiation between flea beetles collected on different plant patches was found as well as a high level of gene flow between the plant patches. The populations on different plant patches do not seem sufficiently differentiated to explain the distribution of the resistance trait, because the microsatellites show only weak differentiation, whereas differences in percentage of resistant flea beetles between populations on different plant patches are substantial. The level of genetic differentiation of neutral loci may be accounted for by genome-wide effects such as random drift and gene flow. When the level of genetic differentiation is substantially larger for the resistance trait than for neutral loci, then this could reflect locus-specific effects such as directional selection or a difference in mutation variation (Balloux and Lugon-Moulin 2002; Hedrick 2005; Noor and Feder 2006). However, with a high rate of gene flow between the subpopulations, there is not a strong effect of mutation rate (Beaumont and Nichols 1996; Balloux and Lugon-Moulin 2002; Hedrick 2005). Certain measurements for differentiation, such as G'_{ST}, also take a possible effect of different mutation rates into account. As high mutation rates would lead to a more different allele composition between subpopulations, differentiation measurements like Jost's *D* (Jost 2008) would be higher for markers with a high mutation rate compared to markers with a lower mutation rate (in contrast to estimators such as G_{ST} that decrease with high mutation rates (Hedrick 1999; Balloux and Lugon-Moulin 2002). Because of the presumably high mutation rate of microsatellites, it could very well differ from the mutation rate of the genetic region causing resistance in the flea beetle. So a difference in mutation rate of the neutral loci compared to the candidate gene may have affected our results. It has not caused an underestimation of

genetic differentiation between subpopulations though. In this case, taking different mutation rates into account or not, leads to the same conclusion: the resistant trait behaves as an outlier. By using multiple differentiation measurements (including mutation models or not) and acknowledging the substantial amount of gene flow between subpopulations, we believe that differences in mutation rate did not influence our findings in any substantial way.

Selection on the resistance trait seems to differ in strength as well as direction between subpopulations. While selection strongly favours resistance on *B. vulgaris* G-type, the ratio of resistant beetles decreases substantially with distance from the *B. vulgaris* G-type patches and the resistance trait behaves as an outlier for other host plant comparisons. Differences in percentage of resistant flea beetles between populations on the same host plant patches are substantial. It seems that there is a cost associated with the resistance trait, so that selection acts against this specific genotype on other host plants than the *B. vulgaris* G-type. This would limit the spread of resistance, even when dispersal is not limited. The difference in strength and direction of selection fits with the hypothesis of the breaking up of coadapted gene complexes (de Jong and Nielsen 2002). When gene complexes break up, the resistance trait becomes less favourable and, therefore, selection will act against it. This is likely to occur at different rates in different subpopulations, leading to a stronger signal of negative selection in subpopulations further away from *B. vulgaris* G-type. That is why we would find a larger genetic differentiation for the resistance trait than expected for neutral markers for comparisons between other host plants at different geographical distances from a G-type patch.

Furthermore, we found higher values of the estimated parameters G_{ST}, G'_{ST}, and D for the resistance trait than for the microsatellites. When we looked at the subpopulations on *S. arvensis*, none of the 95% confidence intervals constructed for the estimators of differentiation among neutral loci had an upper level exceeding the estimator of differentiation for the resistance trait. We cannot conclude from this that the resistance trait is under selection, because these estimators each have their drawbacks. G'_{ST}, and D are not comparable between loci when mutation rates differ substantially. G_{ST} is influenced by high within population heterozygosity (H_s). When H_s is high (as is often the case with microsatellites because they may contain many alleles), G_{ST} estimators can only produce low values, whether there is high differentiation between subpopulations or not.

There are some limitations of this study that should be taken into account. One of the limitations is that the genotype of the resistant and susceptible beetles was not known because resistance is a dominant marker. Therefore, homozygous and heterozygous genotypes cannot be distinguished by our bioassays. When heterozygosity is used as a measure for differentiation, the ratio of heterozygotes and homozygotes is needed. To estimate the genotype of the collected beetles, we assumed this locus to be in HWE like most of the sampled markers in the subpopulations. Of course we know that loci under selection – whether diversifying or balancing selection – are not in HWE according to the neutral theory. We believe that using estimated genotypes for resistance does not affect our conclusions. If the resistant beetles would include more homozygous beetles than estimated using HWE, then differences

between the resistant trait and neutral loci would only become larger. Also showing that the resistance trait is an outlier even while HWE was used to estimate genotype frequencies, shows that between the subpopulations the allele distribution is not what one would expect for neutral loci. The estimated genotypes for beetles from location C (*B. vulgaris* G-type) are all homozygous, even under HWE, because no susceptible beetles were found at the site. This is supported by data from previous studies where nearly all beetles were found to be homozygous for the resistance trait in specific crosses with susceptible beetles to determine their genotype (de Jong et al. 2000).

Different assumptions are made when applying the population genomics approach. One is assuming the neutral theory of molecular evolution (Kimura 1983) to find selection signatures. This approach is widely used (see e.g. review Butlin 2010), but should be done taking into account other evolutionary processes that can cause a departure from predictions (Charlesworth et al. 2003). In our case, we tested for neutrality, recent bottlenecks, inbreeding and linkage before using STRUCTURE and looking for outliers and we used a stringent criterion for the comparisons in LOSITAN and BAYESCAN. In the population genomic approach, multiple comparisons are made which can more easily lead to false positives than a single comparison. Therefore, usually different outlier detection methods are used as well as corrections such as the Bonferroni method (eg. the false discovery rate correction (FDR) of *P*-values by Benjamini and Hochberg (1995). Our findings are robust against variation in the methods.

The present study forms an example of the use of the population genomics approach as proposed in CHAPTER 2, i.e. rather than using it to identify parts of the genome that are likely to be under selection without a priori knowledge of function of these genomic regions, now we used this approach with a candidate for a genetic factor under selection to answer ecological questions (in this case, about the cause of the geographic distribution of resistance).

This appears to be a suitable method to eventually identify coevolutionary hotspots in a geographic context, as proposed by Vermeer et al. in CHAPTER 2. The results warrant further work on the detection of the postulated coadapted gene complexes in the flea beetles. This local genetic coadaptation might be a more general mechanism underlying local ecological adaptation. A potentially fruitful direction for future research is to identify the detailed molecular mechanism of resistance. This might lead to an understanding of why this resistance trait experiences negative selection on other host plants than the *B. vulgaris* G-type.

Acknowledgements

We would like to thank Patrick Nosil for critically reading of, and constructively commenting on, an earlier version of this manuscript. We would also like to thank John Thompson for inspiring us to work with population genomics/genetics to unravel some aspects of his comprehensive geographic mosaic theory of coevolution. We also would like to show our gratitude to Ulrich B. Madsen, Erikholm Skovdistrikt (Kværkeby) for permission to work on the property.

Data Accessibility

DNA sequences: Genbank accesions; FJ217683, FJ217684, FJ217685, FJ217686, FJ217687, FJ217688, FJ217689, FJ217690, FJ217691, DQ507814, FJ217694, FJ217695, FJ217696, FJ217697.

CHAPTER 6

Evaluation of a candidate resistance gene underlying flea beetle resistance to host plant defence with a population genomics approach

Kim M. C. A. Vermeer • Patrick Verbaarschot • Marcel Dicke • Peter W. de Jong

Abstract

The flea beetle *Phyllotreta nemorum* is polymorphic with respect to resistance to the defence of its host plant, winter cress (*Barbarea vulgaris* G-type). We investigate if a candidate gene for resistance correlates with the beetle's phenotype of resistance and by using a population genomics approach whether there is evidence for it being under selection. Because only resistant flea beetles can survive on *B. vulgaris* G-type, we expect strong selection on the part of the genome associated with resistance. If the candidate gene is indeed involved in resistance, then this part of the genome is expected to show evidence of selection in a population genomics analysis as we found in CHAPTER 5 for the resistance phenotype. Fourteen polymorphic microsatellite loci were used with the candidate locus for resistance in a neutrality test to identify outliers that departed from neutral expectations. The results only show evidence of directional selection on this candidate locus for some scenarios, despite the resistance trait being expected to be subject to directional selection for all scenarios. This leads us to the conclusion that it is unlikely that this candidate gene is directly responsible for the flea beetle resistance to host plant defence.

INTRODUCTION

Plants have developed various defence strategies against herbivorous insects, including chemical defence. Chemical defence is mediated by compounds called allelochemicals or secondary metabolites (Fraenkel 1953), the latter because they are not known to play any substantial role in the primary metabolism of the organisms. Despite being called secondary metabolites, however, many of them do play a substantial role in the survival of the plant by providing a first line of defence against herbivores. A well-known example involves the production of glucosinolates and their break-down products in Cruciferae (Brassicaceae) (e.g. Feeny 1977; Nielsen 1977; Chew 1988; Cipollini et al. 2003) that reduce the performance of many generalist herbivores on the plant, because of their toxicity and/or repellence. For specialists, on the other hand, these glucosinolates can stimulate feeding and/or oviposition (Thorsteinson 1953; Fraenkel 1959; Whittaker and Feeny 1971; Nielsen 1978; Renwick 2002; van Leur et al. 2006). These specialists have the advantage that, by specialising on a specific plant species, they are in a niche that is not readily available to non-adapted herbivores. The plant can, however, subsequentially develop a second line of defence that makes them again resistant towards these specialist insects (Feeny 1977; Nielsen 1978; Sachdevgupta et al. 1993a; Sachdevgupta et al. 1993b). An example is that of the wintercress (*Barbarea vulgaris* R. Br.) (Brassicales: Brassicaceae) which is chemically defended against the crucifer specialist diamondback moth (DBM) (*Plutella xylostella* L.) by feeding-deterrent saponins (Shinoda et al. 2002; Agerbirk et al. 2003a). The DBM is a serious pest of various crucifers (Talekar and Shelton 1993). It only feeds on crucifers, where glucosinolates act as feeding and oviposition stimulants to this moth (Thorsteinson 1953; Nayar and Thorsteinson 1963). Saponins are believed to be a second line of defence, because of their restricted distribution in the Cruciferae. Of course, such a defence will only be effective until the insect evolves another mechanism to overcome the plant's second line of defence. This principle is also known as an evolutionary 'arms race' between plant and insect (Edmunds 1974, but see; Bernays and Graham 1988).

Wintercress is the only crucifer believed to contain this type of repellent saponins, and not even all types of *B. vulgaris* do. There are several varieties of *B. vulgaris*. *Barbarea vulgaris* var. *variegata* is not fed on by DBM (Serizawa et al. 2001). *Barbarea vulgaris* var. *arcuata* (Opiz) Simkovics consists of two types distinguishable by several characters such as morphology (Nielsen 1997b; Agerbirk et al. 2001; Agerbirk et al. 2003b; Ørgaard and Linde-Laursen 2007; 2008; Dalby-Brown et al. 2011): a 'P-type' with pubescent leaves and a 'G-type' with glabrous leaves. The P-type is susceptible to DBM and lacks the saponins that are believed to prevent certain herbivores from feeding, while the G-type is chemically defended against these herbivores by the production of specific saponins (Agerbirk et al. 2003a; Kuzina et al. 2009). Furthermore, the saponin levels in the G-type vary during the year. This seasonal variation is correlated with DBM resistance, for in summer high concentrations prevent DBM from feeding while later in the season saponin levels drop and DBM mortality is strongly reduced when feeding on *B. vulgaris* var. *arcuata* G-type (Agerbirk et al. 2003a). The selection pressure of the defence of *B. vulgaris* G-type is, therefore, not always present and it might therefore take a substantial amount of time for DBM to evolve resistance, especially in the presence of alternative food sources.

Interestingly, another insect has been capable of overcoming the chemical defence of *B. vulgaris* var. *arcuata* G-type. The large striped flea beetle (*Phyllotreta nemorum* L.) (Coleoptera: Chrysomelidae), another pest insect of crucifers, has populations that can completely develop on the G-type of this plant variety (Nielsen 1996; de Jong et al. 2000). This means that also during summer, when saponin levels are high and make the plant completely unacceptable for DBM, the flea beetle is capable of feeding on the plant. This is, however, not the case for all *P. nemorum* flea beetles (hereafter indicated by merely "flea beetles"). They are polymorphic with respect to their ability to use *B. vulgaris* as a host plant. Only flea beetles with so called R-gene(s) (R for resistance Nielsen 1997a) can and do feed on *B. vulgaris* G-type (de Jong and Nielsen 1999; Breuker et al. 2005; Nielsen and de Jong 2005). The majority of the flea beetles seems to lack R-genes and show the same kind of susceptibility towards *B. vulgaris* G-type plants as DBM; during summer, when saponin levels are high, susceptible flea beetles do not feed on G-type plants while later on in the season saponin levels drop and both DBM and susceptible flea beetles show less evidence of avoiding *B. vulgaris* G-type (Nielsen 1997b; Agerbirk et al. 2003a).

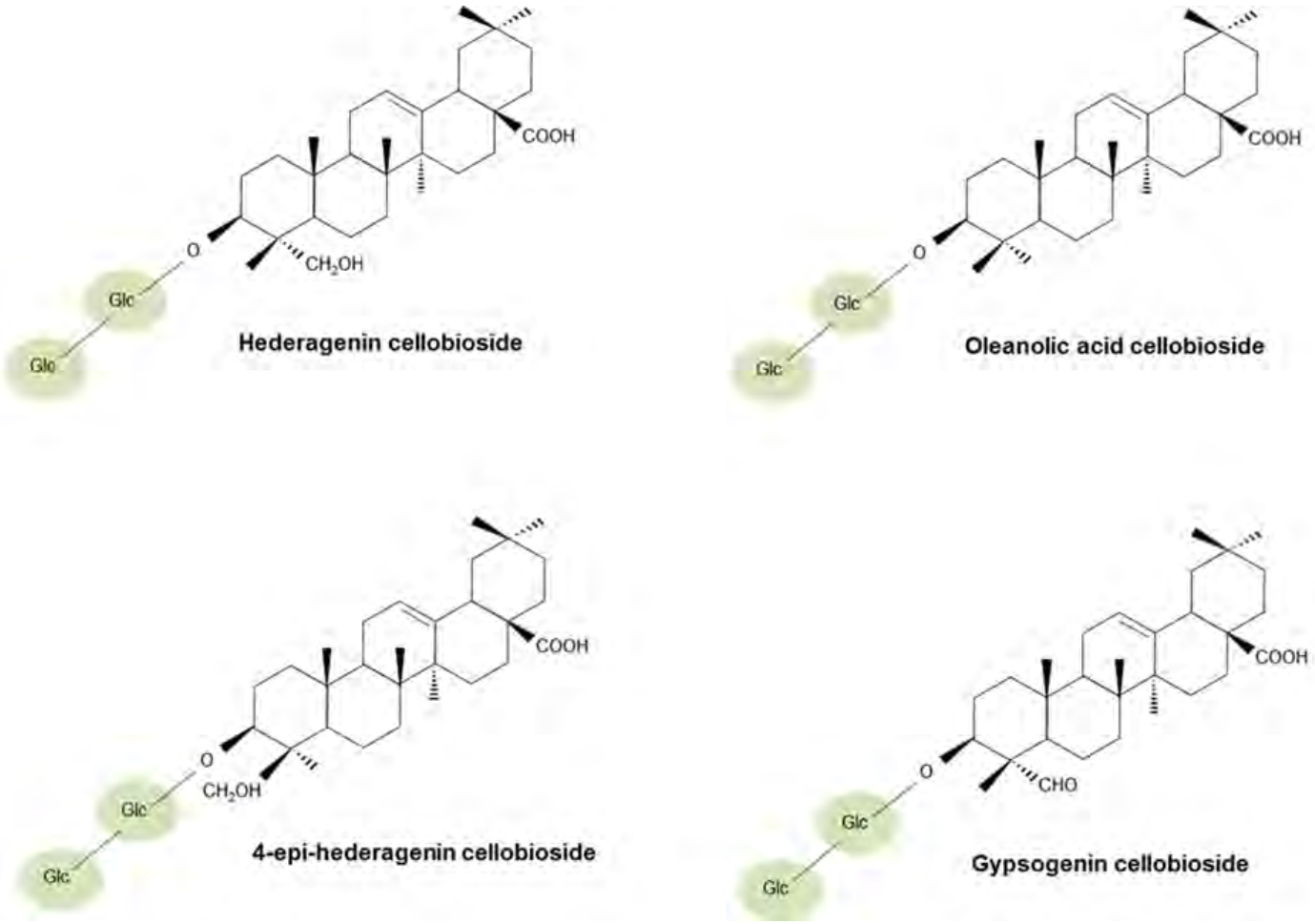


Figure 1 Chemical structure of four triterpenoid saponins in *Barbarea vulgaris* G-type that correlate with resistance to the large striped flea beetle. Glc = glucose.

Understanding the resistance mechanism in the flea beetle might prove helpful in controlling damage on crops by flea beetles, but also closely related species that are considered serious pests of both wild and cultivated *Brassica* species (Andersen et al. 2006). *Phyllotreta nemorum* is closely related to other pest insects such as *P. undulata*, *P. striolata* and *P. cruciferae.* Even control of less related pest insects like the DBM might benefit from knowledge of the development of resistance of *P. nemorum* towards the plant defences. DBM responds the same way to the defence of *Barbarea vulgaris* as the susceptible flea beetle. Thus, they might develop the same resistance mechanism as the flea beetle that is able to develop on G-type *Barbarea*. Despite the difference in selection pressure imposed by natural plant defences and artificial insecticides, cross-resistance between a chemical insecticide and a natural defence chemical of a plant can happen (see for examples Despres et al. 2007, e.g. Feyereisen 1999 and Feyereisen 2005). It is therefore important to understand the resistance mechanism of the insect, not only to insecticides, but also to natural chemical plant defences (Despres et al. 2007).

Saponins are known to function as defence compounds against fungi and herbivores (Osbourn 1996; Weissenberg et al. 1998; Sparg et al. 2004). Because of their effectiveness they are also suggested as natural insecticides or repellents (de Geyter et al. 2007). In crucifers saponins are not common: so far they have only been found in *Barbarea vulgaris*, *Barbarea verna* (Badenes-Perez et al. 2011) and *Brassica napus* (Barron-Yanez et al. 2008). Saponins are believed to be responsible for the defense in *Barbarea vulgaris* G-type, because several triterpenoid saponins in *B. vulgaris* correlated significantly with resistance against flea beetles (Shinoda et al. 2002; Agerbirk et al. 2003a; Kuzina et al. 2009; Nielsen et al. 2010b), namely hederagenin-3-O-(4-O-β-D-glucopyranosyl)-β-D-glucopyranoside (hederagenin cellobioside), oleanolic acid-3-O-(4-O-β-D-glucopyranosyl)-β-D-glucopyranoside (oleanolic acid cellobioside), 3-O-cellobiosyl-4-epihederagenin (4-epihederagenin cellobioside) and 3-O-cellobiosyl-gypsogeni (gypsogenin cellobioside) (Figure 1). Furthermore, hederagenin cellobioside, the most abundant of these saponins in G-type *B. vulgaris* (Kuzina et al. 2009), is known to inhibit feeding of *P. nemorum* individuals that are susceptible to the defence of *B. vulgaris* G-type (Nielsen et al. 2010a). Food consumption of resistant flea beetles was also influenced by hederagenin cellobioside, but to a significantly lesser extent. This suggests that resistant flea beetles are somehow capable of avoiding the effect imposed by this saponin. Of fungi it is known that they are capable of infecting saponin-containing plants by various toxicity-avoiding mechanisms (Osbourn 1996). One of the mechanisms is producing saponin detoxifying enzymes that can remove sugars from the saponin (e.g. Crombie et al. 1986; Wubben et al. 1996; Morrissey et al. 2000). This hydrolysis of the glycosidic bonds in the saponins might increase or decrease the biological activity of the saponin or its aglycone (Crombie et al. 1986; Nisius 1988; Bowyer et al. 1995; Osbourn et al. 1996; Adel et al. 2000; Morrissey et al. 2000; Bouarab et al. 2002). How the flea beetle is capable of rendering the saponins harmless is unknown, but the fact that hederagenin, the aglycone of hederagenin cellobioside (see Figure 2 for the chemical structure), does not inhibit the beetle's feeding (Nielsen et al. 2010a) suggests that the flea beetle may use detoxifying enzymes similar to those in fungi. Further support for this hypothesis was found with α-hederin which only differs from hederagenin cellobioside in the sugar chain. α-

Hederin had as much influence on the consumption rate of susceptible beetles as it had on the consumption rate of resistant beetles, while hederagenin cellobioside inhibited feeding of resistant beetles significantly less than feeding of susceptible ones. This could be explained by specific enzymatic activity targeting the cellobioside sugar chain by resistant beetles. The partial inhibitory effect of hederagenin cellobioside for resistant beetles might be caused by higher rates of intake than degradation, leading partially to the same effect as found for the susceptible individuals. The removal of only one sugar unit from hederagenin cellobioside leads to hederagenin-monoglucoside and glucose. This monoglucoside has the same effect

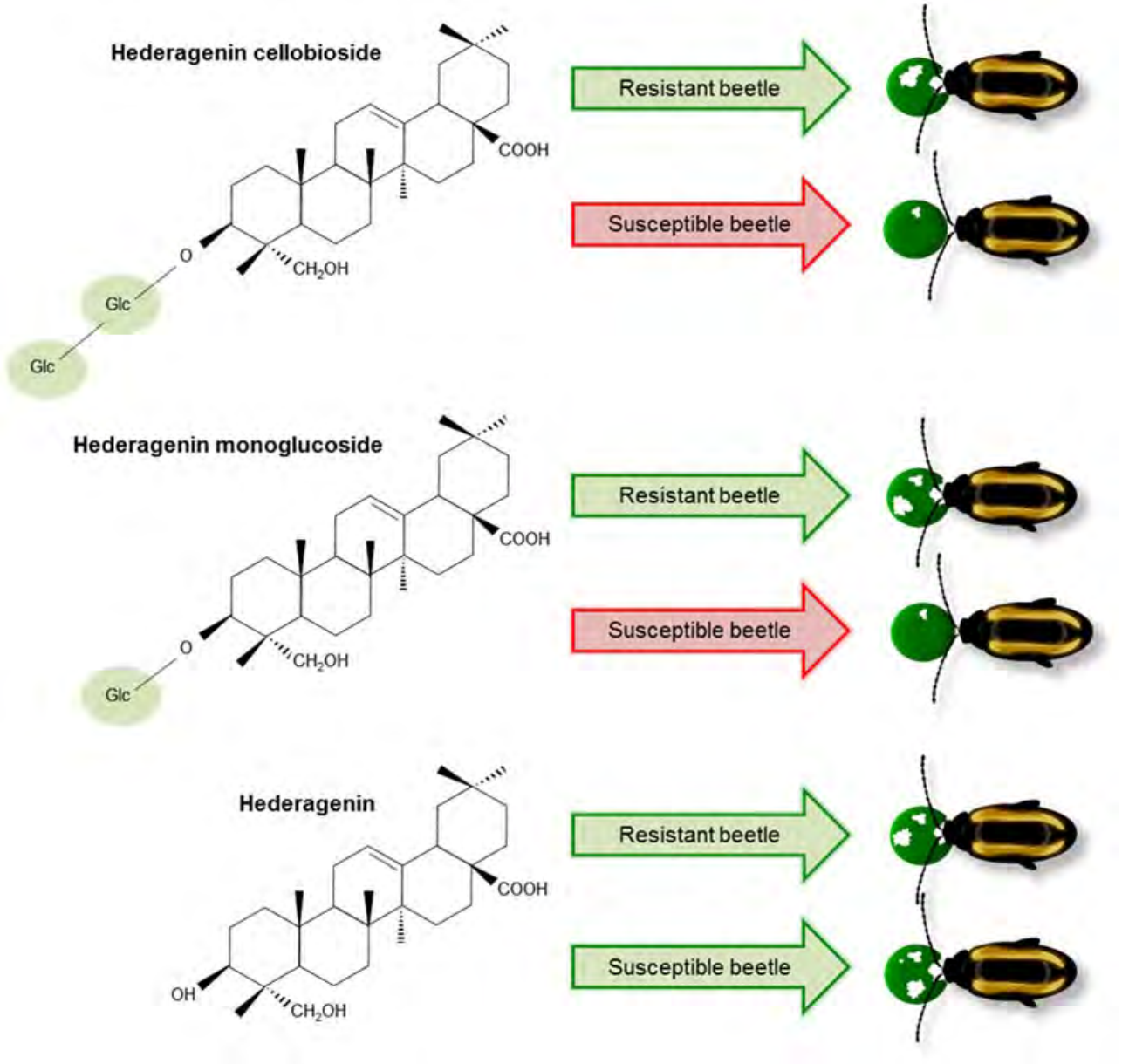


Figure 2 Schematic representation of the large striped flea beetle and its response to saponin hederagenin cellobioside, hederagenin monoglycoside and the aglycone, hederagenin: Green arrows represent the flea beetle's acceptance and the red arrow represents the flea beetle's refusal to feed on plants dipped in the respective compounds. Resistant beetles are beetles capable of developing on *Barbarea vulgaris* G-type despite its plant defences and susceptible beetles are not capable to do so.

on the flea beetle as hederagenin cellobioside; it inhibits feeding less in resistant beetles than in susceptible individuals when used in the same concentrations as hederagenin cellobioside. The concentration of hederagenin cellobioside used in the study of Nielsen et al. 2010a to investigate the effect on flea beetle feeding rate, occurs naturally in *B. vulgaris* (Shinoda et al. 2002; Kuzina et al. 2009; Augustin et al. 2012).

The hydrolysis of glucose units is catalysed by enzymes named β-glucosidases (β-D-glucoside glucohydrolases, EC 3.2.1.21). It is this type of enzymes that fungi use to detoxify the saponins in their host plant (Faure 2002; Morant et al. 2008). An example of an enzyme removing the β-1,4-linked glucose from a saponin is avenacinase. The fungus *Gaeumannomyces graminis* var. *avenae* produces this β-glucosidase which removes the glucose from oat root saponin A-1 making it less toxic to the fungus (Turner 1961; Bowyer et al. 1995). Also *P. nemorum* expresses mRNA of several β-glucosidases (see chapter 3 of this thesis). At least one of these β-glucosidases was capable of degrading hederagenin cellobioside to hederagenin-monoglucoside and glucose when expressed in a cell line. Unfortunately there was not enough evidence on whether subsequently the second sugar unit was hydrolysed and the inactive aglycone was produced. We do not know if hydrolyzing hederagenin cellobioside to hederagenin needs the intermediate step of first hydrolyzing one glucose unit to be able to hydrolyze the second one or that both glucose units can be removed at once by an enzyme. So, even though the flea beetle produces β-glucosidases capable of hydrolyzing the saponin held responsible for the resistance of *B. vulgaris* G-type, we do not know if it also influences the resistance trait. If resistant beetles are capable of deactivating their host plant's saponins, then this is the first known case of an insect using saponin-degrading enzymes to be able to feed on a plant that contains a saponin as defence compound.

Most research on the flea beetles and their resistance to *B. vulgaris* G-type has focussed on populations in Denmark. In populations near Kværkeby the resistance trait in flea beetles seems more straightforward than in other populations in Denmark. Where in other populations also minor genes and sex-linked genes seem to be involved (Nielsen 1997a; de Jong and Nielsen 1999), near Kværkeby only major R-genes confer resistance to the flea beetle (de Jong et al. 2000, but see de Jong and Nielsen 1999). This means that one dominant resistant allele is sufficient to enable the beetle to feed and develop on *B. vulgaris* G-type. The populations from Kværkeby seemed therefore most appropriate to investigate if genes coding for β-glucosidase play a role in the resistance of flea beetles to saponins. Near Kværkeby, populations of flea beetles are found on *B. vulgaris* G-type and other host plants. On *B. vulgaris* G-type virtually all beetles are resistant (de Jong et al. 2000; CHAPTER 4). Segregation patterns of individuals sampled from *B. vulgaris* G-type plants suggested that all were homozygous resistant at one autosomal locus (de Jong et al. 2000). The proportion of resistant beetles in populations on other host plants was surprisingly low, taken genetic differentiation into account (CHAPTER 4). Allozyme studies showed little genetic differentiation between flea beetles on different plant patches ($\theta = 0.009$ at a distance of 100 m to 1 km) (de Jong et al. 2001). Not only geographic distance between plant patches resulted in slight genetic differentiation, but also the resistance phenotype of the beetle (de Jong et al. 2009). So little genetic differentiation is found by neutral markers between subpopulations near

Kværkeby, but for the R-gene genetic differentiation seems to differ more profoundly among subpopulations.

Negative pleiotropy of the R-gene might explain the geographical distribution of the R-alleles. Beside positive selection on *B. vulgaris* G-type, the resistance trait might also be subject to negative selection on other host plants. Despite substantial gene flow, the spread of resistance would then be limited by selection against the trait. Previously, no reduction in fitness was found for resistant beetles compared to susceptible beetles (de Jong et al. 2000) because crosses between two heterozygous beetles would result in a 1:2:1 ratio of RR:Rr:rr genotypes. When the R-allele was crossed into a susceptible line from another location, however, crosses between two heterozygous individuals resulted in a skewed ratio of resistant:susceptible offspring (close to 1:1 instead of the expected 3:1) (Breuker et al. 2007). Earlier, similar results were found for autosomal R-alleles originating from a population in Ejby (Denmark) (de Jong and Nielsen 2000) when the R-allele was crossed into a susceptible line. After several backcrosses the offspring of a cross between two heterozygous beetles suffered high mortality among the homozygous resistant beetles (RR genotype). It seems that the R-allele itself might have a negative effect on the flea beetle unless accompanied with positive modifier genes in the genetic background (de Jong and Nielsen 2002). If these positive modifiers are not tightly linked to the R-gene, offspring can lose these positive modifiers after outcrossing, resulting in lower fitness of beetles carrying the R-allele.

Results from CHAPTER 5 suggest that negative pleiotropic effects of the R-gene may be present in natural populations near Kværkeby. The resistance phenotype was assayed for flea beetles in populations on *B. vulgaris* G-type and on the non-defended host plant *Sinapis arvensis*, a common host plant of *P. nemorum*. Populations on *S. arvensis* were sampled on patches situated at different distances from *B. vulgaris* G-type. As strong selection favours the resistance trait on the *B. vulgaris* G-type patch but not on other host plant patches, we expect larger genetic differentiation for the resistance trait between populations on a defended host plant and a non-defended host plant, than for neutral parts of the genome that only experience genome-wide effects such as gene flow. If, by outbreeding on non-defended host plants farther apart from *B. vulgaris* G-type, the linkage between positive modifier genes and the R-gene gets disrupted, we expect to find more genetic differentiation for the resistance trait than for neutral parts of the genome between populations on non-defended host plants at different distances from *B. vulgaris* G-type. Indeed, a higher level of genetic differentiation was found for the resistance trait than for neutral parts of the genome between a population on *B. vulgaris* G-type and populations on *S. arvensis*. Also between populations on *S. arvensis* a higher, though less pronounced, level of genetic differentiation was found for the resistance trait than for neutral parts of the genome.

The method of distinguishing between genome-wide effects and locus-specific effects by sampling variable loci in the genome of numerous individuals among populations is called a population genomics approach (Black et al. 2001). Black et al. (2001) mentioned population genomics for the first time as an approach for plant-insect interactions to investigate adaptation in insects. Differential selection on an adaptive trait can lead to large allele frequency

differences between subpopulations at the loci that control the involved trait. Because genome-wide effects are more or less the same for every locus, the variance of the estimated values for genetic differentiation is expected not to differ much between neutral loci. Divergent selection, on the other hand, can lead to an unusually high level of genetic differentiation for a locus that is involved in an adaptive trait (and the closely linked area on the genome). When using the population genomics approach, many loci are sampled throughout the genome of individuals derived from different (sub)populations, to draw a sample distribution of genetic differentiation between the (sub)populations at these loci. Significant outliers in the distribution of estimates for genetic differentiation, suggest selection on these outlier loci.

In this study, a variant of the population genomics approach is employed to discover whether our candidate gene for the resistance trait in *P. nemorum*, a gene coding for a β-glucosidase, is likely to be under selection. In CHAPTER 3, three different β-glucosidase sequences were amplified from a cDNA pool synthesized from resistant beetles, coined β-glucosidase A, B and C. Two primer sets were used to amplify these β-glucosidase cDNAs from susceptible and resistant flea beetle lines, one primer set for β-glucosidase A and one for β-glucosidase B and C. No β-glucosidase C cDNA was found in the susceptible flea beetle line, suggesting that this variant is specifically expressed in resistant beetles. β-glucosidase C cDNA was built in insect cell lines and the expressed recombinant protein was able to break down hederagenin cellobioside to hederagenin-monoglucoside and glucose. The DNA sequences of β-glucosidase B and C are highly homologous (94%) and these two sequences may represent two genes or two alleles of the same gene. As the resistance trait follows a Mendelian inheritance pattern, the sequence for β-glucosidase C might correspond to the dominant R-allele and the sequence for β-glucosidase B to the r-allele (susceptible allele). Here, we refer to the β-glucosidase B and C sequences as the candidate gene, the putative PneR-gene (PneR as abbreviation of *P*hyllotreta *ne*morum *R*esistance). To test if these β-glucosidase B and C sequences indeed correspond to a resistance gene, we compared the presence of either β-glucosidase B or C in the flea beetle with the resistant phenotype.

The following hypotheses were tested: 1) if the candidate gene is responsible for resistance, then we expect a correlation between the resistance phenotype and the putative candidate gene alleles. 2) if selection favours resistance of the flea beetle on *B. vulgaris* G-type, then we expect to find a higher level of genetic differentiation for the candidate gene than for neutral markers in comparisons of populations of flea beetles feeding on different host plants.

To this end, we sampled flea beetles on *B. vulgaris* G-type and *S. arvensis* and assayed their resistance phenotype, scored the presence of sequences coding for β-glucosidase B and C, and amplified microsatellites. Microsatellites are repeating short sequence motifs located throughout the genome, mostly outside of coding sequences (Jarne and Lagoda 1996). They are co-dominant, highly variable in the number of repeats, making them powerful markers, especially when genetic differentiation is measured between subpopulations with small differentiation between them (Hughes and Queller 1993; Parker et al. 1998).

MATERIAL AND METHODS

The same collection of beetles was used as in chapter 5: in 2009, on the 11th and 26th of June, leaves with leaf-mining flea beetle larvae were collected on *Sinapis arvensis* and *Barbarea vulgaris* G-type nearby Kværkeby (Denmark) from three plant patches (see CHAPTER 5 for exact locations and description of the area and the patches). Location A consists of an *S. arvensis* plant patch, as does location B. Plant patch A is located respectively 2.5 and 2.3 km away from locations B and C. Location C consists of a *B. vulgaris* G-type plant patch and is located 3.4 km away from location B. In CHAPTER 5 we found a low level of differentiation among the three subpopulations and showed that the sampled locations roughly agreed with the subpopulations derived from the genetic data. Collected larvae were reared as in CHAPTER 5 and used in bioassays as adult beetles within 24 h of emergence from the soil. Each beetle was sexed before being used in a no-choice test bioassay with *B. vulgaris* G-type to determine their resistance phenotype. Individuals were sexed using a stereo microscope (25 x magnification) (de Jong and Nielsen 1999). The sex ratio for each subpopulation was approximately evenly balanced between males and females (Chi square test, DF = 1, $P > 0.05$). For each subpopulation, 96 individuals were tested in a bioassay. The bioassays were set up as described in CHAPTER 5. The beetles had no access to food prior to the bioassay.

Table 1 Forward and reverse primer sequences used for candidate gene (PneR) and microsatellite amplification.

target	forward primer (5'-3')	reverse primer (5'-3')
PneR (β-gluB)	CAGGTAGCCTAGATGAGGATGAC	CCAATAACATTGACGCCATCGTC
PneR (β-gluC)	AGGCTATGATCAGGATTCCGAAG	CCAATAACATTGACGCCATCGTC
PnA03	CAACGAGCAATCGATACAATTCG	ACATTCTGCGCCGAGATTGG
PnA04	AATTACGAGAGCAACATGTCGG	ACTGTTGCTGTTGGGTTTGC
PnA06	CCAGAAATGTCATCGTACA	GTTTGGTCTTTGTGATGGACAAGG
PnA10	ACTCACGCCCGAATCGCTTC	GTTTAGAATGGACATGGTCGGCGG
PnAB12	GAGATTGAGACGATTGCTGGG	CTCAACAGTTGCATTACCAGC
PnB11.311	CTTCGAAATAATCGTCTTC	GTTTAATCTGGAGACGATGATGAC
PnC05.11	TCACCAAATTGTGACATGTACC	GTTTACCATAAACGCACTGTTGA
PnD04	CTCGAGCTTGACTCACTACTGC	CCAGTTCCAGTGATTCGAGC
PnD06	ATCACGTTCGGCACCACCTG	GTTTCTTCAGCAGCCTGATGGGC
PnD09	GCTCCAACTACACCAAACTCG	TCGCGTACCCGTAATAGTGG
PnE08.61	GCAGCAGGTCGAGGCGACTG	GTTTATTCGCCACCGTACCGTTCG
PnE11	GTACAGTCATGTCTTGGAACGC	CTCGATTGCGTAGTAGCCGG
PnF03	AAATCCTTCAAAGGCTAAGCCAGC	CAACGGTTCAGCAGCAACG
PnH09	CGTGAGGCTTGTAGTATTTGG	CTACCATCCGATGATGAACG

Non-feeders were classified as susceptible beetles and feeding beetles were classified as resistant beetles. If we could not distinguish between feeding and non-feeding, the beetle was scored as questionable (<1.5%).

DNA isolation and amplification

We isolated DNA from each beetle by using DNeasy® Blood & Tissue Kit (QIAGEN) and the Purification of Total DNA from Animal Tissues protocol (Spin-Column) (See CHAPTER 5 for details). This method resulted in isolated DNA for each of the bio-assayed 96 beetles per sampled plant patch.

PneR-gene. Primers were designed based on two cDNA variants of β-glucosidase found in CHAPTER 3, namely β-glucosidase B and β-glucosidase C (hereafter abbreviated as β-gluB and β-gluC) (see Table 1 for the primer sequences). Only one reverse primer was developed because it was sufficient to get both desired DNA fragments. The two forward primers (FB and FC) and one reverse primer (RBC) were used for a polymerase chain reaction (PCR) to amplify the DNA fragment corresponding to β-gluB and β-gluC. DNA samples were amplified in a 25 µl PCR mix containing 15.375 µl nuclease-free water, 5 µl 5X green GoTaq® buffer, 1.5 µl 25 mM $MgCl_2$, 0.5 µl 200 µM dNTp's, 0.5 µl 500nM of each primer, 0.125 µl GoTaq® DNA polymerase, 1 µl of DNA (at least 4ng of DNA in water). PCR conditions were as follows: initial denaturation at 94°C for 3 minutes, then 35 cycles of 94°C for 30 seconds, an annealing temperature of 57°C for 30 seconds, 72°C for 1 minute, and a final extension at 72°C for 5 minutes. PCR products were loaded on a 1% w/v agarose gel stained with ethidium bromide in TAE buffer. Each gel was run by electrophoresis for 90 minutes and then fragment sizes were manually scored by UV light visualization. DNA amplification by primer FB and RBC resulted in a fragment of 810-825 bp and amplification by primer FC and RBC resulted in a fragment of 310-325 bp.

Microsatellites. Microsatellites are more variable than other markers such as AFLPs and therefore better suited for genetic differentiation measurements for populations with small genetic differentiation between them (Hughes and Queller 1993; Parker et al. 1998). 14 already available polymorphic microsatellite primer pairs (see Table 1, CHAPTER 5 and Verbaarschot et al. 2007) were used for the amplification of microsatellite loci using polymerase chain reaction (PCR) as described in CHAPTER 5. GenBank accession numbers of the microsatellites are DQ507809, DQ507810, DQ507812-DQ507817, FJ217683-FJ217689, FJ217693, FJ217694, FJ217696, FJ217697. In order to draw a sample distribution of genetic differentiation estimates between subpopulations for neutral loci, we used microsatellites whose allele frequencies did not deviate from Hardy-Weinberg Equilibrium (HWE) (see CHAPTER 5 for a description of the method). Microsatellites PnF03 and PnH09 showed a heterozygote deficit for sample location C and microsatellite PnA06 showed a heterozygote deficit for sample location A. These microsatellite markers were only used in analyses dealing with subpopulations for which they did not deviate significantly from HWE. All other used microsatellites in further analyses behaved as neutral loci in all subpopulations. Also was there no evidence of linkage disequilibrium between any of the used microsatellites (see al-

so CHAPTER 5). Tubes containing the isolated DNA were coded and mixed over 3 plates so that no sample site bias would occur per 96 processed samples. The PCR products were purified and run on an ABI PRISM 3700 sequencer (Applied Biosystems) (as in CHAPTER 5). Results were compiled by manually scoring the size of each microsatellite (fragment) on an electrophoregram visualized with GeneMarker® V1.5 software (company: Softgenetics®, LLC).

Data analyses

Microsatellites, resistance phenotype, and the presence of the β-gluB and β-gluC variant were assessed for 96 beetles per location. Per individual the phenotype derived from the bioassays was compared with the genotype of the candidate gene (PneR). Only individuals of which both the PneR genotype and resistance phenotype were known were used in the comparison.

The PneR-gene was tested for heterozygote deficiency. Expected heterozygosity of the PneR-gene if β-gluB and β-gluC are its alleles and in HWE, was calculated for each subpopulation and we used a chi-square goodness-of-fit test to assess deviations of the observed frequencies of β-gluB and β-gluC from expected frequencies under HWE. Because the counts of some genotypes were less than five, a Yates's correction is usually advised to compensate for high chi-square values. We deemed a Yates's correction unnecessary because Chi-square values were as low as not to indicate significance.

From the resistance phenotype contrived by the bioassays, we calculated the genotype using the Hardy-Weinberg Principle, to also use the resistance genotype in the population genomics approach (see below).

Outlier detection

We tested for outlier loci in the data set composed of microsatellites, the resistance trait and the candidate gene using two different methods, the frequentist method (fdist) and the Bayesian method (Beaumont and Nichols 1996; Beaumont and Balding 2004). Multiple analyses were done under different mutation models for each of the two methods for robustness of the results. The first method we used was the fdist method (Beaumont and Nichols 1996) as implemented in LOSITAN (Antao et al. 2008), running 50 000 simulations under either the infinite allele model (IAM) or the stepwise mutation model (SMM) to test for outlier loci by evaluating the relationship between the $F_{ST,}$ using the estimator θ as in (Weir and Cockerham 1984) and expected H_e for each locus. LOSITAN first estimates a 'neutral' mean F_{ST} for each pair of subpopulations by removing F_{ST} values of potential outliers in this calculation. We set the confidence interval to 95 and 99% when computing the mean F_{ST} value. Then a second round of simulations was done with the 'neutral' mean F_{ST}, and loci whose empirical F_{ST} estimates fell outside the 97.5% quantiles of the simulated distribution were deemed outliers.

We used BayeScan version 2.1 (available from website http://www.cmpg.unibe.ch/software/bayescan/) (Foll and Gaggiotti 2008) for the Bayesian method (Balding 2003; Beaumont and Balding 2004). The program assumes that allele frequencies within a subpopulation follow a Dirichlet distribution (Foll and Gaggiotti 2008). BayeScan runs two models for each locus using a logistic regression; one model that contains only a population-specific component (β) shared by all loci, and one model that also includes a locus-specific component (α) that is shared by all populations. When for a locus, α differs significantly from 0 and thus is needed to estimate F_{ST}, that locus is likely experiencing an effect of selection. For each locus a posterior probability was estimated for both models using a reversible-jump MCMC algorithm. For each run 50 000 burn-in simulations were used, followed by 50 000 repeats. Posterior model probabilities were computed for three pairwise subpopulation comparisons and one comparison involving all three subpopulations together. Because the program uses a Bayesian probability, the posterior probability should not be read as a *p*-value. Instead, by using the "Bayes factor" we can say how likely the model with selection is for that locus. As a result, when the posterior probability of the model with selection for a locus exceeds 0.76, this already suggests evidence for selection, and a posterior probability of more than 0.99 can be interpreted as decisive evidence for selection (Foll 2012).

Genetic differentiation between two subpopulations was further estimated for each microsatellite, the resistance trait and the candidate gene by calculating G_{ST} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005) and *D* (Jost 2008), using SMOGD (Crawford 2010). Microsatellites that showed an excess of homozygotes in a particular subpopulation were left out of the pairwise comparison involving that subpopulation, because we are only interested in neutral microsatellite loci. Both estimators of G_{ST}, G'_{ST} and Jost's *D* are given by SMOGD as well as their 95% confidence interval after 500 bootstrap replicates done for each marker separately. Also the heterozygosities within subpopulations (H_S) and over all subpopulations (H_T) are estimated by SMOGD. G_{ST} and G'_{ST} are actually not genetic differentiation measurements (see (Jost 2008)), but are still used as indicators of standard differentiation among subpopulations (Ryman and Leimar 2009; Whitlock 2011) because they also tell us something about evolutionary or demographic processes that lead to the measured variation between subpopulations. The measurement G_{ST} given by SMOGD is Nei and Chesser's improved version of Wright's F_{ST} (Wright 1951; Nei 1973), accounting for multiple loci with multiple alleles and deviation from HWE within subpopulations and the sample size of the subpopulations (Nei and Chesser 1983). A drawback of using G_{ST} to measure differentiation among subpopulations is that the value of G_{ST} will always be very small with high levels of heterozygosity, even if not many alleles are shared by compared subpopulations. Because microsatellites can have high levels of heterozygosity, the value of G_{ST} cannot be directly compared with that of bi-allelic markers that usually show a much lower level of heterozygosity. G'_{ST} is a standardized measurement of G_{ST} which is less dependent on the level of heterozygosity of a marker. Leng and Zhang (2011) advise to calculate both G_{ST} and Jost's *D* and the relative levels of heterozygosity. Jost's *D* is calculated with the effective number of alleles rather than heterozygosities, so *D* is independent of average within-subpopulation heterozygosity. Furthermore, the estimator *D* is also dependent on the mutation rate of the marker and very sensitive to mutation model assumptions, making it less favourable when used to compare

Table 2 Resistance phenotype of collected flea beetles. *R* is a resistant phenotype and *r* a susceptible phenotype. Behind each phenotype the number of beetles with that phenotype is given from that location.

Location	plant type	phenotype	#
A	*Sinapis arvensis*	R	66
		r	29
B	*Sinapis arvensis*	R	28
		r	65
C	*Barbarea vulgaris* G-type	R	96
		r	0

different types of markers among each other as we do here. All three differentiation measurements can still be used though, when taken into account their limitations, but one has to be careful when trying to interpret these measurements.

RESULTS

Bioassays

Nearly 70% of the beetles (66 individuals) from plant patch A (*Sinapis arvensis* plants) were classified as resistant and from plant patch B (*Sinapis arvensis* plants) only 30% (28 individuals) (Table 2). All beetles from plant patch C (*B. vulgaris* G-type plants) were classified as being resistant. Of one beetle from plant patch A we could not with certainty determine the phenotype and from plant patch B this was the case for 3 beetles. All other beetles were classified as susceptible beetles.

Observed and expected genotype at the PneR-locus, if β-gluB and β-gluC are regarded as two alleles of the same gene, are given in Table 3. For each subpopulation the goodness-of-fit test shows no significant deviation of the observed frequencies from the expected frequencies of the genotypes assuming Hardy-Weinberg Equilibrium (X^2 test, $p >> 0.05$). So considering β-gluB and β-gluC alleles of one gene, there is no deviation of the candidate gene from HWE in any subpopulation.

Comparison with resistance phenotype

Per individual, the resistance phenotype derived from the bioassays was compared with the genotype derived from the amplified β-glucosidase sequences. All individuals from location C were resistant and they all contained β-gluC. So, for this location there was a 100% correlation between resistance phenotype and the possession of β-gluC. However, when data of the other two locations were included, there was no longer any correlation between genotype and phenotype (Figure 3). For location A only 24 individuals had a resistant phenotype

Table 3 Number of beetles in each subpopulation with observed genotype and number of beetles in each subpopulation with expected genotype of the PneR-gene if β-gluB and β-gluC are regarded as two alleles of the same gene following Hardy-Weinberg expectations.

subpopulation	genotype	# observed genotypes	# expected genotypes	X^2
A	only β-gluC present	73	72.41	0.32
	both present	19	20.19	
	only β-gluB present	2	1.41	
B	only β-gluC present	60	59.21	0.24
	both present	30	31.58	
	only β-gluB present	5	4.21	
C	only β-gluC present	92	92.04	0.04
	both present	4	3.92	
	only β-gluB present	0	0.04	

and β-gluC, 2 individuals had a susceptible phenotype and no β-gluC, while 67 beetles were susceptible but also contained β-gluC. For location B only 26 individuals were scored as resistant and containing β-gluC, 3 individuals were scored as susceptible and no β-gluC was present, while 61 beetles were scored as susceptible but also containing β-gluC. Moreover, also resistant individuals were found that did not contain β-gluC, suggesting that β-gluC is not per se needed to confer the resistance genotype.

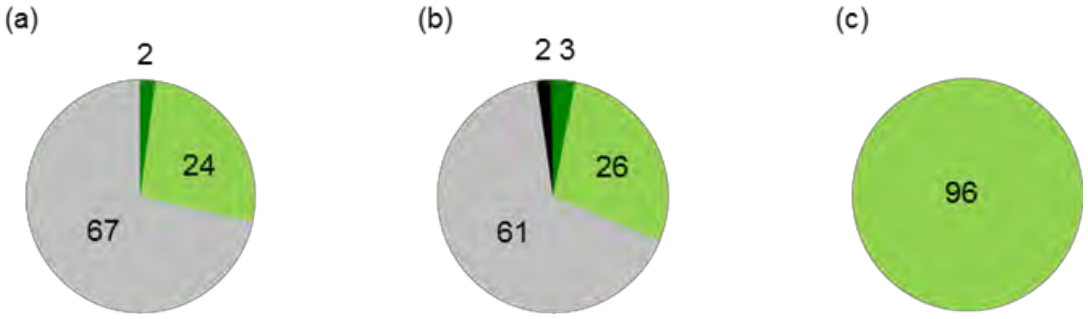


Figure 3 Comparison between resistance phenotype and genotype assigned by the PneR primers for three sampled locations; (a) samples from location A, (b) samples from location B, and (c) samples from location C. Dark green represents the fraction of sampled beetles that had a susceptible phenotype and no β-gluC, light green represents the fraction of sampled beetles that had a resistant phenotype and β-gluC, grey represents the fraction of sampled beetles that had the susceptible phenotype but also β-gluC, and black represents the fraction of sampled beetles that had the resistant phenotype but no β-gluC. Numbers represent the number of individual beetles per category.

Outlier detection

We used the program LOSITAN (Antao et al. 2008) to identify putative loci under selection. LOSITAN compares the 'neutral' mean F_{ST} with the expected level of heterozygosity (H_e) at the loci. We report results from the stepwise mutation model (SMM); the results obtained using the infinite alleles model (IAM) were congruent. The resistance trait appeared as an outlier in all of the subpopulation comparisons (Figure 4), for both a confidence interval of 95 and 99% used to find the simulated F_{ST}. The PneR-gene was deemed an outlier only for the pairwise comparison between subpopulations B and C and only when a confidence interval of 95% was used. None of the microsatellite loci emerged as a candidate for being a locus under selection.

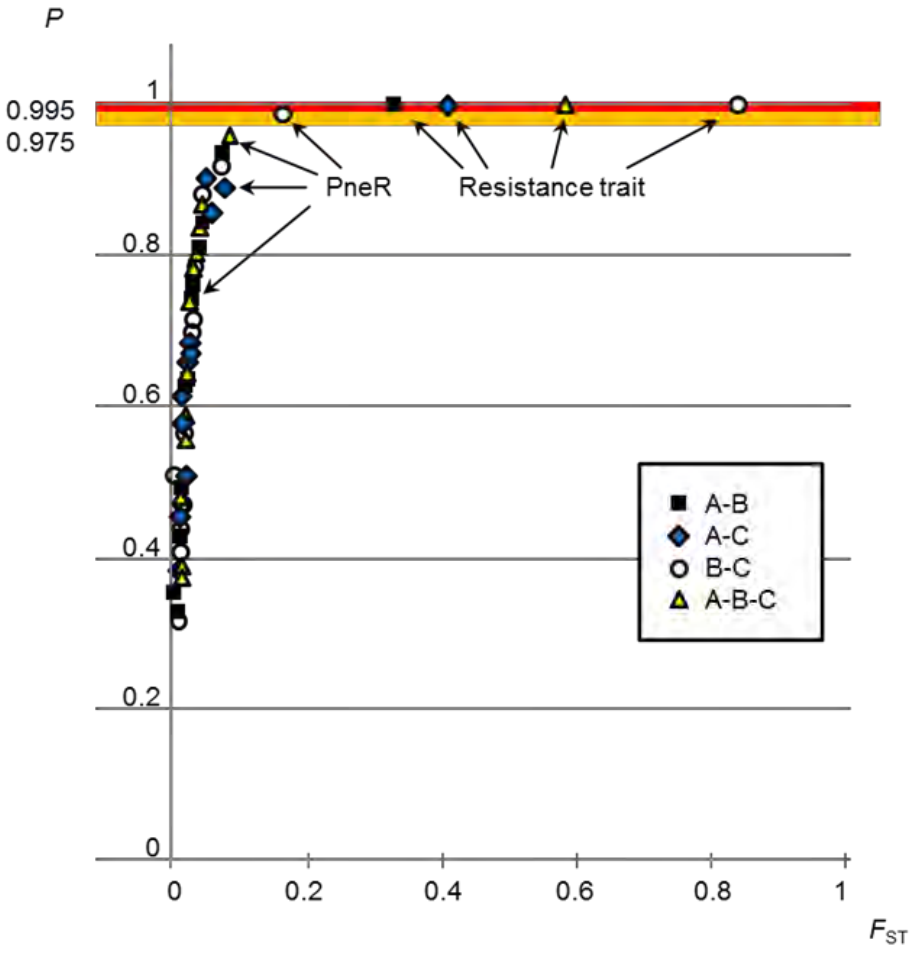


Figure 4 Results of outlier analysis using LOSITAN (Antao et al. 2008), where separate analyses were run for each of the four subpopulation comparisons. F_{ST} values and the probability (P) that the simulated neutral F_{ST} is lower than the empirical F_{ST} of a locus were simulated for pairwise comparisons between subpopulations A and B (black square), subpopulations A and C (blue diamond), subpopulations B and C (blank circle), and all three subpopulations (yellow triangle). Loci in the red area have a probability of more than 0.995 and are candidates for directional selection. The locus in the orange area has a probability of more than 0.975 and is also a candidate for directional selection when a confidence interval of 95% is used. With a confidence interval of 99% only the resistance trait was detected as outlier in any comparison.

The program BayeScan (Foll and Gaggiotti 2008) was used for an F_{ST} outlier test according to the Bayesian method (see Figure 5). This method yielded different results than the method of Beaumont and Nichols (fdist), as it exhibited decisive evidence for the resistance trait to be an outlier (P > 0.99) in all comparisons but not for the pairwise comparison between subpopulations A and B (comparison between subpopulations on *S. arvensis*). For the pairwise analysis of populations A and B, the posterior probability was not indicating evidence

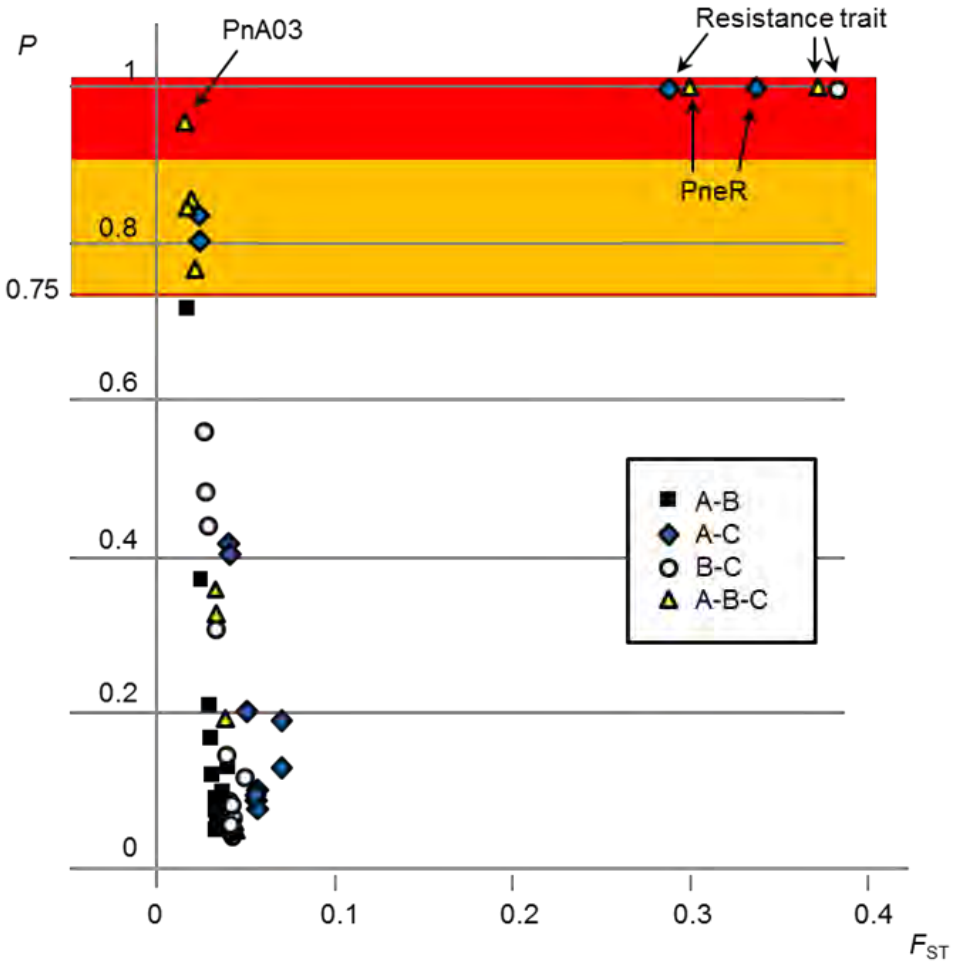


Figure 5 Results of outlier analysis using BayeScan (Foll and Gaggiotti 2008), where separate analyses were run for each of the four subpopulation comparisons. Pairwise comparisons were made between subpopulations A and B (black square), subpopulations A and C (blue diamond), subpopulations B and C (blank circle), and all three subpopulations (yellow triangle). F_{ST} values and the posterior probability (*P*) that a model including selection fits better than a model without were calculated using the Bayesian method. Loci in the coloured area have a posterior model probability high enough to suggest selection ($P > 0.76$). Loci in the red area have a posterior model probability high enough to suggest strong evidence for selection ($P > 0.91$). Both the resistant trait and the PneR-gene were detected as candidate loci for selection in the comparison between subpopulations A and C and the overall subpopulation comparison. The resistant trait was also deemed a candidate for being under selection for the pairwise comparison between subpopulations B and C. Some of the other loci also have a posterior model probability high enough to suggest selection, though the low F_{ST} values for these suggest that it doesn't involve directional selection but balancing selection.

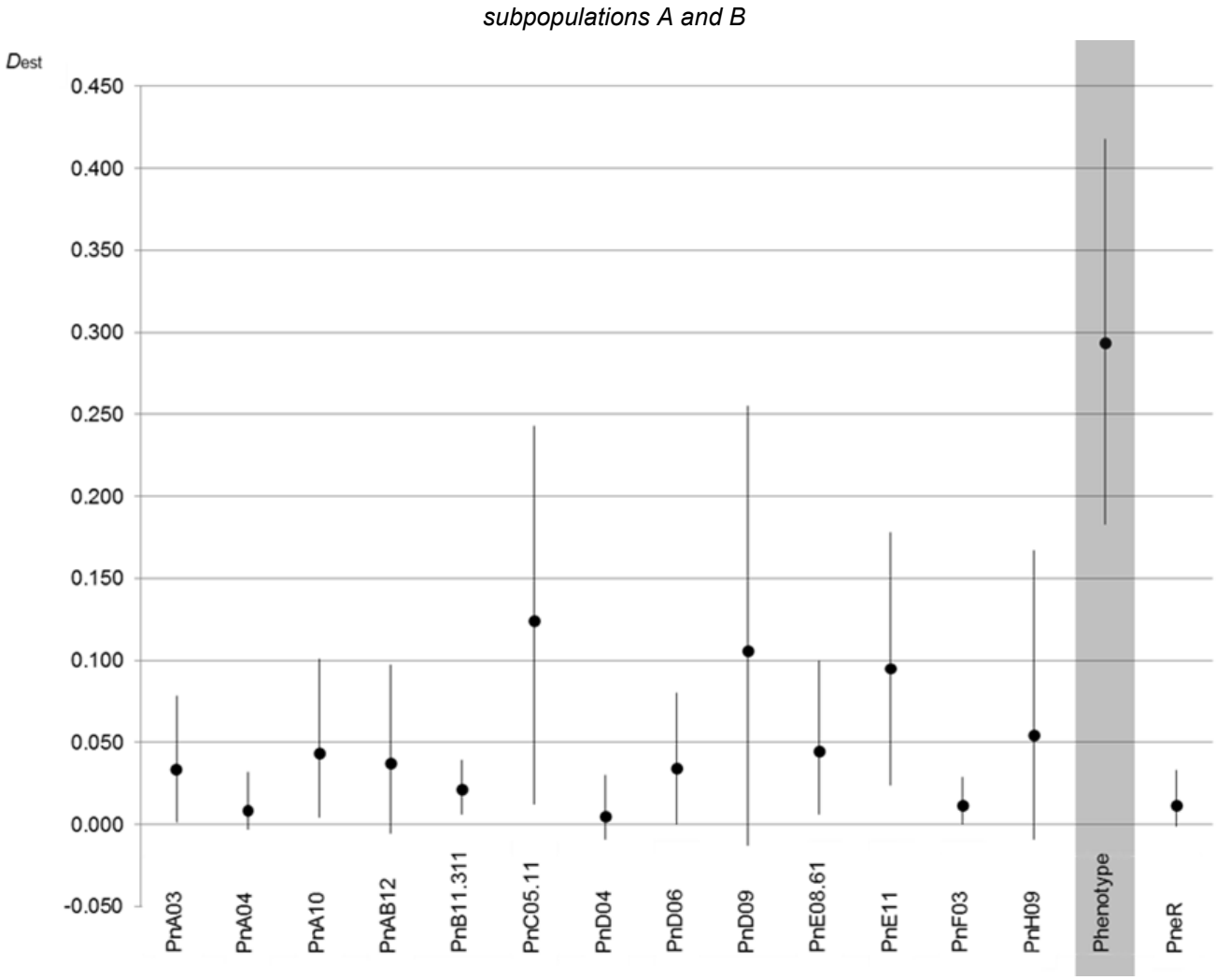


Figure 6a Genetic differentiation estimates (D_{est} (Jost 2008)) between different sets of subpopulations for microsatellite markers, the resistance trait and the PneR-gene. Lines represent the 95% confidence interval of each estimate. The highest estimate is found for the resistance trait and is highlighted in the figure with grey shading.

for selection where the fdist method indicated evidence of the resistance trait being under selection. For the PneR-gene the posterior probability for a model including selection exceeded 0.99 in the pairwise comparison between subpopulations A and C and the comparison among all three subpopulations. Remarkably, just like the resistance trait, the PneR-gene did not show any evidence for selection for the pairwise comparison between subpopulations A and B. Only for microsatellite locus PnA03 in the pairwise comparison between all subpopulations did the posterior probability for a model including selection exceeding 0.91 which can be interpreted as strong evidence for selection on this locus (Foll 2012). Some microsatellite loci had a posterior probability exceeding 0.76 suggesting substantial evidence for selection. When the inclusion of alpha in the model affected the results for a microsatellite locus, the value of alpha was negative suggesting balancing selection. Also the low F_{ST} value of these microsatellites (Figure 5) suggests balancing selection. For the resistance trait and the PneR-gene, however, the average alphas for the comparisons between sub-

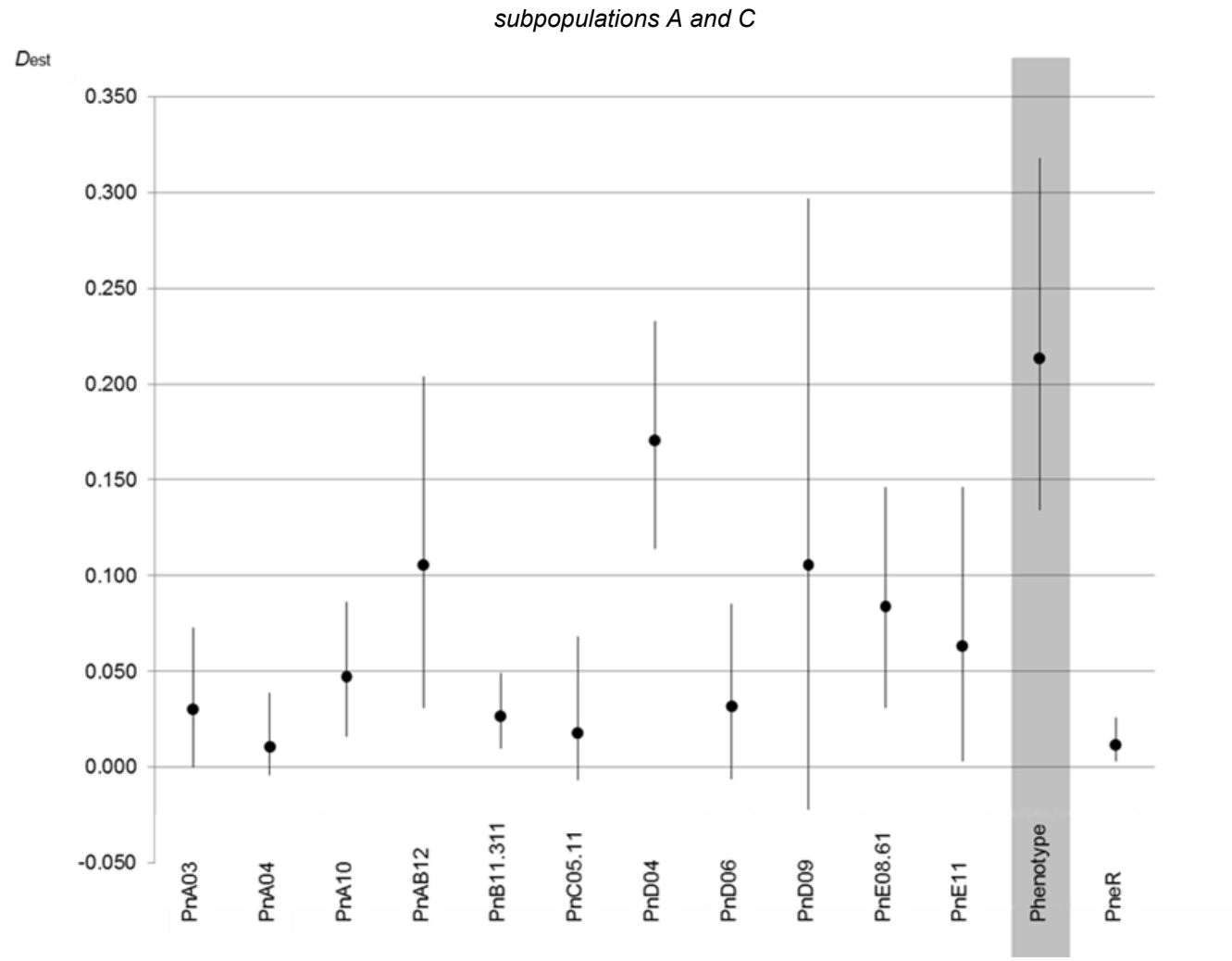


Figure 6b Genetic differentiation estimates (D_{est} (Jost 2008)) between different sets of subpopulations for microsatellite markers, the resistance trait and the PneR-gene. Lines represent the 95% confidence interval of each estimate. The highest estimate is found for the resistance trait and is highlighted in the figure with grey shading.

populations were positive for the comparisons that resulted in a high posterior probability. These positive alphas all indicate that directional selection might be present on the resistance trait and the PneR-gene in these subpopulations.

Though both methods produce somewhat different results, the outcomes are mostly the same. None of the presumed neutral microsatellites came out as a candidate locus for directional selection in any of the analyses. Also, both methods designated the PneR-gene as a potential gene under selection for the comparison between subpopulations B and C, while the gene was found to be neutral for the comparison between subpopulations A and B. The Bayesian method produced a different outcome than the fdist method for the comparison between subpopulations A and C, as the PneR-gene was deemed an outlier in BayeScan but not in LOSITAN. The resistance trait was an outlier locus for either all scenario's or for all comparisons except the one between subpopulations A and B.

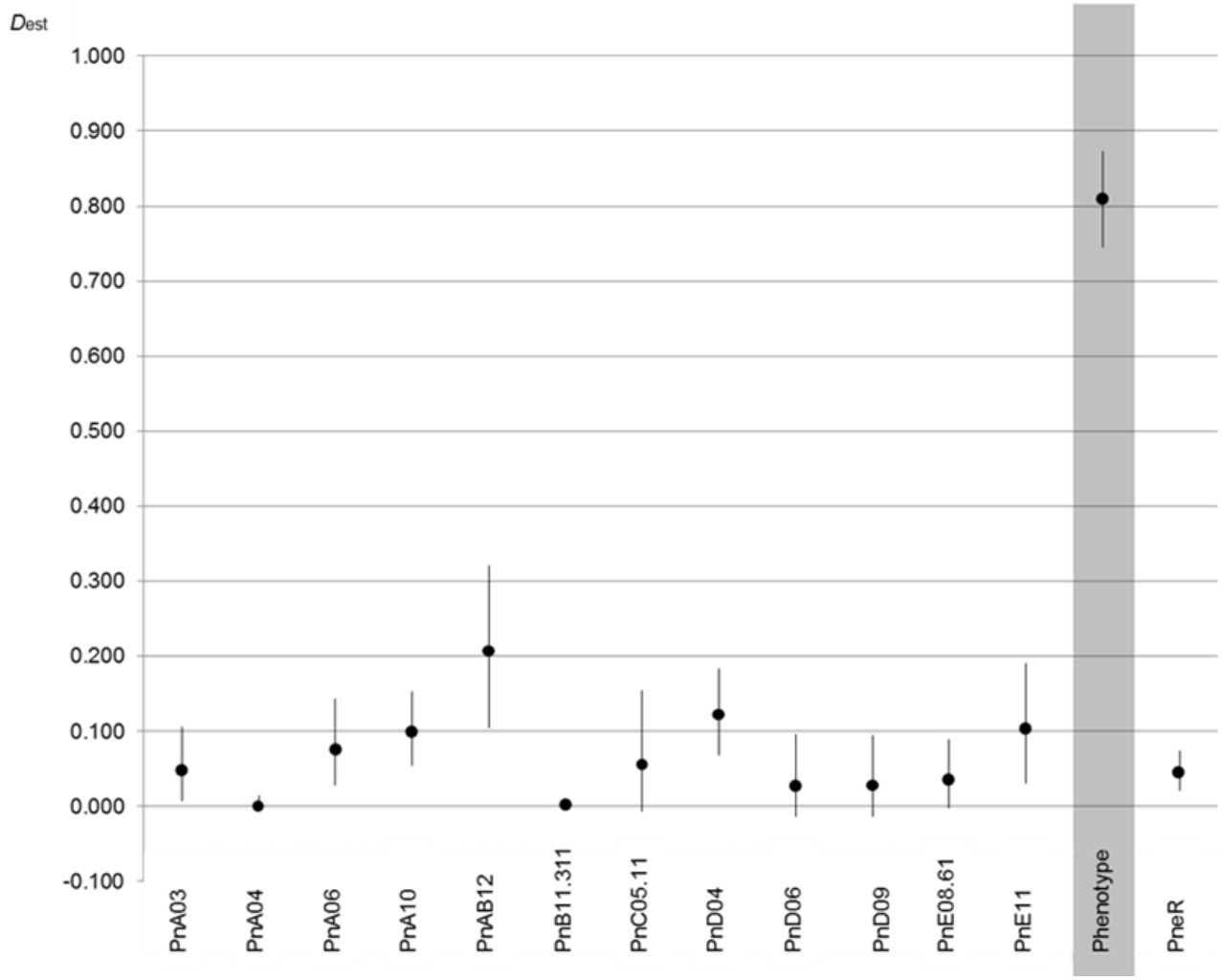


Figure 6c Genetic differentiation estimates (D_{est} (Jost 2008)) between different sets of subpopulations for microsatellite markers, the resistance trait and the PneR-gene. Lines represent the 95% confidence interval of each estimate. The highest estimate is found for the resistance trait and is highlighted in the figure with grey shading.

In the previous chapter (CHAPTER 5) we found a similar outcome for the resistance trait; the trait behaved as an outlier for all subpopulation comparisons when using the fdist method, but not when using the Bayesian method. As in this chapter, no evidence for selection on the resistance trait was found using the Bayesian method for the comparison between subpopulations A and B.

Genetic differentiation between two subpopulations was further estimated by G_{ST} (Nei and Chesser 1983), G'_{ST} (Hedrick 2005) and D_{est} (Jost 2008) for each microsatellite, the resistance trait and the candidate gene, by using SMOGD (Crawford 2010) (see Figure 6 for the estimates of D_{est} for each pairwise subpopulation comparison and supplemental Table S1). As expected the resistance trait showed higher estimates of genetic differentiation than the microsatellites for the comparison between subpopulations B and C and between subpopulation A and C (both comparisons between subpopulations on *B. vulgaris* G-type and *S.*

arvensis). The resistance trait also shows higher estimates than the microsatellites for the comparison between subpopulations A and B (comparisons between subpopulations on *S. arvensis*). For all pairwise subpopulation comparisons the 95% confidence intervals of G_{ST} do not overlap between microsatellites and the resistance trait. The 95% confidence intervals of G'_{ST} and D_{est} do overlap between some microsatellites and the resistance trait, except for the pairwise comparison between subpopulations B and C.

The candidate gene shows no substantially higher estimated parameters than the microsatellite markers for any pairwise comparison between sampled subpopulations (Figure 6 and supplemental Table S1), except for the comparison between locations B and C (host plant *S. arvensis* versus *B. vulgaris* G-type). Though the average value of G_{ST} is here substantially higher than that of the microsatellites (G_{ST} = 0.087 versus G_{ST} = 0 - 0.035), the 95% confidence interval of the PneR-gene overlaps with those of markers PnA10 and PnC05.11. The other two estimated parameters of the candidate gene do not differ from the estimated parameters of the microsatellite markers. The resistance trait shows for all estimators a higher value than the estimator for the PneR-gene in every pairwise comparison between sampled subpopulations. In all cases the 95% confidence interval of the estimated parameter of the resistance trait exceeds the 95% confidence interval of the PneR estimated parameter completely.

Most striking is the difference between estimates of G_{ST}, G'_{ST} and D of the PneR-gene and the resistance trait. Where resistance behaves as an outlier for most estimates, the PneR-gene falls well within the 95% confidence intervals of the microsatellites.

DISCUSSION

In chapter 3 we found three sequences coding for a β-glucosidase enzyme in *Phyllotreta nemorum*. One of these was only present in flea beetle lines which are resistant against the defense of *B. vulgaris* G-type and not in a susceptible flea beetle line. This prompted us to investigate whether this sequence and its homologue were responsible for the difference in the flea beetle's resistance phenotype, knowing that resistance is a Mendelian trait in the populations found near Kværkeby. The β-gluC sequence, which had so far only been found in resistant beetle lines, would correspond to the dominant resistant allele (R-allele) and the similar β-gluB sequence to the susceptible allele (r-allele). Based on the idea that both sequences would code each for an allele of the same gene, we hypothesized that the resistance phenotype of sampled flea beetles would correlate with the possession of the β-gluB or β-gluC sequence. When we directly compared the resistance phenotype of sampled beetles with the presence of β-gluB and β-gluC sequences, the beetle's resistance did not seem to depend on possessing either β-gluB or β-gluC. In fact, nearly all beetles contained the β-gluC variant and the few that did not were either resistant or susceptible (Figure 3). The discrepancy between this finding and earlier results may be because the susceptible line that was used to clone β-glucosidase sequences consisted of beetles originating from the same location in Denmark. When setting up this susceptible line by crossing between

sampled beetles that were not resistant, the β-gluC sequence can have been lost along the process if not already not present in the beetles used to set up the susceptible line.

In addition to the direct comparison between β-glucosidase sequences and resistance phenotype, we tried to detect evidence for selection on both β-glucosidase sequences and resistance phenotype using a molecular population genomics approach. Selection on a locus has as effect that the heterozygosity of that locus will diminish, leading to a deviation of Hardy-Weinberg expectations (HWE). If both β-glucosidase sequences code for alleles of the same PneR-gene, then allele frequencies of this candidate gene are expected to deviate from HWE when they are involved in the flea beetle's ability to use *B. vulgaris* G-type as a host plant and selection pressure acts upon the resistance trait. In our sampled subpopulations, on patches of *B. vulgaris* G-type and *S. arvensis* both the β-gluB and β-gluC sequences were in frequencies according to Hardy-Weinberg expectations. This means that if β-gluB and β-gluC are alleles of a bi-allelic gene, this gene does not experience the selection pressure expected on the gene responsible for resistance. The outlier approach, however, suggests that the candidate PneR-gene is subject to selection in some scenarios albeit not as strong as the resistance trait. Only for the comparison between subpopulations B and C did the PneR-gene emerge as a candidate for being under selection when the fdist method was applied, but not when using the Bayesian method. Vice versa, the PneR-gene emerged as a candidate for being under selection for subpopulations A and C and the comparison among all subpopulations when using the Bayesian method, but not when using the fdist method. Because for none of the subpopulation comparisons both methods gave a clear indication for the PneR-gen to be a candidate for selection, while the resistance trait was a strong candidate for selection, the PneR-gene seems not directly responsible for the difference between susceptible and resistant beetles. On the other hand, the PneR-gene is an outlier compared to the microsatellite loci in some comparisons. Hence, we cannot determine whether β-gluB and β-gluC are under selection or not. It could very well be that selection on these sequences is simply not strong enough to be picked up by the outlier approach. We can, however, rule out that β-gluB and β-gluC are alleles of a bi-allelic gene which is directly responsible for resistance, because the resistance phenotype of flea beetles does not correspond to the presence of β-gluB or β-gluC DNA sequences in the genome.

Because mutation rates, mode of mutation and level of heterozygosity are likely to differ between microsatellites, the resistance trait and the PneR-gene one has to be very careful when trying to directly compare different genetic differentiation measures among these loci. Estimators G'_{ST} and D are, for example, very sensitive to differences in mutation rate and are therefore advised not to be used when different types of loci are used to compare genetic differentiation measures (Leng and Zhang 2011; Whitlock 2011). The assumed underlying mutation model of loci can also influence the value of D_{est} (Leng and Zhang 2011). In our outlier approach we therefore used the infinite allele model (IAM) and the stepwise mutation model (SMM) when we estimated the genetic differentiation between subpopulations. We found that our results in the outlier analyses were the same under IAM and SMM, making it less likely that assuming a wrong mutation model has influenced our results. G_{ST} is more robust under different underlying mutation models, but has another disadvantage. The esti-

mator of G_{ST} is greatly influenced by H_s when highly polymorphic loci are used. Because of the way G_{ST} is estimated, high levels of within population heterozygosity (H_S) always lead to low G_{ST} estimates, whether there is high differentiation between subpopulations or not. A direct comparison of G_{ST} estimates between loci with different and high heterozygosity levels is therefore not recommended. Because G'_{ST} is less influenced by the heterozygosity of a locus and Jost's D is not calculated using heterozygosity, G_{ST} can become much smaller than G'_{ST} and D when H_S is high. Each estimator has different drawbacks and therefore comparisons among estimators may shed light on what causes differences within a distribution of an estimator for different loci. In our data H_S for the microsatellites varied between 0.21 and 0.86, and most microsatellites showed a within-population heterozygosity greater than 0.7. Also, estimates of G_{ST} were consistently smaller than estimates of G'_{ST} and D, except for microsatellite PnB11.311 which also showed the smallest values of H_S (0.21 - 0.32). In that sense, we believe that the estimator of G_{ST} for the microsatellites underestimates genetic differentiation compared to loci with lower H_S like the resistance trait and the PneR-gene (H_S varying between 0.38 and 0.14, and between 0.27 and 0.13 respectively). G'_{ST} approaches D when diversity is high (Jost 2008), and indeed each estimator of G'_{ST} is close to D_{est} (Figure 6 would look more or less the same for G'_{ST}, except for values found for the resistance trait and PneR-gene). Despite the unsuitability of the genetic differentiation estimators to compare their values among different markers, the genetic differentiation estimators can tell us about the processes underlying the subpopulation differentiation when being compared amongst themselves. According to a microsatellite simulation study by Leng and Zhang (2011), very small G_{ST} values and D values larger than G_{ST} - combined with very high H_S values - may indicate a large effective population size and G_{ST} likely strongly underestimates population differentiation. For most microsatellite markers, H_S was estimated to be between 0.7 and 0.9, which can be considered high. G_{ST} values were always smaller than 0.05 (considered very small by Leng and Zhang 2011), and D values were often approximately ten times greater than G_{ST} values. This suggests that the sampled subpopulations of flea beetles near Kvaerkeby have likely a large effective population size. Additionally, an indication of the mutation rate of a locus can be given by looking at the level of within-population heterozygosity (H_S) of a locus. A high H_S usually indicates a high mutation rate. The levels of heterozygosity of the microsatellites used in the outlier approach were not higher than the levels of H_S calculated for the resistance trait assuming HWE (Table S1). The lowest levels of H_S in each subpopulation were found for the candidate gene, indicating that higher estimated values of genetic differentiation for the PneR-gene than for the other loci may be due to a lower mutation rate rather than selection. Nevertheless, no higher values of G_{ST}, G'_{ST}, and D estimators were found for the candidate gene compared to values for the microsatellites. For the resistance trait on the other hand, differentiation estimators were significantly higher than for the candidate gene in all pairwise comparisons. Because we have no direct information about the mutation rates at each locus, we cannot conclude from this that the resistance trait is under selection and the candidate gene is not. But what we can see from the results is that the candidate gene does not show the same amount of genetic differentiation as the resistance trait for any of the estimators and for all pairwise comparisons. Moreover, the genetic differentiation level of the candidate gene falls within the genetic differentiation levels found for the microsatellites which are - apart from mutation rate

- only affected by genome-wide effects, thus not by selection. This also suggests that the candidate gene is not responsible (at least not directly) for the difference in the flea beetle ability to feed on *B. vulgaris* G-type.

Runemark et al. (2010) used a similar approach as we used here to investigate the role of genetic drift and directional selection on the throat colour of Skyros wall lizards, assuming that simple Mendelian inheritance patterns underlie throat colour polymorphism in these lizards. Conform our results for the candidate gene, the assumed allele frequencies underlying the colour morph trait did not deviate from HWE in any of the sampled subpopulations. Another interesting finding in their study is that the F_{ST} value based on colour morph frequencies falls within the distribution of F_{ST} values belonging to each sampled microsatellite. No evidence was found for locus-specific effects on the hypothesized colour morph locus. In contrast, F_{ST} values based on the resistance trait in our study were significantly higher than F_{ST} values found for the microsatellites (chapter 5 and 6). Frequencies of the resistance trait alleles were calculated from the phenotype assuming the alleles to be in HWE in each subpopulation. Even when being assumed in HWE within subpopulations, the differentiation measurement of the resistant trait deviates from neutral expectations when comparing genetic differentiation measurements between subpopulations among loci. For the candidate gene we did not find a deviation from neutral expectations when comparing genetic differentiation measurements among loci.

Based on our results we cannot rule out that these β-glucosidase sequences have something to do with the resistance trait. So far nothing indicates that β-gluB or β-gluC should be considered alleles of the same gene. These sequences may, for example, be homologues located on different parts of the genome. An interesting next step would be to investigate whether β-gluB and β-gluC are located at the same locus by sequencing the part of the genome outside the open reading frame of these sequences. Future research will hopefully reveal if the encoded enzyme of β-gluC also hydrolyses hederagenin-monoglucoside into its aglycone. It will also be interesting to look at levels of upregulation of these β-glucosidase sequences in beetles on *B. vulgaris* G-type and undefended host plant, and look at the differences between upregulation in resistant beetles and susceptible beetles. If these β-glucosidases are involved in the resistance of flea beetles against their host plant defences, our results suggest that they are in any case not directly responsible for the resistance, but rather form part of a pathway of two or more tightly linked genes or are influenced by regulatory changes which cause the difference between resistant and susceptible beetles.

Acknowledgements

We would like to thank Ulrich B. Madsen, Erikholm Skovdistrikt (Kværkeby) for permission to work on the property.

CHAPTER 7

General discussion

Kim M. C. A. Vermeer

The purpose of this chapter is to give an overview over the conclusions drawn in the scientific chapters, discuss the findings in a broader perspective and provide the connection between the chapters. In addition, I will give perspectives for future work. This chapter is organized based on three major questions, linked to the chapters of this thesis: (1) what is the genetic basis of the adaptation studied here, (2) how is this adaptation geographically distributed, and (3) which factors influence this distribution? I will start this chapter with reiterating the incentive for my project, followed by an explanation of how the chapters are logically connected to the three major questions and each other, and finally summarize the findings of each chapter and place them in context and perspective.

Insect adaptation to plant defence

Plants are defended against herbivorous insects in various ways. They produce a broad range of secondary metabolites that may be toxic or deterrent to herbivores or interfere with their development. The herbivorous insects, on the other hand, have ways to circumvent secondary compounds by means of biochemical and behavioural adaptations (Schoonhoven et al. 2005); insects may tolerate, avoid, or detoxify the compounds. In natural populations of the yellow striped flea beetle (*Phyllotreta nemorum* L.) across Denmark, both types of adaptations appear to play a role in the insect's ability to feed on *Barbarea vulgaris* R. Br. G-type. Most individuals of *P. nemorum* cannot develop on *B. vulgaris* G-type because the plant defence levels are high during summer, but for some individuals *B. vulgaris* G-type is suitable throughout the whole year (Nielsen 1997b). The flea beetle's ability to survive on defended *B. vulgaris* G-type depends on dominant resistance genes (R-genes) (Nielsen 1996). Flea beetles with and without resistance genes respond differently to defended *B. vulgaris* G-type in a non-choice bioassay (Nielsen 1996; Nielsen 1997b) (Figure 1). All larvae begin to form a mine, but susceptible larvae stop feeding and die while resistant larvae continue feeding until the third instar and then leave the mine to pupate. Resistant adults feed readily on *B. vulgaris* G-type while mature susceptible beetles take (a)

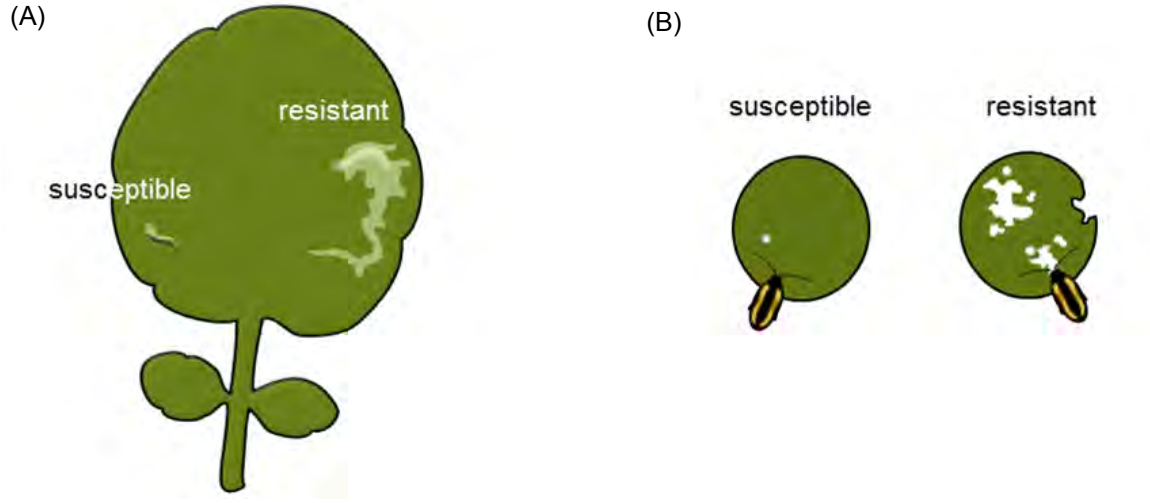


Figure 1 Schematic representation of the response difference of flea beetles in a three-day non-choice bioassay with *Barbarea vulgaris* var. *arcuata* G-type; (A) differences in larval survival of susceptible and resistant larval genotypes after a three-day bioassay with *B. vulgaris* leaves, and (B) differences in the feeding behaviour of susceptible and resistant adult genotypes after a three-day bioassay with leaf discs.

test bite(s) and then stop feeding. This suggests that at least a behavioural component is involved in the response of the flea beetle to different types of host plants. Resistant flea beetles somehow overcome toxic effects of the plant, because they can feed on the plant throughout the year, irrespective of the defence levels that the plant shows to susceptible flea beetles and diamond back moths (DBM). Susceptible flea beetles and DBM do not feed on *B. vulgaris* G-type during summer when saponin levels are high, but accept this plant as a host plant in spring and autumn when saponin levels are low (Agerbirk et al. 2003a; Nielsen 1997b; Shinoda et al. 2002).

The adaptation of insects to a plant defence in a natural environment usually requires changes in both behavioural and metabolic traits (Despres et al. 2007). This is why it is so remarkable that only one dominant R-allele of a resistance gene (R-gene) is enough to convert a susceptible beetle into a resistant one (de Jong et al. 2000; Nielsen 1997a). Despite knowledge of the inheritance patterns of resistance in the flea beetle, the underlying chemical mechanism of flea beetle resistance has, so far, remained unclear. This brings me to the first question I investigated in this thesis:

> 1) Which mechanism underlies the flea beetle adaptation to the defence of *B. vulgaris* G-type?

In CHAPTER 3 I addressed this question by examining the involvement of a possible detoxifying enzyme in *P. nemorum*.

Distribution of insect resistance to a plant defence

Another remarkable aspect of flea beetle resistance to the defence of *B. vulgaris* G-type is that this trait, that intuitively seems beneficial, has not spread across all local flea beetle populations. Moreover, sampling by Nielsen and de Jong (2005) even showed a decline over the years of resistant beetle frequencies on another host plant (*Sinapis arvensis* L.) than *B. vulgaris* G-type. These earlier results suggest that a trade-off is associated with flea beetle resistance, resulting in a lower fitness of resistant flea beetles compared to their non-resistant conspecifics on other host plants than *B. vulgaris* G-type. Indeed, negative effects of the R-allele on the survival of flea beetles have been found in a laboratory set-up (Breuker et al. 2007; de Jong and Nielsen 2000). A disadvantage of studies in the laboratory is that in natural settings in the field selection and fitness are likely to be influenced by factors that are not taken into account in the laboratory (Reznick and Ghalambor 2005). Under natural conditions gene flow between subpopulations can, for example, facilitate the spread of beneficial alleles of modifier genes that counteract negative fitness effects of the R-allele that may be present. The evidence for trade-offs associated with the resistant alleles mentioned above was obtained from flea beetle lines that had been maintained in the laboratory for several generations, and which may have lost such modifiers. Therefore, such laboratory evidence by itself might not provide an explanation for the geographical distribution of resistance genes. The distribution of resistance among the flea beetle populations depends on the structure of the population, the size of subpopulations, the amount

of gene flow among subpopulations and levels of selection within the population. This is why preferably studies of local adaptation should focus on natural conditions in the field (Reznick and Ghalambor 2005). Populations are generally genetically and ecologically structured (Avise 2000; Thompson 2005; Wright 1951). Indeed, in this study system host plants are patchily distributed and flea beetle populations in Denmark show genetic differentiation among each other (de Jong et al. 2001). Thus, the following questions were also investigated during my research:

2) How is the resistance distributed across the flea beetle populations? And

3) Which factors influence this pattern?

These two questions formed the focus of this thesis. More specific, the fundamental question that was investigated is: can the observed geographical distribution of resistant phenotypes of *P. nemorum* to chemically defended *B. vulgaris* in Danish flea beetle populations be explained by factors that are solely associated with effects on the whole genome, such as migration, or do locus-specific factors like selection at the resistance locus explain the geographical distribution of resistant phenotypes? How this could be analysed is described in CHAPTER 2, which reviews and connects the Geographic Mosaic Theory of Coevolution (Thompson 2005) to a recent advance in methods to detect the involvement of selection in the distribution of alleles at presumably adaptive loci, the Population Genomics Approach. Subsequently, CHAPTER 4 presents a field study of the dynamics of resistance of the flea beetles at the phenotypic level. CHAPTER 5 involves a population genomics approach, as laid out in CHAPTER 2, as a means to attribute the observed geographical variation in percentage of resistant and susceptible flea beetle phenotypes to genome-wide versus locus specific effects. In CHAPTER 6, the population genomics approach is applied to the candidate resistance gene found in CHAPTER 3, and the results are compared to those in CHAPTER 5. Below the findings of each chapter separately are summarized.

The potential of a population genomics approach to analyse geographic mosaics of coevolution

Understanding herbivore insect adaptation to its host plant requires the inclusion of spatial aspects that relate to the plant-insect interaction. The geographic mosaic theory of coevolution (GMTC) addresses the spatial aspect of population genetics and evolutionary processes taking place and shaping interspecies interactions (Thompson 2005). The theoretical framework goes beyond local coevolution by regarding evolutionary and ecological processes influencing the evolution of interacting species. Coevolution is not static but a complex process that influences species across highly dynamic landscapes (Thompson 2005). In an interspecies interaction, reciprocal selection only occurs in some localities (coevolutionary hot spots) but these are often connected with localities where the fitness of individuals of a species is not affected by the other interacting species (cold spots). Other processes of coevolution, next to the formation of hot- and cold spots, are the establishment of selection mosaics and trait remixing, according to the GMTC.

In CHAPTER 2 I have shown that processes of the GMTC can be investigated with a population genomics approach:


Detection of variable selection leading to the *identification of evolutionary hot spots and cold spots* that cause spatial variation in traits mediating interactions among species.


Detecting processes that alter the spatial distribution of potentially coevolving alleles and traits (*trait remixing*).

To identify a coevolutionary hot spot, the occurrence of reciprocal selection in that locality should be demonstrated (Gomulkiewicz et al. 2007). Demonstrating reciprocal selection includes identifying traits that mediate an interaction, and inferring reciprocal selection in a statistically rigorous fashion. Population genomics as described by Beaumont and Nichols (1996) distinguishes between genome-wide effects and locus-specific effects that influence the genetic composition of a (sub)population. This enables researchers to disentangle effects by selection from genome-wide effects. By comparing different subpopulations of a species, one can test for the presence of directional selection on a putative adaptive trait.

Trait remixing can be caused by gene flow across coevolutionary hot spots and cold spots, genetic drift within local populations, extinction and recolonization of local populations, and new mutations. Gene flow may cause trait mismatching and local maladaptation by swamping local selection (Slatkin 1985; Thompson et al. 2002). If selection at a certain locus is stronger than the effect of migration, then gene flow at that locus may be weaker than the average gene flow over all loci. Measuring gene flow with neutral loci or by measuring migration is therefore not enough to investigate the effect of gene flow on allele frequencies at a putative adaptive locus. Because population genomics distinguishes genome-wide effects from locus-specific effects at different spatial population levels, a population genomics approach is useful to separate processes of selection and migration/dispersal, which both influence gene flow.

The coevolutionary processes of the GMTC lead to three ecological predictions (Thompson 2005):

1) Populations differ in the traits shaped by an interaction.
2) Traits of interacting species are well matched in only some communities.
3) Few coevolved traits spread across all populations to become fixed traits within the species, because only few coevolved traits are favoured across all populations.

For the interaction between *P. nemorum* and *B. vulgaris*, observations suggest that these predicted patterns of the GMTC are present. Flea beetle resistance to the defence of *B. vulgaris* G-type is only abundant in some flea beetle populations and completely absent in other flea beetle populations (de Jong and Nielsen 1999; Nielsen and de Jong 2005). Populations of *P. nemorum* thus differ in the resistance trait.

The resistance trait of *P. nemorum* is only beneficial in flea beetle populations on *B. vulgaris* G-type (Nielsen 1999). Even so, most *B. vulgaris* G-type patches are not colonized by *P. nemorum* and in some *P. nemorum* populations near non-colonized *B. vulgaris* G-type the resistance trait is absent (de Jong and Nielsen 1999). On the non-defended *B. vulgaris* P-type, susceptible flea beetles thrive as well as resistant flea beetles (Nielsen 1999). Most *B. vulgaris* P-type patches are not colonized by *P. nemorum* and when they are, also resistant flea beetles can be found (J.K. Nielsen, unpubl.). Thus, only in some communities are the resistance trait of *P. nemorum* and the defense trait of *B. vulgaris* well matched.
The resistance trait of the flea beetle is not fixed through all flea beetle populations. Flea beetle resistance to the defence of *B. vulgaris* G-type is only favoured in populations on *B. vulgaris* G-type and not on other host plants. The defence of *B. vulgaris* G-type to *P. nemorum* is only favoured when susceptible flea beetles attempt to colonize a *B. vulgaris* patch, assuming that the flea beetles exert selection pressure on *B. vulgaris* G-type.
To know if these patterns have arisen through what is predicted in the GMTC, one has to investigate the processes behind the patterns (Gomulkiewicz et al. 2007). Indeed, a geographic mosaic of local adaptation and no adaptation is present for the flea beetle and *B. vulgaris* G-type. There is a geographic structure of chemical defense and counter defense in these species across Denmark (Nielsen and de Jong 2005; Thompson 2005). Patchily distributed host plant populations and flea beetle populations can be regarded as coevolutionary hot spots and cold spots, but only if reciprocal selection occurs in at least some of these. So far, we have no information about the issue of whether selection is reciprocal in some subpopulations of the flea beetle and *B. vulgaris* G-type.

Another hypothesis of the GMTC is trait remixing. Processes like gene flow and migration are present in flea beetle populations, which can facilitate trait remixing (De Jong et al. 2009; De Jong et al. 2001). But to know if they lead to trait remixing, one has to know if the force of selection is weaker than drift within patches and how large the selective differences among patches are compared to the force of gene flow (Gomulkiewicz et al. 2007). If we consider the flea beetle – *B. vulgaris* system, the resistance trait of the flea beetle is expected to be under strong directional selection in populations on *B. vulgaris* G-type, while in populations on other host plants the resistance trait is expected not to be under selection or even selected against. The population genomics approach is useful here to find out how selection and other processes influence the resistance trait, not only in flea beetle populations on *B. vulgaris* G-type but also in populations on other host plant patches.

By using a population genomics approach, I have studied the local adaptation of the flea beetle to the defence of *B. vulgaris* G-type in CHAPTERS 5 and 6 by investigating differences in selection pressures on resistance between flea beetle subpopulations. First, I studied the resistance trait by investigating the phenotype of the flea beetles in CHAPTER 5. Then I used a candidate gene for resistance in CHAPTER 6 and investigated if corresponded with the resistant trait and if the candidate gene also experienced locus specific effects. However, before I could apply the candidate gene approach, a candidate gene was needed for flea beetle resistance. In CHAPTER 3 I addressed this.

Identification of saponin-degrading β-glucosidases from the flea beetle *Phyllotreta nemorum*

I have investigated the molecular mechanism that enables the flea beetle to feed on defended *Barbarea vulgaris* in CHAPTER 3. Triterpenoid saponins are responsible for the reduction of the flea beetle food consumption as observed on *B. vulgaris* G-type (Kuzina et al. 2009; Nielsen et al. 2010a; Nielsen et al. 2010b). Hederagenin cellobioside has been shown to deter feeding by susceptible beetles, the mono-glucoside hederagenin also deters feeding by susceptible beetles, but the sapogenin is inactive (Augustin et al. 2012; Nielsen et al. 2010a) (Figure 2). Both hederagenin cellobioside and hederagenin monoglucoside also deter feeding by resistant flea beetles but the effect is significantly smaller than for susceptible beetles (Augustin et al. 2012). The sapogenin also does not deter feeding by resistant beetles. Whether the effect on food consumption is partly due to toxicity of the saponins and/or a deterrent effect of the compound remains unknown.

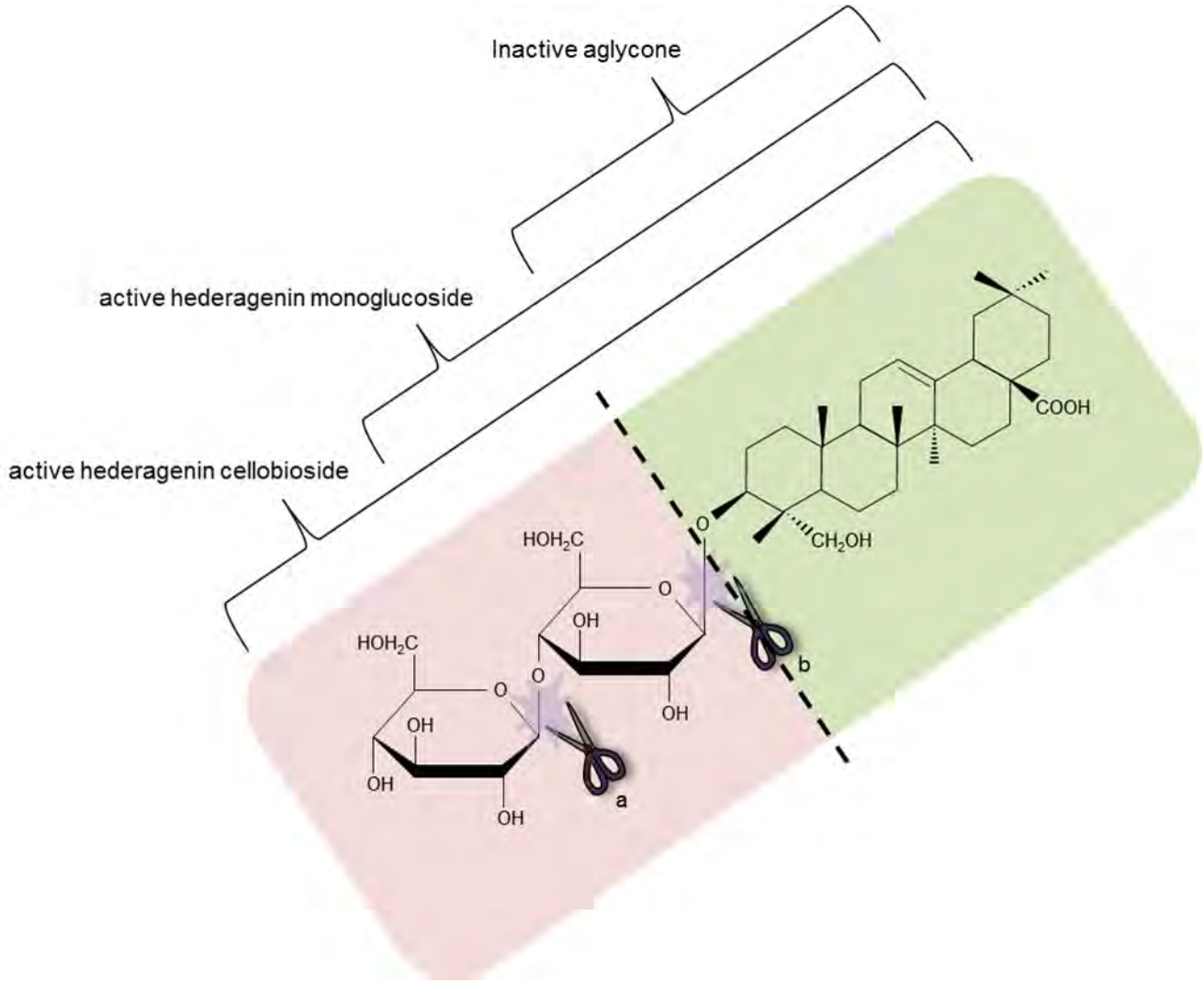


Figure 2 Hypothetical activity of β-glucosidase and the enzyme that is responsible for flea beetle resistance. Hederagenin cellobioside deters feeding. A β-glucosidase is hypothesized to hydrolyse the glucose unit at 'a', producing hederagenin monoglucoside which is still feeding deterrent. To inactivate the saponin, the enzym should hydrolyze the molecule at 'b'.

For the flea beetle to inactivate saponins, we expect β-glucosidases to be involved because this type of enzyme is capable of hydrolyzing glucose units of saponins. Three β-glucosidase cDNA sequences were cloned from resistant flea beetles, named β-glucosidase A, B and C. Recombinant β-glucosidase C protein expressed by an insect cell line hydrolyzed hederagenin cellobioside to glucose and hederagenin monoglucoside. In the light of previous findings from Denmark (Augustin et al. 2012; Nielsen et al. 2010a) where both hederagenin cellobioside and its monoglucoside were proven to be a feeding deterrent for the flea beetle, this enzymatic step does not seem to inactivate the saponin and enable flea beetles to use *B. vulgaris* G-type as a host plant. Moreover, mRNA of all three β-glucosidases was expressed in both susceptible and resistant beetles. This does not rule out that the β-glucosidases we found are involved in the resistance trait of the flea beetle. However, it seems that to find the difference between resistant and susceptible beetles, the ability of resistant beetles to hydrolyze the glycosidic bond between the aglycone and the first linked sugar should be investigated (Augustin et al. 2012). For now, the results in this thesis suggest that expressed β-glucosidase C can hydrolyze the most abundant antifeedant saponin.

Changes in frequencies of flea beetles that are resistant to the defence of *Barbarea vulgaris* vary as much within as between seasons

One of the predictions of the GMTC is that in a plant-insect interaction, a plant defence trait and insect counter defence trait are only well adapted to the interaction in some localities and mismatched in other localities (Thompson 2005). The resistance trait of the flea beetle seems only beneficial in populations on *B. vulgaris* G-type. Earlier research even suggests that a trade-off is associated with flea beetle resistance. In a laboratory set-up, negative effects of the R-allele have been found on the survival of *P. nemorum* (Breuker et al. 2007; de Jong and Nielsen 2000). Furthermore, sampling by Nielsen and de Jong (2005) suggested that the frequency of resistant beetles on *Sinapis arvensis* declines throughout the years. To know if the distribution of the resistance trait can be explained by the hypotheses of the GMTC, such as trait remixing, the processes behind the distribution have to be investigated (Gomulkiewicz et al. 2007). Are genome-wide effects such as gene flow and genetic drift solely responsible for the distribution of R-alleles among populations or also locus-specific effects like differences in selection regimes of the resistance trait? If the resistance trait confers a disadvantage to the beetles when feeding on other host plants than *B. vulgaris* G-type, then we expect selection against resistance in populations on these plants. Depending on how strong gene flow is between populations on *B. vulgaris* G-type and other host plant patches, the strength of selection, population sizes, etc., the frequency of resistant beetles is expected to decline with time on other host plant patches.

An important question therefore is, is there a decline of resistant beetles on other host plants with time? If the frequency of resistance in the flea beetles shows intra-seasonal variation, the results of Nielsen and de Jong (2005) might have been influenced by the timing of sampling, producing the artefactual suggestion that frequencies of resistant flea beetles on other host plants than *B. vulgaris* G-type decreased over the years. To detect whether there is intra-seasonal variation in the frequency of resistant flea beetles, I sampled again populations

from the same study sites and some populations at different times within the season. The frequency of resistant beetles did not decrease over years, but proved to be highly dynamic among years. Even within the season the frequency of resistant beetles varied significantly within a population.

An example of dramatic temporal variation within a year of a phenotypic trait that affects a plant-insect interaction is that in Gols et al. (2007). They showed that under laboratory conditions and controlled temperature conditions, there are dramatic temporal shifts in the level of plant chemicals in *Brassica oleracea* and *Sinapis alba* and that performance of DBM was not strictly correlated with plant defence levels, suggesting that seasonal variation of environmental conditions has consequences for both plant and herbivore. In the field these differences are expected to be even more pronounced. Both the study by Gols et al. (2007) and the present study show that temporal variation of an adaptive trait within a season should not be neglected when comparing changes over years.

The most important conclusion of this chapter is that the frequency of resistance is not systematically decreasing over years on other host plants than *B. vulgaris* G-type, but varying significantly within the season. Nevertheless, frequencies of resistant flea beetles are lower on other host plants than *B. vulgaris* G-type than expected knowing the neutral genetic differentiation between populations. There might still be negative selection on the resistance trait in some populations on other host plants than *B. vulgaris* G-type, for example by outbreeding. Breeding between beetles from *B. vulgaris* G-type that possess both the R-gene and positive modifier-genes that counteract the negative fitness effect of the R-allele, and susceptible beetles on other host plants that do not possess such positive modifier-genes, may disrupt the linkage between R-gene and modifier-genes (de Jong et al. 2001). In the neighbourhood of patches of *B. vulgaris* G-type these co-adapted gene complexes of R-gene and positive modifier-genes may still be present within the flea beetle populations, but further away they might have broken up, only leading to negative selection on resistance in populations further away from *B. vulgaris* G-type. To get more insight in processes leading to trait remixing and the distribution of the resistance trait, a population genomics approach is useful. In CHAPTER 5, I used a population genomics approach to distinguish between genome-wide effects and specific effects influencing the distribution of the resistance trait.

Natural selection against resistance of a flea beetle to host plant defences

In CHAPTER 5, a population genomics approach was used to distinguish signs of selection on the resistance trait from signs produced by genome-wide processes, namely gene flow and genetic drift. Two sites from previous work were sampled and one site that had not been sampled previously. All individuals sampled from the *B. vulgaris* G-type patch (site C) were resistant according to conducted bioassays. This is in agreement with previous work where also nearly all individuals sampled on *B. vulgaris* G-type were resistant (Nielsen and de Jong 2005). Of individuals sampled from the *S. arvensis* patch relatively close to *B. vulgaris* G-type (site A) nearly 70% was resistant and from an *S. arvensis* patch further away from *B. vulgaris* G-type (site B) it was 30%. Because site B is situated further away from *B. vulgaris*

G-type than site A, a lower percentage of resistant individuals is expected in the population if the population is structured and only limited gene flow occurs between the subpopulations. To get more insight in the problem whether in addition to neutral processes such as gene flow, also selection against resistance on other host plants than *B. vulgaris* G-type is resulting in the percentages of resistance recorded, microsatellite loci were scored from each individual as well. First, we analyzed the population structure of the sampled flea beetles by means of neutrally behaving polymorphic microsatellite loci. With the program STRUCTURE we inferred from these microsatellites that the genetic structure of the sampled individuals roughly corresponds with the sampled subpopulations but also that there is a lot of admixture in the genetic composition of individuals among subpopulations. This could point at either a recent common ancestral population that divided into these subpopulations and/or lots of genetic exchange between subpopulations. Whatever the cause of this subtle population structure, it does not correspond to the relatively strong population structure derived from the distribution of the resistance trait.

Subsequently, I used a population genomics approach to study whether the resistance trait behaves as an outlier in pairwise comparisons between flea beetle populations on *B. vulgaris* G-type and *S. arvensis*, and more interestingly, in pairwise comparisons between flea beetle populations on *S. arvensis* nearby *B. vulgaris* G-type and further away. Different methods were used for stronger support and different mutation models were used to take into account the possible nature of the used microsatellites. Results were the same for different mutation models when using the frequentist method (Beaumont and Nichols 1996); none of the microsatellite loci was an outlier, while the resistance trait was an outlier in all comparisons. The Bayesian method (Balding 2003; Beaumont and Balding 2004), however, gave a slightly different outcome. Again, no strong evidence was found for locus-specific effects at any microsatellite locus, but strong evidence for selection on the resistance trait was only found for pairwise comparisons between populations on *B. vulgaris* G-type and *S. arvensis* but not for the pairwise comparison between the two populations on *S. arvensis*.

Evaluation of a candidate resistance gene underlying flea beetle resistance to host plant defense

A population genomics approach as applied in CHAPTER 5 can also be used for finding additional proof for involvement of the β-glucosidases found in CHAPTER 3 with resistance. If the gene coding for β-glucosidase is involved in flea beetle resistance, then we expect it to behave as an outlier like the resistance trait. Also, one can test the involvement directly by comparing the resistance phenotype of individuals with the genotype conferred with β-glucosidase primers.

In CHAPTER 3 we found three candidates for genes underlying flea beetle resistance. We developed primers to identify β-glucosidase B and C in flea beetles sampled at the sites of CHAPTER 5 (sites A, B and C), before recombinant β-glucosidase C protein was expressed in an insect cell line to study the ability of the protein to break down the most active saponin. Because DNA sequences of β-glucosidases B and C are highly homologous (94%) and so

far β-glucosidase C was found in resistant beetles only, these two sequences may represent two alleles of the same gene. If this gene is responsible for the difference between resistant and susceptible beetles, then β-glucosidase C would correspond to the R-allele and β-glucosidase B to the r-allele. In CHAPTER 6 we examined if β-glucosidases B and C matched that idea and if the hypothetical gene was an outlier in the population genomics approach for the same pairwise subpopulation comparisons as in CHAPTER 5. The comparison between the presence of β-glucosidase B and C in the genome of an individual and their resistance phenotype showed no correlation. The hypothetical β-glucosidase gene also did not behave the same as the resistance phenotype in the population genomics approach. Interestingly, the hypothetical β-glucosidase gene did also not completely behave as a neutral part of the genome (microsatellites). When the Baysian method was used, the hypothetical β-glucosidase gene was an outlier for both pairwise comparisons between populations on *B. vulgaris* G-type and *S. arvensis*, as was found for the resistance trait. When the frequentist method was used, however, the hypothetical β-glucosidase gene was only an outlier for one pairwise comparison between populations on *B. vulgaris* G-type and *S. arvensis*, while the resistance trait was an outlier in all pairwise comparisons.

CONCLUSION

The mechanism behind the ability of the flea beetle to use defended *Barbarea vulgaris* G-type still remains unclear (CHAPTER 3). Saponins seem to deter feeding of most flea beetles, but how saponins deter feeding by flea beetles and the location in the body of the beetle where it is effective are unknown. Flea beetles produce enzymes that can hydrolyze one sugar from the saccharide chain of the saponin, but the product of this hydrolysis also has a deterrent effect on flea beetle feeding. These enzymes are also transcribed in both resistant and susceptible beetles, while we expect a difference in transcription between resistant and susceptible flea beetles if the enzyme underlies flea beetle resistance. The gene coding for β-glucosidase C is therefore not the resistance gene that distinguishes resistant flea beetles from susceptible flea beetles.

The genomic make-up of a species and a population is influenced by forces such as selection, genetic drift and mutation rate. These forces also vary spatially, through time and per part of the genome. For example, CHAPTER 4 shows that frequencies of flea beetles that are genetically adapted to the defense of *B. vulgaris* G-type vary highly on other host plants between and within years. Results from CHAPTERS 5 and 6 suggest that this flea beetle adaptation is not only under positive selection in populations on *B. vulgaris* G-type but also under directional selection in populations on other host plants.

Results from previous studies nearby Kværkeby have suggested that the percentage resistant flea beetles is close to 100% in populations on *B. vulgaris* G-type, but much lower (± 50%) in populations on other host plants in geographically distinct patches (de Jong and Nielsen 1999; Nielsen and de Jong 2005). Results from CHAPTERS 4 and 5 add to a better understanding of the distribution of the resistance trait. Not only in *P. nemorum* populations

nearby Kværkeby but also in other populations in Denmark the ratio resistant/susceptible flea beetles is variable, even significantly varying within the reproductive season of the flea beetle (CHAPTER 4). In both CHAPTERS 4 and 5 the percentage resistant flea beetles in populations on *B. vulgaris* G-type was close to 100% as has been found before (de Jong and Nielsen 1999; Nielsen and de Jong 2005). The percentage resistant beetles found on *S. arvensis* nearby Kværkeby differed, however, from previous findings. Up to 80% of the sampled flea beetles from a population on *S. arvensis* was resistant to the defense of *B. vulgaris* G-type (CHAPTER 4) where 71% was the maximum of resistant beetles found on *S. arvensis* in an earlier study (Nielsen and de Jong 2005). The high variability in ratio of resistant/ susceptible beetles in flea beetle populations within a year and among geographically distinct host plant patches may be fuel to a geographic mosaic of coevolution between the flea beetle and its host plants. Vice versa, a geographic mosaic of coevolution may fuel the high variability in ratio of resistant/susceptible beetles in flea beetle populations within a year and among geographically distinct host plant patches. Geographically variable selection on the resistance trait of the flea beetle across different host plant patches may lead to geographic differences in the occurrence and frequency of resistant beetles.

Taking together all evidence from previous studies and this thesis, the following scenario may explain the distribution of the resistance trait in the flea beetle *P. nemorum*. Resistance is highly favourable in *P. nemorum* populations on *B. vulgaris* G-type because the flea beetle cannot complete development on this host plant without resistance to the defense of *B. vulgaris* G-type. Most patches of *B. vulgaris* are not colonized by the flea beetle, despite the occurrence of flea beetle populations nearby with R-alleles (J.K. Nielsen, unpubl.). On the other hand, flea beetle populations on other host plant patches with no *B. vulgaris* G-type in close proximity are carrying R-alleles. Together with the high variation in frequency of resistant beetles in populations on other host plants than *B. vulgaris* G-type and little genetic differentiation between geographically distinct subpopulations at neutral markers, it suggests that at the neutral level there is moderate gene flow between subpopulations. *Barbarea vulgaris* is available to the beetles early in the season and both plant types are then still susceptible to all flea beetles. This may have added to the colonization of this plant by *P. nemorum,* because flea beetles did not have to evolve resistance before starting to use *B. vulgaris* as a host plant. The early availability of *B. vulgaris* may also have negative consequences for the flea beetles. *Barbarea vulgaris* starts wilting relatively early compared to other host plants of the flea beetle. Wilting affects both susceptible and resistant beetles as fewer resources become available during the season. Susceptible flea beetle larvae cannot complete all larval stages in the *B. vulgaris* G-type leaves as the plant's defense level rises through the season (hence the strong selection in favour of the resistance trait in beetles on *B. vulgaris* G-type) and resistant beetles do not have many leaves available, especially later in the season when wilting takes place. This might be the reason why both *B. vulgaris* P-type and G-type are not often used as a host plant in Denmark. As *B. vulgaris* becomes less available through the flea beetle season, flea beetles are expected to migrate to other host plant patches. This migration results in a higher frequency of resistant beetles in populations on other host plants. Nevertheless, the frequency of resistant beetles does not seem to increase on other host plant patches. We now have evidence suggesting that in the field

the resistance trait is also under selection on other host plants than *B. vulgaris* G-type. In CHAPTER 5 more genetic differentiation was found for the resistance trait in the pairwise comparison between a flea beetle population nearby *B. vulgaris* G-type and more distant from *B. vulgaris* G-type compared to genetic differentiation at the neutral level. This indicates that locus-specific effects also influence the resistance trait in populations on other host plants than *B. vulgaris* G-type. This could be because of the breaking up of coadapted gene complexes consisting of the R-gene and positive modifier genes (see also de Jong and Nielsen 2002). Direct negative pleiotropic effects of R-allele possession on other host plants than *B. vulgaris* G-type result then in negative selection. Different selection mosaics are thus believed to be part of the interaction between *P. nemorum* and its host plants. Also *B. vulgaris* P-type and G-type are influenced by the presence of *P. nemorum*. Both resistant and susceptible flea beetles are capable of colonizing *B. vulgaris* P-type, but *B. vulgaris* G-type is mostly colonized by resistant flea beetles. *Barbarea vulgaris* G-type seems to have an advantage over P-type in that respect, because the G-type is only attacked by resistant flea beetles.

The interaction between *P. nemorum* and *B. vulgaris* is more complicated though than what solely a pairwise interaction between the two species would yield, because there are more players in the field. *Phyllotreta nemorum* populations sometimes suffer from heavy parasitization (P. W. de Jong and J. K. Nielsen, unpublished data). Research has shown that flea beetle populations on *B. vulgaris* G-type are season-wide less heavily parasitized than populations on other host plants (Kerstes and de Jong 2011). Another study has also focused on the interaction between *B. vulgaris* and the oomycete *Albugo candida* (Pers. ex Fr.) O. Kuntze (Toneatto 2009). In contrast to susceptible flea beetles, this fungus performs best on *B. vulgaris* G-type and is not capable of infecting *B. vulgaris* P-type (Toneatto 2009). If there is a reciprocal interaction between flea beetle and *B. vulgaris*, then it is not a tight pairwise interaction, but one embedded in a geographical mosaic of coevolution with reciprocal interaction only in hot spots. In this geographical mosaic of coevolution we also find possible trait mismatching where flea beetles carry R-genes on other host plants than *B. vulgaris* G-type and other types of selection pressures on both flea beetle and *B. vulgaris* where other interactors such as the oomycete *A. candida,* also exert selection pressure on *B. vulgaris* and *P. nemorum*.

FUTURE PERSPECTIVES

In this thesis I have mostly focussed on *P. nemorum* populations near Kværkeby where the mode of inheritance of resistance appears to be straightforward. In these populations only one autosomal gene seems responsible for resistance in the beetles (de Jong et al. 2000). In other parts in Denmark, however, inheritance of the ability to develop on *B. vulgaris* G-type is less straightforward; crosses indicate that also sex-linkage and minor genes are involved in resistance (de Jong and Nielsen 1999; Nielsen 1997a; Nielsen 1999). This leads to the question: is the R-gene of Kværkeby the same gene that is responsible for resistance of flea beetles at other sites? Crosses among beetles from different sample sites can give an

answer to that. By means of crosses it may be possible to discriminate between R-genes of different populations, such as the populations of Ejby and Kværkeby located on *B. vulgaris* G-type (Figure 3). If a heterozygous individual from Kværkeby will be crossed with a heterozygous individual from Ejby, we would expect 75% of the offspring (F1) to be resistant and 25% to be susceptible. If it concerns one resistance gene, 1/3 of the resistant offspring would be homozyous for the resistance trait. This means that if the F1 is crossed with susceptible beetles (beetles that possess no R-alleles) the next generation (F2) of 1/3 of these

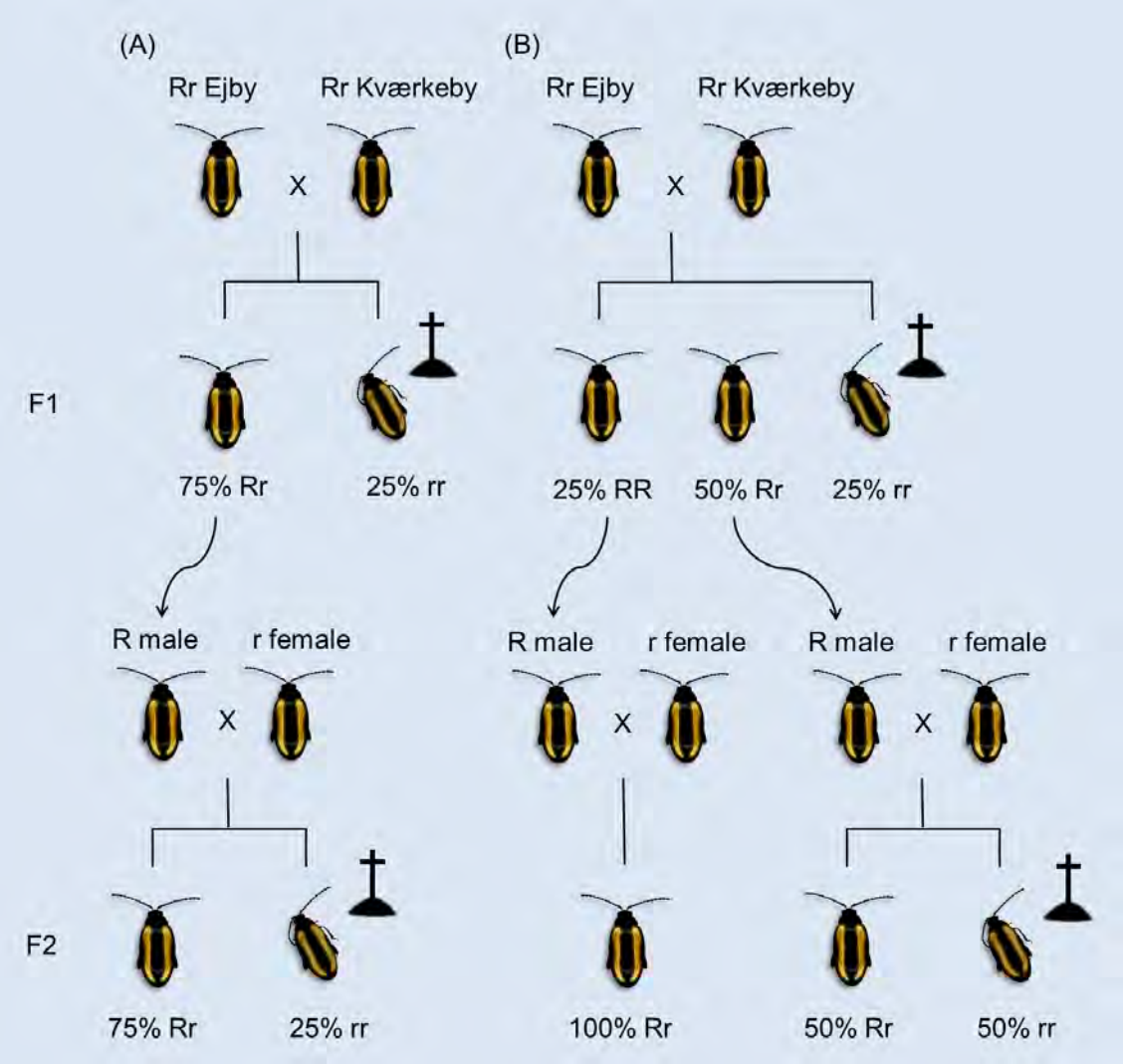


Figure 3 Schematic representation of crosses between flea beetles of different populations to investigate if resistance is inherited by the same R-gene. (A) outcome of the cross when resistance involves different genes, (B) outcome of the cross when resistance in both parents involves the same gene; only in this cross the offspring would also include homozygous resistant individuals. Because both homozygous and heterozygous resistant beetles show the same phenotype, a distinction between homozygous and heterozygous resistant individuals can be made by crossing the resistant beetles of the F1 with a susceptible beetle. Percentages of the F2 are when assumed that the R-genes are not linked. If they are linked then the percentages of the F2 of cross A may differ according to the degree of linkage.

crosses is expected to produce only resistant beetles. If two different R-genes are responsible for resistance in the initial cross, the resistant offspring is expected to consist solely of heterozygous beetles. When the F1 is then crossed with susceptible beetles, the F2 will always produce susceptible and resistant beetles (one would expect a ratio of 1:1 when there is no linkage between the R-genes). Similar crossings can be set-up to investigate different types of inheritance among and within populations.

So far, bioassays and crossings have been used to detect the R-gene and discriminate between resistant and susceptible beetles. Both bioassays and crossings are laborious and therefore the development of primers to detect the R-gene in flea beetles would be a step forward in the study of this system. To be able to develop primers for the R-gene, the gene responsible for flea beetle resistance first needs to be identified. In this thesis we found a candidate for the R-gene, namely a gene coding for the enzyme β-glucosidase C. Even though this enzyme is capable of hydrolyzing one of the saponins responsible for the defense of *B. vulgaris* G-type, the gene coding for β-glucosidase C was not exclusively present in resistant beetles and the transcribed enzyme turned out not to differ consistently between resistant and susceptible flea beetles either. To find the gene responsible for the difference between resistant and susceptible flea beetles, it would be interesting to screen for more genes involved in resistance, including other genes coding for β-glucosidase. The results of CHAPTER 6 make it very unlikely that the DNA sequences β-glucosidase B and C correspond to two alleles of a bi-allelic resistance gene. This does not rule out that any of these DNA sequences are involved in flea beetle resistance. To know more about the sequences of β-glucosidase B and β-glucosidase C, it would be interesting to sequence outside of their reading frames to see whether they are alleles of the same locus or located on different part of the genome.

The activity of the recombinant β-glucosidase C protein was only tested with hederagenin cellobioside as a substrate. *Barbarea vulgaris* G-type produces more saponins that influence the flea beetle feeding rate. Also oleanolic acid cellobioside and its monoglucoside deter feeding by flea beetles while the aglycone of oleanolic acid cellobioside does not (Augustin et al. 2012; Nielsen et al. 2010a). The effect of oleanolic acid cellobioside and its monoglucoside was significantly weaker than that of hederagenin cellobioside and hederagenin monoglucoside. Furthermore oleanolic acid cellobioside is not present in *Barbarea vulgaris* G-type in such high doses as hederagenin cellobioside (Kuzina et al. 2009). Because we don't know whether several saponins with minor influence have an additional effects on flea beetle feeding rates, one could gain more insight in flea beetle resistance by taking other saponins into account as well when searching for the resistance gene.

Research on the interaction of *P. nemorum* and *B. vulgaris* G-type may also benefit from new molecular tools such as quantitative PCR (qPCR). qPCR enables scientists to compare quantitative differences in the transcript levels of genes of interest. In the future this can, for example, be used to investigate the difference in transcript level of one or more β-glucosidase sequences between resistant and susceptible flea beetles. Both susceptible and resistant beetles may possess the gene coding for an enzyme that inactivates the de-

fence compounds produced by *B. vulgaris* G-type, but the transcription level of this gene may be high in resistant beetles compared to the level in susceptible beetles. If resistance of the flea beetle is induced by feeding on *B. vulgaris* G-type, we would expect resistant beetles to show a high transcription level of the gene coding for an enzyme that enables the flea beetle to feed on *B. vulgaris* G-type, when presenting *B. vulgaris* G-type leaves. Susceptible beetles are then expected not to show an increase in transcription level when presented *B. vulgaris* G-type leaves. Combining bioassays with qPCR may facilitate finding the resistance gene that enables flea beetles to completely develop on *B. vulgaris* G-type. This type of approach may also prove helpful when studying candidate genes for traits of interest in other plant-insect interactions.

Previously, genetic differentiation between flea beetle subpopulations in Denmark was measured using allozymes (de Jong et al. 2001). Because allozymes are not as variable as other markers, they are less powerful in detecting genetic differentiation (Hoy 1994). Microsatellites are a more powerful type of marker and genetic differentiation measured with microsatellites is expected to be more pronounced than when using allozymes. Nonetheless, the genetic differentiation found with microsatellites in CHAPTER 5 seems very low and a lot of the genetic composition of individuals among subpopulations overlap. A study of selection on the resistance trait in beetles on other host plants than *B. vulgaris* G-type done at a larger scale may therefore yield clearer results. Field experiments, or experimental evolution designs, may be another way of studying selection on the resistance trait in beetles on other host plants than *B. vulgaris* G-type. Measuring the fitness of resistant beetles and susceptible beetles on other host plants will elucidate if there is negative selection on the resistance trait on other host plants.

Another step forward in research on the interaction of *P. nemorum* and *B. vulgaris* can be made in monitoring frequencies of resistant beetles on other host plants. We found significant variation within the season in the frequency of resistant beetles on other host plants, though only for two sampled populations. In both populations the frequency increased significantly. It would be interesting to study if such an increase in resistance in the frequency of resistant beetles on host plants near colonized *B. vulgaris* G-type patches over the season is more common and if this is caused by migration of beetles from *B. vulgaris* G-type patches.

The interaction between *P. nemorum* and *B. vulgaris* is considered to be a good example of a geographical mosaic of coevolution (Thompson 2005; Thompson 2013). To conclude that coevolution takes place, however, we have to show that there is reciprocal selection involving the plant- and the insect traits. The flea beetle resistance seems to be under selection on *B. vulgaris* G-type, but is *B. vulgaris* experiencing selection pressure by *P. nemorum*? To conclude that coevolution takes place we should first investigate if the flea beetle influences the fitness of *B. vulgaris*. Because other species are also likely to influence the fitness of *B. vulgaris* and *P. nemorum*, studies in the field should also include those other species, such as the oomycete *A. candida* and parasites of *P. nemorum*. Studies about evolution are shifting more towards the evolution of interactions rather than evolution in a single species (Thompson 2013) and the system of the flea beetle and its host plants, natural enemies and

competitors is a promising study system to study evolution through the interactions that it encompasses.

In conclusion, some of the topics that need further investigation in the flea beetle - *B. vulgaris* model system are: the discrimination of resistance gene(s) in different populations, the quantitative expression of candidate genes, the effect of different saponins on flea beetle resistance and reciprocal selection in the flea beetle - *B. vulgaris* interaction. Further research on these topics will hopefully lead to a deeper understanding of the evolution of insect resistance to plant defences in general. One of the biggest challenges in evolutionary ecology is to gain more knowledge of the genetic basis of natural adaptations. Understanding the genetic basis of adaptation is crucial to explain variation in adaptive traits at the level of populations, species or communities (Schluter et al. 2010). The focus in further research on the flea beetle - *B. vulgaris* interaction should therefore be on the genetics of this interaction. This study (CHAPTERS 5 and 6) and others (e.g. Nosil 2009; Rogers and Bernatchez 2005; Runemark et al. 2010) show the potential of using a population genomics approach to study processes behind (co)evolution such as gene flow, genetic drift and selection. These studies illustrate that scientists should not be afraid to implement and combine several techniques and/or theories as was done in this thesis with using a population genomics approach to study geographic mosaics of local adaptation between plant and insect. Building on the knowledge gained by multidisciplinary approaches will hopefully lead to a better understanding of defences and counter-defences in plant-insect interactions. And this will, in turn, benefit applied research, such as studying the potential of *B. vulgaris* as a dead-end trap for pest insects and preventing/diminishing the development of resistance to crop defences by pest insects.

REFERENCES

Adel MM, Sehnal F, Jurzysta M (2000) Effects of alfalfa saponins on the moth *Spodoptera littoralis*. Journal of Chemical Ecology 26:1065-1078

Agerbirk N, Olsen CE, Nielsen JK (2001) Seasonal variation in leaf glucosinolates and insect resistance in two types of *Barbarea vulgaris* ssp. *arcuata*. Phytochemistry 58:91-100

Agerbirk N, Olsen CE, Bibby BM, Frandsen HO, Brown LD, Nielsen JK, Renwick JA (2003a) A saponin correlated with variable resistance of *Barbarea vulgaris* to the diamondback moth *Plutella xylostella*. Journal of Chemical Ecology 29:1417-1433

Agerbirk N, Ørgaard M, Nielsen JK (2003b) Glucosinolates, flea beetle resistance, and leaf pubescence as taxonomic characters in the genus *Barbarea* (Brassicaceae). Phytochemistry 63:69-80

Agrawal AF, Brodie ED, Wade MJ (2001) On indirect genetic effects in structured populations. American Naturalist 158:308-323

Alford DV (1999) A textbook of agricultural entomology. Blackwell Science, Oxford, UK

Althoff DM, Thompson JN (1999) Comparative geographic structures of two parasitoid-host interactions. Evolution 53:818-825

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

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389-3402

Andersen CL, Hazzard R, van Driesche R, Mangan FX (2006) Alternative management tactics for control of *Phyllotreta cruciferae* and *Phyllotreta striolata* (Coleoptera: Chrysomelidae) on *Brassica rapa* in Massachusetts. Journal of Economic Entomology 99:803-810

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. BMC Bioinformatics 9:323

Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, Olsen CE, Halkjær Hansen E, Kuzina V, Ekstrøm CT, Hauser T, Bak S (2012) UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyse sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant Physiology 160:1881-1895

Avise JC (2000) Stability, equilibrium and molecular aspects of conservation in marine species. Hydrobiologia 420:Xi-Xii

Badenes-Perez FR, Reichelt M, Heckel DG (2010) Can sulfur fertilisation improve the effectiveness of trap crops for diamondback moth, *Plutella xylostella* (L.) (Lepidoptera: Plutellidae)? Pest Management Science 66:832-838

Badenes-Perez FR, Reichelt M, Gershenzon J, Heckel DG (2011) Phylloplane location of glucosinolates in *Barbarea* spp. (Brassicaceae) and misleading assessment of host suitability by a specialist herbivore. New Phytologist 189:549-556

Balding DJ (2003) Likelihood-based inference for genetic correlation coefficients. Theoretical Population Biology 63:221-230

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11:155-165

Barron-Yanez M, Villanueva-Verduzco C, Garcia-Mateos R, Colinas-Leon T (2008) Saponin content of sprouts from *Chenopodium nuttalliae* Saff., *Cucurbita pepo* L., *Brassica napus* L. and *Amaranthus leucocarpus* S. Watson syn. *hypochondriacus* L. Planta Medica 74:1174-1174

Barton NH (2000) Genetic hitchhiking. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 355:1553-1562

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B-Biological Sciences 263:1619-1626

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13:969-980

Beebee TJC, Rowe G (2004) Introduction to molecular ecology. Oxford University Press, Oxford

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1998) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II, Montpellier (France)

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B - Methodological 57:289-300

Benkman CW (1999) The selection mosaic and diversifying coevolution between crossbills and lodgepole pine. American Naturalist 153:S75-S91

Benkman CW, Holimon WC, Smith JW (2001) The influence of a competitor on the geographic mosaic of coevolution between crossbills and lodgepole pine. Evolution 55:282-294

Benkman CW, Parchman TL, Favis A, Siepielski AM (2003) Reciprocal selection causes a coevolutionary arms race between crossbills and lodgepole pine. American Naturalist 162:182-194

Berenbaum MR, Zangerl AR (1992) Genetics of physiological and behavioral resistance to host furanocoumarins in the parsnip webworm. Evolution 46:1373-1384

Bermingham E, Avise JC (1986) Molecular zoogeography of fresh-water fishes in the southeastern United-States. Genetics 113:939-965

Bernays E, Graham M (1988) On the evolution of host specificity in phytophagous arthropods. Ecology 69:886-892

Black WC, Baer CF, Antolin MF, DuTeau NM (2001) Population genomics: genome-wide sampling of insect populations. Annual Review of Entomology 46:441-469

Bohonak AJ (1998) Genetic population structure of the fairy shrimp *Branchinecta coloradensis* (Anostraca) in the Rocky Mountains of Colorado. Canadian Journal of Zoology 76:2049-2057

Bouarab K, Melton R, Peart J, Baulcombe D, Osbourn A (2002) A saponin-detoxifying enzyme mediates suppression of plant defences. Nature 418:889-892

Bowyer P, Clarke BR, Lunness P, Daniels MJ, Osbourn AE (1995) Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. Science 267:371-374

Breuker CJ, Victoir K, de Jong PW, van der Meijden E, Brakefield PM, Vrieling K (2005) AFLP markers for the R-gene in the flea beetle, *Phyllotreta nemorum*, conferring resistance to defenses in *Barbarea vulgaris*. Journal of Insect Science 5:38

Breuker CJ, de Jong PW, Victoir K, Vrieling K, Brakefield PM (2007) Pleiotropic effects associated with an allele enabling the flea beetle *Phyllotreta nemorum* to use *Barbarea vulgaris* as a host plant. Evolutionary Ecology 21:13-26

Brito PH, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence -based markers. Genetica 135:439-455

Brodie ED Jr, Ridenhour BJ, Brodie ED III (2002) The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. Evolution 56:2067-2082

Burdon JJ, Thrall PH (1999) Spatial and temporal patterns in coevolving plant and pathogen associations. American Naturalist 153:S15-S33

Butlin RK (2010) Population genomics and speciation. Genetica 138:409-418

Campana MG, Hunt HV, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. Molecular Ecology Resources 11:349-352

Carmody WR (1961) An easily prepared wide range buffer series. Journal of Chemical Education 38:559-560

Charlesworth B, Charlesworth D, Barton NH (2003) The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology Evolution and Systematics 34:99-125

Chew FS (1988) Biological effects of glucosinolates. In: Cutler HC (ed) Biological Active Natural Products: Potential use in Agriculture. American Chemical Society, Washington, DC, pp. 155-181

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry 162:156-159

Cipollini DF, Busch JW, Stowe KA, Simms EL, Bergelson J (2003) Genetic variation and relationships of constitutive and herbivore-induced glucosinolates, trypsin inhibitors, and herbivore resistance in *Brassica rapa*. Journal of Chemical Ecology 29:285-302

Cornell HV, Hawkins BA (2003) Herbivore responses to plant secondary compounds: A test of phytochemical coevolution theory. American Naturalist 161:507-522

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

Courtney S (1988) If it's not coevolution, it must be predation. Ecology 69:910-911

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Molecular Ecology Resources 10:556-557

Crombie WML, Crombie L, Green JB, Lucas JA (1986) Pathogenicity of take-all fungus to oats - Its relationship to the concentration and detoxification of the 4 Avenacins. Phytochemistry 25:2075-2083

Dalby-Brown L, Olsen CE, Nielsen JK, Agerbirk N (2011) Polymorphism for Novel Tetraglycosylated Flavonols in an Eco-model Crucifer, *Barbarea vulgaris*. Journal of Agricultural and Food Chemistry 59:6947-6956

Darwin CR (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London

Darwin CR (1862) On the various contrivances by which British and foreign orchids are fertilised by insects, and on the good effects of intercrossing. John Murray, London

de Geyter E, Lambert E, Geelen D, Smagghe G (2007) Novel advances with plant saponins as natural insecticides to control pest insects. Pest Technology 1:96-105

de Jong PW, Nielsen JK (1999) Polymorphism in a flea beetle for the ability to use an atypical host plant. Proceedings of the Royal Society of London Series B-Biological Sciences 266:103-111

de Jong PW, Nielsen JK (2000) Reduction in fitness of flea beetles which are homozygous for an autosomal gene conferring resistance to defences in *Barbarea vulgaris*. Heredity 84:20-28

de Jong PW, Nielsen JK (2002) Host plant use of Phyllotreta nemorum: do coadapted gene complexes play a role? Entomologia Experimentalis et Applicata 104:207-215

de Jong PW, Frandsen HO, Rasmussen L, Nielsen JK (2000) Genetics of resistance against defences of the host plant Barbarea vulgaris in a Danish flea beetle population. Proceedings of the Royal Society of London Series B-Biological Sciences 267:1663-1670

de Jong PW, de Vos H, Nielsen JK (2001) Demic structure and its relation with the distribution of an adaptive trait in Danish flea beetles. Molecular Ecology 10:1323-1332

de Jong PW, Breuker CJ, de Vos H, Vermeer KMCA, Oku K, Verbaarschot P, Nielsen JK, Brakefield PM (2009) Genetic differentiation between resistance phenotypes in the phytophagous flea beetle, *Phyllotreta nemorum*. Journal of Insect Science 9:69

Degen T, Dillmann C, Marion-Poll F, Turlings TC (2004) High genetic variability of herbivore-induced volatile emission within a broad range of maize inbred lines. Plant Physiol 135:1928-1938

Despres L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. Trends in Ecology & Evolution 22:298-307

Dicke M, van Loon JJA, de Jong PW (2004) Ecogenomics benefits community ecology. Science 305:618-619

Dobzhansky T, Pavlovsky O (1958) Interracial hybridization and breakdown of coadapted gene complexes in *Drosophila paulistorum* and *Drosophila willistoni*. Proceedings of the National Academy of Sciences of the United States of America 44:622-629

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359-361

Edmunds M (1974) Defence in animals. Longman, London

Egan SP, Funk DJ (2006) Individual advantages to ecological specialization: insights on cognitive constraints from three conspecific taxa. Proceedings of the Royal Society of London Series B-Biological Sciences 273:843-848

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution 62:1162-1181

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. Evolution 18:586-608

Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS, Weinstock GM (2007) Creating a honey bee consensus gene set. Genome Biology 8:R13

Endler JA (1977) Geographic variation, speciation and clines. Princeton University press, Princeton, New Jersey

Esen A (1993) Beta-glucosidases - overview. American Chemical Society Symposium Series 533:1-14

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to the human mitochondrial DNA restriction data. Genetics 131:479-491

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

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

Faure D (2002) The family-3 glycoside hydrolases: From housekeeping functions to host-microbe interactions. Applied and Environmental Microbiology 68:1485-1490

Feeny P (1977) Defensive ecology of Cruciferae. Annals of the Missouri Botanical Garden 64:221-234

Fenwick GR, Heaney RK, Mullin WJ (1983) Glucosinolates and their breakdown products in food and food plants. Critical Reviews in Food Science and Nutrition 18:123-201

Feyereisen R (1999) Insect P450 enzymes. Annual Review of Entomology 44:507-533

Feyereisen R (2005) Insect cytochrome P450. In: Gilbert LI, Iatrou K, Gill SS (ed) Comprehensive Molecular Insect Science. Elsevier BV, Amsterdam, pp. 1-77

Foll M (2012) BayeScan v2.1 User Manual, available from http://cmpg.unibe.ch/software/bayescan/files/BayeScan2.1_manual.pdf

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977-993

Forde SE, Thompson JN, Bohannan BJ (2004) Adaptation varies through space and time in a coevolving host-parasitoid interaction. Nature 431:841-844

Foster SP, Harrington R, Dewar AM, Denholm I, Devonshire AL (2002) Temporal and spatial dynamics of insecticide resistance in *Myzus persicae* (Hemiptera: Aphidae). Pest Management Science 58: 895-907

Fox LR (1981) Defense and dynamics in plant-herbivore systems. American Zoologist 21:853-864

Fox LR (1988) Diffuse coevolution within complex communities. Ecology 69:906-907
Fraenkel GS (1953) The nutritional value of green plants for insects. Transactions of the IXth International Congress of Entomology 1951:90-100
Fraenkel GS (1959) Raison d'être of secondary plant substances. Science 129:1466-1470
Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution 52:1744-1759
Funk DJ, Nosil P (2008) Comparative analyses and ecological speciation in herbivorous insects. In: Tilmon KJ (ed) Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects. University of California Press, Berkeley, pp 117-135
Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. Genetica 116:251-267
Futuyma DJ (2009) Evolution. Sinauer Associates Inc., Sunderland
Futuyma DJ, Slatkin M (1983) Coevolution. Sinauer Associates Inc., Sunderland

Gardner SN, Agrawal AA (2002) Induced plant defence and the evolution of counter-defences in herbivores. Evolutionary Ecology Research 4:1131-1151
Gardner SN, Gressel J, Mangel M (1998) A revolving dose strategy to delay the evolution of both quantitative vs major monogene resistances to pesticides and drugs. International Journal of Pest Management 44:161-180
Gillet EM, Gregorius HR (2008) Measuring differentiation among populations at different levels of genetic integration. BMC Genetics 9:60
GLEAN (2006) open source program, available from http://sourceforge.net/projects/gleangene/
Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471
Gols R, Raaijmakers CE, van Dam NM, Dicke M, Bukovinszky T, Harvey JA (2007) Temporal changes affect plant chemistry and tritrophic interactions. Basic and Applied Ecology 8:421-433
Gomez JM, Abdelaziz M, Camacho JPM, Munoz-Pajares AJ, Perfectti F (2009a) Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. Ecology Letters 12:672-682
Gomez JM, Perfectti F, Bosch J, Camacho JPM (2009b) A geographic selection mosaic in a generalized plant-pollinator-herbivore system. Ecological Monographs 79:245-263
Gomulkiewicz R, Drown DM, Dybdahl MF, Godsoe W, Nuismer SL, Pepin KM, Ridenhour BJ, Smith CI, Yoder JB (2007) Dos and don'ts of testing the geographic mosaic theory of coevolution. Heredity 98:249-258
Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate *F*-statistics. Journal of Heredity 86:485-486
Goudet J (2002) FSTAT: a program to estimate and test gene diversities and fixation indices, Lausanne (Switzerland), available from http://www2.unil.ch/popgen/softwares/fstat.htm
Gregorius HR, Degen B, König A (2007) Problems in the analysis of genetic differentiation among populations - a case study in *Quercus robur*. Silvae Genetica 56:190-199

Griffiths AJF, Wessler SR, Carroll SB, Doebley J (2012) Introduction to genetic analysis. W. H. Freeman and Company, New York

Hanski I (1998) Metapopulation dynamics. Nature 396:41-49
Hanski I, Mononen T, Ovaskainen O (2011) Eco-evolutionary metapopulation dynamics and the spatial scale of adaptation. American Naturalist 177: 29-43
Hartl DL, Clark GA (1997) Principles of population genetics. Sinauer Associates, Inc., Sunderland
Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution 53:313-318
Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633-1638
Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Molecular and Cellular Biology 12:2749-2757
Henrissat B (1991) A classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochemical Journal 280:309-316
Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. Molecular Ecology 17:5378-5390
Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010) Phylogeography's past, present and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54:291-301
Hoy MA (1994) Insect molecular genetics; An introduction to principles and applications. Academic Press, Amsterdam
Hu S, Dilcher DL, Jarzen DM, Winship Taylor D (2008) Early steps of angiosperm pollinator coevolution. Proceedings of the National Academy of Sciences of the United States of America 105:240-245
Huang XP, Renwick JAA, Sachdevgupta K (1994) Oviposition stimulants in *Barbarea vulgaris* for *Pieris rapae* and *P. napi-oleracea* - Isolation, identification and differential activity. Journal of Chemical Ecology 20:423-438
Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322-1332
Hughes CR, Queller DC (1993) Detection of highly polymorphic microsatellite loci in a species with little allozyme polymorphism. Molecular Ecology 2:131-137

Idris AB, Grafius E (1996) Effects of wild and cultivated host plants on oviposition, survival, and development of diamondback moth (Lepidoptera: Plutellidae) and its parasitoid *Diadegma insulare* (Hymenoptera: Ichneumonidae). Environmental Entomology 25:825-833
Ikeda H, Setoguchi H (2010) Natural selection on PHYE by latitude in the Japanese archipelago: insight from locus specific phylogeographic structure in *Arcteria nana* (Ericaceae). Molecular Ecology 19:2779-2791
Iwao K, Rausher MD (1997) Evolution of plant resistance to multiple herbivores: quantifying diffuse coevolution. American Naturalist 149:316-335

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806

Jander G, Cui J, Nhan B, Pierce NE, Ausubel FM (2001) The *TASTY* locus on chromosome 1 of Arabidopsis affects feeding of the insect herbivore *Trichoplusia ni*. Plant Physiology 126:890-898

Janzen DH (1980) When is it coevolution? Evolution 34:611-612

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. Trends in Ecology & Evolution 11:424-429

Jermy T (1984) Evolution of insect host plant relationships. American Naturalist 124:609-630

Jolivet P, Petitpierre E, Hsiao TH (1988) Biology of Chrysomelidae. Kluwer Academic Publishers, Dordrecht

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Molecular Ecology 17:4015-4026

Jost L (2009) *D* vs. G_{ST}: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Molecular Ecology 18:2088-2091

Kanamori Y, Hayakawa Y, Matsumoto H, Yasukochi Y, Shimura S, Nakahara Y, Kiuchi M, Kamimura M (2010) A eukaryotic (insect) tricistronic mRNA encodes three proteins selected by context-dependent scanning. Journal of Biological Chemistry 285:36933-36944

Kappers IF, Hoogerbrugge H, Bouwmeester HJ, Dicke M (2011) Variation in herbivory-induced volatiles among cucumber (*Cucumis sativus* L.) varieties has consequences for the attraction of carnivorous natural enemies. Journal of Chemical Ecology 37:150-160

Kerstes NAG, de Jong PW (2011) Detection of refuge from enemies through phenological mismatching in multitrophic interactions requires season-wide estimation of host abundance. Evolutionary Ecology 25:485-498

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725-738

Kuzina V, Ekstrøm CT, Andersen SB, Nielsen JK, Olsen CE, Bak S (2009) Identification of defense compounds in *Barbarea vulgaris* against the herbivore *Phyllotreta nemorum* by an ecometabolomic approach. Plant Physiology 151:1977-1990

Kuzina V, Nielsen JK, Augustin JM, Torp AM, Bak S, Andersen SB (2011) *Barbarea vulgaris* linkage map and quantitative trait loci for saponins, glucosinolates, hairiness and resistance to the herbivore *Phyllotreta nemorum*. Phytochemistry 72:188-198

Laine AL (2009) Role of coevolution in generating biological diversity: spatially divergent selection trajectories. Journal of Experimental Botany 60:2957-2970

Leng L, Zhang DE (2011) Measuring population differentiation using G_{ST} or *D*? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. Molecular Ecology 20:2494-2509

Lerin J (1980) Influence des substances allélochimiques des crucifères sur les insectes. Acta Oecologica-Oecologia Generalis 1:215-235

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. Genetics 74:175-195

Lively CM (1999) Migration, virulence, and the geographic mosaic of adaptation by parasites. American Naturalist 153:S34-S47

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

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. Evolution 45:622-629

Manel S, Conord C, Després L (2009) Genome-scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BMC Evolutionary Biology 9:288

Marana SR, Jacobs-Lorena M, Terra WR, Ferreira C (2001) Amino acid residues involved in substrate binding and catalysis in an insect digestive beta-glycosidase. Biochimica Et Biophysica Acta 1545:41-52

Medrano M, Herrera CM (2008) Geographical structuring of genetic diversity across the whole distribution range of *Narcissus longispathus*, a habitat-specialist, Mediterranean narrow endemic. Annals of Botany 102:183-194

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5-18

Millipore (2000) Sequencing Reaction Purification Protocol, available trough http://www.millipore.com

Minder AM, Widmer A (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. Molecular Ecology 17:1552-1563

Mopper S, Strauss SY (1998) Genetic structure and local adaptation in natural insect populations: effects of ecology, life history, and behavior. Chapman & Hall, New York

Morant AV, Jorgensen K, Jorgensen C, Paquette SM, Sanchez-Perez R, Moller BL, Bak S (2008) beta-glucosidases as detonators of plant chemical defense. Phytochemistry 69:1795-1813

Morrissey JP, Wubben JP, Osbourn AE (2000) *Stagonospora avenae* secretes multiple enzymes that hydrolyze oat leaf saponins. Molecular Plant-Microbe Interactions 13:1041-1052

Mshvildadze V, Favel A, Delmas F, Elias R, Faure R, Decanosidze G, Kemertelidze E, Balansard G (2000) Antifungal and antiprotozoal activities of saponins from *Hedera colchica*. Pharmazie 55:325-326

Nayar JK, Thorsteinson AJ (1963) Further investigations into chemical basis of insect-host plant relationships in an oligophagous insect, *Plutella Maculipennis* (Curtis) (Lepidoptera - Plutellidae). Canadian Journal of Zoology 41:923-929

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70:3321-3323

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics 41:225-233

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, Columbia

Nei M, Chesser RK (1983) Estimation of fixation indexes and gene diversities. Annals of Human Genetics 47:253-259

Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology and Systematics 28:105-128

Neigel JE (2002) Is F_{ST} obsolete? Conservation Genetics 3:167-173

Neuhauser C, Andow DA, Heimpel GE, May G, Shaw RG, Wagenius S (2003) Community genetics: expanding the synthesis of ecology and genetics. Ecology 84:545-558

Nielsen JK (1977) Host plant relationships of *Phyllotreta nemorum* L. (Coleoptera: Chrysomelidae). I. Field studies. Zeitschrift für Angewandte Entomologie 84:396-407

Nielsen JK (1978) Host plant discrimination within *Cruciferae* - Feeding responses of 4 leaf beetles (Coleoptera - Chrysomelidae) to glucosinolates, cucurbitacins and cardenolides. Entomologia Experimentalis et Applicata 24:41-54

Nielsen JK (1988) Crucifer-feeding Chrysomelidae: Mechanisms of host plant finding and acceptance. In: Jolivet P, Petitpierre E, Hsiao TH (eds) Biology of Chrysomelidae. Kluwer Academic Publishers, Dordrecht, pp. 25-40

Nielsen JK (1989) Host plant relations of *Phyllotreta nemorum* L. (Coleoptera, Chrysomelidae) II. Various defensive mechanisms in plants and their role in defining the host plant range. Journal of Applied Entomology 107: 193-202

Nielsen JK (1996) Intraspecific variability in adult flea beetle behaviour and larval performance on an atypical host plant. Entomologia Experimentalis et Applicata 80:160-162

Nielsen JK (1997a) Genetics of the ability of *Phyllotreta nemorum* larvae to survive in an atypical host plant, *Barbarea vulgaris* ssp *arcuata*. Entomologia Experimentalis Et Applicata 82:37-44

Nielsen JK (1997b) Variation in defences of the plant *Barbarea vulgaris* and in counteradaptations by the flea beetle *Phyllotreta nemorum*. Entomologia Experimentalis Et Applicata 82:25-35

Nielsen JK (1999) Specificity of a Y-linked gene in the flea beetle *Phyllotreta nemorum* for defences in *Barbarea vulgaris*. Entomologia Experimentalis et Applicata 91:359-368

Nielsen JK (2012) Non-random segregation of an autosomal gene in males of the flea beetle, *Phyllotreta nemorum*: implications for colonization of a novel host plant. Entomologia Experimentalis et Applicata 143:301-312

Nielsen JK, de Jong PW (2005) Temporal and host-related variation in frequencies of genes that enable *Phyllotreta nemorum* to utilize a novel host plant, *Barbarea vulgaris*. Entomologia Experimentalis et Applicata 115:265-270

Nielsen JK, Nagao T, Okabe H, Shinoda T (2010a) Resistance in the plant, *Barbarea vulgaris*, and counter-adaptations in flea beetles mediated by saponins. Journal of Chemi-

cal Ecology 36:277-285
Nielsen NJ, Nielsen J, Staerk D (2010b) New resistance-correlated saponins from the insect -resistant crucifer *Barbarea vulgaris*. Journal of Agricultural and Food Chemistry 58:5509-5514
Nisius A (1988) The stromacenter in *Avena* plastids - an aggregation of *β*-glucosidase responsible for the activation of oat-leaf saponins. Planta 173:474-481
Noor MA, Feder JL (2006) Speciation genetics: evolving approaches. Nature Reviews: Genetics 7:851-861
Nosil P (2009) Adaptive population divergence in cryptic color-pattern following a reduction in gene flow. Evolution 63:1902-1912
Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: ‘‘Isolation by adaptation’’ and multiple roles for divergent selection. Evolution 62:316-336
Nuismer SL, Thompson JN (2006) Coevolutionary alternation in antagonistic interactions. Evolution 60:2207-2217
Nuismer SL, Gomulkiewicz R, Morgan MT (2003) Coevolution in temporally variable environments. American Naturalist 162:195-204
Nuismer SL, Gomulkiewicz R, Ridenhour BJ (2010) When is correlation coevolution? American Naturalist 175: 525-537

Ohta T, Kimura M (1973) Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. Genetical Research 22:201-204
Ørgaard M, Linde-Laursen I (2007) Cytogenetics of Danish species of *Barbarea* (Brassicaceae): chromocentres, chromosomes and rDNA sites. Hereditas 144:159-170
Ørgaard M, Linde-Laursen I (2008) Meiotic analysis of Danish species of *Barbarea* (Brassicaceae) using FISH: chromosome numbers and rDNA sites. Hereditas 145:215-219
Orr HA (2005) The genetic theory of adaptation: a brief history. Nature Reviews Genetics 6:119-127
Osbourn A (1996) Saponins and plant defence - A soap story. Trends in Plant Science 1:4-9
Osbourn A, Bowyer P, Lunness P, Clarke B, Daniels M (1995) Fungal pathogens of oat roots and tomato leaves employ closely related enzymes to detoxify different host plant saponins. Molecular Plant-Microbe Interactions 8:971-978
Osbourn AE, Bowyer P, Daniels MJ (1996) Saponin detoxification by plant pathogenic fungi. Advances in Experimental Medicine and Biology 404:547-555

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998) What molecules can tell us about populations: Choosing and using a molecular marker. Ecology 79:361-382
Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295
Pellmyr O (2003) Yuccas, yucca moths, and coevolution: A review. Annals of the Missouri Botanical Garden 90:35-55

Peterson MA, Denno RF (1998) The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. American Naturalist 152:428-446

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Pritchard VL, Jones K, Metcalf JL, Martin AP, Wilkinson P, Cowley DE (2007) Characterization of tetranucleotide microsatellites for Rio Grande cutthroat trout and rainbow trout, and their cross-amplification in other cutthroat trout subspecies. Molecular Ecology Notes 7:594-596

Pritchard JK, Wen W, Falush D (2010) Documentation for *structure* software: Version 2.3, available from http://pritch.bsd.uchicago.edu/structure.html

Rausher MD (1988) Is coevolution dead? Ecology 69:898-901

Rausher MD (1996) Genetic analysis of coevolution between plants and their natural enemies. Trends in Genetics 12:212-217

Renwick JAA (2002) The chemical world of crucivores: lures, treats and traps. Entomologia Experimentalis et Applicata 104:35-42

Reznick DN, Ghalambor CK (2005) Selection in nature: experimental manipulations of natural populations. Integrative and Comparative Biology 45:456-462

Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). Molecular Ecology 14:351-361

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137-138

Runemark A, Hansson B, Pafilis P, Valakos ED, Svensson EI (2010) Island biology and morphological divergence of the Skyros wall lizard *Podarcis gaigeae*: a combined role for local selection and genetic drift on color morph frequency divergence? BMC Evolutionary Biology 10:269

Ryman N, Leimar O (2009) G_{ST} is still a useful measure of genetic differentiation - a comment on Jost's *D*. Molecular Ecology 18:2084-2087

Sachdevgupta K, Radke CD, Renwick JAA (1993a) Antifeedant activity of cucurbitacins from *Iberis amara* against larvae of *Pieris rapae*. Phytochemistry 33:1385-1388

Sachdevgupta K, Radke CD, Renwick JAA, Dimock MB (1993b) Cardenolides from *Erysimum cheiranthoides* - Feeding deterrents to *Pieris rapae* larvae. Journal of Chemical Ecology 19:1355-1369

Sagan L (1966) On the origin of mitosing cells. Journal of Theoretical Biology 14:225-274

Schlotterer C (2002) Towards a molecular characterization of adaptation in local populations. Current Opinion in Genetics & Development 12:683-687

Schlotterer C (2003) Hitchhiking mapping - Functional genomics from the population genetics perspective. Trends in Genetics 19:32-38

Schluter D, Marchinko KB, Barrett RD, Rogers SM (2010) Natural selection and the genetics of adaptation in threespine stickleback. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 365:2479-2486

Schmitt TM, Hay ME, Lindquist N (1995) Constraints on chemically mediated coevolution – multiple functions for seaweed secondary metabolites. Ecology 76:107-123

Schneider CJ (2008) Exploiting genomic resources in studies of speciation and adaptive radiation of lizards in the genus *Anolis*. Integrative and Comparative Biology 48:520-526

Schoonhoven LM (2005) Insect-plant relationships: the whole is more than the sum of its parts. Entomologia Experimentalis et Applicata 115:5-6

Schoonhoven LM, van Loon JJA, Dicke M (2005) Insect-plant biology. Oxford University Press, Oxford

Sefc KM, Payne RB, Sorensen MD (2007) Genetic differentiation after founder events: an evaluation of F_{ST} estimators with empirical and simulated data. Evolutionary Ecology Research 9:21-39

Serizawa H, Shinoda T, Kawai A (2001) Occurrence of a feeding deterrent in *Barbarea vulgaris* (Brassicales : Brassicaceae), a crucifer unacceptable to the diamondback moth, *Plutella xylostella* (Lepidoptera : Plutellidae). Applied Entomology and Zoology 36:465-470

Shelton AM, Nault BA (2004) Dead-end trap cropping: a technique to improve management of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). Crop Protection 23:497-503

Shinoda T, Nagao T, Nakayama M, Serizawa H, Koshioka M, Okabe H, Kawai A (2002) Identification of a triterpenoid saponin from a crucifer, *Barbarea vulgaris*, as a feeding deterrent to the diamondback moth, *Plutella xylostella*. Journal of Chemical Ecology 28:587-599

Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics - a guide for the uninitiated. Trends in Genetics 15:354-358

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983-993

Slatkin M (1981) Estimating levels of gene flow in natural populations. Genetics 99:323-335

Slatkin M (1985) Gene flow in natural populations. Annual Review of Ecology and Systematics 16:393-430

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

SMART™ RACE cDNA Amplification Kit User Manual (2007) available from http://www.genex.cl/stock/clon634914.html

Sparg SG, Light ME, van Staden J (2004) Biological activities and distribution of plant saponins. Journal of Ethnopharmacology 94:219-243

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. Genetics 135:367-374

Stephens C (2010) Forces and causes in evolutionary theory. Philosophy of Science 77: 716 -727

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158-170

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14:671-688

Strong DR, Lawton JH, Southwood R (1984) Insects on plants : community patterns and mechanisms. Harvard University Press, Cambridge

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144:389-399

Talekar NS, Shelton AM (1993) Biology, ecology, and management of the diamondback moth. Annual Review of Entomology 38:275-301

Thompson JN (1988) Coevolution and alternative hypotheses on insect plant interactions. Ecology 69:893-895

Thompson JN (1994) The coevolutionary process. University of Chicago Press, Chicago

Thompson JN (1999a) The evolution of species interactions. Science 284:2116-2118

Thompson JN (1999b) Specific hypotheses on the geographic mosaic of coevolution. American Naturalist 153:S1-S14

Thompson JN (2005) The geographic mosaic of coevolution. University of Chicago Press, Chicago

Thompson JN (2009a) The coevolving web of life. American Naturalist 173:125-140

Thompson JN (2009b) Which ecologically important traits are most likely to evolve rapidly? Oikos 118:1281-1283

Thompson JN (2013) Relentless evolution. University of Chicago Press, Chicago

Thompson JN, Cunningham BM (2002) Geographic structure and dynamics of coevolutionary selection. Nature 417:735-738

Thompson JN, Pellmyr O (1992) Mutualism with pollinating seed parasites amid co-pollinators - constraints on specialization. Ecology 73:1780-1791

Thompson JN, Cunningham BM, Seagraves KA, Althoff DM, Wagner D (1997) Plant polyploidy and insect/plant interactions. American Naturalist 150:730-743

Thompson JN, Nuismer SL, Gomulkiewicz R (2002) Coevolution and maladaptation. Integrative and Comparative Biology 42:381-387

Thompson JN, Laine AL, Thompson JF (2010) Retention of mutualism in a geographically diverging interaction. Ecology Letters 13: 1368-1377

Thorsteinson AJ (1953) The chemotactic responses that determine host specificity in an oligophagous insect (*Plutella maculipennis* (Curt.): Lepidoptera). Canadian Journal of Zoology 31:52-72

Toju H, Sota T (2006) Imbalance of predator and prey armament: geographic clines in phenotypic interface and natural selection. American Naturalist 167:105-117

Toju H, Abe H, Ueno S, Miyazawa Y, Taniguchi F, Sota T, Yahara T (2011) Climatic gradients of arms race coevolution. American Naturalist 177: 562-573

Toneatto F (2009) Genetics and distribution of insect resistance polymorphism in *Barbarea vulgaris*. PhD thesis, University of Copenhagen

Toneatto F, Nielsen JK, Ørgaard M, Hauser TP (2010) Genetic and sexual separation between insect resistant and susceptible *Barbarea vulgaris* plants in Denmark. Molecular Ecology 19:3456-3465

Turner EM (1961) An enzymic basis for pathogenic specificity in *Ophiobolus graminis*. Journal of Experimental Botany 12:169-175

van Leur H, Raaijmakers CE, van Dam NM (2006) A heritable glucosinolate polymorphism within natural populations of *Barbarea vulgaris*. Phytochemistry 67:1214-1223

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

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology 14:3623-3642

Verbaarschot P, Calvo D, Esselink GD, Molina JM, Vrieling K, de Jong PW (2007) Isolation of polymorphic microsatellite loci from the flea beetle *Phyllotreta nemorum* L. (Coleoptera: Chrysomelidae). Molecular Ecology Notes 7:60-62

Verschaffelt E (1910) The cause determining the selection of food in some herbivorous insects. Proceedings of the Royal Academy, Amsterdam 13:536-542

Vincken JP, Heng L, de Groot A, Gruppen H (2007) Saponins, classification and occurrence in the plant kingdom. Phytochemistry 68:275-297

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Human Molecular Genetics 2:1123-1128

Weir BS (1996) Genetic data analysis II: Methods for discrete population genetic data. Sinauer Associates, Inc., Sunderland

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370

Weissenberg M, Levy A, Svoboda JA, Ishaaya I (1998) The effect of some *Solanum* steroidal alkaloids and glycoalkaloids on larvae of the red flour beetle, *Tribolium castaneum*, and the tobacco hornworm, *Manduca sexta*. Phytochemistry 47:203-209

Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T (2007) The genetic basis of a plant-insect coevolutionary key innovation. Proceedings of the National Academy of Sciences of the United States of America 104:20427-20431

Whitlock MC (2011) G'_{ST} and D do not replace F_{ST}. Molecular Ecology 20:1083-1091

Whittaker RH, Feeny PP (1971) Allelochemics: chemical interactions between species. Science 171:757-770

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97-159

Wright S (1938) Size of population and breeding structure in relation to evolution. Science 87:430-431

Wright S (1951) The genetical structure of populations. Annals of Eugenics 15:323-354

Wright S (1968) Evolution and the genetics of populations: a treatise in four volumes. University of Chicago Press, Chicago

Wubben JP, Price KR, Daniels MJ, Osbourn AE (1996) Detoxification of oat leaf saponins by *Septoria avenae*. Phytopathology 86:986-992

Xu H, Sarkar B, George V (2009) A new measure of population structure using multiple single nucleotide polymorphisms and its relationship with F_{ST}. BMC Research Notes 2:21

Yamane H, Konno K, Sabelis M, Takabayashi J, Sassa T, Oikawa H (2010) Chemical defence and toxins of plants. In: Mander L, Lui HW (eds) Comprehensive natural products II. Chemistry and biology. Elsevier, Oxford, pp. 339-385

Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution 7: 2123-2129

Zangerl AR, Berenbaum MR (2003) Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. Evolution 57:806-815

Zhang JD, Cao YB, Xu Z, Sun HH, An MM, Yan L, Chen HS, Gao PH, Wang Y, Jia XM, Jiang YY (2005) In vitro and in vivo antifungal activities of the eight steroid saponins from *Tribulus terrestris* L. with potent activity against fluconazole-resistant fungal pathogens. Biological & Pharmaceutical Bulletin 28:2211-2215

Plants are chemically defended against insect herbivory in various ways. They produce a broad range of secondary metabolites that may be toxic or deterrent to insects. Specialist insects, however, are often capable of overcoming these defences. The yellow striped flea beetle (*Phyllotreta nemorum* L.) is a specialist that feeds on crucifers (Brassicaceae) such as *Sinapis arvensis* and *Barbarea vulgaris*. In Denmark, two types of *Barbarea vulgaris* var. *arcuata* are distinguished: one with **p**ubescent leaves (P-type) and one with **g**labrous leaves (G-type). All individuals of *P. nemorum* can feed on *B. vulgaris* P-type. *Barbarea vulgaris* G-type, on the other hand, is chemically defended against most *P. nemorum* individuals during the flea beetle reproductive season. The defence compounds are hypothesized to be saponins, a class of compounds with various biological effects and insecticidal properties. Despite high levels of these saponins during summer, some flea beetles can and do feed on *B. vulgaris* G-type. The ability of *P. nemorum* to feed on *B. vulgaris* G-type is heritable; resistance against the defence of *B. vulgaris* G-type is controlled by dominant major resistance genes (R-genes). One dominant R-allele of an R-gene is enough to convert a susceptible beetle into a resistant one. Despite knowledge of the inheritance patterns of resistance in the flea beetles, which have been demonstrated to be variable, the underlying mechanism of flea beetle resistance has, so far, remained unclear. This prompted me to investigate, as an initial part of my thesis, the genetic basis of the flea beetle adaptation to the defence of *B. vulgaris* G-type.

The interaction between *B. vulgaris* and the flea beetle is a unique natural model system to study chemical defences in plants and counter-adaptations in insects. Plant and insect are both polymorphic with respect to the trait involved in resistance and hereby provide an excellent opportunity to study the geographic aspects of the evolution of the resistance trait in both interacting species. In this thesis, I focus on the resistance of the flea beetle, and take the presence of different genotypes of the plant as a given. *Phyllotreta nemorum* is a major pest, for example in oil seed rape. Understanding how resistance evolves in *P. nemorum* will not only benefit flea beetle control, but also control of other pest insects. Understanding insect resistance includes knowledge of seasonal, geographic and genetic variation in both plant defense and herbivore adaptation. The R-gene has a remarkable distribution. Flea beetle populations living on *B. vulgaris* G-type consist solely of resistant individuals, but on host plant patches nearby *B. vulgaris* G-type lower frequencies of resistant beetles are found than one would expect with the amount of gene flow found at the neutral level between these subpopulations.

The aim of this thesis was to find the gene that is held responsible for the resistance of *P. nemorum* to the defences of *B. vulgaris*, investigate the distribution of this resistance trait and explain the distribution of this trait in natural populations. The following questions were addressed: (1) what is the genetic basis of the adaptation under study? (2) how is the resistance distributed across flea beetle populations in Denmark? and (3) which factors underlie this distribution?

In order to answer these questions, I used an integrated approach. I have combined a candidate gene approach (CHAPTER 3) with an empirical approach via the study of variation in

resistance in flea beetle populations (CHAPTER 4), and a population genomics approach by using molecular markers to gain insight in the genomic make-up of the population and its connection with the resistance trait (CHAPTERS 5 and 6). The population genomics approach is a recent advance in methods to detect the involvement of selection in the distribution of alleles at presumably adaptive loci. Using this approach one can distinguish locus-specific effects, like directional selection, from genome-wide effects, on the distribution of alleles at loci of interest.

The population genomics approach is introduced in CHAPTER 2 together with the Geographic Mosaic Theory of Coevolution. I illustrate how processes underlying this theory of coevolution can be investigated with the population genomics approach. According to the geographic mosaic theory of coevolution, reciprocal selection between interacting species only happens in so-called hot-spots. Hot spots can be identified using population genomics and genetic variation found at specific loci can be attributed to locus-specific processes such as directional selection. For the *B. vulgaris* - flea beetle system this means that with a population genomics approach we can examine whether the distribution of resistant flea beetles on alternative host plants is only influenced by migration, or also by selection (CHAPTER 5). Another valuable utility of the population genomics approach is to investigate whether a candidate gene for the R-gene is under selection, by looking whether a candidate gene is experiencing locus-specific effects beside genome-wide effects when comparing flea beetle populations living on *B. vulgaris* G-type with populations living on alternative host plants (CHAPTER 6).

However, before using a population genomics approach to compare the resistance trait or a candidate gene with parts of the genome that only experience genome-wide effects, I have tried to identify the genetic basis of the flea beetle adaptation to the defence of *B. vulgaris* G-type. In CHAPTER 3, I have addressed this question by using a candidate gene approach to examine the involvement of a possible detoxifying enzyme in *P. nemorum*. Genes coding for β-glucosidase were a candidate for genes underlying the difference between resistant and susceptible beetles, because β-glucosidase is used as detoxifying enzyme by other organisms resistant to saponin defence. Three different β-glucosidase cDNA sequences were cloned from Danish flea beetle lines. We named them β-glucosidase A, B and C. β-glucosidase C was only found in resistant lines and not in the susceptible line. We then tested if recombinant β-glucosidase C breaks down the most abundant and most effective defence compound in *B. vulgaris* G-type, hederagenin cellobioside. β-glucosidase C was able to deglycosylate one glucose unit of hederagenin cellobioside, when expressed in an insect cell line. This suggests that expressed β-glucosidase C can deglycosylate antifeedant saponins and may play a role in the resistant flea beetle's ability to overcome the defence of *B. vulgaris*. Next, a segregating family was created in which offspring differed in resistance genotype. Again β-glucosidase cDNA sequences were cloned to find a difference in the presence of these β-glucosidases between resistant and susceptible individuals. This time cDNA sequences of β-glucosidases A, B and C were present in both resistant and susceptible individuals although significantly fewer β-glucosidase C cDNA sequence variants were found in susceptible individuals than in resistant individuals. Thus, the genet-

ic basis of flea beetle resistance remains unclear. Further investigation is needed to explore if the β-glucosidase C protein is also capable of inactivating hederagenin cellobioside by hydrolysizing the second glucose unit from the saponin and if there is a difference in enzyme activity of β-glucosidase C between resistant and susceptible beetles.

Subsequently, in CHAPTER 4 I have investigated whether the frequency of resistant beetles decreased in populations living on other host plant patches than *B. vulgaris* G-type and whether the change in frequency was significant within the flea beetle season. I found that the frequency of resistant beetles varied significantly among years, but there was no evidence for a decrease in the frequency of resistant beetles, the latter being expected if selection acts against the resistance on other host plants than *B. vulgaris* G-type. Furthermore, I found that the frequency of resistant beetles varied significantly within a flea beetle season. This study demonstrates that relative frequencies of different resistance phenotypes of *P. nemorum* on other host plants than *B. vulgaris* G-type are highly dynamic, both within and across years. It is, therefore, important to sample season-wide when one wants to monitor the changes in frequencies of insect resistance in natural systems.

In CHAPTERS 5 and 6 I took a population genomics approach to investigate if the observed geographical distribution of resistance of *P. nemorum* to chemically defended *B. vulgaris* in flea beetle populations could be explained by factors that are solely associated with genome-wide effects, such as migration, or also by locus-specific factors like selection at the resistance locus. First, neutral microsatellites were used to reveal the genetic differentiation at parts of the genome that are only influenced by genome-wide processes. Next, the level of neutral genetic differentiation was compared with the genetic differentiation found for the resistance trait. The resistance trait was an outlier in pairwise comparisons between flea beetle populations on *B. vulgaris* and *S. arvensis*, meaning that the level of genetic differentiation was significantly higher than expected if the resistance trait experiences only genome-wide effects. The resistance trait was also an outlier in the pairwise comparison between populations on *S. arvensis*, which suggests that the resistance trait is also under directional selection on other host plants than *B. vulgaris* G-type.

Additionally, I examined in CHAPTER 6 if the homologous β-glucosidases B and C sequences found in CHAPTER 3 correspond to two alleles of the major resistance gene, because of their similarity and their presence in flea beetle lines. The sequence of β-glucosidases C had so far only been found in resistant individuals, so we hypothesized it to be the dominant resistance allele and the sequence of β-glucosidases B would then correspond to the susceptible allele. In order to find out if this hypothesized PneR-gene (***P**hyllotreta **ne**morum* **R**-gene) is the resistance gene, we first directly compared resistance phenotypes of beetles collected from populations on *B. vulgaris* G-type and *S. arvensis* with genotypes derived with primers developed for β-glucosidase B and C. The phenotype of the flea beetles did not match the genotype derived with the β-glucosidase primers. Additionally, the frequency of heterozygotes and homozygotes of the PneR-gene genotype was not significantly deviating from Hardy-Weinberg Equilibrium which implies that there are no locus-specific effects involved when both sequences are seen as one gene with two alleles. A population approach was

taken like in CHAPTER 5, this time including the genetic differentiation estimated for the candidate gene as well. The candidate gene behaved similar to the neutral loci while the resistance trait was an outlier in most pairwise comparisons between flea beetle populations. If both sequences are alleles of the same gene, then the candidate gene is not directly responsible for the flea beetle resistance to *B. vulgaris* G-type defence.

The results presented in this thesis show the complexity of genetic processes (either genome-wide or locus specific) affecting local adaptation and the distribution of a resistance trait in insects in natural populations. Furthermore, the present study shows that when studying coevolution between insect and host plant by means of adaptive traits, also geographical and seasonal variation in allele frequencies should be considered. A multidisciplinary approach to study adaptation in plant-insect interactions such as used in this thesis, will benefit research on plant-insect interactions, including applied research such as studying the potential of host plants as dead-end traps for pest insects and preventing/diminishing the development of resistance by pest insects to crop defences.

Planten zijn op verschillende manieren beschermd tegen insectenvraat door herbivoren. Ze produceren o.a. een scala aan secundaire metabolieten die toxisch of insectenafwerend kunnen zijn. Specialistische insecten zijn vaak in staat om deze plantenafweer te omzeilen of ongedaan te maken. De grote gestreepte aardvlo (*Phyllotreta nemorum* L.) is een specialist die zich voedt met kruisbloemigen (Brassicaceae) zoals *Sinapis arvensis* en *Barbarea vulgaris*. In Denemarken worden twee types *Barbarea vulgaris* var. *arcuata* onderscheiden: de één heeft harige bladeren (P-type genaamd) en de andere heeft gladde bladeren (G-type genaamd). Alle *P. nemorum* individuen kunnen zich voeden met *B. vulgaris* P-type. *Barbarea vulgaris* G-type, daarentegen, is chemisch beschermd tegen de meeste *P. nemorum* kevers tijdens het kevervoortplantingsseizoen. De stoffen waarmee de plant zich verdedigt tegen onder andere *P. nemorum* zijn saponinen. Dit zijn stoffen met verscheidene biologische effecten waaronder insecticideachtige. Ondanks de hoge saponinegehaltes tijdens de zomer, kunnen sommige aardvlooien zich toch tegoed doen aan *B. vulgaris* G-type. De mogelijkheid van de aardvlo om zich te kunnen voeden met *B. vulgaris* G-type is erfelijk; resistentie tegen de afweer van *B. vulgaris* G-type wordt gecontroleerd door dominante “major” resistentie genen (R-genen). Eén dominant R-allel van een R-gen is al voldoende om een anders gevoelige kever resistent te maken; vandaar dat ze “major” genoemd worden. Ondanks kennis van de overervingspatronen van resistentie van de aardvlooien, is het onderliggend mechanisme van de aardvlo resistentie dusverre nog niet bekend. Dit heeft mij ertoe gezet te onderzoeken wat de genetische basis is van de aardvlo-adaptatie aan de afweer van *B. vulgaris* G-type.

De interactie tussen *B. vulgaris* en de aardvlo is een uniek natuurlijk modelsysteem waarmee de chemische afweer van planten en de adaptatie van insecten hieraan te bestuderen. Plant en insect zijn beiden polymorf wat betreft de eigenschap die ten grondslag ligt aan hun resistentie en bieden zodoende een excellente kans om de geografische aspecten van de evolutie van resistentie te bestuderen in beide soorten van de interactie. In dit proefschrift richt ik me op de resistentie van de aardvlo en beschouw daarbij de verschillende resistentie genotypen van de plant als gegeven. *Phyllotreta nemorum* is een belangrijk plaaginsect, bijvoorbeeld van koolzaad. Begrijpen hoe resistentie zich ontwikkelt in *P. nemorum* zal niet alleen voordelig zijn voor de bestrijding van aardvlooien, maar ook helpen bij de bestrijding van andere plaaginsecten. Om het resistentiemechanisme van insecten te begrijpen is kennis over seizoensbrede, geografische en genetische variatie in zowel de afweer van planten als de adaptatie van insecten nodig. Het R-gen heeft een opmerkelijke distributie. Zo bestaan aardvlopopulaties die zich bevinden op *B. vulgaris* G-type volledig uit resistente individuen, maar op andere soorten waardplanten nabij *B. vulgaris* G-type bevinden zich lagere frequenties resistente kevers dan men verwacht op basis van de hoeveelheid genetische uitwisseling tussen deze populaties die gevonden is voor neutrale delen van het aardvlo-genoom.

Het doel van dit proefschrift was het vinden van het gen dat verantwoordelijk is voor de resistentie van *P. nemorum* tegen de afweer van *B. vulgaris*, de geografische, en waardplantgerelateerde verspreiding van dit resistentiegen te onderzoeken en dit te verklaren in natuurlijke populaties. De volgende onderzoeksvragen werden daarbij geformuleerd: (1) wat is de

genetische basis van de bestudeerde adaptatie? (2) hoe is resistentie tegen de afweer van *B. vulgaris* G-type verspreid over aardvlopopulaties in Denemarken? en (3) welke factoren liggen ten grondslag aan deze distributie?

Om tot een antwoord te komen op deze vragen, heb ik een geïntegreerde aanpak gebruikt. Ik heb een zogenoemde "kandidaatgen aanpak" (HOOFDSTUK 3, vrije vertaling van de term *candidate gene approach*) gecombineerd met veldonderzoek en bioassays door middel van de studie van variatie in resistentie in de aardvlopopulaties (HOOFDSTUK 4), en een zogenaamde "populatiegenomics aanpak" (vrije vertaling van de term *population genomics approach*) met behulp van moleculaire markers om inzichten te verwerven in de genomische opmaak van de populatie en zijn connectie met de resistentie-eigenschap (HOOFDSTUKKEN 5 en 6). De populatiegenomics aanpak is een recente ontwikkeling in de methodiek die gebruikt wordt om de betrokkenheid van selectie bij de distributie van allelen van veronderstelde adaptieve loci te onderzoeken. Door deze aanpak te gebruiken, kan men onderscheid maken tussen locus-specifieke effecten, zoals gerichte selectie, en genoom-wijde effecten op de distributie van allelen van loci waarin men geïnteresseerd is.

De populatiegenomics aanpak wordt geïntroduceerd in HOOFDSTUK 2 samen met de Geografische Mozaïek Theorie over Coevolutie. Ik illustreer hoe processen die ten grondslag liggen aan deze theorie over coevolutie kunnen worden onderzocht met de populatiegenomics aanpak. Volgens de geografische mozaïek theorie over coevolutie vindt wederzijdse selectie tussen op elkaar inwerkende soorten alleen plaats in zogenoemde "hot-spots". Deze "hot-spots" kunnen worden geïdentificeerd door middel van populatiegenomics en ook genetische variatie die gevonden wordt op specifieke loci kan met deze methode worden toegeschreven aan locus-specifieke processen zoals gerichte selectie. Voor het *Barbarea vulgaris – Phyllotreta nemorum* systeem betekent dit dat we met een populatiegenomics aanpak kunnen onderzoeken of de distributie van resistente kevers over andere waardplanten alleen beïnvloed wordt door migratie of ook door selectie (HOOFDSTUK 5). Een ander waardevol gebruik van de populatiegenomics aanpak is het onderzoeken of een kandidaatsgen voor het R-gen onder selectie staat door te onderzoeken of het kandidaatsgen beïnvloed wordt door alleen genoom-brede effecten of ook locus-specifieke effecten wanneer aardvlopopulaties op *B. vulgaris* G-type worden vergeleken met populaties op andere waardplanten (HOOFDSTUK 6).

Voordat echter een populatiegenomics aanpak kon worden toegepast om de resistentie-eigenschap of een kandidaatsgen te kunnen vergelijken met delen van het genoom die alleen genoom-wijde effecten ondervinden, probeerde ik te achterhalen wat de genetische basis is van de aardvlo adaptatie aan de afweer van *B. vulgaris* G-type. Dit deed ik in HOOFDSTUK 3 door middel van een kandidaatsgen aanpak, om te onderzoeken of een mogelijk saponine-onschadelijk makend enzym is betrokken bij de resistentie van *P. nemorum*. Genen die coderen voor β-glucosidase waren kandidaten voor genen die aan het verschil tussen resistente en gevoelige kevers ten grondslag konden liggen, omdat β-glucosidase door andere organismen wordt gebruikt als gif-onklaar makend enzym tegen afweer door saponinen. Drie verschillende β-glucosidase cDNA sequenties werden geklo-

neerd uit Deense aardvlolijnen. Deze werden β-glucosidase A, B en C genoemd. β-glucosidase C werd alleen aangetroffen in de resistente lijnen en niet in de gevoelige lijn. Vervolgens testten we of recombinante β-glucosidase C in staat was om de meest voorkomende en meest effectieve stof in *B. vulgaris* G-type, hederagenin cellobioside, af te breken. Wanneer het eiwit β-glucosidase C tot expressie werd gebracht in een insecten cellijn, bleek β-glucosidase C in staat om één glucose eenheid te hydrolyseren van hederagenin cellobioside. Dit suggereert dat de tot expressie gebrachte β-glucosidase C de insectenafwerende saponinen kan deglycosyleren en wellicht een rol speelt in het vermogen van de resistente aardvlooien om *B. vulgaris* G-type het hele jaar door te gebruiken als waardplant. Daarna is een segregerende aardvlofamilie gecreëerd waarin het nageslacht verschilde in resistentiegenotype. Wederom werden β-glucosidase cDNA sequenties gekloneerd om te onderzoeken of er een duidelijk verschil bestond in het voorkomen van deze β-glucosidasen in resistente en gevoelige individuen van één familie. Ditmaal waren cDNA sequenties van β-glucosidasen A, B and C allen present in zowel resistente als gevoelige individuen, al werden er significant minder varianten van de β-glucosidase C cDNA sequentie gevonden in gevoelige individuen dan in resistente individuen. Derhalve blijft de genetische basis van de aardvloresistentie nog onopgehelderd. Voor meer duidelijkheid hieromtrent is er meer onderzoek nodig om erachter te komen of het β-glucosidase C eiwit ook in staat is hederagenin cellobioside te inactiveren door de andere glucose eenheid ook te hydrolyseren en of er een verschil is in β-glucosidase C enzymactiviteit tussen de resistente en gevoelige aardvlo types.

Vervolgens heb ik in HOOFDSTUK 4 onderzocht of de frequenties van resistente kevers daalden in populaties die zich bevinden op andere waardplanten dan *B. vulgaris* G-type, zoals eerder werd gesuggereerd door bevindingen in het veld. Ook onderzocht ik of de frequenties van resistente kevers op andere waardplanten dan *B. vulgaris* G-type significant veranderden binnen een aardvloseizoen. De frequentie resistente kevers bleek significant te variëren tussen de jaren, maar ik vond geen bewijs voor een verdere daling in de frequentie resistente kevers op andere waardplanten dan *B. vulgaris* G-type, wat men zou verwachten als er selectie plaatsvindt tegen resistentie op andere waardplanten dan *B. vulgaris* G-type. Verder bleek dat de frequentie resistente kevers significant varieerde binnen een aardvloseizoen. Dit onderzoek laat zien dat relatieve frequenties van verschillende aardvlo resistentie fenotypes op andere waardplanten dan *B. vulgaris* G-type erg dynamisch zijn, zowel binnen als tussen jaren. Het is daarom belangrijk dat men door het seizoen heen bemonstert als men de veranderingen in resistentiefrequenties in insecten wil monitoren in natuurlijke systemen.

In HOOFDSTUKKEN 5 en 6 gebruikte ik een populatiegenomics aanpak om te onderzoeken of de waargenomen geografische verdeling van de aardvloresistentie tegen de afweer van *B. vulgaris* G-type in aardvlopopulaties geheel kan worden verklaard door factoren die alleen betrekking hebben op genoombrede invloeden zoals migratie, of dat locus-specifieke factoren zoals selectie ook een rol spelen in de waargenomen geografische verspreiding van resistente fenotypes. Eerst werden neutrale microsatellieten gebruikt om de genetische differentiatie van de delen van het genoom te achterhalen die alleen onder invloed staan van ge-

noombrede processen. Aansluitend werd het niveau van neutrale genetische differentiatie vergeleken met de genetische differentiatie die gevonden werd voor de resistentie eigenschap. De resistentie-eigenschap vertoonde een uitschieter in een index voor differentiatie in paarsgewijze vergelijkingen tussen aardvlo populaties op *B. vulgaris* en die op *S. arvensis*, wat betekent dat het niveau van genetische differentiatie significant hoger was dan verwacht werd als de resistentie eigenschap alleen genoombrede effecten zou ondervinden. Daarnaast was de resistentie-eigenschap ook een uitschieter in de paarsgewijze vergelijking tussen populaties op *S. arvensis*, wat een aanwijzing is dat de resistentie-eigenschap ook onder selectie staat in populaties die zich bevinden op andere waardplanten dan *B. vulgaris* G-type.

Tevens onderzocht ik in HOOFDSTUK 6 of de homologe β-glucosidase B en C sequenties die gevonden waren in HOOFDSTUK 3 overeen kwamen met twee allelen van het resistentiegen, vanwege hun overeenkomst en hun aanwezigheid in de onderzochte aardvlolijnen. De sequentie van β-glucosidases C was tot dan toe alleen aangetroffen in resistente individuen, dus we veronderstelden dat dit het dominante resistentie allel was en de sequentie van β-glucosidase B werd verondersteld overeen te komen met het gevoelige allel. Om te onderzoeken of dit veronderstelde PneR-gen (van "**P**hyllotreta **ne**morum **R**-gen") het resistentie gen is, vergeleken we eerst direct het resistentie fenotype van verzamelde kevers van populaties op *B. vulgaris* G-type en *S. arvensis* met het PneR-gen genotype dat we afleidden met behulp van primers ontwikkeld om β-glucosidase B en C aan te tonen. Het fenotype van de aardvlooien kwam niet overeen met het genotype dat we afleidden met behulp van de β-glucosidase primers. Ook week de frequentie van heterozygoten en homozygoten voor het PneR-gen genotype niet significant af van het Hardy-Weinberg Evenwicht wat impliceert dat er geen locus specifieke effecten invloed hebben op het veronderstelde PneR-gen. Daaropvolgend werd een populatie genomics aanpak gebruikt zoals in HOOFDSTUK 5, maar ditmaal werd ook de geschatte genetische differentiatie van het kandidaatsgen meegenomen in de analyse. Het veronderstelde PneR-gen gedroeg zich net als de neutrale loci in de analyse terwijl de resistentie eigenschap een uitschieter was in de meeste paarsgewijze vergelijkingen tussen aardvlo populaties. Hieruit volgt dat als beide sequenties allelen zijn van hetzelfde gen, dit kandidaat gen dan niet direct verantwoordelijk is voor aardvlo resistentie tegen de afweer van *B. vulgaris* G-type.

De in dit proefschrift gepresenteerde resultaten laten zien dat complexiteit van genetische processen (of het nu genoomwijde of locus specifieke zijn), lokale adaptatie en de distributie van een resistentie eigenschap in insecten in natuurlijke populaties beïnvloedt. Verder laat dit onderzoek zien dat ook geografische en seizoensbrede variatie in allelfrequenties in acht moet worden genomen wanneer men coevolutie tussen plant en insect bestudeert aan de hand van adaptieve eigenschappen. Een multidisciplinaire aanpak zoals in dit proefschrift is gebruikt, zal een toegevoegde waarde hebben voor onderzoek aan plant-insect interacties, inclusief toegepast onderzoek zoals naar de potentie van waardplanten als "doodspoor vallen" (vrije vertaling van de term *dead-end traps*) voor insecten en het voorkomen en inperken van resistentie-ontwikkeling door plaaginsecten tegen de afweer van gewassen.

Allereerst wil ik mijn dagelijkse begeleider bedanken voor de tijd die hij in me stak, zijn geduld en altijd positieve instelling. Peter, bedankt. Ik wil ook Patrick Verbaarschot bedanken, voor zijn expertise in the moleculair laboratorium en vruchtvolle discussies over het aardvlo project. En natuurlijk wil ik graag het grote opperhoofd Marcel Dicke bedanken voor de mogelijkheid die hij me heeft gegund om hier op Entomologie een PhD te doen. Bedankt voor jullie begeleiding!

In general I would like to thank the people at Entomology. Present and past colleagues and also the students that did their project under mine and Peter's supervision during my PhD traject. Because of the people at Entomology there was always a nice working atmosphere. I want to thank separately, Nelson, who supported me not only in the lab but also outside. Although you are still in the mids of your PhD traject, you have to know that I look up to you and your professional attitude. You helped me focus on my thesis and get my priorities set right.

I would like to thank the people in Denmark as well that helped me during my thesis traject. Especially both Jens and Jens I would like to thank, not only for their support but also their hospitality and good food. Whenever I was in Denmark, I would find a warm welcome in Frederiksen 's house in Copenhagen and I could always count on Kvist 's kindness, helpfulness and knowledge. Also would I like to express my gratitude to Manabu Kamimura for his comments, collaboration and the fruitful discussions we had about the flea beetle project.

En natuurlijk! Ik ben ook mijn vrienden dankbaar voor het vertier tijdens mijn promotietraject. In het bijzonder mijn oud-studiegenootjes Gijs, Minou, Dianne, Gwen en Fedor. En ook de Warhammermannen, Joep, Mark, Frans-Jan, Sjaak en Martijn.

Ik wil ook graag mijn ouders bedanken. Ze hebben me altijd gesteund in wat ik dan ook wilde doen. Dit boek was nooit tot stand gekomen zonder jullie. Laat ik ook vooral niet vergeten mijn zusters te bedanken. Anouk en Yvette, bedankt!

Kim Maria Cornelia Annette Vermeer was born on 31 October, 1982 in Tilburg, The Netherlands. After attending secondary school Theresia lyceum (1995 to 2001) in Tilburg, she started the study Biology at Wageningen University in the Netherlands. After the first year, she received her propedeuse and in 2007 she graduated at Wageningen University in Biology with the specialization Ecology and Biodiversity. During her Master she decided to do her MSc thesis at the laboratory of Entomology under the supervision of Dr. P.W. de Jong. She investigated if flea beetles (*Phyllotreta nemorum*) had co-adapted gene complexes that enabled them to survive on an otherwise defended variety of bittercress (*Barbarea vulgaris*). She was then asked to join the first edition of the Battle of the Universities in the Netherlands in the team led by Prof. Marcel Dicke. Their team successfully won the Academic Year Prize for translating excellent scientific research to the general public (organized by NRC-Handelsblad, Netherlands Organization for Scientific Research and Royal Netherlands Academy of Sciences). Immediately after graduation she could start a PhD project at the laboratory of Entomology with Prof. M. Dicke as promoter and Dr. P.W. de Jong as co-promoter. The results of this project are presented in this thesis.

Vermeer KMCA, Dicke M and De Jong PW (2011) The potential of a population genomics approach to analyse geographic mosaics of plant--insect coevolution. Evolutionary Ecology 25:977-992

Vermeer KMCA, Verbaarschot P and De Jong PW (2012) Changes in frequencies of genes that enable *Phyllotreta nemorum* to utilize its host plant, *Barbarea vulgaris*, vary in magnitude and direction as much within as between seasons. Entomologia Experimentalis et Applicata 144:37-44

De Jong PW, Breuker CJ, Vos H, **Vermeer KMCA**, Oku K, Verbaarschot P, Nielsen JK and Brakefield PM (2009) Genetic differentiation between resistance phenotypes in the phytophagous flea beetle, *Phyllotreta nemorum*. Journal of Insect Science 9:1-8

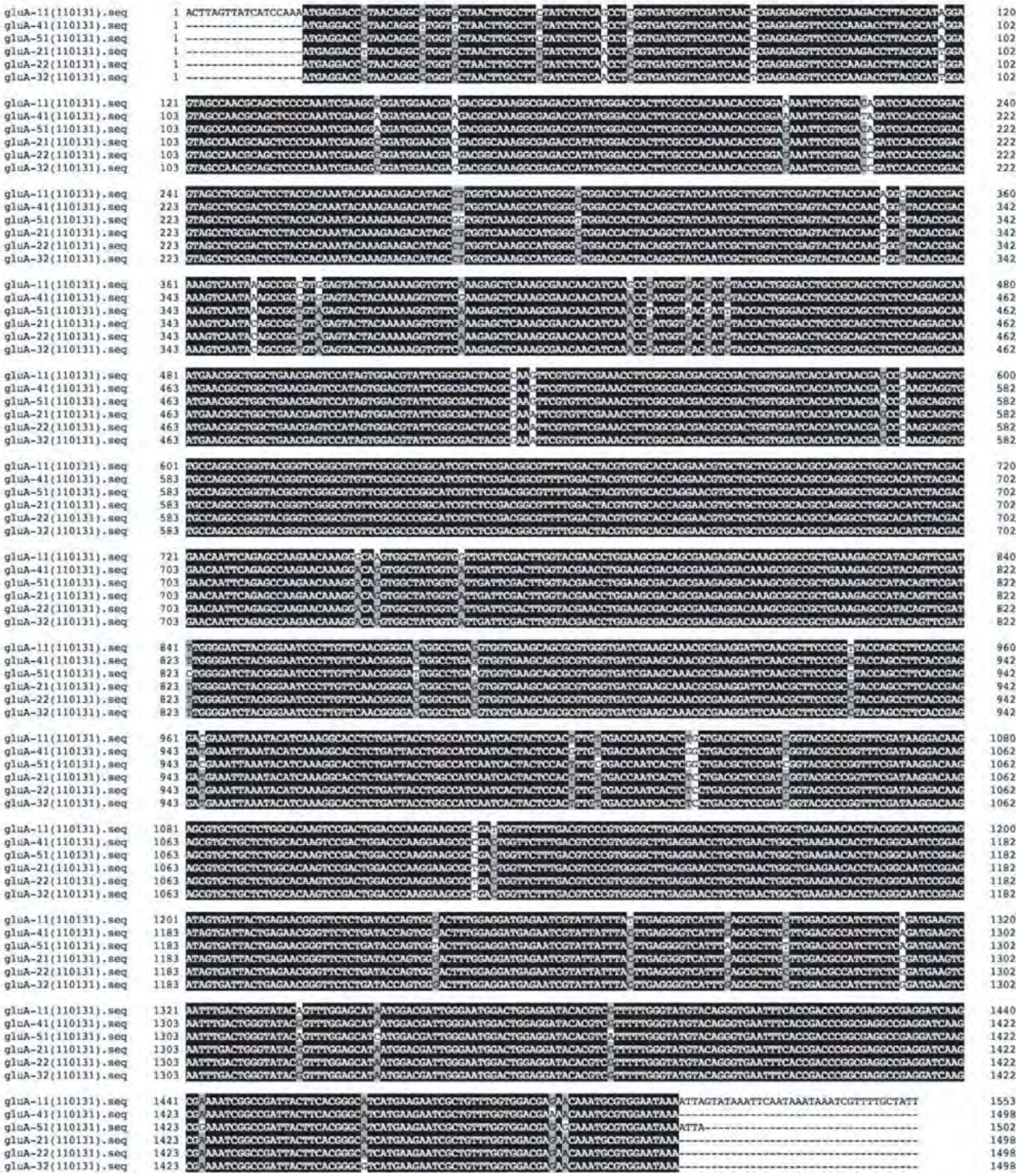
Oku K, **Vermeer KMCA**, Verbaarschot P and De Jong PW (2010) Effects of starvation and mating status on the activity of the flea beetle, *Phyllotreta nemorum* (Coleoptera: Chrysomelidae). European Journal of Entomology 107:549-551

To be submitted

Vermeer KMCA, Ikink G, Verbaarschot P, De Jong P and Kamimura M. Identification of saponin degrading beta-glucosidases from the flea beetle *Phyllotreta nemorum*.

Vermeer KMCA, Verbaarschot P, Dicke M and De Jong P. Detection of selection against resistance of a flea beetle to host plant defences.

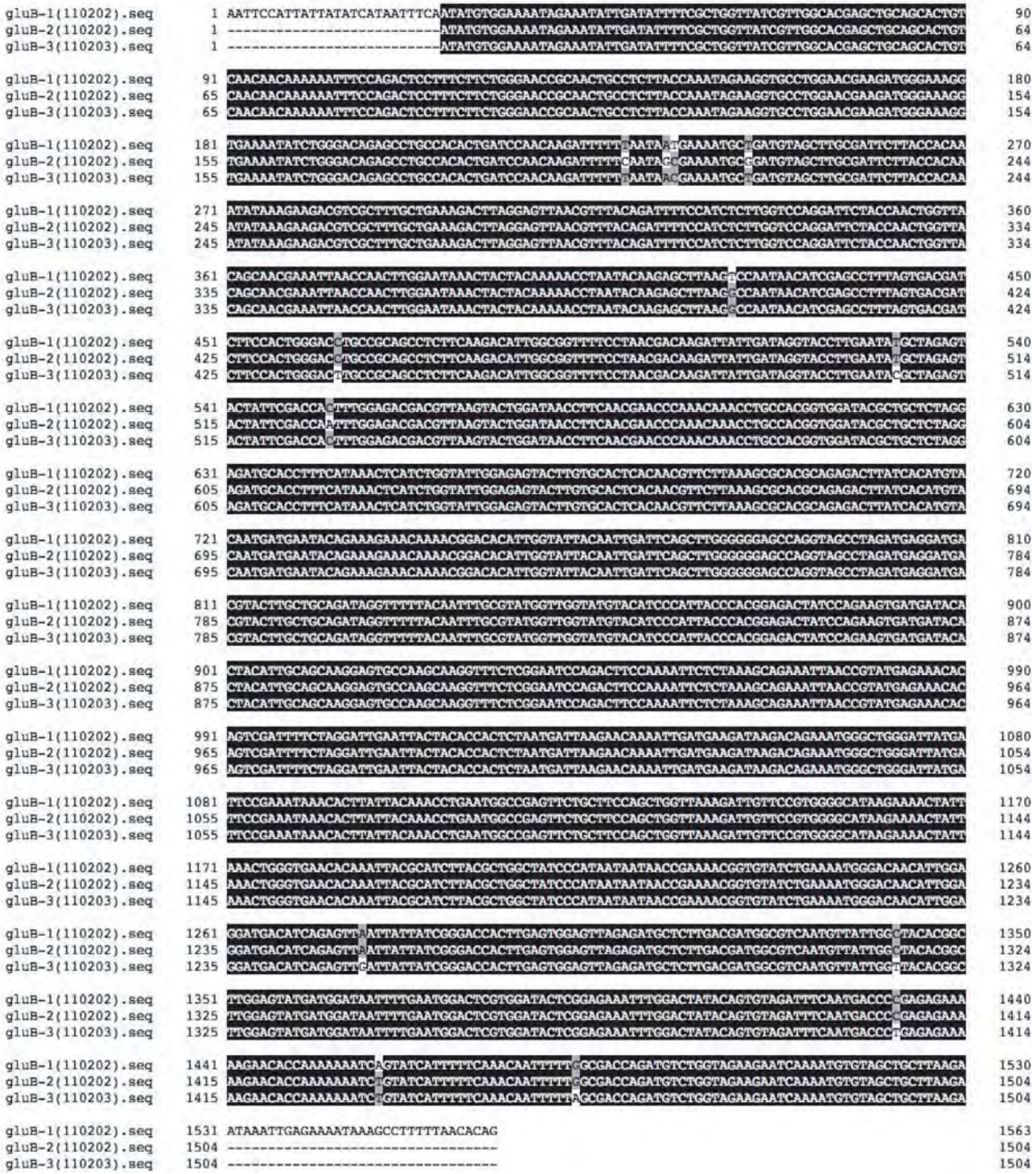
a)


gluA-11(110131).seq
gluA-41(110131).seq
gluA-51(110131).seq
gluA-21(110131).seq
gluA-22(110131).seq
gluA-32(110131).seq

b)


gluA-11(110131).pep 1 MRTLTGLVLTCLLSLSLGDGSINPRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-41(110131).pep 1 MRTLTGLVLTCLLSLSLGDGSINPRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-51(110131).pep 1 MRTLTGLVLTCLLSLSLGDGSINPRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-21(110131).pep 1 MRTLTGLVLTCLLSLNLGDGSINSRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-22(110131).pep 1 MRTLTGLVLTCLLSLNLGDGSINSRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-32(110131).pep 1 MRTLTGLVLTCLLSLNLGDGSINSRRFPKTLRIGVANAAPQIEGGWNEDGKGETIWDHFA 60
gluA-11(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNKAG 120
gluA-41(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNKAG 120
gluA-51(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNKAG 120
gluA-21(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNTAG 120
gluA-22(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNTAG 120
gluA-32(110131).pep 61 HKHPEKFVDRSTPDVACDSYHKYKEDIALVKAMGLDHYRLSIAWSRVLPTGYTDKVNTAG 120
gluA-11(110131).pep 121 VEYYKKVFKELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-41(110131).pep 121 VEYYKKVFEELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-51(110131).pep 121 VEYYKKVFKELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-21(110131).pep 121 VEYYKKVFKELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-22(110131).pep 121 VEYYKKVFKELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-32(110131).pep 121 VEYYKKVFKELKANNIKPMVTIYHWDLPQPLQEQMNGWLNESIVDVFGDYAKFVFETFGD 180
gluA-11(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-41(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-51(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-21(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-22(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-32(110131).pep 181 DADWWITINEPKQVCQAGYGSGVFAPGIVSDGVLDYVCTRNVLLAHARAWHIYDEQFRAK 240
gluA-11(110131).pep 241 NKGQVAMVVDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-41(110131).pep 241 NKGQVAMVIDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-51(110131).pep 241 NKGQVAMVIDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-21(110131).pep 241 NKGQVAMVIDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-22(110131).pep 241 NKGQVAMVIDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-32(110131).pep 241 NKGQVAMVIDSTWYEPGSDSEEDKAAAERAIQFDLGIYGNPLFNGDWPEVVKQRVGDRSK 300
gluA-11(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSADAPIGTPGFDKDKSVLLWH 360
gluA-41(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSADAPIGTPGFDKDKSVLLWH 360
gluA-51(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSADAPIGTPGFDKDKSVLLWH 360
gluA-21(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSEDAPIGTPGFDKDKSVLLWH 360
gluA-22(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSPDAPIGTPGFDKDKSVLLWH 360
gluA-32(110131).pep 301 REGFNASRLPAFTEDEIKYIKGTSDYLAINHYSTLLTNHSPDAPIGTPGFDKDKSVLLWH 360
gluA-11(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-41(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-51(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-21(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-22(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-32(110131).pep 361 KSDWTQGSADWFFDVPWGLRNLLNWLKNTYGNPEIVITENGFSDTSGTLEDENRIIYLRG 420
gluA-11(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-41(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-51(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-21(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-22(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-32(110131).pep 421 HLSACLDAIFSDEVNLTGYTVWSIMDDWEWTGGYTSFLGMYRVNFTDPARPRIKRKSADY 480
gluA-11(110131).pep 481 FTGIMKNRCLVDENKCVE 498
gluA-41(110131).pep 481 FTGIMKNRCLVDENKCVE 498
gluA-51(110131).pep 481 FTGIMKNRCLVDESKCVE 498
gluA-21(110131).pep 481 FTGIMKNRCLVDENKCVE 498
gluA-22(110131).pep 481 FTGIMKNRCLVDENKCVE 498
gluA-32(110131).pep 481 FTGVMKNRCLVDENKCVE 498

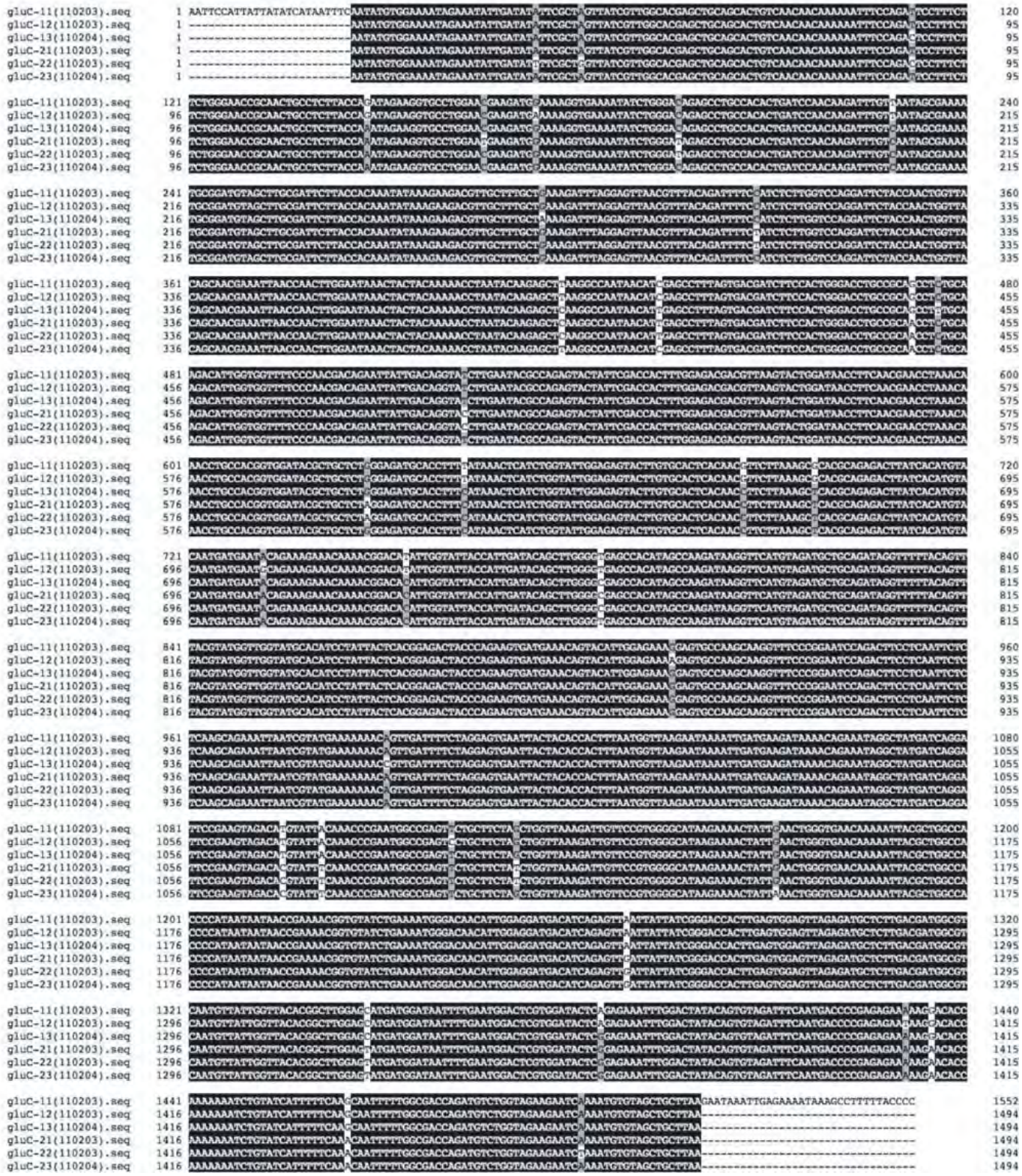
c)


gluB-1(110202).seq 1 AATTCCATTATTATATCATAATTTCAATATGTGGAAAATAGAAATATTGATATTTTCGCTGGTTATCGTTGGCACGAGCTGCAGCACTGT 90
gluB-2(110202).seq 1 ----------ATATGTGGAAAATAGAAATATTGATATTTTCGCTGGTTATCGTTGGCACGAGCTGCAGCACTGT 64
gluB-3(110203).seq 1 ----------ATATGTGGAAAATAGAAATATTGATATTTTCGCTGGTTATCGTTGGCACGAGCTGCAGCACTGT 64
gluB-1(110202).seq 91 CAACAACAAAAAATTTCCAGACTCCTTTCTTCTGGGAACCGCAACTGCCTCTTACCAAATAGAAGGTGCCTGGAACGAAGATGGGAAAGG 180
gluB-2(110202).seq 65 CAACAACAAAAAATTTCCAGACTCCTTTCTTCTGGGAACCGCAACTGCCTCTTACCAAATAGAAGGTGCCTGGAACGAAGATGGGAAAGG 154
gluB-3(110203).seq 65 CAACAACAAAAAATTTCCAGACTCCTTTCTTCTGGGAACCGCAACTGCCTCTTACCAAATAGAAGGTGCCTGGAACGAAGATGGGAAAGG 154
gluB-1(110202).seq 181 TGAAAATATCTGGGACAGAGCCTGCCACACTGATCCAACAAGATTTTTTAATAATGAAAATGCTGATGTAGCTTGCGATTCTTACCACAA 270
gluB-2(110202).seq 155 TGAAAATATCTGGGACAGAGCCTGCCACACTGATCCAACAAGATTTTTCAATACGAAAATGCGGATGTAGCTTGCGATTCTTACCACAA 244
gluB-3(110203).seq 155 TGAAAATATCTGGGACAGAGCCTGCCACACTGATCCAACAAGATTTTTTAATACGAAAATGCTGATGTAGCTTGCGATTCTTACCACAA 244
gluB-1(110202).seq 271 ATATAAAGAAGACGTCGCTTTGCTGAAAGACTTAGGAGTTAACGTTTACAGATTTTCCATCTCTTGGTCCAGGATTCTACCAACTGGTTA 360
gluB-2(110202).seq 245 ATATAAAGAAGACGTCGCTTTGCTGAAAGACTTAGGAGTTAACGTTTACAGATTTTCCATCTCTTGGTCCAGGATTCTACCAACTGGTTA 334
gluB-3(110203).seq 245 ATATAAAGAAGACGTCGCTTTGCTGAAAGACTTAGGAGTTAACGTTTACAGATTTTCCATCTCTTGGTCCAGGATTCTACCAACTGGTTA 334
gluB-1(110202).seq 361 CAGCAACGAAATTAACCAACTTGGAATAAACTACTACAAAAACCTAATACAAGAGCTTAAGCCCAATAACATCGAGCCTTTAGTGACGAT 450
gluB-2(110202).seq 335 CAGCAACGAAATTAACCAACTTGGAATAAACTACTACAAAAACCTAATACAAGAGCTTAAGCCCAATAACATCGAGCCTTTAGTGACGAT 424
gluB-3(110203).seq 335 CAGCAACGAAATTAACCAACTTGGAATAAACTACTACAAAAACCTAATACAAGAGCTTAAGCCCAATAACATCGAGCCTTTAGTGACGAT 424
gluB-1(110202).seq 451 CTTCCACTGGGACCTGCCGCAGCCTCTTCAAGACATTGGCGGTTTTCCTAACGACAAGATTATTGATAGGTACCTTGAATATGCTAGAGT 540
gluB-2(110202).seq 425 CTTCCACTGGGACCTGCCGCAGCCTCTTCAAGACATTGGCGGTTTTCCTAACGACAAGATTATTGATAGGTACCTTGAATATGCTAGAGT 514
gluB-3(110203).seq 425 CTTCCACTGGGACTTGCCGCAGCCTCTTCAAGACATTGGCGGTTTTCCTAACGACAAGATTATTGATAGGTACCTTGAATACGCTAGAGT 514
gluB-1(110202).seq 541 ACTATTCGACCACTTTGGAGACGACGTTAAGTACTGGATAACCTTCAACGAACCCAAACAAACCTGCCACGGTGGATACGCTGCTCTAGG 630
gluB-2(110202).seq 515 ACTATTCGACCAATTTGGAGACGACGTTAAGTACTGGATAACCTTCAACGAACCCAAACAAACCTGCCACGGTGGATACGCTGCTCTAGG 604
gluB-3(110203).seq 515 ACTATTCGACCACTTTGGAGACGACGTTAAGTACTGGATAACCTTCAACGAACCCAAACAAACCTGCCACGGTGGATACGCTGCTCTAGG 604
gluB-1(110202).seq 631 AGATGCACCTTTCATAAACTCATCTGGTATTGGAGAGTACTTGTGCACTCACAACGTTCTTAAAGCGCACGCAGAGACTTATCACATGTA 720
gluB-2(110202).seq 605 AGATGCACCTTTCATAAACTCATCTGGTATTGGAGAGTACTTGTGCACTCACAACGTTCTTAAAGCGCACGCAGAGACTTATCACATGTA 694
gluB-3(110203).seq 605 AGATGCACCTTTCATAAACTCATCTGGTATTGGAGAGTACTTGTGCACTCACAACGTTCTTAAAGCGCACGCAGAGACTTATCACATGTA 694
gluB-1(110202).seq 721 CAATGATGAATACAGAAAGAAACAAAACGGACACATTGGTATTACAATTGATTCAGCTTGGGGGGAGCCAGGTAGCCTAGATGAGGATGA 810
gluB-2(110202).seq 695 CAATGATGAATACAGAAAGAAACAAAACGGACACATTGGTATTACAATTGATTCAGCTTGGGGGGAGCCAGGTAGCCTAGATGAGGATGA 784
gluB-3(110203).seq 695 CAATGATGAATACAGAAAGAAACAAAACGGACACATTGGTATTACAATTGATTCAGCTTGGGGGGAGCCAGGTAGCCTAGATGAGGATGA 784
gluB-1(110202).seq 811 CGTACTTGCTGCAGATAGGTTTTTACAATTTGCGTATGGTTGGTATGTACATCCCATTACCCACGGAGACTATCCAGAAGTGATGATACA 900
gluB-2(110202).seq 785 CGTACTTGCTGCAGATAGGTTTTTACAATTTGCGTATGGTTGGTATGTACATCCCATTACCCACGGAGACTATCCAGAAGTGATGATACA 874
gluB-3(110203).seq 785 CGTACTTGCTGCAGATAGGTTTTTACAATTTGCGTATGGTTGGTATGTACATCCCATTACCCACGGAGACTATCCAGAAGTGATGATACA 874
gluB-1(110202).seq 901 CTACATTGCAGCAAGGAGTGCCAAGCAAGGTTTCTCGGAATCCAGACTTCCAAAATTCTCTAAAGCAGAAATTAACCGTATGAGAAACAC 990
gluB-2(110202).seq 875 CTACATTGCAGCAAGGAGTGCCAAGCAAGGTTTCTCGGAATCCAGACTTCCAAAATTCTCTAAAGCAGAAATTAACCGTATGAGAAACAC 964
gluB-3(110203).seq 875 CTACATTGCAGCAAGGAGTGCCAAGCAAGGTTTCTCGGAATCCAGACTTCCAAAATTCTCTAAAGCAGAAATTAACCGTATGAGAAACAC 964
gluB-1(110202).seq 991 AGTCGATTTTCTAGGATTGAATTACTACACCACTCTAATGATTAAGAACAAAATTGATGAAGATAAGACAGAAATGGGCTGGGATTATGA 1080
gluB-2(110202).seq 965 AGTCGATTTTCTAGGATTGAATTACTACACCACTCTAATGATTAAGAACAAAATTGATGAAGATAAGACAGAAATGGGCTGGGATTATGA 1054
gluB-3(110203).seq 965 AGTCGATTTTCTAGGATTGAATTACTACACCACTCTAATGATTAAGAACAAAATTGATGAAGATAAGACAGAAATGGGCTGGGATTATGA 1054
gluB-1(110202).seq 1081 TTCCGAAATAAACACTTATTACAAACCTGAATGGCCGAGTTCTGCTTCCAGCTGGTTAAAGATTGTTCCGTGGGGCATAAGAAAACTATT 1170
gluB-2(110202).seq 1055 TTCCGAAATAAACACTTATTACAAACCTGAATGGCCGAGTTCTGCTTCCAGCTGGTTAAAGATTGTTCCGTGGGGCATAAGAAAACTATT 1144
gluB-3(110203).seq 1055 TTCCGAAATAAACACTTATTACAAACCTGAATGGCCGAGTTCTGCTTCCAGCTGGTTAAAGATTGTTCCGTGGGGCATAAGAAAACTATT 1144
gluB-1(110202).seq 1171 AAACTGGGTGAACACAAATTACGCATCTTACGCTGGCTATCCCATAATAATAACCGAAAACGGTGTATCTGAAAATGGGACAACATTGGA 1260
gluB-2(110202).seq 1145 AAACTGGGTGAACACAAATTACGCATCTTACGCTGGCTATCCCATAATAATAACCGAAAACGGTGTATCTGAAAATGGGACAACATTGGA 1234
gluB-3(110203).seq 1145 AAACTGGGTGAACACAAATTACGCATCTTACGCTGGCTATCCCATAATAATAACCGAAAACGGTGTATCTGAAAATGGGACAACATTGGA 1234
gluB-1(110202).seq 1261 GGATGACATCAGAGTTAATTATTATCGGGACCACTTGAGTGGAGTTAGAGATGCTCTTGACGATGGCGTCAATGTTATTGGCTACACGGC 1350
gluB-2(110202).seq 1235 GGATGACATCAGAGTTAATTATTATCGGGACCACTTGAGTGGAGTTAGAGATGCTCTTGACGATGGCGTCAATGTTATTGGCTACACGGC 1324
gluB-3(110203).seq 1235 GGATGACATCAGAGTTGATTATTATCGGGACCACTTGAGTGGAGTTAGAGATGCTCTTGACGATGGCGTCAATGTTATTGGTTACACGGC 1324
gluB-1(110202).seq 1351 TTGGAGTATGATGGATAATTTTGAATGGACTCGTGGATACTCGGAGAAATTTGGACTATACAGTGTAGATTTCAATGACCCCGAGAGAAA 1440
gluB-2(110202).seq 1325 TTGGAGTATGATGGATAATTTTGAATGGACTCGTGGATACTCGGAGAAATTTGGACTATACAGTGTAGATTTCAATGACCCCGAGAGAAA 1414
gluB-3(110203).seq 1325 TTGGAGTATGATGGATAATTTTGAATGGACTCGTGGATACTCGGAGAAATTTGGACTATACAGTGTAGATTTCAATGACCCTGAGAGAAA 1414
gluB-1(110202).seq 1441 AAGAACACCAAAAAAATCAGTATCATTTTTCAAACAATTTTTTGCGACCAGATGTCTGGTAGAAGAATCAAAATGTGTAGCTGCTTAAGA 1530
gluB-2(110202).seq 1415 AAGAACACCAAAAAAATCTGTATCATTTTTCAAACAATTTTTTGCGACCAGATGTCTGGTAGAAGAATCAAAATGTGTAGCTGCTTAAGA 1504
gluB-3(110203).seq 1415 AAGAACACCAAAAAAATCTGTATCATTTTTCAAACAATTTTTTAGCGACCAGATGTCTGGTAGAAGAATCAAAATGTGTAGCTGCTTAAGA 1504
gluB-1(110202).seq 1531 ATAAATTGAGAAAATAAAGCCTTTTTAACACAG 1563
gluB-2(110202).seq 1504 ---------- 1504
gluB-3(110203).seq 1504 ---------- 1504

d)


gluB-1(110202).pep 1 MWKIEILIFSLVIVGTSCSTVNNKKFPDSFLLGTATASYQIEGAWNEDGKGENIWDRACH 60
gluB-2(110202).pep 1 MWKIEILIFSLVIVGTSCSTVNNKKFPDSFLLGTATASYQIEGAWNEDGKGENIWDRACH 60
gluB-3(110203).pep 1 MWKIEILIFSLVIVGTSCSTVNNKKFPDSFLLGTATASYQIEGAWNEDGKGENIWDRACH 60
gluB-1(110202).pep 61 TDPTRFFNNENADVACDSYHKYKEDVALLKDLGVNVYRFSISWSRILPTGYSNEINQLGI 120
gluB-2(110202).pep 61 TDPTRFFNSENADVACDSYHKYKEDVALLKDLGVNVYRFSISWSRILPTGYSNEINQLGI 120
gluB-3(110203).pep 61 TDPTRFFNNENADVACDSYHKYKEDVALLKDLGVNVYRFSISWSRILPTGYSNEINQLGI 120
gluB-1(110202).pep 121 NYYKNLIQELKSNNIEPLVTIFHWDLPQPLQDIGGFPNDKIIDRYLEYARVLFDHFGDDV 180
gluB-2(110202).pep 121 NYYKNLIQELKANNIEPLVTIFHWDLPQPLQDIGGFPNDKIIDRYLEYARVLFDQFGDDV 180
gluB-3(110203).pep 121 NYYKNLIQELKANNIEPLVTIFHWDLPQPLQDIGGFPNDKIIDRYLEYARVLFDHFGDDV 180
gluB-1(110202).pep 181 KYWITFNEPKQTCHGGYAALGDAPFINSSGIGEYLCTHNVLKAHAETYHMYNDEYRKKQN 240
gluB-2(110202).pep 181 KYWITFNEPKQTCHGGYAALGDAPFINSSGIGEYLCTHNVLKAHAETYHMYNDEYRKKQN 240
gluB-3(110203).pep 181 KYWITFNEPKQTCHGGYAALGDAPFINSSGIGEYLCTHNVLKAHAETYHMYNDEYRKKQN 240
gluB-1(110202).pep 241 GHIGITIDSAWGEPGSLDEDDVLAADRFLQFAYGWYVHPITHGDYPEVMIHYIAARSAKQ 300
gluB-2(110202).pep 241 GHIGITIDSAWGEPGSLDEDDVLAADRFLQFAYGWYVHPITHGDYPEVMIHYIAARSAKQ 300
gluB-3(110203).pep 241 GHIGITIDSAWGEPGSLDEDDVLAADRFLQFAYGWYVHPITHGDYPEVMIHYIAARSAKQ 300
gluB-1(110202).pep 301 GFSESRLPKFSKAEINRMRNTVDFLGLNYYTTLMIKNKIDEDKTEMGWDYDSEINTYYKP 360
gluB-2(110202).pep 301 GFSESRLPKFSKAEINRMRNTVDFLGLNYYTTLMIKNKIDEDKTEMGWDYDSEINTYYKP 360
gluB-3(110203).pep 301 GFSESRLPKFSKAEINRMRNTVDFLGLNYYTTLMIKNKIDEDKTEMGWDYDSEINTYYKP 360
gluB-1(110202).pep 361 EWPSSASSWLKIVPWGIRKLLNWVNTNYASYAGYPIIITENGVSENGTTLEDDIRVNYYR 420
gluB-2(110202).pep 361 EWPSSASSWLKIVPWGIRKLLNWVNTNYASYAGYPIIITENGVSENGTTLEDDIRVNYYR 420
gluB-3(110203).pep 361 EWPSSASSWLKIVPWGIRKLLNWVNTNYASYAGYPIIITENGVSENGTTLEDDIRVDYYR 420
gluB-1(110202).pep 421 DHLSGVRDALDDGVNVIGYTAWSMMDNFEWTRGYSEKFGLYSVDFNDPERKRTPKKSVSF 480
gluB-2(110202).pep 421 DHLSGVRDALDDGVNVIGYTAWSMMDNFEWTRGYSEKFGLYSVDFNDPERKRTPKKSVSF 480
gluB-3(110203).pep 421 DHLSGVRDALDDGVNVIGYTAWSMMDNFEWTRGYSEKFGLYSVDFNDPERKRTPKKSVSF 480
gluB-1(110202).pep 481 FKQFLATRCLVEESKCVAA 499
gluB-2(110202).pep 481 FKQFLATRCLVEESKCVAA 499
gluB-3(110203).pep 481 FKQFLATRCLVEESKCVAA 499

e)


gluC-11(110203).seq
gluC-12(110203).seq
gluC-13(110204).seq
gluC-21(110203).seq
gluC-22(110203).seq
gluC-23(110204).seq

f)

Figure S1 Alignments of cDNA and amino acid sequences of the three *Phyllotreta nemorum* β-glucosidases:
a) alignments of cDNA sequences of β-glucosidase A,
b) alignment of amino acid sequences of β-glucosidase A,
c) alignments of cDNA sequences of β-glucosidase B,
d) alignment of amino acid sequences of β-glucosidase B,
e) alignments of cDNA sequences of β-glucosidase C,
f) alignment of amino acid sequences of β-glucosidase C.

Table S1 Genetic differentiation estimators and their 95% confidence interval (CI) per pair of subpopulations for each microsatellite, for the resistance phenotype and for the PneR-gene. Within subpopulation heterozygosity is given as well.

	pairwise comparison between location A and B														
	microsatellite marker														
	PnA03	PnA04	PnA10	PnAB12	PnB11.311	PnC05.11	PnD04	PnD06	PnD09	PnE08.61	PnE11	PnF03	PnH09	resistance	PneR
H_S	0.799	0.443	0.555	0.862	0.324	0.556	0.726	0.752	0.75	0.744	0.824	0.376	0.429	0.381	0.274
G_{ST}	0.004	0.005	0.017	0.003	0.021	0.047	0.001	0.006	0.018	0.008	0.010	0.008	0.031	0.193	0.014
CI min	0	-0.002	0.002	-0.001	0.006	0.005	-0.002	0	-0.002	0.001	0.003	0	-0.008	0.119	-0.002
CI max	0.010	0.021	0.040	0.008	0.039	0.107	0.006	0.013	0.050	0.017	0.019	0.022	0.093	0.282	0.043
G'_{ST}	0.035	0.013	0.058	0.038	0.040	0.162	0.005	0.039	0.120	0.051	0.103	0.017	0.055	0.426	0.024
CI min	0.001	-0.005	0.006	-0.007	0.012	0.016	-0.011	0	-0.015	0.007	0.027	0	-0.017	0.280	-0.003
CI max	0.087	0.054	0.137	0.105	0.078	0.324	0.035	0.092	0.296	0.116	0.195	0.051	0.239	0.583	0.075
D_{EST}	0.032	0.008	0.042	0.036	0.020	0.123	0.004	0.033	0.105	0.044	0.094	0.010	0.054	0.293	0.011
CI min	0.001	-0.003	0.004	-0.006	0.006	0.012	-0.009	0	-0.013	0.006	0.024	0	-0.009	0.183	-0.001
CI max	0.078	0.032	0.101	0.097	0.039	0.243	0.030	0.080	0.255	0.100	0.178	0.029	0.167	0.418	0.033

Continuation of Table S1 Genetic differentiation estimators and their 95% confidence interval (CI) per pair of subpopulations for each microsatellite, for the resistance phenotype and for the PneR-gene. Within subpopulation heterozygosity is given as well.

pairwise comparison between location A and C

microsatellite marker	PnA03	PnA04	PnA10	PnAB12	PnB11.311	PnC05.11	PnD04	PnD06	PnD09	PnE08.61	PnE11	resistance	PneR
H_S	0.817	0.446	0.65	0.823	0.308	0.555	0.765	0.766	0.743	0.762	0.813	0.405	0.138
G_{ST}	0.003	0.007	0.012	0.011	0.029	0.007	0.025	0.005	0.019	0.013	0.007	0.250	0.037
CI min	0	-0.002	0.004	0.003	0.011	-0.003	0.017	-0.001	-0.003	0.005	0	0.191	0.011
CI max	0.008	0.023	0.023	0.023	0.052	0.028	0.036	0.013	0.061	0.022	0.017	0.322	0.067
G'_{ST}	0.033	0.017	0.058	0.116	0.054	0.025	0.190	0.035	0.122	0.096	0.070	0.408	0.048
CI min	0	-0.006	0.019	0.034	0.020	-0.010	0.128	-0.007	-0.025	0.036	0.003	0.299	0.013
CI max	0.080	0.061	0.110	0.222	0.098	0.092	0.260	0.096	0.344	0.164	0.161	0.538	0.092
D_{EST}	0.030	0.011	0.047	0.106	0.026	0.018	0.170	0.031	0.106	0.084	0.064	0.214	0.012
CI min	0	-0.004	0.016	0.031	0.010	-0.007	0.114	-0.006	-0.022	0.031	0.003	0.134	0.003
CI max	0.073	0.039	0.086	0.204	0.049	0.068	0.233	0.085	0.297	0.146	0.146	0.318	0.026

pairwise comparison between location B and C

microsatellite marker	PnA03	PnA04	PnA06	PnA10	PnAB12	PnB11.311	PnC05.11	PnD04	PnD06	PnD09	PnE08.61	PnE11	resistance	PneR
H_S	0.809	0.493	0.799	0.577	0.829	0.213	0.539	0.783	0.75	0.81	0.783	0.797	0.137	0.188
G_{ST}	0.006	0	0.009	0.035	0.021	0.002	0.023	0.017	0.004	0.003	0.005	0.013	0.718	0.087
CI min	0.001	-0.003	0.004	0.019	0.010	-0.002	-0.003	0.009	-0.002	-0.002	0	0.004	0.649	0.051
CI max	0.013	0.008	0.018	0.057	0.036	0.009	0.062	0.026	0.017	0.012	0.013	0.024	0.794	0.127
G'_{ST}	0.053	0	0.085	0.131	0.224	0.003	0.077	0.136	0.030	0.031	0.039	0.115	0.944	0.127
CI min	0.008	-0.008	0.032	0.072	0.113	-0.003	-0.009	0.077	-0.015	-0.015	-0.004	0.035	0.911	0.073
CI max	0.117	0.022	0.158	0.201	0.346	0.014	0.208	0.205	0.112	0.105	0.101	0.210	0.974	0.189
D_{EST}	0.048	0	0.076	0.099	0.207	0.001	0.056	0.122	0.026	0.028	0.034	0.103	0.809	0.045
CI min	0.007	-0.005	0.028	0.054	0.104	-0.001	-0.006	0.068	-0.013	-0.014	-0.003	0.031	0.745	0.021
CI max	0.106	0.015	0.143	0.153	0.321	0.005	0.154	0.184	0.096	0.095	0.089	0.191	0.873	0.074

PE&RC PhD Training Certificate

With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

The potential of a population genomics approach to analyse geographic mosaics of plant-insect coevolution

Post-graduate courses (2.6 ECTS)

Linear mixed models; PE&RC (2009)
Evolutionary dynamics; PE&RC (2009)

Invited review of (unpublished) journal manuscript (1 ECTS)

International Journal of Medicinal Plants Research: phytochemical defence against insects (2013)

Deficiency, refresh, brush-up courses (0.3 ECTS)

Lab safety course (2010)

Competence strengthening / skills courses (4.5 ECTS)

The art of writing; Language services Wageningen (2007)
Competence assessment; Maas Assessment & Development Amersfoort (2007)
Techniques for writing and presenting scientific papers; WGS (2008)
Effective behaviour in your professional surroundings; WGS (2009)
Basic statistics; PE&RC (2009)
Project and time management; WGS (2010)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.7 ECTS)

PE&RC Day (2007-2009)
PE&RC Weekend; beginning PhD students (2008)
PE&RC Weekend; end of PhD traject (2012)

Discussion groups / local seminars / other scientific meetings (7.5 ECTS)

Annual meeting of the Netherlands Entomology Society (2007-2010)
PhD Candidate discussion group; Entomology (2007-2011)
Insect Plant Interaction discussion group (2007-2011)
EPS Symposium: Ecology and Experimental Plant Sciences II (2009)
Plant-Insect workshop (2009)
EPS Symposium: Interaction between plants and biotic agents (2010)

International symposia, workshops and conferences (7.8 ECTS)

Mini-symposium on the interaction between Barbarea, flea beetles and pathogens; oral presentations; Copenhagen (2008, 2009 and 2011)
The Netherlands Annual ecology Meeting; oral and poster presentation; Lunteren (2009-2011)
Plant Insect Interactions Symposium XIIII; poster presentation; Wageningen (2011)

Supervision of 3 MSc students

Mapping of resistance genes in the flea beetle *Phyllotreta nemorum* with microsatellite markers
Food-choice of resistant *Phyllotreta nemorum* beetles between *Barbarea vulgaris* and *Raphanus sativus* plants
The mapping of a resistance gene in the flea beetle *Phyllotreta nemorum*

Cover design and thesis lay-out by author