Brassicaceae plant - *Pieris* butterfly interactions mediated by an arms-race between glucosinolates and the nitrile specifier protein gene family

February 2018

Yu Okamura

Graduate School of Science

Chiba University

Brassicaceae plant - *Pieris* butterfly interactions mediated by an arms-race between glucosinolates and the nitrile specifier protein gene family

General Introduction **1**
References 5

Chapter 1 The effects of different secondary metabolite profiles in plant defense syndromes on specialist and generalist herbivores **8**
Abstract 8
Introduction 9
Materials and Methods 11
Results 14
Discussion 17
References 22
Figures and Tables 27

Chapter 2 Differential larval performance of *Pieris* butterfly species correlates to Glucosinolate diversity and side chain length **33**
Abstract 33
Introduction 35
Materials and Methods 38
Results 45
Discussion 49
References 55
Figures and Tables 64

Chapter 3 *Pieris* butterflies regulate two genes differently for dealing with wider range of glucosinolate profiles in host plants **81**
Abstract 81
Introduction 82
Materials and Methods 84
Results 89
Discussion 91
References 95
Figures and Tables 101

Chapter 4 Microevolution of NSP in *Pieris* butterflies in Japan **108**
Abstract 108
Introduction 110
Materials and Methods 113
Results 121
Discussion 126
References 135
Figures and Tables 142

General Discussion **164**
Acknowledgement 169
References 170

General Introduction

The tremendously high biodiversity observed today on the earth is a consequence of diversification of organisms. Understanding the mechanisms, which have driven this diversification, is one of the most important propositions in ecology. Plants and herbivorous insects contribute to this high biodiversity, and they also have crucial roles in terrestrial ecosystems. Plant–herbivore coevolution theory was raised in 1960s (Ehrlich & Raven 1964), for explaining the mechanisms which have caused this higher diversity of plants and herbivores. According to this theory, plants have defended themselves from herbivores with diversifying its defense traits, and herbivores also simultaneously have evolved to adapt to these defenses. This theory explained both plants and herbivores have experienced the exponential diversification mediated by this arms-race of "defense" and "adaptation". After this theory was launched, a number of researches have assessed this hypothesis (Futuyma & Agrawal 2009).

Brassicales plants and pierid butterfly interaction system have been focused as one of the useful systems to test this hypothesis (Wheat et al. 2007; Edger et al. 2015). Since brassicales contains many agriculturally important crops, such as cabbage or broccoli, and the model plant species *Arabidopsis thaliana*, huge quantity of molecular or genetic information have been accumulated (Halkier & Gershenzon 2006; Wittstock & Halkier 2002). Brassicales plants possess glucosinolate (GLS), one of the secondary metabolites, as a specific chemical defense (Winde & Wittstock 2011). Until recently, at least 140 types of GLS are identified and a part of them are known to have strong defensive abilities against general herbivores (Fahey et al. 2001; Agerbirk & Olsen 2012; Olsen et al. 2016). However, pierid butterflies, a member of small cabbage white, can use brassicales plants as hosts, by detoxifying GLS defense system with larval gut express enzyme called nitriles specifier protein (NSP)

(Wittstock et al. 2004). For understanding the evolutionary consequences of arms-race between plant defense and herbivore adaptation, this brassicales – pierid butterfly system is practically useful because we at least have known a certain part of both defense and adaptation mechanisms in this interaction (Wheat et al. 2007).

Recent phylogenetic comparative researches revealed that brassicales plants diversified their GLS types dramatically via experiencing several whole genome duplication events, and the speciation rates of pierid butterflies also simultaneously raised with these events (Edger et al. 2015). This suggested co-diversification between brassicales plants and pierid butterflies mediated by GLS diversification, and proposed first body of evidence to explain the mechanism of co-diversification. However, important information is still lacking even in this system to confirm this co-diversification, such as ecological consequences of GLS diversification and their differential role against herbivores, impacts of GLSs on host utilization of pierid butterflies, or, evolution of adaptation traits in pierid butterflies and its effect on speciation. Here, in this thesis, I focus on these important questions to understand mechanisms underlying on the bases of arms-race between brassicales and pierid butterflies.

In Chapter 1, I focus on multiple defense strategy of Brassicaceae plants with GLS distribution among them. Plants use several types of defenses as a multiple defense (Agrawal & Fishbein 2006; Silva & Batalha 2010; Travers-Martin & Müller 2008). Assuming higher diversity of GLSs as secondary metabolites as a consequence of defense evolution, each type of GLS have their own ecological role. Defensive role of some types of GLSs have been confirmed independently (Müller et al. 2010; Kos et al. 2012), however, how a variety of GLS shape multiple defense coupled with non-GLS defenses have not been well understood (Travers-Martin & Müller 2008). I approach this question by focusing on Brassicaceae plants, which is one of the most diversified family in brassicales and also possesses highest diversity of GLSs (Hofberger et al. 2013; Edger et al. 2015). I further analyze the evolutionary

background of these multiple defense strategies with GLSs and test its defensive ability against specialist and generalist herbivores. This approach would give us important insight to understand how GLS potentially have been diversified in a context of multiple defense strategies against herbivores.

In Chapter 2, I focus on *Pieris* butterflies species and test whether their host utilization can be explained by GLS profiles of plants they utilize. Assuming the co-diversification relationship between brassicales plants and pierid butterflies is exist, different host utilization of pierid butterflies can be explained by differential adaptation to other types or species of GLSs among pierid species (Althoff et al. 2014). However, although differential host utilizations even in some closely related species among pierid butterflies were documented (Chew 1980), whether these differences can be explained by GLS profiles of plants has never been tested. I use four *Pieris* butterflies and test this question utilizing comprehensive feeding experiments coupled with defense trait measurement. This would give us a better understanding whether host range shift or host specialization can be correlated with plant secondary metabolite profiles in plant even in closely related specialist herbivores. In addition, this also can produce indirect evidence whether differential chemical adaptation can cause potential speciation in specialist herbivores.

I go more in detail on detoxification system of pierid butterflies in Chapter 3. In this chapter, I test whether *Pieris* butterflies deal with wider range of GLS only by one detoxification enzyme, namely NSP (Wittstock et al. 2004). Until today, at least 140 types of GLSs are identified in brassicales, and higher diversity of GLSs is basically observed in Brassicaceae plant family (Olsen et al. 2016; Hofberger et al. 2013). *Pieris* butterfly, which is one of the main pest on Brassicaceae plants, use relatively wider range (several genera) of Brassicaceae plants as hosts (Ohsaki & Sato 1994). This means, *Pieris* butterflies are exposed to wider chemical range of GLSs as well and need to deal with them to utilize Brassicaceae

as hosts. *Pieris* butterflies acquired NSP as a detoxification enzyme against GLS defense system, however, we still can not explain its substitute specificity or concrete detoxification mechanism (Wittstock et al. 2004; Heidel-Fischer & Vogel 2015). In addition, it still remain unclear whether *Pieris* butterfly use only NSP to disarm GLS defense system, even this respect would be really important to understand the arms-race between GLS and NSP. Here, I let *Pieris* larvae feed on different plant species which have dramatically contrasting GLS profiles and see expression levels of NSP gene family. NSP gene family contains two related genes whose functions have never been confirmed (Fischer et al. 2008). Assuming these sister gene can also work as a detoxification gene, it would be a huge breakthrough to understand the evolution of adaptation traits in pierid butterflies associated with its speciation and host utilization.

In the last chapter, Chapter 4, I focus on NSP microevolution in the wild. The coevolution hypothesis between brassicales plants and pierid butterflies insisted that GLS diversification mediated rapid speciation of both brassicales and pierid butterflies (Hofberger et al. 2013; Edger et al. 2015). NSP, which is known as a counter adaptation to this GLSs defense in pierid butterflies, is also expected to evolve for dealing with this diversifying GLS defenses. Actually, evolutionary dynamics of NSP gene family in pierid butterflies have been observed (Edger et al. 2015). However, we still lack important information whether different host utilization can cause different microevolutional consequences on NSP. Here, in the Chapter 4, I focus on three *Pieris* butterfly species in Japan and its NSP microevolutionary dynamics when they are exposed to different host plant community. This would give us more concrete evidence whether this arms-race between GLS and NSP can be actually occurred in the wild.

In this thesis, overall, I aim to reveal underling mechanism of arms-race between GLS and NSP which potentially drove diversification of both brassicales and pierid

butterflies. Revealing mechanisms underlying this arms-race among Brassicaceae and pierid butterflies would be a big milestone for understanding whether interaction between plants herbivore can be a driver which contributes to shape tremendous biodiversity observed today.

References

Agerbirk, N. & Olsen, C.E., 2012. Glucosinolate structures in evolution. *Phytochemistry*, 77, pp.16–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22405332.

Agrawal, A. a & Fishbein, M., 2006. Plant defense syndromes. *Ecology*, 87(7 Suppl), pp.S132-49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16922309.

Althoff, D.M., Segraves, K. a & Johnson, M.T.J., 2014. Testing for coevolutionary diversification: linking pattern with process. *Trends in ecology & evolution*, 29(2), pp.82–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24314843.

Chew, F.S., 1980. Foodplant Preferences of Pieris Caterpillars (Lepidoptera). *Oecologia*, pp.347–353.

Edger, P.P. et al., 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, p.201503926. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1503926112.

Ehrlich, P. & Raven, P., 1964. Butterflies and plants: a study in coevolution. *Evolution*, 18(4), pp.586–608. Available at: http://www.jstor.org/stable/2406212.

Fahey, J.W., Zalcmann, a T. & Talalay, P., 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), pp.5–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11198818.

Fischer, H.M. et al., 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular biology and evolution*, 25(5), pp.809–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18296701.

Futuyma, D.J. & Agrawal, A.A., 2009. Macroevolution and the biological diversity of plants and herbivores. *October*, 106(43), pp.18054–18061.

Halkier, B.A. & Gershenzon, J., 2006. Biology and biochemistry of glucosinolates. *Annual review of plant biology*, 57, pp.303–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16669764.

Heidel-Fischer, H.M. & Vogel, H., 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, pp.8–14. Available at: http://dx.doi.org/10.1016/j.cois.2015.02.004.

Hofberger, J. a. et al., 2013. Whole Genome and Tandem Duplicate Retention Facilitated Glucosinolate Pathway Diversification in the Mustard Family. *Genome Biology and Evolution*, 5(11), pp.2155–2173. Available at: http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt162.

Kos, M. et al., 2012. Effects of glucosinolates on a generalist and specialist leaf-chewing herbivore and an associated parasitoid. *Phytochemistry*, 77, pp.162–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22281379.

Müller, R. et al., 2010. Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. *Journal of chemical ecology*, 36(8), pp.905–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20617455.

Ohsaki, N. & Sato, Y., 1994. Food plant choice of Pieris butterflies as a trade-off between parasitoid avoidance and quality of plants. *Ecology*, 75(1), pp.59–68.

Olsen, C.E. et al., 2016. Glucosinolate diversity within a phylogenetic framework of the tribe??Cardamineae (Brassicaceae) unraveled with HPLC-MS/MS and??NMR-based

analytical distinction of 70 desulfoglucosinolates. *Phytochemistry*, 132, pp.33–56. Available at: http://dx.doi.org/10.1016/j.phytochem.2016.09.013.

Silva, D.M. & Batalha, M.A., 2010. Defense syndromes against herbivory in a cerrado plant community. *Plant Ecology*, 212(2), pp.181–193. Available at: http://link.springer.com/10.1007/s11258-010-9813-y.

Travers-Martin, N. & Müller, C., 2008. Matching plant defence syndromes with performance and preference of a specialist herbivore. *Functional Ecology*, 22(6), pp.1033–1043. Available at: http://doi.wiley.com/10.1111/j.1365-2435.2008.01487.x.

Wheat, C.W. et al., 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(51), pp.20427–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2154447&tool=pmcentrez&rendertype=abstract.

Winde, I. & Wittstock, U., 2011. Insect herbivore counteradaptations to the plant glucosinolate-myrosinase system. *Phytochemistry*, 72(13), pp.1566–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21316065.

Wittstock, U. et al., 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4859–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=387339&tool=pmcentrez&rendertype=abstract.

Wittstock, U. & Halkier, B. a, 2002. Glucosinolate research in the Arabidopsis era. *Trends in plant science*, 7(6), pp.263–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12049923.

Chapter 1 The effects of different secondary metabolite profiles in plant defense syndromes on specialist and generalist herbivores

Abstract

Plants defend themselves against herbivores not only by a single trait but also by diversified multiple defense strategies. It remains unclear how these multiple defense mechanisms are effectively organized against herbivores. In this study, I focused on Brassicaceae plants that have one of the most diversified secondary metabolites, glucosinolates (GLSs), as a defense against herbivores. By analyzing various defense traits including GLS profiles among 12 species (11 genera) of Brassicaceae plants, it is revealed that their defense strategies can be divided into three categories as multiple defenses. The GLS profiles differed between these three categories: "high nutritional level with long chain aliphatic GLSs", "low nutritional level and high physical defenses with short chain aliphatic GLSs" and "high nutritional level and low defense". The feeding experiment was conducted using two types of herbivore, *Pieris rapae* (Lepidoptera: Pieridae) as a specialist herbivore and the Eri silkmoth *Samia cynthia ricini* (Lepidoptera: Saturniidae) as a generalist, to assess the ability of each plant in multiple defense strategy. It was observed that the Eri silkmoth's performance differed according to which defense strategy it was exposed to. However, the growth rate of *P. rapae* did not vary among the three categories of defense strategy. These results suggest that the diversified defense strategies of Brassicaceae species have evolved to cope with diversified herbivores.

Introduction

Plants have evolved a variety of defense traits to cope with herbivory. Plant species may possess defense traits as part of their overall defensive strategy. According to the plant defense syndrome hypothesis (Agrawal & Fishbein 2006), these multiple defense traits can be grouped into several syndromes. A novel study of plant multiple defenses using 23 milkweed species (*Asclepias* spp., Apocynaceae) revealed that the multiple defense traits of milkweeds can be grouped into three categories; "low nutritional quality", "nutrition and defense", and "tolerance or escape" (Agrawal & Fishbein 2006). This grouping is called the "defense syndrome triangle". However, the generality of plant defense syndrome has been insufficiently tested and the relative effects of these defense categories on a wide range of herbivores, including generalist and specialist species, are still unclear.

Brassicaceae plants contain the most widely studied group of secondary metabolites, the glucosinolates (GLSs), as a chemical defense against herbivores (Hopkins *et al.* 2009). When plant tissue is damaged, the GLSs stored in the vacuoles are exposed to the enzyme myrosinase, which is stored separately in plant cells, and thus referred to as the glucosinolate-myrosinase system (Wittstock & Halkier 2002). This process enables GLSs to be hydrolyzed into several toxic compounds such as isothiocyanates and nitriles (Rask *et al.* 2000). These products negatively affect a wide variety of generalist herbivores (Kos *et al.* 2012). In addition, it is assumed that simple nitriles are less toxic than more complex isothiocyanates (Lambrix *et al.* 2001). However, specialist herbivores of Brassicaceae plants have evolved specific detoxification systems against this glucosinolates-myrosinase system. *Pieris*

(Lepidoptera: Pieridae) larvae not only detoxify the breakdown products of GLS to adapt to Brassicaceae plants, but also use GLS as their own defense to predators and as a stimulant for feeding and oviposition (De Vos *et al.* 2008).

The basic structure of GLS consists of three building blocks: a β-thioglucose moiety, a sulfonated oxime moiety, and a variable side chain (Mithen 2001). Mainly due to the variability of the side chain, at least 120 different GLSs have been identified, mostly in Brassicaceae (Fahey *et al.* 2001). The different effects of indole *vs.* aliphatic GLSs on two specialist and four non-specialist herbivores have been reported in studies on the defensive abilities of GLSs (Müller *et al.* 2010; Abdalsamee & Müller 2012). These findings highlight the effects of GLSs for protection against different kinds of herbivores. However, analyses of the GLSs have mostly been conducted on the model species *Arabidopsis thaliana* and one of the most important crop plants *Brassica oleracea* (Kushad *et al.* 1999; Bidart-Bouzat & Kliebenstein 2008; Poelman *et al.* 2009), with fewer studies focusing on wild Brassicaceae plants. Furthermore, the role of variations in GLSs in these defense strategies has rarely been investigated in the context of plant defense syndromes. Although a defense syndrome triangle was observed in Brassicaceae plants (Travers-Martin & Müller 2008), it is still not clear how these diversified secondary metabolite GLSs are used strategically in the defense syndromes triangle. By revealing the relationships between plant defense strategies and the profile of the GLSs that are involved, the ecological reason for the diversification of GLS will be clarified. This approach will also provide a better understanding of the factors that have driven and maintained the biodiversity of herbivores.

The small cabbage butterfly, *Pieris rapae* (Lepidoptera: Pieridae), is one of the most abundant specialist herbivores of Brassicaceae plants. It is known that they

possess nitrile-specifier protein (NSP) as counter-adaptation to GLSs. NSP redirects GLS hydrolysis towards the formation of nitriles instead of highly toxic isothiocyanates when plant tissues are ingested (Wittstock *et al.* 2004). Only members of the Pieridae possess this protein (Fischer *et al.* 2008). Whereas *P. rapae* shows specific adaptation to Brassicaceae plants, the Eri silkmoth, *Samia cynthia ricini* (Lepidoptera: Saturniidae), is polyphagous herbivore with their host plants including the castor oil plants, ailanthus, cassava, kesseru, and plumeria (Konno *et al.* 2004). In addition, this moth eventually eats the leaves of any plant unless the leaves are too hard or hairy. As a result, this moth can be killed by poisoning or show symptoms of growth inhibition in response to the respective food plant species. Thus, the larvae of this moth have been frequently used to perform bioassays to evaluate plant defense levels against herbivorous insects (Fukui *et al.* 2002; Konno *et al.* 2004).

In this study, I focused on the GLS profiles in plant multiple defense strategies and the effect of this synergetic defense on insect herbivore performance. Twelve Brassicaceae species in 11 genera, which are the potential hosts of *P. rapae*, were investigated. I used *P. rapae* and Eri silkmoth to analyze the differential effects of the defense strategy of Brassicaceae plants on specialist and non-specialist herbivorous insects.

Materials and Methods

Plants

The seeds of 12 Brassicaceae plant species were originally collected from wild populations at Chiba, Nagano, and Hokkaido in Japan (Table 1-1). These are all annual plants and most of them bloom in every spring or summer. For plant cultivation, the seeds were sown in pots (40 mm in diameter) with vermiculite and grown in a laboratory at 24°C, with 60% relative humidity and a photoperiod of L16:D8. Plants were watered and fertilized weekly with a 2000× diluted solution of Hyponex (N:P:K = 6:10:5, Hyponex, Japan).

After cultivation for three months, the following defense traits were measured in leaves: (1) 20 different GLS concentrations, (2) leaf toughness, (3) trichome density, (4) water content, (5) specific leaf area (SLA) and (6) the C:N ratio. GLS concentrations in one individual plant for each species were measured using a widely targeted metabolome analysis (Sawada *et al.* 2009). While the indolic GLSs were derived from tryptophane, aliphatic GLSs were derived from methionine and were detected as Methylsulfinyl GLS and Methylthio GLS in this study. I treated the length of the aliphatic GLS side chain C6-8 as a long type and C3-5 as a short type (Beekwilder *et al.* 2008). For the other five defense traits, five individuals were used for the measurements. Leaf toughness was measured for two leaves of each individual plant with a force gauge penetrometer (1 mm in radius) that measured the mass (g) needed to penetrate a leaf surface (Feeny & Jul 2007). I also took two leaf discs (7.8 mm^2) from each individual plant and calculated the overall density of trichome (n/mm^2) by averaging the density present on both sides of each disc. Water content was assessed by first weighing leaf discs wet and again after drying in an oven (60°C). Specific leaf area (SLA) was calculated as the area of the leaf disc divided by the dry mass. Total leaf carbon (C) and nitrogen (N) concentrations were determined from a mixed sample of five replicates

from each species by a CN coder (NC-220F, Sumika, Tokyo, Japan) using 20 mg of dried ground leaf material.

Insect herbivores

P. rapae

Female adults were originally collected in the field at Chiba, Japan. Three mother butterflies were collected in April 2014 and allowed to lay eggs on cabbage in a chamber. The newly hatched larvae (five individuals) were introduced onto each intact plant individual of each species, and this experiment was repeated twice with two plants. The feeding experiments were performed in a 25m^2 greenhouse at 24°C, 60% relative humidity, and a photoperiod of L16:D8. After five days of feeding, the weights of individual larvae were measured to the nearest 0.01 mg. Because the initial larval mass was very small (< 0.0001 g), the larval performances were determined using the weight (g) of the individuals.

Eri silkmoth

I obtained eggs from the National Institute of Agrobiological Sciences, Tsukuba. Two egg batches were prepared. Five newly hatched larvae from each batch (in total 10 individuals) were introduced onto each intact plant individual of each species, and this experiment was repeated twice in the same way as in the *P. rapae* feeding experiment. Two days after introduction, the larval weights were measured.

Statistical analysis

To determine the tradeoff relationships between each pair of plant defense traits, I performed a Pearson's pairwise correlation analysis for the measured defensive traits, with the exception of data for the GLS concentration (Agrawal & Fishbein 2006). To establish whether the plant species showed defense syndromes, principal component analysis (PCA) and cluster analysis were performed (Agrawal & Fishbein 2006) using R packages (prcomp, kmeans, package stats). The best clustering number was evaluated by the pseudo F-statistic (Calinski & Harabasz 1974; Silva & Batalha 2010). After dividing the plant species into defense syndrome clusters, the values of each defense trait were compared by using a generalized linear model (GLM) with a normal distribution and identity link, based on Akaike's Information Criterion (AIC) to determine whether the defense trait differed among the defense clusters. In this model, each plant trait value was set as the response variable with the clusters being treated as explanatory variables. The model with the lowest AIC value was selected as the preferred grouping of the defense trait. The variation of the herbivore growth rate between plant species was also analyzed with GLM based on AIC. The best model, in which the variation of growth rates among plant species was explained most successfully, was selected. All statistical analyses were performed using R software version 3.0.2 (R Core Team 2014).

Results

Plants

A total of 20 different GLSs were detected from 12 plant species as follows: 2-propenyl (Sinigrin), 3-(Methylsulfinyl)propyl (3MSOP), 4-(Methylsulfinyl)butyl (4MSOB), 5-(Methylsulfinyl)pentyl (5MSOP), 6-(Methylsulfinyl)hexyl (6MSOH), 7-(Methylsulfinyl)heptyl (7MSOH), 8-(Methylsulfinyl)octyl (8MSOO), 3-(Methylthio)propyl (3MTP), 4-(Methylthio)butyl (4MTB), 5-(Methylthio)pentyl (5MTP), 6-(Methylthio)hexyl (6MTH), 7-(Methylthio)heptyl (7MTH), 8-(Methylthio)octyl (8MTO), 3-(Hydroxy)propyl (3OHP), 4-(Hydroxy)butyl (4OHB), 3-(Benzoyloxy)propyl (3BZOP), 4-(Benzoyloxy)butyl (4BZOB), Indol-3-ylmethyl (I3M), 1-Methoxyindol-3-ylmethyl (IMOI3M), 4-Methoxyindol-3-ylmethyl (4MOI3M)(Table 1-S1).

Among the five leaf traits other than GLSs, three significant correlations were identified by the pairwise correlation analyses. Toughness *vs.* SLA, and SLA *vs.* C:N ratio were negatively correlated, and toughness *vs.* C:N ratio were positively correlated (Table 1-2).

The results of the PCA allowed us to reduce the number of axes to five (cumulative proportion of variance = 79.4%, PC1 = 33.4%, PC2 = 19.1%, PC3 = 10.7%, PC4 = 8.9% and PC5 = 7.3%), and the cluster analysis by k-means method and the clustering evaluation supported the plant species to be divided into three clusters (pseudo F-values for the given number of clusters, $F_2 = 14.18$, $F_3 = 16.28$, $F_4 = 15.37$, Fig. 1-1). In addition, in the analysis of the values of defense traits among these three clusters by GLMs, 20 out of 25 (20 GLSs and the five other traits) defense traits differed among clusters (Table 1-S2). In these three clusters, plant species were represented by combinations of the concentrations of GLSs and leaf characters. Cluster 1 is characterized by a high concentration of GLSs, containing a high level of long-

chained aliphatic GLSs (C6-8) and a low C:N ratio. Cluster 2 has high concentrations of short-chained aliphatic GLSs (C3-5) with intermediate concentrations of GLSs and a higher level of toughness and C:N ratio with relatively high trichome density (ΔAIC = +1.31), and Cluster 3 is characterized by high SLA, low physical and chemical defenses and a low C:N ratio (Fig. 1-1, Table 1-S2). Although Cluster 3 was characterized by the presence of sinigrin (Fig. 1-1), the amount of sinigrin was not significantly higher than that present in the species of other clusters (Table 1-S2).

Herbivore performance

GLM showed a difference in the growth rates of *P. rapae* larvae among Brassicaceae plants (AIC_{plant}, -935.8 *vs.* AIC_{null}, -117.2). The larvae fed on *Cardamine scutata* and *Thlaspi arvense* developed better than those fed on any of the other plants. In contrast, they did not grow by feeding on *Erysimum cheiranthoides* and *Capsella bursa-pastoris* (Fig. 1-2). No relationships between the *P. rapae* growth rates and the clusters of plant defense traits were detected (AIC_{null}, -653.33 *vs.* $AIC_{cluster,}$ -649.72).

The growth rates of the Eri silkmoth also differed among plant species (AIC_{plant}, -941.5 *vs.* AIC_{null}, -142.2). Eri silkmoth larvae grew well on *Cardamine scutata* and *Capsella bursa-pastoris*. The best model identified by GLM indicated that the larvae which fed on Cluster 3 plants (high SLA, low physical and chemical defenses) grew better than those that fed on plants in the other two clusters (AIC_{best}, -1292.8 *vs.* AIC_{null}, -1280.6). The second best model pointed out that Cluster 3 was best and Cluster 1 was worst for larvae growth (AIC_{second}, -1291.7).

Discussion

The results revealed three distinct clusters of plant defense strategies among the species investigated (Fig.1-1), a conclusion similar to Agrawal and Fishbein (2006) who developed the theory of "defense syndromes". Agrawal and Fishbein (2006) reported that plant defense could be classified into three converged strategies, known as the "defense syndrome triangle": (A) low nutritional quality, (B) high nutritional quality with defense, (C) tolerance or escape. My results showed that the plant species in Cluster 1 had a higher GLS concentration, especially long-chained aliphatic GLSs (C6-8; Table 1-S1, S2) that are more effective against specific herbivores (Dicke 2012; Beekwilder *et al.* 2008). For the species in Cluster 1, a higher nutrition level (low C:N ratio) was also observed, suggesting the adoption of "high nutritional quality with defense" strategy (see Agrawal & Fishbein 2006). In contrast, a higher leaf toughness and lower nutrition concentration (high C:N ratio) were observed in the species of Cluster 2. A relatively high trichome density was also observed in this cluster. Following the categories of Agrawal and Fishbein (2006), the species in this cluster can be classified as "low nutritional quality". Unlike the species of these two clusters, no strong defenses were detected in the species of Cluster 3. The species in this cluster had a higher nutrition concentration and higher SLA, both of which are often regarded as an indicator of rapid growth and high palatability to herbivores (Agrawal & Fishbein 2006). The possession of both characters is the typical strategy of "tolerance or escape" species (Kursar & Coley 2003). In the theory of the "defense syndromes triangle", plants with low nutritional level do not need to invest in toxin production because of their low palatability for herbivorous insects. However, plants with high nutritional level do have

to defend themselves against herbivores by the production of high levels of toxin or "tolerance or escape" ability.

In contrast to this concordance, some differences were observed between my results and the defense syndrome in Agrawal and Fishbein (2006). In the Cluster 2, the co-occurrence of the physical defenses and short chain GLSs were observed. No such specific chemical compounds concentration with high physical defenses was detected in the milkweeds study by Agrawal and Fishbein (2006). This co-occurrence might be the characteristic of Brassicaceae plant species in contrast with milkweed. Moles *et al.* (2013) suggested that "defense syndromes" could not be observed among multiple plant families as a whole. My results and the other previous research (Travers-Martin & Müller 2008) suggest that each plant group shows a specific defense strategy and that this will hide the defense syndromes when I focus on plants as a whole.

The specialist herbivore *P. rapae* did not display any clear trends in growth rates against the defense strategies among Brassicaceae plant species (Fig. 1-2). This might indicate that the specialist herbivores overcome the defenses of Brassicaceae plants, including the GLS based defense. However, it was notable that the larvae of *P. rapae* that were transferred onto *E. cheiranthoides* and *C. bursa-pastoris,* the members of "tolerance or escape" cluster, all died without feeding these plant at all, even though the Eri silkmoth did well on these plant species. It is known that some specialist herbivores use specific chemical compounds of host plants to recognize the plants as their host (Del Campo *et al.* 2001). The result may be due to by the lack of key chemical compounds that would be used as host recognition by *Pieris* larvae.

The results for the generalist herbivore, the Eri silkmoth, which does not adapt to GLS, corresponded well to the defense syndrome among Brassicaceae plant species.

The larvae of Eri silkmoth grew well on species in Cluster 3, where the strategy of "tolerance or escape" was adopted. However, the growth rates of the Eri silkmoth feeding on plant species in Cluster 1 and Cluster 2 were low, indicating that the Eri silkmoth could not grow by feeding on plants containing high GLSs and with physical defenses. It is reported that generalists tend to be more sensitive to GLSs than specialists by comparing their performance on *A. thaliana* with different GLS profiles (Kliebenstein *et al.* 2002). These results basically indicate that the multiple defense strategies of Brassicaceae plants have differential effects on different herbivores. In addition, it was also observed that even among the plant species in Cluster 3, the performance of Eri silkmoth was poor on *D. nemorosa*. This result might have been caused by toxic materials that were not detected in the present study. In order to indicate the presence of such undetected toxins, it is therefore necessary to observe and compare the herbivore performances as well as levels of plant toxicity when evaluating plant multiple defense.

Although many previous studies have revealed the negative effects of GLSs on herbivore performance (Bidart-Bouzat & Kliebenstein 2008; Müller *et al.* 2010), the effects of the diversity of GLSs in the context of a multiple defense strategy in plants had not been previously identified. In this study, I revealed a measurable relationship between the distribution of various GLSs and defense strategies among plant species (Fig. 1-1). The aliphatic GLSs are reported to have an effective defensive ability against herbivores, e.g. *Mamestra brassicae* (Lepidoptera: Noctuidae) on *A. thaliana* (Beekwilder *et al.* 2008). Higher concentrations of GLSs with long chained aliphatic GLSs (C6-8) were observed in the plant species of Cluster 1. Short chained aliphatic GLSs (C3-5) with high physical defense traits were observed in the species of Cluster 2

(Table 1-S2). The reason for this co-occurrence could be explained by presumed trade-offs in which the plants that have high physical defenses may have impose higher costs on chain elongation of GLSs. The result of the Eri silkmoth experiment showed that the growth rates of larvae decreased when feeding on the species of Cluster 1 and 2, both of which have high defenses, compared to those on the species of Cluster 3 with very low concentrations of GLSs. Thus, both of these defense strategies, high GLSs concentrations with long chained aliphatic GLSs and high physical defenses with short chained aliphatic GLSs, worked efficiently against generalist herbivore. The exact defensive functions of the chain length of aliphatic GLSs are still unknown, although the genes that have a role in inducing chain elongation of aliphatic GLSs have been identified (Kliebenstein *et al.* 2001; Kroymann *et al.* 2001; Windsor *et al.* 2005). Previous research has also estimated the costs of GLS synthesis on Brassicaceae plant species (Bekaert *et al.* 2012). This would explain the "trade-offs" among the defense strategies by long-chained GLS, short-chained GLS, and physical traits observed in the present study. Integration of genetic, phylogenetic and physiological information on Brassicaceae plant including *A. thaliana* would reveal the mechanisms underlying the deviation of GLS profiles in the defense strategies of the Brassicaceae plants.

In this study, I analyzed the defensive secondary metabolite profiles in multiple defense strategies, focusing on both specialist and generalist herbivores, Brassicaceae plants, and GLSs. I successfully observed the defense syndromes in this set of plants and deviations in the unique GLS profiles among them. In addition, I found that the generalist and specialist herbivore showed different responses against the defense strategies. These results suggest the variety of GLSs have been needed for Brassicaceae plants to evolve plural stable defense strategies in order to resist against

several types of herbivores. However, a consideration of the phylogenetic relationship between Brassicaceae plants and *Pieris* butterflies and the distributions of GLSs on the Brassicaceae phylogeny are essential to confirm the GLS diversification mechanism.

References

Abdalsamee MK, Müller C (2012) Effects of indole glucosinolates on performance and sequestration by the sawfly *Athalia rosae* and consequences of feeding on the plant defense system. *Journal of Chemical Ecology* **38,** 1366–1375.

Agrawal AA, Fishbein M (2006) Plant defense syndromes. *Ecology* **87,** 132–149.

Beekwilder J, van Leeuwen W, van Dam NM, *et al.* (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS ONE* **3,** e2068.

Bekaert M, Edger PP, Hudson CM, Pires JC, Conant GC (2012) Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. *The New Phytologist* **196,** 596–605.

Bidart-Bouzat MG, Kliebenstein DJ (2008) Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in *Arabidopsis thaliana*. *Journal of Chemical Ecology* **34,** 1026–1037.

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Communications in Statistics* **3,** 1–27

Del Campo ML, Miles CI, Schroeder FC, *et al.* (2001) Host recognition by the tobacco hornworm is mediated by a host plant compound. *Nature* **411,** 186–189.

De Vos M, Kriksunov KL, Jander G (2008) Indole-3-acetonitrile production from indole glucosinolates deters oviposition by *Pieris rapae*. *Plant Physiology* **146,** 916–926.

Dicke M, Gols R, Poelman ER (2012) Dynamics of plant secondary metabolites and consequences for food chains and community dynamics. In Iason GR, Dicke M,

Hartley SH (eds) *The Ecology of Plant Secondary Metabolites From Genes to Global Processes*, pp. 308-328. Cambridge University Press, Cambridge.

Fahey JW, Zalcmann T, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* ***56,*** 5–51.

Feeny P, Jul N (2007) Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. *Ecology* **51,** 565–581.

Fischer HM, Wheat CW, Heckel DG, Vogel H (2008) Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular Biology & Evolution* **25,** 809–820.

Fukui A, Murakami M, Konno K, Nakamura M, Ohgushi T (2002) A leaf-rolling caterpillar improves leaf quality. *Entomological Science* **5,** 263–266.

Hopkins RJ, van Dam NM, van Loon JJ (2009) Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual Review of Entomology* **54,** 57–83.

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *The Plant Cell* **13,** 681–693.

Kliebenstein D, Pedersen D, Barker B, Mitchell-olds T (2002) Comparative analysis of quantitative trait loci controlling glucosinolates, myrosinase and insect resistance in *Arabidopsis thaliana*. *Genetics* **332,** 325–332.

Konno K, Hirayama C, Nakamura M, *et al.* (2004) Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *The Plant Journal* **37,** 370–378.

Kos M, Houshyani B, Wietsma R, *et al.* (2012) Effects of glucosinolates on a generalist and specialist leaf-chewing herbivore and an associated parasitoid. *Phytochemistry* **77,** 162–170.

Kroymann J, Textor S, Tokuhisa JG, *et al.* (2001) A gene controlling variation in *Arabidopsis* glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiology* **127,** 1077–1088.

Kursar T, Coley P. (2003) Convergence in defense syndromes of young leaves in tropical rainforests. *Biochemical Systematics & Ecology* **31,** 929–949.

Kushad MM, Brown AF, Kurilich AC, *et al.* (1999) Variation of glucosinolates in vegetable crops of *Brassica oleracea*. *Journal of Agricultural & Food Chemistry* **47,** 1541–1548.

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The *Arabidopsis* epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences *Trichoplusia ni* herbivory. *The Plant Cell* **13,** 2793–2807.

Mithen R (2001) Glucosinolates – biochemistry, genetics and biological activity. *Plant Growth Regulation* **34,** 91–103.

Moles AT, Peco B, Wallis IR, *et al.* (2013) Correlations between physical and chemical defences in plants: tradeoffs, syndromes, or just many different ways to skin a herbivorous cat? *The New Phytologist* **198,** 252–263.

Müller R, de Vos M, Sun JY, *et al.* (2010) Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. *Journal of Chemical Ecology* **36,** 905–913.

Poelman EH, Van Dam NM, Van Loon JJA, Vet LEM, Dicke M (2009) Chemical diversity in *Brassica oleracea* affects biodiversity of insect herbivores. *Ecology* **90,** 1863–1877.

Rask L, Andréasson E, Ekbom B, Eriksson S, Pontoppidan B, Meijer J (2000) Myrosinase: gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology* **42,** 93–113.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Sawada Y, Akiyama K, Sakata A, *et al.* (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant & Cell Physiology* **50,** 37–47.

Silva DM, Batalha MA (2010) Defense syndromes against herbivory in a cerrado plant community. *Plant Ecology* **212,** 181–193.

Travers-Martin N, Müller C (2008) Matching plant defence syndromes with performance and preference of a specialist herbivore. *Functional Ecology* **22**, 1033–1043.

Windsor AJ, Reichelt M, Figuth A, *et al.* (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry* **66,** 1321–1333.

Wittstock U, Agerbirk N, Stauber EJ, *et al.* (2004) Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 4859–4864.

Wittstock U, Halkier B (2002) Glucosinolate research in the *Arabidopsis* era. *Trends in Plant Science* **7,** 263–270.

Figures and Tables

Table 1-1 The plant species used in this experiment. The collection sites and the location for the seeds of each plant species were shown.

Plant species	Seed collection site
Arabidopsis lyrata	Ueda, Nagano, Japan (36° 31' N, 138° 20' E)
Arabis hirsuta	Ueda, Nagano, Japan (36° 31' N, 138° 20' E)
Brassica napus	Ichihara, Chiba, Japan (35° 30' N, 140° 50' E)
Brassica tournefortii	Kashiwa, Chiba, Japan (35° 54' N, 139° 56' E)
Capsella bursa-pastoris	Ichihara, Chiba, Japan (35° 30' N, 140° 50' E)
Cardamine scutata	Ichihara, Chiba, Japan (35° 30' N, 140° 50' E)
Draba nemorosa	Hotaka, Nagano, Japan (36° 20' N, 137° 50' E)
Erysimum cheiranthoides	Chitose, Hokkaido, Japan (42° 50' N, 141° 41' E)
Lepidum virginicum	Ichihara, Chiba, Japan (35° 30' N, 140° 50' E)
Raphanus sativus	Kamogawa, Chiba, Japan (35° 07' N, 140° 11' E)
Thlaspi arvense	Yubari, Hokkaido, Japan (42° 56' N, 142° 01' E)
Turritis glabra	Ueda, Nagano, Japan (36° 31' N, 138° 20' E)

Table 1-2 The results of pairwise correlations between defense traits.

Pearson's pairwise correlation analyses were performed.

defense trait	Trichomes	Water	SLA	C:N
Toughness	0.384	-0.007	-0.598 *	0.649 *
Trichomes		0.375	0.353	-0.255
Water			0.535	-0.240
SLA				-0.638 *

* $P < 0.05$

Figure and Tables

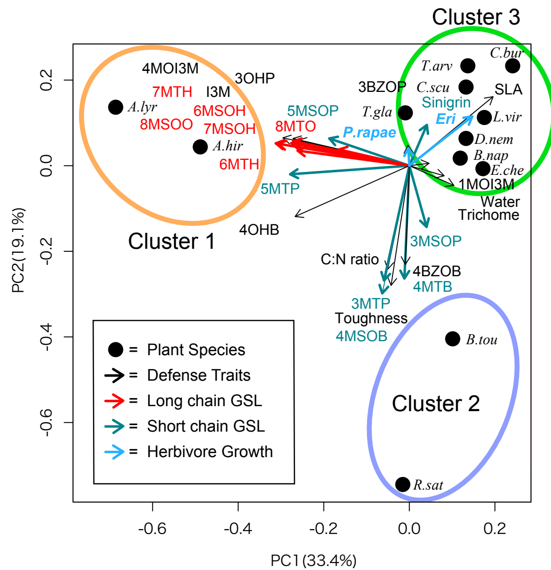
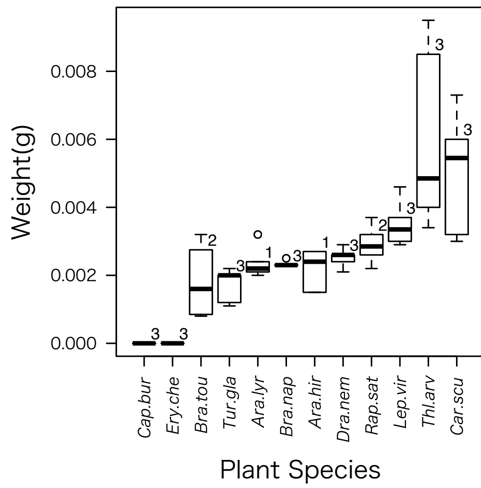


Fig. 1-1 The biplot of the principal component analysis (PCA) of plant defense traits. The black points indicate each plant species. The red and blue arrows indicate the concentrations of long chain aliphatic GLSs and short chain aliphatic GLSs. Black arrows show the strength of other defense traits, and sky blue allows show the results of the herbivore feeding experiment. The percentage of total variance explained is given by PC1 and PC2. PC1 shows long chain concentration and PC2 shows the physical characteristic of plants (Toughness, SLA). The three circles indicate: Cluster 1 (orange) high nutritional level containing long chain aliphatic GLSs, Cluster 2 (blue) low nutritional level and high physical defenses with short chain aliphatic GLSs, and Cluster 3 (green) a higher SLA and nutrition concentration with low defenses.

(a) *P.rapae* growth



(b) Eri silkmoth growth

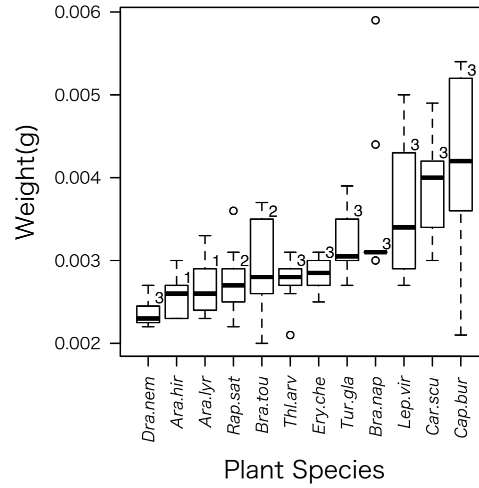


Fig. 1-2 Results of the feeding experiment using 12 species of Brassicaceae plants; (a) *P. rapae* growth and (b) Eri silkmoth growth. All *P. rapae* larvae that fed on *E. cheiranthoides* and *C. bursa-pastoris* died. The different letters beside each box indicate defense clusters (1: high nutritional level containing long chain aliphatic GLSs, 2: low nutritional level and high physical defenses with short chain aliphatic GLSs, 3: higher SLA and nutrition concentration with low defense)

Supplementary information

Table 1-S1 The glucosinolate contents of Brassicaceae plant species.

Plant / Glucosinolate abbreviation	*Ara.lyr*	*Ara.hir*	*Bra.nap*	*Bra.tou*	*Cap.bur*	*Car.scu*	*Dra.nem*	*Ery.che*	*Lep.vir*	*Rap.sat*	*Thl.arv*	*Tur.gla*
Sinigrin	0.1111 ± 0.0073	0.0547 ± 0.0009	0.0002 ± 0.0002	0.0001 ± 0.0000	0.1926 ± 0.0101	0.0068 ± 0.0004	n.d.	n.d.	n.d.	0.0089 ± 0.0015	2.0470 ± 0.0396	0.0002 ± 0.0002
3MSOP	0.1146 ± 0.0015	0.0298 ± 0.0032	0.0002 ± 0.0000	4.8296 ± 0.0059	0.0033 ± 0.0005	n.d.	n.d.	0.5797 ± 0.0070	n.d.	0.0138 ± 0.0026	0.0011 ± 0.0004	n.d.
4MSOB	0.0060 ± 0.0009	0.0043 ± 0.0002	n.d.	0.0115 ± 0.0008	0.0012 ± 0.0004	n.d.	0.0001 ± 0.0000	0.0006 ± 0.0001	n.d.	0.0305 ± 0.0038	n.d.	0.0013 ± 0.0005
5MSOP	0.0753 ± 0.0005	0.0667 ± 0.0052	n.d.	0.0002 ± 0.0000	0.0006 ± 0.0000	n.d.	0.0012 ± 0.0000	0.0001 ± 0.0000	n.d.	0.0031 ± 0.0026	n.d.	0.1458 ± 0.0067
6MSOH	3.0955 ± 0.1427	0.4634 ± 0.0189	n.d.	0.0006 ± 0.0002	n.d.	0.0008 ± 0.0003	0.0031 ± 0.0004	0.0014 ± 0.0003	n.d.	n.d.	0.0003 ± 0.0002	0.3524 ± 0.0159
7MSOH	6.1511 ± 0.2868	1.1906 ± 0.0470	0.0005 ± 0.0002	n.d.	0.0010 ± 0.0008	0.0002 ± 0.0001	0.0140 ± 0.0021	0.0002 ± 0.0001	n.d.	n.d.	n.d.	0.0166 ± 0.0000
8MSOO	0.1727 ± 0.0102	0.0281 ± 0.0035	0.0002 ± 0.0001	n.d.	0.0047 ± 0.0008	n.d.	0.0403 ± 0.0030	n.d.	n.d.	0.0003 ± 0.0000	n.d.	0.0001 ± 0.0000
3MTP	0.0214 ± 0.0019	0.0066 ± 0.0007	0.0001 ± 0.0000	0.0628 ± 0.0046	n.d.	n.d.	n.d.	0.0002 ± 0.0001	n.d.	0.0453 ± 0.0058	0.0003 ± 0.0000	n.d.
4MTB	0.0024 ± 0.0013	0.0029 ± 0.0004	n.d.	0.0003 ± 0.0001	0.0001 ± 0.0001	0.0004 ± 0.0000	n.d.	0.0007 ± 0.0003	n.d.	2.1692 ± 0.0500	n.d.	0.0003 ± 0.0000
5MTP	0.0061 ± 0.0010	0.0084 ± 0.0010	0.0002 ± 0.0002	n.d.	0.0001 ± 0.0000	n.d.	n.d.	0.0002 ± 0.0001	n.d.	0.0034 ± 0.0006	n.d.	0.0058 ± 0.0007
6MTH	0.1284 ± 0.0086	0.1506 ± 0.0118	0.0001 ± 0.0001	0.0008 ± 0.0004	0.0001 ± 0.0000	n.d.	n.d.	n.d.	n.d.	0.0003 ± 0.0003	n.d.	0.0203 ± 0.0006
7MTH	0.1638 ± 0.0121	0.4021 ± 0.0161	0.0003 ± 0.0002	n.d.	0.0001 ± 0.0000	n.d.	0.0002 ± 0.0001	n.d.	n.d.	n.d.	n.d.	0.0008 ± 0.0003
8MTO	0.0012 ± 0.0004	0.0058 ± 0.0008	0.0001 ± 0.0000	n.d.	n.d.	n.d.	0.0042 ± 0.0020	n.d.	n.d.	n.d.	n.d.	n.d.
3OHP	3.8557 ± 0.0799	2.6949 ± 0.1139	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0004 ± 0.0002	n.d.	n.d.
4OHB	0.0023 ± 0.0007	0.0010 ± 0.0005	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0017 ± 0.0005	n.d.	n.d.
3BZOP	n.d.	n.d.	0.0001 ± 0.0000	n.d.	n.d.	n.d.	n.d.	0.0001 ± 0.0001	0.0001 ± 0.0000	n.d.	n.d.	0.0013 ± 0.0007
4BZOB	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0003 ± 0.0002	n.d.	n.d.	0.0005 ± 0.0000	n.d.	n.d.
I3M	0.1133 ± 0.0072	0.1186 ± 0.0061	0.0015 ± 0.0006	n.d.	n.d.	0.0788 ± 0.0052	n.d.	0.0007 ± 0.0003	n.d.	0.0098 ± 0.0020	0.0019 ± 0.0011	0.0010 ± 0.0003
1MOI3M	n.d.	n.d.	0.0168 ± 0.0019	n.d.	n.d.	0.0001 ± 0.0000	n.d.	0.0005 ± 0.0000	n.d.	n.d.	n.d.	n.d.
4MOI3M	0.2229 ± 0.0096	0.2119 ± 0.0081	0.0272 ± 0.0009	0.0010 ± 0.0002	n.d.	0.0008 ± 0.0002	0.0158 ± 0.0013	0.0125 ± 0.0014	0.0411 ± 0.0013	0.0194 ± 0.0010	0.0781 ± 0.0024	0.0010 ± 0.0016

Table 1-S2 The relationship between 25 defensive traits and the defensive clustering of Brassicaceae species. The significance of the differences among traits was tested by using GLM based AICs. Twenty defense traits were significantly different as grouped below. Different letters beside each value indicate differences in defense traits between clusters based on the grouping with the model selection using GLM.

CLUSTER	Toughness	Trichome	Water	SLA	C:N ratio	Sinigrin	3MSOP	4MSOB	5MSOP
CLS1	0.187 ± 0.037 **a**	124.362 ± 124.362 **a**	0.765 ± 0.026 **a**	176.678 ± 21.851 **a**	23.193 ± 2.605 **a**	0.083 ± 0.028 **a**	0.072 ± 0.042 **a**	0.005 ± 0.001 **a**	0.071 ± 0.004 **a**
CLS2	0.278 ± 0.010 **b**	243.782 ± 21.843 **a**	0.786 ± 0.013 **a**	178.849 ± 9.072 **a**	36.438 ± 1.084 **b**	0.004 ± 0.004 **a**	2.422 ± 2.408 **b**	0.021 ± 0.009 **b**	0.002 ± 0.001 **b**
CLS3	0.155 ± 0.015 **a**	265.955 ± 94.432 **a**	0.771 ± 0.015 **a**	322.640 ± 32.468 **b**	18.092 ± 3.092 **a**	0.281 ± 0.253 **a**	0.073 ± 0.072 **a**	0.000 ± 0.000 **c**	0.018 ± 0.018 **b**
CLUSTER	6MSOH	7MSOH	8MSOO	3MTP	4MTB	5MTP	6MTH	7MTH	8MTO
CLS1	1.779 ± 1.316 **a**	3.671 ± 2.480 **a**	0.100 ± 0.072 **a**	0.014 ± 0.007 **a**	0.003 ± 0.000 **a**	0.007 ± 0.001 **a**	0.139 ± 0.011 **a**	0.283 ± 0.119 **a**	0.004 ± 0.002 **a**
CLS2	0.000 ± 0.000 **b**	0.000 ± 0.000 **b**	0.000 ± 0.000 **b**	0.054 ± 0.009 **b**	1.085 ± 1.084 **b**	0.002 ± 0.002 **b**	0.001 ± 0.000 **b**	0.000 ± 0.000 **b**	0.000 ± 0.000 **b**
CLS3	0.045 ± 0.044 **b**	0.004 ± 0.002 **b**	0.006 ± 0.005 **b**	0.000 ± 0.000 **c**	0.000 ± 0.000 **a**	0.001 ± 0.001 **b**	0.003 ± 0.003 **b**	0.000 ± 0.000 **b**	0.001 ± 0.001 **b**
CLUSTER	3OHP	4OHB	3BZOP	4BZOB	I3M	1MOI3M	4MOI3M		
CLS1	3.275 ± 0.580 **a**	0.002 ± 0.001 **a**	2.1E-05 ± 0.000 **a**	2.4E-05 ± 0.000 **a**	0.116 ± 0.003 **a**	0.000 ± 0.000 **a**	0.217 ± 0.005 **a**		
CLS2	0.000 ± 0.000 **b**	0.001 ± 0.001 **b**	0 ± 0.000 **a**	0.00026 ± 0.000 **b**	0.005 ± 0.005 **b**	0.000 ± 0.000 **a**	0.010 ± 0.009 **b**		
CLS3	0.000 ± 0.000 **b**	0.000 ± 0.000 **c**	0.00019 ± 0.000 **a**	3.3E-05 ± 0.000 **a**	0.011 ± 0.010 **b**	0.002 ± 0.002 **a**	0.022 ± 0.009 **b**		

Chapter 2 Differential larval performance of *Pieris* butterfly species correlates to Glucosinolate diversity and side chain length

Abstract

The tremendous diversity of plants and herbivores has been explained by the plant and herbivore coevolutionary theory. Recent studies reveal that pierid butterflies and Brassicales plants have co-diversified through an arms-race mediated by glucosinolates, the main chemical defense of Brassicales. Although different host preferences of pierid butterflies responding to glucosinolates profiles of their hosts are expected based on this hypothesis, empirical evidence have not been obtained. Here, I tested whether the larval performance of *Pieris* butterfly species correlate to plant defense traits especially glucosinolates profiles. I chose four *Pieris* species all of which utilize Brassicaceae plants as hosts and share the same general adaptation mechanisms against glucosinolate-based chemical defenses. Comprehensive feeding experiments utilizing 25 Brassicaceae plants revealed that larval performance patterns of these four species could be classified into two groups. The different larval performance could be explained by plant glucosinolate profiles; one of these groups grew better on plants with lower short-chain aliphatic glucosinolates, higher long-chain aliphatic glucosinolate concentration and higher glucosinolate diversity. These results suggest that individual *Pieris* species have likely evolved to feed on a subset of Brassicaceae plants and are not always capable of adapting to the complete range of glucosinolate defenses. Furthermore, I found that the larval performance of *Pieris* species correlated to their phylogeny on a finer taxonomical scale. My results are still tentative but offer a correlation-type support for

the idea that pierid butterflies and Brassicales have coevolved based on glucosinolate defense system diversification.

Introduction

Ehrlich and Raven (1964) introduced the 'plant–herbivore coevolutionary theory' to explain the remarkable diversity of plants and herbivores. Plants utilize a variety of chemical defenses against herbivores and herbivores evolve adaptive traits to better utilize these host plants. The profiles of chemical defenses of plants are distinguishable depending on plant species or higher plant taxa (Futuyma & Agrawal, 2009). Moreover, these defenses are thought to affect host selection and host ranges of herbivores, because herbivores can detoxify only a restricted range of plant chemical defenses (Janz, 2010). The 'plant–herbivore coevolutionary theory' suggests that plants evolved novel chemicals to deter or poison phytophagous insects, and this arms-race contributed to the diversification of both plants and herbivores.

Plants are known to produce an incredible variety of secondary metabolites as specific chemical defenses against herbivores (Fraenkel, 1959; Ehrlich & Raven, 1964; Futuyma & Agrawal, 2009). To date, a number of chemically mediated interactions between specific plants and herbivore groups have been revealed (Futuyma & Agrawal, 2009). A prominent example of these interactions is between Brassicales plants and pierid butterflies (Edger et al., 2015). As a major defense against herbivores, Brassicales plants produce secondary metabolites called glucosinolates (GLSs) (Hopkins et al., 2009). Upon plant tissue damage, GLSs are hydrolysed by plant myrosinase enzymes and, depending on the GLS type, a number of breakdown products such as epithionitriles, nitriles, or toxic isothiocyanates are produced (Wittstock & Halkier, 2002). The basic structure of GLSs consists of three building parts: a β-thioglucose moiety, a sulphonated oxime moiety, and a variable side chain (Mithen, 2001). More

than 140 GLSs are known and the variations in GLSs are due mainly to side chain differences (Fahey et al., 2001; Olsen et al., 2016). In general, GLSs are grouped into aliphatic-, benzyl-, and indole GLSs, because these are biosynthetically derived from different amino acids (Wittstock & Halkier, 2002). In addition, side chain elongation or modification occurs in the biosynthesis process, leading to the high diversity of GLSs observed in Brassicales. As a result of these chemically diverse side chains, breakdown products are also variable with a variety of functions, and its differential effect against herbivores has been intensively tested in previous studies (Beekwilder et al., 2008; Müller et al., 2010; Winde & Wittstock, 2011).

Specialized Pierinae butterflies overcome this GLS defense system by larval gut-expressed nitrile specifier proteins (NSPs) which redirect the GLS hydrolysis reaction to non-toxic nitriles rather than toxic isothiocyanates (Wittstock et al., 2004). It has been suggested that the increased diversification rate of Pieridae butterfly lineage was a consequence of the acquisition of NSP at the base of Pieridae evolution (Wheat et al., 2007). This event suggests that the evolution of NSP is a key innovation for the diversification of Pieridae butterflies (Heidel-Fischer et al., 2010). According to a recent phylogenetic work, the speciation rate of Pieridae butterflies rose simultaneously with the GLS diversification events in Brassicales, which implies the coevolution process mediated by GLS diversification (Edger et al., 2015).

Host plant shifts can represent a first step towards reproductive isolation and ultimately speciation of phytophagous insects, known as ecological speciation (Ohshima, 2010; Matsubayashi et al., 2010). This speciation process should be detectable in macroevolutionary patterns between Brassicales and pierid butterflies. Reduced performance of herbivore species on novel host plants of coevolved taxa could

be one evidence for coevolutionary diversification (Althoff et al., 2014). If Pieridae and Brassicales co-diversification is being mediated by GLS diversification as suggested, it is expected that each pierid butterfly show different host plant preference and larval performance corresponding to specific plant GLS profiles. Although specialist Pierinae butterflies utilize the same general mechanisms, namely NSPs, to overcome the GLS-based defenses of their host plants, they display different host preferences (Chew, 1980). Despite the host preferences of Pierinae butterfly were examined utilizing several host plant species, the degree to which their host preferences as well as larval performance correlate with GLS profiles has not been well tested (Renwick & Lopez, 1999).

In the present study, I aim to identify plant defense traits which explain differential larval performances among Brassicales feeding *Pieris* butterflies. Specifically, I focus on GLS profiles and the other types of defense traits of each host plant species. For addressing this question, I utilize four closely related Pierinae butterfly species (*Pieris napi*, *P. melete, P. rapae* and *P. brassicae*; Lepidoptera, Pieridae) all of which use Brassicaceae plants as host plants. Three of the four *Pieris* butterfly species are native to Japan while *P. brassicae* is native to Europe but was recently introduced and can only be found in northern Japan. 25 Brassicaceae plant species that covered a broader taxonomical range in Brassicaceae were utilized to test whether larval performance correlates with any specific plant chemical or physical defense traits. I conduct comprehensive feeding experiments, measured defense traits, and analyze correlations between host use and plant defense traits in order to test potential factors which determine differential larval performances among *Pieris* butterflies. Furthermore, I additionally reconstruct the Pierinae butterfly species

phylogeny to test whether larval performance of the four *Pieris* butterflies correlates with their phylogenetic relationships.

Materials and Methods

Pieris butterflies and Brassicaceae plants

The current research included the feeding experiments utilizing four *Pieris* butterfly species and 25 Brassicaceae plants. *P. napi* (green-veined white butterfly) can be found from Europe to Asia (Porter & Geiger, 1995; Chew & Watt, 2006). This species is frequently observed in mountain sites and it is documented that they mainly use the genus *Arabis* as a host in Japan (Ohsaki & Sato, 1994). *P. melete* is widely distributed in Japan and Asia, and mostly utilizes plants of the genus *Cardamine* (Ohsaki & Sato, 1994). *P. rapae* is one of the most common Pieridae butterflies in Northern and Central Europe to Asia and often observed on cultivated Brassicaceae plants (Schoonhoven et al., 2005; Braby & Trueman, 2006). Both *P. rapae* and *brassicae* are common pests of cabbage plants (Hasan & Ansari, 2011).

I collected female butterflies of the four *Pieris* butterfly species from wild populations in Chiba and Hokkaido, Japan (Table 2-S1). I prepared 7–10 female butterflies for each species and reared them by feeding with a 2% glucose solution. Wild caught female butterflies were mostly fertilized already. I released the female butterflies to allow egg-laying in cages containing cabbage (*Brassica oleracea* var. *capitata*) or *Cardamine scutata* plants with high intensity light conditions. For *P. brassicae*, final instar larvae were caught in the wild. They were fed on cabbage and reared to adults. After eclosion, 10 mother butterflies were prepared by hand pairing.

For all of the four *Pieris* butterfly species, the eggs were incubated at 25°C until they hatched.

For experimental plants, I collected seeds of 25 Brassicaceae plant species. For the sake of analysing broader range of physical and chemical defense traits, I covered all three main clades known in Brassicaceae (Lineage1, 2 and 3, Table 2-S2, Beilstein et al., 2008; Couvreur et al., 2010; Franzke et al., 2011). Seeds of 21 of these species were originally collected from wild populations in Japan, Inner Mongolia, and Canada, and the others were purchased. I grew the plants in the greenhouse at 25°C, with 60% relative humidity and L16:D8. Plants were watered and fertilized every week with a 2000× diluted solution of Hyponex (N:P:K = 6:10:5; Hyponex, Japan). After two months of cultivation, plants were used for the feeding experiments or defense measurement experiments.

Feeding experiment

Neonate larvae were collected within 12 hours after they hatched and introduced to the plants for the feeding experiment. I prepared two plant individuals from each species for each larval species, and applied three neonate larvae for each plant by a soft-haired brush. To minimize changes in the condition of the experimental plants, all experimental trials were carried out within 5 days for all the four *Pieris* species. I conducted feeding experiments under the same conditions for the plants to be grown. I measured the weight of each larva individually (within 0.1 mg) after 120 hours of feeding, and used the average weights of larval individuals on each plant species as the index of the performance of each *Pieris* butterfly species. I set the weights of dead larvae as 0, because neonates of *Pieris* butterflies were too small to be measured. I

conducted these experiments using 25 Brassicaceae plant species in August 2014 and repeated the same experimental setup using 17 plant species in June 2015 (Table 2-S2). To investigate the reproducibility of these two feeding trials, the average larval weights of each butterfly species in these two feeding trials were compared by Pearson correlation. Because significant correlation between first and second feeding trials was confirmed for each species ($P < 0.05$, $r > 0.65$), the individual weights of the second trial (17 plant species) were standardized to the first trial based on a linear regression model, and pooled with the data of the first trial. From these pooled data, average weights of larvae on each plant were calculated for each butterfly species again and used as representative values after standardized as z-score for further analyses.

Defense traits measurement

To examine the effect of plant defense traits on herbivore performances, I measured six leaf traits: concentrations of each of 21 individual GLSs (if detectable), leaf toughness, trichome density, water content, specific leaf area (SLA), and C:N ratio. For measuring the 21 types of GLSs, I conducted a widely targeted metabolome analysis by tandem quadrupole mass spectrometry (TQMS) coupled with ultra performance liquid chromatography (UPLC) (Sawada et al., 2009). In my measuring system, I used selected reaction monitoring (SRM) by UPLC-TQS. This analytical method enabled us to detect a broader range of GLSs with higher sensitivity. I sampled three undamaged young rosette leaves from three different plant individuals. Sampled leaves were immediately flash frozen in liquid nitrogen and freeze dried. The dried leaves were grounded and a unit of the leaf powder from the three individuals was mixed as a representative of one species. For the chemical analyses, I used 2 mg of mixed sample

for each species and analyzed one time for each species. Detected peaks of each GLS were evaluated with negative controls which were measured in triplicates including the internal standard only. I extracted peaks that showed > 30 signal/noise ratios as detected peaks and served for further analyses. The relative concentrations of each GLS among samples were calculated by comparing the peak areas with the internal standard (10-camphorsulfonic acid).

To examine the effect of GLS profiles of each plant on larval performance, I classified the detected GLSs and calculated the total concentration of each classified GLS category. Detected GLSs were classified depending on their major chemical class: aliphatic-, benzyl-, and indole GLS. Because aliphatic GLSs are known to contain short-chained and long-chained types depending on side chain length (Beekwilder et al., 2008), I treated the length of the GLS side chain C_2-C_5 as the short-chain aliphatic GLS and that of C_6-C_8 as the long-chain aliphatic GLS. Concentrations of each GLSs were log transformed and converted into *z*-scores. Concentrations of each GLS group were calculated by totalling and standardizing the concentrations of GLSs in each group. In addition, I measured GLS richness and GLS diversity of each plant. GLS richness refers to the numbers of detected types of GLS of each plant. GLS diversity was measured as Shannon index based on the relative concentration of each GLS compounds in each species (Becerra et al., 2009).

The other five defense traits were measured by the same methods as in Okamura et al. (2016) using five plant replicates. I measured leaf toughness by force gauge penetrometer (1 mm in radius) (Feeny & Jul, 2007), and I used a CN coder (NC-220F, Sumika, Tokyo, Japan) for 20 mg of dried ground leaf material to measure C:N ratios. For the other defense traits, I prepared 10 leaf discs (7.8 mm^2) for each plant

species from the five replicates (2 discs from each individual). Trichome density was measured by counting the trichomes on both sides of each disc under a stereo microscope (×10 - 40) with counters and the average trichome density for each plant species was calculated (trichome count / mm^2). To measure the trichome density of *Aurinia saxatilis* and *Descurainia sophia*, which have dense trichomes, I used scanning electron microscopy (JSM-7500F, JOEL, Japan) and counted the numbers of trichomes in scope visual field (×300). I also measured specific leaf area (SLA) and water content by comparing wet and dry weight of the 10 discs for each species. I log transformed these defense traits values when needed to obtain normality for statistical analyses.

Analysis of potential defense traits that affect host usage differences among the four Pieris butterfly species

In this study, I analysed the larval performance data in several scales, such that (1) species scale, (2) interspecies scale, (3) inter-performance group scale (if performance groups are detected). In species scale, I simply tested whether larval performances of each species correlated to specific plant defense traits of each plant species. Since I utilize closely related butterfly species, similar patterns of their performances are expected. Therefore, I additionally focused on interspecific performance difference (interspecies scale). I regressed the results of the feeding experiments for each pair of *Pieris* butterfly species. Here, I utilized Deming regression because it accounts for errors in observations on both the x and y axes (Cornbleet & Gochman, 1979; Linnet, 1998). Then the residuals (observed values minus estimated values from regression) of each pair of butterfly species were treated as the contrast of larval performance between these two species. Note that this performance contrast could take both positive and

negative values. Positive values show that one butterfly species outperforms the other species on the host plant species, and negative values show the inverse. After calculating these different values, I tested the correlations between these contrast and plant defense values to examine whether there are any specific plant traits that could explain the interspecies contrast in larval performance. I represented the comparative sets of butterfly species as follows: NM, *P. napi* vs. *P. melete*; NR, *P. napi* vs. *P. rapae*; NB, *P. napi* vs. *P. brassicae*; MR, *P. melete* vs. *P. rapae*; MB, *P. melete* vs. *P. brassicae*; and RB, *P. rapae* vs. *P. brassicae*.

Furthermore, I assumed the larval performances of several species pair could be similar than any other pair of species. Therefore, I conducted clustering analysis to see whether any performance groups in which species show similar larval performances. If I could detect such performance groups, I checked whether any leaf trait can explain the performance differences between these groups. I calculated the differences of larval performance for each species pair and conducted hierarchical cluster analyses. Clustering numbers were evaluated using the pseudo F statistic (Calinski & Harabasz, 1974). When any performance group was detected, I calculated the inter-group performance contrast that crossed the performance group by pooling the values of the species pairs. This calculated inter-group performance contrast was compared with each defense traits.

For the correlation tests, the number of possible test could be greater than the degrees of freedom and likely to cause type I errors. Therefore, I performed the false discovery rate (FDR) adjustment, which enabled us to evaluate the significance of each detected correlation in a good balance with type I and type II errors (Benjamini and Hochberg, 1995, Nakagawa, 2004).

Pierinae phylogeny

To test whether larval performances correlate with a *Pieris* butterfly species phylogeny, I reconstructed the phylogeny of 11 Pierinae species from wild populations in Japan (Table 2-S1), because sequences of some species I used in this study were missing in current published Pierinae tree (Chew & Watt, 2006; Wahlberg, et al. 2014). I used three genes for the phylogenetic analyses: the mitochondrial cytochrome c oxidase subunit I (CO1), mitochondrial NADH dehydrogenase subunit 5 (ND5), and elongation factor-1 alpha (EF-1α) (Yagi et al., 1999; Braby et al., 2006). I used the CO1 sequences registered on the BOLD system (Ratnasingham & Hebert, 2007; Ratnasingham & Hebert, 2013). For the other two regions, DNA was extracted using Chelex 100 (Bio-Rad) according to the manufacturer's protocol, and I conducted a polymerase chain reaction (PCR) using two primer sets (Yagi et al., 1999; Monteiro & Pierce, 2001; Kandul et al., 2004; Table 2-S3). I purified the PCR products using magnetic beads (AMPure XP Kit) and performed cycle sequencing with BigDye Terminator Cycle Sequencing kit (Applied Biosystems). I removed the dye terminators via ethanol precipitation and conducted purified sequencing reactions on an ABI 3500 automated sequencer (Applied Biosystems). Forward and reverse reads were obtained for all samples, and sequences were edited and aligned with MEGA (version 6.0; Tamura et al., 2013). I selected adequate substitution models using jModelTest ver. 2.1.6 (Guindon & Gascuel, 2003; Darriba et al., 2012); I utilized GTR+I+G, HKY+I+G, and SYM+G substitution models for the CO1, ND5, and EF-1α regions, respectively. I performed Bayesian analyses using MrBayes version 3.2.1 (Huelsenbeck & Ronquist, 2001; Ronquist et al., 2012) with 500,000 Markov chain Monte Carlo generations and

sampling of one tree every 100 generations. Then, 25% burn-in trees were discarded, and the remaining trees were used to estimate Bayesian topology and branch posterior probabilities. In addition, I reconstructed maximum-likelihood (ML) trees using the GTR+I+G models selected by JModelTest and calculated bootstrap probabilities. To make a confidential phylogenetic tree, I additionally used *Gonepteryx maxima* (GenBank accession nos. KF491764.1, GU372649.1, AB855818.1), *Colias erate* (for the ND5 sequence only, GenBank accession no. AB855816.1), *Papilio memnon* (GenBank accession nos. AB969797.1, AY457623.1, AB084426.1), and *Papilio xuthus* (GenBank accession nos. EU105358.1, GU372634.1, AB013149.1) sequences from GenBank and rooted the tree using the *Papilio* clade.

To see how the four *Pieris* butterfly species utilized in the present study distribute in *Pieris* clades, I also reconstructed *Pieris* phylogeny from CO1 region by combining my sequence data and the sequences of other *Pieris* butterfly species stored in GenBank. I used the CO1 sequences of 11 additional *Pieris* butterfly species (Table 2-S1) and reconstructed a ML tree using MEGA. The obtained phylogenies were used to compare host usage differences of *Pieris* butterflies.

Results

Larval performance of the four Pieris butterfly species

My results showed that the four *Pieris* butterfly species displayed similar but different host plant preferences based on larval performance levels. Their growth varied depending on the 25 Brassicaceae plant species (Table 2-S4, Fig. 2-S1). As a general trend, all of the four *Pieris* species showed a better performance on the plant genus

Cardamine (*C. hirsuta, C. regeliana,* and *C. scutata*) compared to the other plant species (Table 2-S4, Fig. 2-S1). In contrast, I also observed that the majority of larvae tested here did not perform well on some plant species, e.g. *Erysimum cheiranthoides* or *Berteroa incana* (Fig. 2-S1).

In addition to this general performance trend for all the four species, the cluster analysis supported that the performance of the four *Pieris* species could be clustered into two groups (pseudo *F* values for the given number of clusters, $F_2 = 1.37$, $F_3 = 1.31$): *P. napi*-*P. melete* (*napi*/*melete*) group and *P. rapae*-*P. brassicae* (*rapae*/*brassicae*) group (Fig. 2-1a). This simply showed species in the same group showed similar larval performances. For example, *P. napi* and *melete* grew relatively better on *Arabis hirsuta* or *Arabidopsis kamchatica* than *P. rapae* and *brassicae* (Fig. 2-S1).

Defense traits correlated with larval performance differences among Pieris butterflies

I measured physical and chemical defense traits of 25 Brassicaceae plants coupled with a widely targeted metabolome analysis. A total of 21 GLSs were detected from the Brassicaceae plants analysed (13 aliphatic-, 5 benzyl-, and 3 indole GLS, Table 2-S5, S6). I compared defense traits (GLS profiles and measured physical defense traits) with larval performance in different scales (species, interspecies and inter-performancegroup scale). In species level, correlation analyses revealed that trichome density was negatively correlated with the performance of all the four species of *Pieris* larvae (Table 2-1, Fig. 2-S2). Some significant correlations were also detected between certain type of GLSs concentrations and larval performances. Short-chain aliphatic GLS concentration showed defense abilities against *P. napi* and *melete* (Table 2-1), and the aliphatic GLS

concentration showed defensive abilities against *P. melete* and *rapae* growth (Table 2-1). In contrast to these negative correlations, benzyl GLS concentration showed a positive correlation with larval performance of *P. melete* even this was not significant after the FDR adjustment (Table 2-1).

To test whether larval performance differences correlated with plant defense traits, I also examined the interspecific larval performance contrasts. I observed that MB (*P. melete vs. P. brassicae*) was negatively correlated with short chain aliphatic GLS concentration (Table 2-S7). This indicated *P. melete* performed worse than *P. brassicae* on the plants with higher short chain aliphatic GLS concentration. Although it was not significant after the FDR adjustment, I could find weak effects of GLS profiles but not those of physical defense traits at the interspecific level (Table 2-S7).

I also focused on the host utilization differences between two detected performance groups (*napi*/*melete* group vs. *rapae*/*brassicae* group) (Fig. 2-1a). The larval performance differences between members of the same performance groups showed no significant correlation with GLS profiles or any other plant trait (Table 2-S7; NM, RB, however NM showed marginal correlation with short chain aliphatic GLS). Regarding the detected performance groups (*napi*/*melete* group and *rapae*/*brassicae* group), NM-RB was calculated by pooling the values of NR, NB, MR, and MB, all of these were interspecific larval performance contrast between species from different performance group. Positive NM-RB values indicated that the *napi*/*melete* group showed greater growth than the *rapae*/*brassicae* group along with the variable, and negative values of NM-RB indicated the *rapae*/*brassicae* did better than the *napi*/*melete* group. I compared this contrast with each defense trait and observed significant correlations with short chain aliphatic GLSs and long-chain aliphatic GLSs and GLS

diversity (Table 2-2, Fig. 2-2a,b,c). The significant negative correlation with short chain aliphatic GLS concentration (Fig. 2-2a) showed that *P. rapae* and *brassicae* grew better than *P. napi* and *melete* under higher concentrations of short-chain aliphatic GLSs, and *P. napi* and *melete* did better on plants with low short-chain aliphatic GLS. The opposite situation was observed in case of long chain aliphatic GLS (Fig. 2-2b). The positive correlation of NM-RB with GLS diversity supported that the *napi*/*melete* group has more resistance against plants with varied GLSs than the *rapae*/*brassicae* group (Fig. 2-2c). This was supported by the marginal correlation found in GLS richness as well (Table 2-2). I also focus on relative concentrations of each of C_2-C_8 aliphatic GLS respectively as an ad-hoc analysis. However, because of the effect of non-detected GLS could be greater when I focused on finer scale, I could not detect any significant correlation at this clustering (data not shown). In any other defense trait such as trichome density or leaf toughness, no significant correlation was detected (Table 2-2, Fig. 2-2d).

Correlation with larval performance and their phylogeny

I reconstructed the phylogeny including Japanese *Pieris* butterfly species to test whether larval performance correlates with their phylogeny. I determined the CO1, EF-1α, and ND5 sequences of 11 Pierinae species (except the ND5 sequence of *Colias erate*) and submitted them to GenBank under accession numbers LC090556–LC090590 (Table 2-S1). The phylogenetic tree of Pierinae butterflies, obtained by Bayesian and ML analyses, showed that *P. napi* and *P. melete* form a sister group and are phylogenetically closely related (Fig. 2-1b). This indicated that the performance of the four *Pieris* butterfly species somehow correlates with their phylogeny. The phylogenetic tree

constructed using the CO1 region also supported this clade and justified that the species I used in this research covered a broader taxonomical range of the genus *Pieris* (Fig. 2-S3). However, because of the limited number of species that I could use for feeding experiments, it is still difficult to conclude that host preferences were correlated with their phylogeny.

Discussion

In this study, I performed feeding experiments of *Pieris* butterflies combining with defense traits measurements of Brassicaceae plants especially focusing on GLSs and physical defenses. The results of feeding experiments showed concordance with previous researches which compared host utilizations or performances of *Pieris* butterfly species among several plant species (Chew, 1980; Ohsaki & Sato, 1994). A number of these reports showed that *Cardamine* are relatively suitable host plants for most *Pieris* butterfly species as I observed (Table 2-S4, Fig. 2-S1). At the species scale, I found that trichome density strongly reduced the growth rates of all the four *Pieris* species (Table 2-1, Fig. 2-S2). In previous work, higher trichome density of *A. thaliana* has been shown to negatively affect the oviposition preference of the specialist herbivore, *Plutella xylostella* (Lepidoptera: Plutellidae), but not its larval performance (Handley et al., 2005). In contrast to this report, my results suggest that a high trichome density is an efficient defense against specialist herbivore larvae of *Pieris* butterflies. This disparity may be partly due to differences in the taxonomic scale of the host plant species examined. The present experiment focused on a family scale, whereas Handley et al. (2005) examined intraspecific variation within *A. thaliana*. Smaller variance of

trichome density within species might cause a smaller effect on herbivore performance. In the present study, responses to trichome density did not differ among *Pieris* butterfly species (Table 2-2, S7, Fig. 2-2d), and trichomes comprise an efficient form of defense in general (Fig. 2-S2). These results indicate that trichomes could not be the key trait corresponding to the interspecific difference in larval performances observed among *Pieris* butterfly species.

In contrast to trichomes, I found that some GLS concentrations showed correlation with larval performances of a certain specific species (Table 2-1). Short-chain aliphatic GLSs showed correlation with *P. napi* and *melete*, but not against the other species (Table 2-1). This suggests that short-chain aliphatic GLSs can be a candidate defense trait underlying the differences in larval performance among *Pieris* butterfly species. Larval performances of *P. melete* and *rapae* were also negatively correlated with aliphatic GLS concentration. Interestingly, these species specific trends could only be found in GLS profiles but not in plant physical traits (Table 2-1). This suggests that sensitivities of the four *Pieris* butterflies against a range of GLSs might be different, but they are equally affected by physical defenses of plants. In *Pieris* butterflies, GLSs are detoxified by the NSP. Potentially different affinity of NSP for individual GLSs and their breakdown products would be the expected mechanism of the observed inter-species differences.

The results of my feeding experiments and subsequent clustering analysis showed that the larval performance patterns of the four *Pieris* butterfly species could be clustered into two groups (*napi*/*melete* group and *rapae*/*brassicae* group; Fig. 2-1a). Although different host preferences among a few plant species were shown in *Pieris* butterfly relatives (Chew, 1980; Ohsaki & Sato, 1994), a striking difference in host use

patterns in *Pieris* butterfly species were observed by exemplifying the larval performance in my broader plant sets. This pattern of larval performance level should confirm the observation of host preference pattern among the four *Pieris* butterflies, in which *P. rapae* and *brassicae* prefer Brassicaceae crops than the other two species (Benson et al., 2003; Harvey et al., 2010). Furthermore, the contrast in larval performance between the two performance groups was only explained by GLS profiles but not by plant physical defenses. I observed that *P. napi* and *melete* group showed better growth than *P. rapae* and *brassicae* when reared on plants with lower short chain aliphatic GLS and higher long chain aliphatic GLS concentration (Table 2-2). These trends were also supported by the results of interspecific performance comparisons, which shows *P. melete* performed better than *P. brassicae* on the plants with lower short chain aliphatic GLS concentration, whereas physical defenses had no effect on the interspecific performance contrasts (Table 2-S7). In addition, I also found that both GLS diversity and richness can explain their performance difference, such that *P. napi* and *melete* group shows more resistance than the other group against plants with higher GLS diversity. These significant correlations suggest that GLS profiles, especially concentrations of short vs. long-chain aliphatic GLS or GLS diversity of plants, might be a plausible candidate that shapes differences in larval performance between these two performance groups. Regarding the physical defense, on the other hand, I could not detect any correlation with larval performances. These results propose an important insight that the differences in host utilization of *Pieris* butterfly species might rely on the differences in GLS profile but not on leaf physical traits.

My results also suggest that short- and long-chain aliphatic GLSs might have different defensive functions against herbivores (Table 2-2). Although the defensive

functions of GLSs have been repeatedly tested against herbivores and pathogens (Mithen, 1992; Beekwilder et al., 2008; Prasad et al., 2012; Abdalsamee & Müller, 2012), only few studies have detected variations in the function of aliphatic GLSs in conjunction with side chain length (Huang & Renwick, 1994; Raybould & Moyes, 2001). Here, I observed that aliphatic GLS with different chain length showed different correlation with larval performances of *Pieris* butterfly species (Table 2-2, Fig. 2-2a, b). Although which kinds of GLS breakdown product are produced in each plant is still unrevealed, it would be interesting if substrate specificity of NSP is strongly affected by side chain length of GLSs. Moreover, I also observed a significant correlation between GLS diversity and larval performance (Table 2-2, Fig. 2-2c). The defensive ability of each single type of GLSs have been documented (Beekwilder et al., 2008; Rohr et al., 2009; Müller et al., 2010). This result additionally leads to the idea that the possession of multiple types of GLSs could be beneficial for plant defense against diversified herbivores. Although GLS diversity was measured from a limited number of detected GLSs in my analysis, these results suggest that there might be a synergistic effect of the variety of chemical defense compounds against herbivores. This could be the focal topic to understand the evolution of diverse chemical defenses in the future.

As a caveat, I have realized that I still have some limitations in the present GLS analyses for concluding defensive effect of each detected GLS. The obtained GLS profiles were consistence with previous observations. For example, all of the detected GLSs in *Arabidopsis thaliana* in my measurement were covered by previous observations on this plant species (Fahey et al., 2001). The characteristic GLS profile of *Rorippa indica* was also captured with detecting long chain aliphatic GLS (Fahey et al., 2001). However, there were a few GLS that were not detected in previous work in some

plant species. In *Brassica napus*, I detected a small peak of Benzyl GLS, which has not been detected in well-established previous GLS profile analysis (Velasco et al., 2008). Similarly, in *Nasturtium officinale*, I detected 4-Hydroxybutyl GLS, which have never been detected in this species (Lockwood & Belkhiri, 1991, Fahey et al., 2001). I used sensitive metabolome analysis (SRM by UPLC -TQS) for acquiring a broader range of GLS profile of Brassicaceae species, in order to test its potential differential defensive ability against the four *Pieris* butterflies. Sensitivity of SRM is quite high even a single seed metabolomics in *A. thaliana* can be performed (Sawada et al., 2017), and useful for acquiring broader GLS profiles of plants. However, since replicate number was not enough in the present study (3 mixed sample measured one time), I are unable to fully confirming each detected GLSs, even so the data was useful for acquiring broader GLS profiles of each plants. In addition, I realized that there are still a number of GLSs that could not be identified in my analytical methods. Currently, more than 140 types of GLSs have been identified, and many additionally detected GLSs are waiting to be elucidated (Olsen et al., 2016). Actually, the inclusions of several types of GLS which could not be identified in my metabolome analyses were reported. For example, *Arabis hirsuta* is known to dominantly include at least 4 types of chain-elongated aromatic GLSs (Agerbirk et al., 2010). Not only un-target GLSs, but also different types of GLS breakdown products can affect the functional outcome as well (Burow et al., 2006). GLSs can be hydrolysed into different breakdown forms depending on the reaction conditions or types of specific enzymes they encountered (Winde & Wittstock, 2011). Therefore, I could not reject potential effects of the non-detected GLSs as well as the various breakdown products on larval performances in the present research.

In addition, other non-GLS defense chemicals would also affect the larval performances. At least *Erysimum cheiranthoides*, which was used in the present study, is known to contain strong non-GLS chemical defense, cardenolide, against herbivores (Sachdev-Gupta et al., 1993). I observed that most of all the four *Pieris* species larvae died when they fed on *E. cheiranthoides* in my feeding assay (Fig. 2-S1, Table 2-S4). In this case, all the four *Pieris* butterfly species were equally affected by the cardenolides. However, it is possible that several non-GLS chemical defense compounds will show different effects on specific *Pieris* species. Therefore, the test of the effects of chemical defense compounds including non-GLS on the *Pieris* performances should be needed as a further research.

In the present research, I found that differential larval performances of the four *Pieris* butterfly species can be explained by plants GLS profiles but not by any other physical defense traits. Although I examined only a subsection of the GLS members, this correlation indicated that the differences in interspecific detoxification mechanism against different types of GLS can be one candidate to explain their host utilization rather than resistances against physical defenses. Furthermore, my phylogenetic analysis partly insisted a phylogenetic correlation of larvae performances in *Pieris* butterfly species. These results supported the previous research which appealed the coevolution between Brassicales plants and pierid butterfly mediated by GLS diversification (Wheat et al., 2007; Heidel-Fischer et al., 2010; Edger et al., 2015). If the molecular detoxification mechanisms and substrate specificity of NSP and its interspecific differences can be revealed, this would propose more direct evidence. This would enable us to understand "Brassicales-pierid coevolution" by examining their co-diversification mechanisms mediated by the GLS and NSP arms-race.

References

Abdalsamee MK & Müller C (2012) Effects of indole glucosinolates on performance and sequestration by the sawfly *Athalia rosae* and consequences of feeding on the plant defense system. Journal of Chemical Ecology 38: 1366–1375. doi: 10.1007/s10886-012-0197-4.

Agerbirk N, Olsen CE, Poulsen E, Jacobsen N & Hansen PR (2010) Complex metabolism of aromatic glucosinolates in *Pieris rapae* caterpillars involving nitrile formation, hydroxylation, demethylation, sulfation, and host plant dependent carboxylic acid formation. Insect Biochemistry and Molecular Biology 40: 126–137. doi:10.1016/j.ibmb.2010.01.003.

Althoff DM, Segraves KA & Johnson MTJ (2014) Testing for coevolutionary diversification: linking pattern with process. Trends in Ecology and Evolution 29: 82–89. doi: 10.1016/j.tree.2013.11.003.

Becerra JX, Noge K & Venable DL (2009) Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. Proceedings of the National Academy of Sciences of the USA 106: 18062–18066. doi: 10.1073/pnas.0904456106.

Beekwilder J, van Leeuwen W, van Dam NM, et al. (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. PLoS One 3:e2068. doi: 10.1371/journal.pone.0002068.

Beilstein MA, Al-Shehbaz IA, Mathews S & Kellogg EA (2008) Brassicaceae phylogeny inferred from phytochrome A and *ndhF* sequence data: tribes and trichomes revisited. American Journal of Botany 95:1307–1327. doi: 10.3732/ajb.0800065.

Benjamini Y& Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289–300.

Benson J, Pasquale A, Driesche RV& Elkinton J (2003) Assessment of risk posed by introduced braconid wasps to Pieris virginiensis, a native woodland butterfly in New England. Biological Control 26 : 83–93.

Braby MF & Trueman JWH (2006) Evolution of larval host plant associations and adaptive radiation in pierid butterflies. Journal of Evolutionary Biology 19: 1677–1690. doi: 10.1111/j.1420-9101.2006.01109.x.

Braby MF, Vila R & Pierce NE (2006) Molecular phylogeny and systematics of the Pieridae (Lepidoptera : Papilionoidea): higher classification and biogeography. Zoological Journal of the Linnean Society 147: 239–275.

Burow M, Müller R, Gershenzon J & Wittstock U (2006) Altered glucosinolate hydrolysis in genetically engineered *Arabidopsis thaliana* and its influence on the larval development of *Spodoptera littoralis*. Journal of Chemical Ecology 32: 2333–2349. doi:10.1007/s10886-006-9149-1.

Calinski T & Harabasz J (1974) A dendrite method for cluster analysis. Communications in Statistics 3: 1–27.

Chew FS (1980) Foodplant preferences of *Pieris* caterpillars (Lepidoptera). Oecologia 46: 347–353.

Chew FS & Watt WB (2006) The green-veined white (*Pieris napi* L.), its Pierine relatives, and the systematics dilemmas of divergent character sets (Lepidoptera, Pieridae). Biological Journal of the Linnean Society 88: 413–435. doi: 10.1111/j.1095-8312.2006.00630.x.

Cornbleet PJ & Gochman N (1979) Incorrect least-squares regression coefficients in method-comparison analysis. Clinical Chemistry 25: 432–438.

Couvreur TLP, Franzke A, Al-Shehbaz IA, et al. (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Molecular Biology and Evolution 27: 55–71. doi: 10.1093/molbev/msp202.

Darriba D, Taboada G, Doallo R & Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772. doi: 10.1038/nmeth.2109.

Edger PP, Heidel-Fischer HM, Bekaert M, et al. (2015) The butterfly plant arms-race escalated by gene and genome duplications. Proceedings of the National Academy of Sciences of the USA 112: 8362–8366. doi: 10.1073/pnas.1503926112.

Ehrlich P & Raven P (1964) Butterflies and plants: a study in coevolution. Evolution 18: 586–608.

Fahey JW, Zalcmann AT & Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56: 5–51.

Feeny P & Jul N (2007) Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. Ecology 51: 565–581.

Fraenkel GS (1959) The raison d'être of secondary plant substances. Science 29: 1466–1470.

Franzke A, Lysak MA, Al-Shehbaz IA, et al. (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. Trends in Plant Science 16: 108–116. doi: 10.1016/j.tplants.2010.11.005.

Futuyma DJ & Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences of the USA 106: 18054–18061.

Guindon S & Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology 52: 696–704.

Handley R, Ekbom B & Ågren J (2005) Variation in trichome density and resistance against a specialist insect herbivore in natural populations of *Arabidopsis thaliana*. Ecological Entomology 30: 284–292. doi: 10.1111/j.0307-6946.2005.00699.x.

Harvey JA, Poelman EH & Gols R (2010) Development and host utilization in Hyposoter ebeninus (Hymenoptera: Ichneumonidae), a solitary endoparasitoid of Pieris rapae and P. brassicae caterpillars (Lepidoptera: Pieridae). Biological Control 53 : 312–318.

Hasan F & Ansari MS (2011) Effects of different brassicaceous host plants on the fitness of *Pieris brassicae* (L.). Crop Protection 30: 854–862. doi: 10.1016/j.cropro.2011.02.024.

Heidel-Fischer HM, Vogel H, Heckel DG & Wheat CW (2010) Microevolutionary dynamics of a macroevolutionary key innovation in a Lepidopteran herbivore. BMC Evolutionary Biology 10: 60. doi: 10.1186/1471-2148-10-60.

Hopkins RJ, van Dam NM & van Loon JJA (2009) Role of glucosinolates in insect-plant relationships and multitrophic interactions. Annual Review of Entomology 54: 57–83. doi: 10.1146/annurev.ento.54.110807.090623.

Huang X & Renwick JAA (1994) Relative activities of glucosinolates as oviposition stimulants for *Pieris rapae* and *P.napi oleracea*. Journal of Chemical Ecology 20: 1025–1037. doi: 10.1007/BF02059739.

Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755. doi: 10.1093/bioinformatics/17.8.754.

Janz N (2010) Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. Annual Review of Ecology Evolution and Systematics 42: 71-89. doi: 10.1146/annurev-ecolsys-102710–145024.

Kandul N, Lukhtanov V, Dantchenko A, et al. (2004) Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of *COI* and *COII* and nuclear sequences of *EF1*-alpha: karyotype diversification and species radiation. Systematic Biology 53: 278–298. doi: 10.1080/10635150490423692.

Linnet K (1998) Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clinical Chemistry 44: 1024–1031.

Lockwood GB & Belkhiri A (1991) Glucosinolate spectrum of some Algerian Cruciferae. Plant Systematics and Evolution 176: 11–20. doi:10.1007/BF00937942.

Matsubayashi KW, Ohshima I & Nosil P (2010) Ecological speciation in phytophagous insects. Entomologia Experimentalis et Applicata 134: 1–27. doi: 10.1111/j.1570-7458.2009.00916.x.

Mithen R (1992) Leaf glucosinolate profiles and their relationship to pest and disease resistance in oilseed rape. Euphytica 63: 71–83. doi: 10.1007/BF00023913.

Mithen R (2001) Glucosinolates – biochemistry, genetics and biological activity. Plant Growth Regulation 34: 91–103.

Monteiro A & Pierce NE (2001) Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from *COI*, *COII*, and *EF-1α* gene sequences. Molecular Phylogenetics and Evolution 18: 264–281. doi: 10.1006/mpev.2000.0872.

Müller R, de Vos M, Sun JY, et al. (2010) Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. Journal of Chemical Ecology 36: 905–913. doi: 10.1007/s10886-010-9825-z.

Nakagawa S (2004) A farewell to Bonferroni: The problems of low statistical power and publication bias. Behavioral Ecology 15: 1044–1045. doi:10.1093/beheco/arh107.

Ohsaki N & Sato Y (1994) Food plant choice of *Pieris* butterflies as a trade-off between parasitoid avoidance and quality of plants. Ecology 75: 59–68. doi: 10.2307/1939382.

Ohshima I (2010) Host-associated pre-mating reproductive isolation between host races of *Acrocercops transecta*: Mating site preferences and effect of host presence on mating. Ecological Entomology 35: 253–257. doi:10.1111/j.1365-2311.2009.01161.x

Okamura Y, Sawada Y, Hirai MY & Murakami M (2016) Effects of different secondary metabolite profiles in plant defense syndromes on specialist and generalist herbivores. Entomological Science 19: 97–103. doi: 10.1111/ens.12172.

Olsen CE, Huang XC, Hansen CIC, Cipollini D, Ørgaard M, Matthes A, Geu-Flores F, Koch MA & Agerbirk N (2016) Glucosinolate diversity within a phylogenetic framework of the tribe Cardamineae (Brassicaceae) unraveled with HPLC-MS/MS and NMR-based analytical distinction of 70 desulfoglucosinolates. Phytochemistry 132: 33–56. doi: 10.1016/j.phytochem.2016.09.013

Porter AH & Geiger H (1995) Limitations to the inference of gene flow at regional geographic scales-an example from the *Pieris napi* group (Lepidoptera: Pieridae)

in Europe. Biological Journal of the Linnean Society 54: 329–348. doi: 10.1016/0024-4066(95)90014-4.

Prasad KVSK, Song BH, Olson-Manning C, et al. (2012) A gain-of-function polymorphism controlling complex traits and fitness in nature. Science 337: 1081–1084. doi: 10.1126/science.1221636.

Ratnasingham S & Hebert PDN (2007) Barcoding Bold : The barcode of life data system. Molecular Ecology Notes 7: 355–364. doi: 10.1111/j.1471-8286.2006.01678.x.

Ratnasingham S & Hebert PDN (2013) A DNA-based registry for all animal species: the barcode index number (BIN) system. PLoS One 8: e66213. doi: 10.1371/journal.pone.0066213.

Raybould AF & Moyes CL (2001) The ecological genetics of aliphatic glucosinolates. Heredity 87: 383–391. doi: 10.1046/j.1365-2540.2001.00954.x.

Renwick JAA & Lopez K (1999) Experience-based food consumption by larvae of *Pieris rapae*: addiction to glucosinolates? Entomologia Experimentalis et Applicata 91:51–58. doi: 10.1046/j.1570-7458.1999.00465.x

Rohr F, Ulrichs C & Mewis I (2009) Variability of aliphatic glucosinolates in *Arabidopsis thaliana* (L .) – Impact on glucosinolate profile and insect resistance. Journal of Applied Botany and Food Quality 82: 131–135.

Ronquist F, Teslenko M, van der Mark P, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542. doi: 10.1093/sysbio/sys029.

Sachdev-Gupta K, Radke CD, Renwick JAA & Dimock MBD (1993) Cardenolides from *Erysimum cheiranthoides*: feeding deterrents to *Pieris rapae* larvae. Journal of Chemical Ecology 19: 1355–1369.

Sawada Y, Akiyama K, Sakata A, et al. (2009) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. Plant and Cell Physiology 50: 37–47. doi: 10.1093/pcp/pcn183.

Sawada Y, Tsukaya H, Li Y, Sato M, Kawade K, Hirai MY (2017) A novel method for single-grain-based metabolic profiling of *Arabidopsis* seed. Metabolomics 13: 75. doi:10.1007/s11306-017-1211-1

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

Tamura K, Stecher G, Peterson D, et al. (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725–2729. doi: 10.1093/molbev/mst197.

Velasco P, Soengas P, Vilar M, Cartea ME & Rio M, Del (2008) Comparison of glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. Journal of the American Society for Horticultural Science 133: 551–558.

Wahlberg N, Rota J, Braby MF, Pierce NE & Wheat CW (2014) Revised systematics and higher classification of pierid butterflies (Lepidoptera: Pieridae) based on molecular data. Zoologica Scripta 43: 641-650.

Wheat CW, Vogel H, Wittstock U, et al. (2007) The genetic basis of a plant-insect coevolutionary key innovation. Proceedings of the National Academy of Sciences of the USA 104: 20427–20431. doi: 10.1073/pnas.0706229104.

Winde I & Wittstock U (2011) Insect herbivore counteradaptations to the plant glucosinolate-myrosinase system. Phytochemistry 72: 1566–1575. doi:10.1016/j.phytochem.2011.01.016

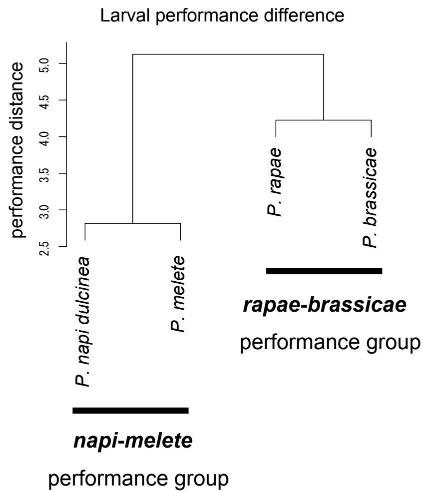
Wittstock U, Agerbirk N, Stauber EJ, et al. (2004) Successful herbivore attack due to metabolic diversion of a plant chemical defense. Proceedings of the National Academy of Sciences of the USA 101: 4859–4864. doi: 10.1073/pnas.0308007101.

Wittstock U & Halkier BA (2002) Glucosinolate research in the *Arabidopsis* era. Trends in Plant Science 7: 263–270.

Yagi T, Sasaki G & Takebe H, (1999) Phylogeny of Japanese papilionid butterflies inferred from nucleotide sequences of the mitochondrial *ND5* gene. Journal of Molecular Evolution 48: 42–48. doi: 10.1007/PL00006443.

Figures and Tables

(a)



(b)

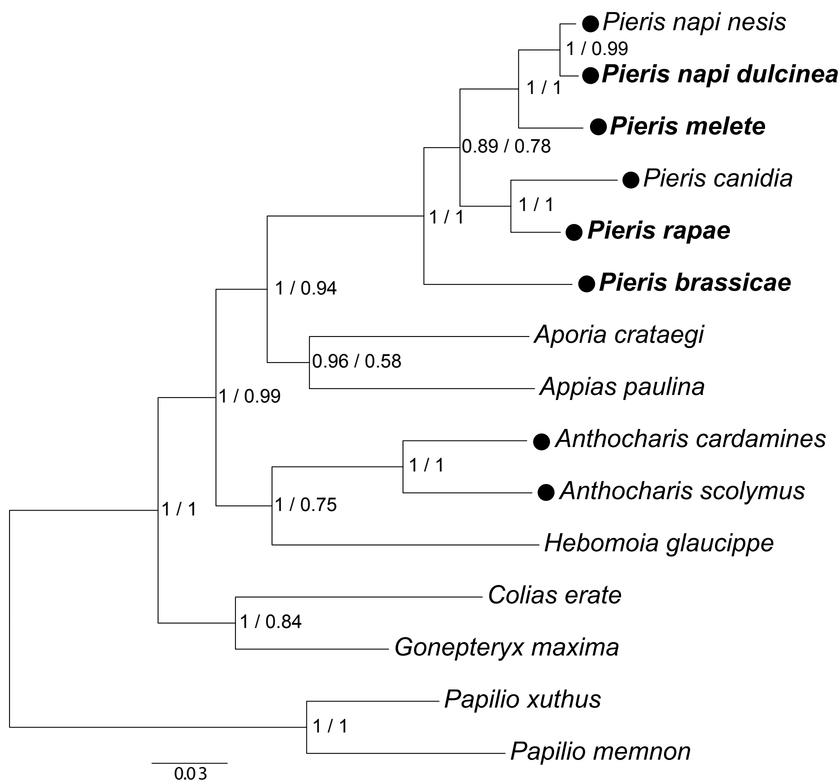


Fig. 2-1 (a) The host preference cladogram, constructed based on the results of feeding experiments. This figure shows that the larval performance levels of the four *Pieris*

butterfly species could be classified into two groups (*napi*/*melete* and *rapae*/*brassicae* groups). (b) The Bayesian phylogeny of Pierinae butterflies is shown. The clades are supported by posterior probabilities and bootstrap values (posterior probabilities/bootstrap). In this phylogeny, the species used for feeding experiments are in bold, and those that use Brassicales as host plants are marked with filled circles. Two *P. napi* subspecies (*P. napi dulcinea* and *P. napi nesis*) exist in Japan, although this classification is not without controversy. In this study, I treated *P. napi dulcinea* as *P. napi*. Comparing the cladogram and phylogeny, it shows that the host preference cladogram is reflected by their phylogeny.

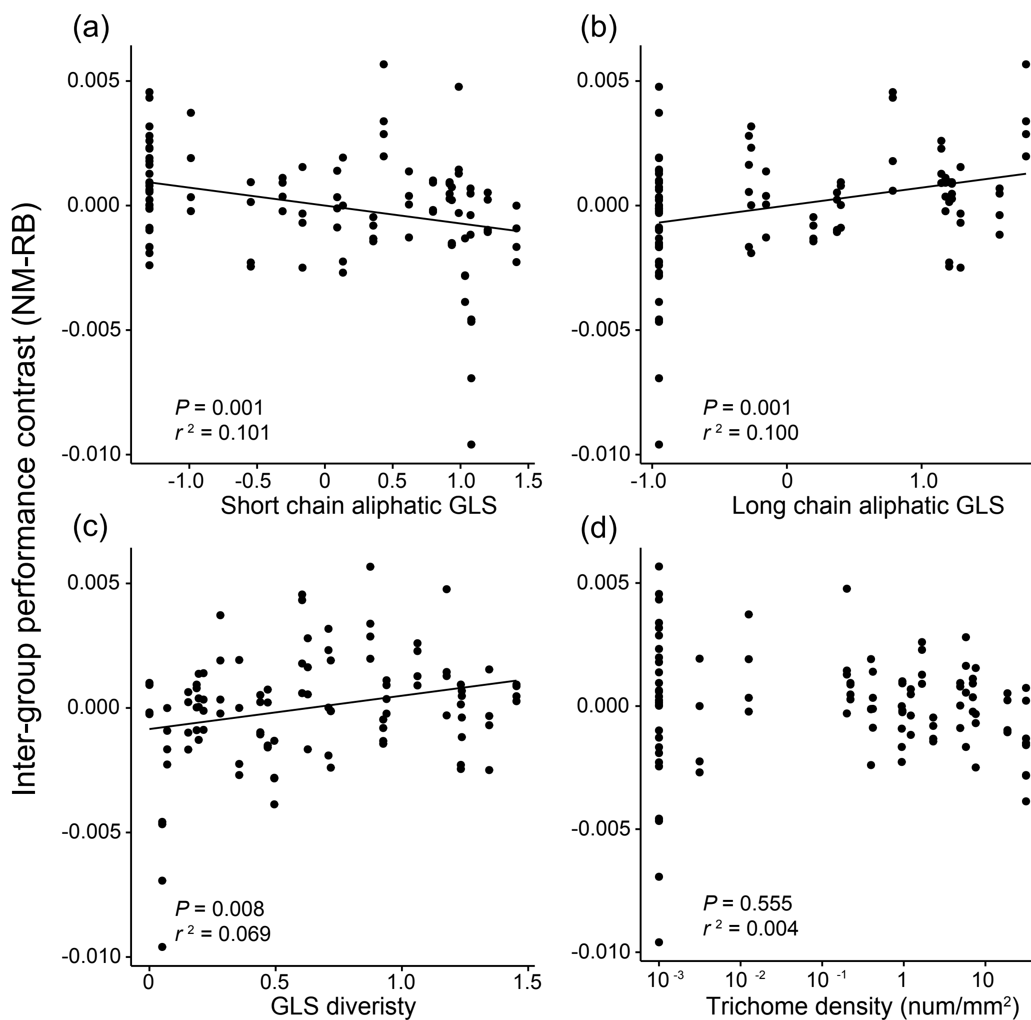


Fig. 2-2 The relationships between plant defense trait values and inter-group larval performance contrast (NM-RB: *napi*/*melete* vs. *rapae*/*brassicae* residuals) are shown. Positive NM-RB values indicate that the *napi*/*melete* group show a better performance than the *rapae*/*melete* group and negative values refer vice versa. (a) Relative concentration of short chain aliphatic GLS correlated negatively with NM-BR whereas (b) that of long chain aliphatic GLS and (c) GLS diversity (Shannon diversity indices based on relative glucosinolate concentration of each plant species) positively correlated with NM-RB. However, (d) no physical traits showed significant relationships with NM-RB including trichome density.

Table 2-1 Leaf traits that potentially affect the larval performance of the *Pieris* butterfly species

	P. napi		*P. melete*		*P. rapae*		*P. brassicae*	
Leaf Trait	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*
Tough	0.637	-0.099	0.735	-0.071	0.331	-0.203	0.172	-0.282
Trichome	**0.003***	**-0.567**	**0.004***	**-0.552**	**≤0.001***	**-0.680**	**≤0.001***	**-0.625**
Water	0.811	-0.051	0.936	0.017	0.483	0.147	0.664	-0.091
SLA	0.537	-0.130	0.957	-0.011	0.953	0.012	0.678	0.087
C:N	0.526	0.133	0.661	-0.092	0.305	0.214	0.274	0.228
Short	**0.015***	**-0.482**	**≤0.001***	**-0.680**	0.061	-0.380	0.087	-0.349
Long	0.509	0.138	0.274	0.228	0.350	-0.195	0.710	-0.078
Aliphatic	0.101	-0.336	**0.023***	**-0.454**	**0.017***	**-0.474**	0.186	-0.273
Indole	0.344	0.198	0.954	0.012	0.995	-0.001	0.815	0.049
Benzyl	0.109	0.328	**0.042**	**0.410**	0.070	0.369	0.076	0.361
GLS diversity	0.564	0.121	0.542	0.128	0.695	-0.082	0.269	-0.230
GLS richness	0.673	0.089	0.899	-0.027	0.382	-0.183	0.380	-0.184

Significant correlations from correlation-tests are represented in bold ($P \leq 0.05$). "*" shows significant P-values after the FDR adjustment ($P \leq 0.05$). Toughness; leaf toughness. SLA; specific leaf area. Short chain; short chain aliphatic glucosinolate concentration. Long chain; long chain aliphatic glucosinolate concentration. GLS diversity; Shannon diversity indices based on relative glucosinolate concentration of each plant species. Trichome density is shown as a high defensive ability against all the four *Pieris* butterfly species. Additionally, the aliphatic glucosinolate concentration (short-chain aliphatic or aliphatic) also exhibited defense against some of the tested species.

Table 2-2 The correlations between inter-group performance contrast (NM-RB: *napi*/*melete* group vs. *rapae*/*brassicae* group) and defense trait values.

	NM-RB	
	P	*r*
Toughness	0.288	0.107
Trichome	0.555	-0.060
Water	0.586	-0.055
SLA	0.223	-0.123
C:N	0.130	-0.153
Short	**0.001**	**-0.317**
Long	**0.001**	**0.317**
Aliphatic	0.311	-0.102
Indole	0.292	0.106
Benzyl	0.367	0.091
GLS diversity	**0.008**	**0.263**
GLS richness	**0.052†**	**0.195**

The correlations between inter-group performance contrast (NM-RB: *napi*/*melete* group vs. *rapae*/*brassicae* group) and defense trait values are shown. Toughness; leaf toughness. SLA; specific leaf area. Short chain; short chain aliphatic glucosinolate concentration. Long chain; long chain aliphatic glucosinolate concentration. GLS diversity; Shannon chemical diversity indices based on relative glucosinolate concentration of each plant species. NM-RB values were calculated as residuals of Deming regression of this two paired performance group. Positive NM-RB values refer *napi*/*melete* group grew better than *rapae*/*brassicae* group on that plant, and negative values represent the opposite situation. Significant values are in bold ($P \leq 0.05$, "†"; $P \leq 0.06$)

Supplementary information

Table 2-S1 Butterfly species used in this study for feeding experiments or phylogenetic analysis. The collection site and Genbank ID of utilized butterflies species are shown. The species with collection sites are collected by the authors, and the sequences of them were identified except ND5 region of *Colias erate*. The sequences of the species without collection sites were obtained from Genbank. The four *Pieris* species, which I used in feeding experiments, are in bold.

Species	Collection Site	*CO1*	*ND5*	*EF-1α*
For Pieirnae Phylogeny				
Anthocharis cardamines	Yamanashi,	LC090556	LC090580	LC090568
Anthocharis scolymus	Chiba, Japan	LC090557	LC090581	LC090569
Pieris brassicae	Hokkaido, Japan	LC090562	LC090585	LC090574
Pieris rapae	Hokkaido, Japan	LC090567	LC090590	LC090579
Pieris melete	Chiba, Japan	LC090564	LC090587	LC090576
Pieris napi dulcinea	Hokkaido, Japan	LC090565	LC090588	LC090577
Pieris napi nesis	Nagano, Japan	LC090566	LC090589	LC090578
Pieris canidia	Okinawa, Japan	LC090563	LC090586	LC090575
Hebomoia glaucippe	Okinawa, Japan	LC090561	LC090584	LC090573
Appias paulina	Okinawa, Japan	LC090559	LC090583	LC090571
Aporia crataegi	Hokkaido, Japan	LC090558	LC090582	LC090570
Colias erate	Chiba, Japan	LC090560	(AB855816)	LC090572
For *Pieris* Phylogeny				
Pieris bryoniae	-	GU707090	-	-
Pieris cheiranthi	-	EU143662	-	-
Pieris davidis	-	JQ922060	-	-
Pieris deota	-	FJ663930	-	-
Pieris ergane	-	KP870437	-	-
Pieris mannii	-	KP870867	-	-
Pieris marginalis	-	KM540739	-	-
Pieris narina	-	FJ663934	-	-
Pieris ochsenheimeri	-	FJ663939	-	-
Pieris oleracea	-	GU097047	-	-
Pieris pseudorapae	-	FJ663941	-	-

Table 2-S2 25 Brassicaceae plant species I utilized in feeding experiments. The plants utilized for replicate experiments in 2015 are in bold. The seeds collection sites of plants that purchased as cultivars are without collection sites. The main lineage numbers are shown from Beilstein et al. (2008).

Plant species	Seed collection site	Lineag
Arabidopsis kamchatica	Nagano, Japan (36° 31' N, 138° 20' E)	1
Arabidopsis thaliana	Hokkaido, Japan (42° 50' N, 141° 41' E)	1
Arabis hirsuta	Nagano, Japan (36° 31' N, 138° 20' E)	-
Aubrieta deltoidea	-	-
Aurinia saxatilis	-	-
Berteroa incana	Burlington, Canada (43° 30' N, 79° 79' W)	-
Brassica napus	Chiba, Japan (35° 30' N, 140° 50' E)	2
Brassica tournefortii	Chiba, Japan (35° 54' N, 139° 56' E)	2
Capsella bursa-pastoris	Chiba, Japan (35° 30' N, 140° 50' E)	1
Cardamine hirsuta	Chiba, Japan (35° 30' N, 140° 50' E)	1
Cardamine regeliana	Chiba, Japan (35° 30' N, 140° 50' E)	1
Cardamine scutata	Chiba, Japan (35° 30' N, 140° 50' E)	1
Descurainia sophia	Inner Mongolia, China (43° 37' N, 116° 42' E)	-
Diplotaxis tenuifolia	-	2
Dontostemon dentatus	Inner Mongolia, China (43° 37' N, 116° 42' E)	3
Draba nemorosa	Nagano, Japan (36° 20' N, 137° 50' E)	-
Eruca sativa	-	2
Erysimum cheiranthoides	Hokkaido, Japan (42° 50' N, 141° 41' E)	1
Lepidium virginicum	Chiba, Japan (35° 30' N, 140° 50' E)	1
Nasturtium officinale	Chiba, Japan (35° 30' N, 140° 50' E)	1
Raphanus sativus var.	Chiba, Japan (35° 07' N, 140° 11' E)	2
Rorippa indica	Chiba, Japan (35° 30' N, 140° 50' E)	1
Sisymbrium orientale	Chiba, Japan (35° 54' N, 139° 56' E)	2
Thlaspi arvense	Hokkaido, Japan (42° 56' N, 142° 01' E)	2
Turritis glabra	Nagano, Japan (36° 31' N, 138° 20' E)	1

Table 2-S3 The list of primers I used in this study for reconstructing Pierinae phylogeny.

Gene name	Primer sequence	Reference
ND5	F 5'-CCTGTTTCTGCTTTAGTTCA-3'	Yagi et al. 1999
	R 5'-AATATDAGGTATAAATCATAT-3'	Yagi et al. 1999
EF-1α	F 5'-GCYGARCGYGARCGTGGTATYAC-3'	Monteiro and Pierce 2001
	R 5'-ACAGCVACKGTYTGYCTCATRTC-3'	Kandul et al. 2004

Table 2-S4

Larval weights of *Pieris* butterfly species larvae from the feeding experiments.

	P. napi		*P. melete*		*P. rapae*		*P. brassicae*	
A. kamchatica	11.33	±0.71	6.95	±0.76	5.65	±0.68	5.71	±1.04
A. thaliana	5.97	±0.66	4.29	±0.26	4.53	±0.31	6.57	±0.74
A. hirsuta	7.45	±1.41	6.10	±1.10	5.66	±0.69	5.53	±0.37
A. deltoidea	2.23	±0.27	0.15	±0.10	0.53	±0.38	1.50	±0.15
A. saxatilis	0.53	±0.53	1.02	±0.61	0.18	±0.11	0.90	±0.31
B. incana	1.17	±0.44	1.56	±0.58	0.95	±0.24	0.87	±0.28
B. napus	8.59	±1.11	4.08	±0.71	5.65	±0.60	5.49	±1.36
B. tournefortii	3.31	±0.62	4.23	±0.67	3.59	±0.33	3.36	±0.83
C. bursa-pastoris	0.39	±0.18	4.03	±0.72	0.75	±0.15	0.00	±0.00
C. hirsuta	5.85	±0.88	6.38	±1.06	8.63	±0.62	8.49	±0.77
C. regeliana	9.15	±1.87	8.06	±1.41	10.70	±0.99	9.63	±1.28
C. scutata	9.38	±0.92	6.83	±1.13	10.68	±1.45	15.10	±2.42
D. sophia	2.97	±1.50	0.37	±0.29	4.55	±0.31	7.96	±1.50
D. tenuifolia	8.03	±1.33	4.40	±0.66	10.22	±0.81	8.25	±1.26
D. dentatus	9.63	±1.41	4.94	±0.72	6.75	±1.05	5.03	±2.53
D. nemorosa	4.70	±0.55	4.48	±0.22	4.94	±0.65	3.92	±0.62
E. sativa	5.00	±0.67	5.16	±0.43	6.53	±0.56	5.55	±0.30
E. cheiranthoides	0.30	±0.19	0.41	±0.14	0.75	±0.15	0.00	±0.00
L. virginicum	6.91	±1.18	5.64	±0.49	5.59	±0.51	16.79	±1.85
N. officinale	3.53	±0.53	5.50	±1.16	6.78	±1.12	8.63	±3.73
R. sativus	3.05	±0.39	2.23	±0.68	4.69	±0.84	4.95	±0.95
R. indica	5.73	±1.06	8.70	±0.70	4.05	±1.43	17.82	±3.70
S. orientalis	2.70	±0.47	3.24	±0.27	3.18	±0.45	7.12	±0.41
T. arvense	0.50	±0.21	0.41	±0.14	7.13	±0.95	11.84	±3.85
T. glabra	2.78	±0.15	3.10	±0.38	2.37	±0.23	4.55	±0.27

Average larval weights (mg, ±SE) of *Pieris* butterfly species after 120 hours of feeding (n = 6-12) against 25 Brassicaceae plant species are shown.

Table 2-S5 The detected glucosinolate compounds detected from 25 Brassicaceae plants with their category and abbreviation.

GLS name	Abbreviation	Group	Chain length
2-propenyl	2P	aliphatic	short
3-(Methylsulfinyl)propyl	3MSOP	aliphatic	short
4-(Methylsulfinyl)butyl	4MSOB	aliphatic	short
5-(Methylsulfinyl)pentyl	5MSOP	aliphatic	short
6-(Methylsulfinyl)hexyl	6MSOH	aliphatic	long
7-(Methylsulfinyl)heptyl	7MSOH	aliphatic	long
8-(Methylsulfinyl)octyl	8MSOO	aliphatic	long
3-(Methylthio)propyl	3MTP	aliphatic	short
4-(Methylthio)butyl	4MTB	aliphatic	short
5-(Methylthio)pentyl	5MTP	aliphatic	short
6-(Methylthio)hexyl	6MTH	aliphatic	long
7-(Methylthio)heptyl	7MTH	aliphatic	long
8-(Methylthio)octyl	8MTO	aliphatic	long
1-Methoxyindol-3-ylmethyl	1MOI3M	indole	-
4-Methoxyindol-3-ylmethyl	4MOI3M	indole	-
Indol-3-ylmethyl	I3M	indole	-
(S)-2-Hydroxy-2-phenethyl	S2H2P	benzyl	-
Benzyl	Bz	benzyl	-
Phenethyl	PE	benzyl	-
3-(Benzoyloxy)propyl	3BZOP	benzyl	-
4-(Benzoyloxy)butyl	4BZOB	benzyl	-

Table 2-S6

Standardized peak areas of detected glucosinolates and calculated glucosinolae diversity indices in this study

GLS	*A.ka*	*A.th*	*A.hi*	*A.de*	*A.sa*	*B.in*	*B.na*	*B.to*	*C.bu*	*C.hi*	*C.re*	*C.sc*	*D.so*	*D.te*	*D.de*	*D.ne*	*E.sa*	*E.ch*	*L.vi*	*N.of*	*R.sa*	*R.in*	*S.or*	*T.ar*	*T.gl*
2P	0.0029	0.0069	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.2859	n.d.	0.0008	n.d.	n.d.	0.0002	n.d.	n.d.	n.d.	n.d.	n.d.	0.3499	n.d.
3MSOP	0.0080	0.1565	n.d.	n.d.	n.d.	n.d.	n.d.	0.1008	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.5067	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
4MSOB	0.0067	0.0096	n.d.	n.d.	0.0011	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0046	0.1719	n.d.	0.0452	0.0136	n.d.	n.d.	0.0013	n.d.	0.0044	n.d.	0.0055
5MSOP	0.0027	0.0001	n.d.	0.0014	0.1668	0.5468	<0.0001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0059	n.d.	0.0001	0.0000	n.d.	0.0002	0.0004	n.d.	0.0001	n.d.	0.3182
6MSOH	0.1627	0.0001	n.d.	n.d.	n.d.	0.0036	n.d.	n.d.	0.0002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0061	n.d.	n.d.	n.d.	0.0055	n.d.	n.d.	n.d.	n.d.	0.7450
7MSOH	1.4541	0.0173	n.d.	0.0017	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0323	n.d.	n.d.	0.0018	0.0960	n.d.	0.0027	n.d.	n.d.	0.0216
8MSOO	0.3624	0.1497	0.1107	0.2073	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0969	n.d.	n.d.	n.d.	0.0256	n.d.	0.0206	n.d.	n.d.	n.d.
3MTP	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0000	n.d.	n.d.	0.0001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
4MTB	n.d.	n.d.	n.d.	<0.0001	0.0002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0008	0.0527	n.d.	0.0007	0.0001	n.d.	n.d.	0.0121	n.d.	n.d.	n.d.	0.0124
5MTP	n.d.	n.d.	n.d.	n.d.	0.0176	0.0504	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0014	0.0008	n.d.	n.d.	n.d.	0.0001	0.0007	n.d.	n.d.	n.d.	0.0054
6MTH	n.d.	n.d.	n.d.	n.d.	n.d.	0.0001	n.d.	n.d.	n.d.	n.d.	0.0002	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0004	n.d.	n.d.	n.d.	0.0559
7MTH	0.0001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0210	n.d.	n.d.	n.d.	n.d.	n.d.
8MTO	0.0007	0.0015	0.0079	0.0157	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0009	0.0004	n.d.	0.0025	0.0061	0.0013	0.0003	n.d.	n.d.	n.d.

GLS	*A.ka*	*A.th*	*A.hi*	*A.de*	*A.sa*	*B.in*	*B.na*	*B.to*	*C.bu*	*C.hi*	*C.re*	*C.sc*	*D.so*	*D.te*	*D.de*	*D.ne*	*E.sa*	*E.ch*	*L.vi*	*N.of*	*R.sa*	*R.in*	*S.or*	*T.ar*	*T.gl*
1MOI3M	n.d.	0.0210	n.d.	n.d.	0.0009	n.d.	0.0058	n.d.	n.d.	0.0008	n.d.	n.d.	n.d.	n.d.	0.0133	0.0007	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
4MOI3M	0.0307	0.4676	0.0164	0.1514	0.0032	0.0160	n.d.	n.d.	n.d.	n.d.	0.0566	n.d.	0.0534	0.0514	0.0175	0.2504	0.0002	0.0039	0.0212	0.0068	n.d.	0.1084	n.d.	0.0031	0.1891
I3M	n.d.	0.1247	n.d.	0.1187	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
S2H2P	n.d.	n.d.	0.0725	n.d.	n.d.	0.0001	n.d.	n.d.	0.0001	n.d.	n.d.	0.0015	0.0039	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Bz	n.d.	n.d.	n.d.	0.0187	n.d.	n.d.	0.0004	n.d.	n.d.	0.0208	0.0439	0.1307	n.d.	n.d.	n.d.	n.d.	0.0001	n.d.	0.6296	n.d.	n.d.	0.0015	n.d.	<0.0001	0.0017
PE	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0366	0.0002	0.0025	n.d.	n.d.	0.0094	n.d.	n.d.	n.d.	0.0001	0.2068	n.d.	n.d.	n.d.	n.d.	n.d.
3BZOP	n.d.	<0.0001	n.d.	0.0018	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	<0.0001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
4BZOB	n.d.	n.d.	0.0014	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	<0.0001	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.0001	0.0001	n.d.	n.d.
GLS div.	0.87	1.45	1.06	1.35	0.47	0.44	0.28	0.00	0.63	0.72	0.71	0.15	0.49	0.36	1.18	0.94	0.20	0.07	0.19	1.23	0.93	0.60	0.21	0.05	1.24
Short	0.4213	1.0502	0.0022	0.0027	0.2568	0.6632	0.1445	0.1008	n.d.	0.3116	n.d.	n.d.	0.2922	0.0054	2.3056	0.0044	0.0464	1.5205	0.0002	0.0039	0.1483	n.d.	0.4346	0.3509	0.3422
Long	1.9801	0.1686	0.1186	0.2247	n.d.	0.0037	n.d.	n.d.	0.0002	n.d.	0.0002	n.d.	n.d.	n.d.	n.d.	0.1362	0.0004	n.d.	0.0042	0.1541	0.0017	0.0236	n.d.	n.d.	0.8225
Aliph	2.4014	1.2188	0.1208	0.2274	0.2568	0.6669	0.1445	0.1008	0.0002	0.3116	0.0002	n.d.	0.2922	0.0054	2.3056	0.1406	0.0467	1.5205	0.0044	0.1581	0.1500	0.0236	0.4346	0.3509	1.1647
Indol	0.0307	0.6133	0.0164	0.2701	0.0041	0.0160	0.0058	n.d.	n.d.	0.0008	0.0566	n.d.	0.0534	0.0514	0.0308	0.2511	0.0002	0.0039	0.0212	0.0068	n.d.	0.1084	n.d.	0.0031	0.1891
Benz	n.d.	<0.0001	0.0739	0.0205	n.d.	0.0001	0.0004	n.d.	0.0001	0.0574	0.0441	0.1347	0.0040	n.d.	0.0094	n.d.	0.0001	n.d.	0.6296	0.2068	n.d.	0.0016	0.0001	<0.0001	0.0017

The standardized peak areas of each detected GLS from 25 Brassicaceae plant species and calculated GLS diversity are shown. Plat species name and the abbreviations of each glucosinolate are listed in Table 2-S2, S5. n.d.; not detected. GLS div.; Shannon glusocinolate diversity based on relative glucosinolate concentration of each plant species. The totalized concentrations of each class of glucosinolate are shown as Short (short chain aliphatic), Long (long chain aliphatic), Aliph (aliphatic), Indol (indole), and Benz (Benzyl).

Table 2-S7 Defense traits that affect interspecific larval performance difference

Defense	NM		NR		NB		MR		MB		RB	
	P	*r*	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*	*P*	*r*
Toughness	0.866	-0.035	0.522	0.134	0.608	0.108	0.565	0.121	0.706	0.079	0.912	0.023
Trichome	0.840	0.042	0.377	0.185	0.499	-0.142	0.847	0.041	0.221	-0.254	0.244	-0.242
Water	0.604	-0.109	0.255	-0.237	0.944	0.015	0.540	-0.129	0.741	0.069	0.180	0.277
SLA	0.373	-0.186	0.438	-0.162	0.315	-0.209	0.903	-0.026	0.772	-0.061	0.735	-0.071
C:N	0.071	0.368	0.603	-0.109	0.891	-0.029	0.115	-0.323	0.274	-0.228	0.831	0.045
Short chain	**0.051**	**0.395**	0.690	-0.084	0.209	-0.26	**0.041**	**-0.411**	**0.005***	**-0.547**	0.519	-0.135
Long chain	0.421	-0.168	**0.051**	**0.395**	0.309	0.212	**0.020**	**0.463**	0.157	0.292	0.418	-0.170
Aliphatic	0.245	0.241	0.347	0.196	0.451	-0.158	0.804	-0.052	0.094	-0.342	0.107	-0.330
Indole	0.157	0.292	0.277	0.226	0.392	0.179	0.941	0.016	0.946	-0.014	0.810	-0.051
Benzyl	0.395	-0.178	0.717	-0.076	0.695	0.082	0.610	0.107	0.237	0.245	0.605	0.109
GLS diversity	0.900	-0.027	0.252	0.238	0.133	0.309	0.263	0.233	0.195	0.268	0.556	0.124
GLS richness	0.373	0.186	0.115	0.323	0.251	0.238	0.466	0.153	0.728	0.073	0.815	-0.049

Larval performance differences between pair of *Pieris* butterflies are compared with defense traits of plants by correlation test. Performance differences were calculated as residuals from Deming regression of the pair. Comparative sets of butterfly species as follows: NM, *P. napi* vs. *P. melete*; NR, *P. napi* vs. *P. rapae*; NB, *P. napi* vs. *P. brassicae*; MR, *P. melete* vs. *P. rapae*; MB, *P. melete* vs. *P. brassicae*; and RB, *P. rapae* vs. *P. brassicae.* Toughness; leaf toughness. SLA; specific leaf area. Short chain; short chain aliphatic glucosinolate concentration. Long chain; long chain aliphatic glucosinolate concentration. GLS diversity; Shannon

glucosinolate diversity based on relative glucosinolate concentration of each plant species. Significant correlations refer that larval performance of two paired species differ along with the defense traits. Significant and marginal values from pairwise correlation tests are represented in bold ($P \leq 0.05$, "†"; $P \leq 0.055$). "*" shows significant P-values after the FDR correction ($P \leq 0.05$).

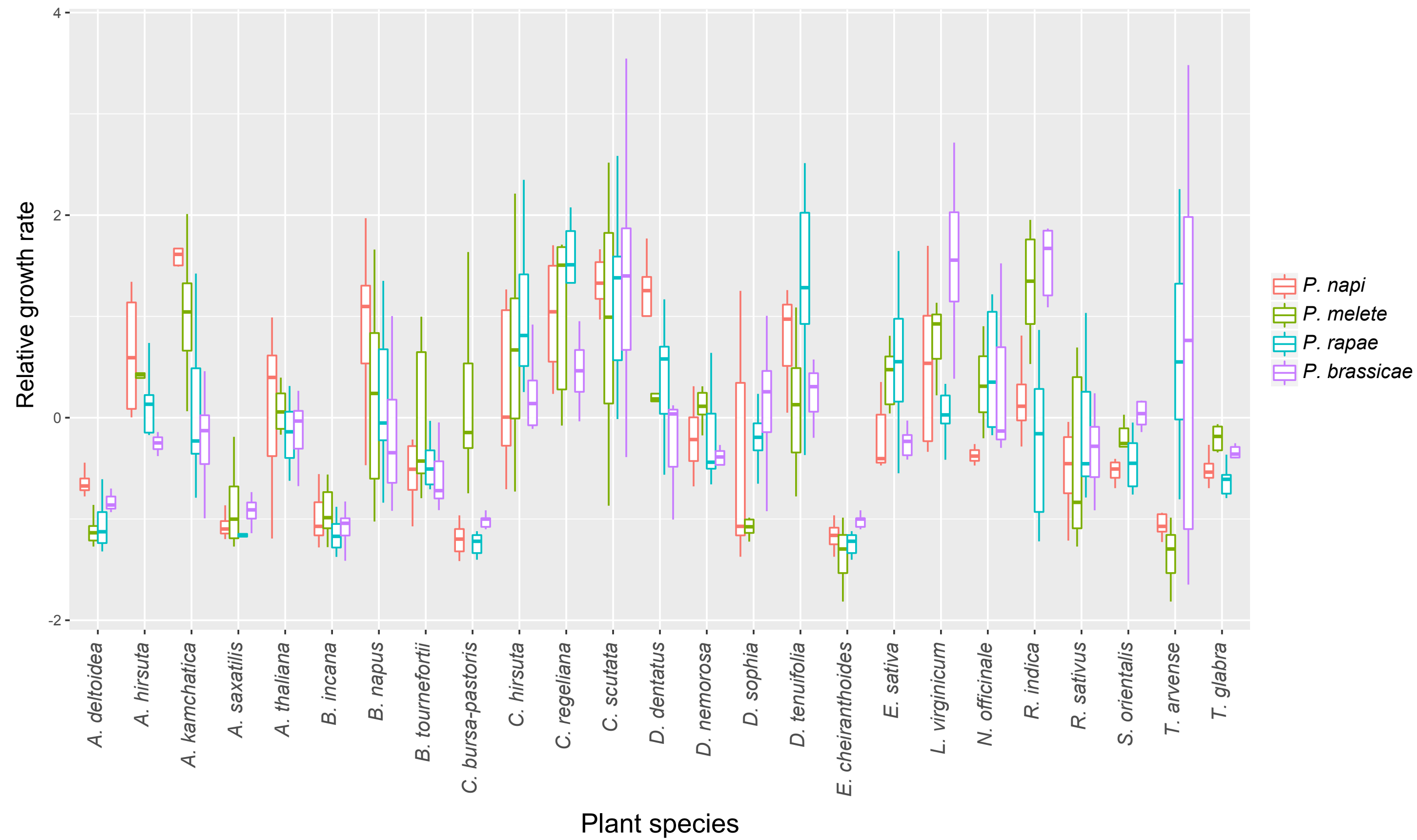


Fig. 2-S1 The result of feeding experiment

All of the four *Pieris* butterfly species grew well on the plant genus *Cardamine* (*C. hirsute*, *C. regeliana*, *C. scutata*), however, all of the species could not grew on *Erysimum cheiranthoides* or *Berteroa incana*.

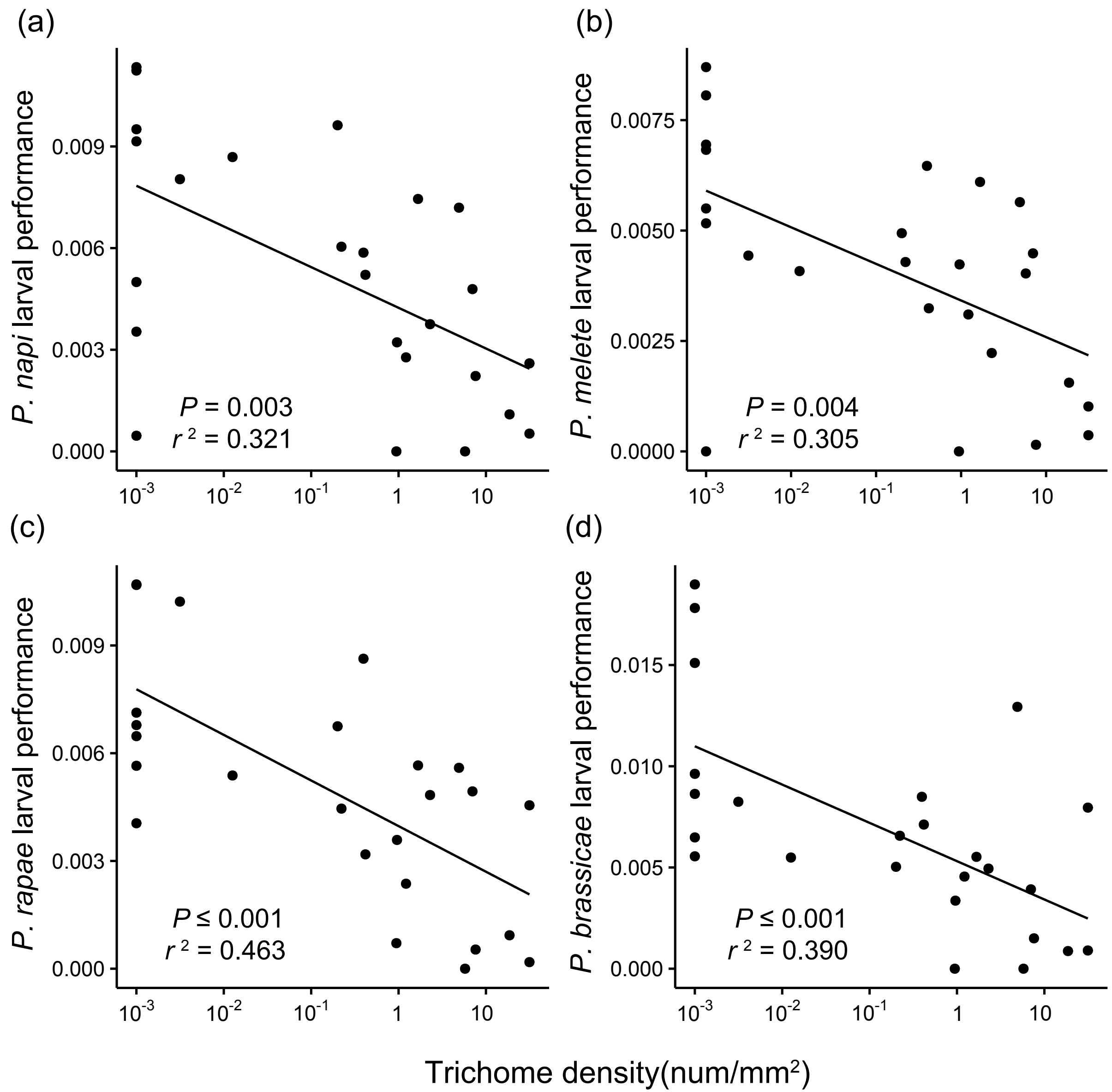


Fig. 2-S2 The relationship between plant trichome density and larval performance of the four *Pieris* species. Trichome density worked as an efficient defense against all the four *Pieris* species I utilized in this study, however, this defense traits could not explain the larval performance differences among the four species.

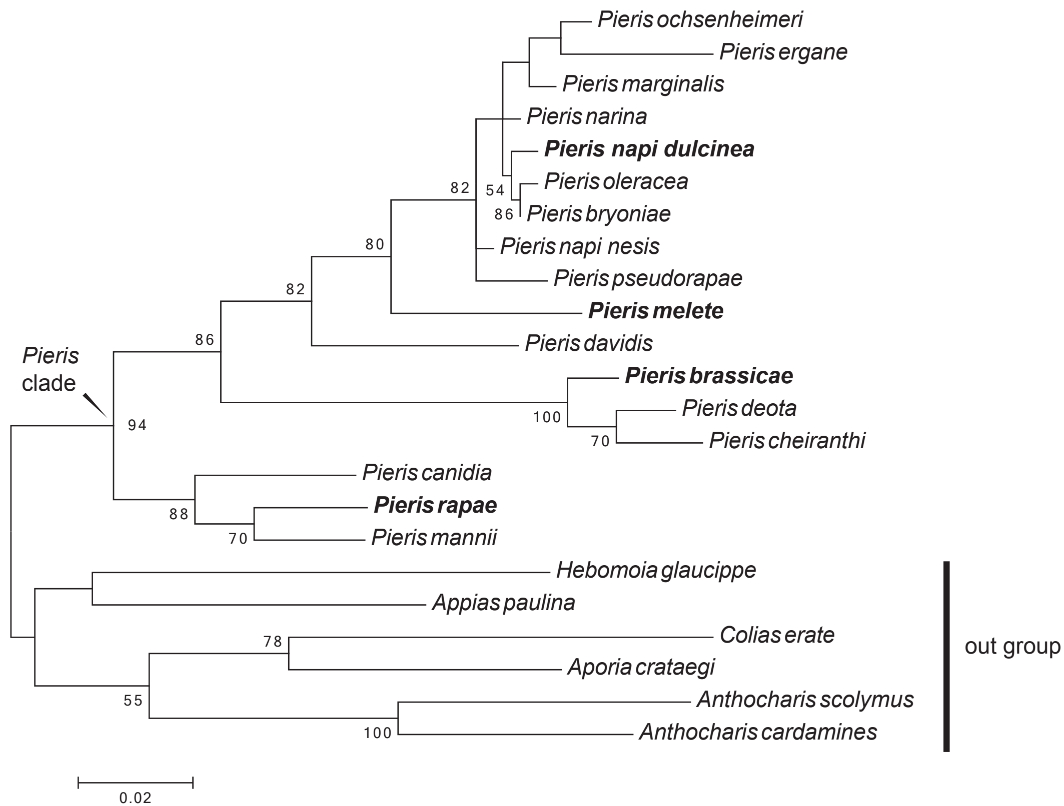


Fig. 2-S3 The ML tree of genus *Pieris* generated by the CO1 region. The species I used for feeding experiments are in bold, and they spread widely in genus *Pieris*. Bootstrap values over 50% are shown in nodes.

Chapter 3 *Pieris* butterflies regulate two genes differently for dealing with wider range of glucosinolate profiles in host plants

Abstract

Herbivorous insects have to deal with a wide range of chemical compounds produced by their host plants to utilize them as hosts. Resistance to a certain range of secondary metabolites in herbivores would be strongly associated with their host ranges. However, the adaptation mechanisms of these resistances have not been well understood especially in specialist herbivores, which has narrower host range as herbivores, even they as well are exposed to a certain range of chemical compounds in their narrower hosts. Here, I focused on *Pieris* butterflies which are known as one of specialist herbivores of Brassicaceae plants, the plant family which possesses more than 140 identified types of glucosinolates as specific chemical defenses. I focused on nitrile specifier protein (NSP), which is known to act as a glucosinolate detoxification in *Pieris* larvae, and its gene family expression when *Pieris* larvae fed on plants with different GLS profiles. I found expression level of NSP was dramatically changed relying on the plant species the larvae fed on. Furthermore, lower expression of NSP was complemented with higher expression of its sister gene called major allergen (MA). These dynamic expression regulation patterns of NSP and MA were confirmed by *Arabidopsis thaliana* mutants with different GLS profiles, and were also general among four *Pieris* butterflies used in this study. My results suggest *Pieris* butterflies handle two different genes under fine regulations to deal with plants with different glucosinolate profiles. This proposes the first glimpse to understand the mechanism of host range evolution in specialist herbivores which can have more restricted interaction with plants and can give us more clear evidence in future research.

Introduction

Herbivorous insects are exposed to a variety of toxic secondary metabolites produced by host plants. As a coevolutionary consequence, herbivores should have their own specific adaptation strategies to deal with these toxic compounds for utilizing plants as hosts (Futuyma & Agrawal 2009). Recently, several adaptation traits or mechanisms of herbivores have been detected such as cytochrome P450 proteins adapted to furanocoumarins by *Papilio* butterflies or UDP-glycosyltransferases from *Helicoverpa armigera* or *Helicoverpa zea* to capsaicin (Heidel-Fischer & Vogel 2015; Futuyma & Agrawal 2009).

Glucosinolate (GLS) is one of the secondary metabolites which is mainly observed in brassicales (Wittstock & Halkier 2002). GLS is stored in plant cells isolated from its specific enzymes called myrosinase. Upon plant tissue damage, GLS contacts with myrosinase and is immediately hydrolyzed with forming several breakdown products (Halkier & Gershenzon 2006). Isothiocyanates (ITCs) are spontaneously formed dominant breakdown products from this reaction and have strong toxicity against herbivores (Wittstock & Halkier 2002). GLSs are highly diversified because of its variable side chain form (more than 140 types are known) and are classified into three distinct classes depending on its differential biosynthesis origins (aliphatic GLS, benzyl GLS and Indol GLS, Fahey et al. 2001; Olsen et al. 2016). Brassicales plants, especially core Brassicaceae plants, possess several types of GLS from different classes as multiple GLS profiles which can be species specific (Hofberger et al. 2013; Fahey et al. 2001).

GLS defense system is known as “mustard oil bomb” which can be a strong defense system against herbivores (Agrawal & Kurashige 2003; Wittstock & Burow 2010). However, some of insect herbivores can utilize brassicales plants as hosts by dealing with this GLS defense system (Heidel-Fischer & Vogel 2015; Futuyma & Agrawal 2009). One of the well-known

Brassicaceae crop pests, *Plutella xylostella,* can cleave off the reactive sulfate of GLSs by specific protein called glucosinolate sulfatase (GSS) before the toxic ITCs are formed (Ratzka et al. 2002). GSS can act on distinct classes of GLSs which enables *Plutella xylostella* be common pest on variety of Brassicaceae crops (Ratzka et al. 2002). On the other hand, Pierid butterflies can also use Brassicaceae as host by dealing with GLS defense by larval gut expressed protein called nitrile speicfier protein (NSP) (Wittstock et al. 2004). NSP can redirect the spontaneous GLS - myrosinase reaction to form less toxic nitriles rather than toxic ITCs. Major allergen (MA) and single domain major allergen (SDMA) genes are known to be included in NSP gene family, which is specific gene family as any related gene family has never been found (Fischer et al. 2008). SDMA is generally possessed by lepidopteran insects, and NSP and MA are known to be derived from SDMA by gene duplications. NSP and MA are sister genes and have three replicated domains which are originated from SDMA (Fischer et al. 2008). Although the function of MA is not defined, MA likely has functions which relates to GLS detoxification because NSP and MA are both specific in brassicales feeding pierid butterfly (Fischer et al. 2008). In addition, the substrate specificity of NSP are not well investigated as that of GSS are well reported in previous research, because the concrete molecular reaction mechanism of NSP is still unclear (Hofberger et al. 2013; Wittstock et al. 2004; Ratzka et al. 2002).

Since one pierid butterfly species tend to use a certain range of brassicales plants as host plants, such species are expected to be exposed to a range of GLSs from different plants (Chew 1980). Especially, some of *Pieris* butterflies rely on several different genus in core Brassicaceae, the family which has highest GLS diversity in brassicales (Edger et al. 2015; Hofberger et al. 2013). For example, *Pieris melete* utilizes at least 5 genus of Brassicaceae plants (Ohsaki & Sato 1994). However, it remains still unknown whether pierid butterflies deal with broader range of

GLSs only by NSP. Here, I address this question by focusing on entire NSP gene family regulation when larvae of *Pieris* butterflies are exposed to completely different GLS profiles of plants.

In the present study, I conduct feeding experiments utilizing *Pieris* butterfly species and two Brassicaceae plants which has completely different GLS background, coupled with gene expression level analysis. I use four closely related *Pieris* butterfly species which have different host range, aiming to see the potential differences in NSP gene expression patterns among *Pieris* species associated with their host ranges. NSP gene family sequences have only been identified in *Pieris rapae* and *P. brassicae* in this genus (Edger et al. 2015; Heidel-Fischer et al. 2010; Fischer et al. 2008). Therefore, I combine RNA sequencing (RNA-seq) method with this feeding experiment to identify NSP gene family sequences and their expression levels at once from the four *Pieris* butterfly species.

Different plant species should represent different GLS profiles, therefore, I expect to observe distinct gene expression responses of *Pieris* larvae against the two Brassicaceae plants. However, since different plant species differ in their chemical background in variety of ways, this background difference can also affect gene regulations of larvae strongly. Therefore, I also conduct same sets of feeding experiments with *Arabidopsis thaliana* mutants, which have different GLS profiles with sharing same species chemical background. Combining these two approaches, I can observe how *Pieris* butterflies regulate NSP gene family to deal with broader range of GLS defenses from their wider range of host plants.

Materials and Methods

1. 1 Feeding experiments utilizing two wild Brassicaceae plant species

I conducted feeding experiments utilizing four closely related *Pieris* butterfly species (*Pieris napi*, *P. melete*, *P. rapae*, and *P. brassicae*) and two Brassicaceae plants from different genera (*Arabidopsis kamchatica* and *Cardamine scutata*). According to previous research (Chapter 3), these two Brassicaceae plants are known to possess quite different GLS profiles. I collected mother butterflies of these four species from the wild population. For *P. brassicae*, I collected final instar larvae and reared them to adults. After eclosion, several pairs of adult were made by hand-pairing. I place the mother butterflies for egg-laying in a chamber with their host plants (*Brassica oleracea* var. *capitata* or *Cardamine leucantha*) under high intensity light condition. Acquired eggs were incubated at 25°C and neonates were utilized for feeding experiments immediately after hatching. I collected seeds of the two different Brassicaceae plants from the wild population. The collected seeds were watered, and germinated seeds were transplanted to vermiculite soil. I watered plants once a week with optimally diluted Hyponex solution (N:P:K = 6:10:5; Hyponex, Japan). I reared the plants for 2 months under the condition 25°C, with 60% relative humidity and L16:D8 before the feeding experiment.

For feeding experiments, 3 neonates were applied to one plant individual by soft-haired brush. I replicated this sets twice for each butterfly species and harvested totally 6 individuals from each plant species after 120 hours of feeding. I measured weight of harvested larvae individually (within 0.1 mg). Harvested larvae were flash frozen in liquid nitrogen immediately after the weight measurement and stored at -80°C until RNA extraction.

1. 2 RNA extraction, RNA-seq, de novo assembly, NSP gene family sequence identification and expression level analysis.

For each butterfly species, I selected one representative larva from each plant species. In total, I chose 8 larvae for RNA-seq (4 *Pieris* species against 2 plant species). I extracted RNA by using RNeasy Mini Kit (QIAGEN). Extracted RNA samples were quality controlled with Agilent 2100 Bioanalyzer and all samples were confirmed as having higher quality score for sequencing (RIN > 9). Library for RNA-seq was prepared by SureSelect Strand-Specific RNA Library Preparation Kit for Illumina Multiplexed Sequencing. I sequenced the samples with HiSeq 1500 (100bp PE). Acquired reads were trimmed by trimmomatic with following options (LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:40 --normalize_reads) (Bolger et al. 2014). For de-novo assembly, I pooled the trimmed data from the same species as one read data. I conducted de-novo assembly with Trinity ver. 2.0.6 (Grabherr et al. 2011). For identifying NSP gene family sequences from each species, I conducted tblastn with setting the assembled contigs (backbone) as databases and NSP gene family protein sequences from *P. rapae* as queries (Altschul et al. 1990). I extracted best hits for each query from each species and they were aligned with MEGA6 to reference sequence and trimmed in ORF (Tamura et al. 2013). For measurements of relative expression level of each extracted genes, I excluded redundant isoforms of NSP gene family observed in assembled contig backbone and replaced them by trimmed representative sequence in ORF. Expression levels of each gene were estimated by mapping trimmed reads on assembled backbone by RSEM (Li & Dewey 2011). Fragments per kilobase of exon per million reads mapped (FPKM) were used as relative expression levels of each gene.

2. 1 Feeding experiments with Arabidopsis thaliana mutants with different GLS background

I prepared four *Arabidopsis thaliana* mutant lines which have different GLS profiles, such that Col-0; wild type, MAM1; lacking short chain aliphatic GLS synthesizing gene, MAM3; lacking long chain aliphatic GLS synthesizing gene, quad-GLS; lacking myb28, myb29, cyp79B2, and cyp79B3 with no GLS (Kroymann et al. 2001; Textor et al. 2007; Müller et al. 2010). I grew these four lines in short day condition (25°C, 8L16D, 60% humidity), and utilized them for feeding experiments after 5weeks from germination. In this experiment, I used *P. napi* as a representative species. I collected *P. napi* mother butterflies in Fukushima, Japan, and reared them to adults. Adults were paired by hand pairing and acquired neonates were used for the feeding assay. I followed the same protocol as I used for the feeding experiments with two wild Brassicaceae plants described above. I applied 5 larvae for each mutant individual and replicate this set 4 times (n = 20). I harvested larvae after 120h feeding and measured their weights. 5 larvae from each treatment (4 mutant lines) were randomly chosen and dissected for further expression level analysis. Mid-gut samples were stored in RNA later (QIAGEN) and placed in -80°C until RNA extraction. RNA was extracted with innuPREP RNA Mini Kit (Analytik Jena).

2.2 Expression analysis by RT-qPCR

I designed primers for RT-qPCR analysis as follows (production size = 70-180bp, Tm = 59-61°C, GC% = 40-60%, Max Poly-base = 3) with Primer3Plus for each NSP gene family gene (Steve Rozen & Helen J. Skaletsky 2000). I also designed primers for three common house keeping genes (*EF1a*, *eiF4a*, and *RPS5*) which are frequently used as expression standards in insects (Schweizer et al. 2017). Designed primers are listed in Table 3-1. After confirming quality of RNA by Agilent 2100 Bioanalyzer, I digested gDNA from each extracted RNA samples by using TURBO DNA-free Kit (QIAGEN). I synthesized cDNA with PrimeScript RT reagent Kit with

gDNA Eraser (Perfect Real Time) (TAKARA). I run quantitative real time PCR (RT-qPCR) reaction with CFX Connect Real-Time PCR Detection System (BIO-RAD) using SYBR Premix Ex Taq（Tli RNaseH Plus）. Acquired expression levels were standardized based on house keeping genes expression level. I analyzed relative expression levels of NSP gene families with software Rstudio ver. 1.0.136 (RStudio Team, 2016). I conducted one-way ANOVA and FDR adjusted pairwise t test as ad-hoc analysis to see expression level differences.

2. 3 GLS analysis of A. thaliana mutant lines

I harvested entire rosette of 5 individuals from each mutant line and froze them with liquid nitrogen. The samples were freeze-dried and grounded by metal ball with a shaker. 10mg of grounded leave powder was used for chemical analysis. I added 80% of Methanol with 50 µM of 4-hydroxybenzyl GLS (Sinalbin), which is absent in *A. thaliana*, to each mix as an internal standard. After 5 minutes of incubation with 230 rpm of shaking, I spin-downed the samples with 130,000 rpm for 10 minutes. I added the supernatant to filters conditioned with DEAE sephadex A-25. I washed the filter columns with 500 µl of 80% MeOH and with 1 ml of water two times. After final washing step with 1 ml of MES buffer pH5.2, I added 30 µl sulfate to convert GLS into desulfo GLS and incubated over night at room temperature. I eluted the column with 0.5ml water and analyzed by HPLC-UV with reversed phase C-18 column (Nucleodur Sphinx RP, 250 mm × 4.6 mm, 5 µm, Machrey-Nagel, Düren, Germany). For identification of GLSs, I followed Burow et al. (2006). In brief, desulfo GLS were identified based on the retention time and UV spectra with know standard libraries (Reichelt et al. 2002).

Results

NSP gene family sequence in Pieris butterflies

I got 32-40 million 100 bp paired end reads from each sample and 64,279, 62,054, 59,327, and 53,004 contigs were assembled (*P. napi*, *P. melete*, *P. rapae*, *P. brassicae*) by Trinity. These have 2,048 bp, 2,132 bp, 2,060 bp and 2,594 bp N50 values respectively.

I used tblastn for searching NSP gene family sequences from assembled contigs of each species. *P. rapae* NSP gene family protein sequences were used as queries (GenBank accession number AAR84202, ABY88945, ABY88946) and I made blast database from assembled backbone by "makeblastdb" program in blast. I identified NSP, MA and SDMA sequences from all the four *Pieris* butterflies. The molecular phylogeny of NSP gene family are shown in Fig. 3-1 with reference sequences from *P. rapae*. Newly acquired NSP sequences from *P. napi* and *P. melete* have 86% and 84% amino acid sequence identities with that of *P. rapae*. MA also showed similar similarities (89%) and SDMA showed slightly higher identity (92%) (Table 3-1).

Feeding experiments with two wild Brassicaceae plant species

GLS profiles of both of the plant species were acquired from previous research (Chapter 3). *A. kamchatica* have higher amount of aliphatic GLS with indol GLS, whereas *C. scutata* contains benzyl GLS which is not detected in *A. kamchatica* (Fig. 3-2a). The Growth rates of the four *Pieris* butterflies were acquired in both of the plant species (Fig. 3-2b). The larval condition of all the four species from the two plant species was similar, not as having strongly poisoned or infected. *P. rapae* and *P. brassicae* grew better on *C. scutata* than on *A. kamchatica* ($P \leq 0.05$, t-test,), however, *P. napi* and *P. melete* did not show any difference ($P = 0.08$ and 0.86 each, Fig.

3-2b). I acquired expression levels of NSP gene family from all the four *Pieris* species larvae against two Brassicaceae plants with RNA-seq and RSEM (Fig. 3-2c). The expression levels of NSP were increased when larvae feed on *C. scutat*a than *A. kamchatica*, however inverse patterns were observed in that of MA. These trends were observed in all the four species, whereas expression levels of SDMA were rather similar between the two Brassicaceae plants among all the pairs of four *Pieris* species (Fig. 3-2c). Although the number of replicate of each treatment were too small for statistical analysis ($n = 1$), t-test revealed that the expression levels of NSP and MA are significantly different between the two plant treatments when I treated the four species as replicates ($P \leq 0.05$).

Feeding experiments with the four A. thaliana mutants

I measured the GLS profiles of each mutant by LC-UV (Fig. 3-3a). GLS profiles of Col-0 and quad-GLS mutant were well consistent with previous research, such that Col-0 had higher short chain aliphatic GLSs and quad-GLS mutant did not have any GLSs. In MAM1 mutant, which lacks a part of short chain aliphatic GLS synthesis genes, I detected higher amount of 3-(Methylsulfinyl)propyl GLS (3MSOP) but less 4-(Methylsulfinyl)butyl GLS (4MSOB) comparing to the wild type (Col-0). Both of 3MSPO and 4MSOB are short chain aliphatic GLSs and entire amount of short chain aliphatic GLS did not differ from Col-0, however, the proportion of these GLSs was completely different from that of wild type (Fig. 3-3a). In addition, amount of one long chain aliphatic GLS, 8-(Methylsulfinyl)octyl GLS (8MSOO), was obviously increased from wild type. Regarding MAM3, which losses long chain aliphatic GLS synthesis gene, lacked long chain aliphatic GLSs (8MSOO or 7MSOH; 7-(Methylsulfinyl)heptyl GLS).

Although extremely small peaks were still detected, basically, I could not detect any GLSs in quad-GLS mutant which losses a set of GLS synthesizing genes.

Growth rates of each larva were measured after 120 hours of feeding. One-way ANOVA revealed that the growth rates of *P. napi* larvae were significantly different among some mutant lines. Ad-hoc pairwise t test (FDR adjustment) showed that larvae that fed on MAM1 grew worth than that from Col-0 and MAM3 (Fig. 3-2b).

I measured expression levels of NSP gene family by RT-qPCR and standardized each gene expression levels based on that of house keeping genes. Regarding the expression patterns of NSP gene family, I observed significant regulation difference in NSP and marginal in MA but not in SDMA (Fig. 3-2c). In quad-GLS mutant, I observed NSP was significantly down regulated comparing to that of larvae from Col-0. I also observed MA showed similar down regulation trend in quad-GLS mutant. In MAM1 feeding larvae, NSP was significantly down regulated, whereas MA did not show this trend. The expression level of NSP was rather similar in MAM3 with Col-0, whereas, that of MA was higher as it had significantly higher expression level than that of larvae from quad-GLS mutant.

Discussion

In this research, I conducted feeding experiments utilizing Brassicaceae plants and *Pieris* butterflies coupled with gene expression analysis in order to reveal how *Pieris* butterflies utilize NSP gene families against broad range of GLS profiles. I identified NSP gene family sequences from the four *Pieris* butterflies by RNA-seq. I observed NSP, MA and SDMA genes from all the four species examined here (Fig. 3-1). As previous study revealed, NSP gene family experienced

dynamic evolutionary process in this taxa including frequent gene birth, death or duplicating process (Edger et al. 2015). Indeed, I found several amino acid insertion or deletion between newly acquired NSP or MA and its reference from *P. rapae* even all of the species analyzed are closely related as they are from the same genus. This might suggest interspecific different functionalization of NSP and MA which can associate with their host plant utilizations.

In feeding experiments with two wild Brassicaceae plants, I found that NSP and MA were under dramatically different gene regulations against the two plants. NSP was expressed higher when larvae fed on *C. scutata*, and MA showed quite high expression in larvae which fed on *A. kamchatica* (Fig. 3-2c). Interestingly, this trend was rather common in all the four *Pieris* species. It is known that at least *P. napi* and *melete* can utilize both of these plant species as host in the wild (Ohsaki & Sato 1994), therefore, this expression differences can not be explained as responses against host or non host plants, which can cause strong stress responses. This was also supported by the results of growth rates of the larvae, in which I could not detect any general trend as I saw in expression level measurement, even some species grew significantly lower on *A. kamchatica* (Fig. 3-2b). Assuming background non–GLS chemical differences between the plant species did not affect the expression patterns of NSP gene family, this differential regulation might be explained by GLS profile differences of the two plants. According to a previous research coupled with widely targeted metabolome analysis, GLS profiles of these two Brassicaceae plants showed distinct ones each other (Fig. 3-3a). Although these results were still from a restricted number of detectable GLS, it is likely that they have completely different GLS profiles. The observed dramatic differential regulation between NSP and MA suggests that NSP and MA have at least different functions or roles for dealing with plants which have completely different GLS profiles (Fig. 3-2ac). The regulation pattern might insist that NSP would be

responsible for benzyl GLS which was highly detected in *C. scutata,* and MA, on the other hand, might be responding to aliphatic GLSs. Regarding the activities of SDMA, it seems they were not responding to host plant differences and showing rather general expression. This observation suggests SDMA might have more general functions or roles than that of NSP or MA.

However, since I utilized different plant "species" in this feeding experiment, there should be non-GLS chemical background difference between the two plant species. This difference can strongly affect gene regulations of larvae. In order to test whether observed differential regulations of NSP and MA are consequences of response against different types of GLSs in plants, I utilized *A. thaliana* mutants which have different GLS profiles (Fig. 3-3a). From the feeding experiments with the four mutant lines, I observed differential gene expression patterns in NSP and MA again, but not in SDMA (Fig. 3-3b). I found NSP and MA were down regulated in larvae from quad-GLS mutant which has no GLS (MA was statistically marginal $P = 0.081$). Recent study also reported *P. brassicae* as well down regulated NSP while they were exposed to no GLS containing *A. thaliana* mutant and this supports my findings in *P. napi* (Schweizer et al. 2017). My result shows that NSP and MA responded to GLS content of plants, and suggests that they have some functions which relate to GLS detoxification or digestion in larvae. Although functions of MA have never been confirmed, it is likely that MA can have similar functions as NSP, because NSP and MA are sister genes and have quite similar structure (having 3 repeat domain with around 2kb gene size) (Fischer et al. 2008). Interestingly, my results also suggest the functions of NSP and MA would be different each other, because they showed different regulation patterns against different GLS profiles. I observed NSP was down regulated when larvae fed on MAM1 or quad-GLS but kept its expression level on MAM3 (Fig. 3-3c). This suggests NSP at least did not respond to several GLS which are highly concentrated

in MAM1 such as 3MSOP or 8MSOO, and lower concentrated GLS in MAM3 such as 7MSOH (Fig. 3-3ac). However, this was not the case in MA (Fig. 3-3c). MA kept high expression level among Col-0, MAM1, and MAM3. This suggests MA can respond other types of GLSs that NSP did not respond. Since *A. thaliana* does not contain any benzyl GLS which are highly concentrated in *C. scutata*, I could not imitate the GLS profiles observed in *C. scutata* in *A. thalaiana* mutant lines and could only detect more mild regulation differences among NSP gene family. However, both of the feeding experiments indicate that both of NSP and MA are likely to act as a part of GLS detoxification and their functions are not identical. For SDMA, experiments with mutants also showed the similar patterns as I observed in experiments with two Brassicaceae plants (Fig. 3-2c, 3c). SDMA appears not to respond to GLS profile differences and would have more general functions.

Gene regulations of detoxification related genes in herbivores against different host plants were compared in several previous researches (Heidel-Fischer et al. 2009; Schweizer et al. 2017; de la Paz Celorio-Mancera et al. 2013; Celorio-Mancera et al. 2016). These mainly focused on the differential gene regulation response in generalist herbivores, for understanding the molecular mechanisms which enables them to have wider host ranges. However, gene regulation responses of specialist herbivores against different types of host plants have not been well tested (Schweizer et al. 2017). My research produces an exciting finding that in even specialist herbivores, they used different genes under fine tuned regulation to deal with a broad range of plant chemical defenses. This gives us an important insight to understand how specialist herbivores have evolved its host specificity and changed their host range as an evolutionary consequence.

Pieris butterflies appear to regulate NSP and MA differently to deal with broad range of GLSs (Fig. 3-2c, 3c). Their regulation pattern suggests NSP and MA seem to respond different types of GLSs. This result also gives us an insight which can be a key to understand the evolutionary history of NSP gene family. My results showed that functions of NSP gene family could be changed and differentiated by mutations to act for different types of GLSs. These changes on NSP gene family might enable *Pieris* butterflies utilize and adapt wider range of Brassicaceae plants. NSP gene family is known to be under a dramatic evolutionary process, and this would connect to frequent differentiation of NSP gene family and host range change of pierid butterflies (Edger et al. 2015). For example, I observed NSP or MA was completely down regulated in larvae that fed on a type of Brassicaceae plants even they have a certain amount of GLSs (Fig. 3-2c). This would suggest, these sister genes would be under different selection in case *Pieris* butterfly can only use one of the types of plants as host for a certain evolutionary time and this could cause dramatically different evolutionary consequences of these genes. Actually, in *Anthocharis cardamines*, which is a specialist pierid butterflies of a restricted Brassicaceae plants possess only MA but not NSP (Edger et al. 2015). Further detailed research focusing on substrate specificity or confirming functional differentiation of each genes from different pierid butterflies by heterologous expression would give us convincible insight to understand the evolutionary background of these high diverse gene family and its affect on host range evolution in pierid butterflies.

References

Agrawal, A. a & Kurashige, N.S., 2003. A role for isothiocyanates in plant resistance against the specialist herbivore Pieris rapae. *Journal of chemical ecology*, 29(6), pp.1403–15. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12918924.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10. Available at: http://www.sciencedirect.com/science/article/pii/S0022283605803602.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.

Burow, M. et al., 2006. Altered glucosinolate hydrolysis in genetically engineered Arabidopsis thaliana and its influence on the larval development of Spodoptera littoralis. *Journal of Chemical Ecology*, 32(11), pp.2333–2349.

Celorio-Mancera, M. de la P. et al., 2016. Evolutionary history of host use, rather than plant phylogeny, determines gene expression in a generalist butterfly. *BMC Evolutionary Biology*, 16(1), p.59. Available at: http://www.biomedcentral.com/1471-2148/16/59.

Chew, F.S., 1980. Foodplant Preferences of Pieris Caterpillars (Lepidoptera). *Oecologia*, pp.347–353.

Edger, P.P. et al., 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, p.201503926. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1503926112.

Fahey, J.W., Zalcmann, a T. & Talalay, P., 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), pp.5–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11198818.

Fischer, H.M. et al., 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular biology and evolution*, 25(5), pp.809–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18296701.

Futuyma, D.J. & Agrawal, A.A., 2009. Macroevolution and the biological diversity of plants and herbivores. *October*, 106(43), pp.18054–18061.

Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), pp.644–652.

Halkier, B.A. & Gershenzon, J., 2006. Biology and biochemistry of glucosinolates. *Annual review of plant biology*, 57, pp.303–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16669764.

Heidel-Fischer, H. & Vogel, H., 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, pp.2–8. Available at: http://linkinghub.elsevier.com/retrieve/pii/S221457451500022X.

Heidel-Fischer, H.M. et al., 2010. Microevolutionary dynamics of a macroevolutionary key innovation in a Lepidopteran herbivore. *BMC evolutionary biology*, 10, p.60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2841170&tool=pmcentrez&rendertype=abstract.

Heidel-Fischer, H.M. et al., 2009. Phylogenetic relatedness and host plant growth form influence gene expression of the polyphagous comma butterfly (Polygonia c-album). *BMC genomics*, 10, p.506. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2775755&tool=pmcentrez&rendertype=abstract.

Hofberger, J. a. et al., 2013. Whole Genome and Tandem Duplicate Retention Facilitated Glucosinolate Pathway Diversification in the Mustard Family. *Genome Biology and Evolution*, 5(11), pp.2155–2173. Available at: http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt162.

Kroymann, J. et al., 2001. A Gene Controlling Variation in Arabidopsis Glucosinolate Composition Is Part of the Methionine Chain Elongation Pathway 1. , 127(November), pp.1077–1088.

de la Paz Celorio-Mancera, M. et al., 2013. Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular ecology*, pp.4884–4895. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23952264.

Li, B. & Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), p.323. Available at: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-323.

Müller, R. et al., 2010. Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. *Journal of chemical ecology*, 36(8), pp.905–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20617455.

Ohsaki, N. & Sato, Y., 1994. Food plant choice of Pieris butterflies as a trade-off between parasitoid avoidance and quality of plants. *Ecology*, 75(1), pp.59–68.

Ratzka, A. et al., 2002. Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), pp.11223–11228.

Reichelt, M et al., 2002. Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry* 59, 663–671

Schweizer, F. et al., 2017. Arabidopsis glucosinolates trigger a contrasting transcriptomic response in a generalist and a specialist herbivore. *Insect Biochemistry and Molecular Biology*, 85, pp.21–31.

Tamura, K. et al., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12), pp.2725–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24132122.

Textor, S. et al., 2007. MAM3 Catalyzes the Formation of All Aliphatic Glucosinolate Chain Lengths in Arabidopsis. *Plant Physiology*, 144(1), pp.60–71. Available at: http://www.plantphysiol.org/cgi/doi/10.1104/pp.106.091579.

Vogel, H., Kroymann, J. & Mitchell-Olds, T., 2007. Different transcript patterns in response to specialist and generalist herbivores in the wild Arabidopsis relative Boechera divaricarpa. *PloS one*, 2(10), p.e1081. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2031921&tool=pmcentrez&rendertype=abstract.

Wittstock, U. et al., 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4859–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=387339&tool=pmcentrez&rendertype=abstract.

Wittstock, U. & Burow, M., 2010. Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *The Arabidopsis book / American Society of Plant Biologists*, 8, p.e0134. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3244901&tool=pmcentrez&rendertype=abstract.

Wittstock, U. & Halkier, B. a, 2002. Glucosinolate research in the Arabidopsis era. *Trends in plant science*, 7(6), pp.263–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12049923.

Figures and Tables

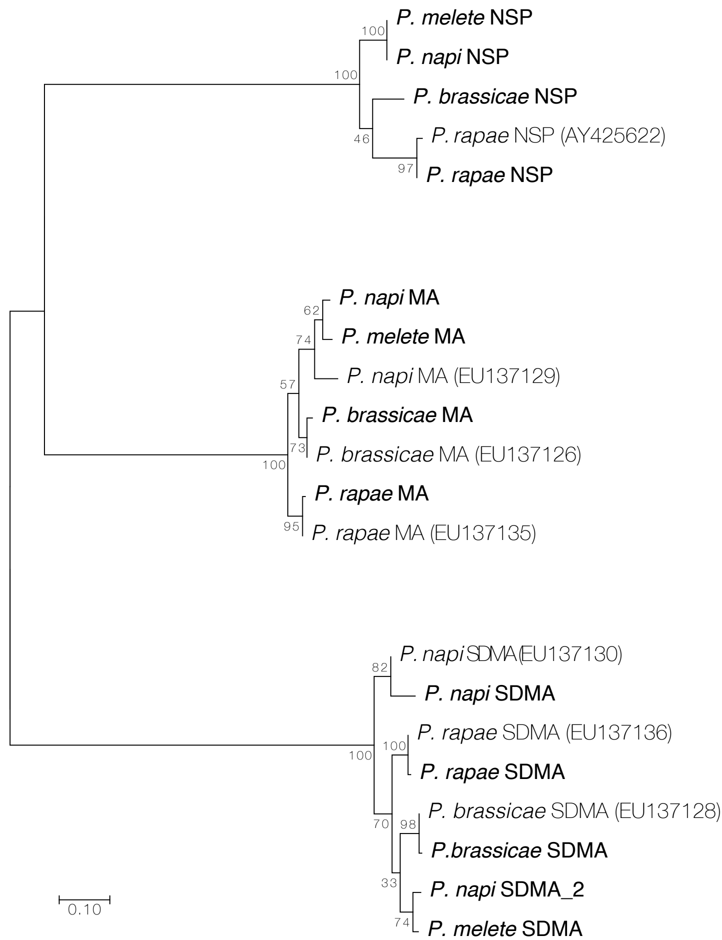
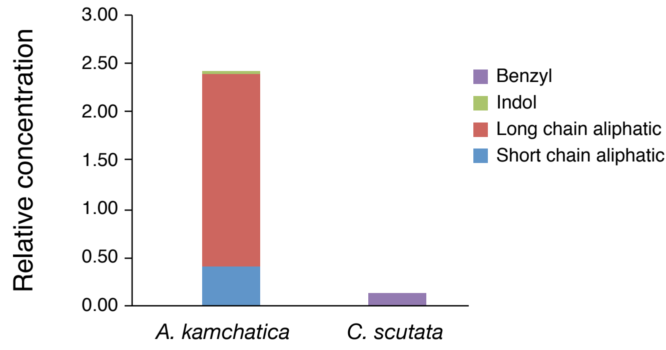
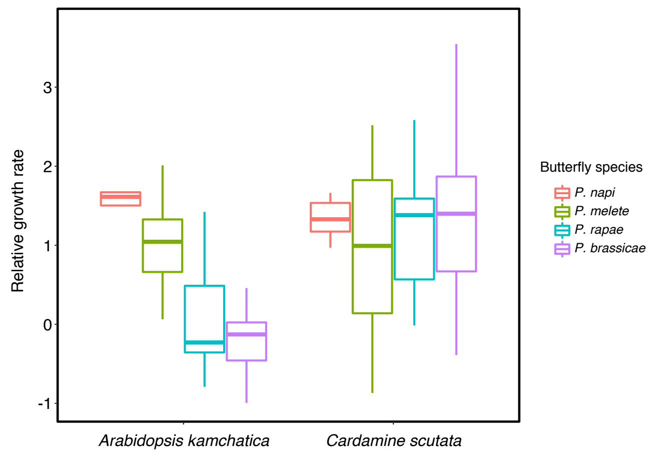


Fig. 3-1 Molecular phylogeny of NSP gene family in *Pieris* butterflies from amino acid sequences. In the present study, all of NSP gene family member (NSP, MA and SDMA) were found in all the four *Pieris* butterflies used.

(a)



(b)



(c)

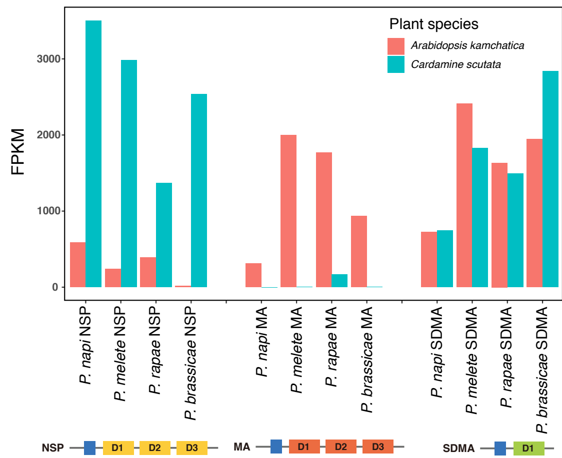
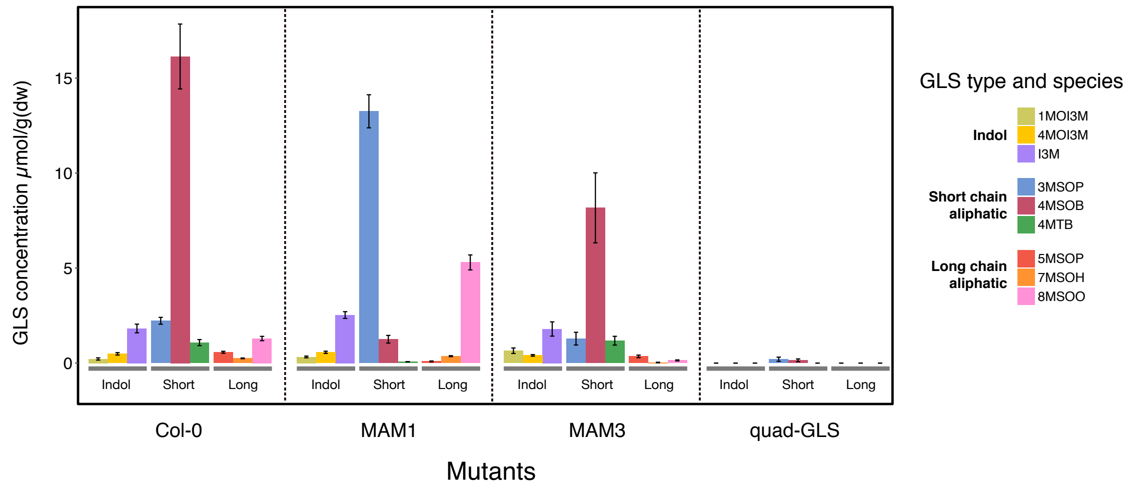
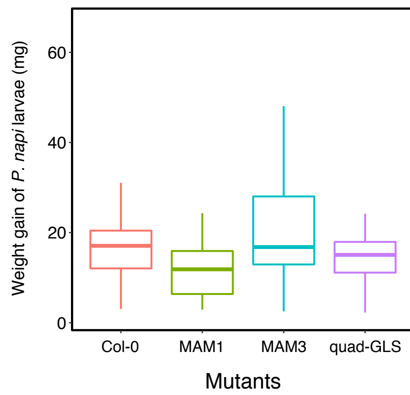


Fig. 3-2 (a) Categorized GLS profiles of *Arabidopsis kamchatica* and *Cardamine scutata* measured by UPLC-TQMS. *Arabidopsis kamchatica* had higher aliphatic GLS concentration and indol GLS, whereas, *C. scutata* had benzyl GLS. (b) Relative growth rates of *Pieris* butterflies against the two plant species from the feeding experiment in Chapter 2. All the four species grew on both of the plants, although growth level differences were observed among the four species. (c) Relative gene expression levels of NSP gene family in *Pieris* butterflies against *Cardamine scutata* and *Arabidopsis kamchatica.* SDMA showed rather equal expression level, however, NSP and MA showed nearly opposite regulation against two different host plants they fed on.

(a)



(b)



(c)

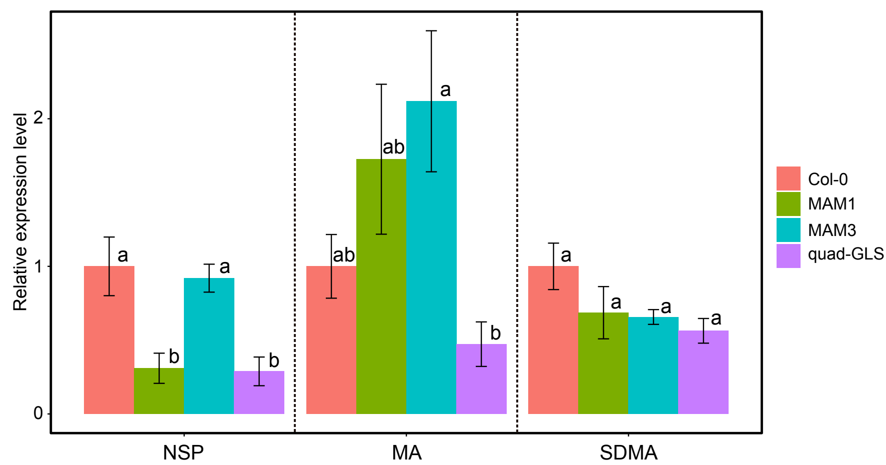


Fig. 3-3 (a) GLS profiles of *A. thaliana* mutants measured by LC-UV (n = 5). Col-0: wild type, MAM1: lacking a part of short chain synthesis genes, MAM3: lacking long chain aliphatic GLS synthesis genes, quad-GLS: lacking GLS synthesis genes. MAM1 showed lower 4MSOB concentration, but accumulate 3MSOP as an alternative. MAM3 has lower long chain aliphatic GLS concentration, and quad-GLS doesn't have any GLSs. No Benzyl GLS was detected from these four lines. (b) Growth rates of *P. napi* that fed on the four *A. thaliana* mutant lines. (c) Relative expression levels of NSP gene family of *Pieris napi* against the four *A. thaliana* mutants (pairwise t test with FDR adjustment). Expression level was standardized based on Col-0. Different letters on each box show significance. (n = 5)

Table 3-1 Primer sequences for RT-qPCR in this research.

Target Gene	Primer Name	Sequences
NSP	P.napi-NSP-RT-F	AATTGGCGGCTTTATACACG
NSP	P.napi-NSP-RT-R	TTCTTTCCTTCGGCACTTGT
MA	P.napi-MA-RT-F	TGTTGCTAACGCACTGGAAG
MA	P.napi-MA-RT-R	CCCTCCAACGCAGTAATGAT
SDMA	P.napi-SDMA-RT-F	CCACGAGCTAAGCGGTAGAG
SDMA	P.napi-SDMA-RT-R	CCATATTTCCTGCCATTCGT
RPS5	ALL-RPS5_F	TTGAGCGCCTTACCAACTCT
RPS5	ALL-RPS5_R	ATCTTCCCGAGGACCAGAAT
EF1α	ALL-EF1a_F	AGGAATTGCGTCGTGGTTAC
EF1α	ALL-EF1a_R	GCAAGCAATGTGAGCTGTGT

Table 3-2

Sequence identity of NSP gene family among the four *Pieris* butterflies. Upper right shows nucleotide identity and lower left shows amino acid identity. (a) NSP, (b) MA, and (c) SDMA.

(a)

	P. napi	*P. melete*	*P. rapae*	*P. brassicae*
*P. napi*_NSP		0.996	0.84	0.875
*P. melete*_NSP	0.998		0.838	0.873
*P. rapae*_NSP	0.894	0.894		0.909
*P. brassicae*_NSP	0.904	0.903	0.876	

(b)

	P. napi	*P. melete*	*P. rapae*	*P. brassicae*
*P. napi*_MA		0.985	0.887	0.913
*P. melete*_MA	0.978		0.892	0.917
*P. rapae*_MA	0.907	0.906		0.911
*P. brassicae*_MA	0.92	0.92	0.903	

(c)

	P. napi	*P. melete*	*P. rapae*	*P. brassicae*
*P. napi*_SDMA		0.977	0.922	0.925
*P. melete*_SDMA	0.976		0.922	0.93
*P. rapae*_SDMA	0.931	0.936		0.917
*P. brassicae*_SDMA	0.929	0.933	0.92	

Chapter 4

Microevolution of NSP in *Pieris* butterflies in Japan

Abstract

Herbivores are exposed to plant chemical defenses of their host, and the chemical profile is specific to each host plant. Differential host plant utilization of herbivores can lead speciation in terms of ecological speciation, since herbivores need to deal with different chemical environment of the novel host plants. Understanding microevolutionary dynamics of detoxification traits of herbivores to different host plants would shed light on the mechanisms of the initial phase of herbivore speciation associated with different host utilization. However, despite several detoxification mechanisms of herbivores against plant chemical defense have been revealed in several insect taxa, its microevolution and ecological consequences in the field associated with host plant difference is unclear. Here, I focused on three *Pieris* butterflies which possess specific larval gut expressed enzyme (NSP) to deal with their Brassicaceae host plant chemical defense (glucosinolates). I sampled larvae from populations which are exposed to different Brassicaceae plant community. I compared the observed plant diversity and nucleotide diversity of NSPs in each population for the three species, and found a positive correlation among them for the two of the three species. Furthermore, purifying selection was observed on the gene in populations which were exposed to lower plant diversity. Analysis of genome wide genetic diversity for each sample showed no correlation with plant community diversity, supporting observed correlation on the gene was not from genetic background. These results clearly suggest that utilizing narrow host plant range consequently give purifying selection on NSPs, on the other hand, populations with higher potential host plant species can maintain diversity

of NSPs. This would give us an important support that differential host use in the field can cause different genetic consequences of adaptation traits of herbivores which might be connected with the initial phase of ecological speciation.

Introduction

Genetic diversity accumulated in a species can lead speciation as a consequence. In herbivorous insects, speciation can be mediated by host utilization difference in a species followed by divergent selection (Futuyma & Agrawal 2009). Since phytophagous insects generally have strong interaction with their hosts, physical and chemical environments that they are exposed to can be dramatically changed depending on which host plants they utilize (Matsubayashi et al. 2010). Recently, a certain number of researches examined whether adaptation to different host plants can cause reproductive isolation (Fujiyama et al. 2013; Matsubayashi et al. 2013; Powell et al. 2014). In some cases, genetic structure of herbivore associated with host plant utilization difference of populations were observed (Forbes et al. 2017). For example, in *Timema* stick insects, populations which use different host plant changed their body color and reproductive isolation have also been found between the populations (Nosil et al. 2002; Riesch et al. 2017). In this context, different host utilization can cause divergent selection on adaptation traits or gene, which can finally cause reproductive isolation. However, the genetic basis or evolutionary consequences of these adaptation traits associated with different host utilization in herbivores are still unclear.

Plants generally defend themselves by several types of defense strategies. Chemical defenses involving plant secondary metabolites can be a primary defense among them as its specificity and toxicity could be so diverged in the plant kingdom (Futuyma & Agrawal 2009). Each plants species has their own specialized chemical defenses and herbivores need to disarm these for using them as host plants by acquiring specific adaptation traits. Until recently, genetic bases of several detoxification

mechanisms against plant secondary metabolites in herbivores have been revealed and some of them are known as key innovations, which enable herbivores to utilize novel types of hosts and to experience explosive diversification (Heidel-Fischer & Vogel 2015; Wheat et al. 2007).

In case differential host use initiate speciation in herbivores, evolution of detoxification mechanisms against plant defense can be involved in the speciation (Matsubayashi et al. 2010). However, it still remains unknown whether adapting to different host plants causes different microevolutional consequences on these detoxification traits or genes. Moreover, evidence of natural selection in the wild especially on these detoxification mechanisms against plant chemical defenses has never been observed. Understanding microevolutional dynamics of detoxification genes in herbivores to different host plant would shed a light on the mechanisms of the initial phase of herbivore speciation responding to different chemical defenses from different host plant (Heidel-Fischer et al. 2010). Here, I focused on pierid butterfly's glucosinolate detoxification genes and its microevolutionary consequences along with different host plant community, in order to reveal whether evolution on the gene respond to different host plant utilization.

Brassicaceae and pierid butterfly interaction has been used as one of models for understanding arms-race between plant defense and herbivore adaptation (Edger et al. 2015; Fischer et al. 2008; Wheat et al. 2007). Brassicaceae plants have diverged glucosinolate (>140 species) as a strong defense against herbivores, whereas a part of pierid butterflies overcome this defense by acquiring gut expressed protein called nitrile specifier proteins (NSPs) (Wittstock & Halkier 2002; Wittstock et al. 2004). NSP is known as a key innovation for adapting to brassicales plants, and pierid butterflies

experience rapid speciation event mediated by the arms-race between NSP evolution and glucosinolate diversification (Edger et al. 2015; Fischer et al. 2008; Wheat et al. 2007). The dynamic evolution of NSPs in pierid family was observed in previous research, however, the microevolution of NSPs responding to different host utilization have never been tested in the wild (Fischer et al. 2008; Edger et al. 2015). A population genetic work on NSP was conducted with small cabbage white (*Pieris rapae*), however, this only focused on selection on NSPs among different continent and could not detect any consistent natural selection on NSPs (Heidel-Fischer et al. 2010). *P. rapae* mainly relies on Brassica crops as hosts in all over the world (Cameron & Walker 2002; Oh et al. 2013; Hasan & Ansari 2011), therefore, it would be difficult to detect selection on NSPs since they are supposed to be exposed more uniform host plant community in every population.

Here I focus on Japanese two *Pieris* butterfly species, *P. melete* and *P. napi* (Fig. 4-1ab), and evaluate microevolution of NSPs exposed to different host plant communities. Since both of the two species use Brassicaceae weeds as host plants, the potential host plant community are supposed to be more ununiformed even in entire Japan (Ohsaki & Sato 1994). Therefore, these two species would be useful for testing NSP evolution than *P. rapae*. I also use *P. rapae* as a control and compare microevolutionary consequences of NSPs to evaluate whether completely different form of host utilization (crop-dependent and weed dependent) affects NSP microevolutionary dynamics.

In the present study, I compare NSP diversity of Pieris butterfly populations with host plant diversity they exposed to, in order to reveal whether different host utilization could cause different microevolutional consequences on this detoxification

gene. I sample the three *Pieris* species from sampling sites in all over Japan for NSP sequencing. These sites are supposed to have different Brassicaceae plant community and I collect larva with host plant utilization data for each species. I decide sampling sites by Brassicaceae plant community diversity estimation based on herbarium data in Japan, to cover a wider range of diversity of Brassicaceae plants. I also measure observed host plant diversity in each sampling sites and compared them with NSP diversity of the three species in each sampling sites. For resolving genetic background diversity and its demographic history of the three *Pieri*s species in each population, genome-wide single-nucleotide polymorphisms (SNPs) are obtained through restriction site associated (RAD) sequencing. I use this data to evaluate NSP diversity in each sampling site by comparing them with genome based genetic diversity. These data would give us concrete evidence to understand microevolutionary dynamics of NSPs when they are exposed to an environment with a different host plant community.

Materials and Methods

Pieris **butterflies used in this study**

I used three *Pieris* butterfly species in this study, such that *Pieris melete*, *napi* and *rapae*. All of these three *Pieris* species can be basically observed in all over Japan. *Pieris melete* and *napi* are sister species and resemble as it is difficult to fully distinguish one from the other by their morphological trait (Fig. 4-1 ab). However, adult males of *P. melete* and *P. napi* are distinguishable because of the dramatically different

shape of androconium (scent scale) between the species (Fig. 4-1 cd) (Fujimori 2012). Both of them generally use wild Brassicaceae weeds such as *Cardamine* and *Arabis* as hosts, however, their host preferences in field are known to be slightly different from each other (Ohsaki & Sato 1994). Both of the species appear to prefer mountain sites, but *P. melete* can also be observed in lowlands or urban sites. Regarding *Pieris napi* group in Japan, there are controversially two species; *Pieris nesis* and *Pieris dulcinea* (Fig. 4-2). Although, these two were grouped as *P. napi* and recently identified as different species based on ND5 sequence, it is still controversial and also nearly impossible to distinguish them completely by their morphology even from the shape of androconium (Shirouzu 2006). They appear to have hybrid zone in Hokkaido and *Pieris dulcinea* can only be found in the eastern part of this hybrid zone, whereas *Pieris nesis* are thought to live from west of this hybrid zone to the rest part of Japan (Fig. 4-2) (Shirouzu 2006). In this study, I handled them as subspecies of *P. napi*, and covered the east side and west side of this hybrid zone in the sampling sites (Fig. 4-2). *P. rapae* is known to be one of the main pests of Brassicaceae crops and they mainly rely on plants of genus *Brassica* which includes cabbage or oil seed rapa (Cameron & Walker 2002; Oh et al. 2013; Hasan & Ansari 2011). *P. rapae* can be observed from low to high lands in Japan, and its habitat range seems to be bigger than the other two species in Japan (Shirouzu 2006).

Brassicaceae plant community diversity estimation with Maxent

I estimated Brassicaceae plant community diversity by Maxent ver. 3.4.1 distribution modeling (Phillips et al. 2017; Phillips et al. 2004). The collection data with locality information of Japanese Brassicaceae plants were gathered from Global Biodiversity

Information Facility (GBIF) and S-Net (http://science-net.kahaku.go.jp/). I modified acquired data into genus level and locality data were converted into latitude and longitude format. In total, 11325 individual data were collected from 44 Brassicaceae genera. I also collected climate data of Japan from WorldClim (Fick & Hijmans 2017) with 19 variables and estimated potential distribution range of each Brassicaceae genus in Japan by Maxent. Predicted probabilities of presence of each genus were treated as potential densities of each genus (0 to 1 present probability are treated as estimated individual number of each genus) and Shannon diversity index was calculated as estimated Brassicaceae community diversity by R package "vegan" (R Core Team 2015; Oksanen et al. 2017). I selected 7 to 8 sampling sites for the three butterfly species respectively from all over Japan based on this estimated host plant diversity, covering sites which have varied estimated diversity index. Four of the sampling sites are shared by all the three species.

Sampling and measurement of Brassicaceae plant community diversity at sampling sites

I conducted field sampling on April to August in 2017. In each sampling sites, I did transect method to find Brassicaceae plants along with the sampling path, which is at least 2 km long each. All of the three species lay eggs individually on their host plants. Wild larvae or eggs were collected from host plants at sampling sites. I recorded the number and species of Brassicaceae plants and numbers of eggs or larvae of *Pieris* butterfly on each plant species. To minimize the possibility to collect sibling larvae or eggs, I collected samples from at least 8 host plant batches in each sampling sites (except *P. rapae* from Fukushima with only two batches).

Sampled larvae bigger than 3rd instar were dissected for gut RNA extraction. Eggs or small larvae were reared by feeding on their host plants until around 3rd instar for dissection. Dissected larval guts were immediately processed with RNA later (QIAGEN) and stored in -80°C until RNA extraction. The rest bodies were also stored in -80°C as samples for gDNA extraction.

Community diversity of Brassicaceae plants in each sampling sites were measured based on the recorded individual number of each Brassicaceae plant species. I excluded 2 Brassicaceae plant genus (*Capsella* and *Erysimum*) from these sampling, since it is known that *Pieris* butterfly larvae cannot use these genus as hosts. In addition, Brassica crops (cabbage or broccoli) which are basic host plants of *P. rapae* were found in all the sampling sites except Okinawa (south Japan). I also did not include this data, since these crops can be found nearly every sampling site and also cannot be correctly measured in all the sampling sites. I calculated Shannon diversity index for each sampling site based on the collected data as described above.

Pieris butterfly species identification

Since larvae of *Pieris* butterfly species are impossible to be morphologically identified, I used PCR-RFLP method for the identification. I used mitochondrion ND5 region for species identification following the sequence data in GenBank (LC090587- LC090590). The dissected rest bodies were used as PCR templates and I amplified ND5 region directly from the templates with MightyAmp DNA Polymerase Ver.3 (Takara). For PCR, ND5 universal primers were used (: V1, 5'-CCTGTTTCTGCTTTAGTTCA-3'; A1, 5'-AATATDAGGTATAAATCATAT-3';). The amplified ND5 PCR products were processed by several restriction enzymes (Fig. 4-3). For identifying *P. melete*, I used

HincII, and HinfI for *P. napi nesis*, and HindIII was used for identifying *P. napi dulcinea*. Larvae which did not have any digested ND5 PCR product were identified as *P. rapae*. I identified species based on loading patterns of digested PCR products on 2% agarose gel (TAE). I also confirmed this identification methodology by using male adult samples, which can be easily identified with androconium (Fig. 4-1cd), combined with ND5 sequencing and confirmed that the identification rate was 100 % for 64 individuals (24 *P. melete*, 16 *P. napi*, and 24 *P. rapae*) of the three species.

RNA extraction, cDNA synthesis, NSP amplification, cloning and sequencing

The identified larvae specimens were served for the RNA sequences. I randomly choose 10 larvae from each sampling site for all the three species respectively. In total, I selected 70 larvae (7 sampling sites) for *P. melete* and *napi* respectively, and 78 larvae for *P. rapae* (10 larvae from 7 sampling sites and 8 larvae from Yonaguni island). I extracted RNA from dissected larval gut samples stored in RNA later. I used RNeasy Mini Kit (QIAGEN) for extracting RNA. cDNA was synthesized by ReverTra Ace qPCR RT Master Mix (TOYOBO). I used TaKaRa Ex Taq (Takara) for amplifying NSP from each cDNA samples and amplified PCR products were gel purified with Favorgen GEL/PCR Purification Mini Kit (Favorgen). PCR was done with primer sets; 5’-ATGAAAGCTGTTGTAGTCTTATTAGC-3’ and 5’-CTGTCCGTAAAGAGCAG GTAC-3’ for *P. melete* and *napi*, 5’-ATGAAAGGTGTTGTAGTCTTCTTAG-3’ and 5’-TTACTGTCCGTAAAGGGCA-3’ for *P. rapae*. Purified PCR products were diluted and used for cloning reaction. For cloning reaction, I used Mighty TA-cloning Kit (Takara). Colonies were selected with ampicillin and blue-white selection. I picked white colonies and confirmed inserted fragment length with colony PCR by using

EmeraldAmp MAX PCR Master Mix (Takara) and M13 primers. Colonies having plasmid with around 2kb insertion size were gown in LB medium with ampicillin, and plasmids were extracted and purified with NucleoSpin Plasmid EasyPure (Takara) after harvesting over-night culture. Finally, 2 plasmids were prepared for each 218 larval individual samples respectively. I did cycle-sequencing reaction with M13 and 2 primer sets for each plasmid samples with BigDye Terminator Cycle Sequencing kits (Applied Biosystems) and inserted NSPs were sequenced by ABI 3730xl DNA Analyzer (Applied Biosystems). Used primer sequences were 5'-GCTTAGATGCCTTGTCAAAGACT-3' and 5'-AATAGCGTGGTCGTTCTTAGC -3' for P. *melete* and *napi*, and 5'-CTCTGGAAGAACGAAGCATT-3' and 5'-AACTCGGCTAGTCCTGCTTTC-3' for *P. rapae*. Acquired reads were trimmed and aligned with MEGA6 (Tamura et al. 2013) and used for further analysis.

One representative NSP full sequence per sample individual was randomly selected and used for genetic analysis. I made molecular phylogeny of NSP sequence with ML method in nucleotide and amino acid level. I used MEGA6 for constructing the tree combined with 500 bootstrap tests. I also conducted principal coordinate analysis (PCoA) on NSP sequences in nucleotide and amino acid scale. For calculating population genetics statistics, such that nucleotide diversity (π), F_{st} or Tajima's *D* (Tajima 1989), I used Arlequin ver. 3.5 (Excoffier & Lischer 2010). For estimating amino acid divergence of NSPs in each population, I calculated mean pairwise amino acid difference of NSPs in each population by MEGA.

NSP type identification from gDNA in *P. melete*

From NSP sequencing, apparently two types of NSP were found in *P. melete* and the frequencies of each type were different among sampling sites (see results). In this study, I sequenced NSPs from mRNA, therefore, observed biased proportion of NSP types can be generated by expression bias on two different gene copies. Therefore, I additionally conducted PCR-RFLP to identify whether these two types of NSPs were two copies or two alleles. These two types of NSP had 11 fixed SNPs in between, and 9 of them were observed in the first exon (1-252bp). I designed primers on the first exon to be able to amplify 228 bp fragments on this first exon (5'-GTCTTGTACTTCGGACTCCTTTT-3'). I used extracted *P. melete* gDNA from all the 70 individulas that I used for NSP sequencing as PCR templates, and PCR was done with EmeraldAmp MAX PCR Master Mix (Takara). Then, I utilized HinCII restriction enzyme. HinCII can distinguish these two types of NSP and only digest only one of the two types of NSP generating about 180 bp fragments. After 5 hours of digestion, I loaded the digested PCR products on 3% TBE agarose gel for 20 minutes and checked observed band pattern. If these two observed types of NSP were two gene copies, each individual would have two bands. On the other hand, if these were two alleles, each individual was supposed to have homo- or heterozygous band pattern depending on their genotype.

RAD sequencing and analysis

For RAD sequencing, I used dissected rest bodies of larva, which was used for RNA extraction for NSP sequencing, as gDNA samples. I used the same larvae that I utilized for RNA extraction and NSP sequencing. gDNA was extracted with Maxwell 16 LEV Plant DNA Kit (Promega). Extracted DNA was quantified with Qubit 2 Fluorometer

(Invitrogen). In total, 218 samples in total were used for RAD-seq library preparation. I used EcoR1 as restriction enzyme for RAD-seq library preparation and 218 samples were run in 1 lane in HiSeq2500 (Illumina). Acquired reads were trimmed with trimmomatic with following options (ILLUMINACLIP:2:10:10 TRAILING:20 SLIDINGWINDOW:4:15 MINLEN:30) (Bolger et al. 2014). Samples with less than 500,000 reads were excluded from further analysis. SNPs were called by Stacks ver. 1.48 (Catchen 2013). For ustacks, I set n = 3 and M = 3 option, and for cstacks, n = 3 was set and I did this analysis not only for each species independently but also for multi species scales; 3 species (*melete* + *napi* + *rapae*) or closely related 2 species (*melete* + *napi*). For population analysis, I used "populations" in Stacks and set parameter as p = 4, r = 0.75 for species analysis, and set p = 1, r = 0.85 in multiple species scales without involving population information. I used LOSITAN (Antao et al. 2008) for excluding SNPs under selections, and population structure estimations were done with Structure ver. 2.3.4 (Falush et al. 2007) as setting burn in = 100,000 and replicates as 500,000 times after burn in, from K = 1 to 10 with 5 times iteration. PCoA analysis was also performed to compare the result with that of NSP sequences. PCoA was conducted with R based on gower's distance matrix.

Comparison of NSP diversity in populations and Brassicaceae plant community diversity in each sampling sites

In order to reveal microevolutionary dynamics of NSPs in the three *Pieris* butterflies potentially responding to the host plant diversity, I compared NSP nucleotide diversity (π) and whole genome associated nucleotide diversity (from RAD-seq) with estimated plant community diversity in each population by linear regression. In addition, NSP-

RAD diversity contrast was used to evaluate the correlation of NSP and plant diversity excluding back ground genetic diversity. The contrast was calculated as differences of scaled diversity of both NSP (π) and RAD genetic diversity. All of these statistical comparisons were conducted in R.

Results

Brassicaceae plant diversity estimation and sampling

Diversity estimation based on Maxent showed Brassicaceae community diversity would be higher in Honshu area in Japan, and tend to be lower in northern and southern area (Fig. 4-4a). Based on this result, I decided 11 sampling sites from Hokkaido to southern area which can potentially cover broader Brassicaceae community diversity.

At the 11 sampling sites, I collected 4777 individual Brassicaceae plants from 25 species of 14 genera (Fig. 4-4a). Brassicaceae plant diversity was measured from this data set for each sampling site (Table 4-1). Measured Brassicaceae plant diversity also insisted lower diversity in Hokkaido and was significantly correlated with the Maxent diversity estimation (Fig. 4-4b, $P = 0.0019$).

Larvae sampling and NSP sequencing

In total, 945 larvae from 11 sampling sties were collected (Table 4-1). The species of these larvae were identified by PCR-RFLP. I identified 483 individuals of *P. melete*, 253 of *P. napi* and 209 of *P. rapae* larvae. The butterfly–host plant association of *P. melete* and *P. napi* in each sampling sites were shown in Fig. 4-5. *P. melete* tend to use

Cardamine leucantha in northern part of Japan, whereas they use *Rorippa indica* in southern area. *P. napi* in Hokkaido populations hardly rely on *Rorippa sylvestris*, however, populations from Honshu use *Arabis hirsuta* or *Arabis flagellosa* as main hosts. *P. rapae* also use *R. sylvestris* in Hokkaido, whereas this species strongly relies on *Brassica* crops (mainly cabbage, broccoli, kale or oil seed rapa) and a number of larvae and eggs were found in crop field in each sampling sites except Okinawa (eggs and larvae of *P. rapae* are morphologically distinguishable from the other two closely related species). I sequenced NSPs from each larva, and finally got 1870 bp of NSP sequences from 214 individuals after excluding low quality sequences.

The acquired molecular phylogenies of NSP based on nucleotide sequences and amino acid sequences are shown in Fig. 4-6. The nucleotide phylogeny supported the species clustering, however that of amino acid sequence was inconsistent with the species phylogeny. In the phylogeny from nucleotide sequences, *P. rapae* represents a monophyletic clade and *P. napi* made a species clade as well, whereas that of *P. melete* was not the case (Fig. 4-6a). Furthermore, *P. napi* clade was split into two clade; samples from Honshu or samples from Hokkaido. Similar structure was also found in *P. melete*, such that, one clade was composed only by the members from Honshu and the other group was mainly occupied by Hokkaido samples. In amino acid sequence phylogeny, the species clades were collapsed both in *P. napi* and *P. melete* clade (Fig. 4-6b). *P. rapae* clade was still fixed, however, *P. napi* Hokkaido clade was included in *P. melete* clade. Furthermore, *P. napi* Hokkaido clade made a clade with *P. melete* Hokkaido clade, although bootstrap support value was low.

PCoA revealed similar trends as the phylogenetic analysis on NSP sequences from the three species. PCoA divided *P. rapae* NSPs from *P. melete* and *P. napi* (Fig.

4-7a). NSPs from *P. napi* Honshu clade also showed differences from that of Hokkaido clade in both nucleotide and amino-acid scale (Fig. 4-7 ab). PCoA on nucleotide NSP sequences focusing on the two closely related species (*P. melete* and *P. napi*) showed that both species had 2 types of NSPs but its distribution patterns among population were different between the two species (Fig. 4-7c). *P. melete* had two NSP types; one was composed only by Honshu samples (type A), and the other type included samples from both Honshu and Hokkaido (type B). However, two observed NSP types from *P. napi* were clearly split between Honshu and Hokkaido populations. *P. napi* from Honshu had more distant NSPs from the other, and that from Hokkaido had closer NSPs to *P. melete* type B NSP based on this analysis. In amino acid sequence level, *P. napi* NSPs from Hokkaido were completely combined with *P. melete* type B NSP group, which was composed by *P. melete* from both Honshu and Hokkaido (Fig. 4-7d).

The calculated population genetic statistics with Arlequin based on NSP sequences and amino acid diversity by MEGA were shown in Table 4-2. For both of the *P. melete* and *P. napi*, Tajima's *D* of NSPs in Hokkaido populations were significantly lower than zero. In *P. rapae*, populations from Yubari (Hokkaido), and Nagano (Honshu) also showed significantly low Tajima's *D* values. Observed nucleotide diversity (π) of NSP in Hokkaido populations were apparently lower in both of *P. melete* and *P. napi*, but this was not the case in *P. rapae.* Regarding amino acid diversity, this was basically correlate to nucleotide diversity, however, *P. napi* population from Miyazaki showed lower amino acid diversity comparing to the other populations.

Two apparent types of NSPs were found in *P. melete*; type A which is only observed Honshu, and type B which is dominant in Hokkaido but can be also found in

Honshu (Fig. 4-8a). According to PCR-RFLP identification, each sample showed one of the three types of band pattern, such that one band with type A size, two bands with type A and B size, one band with type B size (Fig. 4-8b). Given these band pattern showed allelic homo- heterozygote band pattern, frequency of each types of NSP in each site was in Table 4-3, and these band patterns correspond to observed types from mRNA full-length sequences. Moreover, Hokkaido *P. melete* populations did not have type A NSPs.

RAD-seq analysis

On average, 1,062,550 reads were acquired for each sample in RAD-seq analysis. 183 samples were used for population genetics analysis after excluding samples with low read counts. Stacks called 4821, 3697, and 4334 SNPs for *P. melete*, *P. napi*, and *P. rapae* respectively. 543, and 4587 SNPs were also found in the analysis including all the three species or closely related two species (*P. melete and P. napi*). Population statistics calculated by "populations" in Stacks were shown in Table 4-4. Genome wide associated nucleotide diversity was higher in *P. rapae* and lower in *P. melete* (*P. melete*; π_{mean} = 0.068, *P. napi*; π_{mean} = 0.093, *P. rapae*; π_{mean} = 0.109). Mean F_{st} in each species pair calculated by shared SNPs among all the three species were as follows; F_{st} *melete-napi* = 0.380, F_{st} *melete-rapae* = 0.821, F_{st} *napi-rapae* = 0.856. F_{st} from closely related two species were also separately analyzed in Stacks and compared by clustering analysis and it showed samples from Hokkaido made cluster in both of the species (Table 4-5).

LOSITAN excluded 30 to 41 % of SNPs as outliers and 2798 (*P. melete*), 2296 (*P. napi*), 3010 (*P. rapae*), 347 (3 species), and 2844 SNPs (*P. melete* + *P. napi*)

were used for populations structure analysis. STRUCTURE showed that the optimal clustering numbers were $K = 3$ (*P. melete*), 2 (*P. napi*), and 2 (*P. rapae*) in species scale (Fig. 4-9, 10 abc). In *P. melete*, STRUCTURE showed weak distinction from north to south, whereas that of *P. napi* was apparent. *P. napi* apparently showed genetic structure between Hokkaido and Honshu populations, however, this trend was inconsistence with ND5 species identification, which indicated subspecies border in the middle of Hokkaido (Fig. 4-10b). *P. rapae* result indicated population in Fukushima was slightly diverged from the other when $K = 2$. STRUCTURE analysis in the closely related two species (*P. melete* and *P. napi* together) showed apparent distinction between *P. melete* and *P. napi*. In addition, this result also suggested *P. napi* samples from Hokkaido showed different population structure from that of Honshu (Fig. 4-9d).

PCoA on RAD-seq from 3 species data clearly showed species grouping (data not shown). PCoA on *P. melete* did not show apparent cluster which was observed in STRUCTURE analysis (Fig. 4-11ab). On the other hand, *P. napi* samples from Hokkaido were distant from the rest in PCoA and this supported the genetic structure observed in STRUCTURE analysis (Fig. 4-11cd). Furthermore, PCoA showed samples from Rusutsu (Southern Hokkaido) are closer to Honshu populations than that of Yubari (more northern population) (Fig. 4-11d). Regarding *P. rapae*, the observed genetic structure in STRUCTURE analysis was not supported in PCoA, which only showed samples from south most population was slightly distant from the other (Fig. 4-11ef).

Comparison between NSP diversity and Brassicaceae community diversity

NSP nucleotide diversity of each species from each sampling site was compared with measured Brassicaceae community diversity by linear regression (Fig. 4-12ab). In *P.*

melete and *P. napi*, NSP diversity showed significant positive correlation with plant diversity (*P. melete*: P = 0.0234, *P. napi*: P = 0.0282), whereas this trend was not significant in *P. rapae* (P = 0.627). Amino acid diversity of NSP also correlated with host plant diversity in *P. melete* and *P. napi* but not in *P. rapae* (*P. melete*: P = 0.0170, *P. napi*: P = 0.0166, *P. rapae*: P = 0.850). On the other hand, genome wide associated genetic diversity from RAD-seq was also compared, but no significant correlation was found with plant diversity in all the three species (Fig. 4-12c). Both of the genetic diversity (NSP and RAD) were also compared with Maxent Brassicaceae community diversity inference, but no significant correlation was observed (data not shown). In addition to these, calculated NSP – genome wide nucleotide diversity contrast was compared with plant diversity by linear regression as well (Fig. 4-12d). *P. melete* and *P. napi* again showed positive correlation with measured plant diversity (*P. melete* P = 0.0346, *P. napi*: P = 0.0328), whereas no correlation was found in *P. rapae* (P = 0.624) (Fig. 4-12c). As a general trend, both of NSP diversity and plant community diversity in Hokkaido population was lower in each found correlation (Fig. 4-12abd).

Discussion

Brassicaceae plant community diversity in Japan and NSP diversity in *Pieris* butterflies

The result of Brassicaceae diversity estimation applying Maxent genus distribution probability inference showed that Brassicaceae community diversity would be higher in center part of Honshu area in Japan, whereas it would be lower in Hokkaido and

southern area (Fig. 4-4). The global diversity pattern of Brassicaceae showed that center of Brassicaceae diversity is in northern hemisphere (Irano-Turanian region) and becomes lower in tropic as they can only be found in mountainous and alpine regions in tropic area (Lysak & Koch 2011). Although Maxent only used genus data, the Maxent estimation was along with this general distribution trend in Brassicaceae and showed lower diversity in south most islands in Japan.

The collected host plant data from the fields was correlated with Maxent estimation and showed lower diversity in Hokkaido or south regions (Fig. 4-4ab). This field sampling was done only in a restricted seasonal period (April to August) and restricted regions, therefore, this cannot cover all the Brassicaceae species diversity in each sampling site. However, observed Brassicaceae species diversity cline from the field sampling was not only correlate with Maxent estimation but also consistent with general plant diversity cline found in Japan, which showed plant diversity is higher in center of Honshu and lower in Hokkaido region (Kubota et al. 2015). Therefore, observed Brassicaceae species diversity appeared to reflect actual Brassicaceae diversity in each sampling site even this was restricted in several aspects.

Three *Pieris* species and its host utilizations

In this study, I focused on three *Pieris* species in Japan, *P. melete*, *P. napi* and *P. rapae*. Although *P. melete* and *P. napi* are resemble and cannot be distinguished fully by their morphology in some cases (Fig. 4-1ab, Shirouzu 2006), the result of RAD-seq showed it seems these three species can be treated as different species, since they showed higher F_{st} values among each species pair. Furthermore, RAD-seq analysis indicated *P. napi* have two genetically distant groups which are inconsistence with ND5 species

identification (Shirouzu 2006). From ND5 identification, Japanese *P. napi* can be divided into *P. nesis* and *P. dulcinea* at the border lying on middle of Hokkaido, so called the Ishikari depression (Shirouzu 2006, Fig. 4-10b). However, RAD-seq analysis indicated that most of the genetic background of south population in Hokkaido (Rusutsu) was also have similar genetic structure as that of northern population (Yubari) (Fig. 4-10b). Discordance between mitochondria and nuclear genetic structure can be led by hybridization event in the past (Bernal et al. 2017). In this study, although I used RAD-seq data which included both nuclear and mitochondrion genome, I still could find this discrepancy. Furthermore, I also observed that southern Hokkaido population (Rusutsu) was closer to Honshu population than that of northern population (Yubari) in PCoA (Fig. 4-11cd), and these may indicate hybridization event between Hokkaido and Honshu population in the past in *P. napi*.

Similar to the genetic structure found in *P. napi, P. melete* also showed similar trend from north to south although this was not strong enough to be detected in PCoA as apparent groupings (Fig. 4-10a, 11ab). The exciting result that both of the closely related species had similar genetic structure in Japanese archipelago (Fig. 4-10ab), indicates these two species experienced similar biogeographic event in the past. These different genetic patterns between Hokkaido and Honshu were also observed in different organisms in Japan (Tsuda et al. 2015). Although further data and analysis are required, it might be consequences of secondary contact of northern and southern populations in both of the species at the same geological event. In spite of this interesting genetic structure in these two species, *P. rapae* did not show any apparent genetic structure. Optimal K from STRUCTURE analysis (K = 2) was not supported from PCoA result in *P. rapae* and it seems they did not have strong genetic divergence

in Japan (Fig. 4-10c, 11ef). This result is somehow plausible because *P. rapae* is known to be one of cabbage crop pests and likely to have higher dispersal rates because of human activity. For instance, this species was recently introduced to north America from Eurasia continent (Scudder 1889, Keeler et al. 2006).

Host utilization of Japanese butterflies have been studied for a while (Muto-Fujita et al. 2017), however, the data was fragmented and potentially include species miss-identification especially in *Pieris* butterflies, since young instar larvae of *Pieris* are nearly impossible to be distinguished from their morphology. I collected larvae from the wild and also recorded host plant species at the same time. Furthermore, I also identified species by PCR-RFLP based on ND5 sequences. Previous study insisted that *P. melete* and *P. napi* both used Brassicaceae plants as hosts but their host utilizations were slightly different (Ohsaki & Sato 1994). According to the host utilization data from 736 individuals of *P. melete* and *napi*, their host utilizations are different in each other (Fig. 4-5). For example, in southern part of Japan, *P. napi* mainly used *Arabis*, whereas *P. melete* utilized *Rorippa* plants. Moreover, their host use is also different from north to south even in one species, although this would be along with plant community difference (Fig. 4-5). Regarding *P. rapae*, I also observed slight host plant utilization difference in each sampling sites, however, they mainly used cabbage crops as hosts in most of the sampling sites except Okinawa.

Nucleotide diversity of NSPs along with Brassicaceae plant community diversity

I compared NSP sequence diversity with observed Brassicaceae plant community diversity in each sampling site, in order to test microevolutionary dynamics of adaptation gene of *Pieris* butterflies to its Brassicaceae host plant. Surprisingly, NSP

nucleotide diversity and amino acid diversity was correlated with observed Brassicaceae plant community diversity in both of *P. melete* and *P. napi* (Fig. 4-12ab). In addition, genome wide genetic diversity did not correlate with this observed plant diversity index in the two species (Fig. 4-12c). This indicates NSP diversity can respond to host plant diversity and this was not from genetic background diversity of each population. Furthermore, I also found populations of both of the species in Hokkaido, where lower host plant diversity was observed, had NSPs with significantly lower Tajima's *D* value than zero (Table 4-2). Negative Tajima's *D* value insists recent bottleneck event or purifying selection (Tajima 1989, Van Belleghem 2015). Based on the observations in RAD-seq results in which lower genome wide genetic diversity was not observed even in Hokkaido population, this may insists NSPs in Hokkaido populations are under purifying selection. According to the host utilization analysis, *P. melete* and *P. napi* appear to use one plant species dominantly in Hokkaido (Fig. 4-5). These results suggest that diversity of NSPs is maintained on a certain level in populations that use various host plants, whereas its diversity is dramatically decreased in population that use only a few plant species as host because of purifying selection to optimize narrow range of host plants.

On the contrary to these, I could not find any significant correlation between NSP diversity and plant community diversity in *P. rapae* (Fig. 4-12ab). As I observed in each sampling site, *P. rapae* strongly rely on Brassica cops in Japan, and Brassicaceae plant community difference in each site might not reflect to host plant utilization of *P. rapae*. Therefore it is plausible that the NSP sequence diversity did not correlate to observed Brassicaceae plant community diversity in *P. rapae*, since they use rather uniform hosts in Japan. In previous study, which focused on NSP

microevolutionary dynamics in *P. rapae* in continent level, also could not find any evidence of local adaptation, purifying or divergent selection on NSPs (Heidel-Fischer et al. 2010). The results of this present study was not in contradiction with the previous study. However, we also observed significant negative Tajima's *D* value in two *P. rapae* populations as well (Table 4-2). Although these two populations did not show lower NSP nucleotide diversity comparing to the other populations, further analysis including local crop utilization data would be a help to understand selection on NSP in *P. rapae*.

The potential factors and mechanisms that cause NSP microevolution in *P. melete* and *P. napi*

In this study, I found lower nucleotide diversity of NSPs in *P. melete* and *P. napi* Hokkaido populations along with lower host plant community diversity.

NSP molecular phylogeny in nucleotide sequence supported species clades, on the other hand, that of amino acid sequences made one Hokkaido clade including both *P. melete* and *P. napi* samples and inconsistent with species clades (Fig. 4-6ab). The inconsistence between the species phylogeny and the molecular phylogeny of the gene sometimes can be caused by convergence event on the gene (Li et al. 2008, Liu et al 2010). Although support of this Hokkaido clade was quite low, this clade might show an aspect of NSP convergence in Hokkaido population between these two species. This potential NSP convergence was also supported by the result of PCoA. NSP PCoA in nucleotide sequences showed difference between *P. melete* and *P. napi* in Hokkaido populations, however, these could not be distinguished in PCoA with amino acid sequence (Fig. 4-7cd). This suggests these two NSPs from *P. melete* and *P. napi* in

Hokkaido population have quite similar functions but both of them were experienced different evolutional history which is associated with their nucleotide sequence differences. The results of Tajima's *D* test also supported this convergent scenario, since Tajima's *D* test suggested NSPs from these Hokkaido populations are under purifying selection. This might mean NSP of both of the species have been under strong purifying selection and had similar protein structure as a consequence of convergence. It is interesting that this trend was not observed in Honshu populations which have been exposed to more diverged host plant communities. This result would also supported an idea that NSPs lose its diversity if it was exposed to lower host plant diversity even in interspecies scales, however, can be diverged when they encounter to a certain levels of host plant diversity.

I also found apparent two types of NSP in *P. melete* (Fig. 4-6ab, 8a). One type (namely Type A) can only be found in population from Honshu area, and the other types (namely Type B) can be observed in both of Hokkaido and Honshu. In Hokkaido, only type B NSP were found (Fig. 4-8a). The nucleotide diversity cline in *P. melete* NSP was generated by this unequal distribution of these two types of NSPs, such that Hokkaido has lower π since they expressed only type B NSPs. In this study I sequenced cDNA from mRNA, therefore, the sequence results can be affected by regulations of gene expression. In other words, the observed different distribution patterns of NSP types between Hokkaido and Honshu can be shaped only by expression regulation difference on two gene copies not by different alleles. In this case, the evolutionary scenario of NSPs can be more complex and difficult to understand. However, PCR-RFLP did not support this scenario and strongly suggested these two types of NSPs were alleles (Fig. 4-8). Interestingly, these two types of NSPs showed frequency cline

from north to south (Fig. 4-8b, Table 4-3). Type B NSP is fixed in Hokkaido and type A can be found most in south Japan (Nara and Tokushima)(Fig. 4-8, Table 4-3). Considering the results of RAD-seq analysis, it is likely these two alleles are originated from two diverged ancestral populations. Given RAD-seq results showed secondary contact of northern and southern populations in Japan, this asymmetric proportion of two NSP alleles from north to south would suggest these two population have their own NSP types and the secondary contact triggered let them mix in Japan. It is still unclear whether these different NSP alleles can have different function, however, it would be interesting that this ununiformed allele frequency in Japan can be maintained by its functional differences and host plant association. In other words, it should be tested whether Hokkaido fixed NSP type (type B) is more adaptive to *C. leucantha*, dominant host in Hokkaido, and type A is more adaptive to *R. indica* or its inverse situation considering its host association (Fig. 4-5). This would give us more concrete insight for understanding microevolution of detoxification enzyme and its potential effect on speciation.

Conclusion

The present research revealed that herbivorous adaptation key gene which enable them to utilize its host plants can respond to the host plant community diversity. Furthermore, purifying selection observed in lower host plant diversity area, which may also occur convergent evolution on the genes between two closely related species. Until recently, several key innovations for plant adaptation in herbivore insects were found in several phytophagous insects. For example, cytochrome P450 family was revealed as detoxification mechanism in *Papilio* butterflies against *Apeaceae* containing

furanocoumarin (Berenbaum et al. 1996; Li et al. 2003; Wen et al. 2003), and Glucosinolate Sulfatase (GSS) was known as another glucosinolate detoxification mechanisms in diamond back moth (Ratzka et al. 2002). These genes appear to be under selection and respond sensitively to its host plant chemical composition, otherwise herbivores can not track evolution of plant chemical defenses in terms of plant-herbivore arms-race (Futuyma & Agrawal 2009). However, the microevolutionary dynamics of these genes in the wild was not tested and difficult to be tracked, since still a number of detoxification mechanism in herbivores are unclear and hard to be identified (Heidel-Fischer et al. 2015). In this respect, this study give us the first glimpse to understand how these genes respond to host plant community they are exposed to. At least in two *Pieris* butterflies in Japan, NSP were supposed to be under purifying selection when the butterfly species use low number of host plant species and its diversity was maintained with higher host plant diversity. This indicates, it would be plausible that ecological speciation triggered by differential host utilization can generate different evolutionary consequences on the adaptation genes, as higher divergence of NSP was observed in pierid butterflies in previous research (Edger et al. 2015). In this study, I prepared three species sets; *P. rapae* has no genetic structure in Japan, *P. melete* which have slight diverged structure between Hokkaido and Honshu populations, and *P. napi* has stronger genetic border at the Honshu-Hokkaido boundary. Interestingly, the host utilizations of these species also correlate to this genetic structure (Fig. 4-5). Although it is difficult to decide cause or effect between this genetic structure and host association (Bagley et al. 2017, Orsini et al. 2013), our results suggest host use can at least give selection on adaptive genes. If mechanisms how *P. melete* type B NSP is fixed in Hokkaido population is revealed, it would give us to insightful evidence to

justify whether different host use can generate adaptation gene evolution and successive entire genetic structure change in the end, and can cause ecological speciation in phytophagous insect.

References

Antao, T. et al., 2008. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, 9, pp.1–5.

Bagley, R.K. et al., 2017. History, geography and host use shape genomewide patterns of genetic variation in the redheaded pine sawfly (Neodiprion lecontei). Molecular Ecology, 26(4), pp.1022–1044.

Berenbaum, M.R., Favret, C. & Schuler, M. a., 1996. On Defining "Key Innovations" in an Adaptive Radiation: Cytochrome P450S and Papilionidae. The American Naturalist, 148(May), p.S139. Available at: http://www.jstor.org/stable/2463053.

Bernal, M.A. et al., 2017. Introgression and selection shaped the evolutionary history of sympatric sister-species of coral reef fishes (genus: Haemulon). *Molecular Ecology*, 26(2), pp.639–652.

Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–2120.

Cameron, P.J. & Walker, G.P., 2002. Field evaluation of Cotesia rubecula (Hymenoptera: Braconidae), an introduced parasitoid of Pieris rapae (Lepidoptera: Pieridae) in New Zealand. *Environmental Entomology*, 31(2), pp.367–374. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-0036557217&partnerID=40&md5=2785c31153ead5362744dfc3e7983654.

Catchen, J.M., 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11), pp.3124–3140.

Edger, P.P. et al., 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, p.201503926. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1503926112.

Excoffier, L. & Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), pp.564–567.

Falush, D., Stephens, M. & Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), pp.574–578.

Fick, S.E. & Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), pp.4302–4315.

Fischer, H.M. et al., 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular biology and evolution*, 25(5), pp.809–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18296701.

Forbes, A.A. et al., 2017. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, 71(5), pp.1126–1137.

Fujimori, S. 2012. Androconium. Japanese *Pieris* butterflies. Yuho-sha. (in Japanese)

Fujiyama, N. et al., 2013. Distribution and Differentiation of Henosepilachna diekei (Coleoptera: Coccinellidae) on Two Host-Plant Species Across Java,. *Entomological Society of America*, 106(6), pp.741–752.

Futuyma, D.J. & Agrawal, A.A., 2009. Macroevolution and the biological diversity of plants and herbivores. *October*, 106(43), pp.18054–18061.

Hasan, F. & Ansari, M.S., 2011. Effects of different brassicaceous host plants on the fitness of Pieris brassicae (L.). *Crop Protection*, 30(7), pp.854–862. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0261219411000676.

Heidel-Fischer, H.M. et al., 2010. Microevolutionary dynamics of a macroevolutionary key innovation in a Lepidopteran herbivore. *BMC evolutionary biology*, 10, p.60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2841170&tool=pmcentrez&rendertype=abstract.

Heidel-Fischer, H.M. & Vogel, H., 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, pp.8–14. Available at: http://dx.doi.org/10.1016/j.cois.2015.02.004.

Keeler, M.S. et al., 2006. Modelling the impacts of two exotic invasive species on a native butterfly: Top-down vs. bottom-up effects. Journal of Animal Ecology, 75(3), pp.777–788.

Kubota, Y., Shiono, T. & Kusumoto, B., 2015. Role of climate and geohistorical factors in driving plant richness patterns and endemicity on the east Asian continental islands. *Ecography*, 38(6), pp.639–648.

Li, G. et al., 2008. The hearing gene Prestin reunites echolocating bats. Proceedings of the National Academy of Sciences, 105(37), pp.13959–13964. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0802097105.

Li, W., Schuler, M. a & Berenbaum, M.R., 2003. Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity

and substrate encounter rate. *Proceedings of the National Academy of Sciences of the United States of America*, 100 Suppl, pp.14593–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=304124&tool=pmcentrez&rendertype=abstract.

Liu, Y. et al., 2010. Congergent sequence evlution between echolocating bats and dolphins. Current Biology, 20(2), pp.R53–R54. Available at: http://ac.els-cdn.com/S0960982209020739/1-s2.0-S0960982209020739-main.pdf?_tid=e5d30652-be58-11e4-8b40-00000aacb35d&acdnat=1425025155_a3c25fe34270a1a72e3f0b0a93e8786a.

Lysak, M.A. & Koch, M.A., 2011. Phylogeny, Genome, and Karyotype Evolution of Crucifers (Brassicaceae) BT - Genetics and Genomics of the Brassicaceae. In R. Schmidt & I. Bancroft, eds. New York, NY: Springer New York, pp. 1–31. Available at: https://doi.org/10.1007/978-1-4419-7118-0_1.

Matsubayashi, K.E.I.W., Kahono, S.I.H. & Katakura, H., 2013. Divergent host plant preference causes assortative mating between sympatric host races of the ladybird beetle , Henosepilachna diekei. , pp.606–614.

Matsubayashi, K.W., Ohshima, I. & Nosil, P., 2010. Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, 134(1), pp.1–27.

Muto-Fujita, A. et al., 2017. Data integration AIDS understanding of butterfly-host plant networks. Scientific Reports, 7(January), pp.1–14.

Nosil, P., Crespi, B.J. & Sandoval, C.P., 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*, 417(6887), pp.440–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12024213.

Oh, S. et al., 2013. Molecular and immunohistochemical characterization of granulin gene encoded in Pieris rapae granulovirus genome. *Journal of invertebrate pathology*, 113(1), pp.7–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23295682.

Ohsaki, N. & Sato, Y., 1994. Food plant choice of Pieris butterflies as a trade-off between parasitoid avoidance and quality of plants. *Ecology*, 75(1), pp.59–68.

Oksanen, J. et al., 2017. Community Ecology Package, Package " vegan ."

Orsini, L. et al., 2013. Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular Ecology, 22(24), pp.5983–5999.

Phillips, S., Dudík, M. & Schapire, R., 2004. A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, pp.655–662. Available at: http://dl.acm.org/citation.cfm?id=1015412.

Phillips, S.J. et al., 2017. Opening the black box: an open-source release of Maxent. *Ecography*, 40(7), pp.887–893.

Powell, T.H.Q. et al., 2014. Ecological adaptation and reproductive isolation in sympatry: Genetic and phenotypic evidence for native host races of Rhagoletis pomonella. *Molecular Ecology*, 23(3), pp.688–704.

Ratzka, A. et al., 2002. Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), pp.11223–11228.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Riesch, R. et al., 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology and Evolution*, 1(4).

Scudder, S.H. (1889) The Butterflies of the Eastern United States and Canada. Houghton, Mifflin, Boston.

Shirouzu, T. 2006. Standard book of Japanese butterflies. Gakken. (in Japanese).

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123(3), pp.585–595.

Tamura, K. et al., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12), pp.2725–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24132122.

Tsuda, Y. et al., 2015. The population demography of Betula maximowicziana, a cool-temperate tree species in Japan, in relation to the last glacial period: Its admixture-like genetic structure is the result of simple population splitting not admixing. Molecular Ecology, 24(7), pp.1403–1418.

Van Belleghem, S.M., Roelofs, D. & Hendrickx, F., 2015. Evolutionary history of a dispersal-associated locus across sympatric and allopatric divergent populations of a wing-polymorphic beetle across Atlantic Europe. Molecular Ecology, 24(4), pp.890–908.

Wen, Z. et al., 2003. Metabolism of linear and angular furanocoumarins by Papilio polyxenes CYP6B1 co-expressed with NADPH cytochrome P450 reductase. *Insect Biochemistry and Molecular Biology*, 33(9), pp.937–947.

Wheat, C.W. et al., 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(51), pp.20427–31. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2154447&tool=pmcentrez&rendertype=abstract.

Wittstock, U. et al., 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4859–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=387339&tool=pmcentrez&rendertype=abstract.

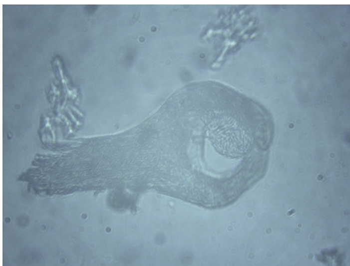
Wittstock, U. & Halkier, B. a, 2002. Glucosinolate research in the Arabidopsis era. *Trends in plant science*, 7(6), pp.263–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12049923.

Figures and Tables

(a)

(b)

(c)



(d)

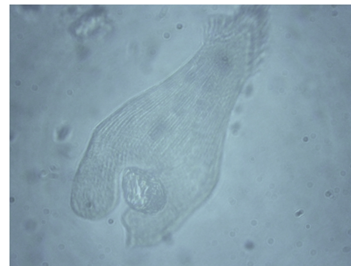


Fig. 4-1 (a) *Pieris melete* male spring form, (b) *Pieris napi* (*nesis*) male spring form, (c) Typical shape of androconium of *Pieris melete*, (d) Typical shape of androconium of *Pieris napi* In Brief, *P. melete* has bigger scent bag than *P. napi.*

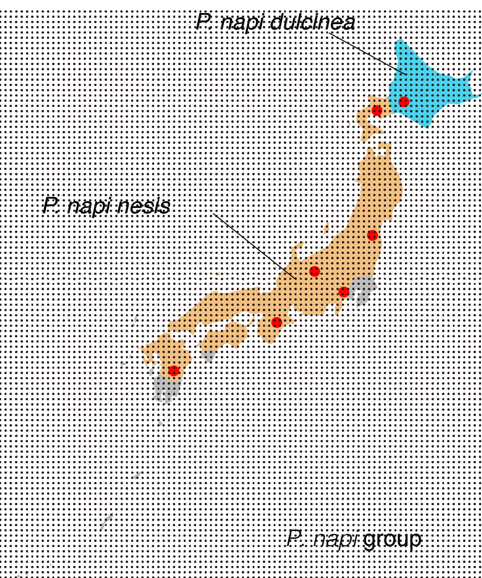


Fig. 4-2 Distribution of *P. napi* group in Japan. Two controversial species (*P. napi dulcinea* and *P. napi nesis*) exist in this group and they have species border at the middle of Hokkaido, which is revealed based on ND5 sequences. The red points indicate sampling sites in this study.

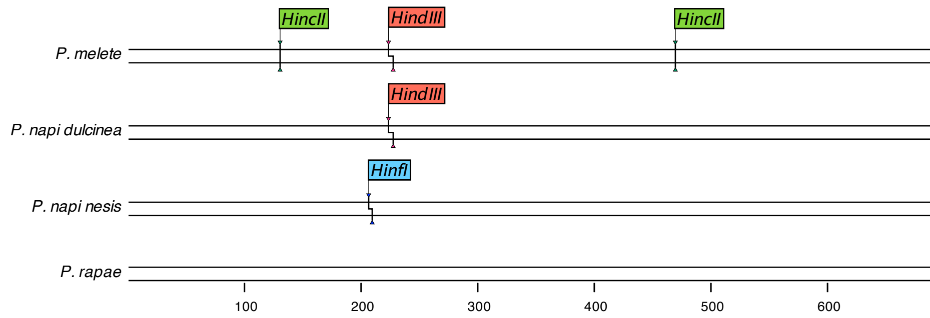
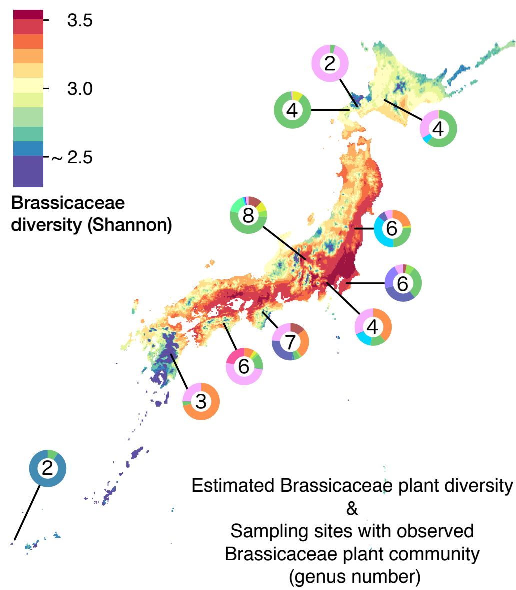


Fig. 4-3 PCR-RFLP on ND5 region was used for species identification in the three *Pieris* species. Digestion sites of each restriction enzyme are shown for each species.

(a)



(b)

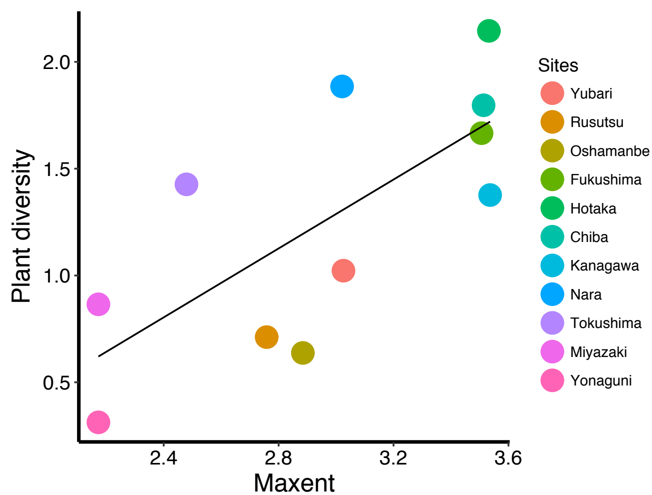


Fig. 4-4 (a) Estimated Brassicaceae plant community diversity applying Maxent distribution modeling and actual sampling sites with observed Brassicaceae genus composition. The number in each sampling sites shows observed number of Brassicaceae genus in each site. (b) Correlation between estimated Maxent diversity and

observed diversity of Brassicaceae plant community at 11 sampling sites (P = 0.0119 by Pearson's correlation test).

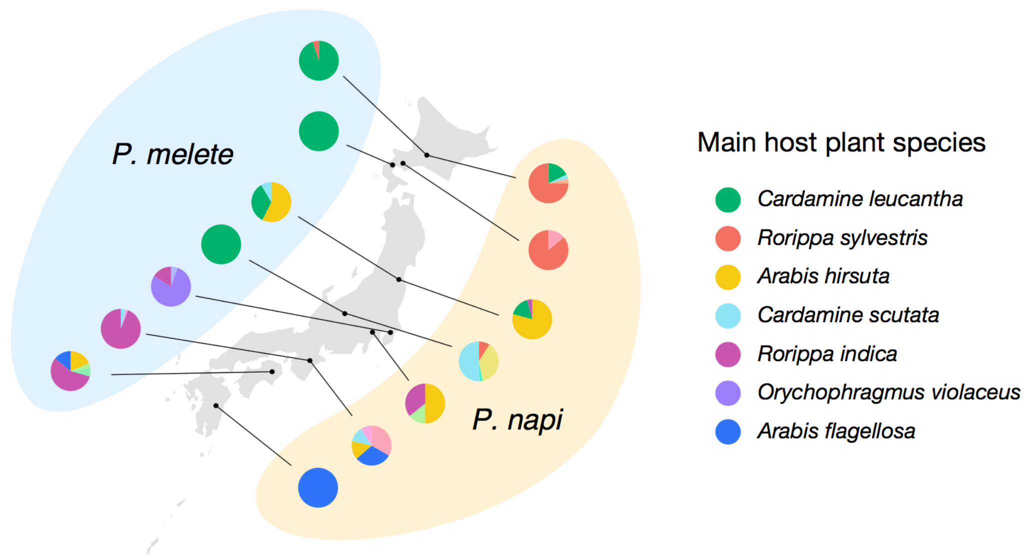
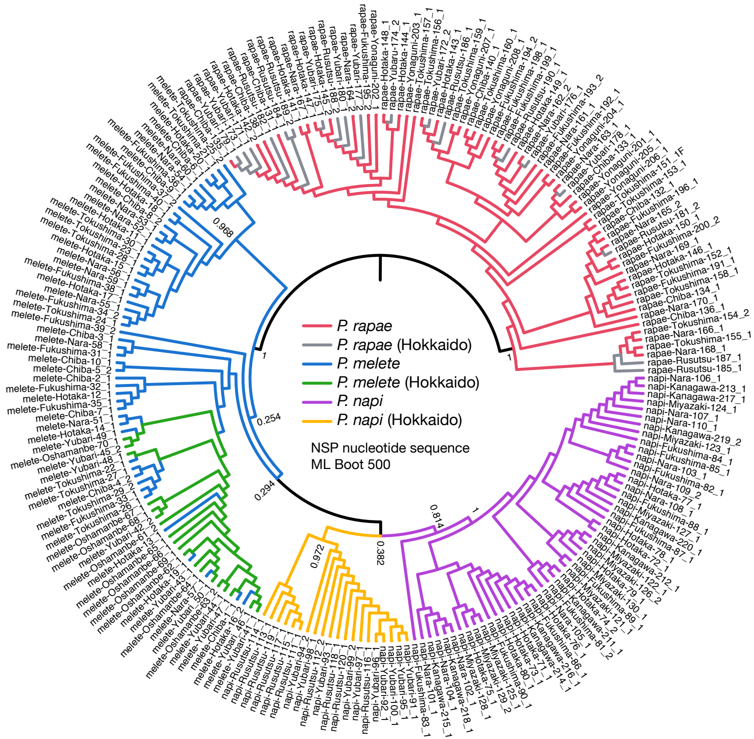


Fig. 4-5 Herbivore and host plant association in each sampling sites regarding two closely related *Pieris* species. These two species tend to use different plan even in the same sites. Moreover, host plant utilizations are different between Hokkaido and Honshu populations even in one species.

(a)



(b)

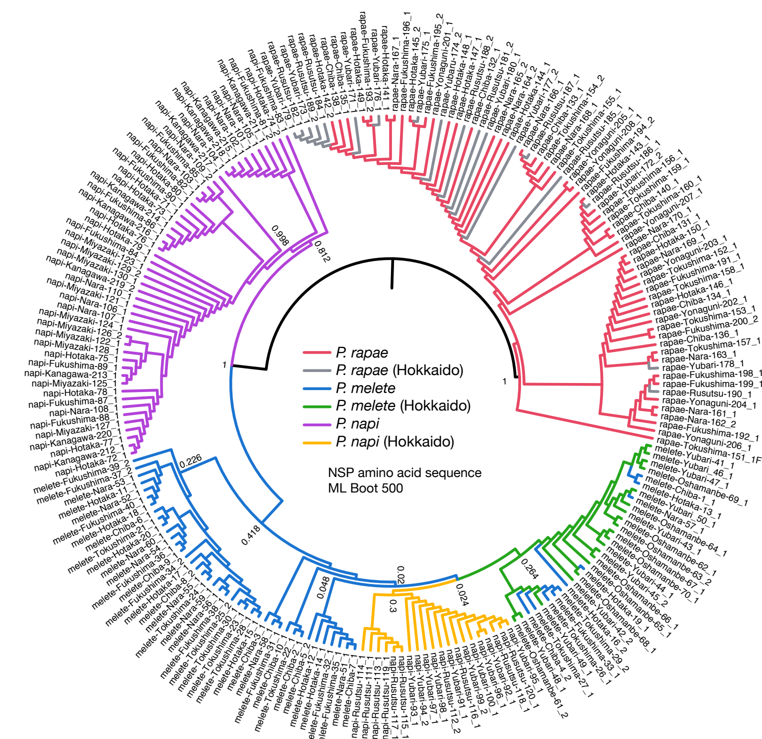


Fig. 4-6 (a) ML molecular phylogenetic tree of NSP nucleotide sequences from three *Pieris* butterflies species (214 individuals). The numbers on branches are bootstrap values based on 500 times reputation. The tree topology contradicts with species phylogeny of these three species. NSPs of *P. napi* show clear differentiation between Hokkaido and populations from the rest of Japan. That of *P. melete* also shows similar trend, whereas *P. rapae* does not show the trend.

(b) ML molecular phylogeny of NSP amino acid sequences from the 214 individuals. The values on each node show bootstrap values from 500 reputations. Amino acid phylogeny shows NSP is apparently different between *P. rapae* and the rest of the two species (*P. melete* and *P. napi*). NSPs from *P. napi* can be distinguishable among Hokkaido and rest of the sites, further more, they lose species clade topology in amino acid level. Furthermore, NSPs of *P. melete* and *P. napi* make a clade, although they have lower support and also contain some NSPs from *P. melete* collected outside Hokkaido.

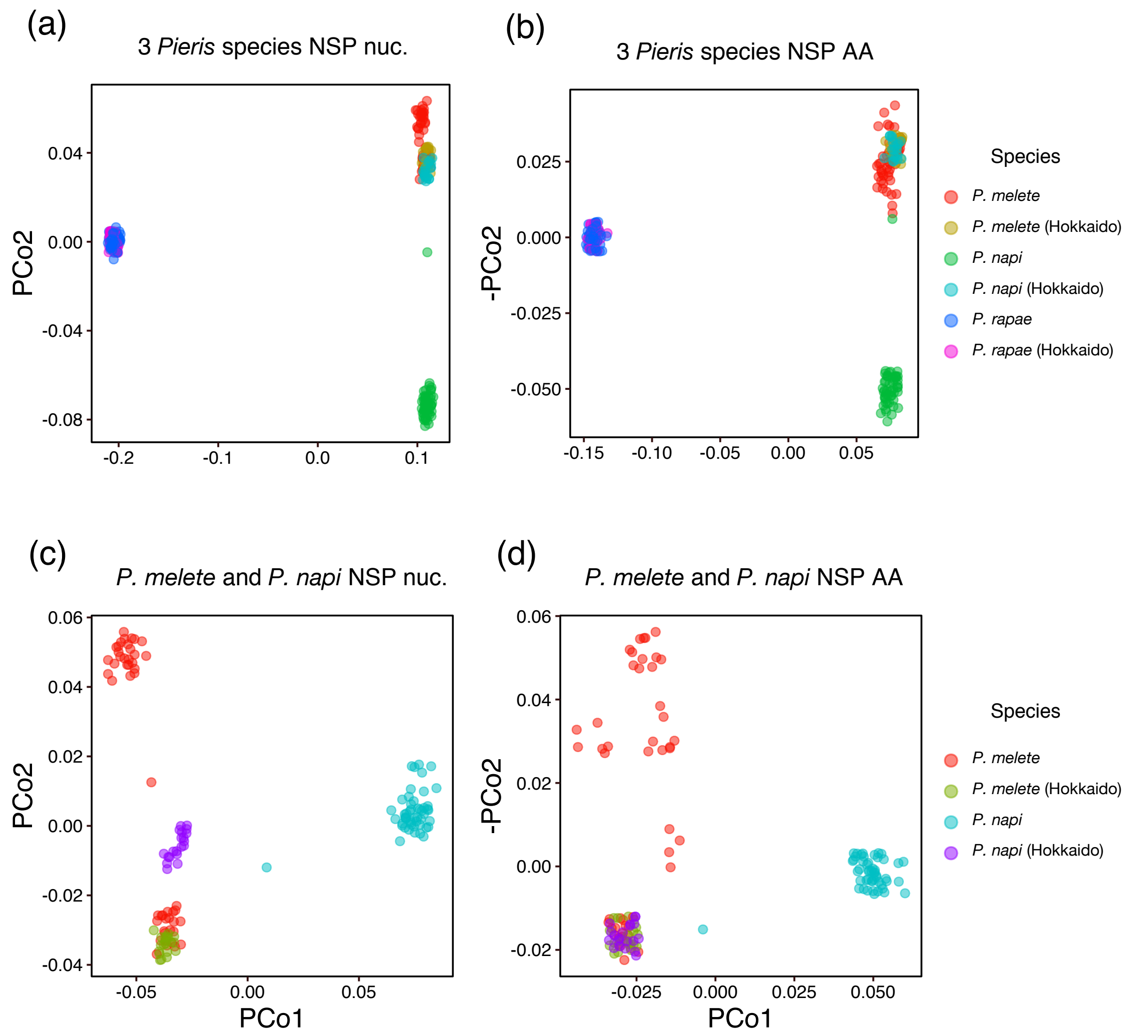
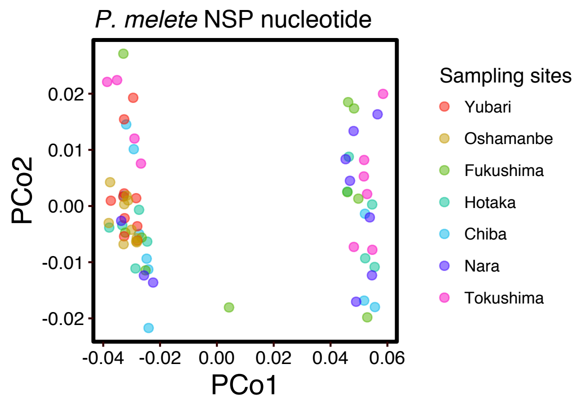


Fig. 4-7 PCoA on NSP sequences. (a)PCoA on nucleotide sequences from the three species, (b)PCoA on amino acid sequences from the three species, (c)PCoA on nucleotide sequences from *P. melete* and *P. napi*, (d)PCoA on amino acid sequences from the *P. melete* and *P. napi*.

(a)



(b)

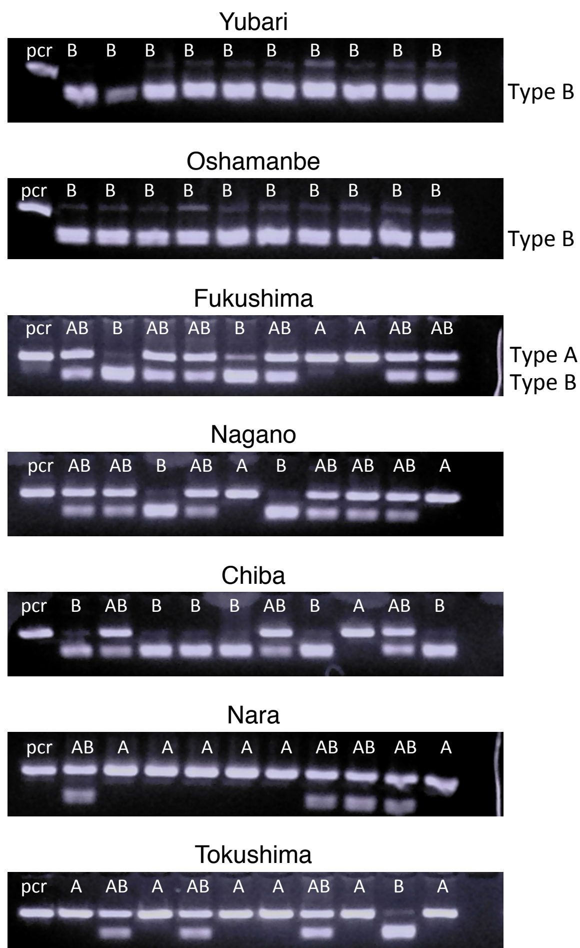


Fig. 4-8 (a) The results of PCoA of *P. melete* NSP nucleotide sequences. This clearly shows apparent two types of NSPs exist in *P. melete* and one of them is not observed in Hokkaido populations. (b)The results of PCR-RFLP for identifying two NSP types in each individual and its genotype are shown. “pcr” shows non-digested product. HinCII digested only type B NSP and all the individuals in Hokkaido populations have type B NSP.

(a)

P. melete

K=2
1.00
0.75
0.50
0.25
0.00
Yubari Oshamanbe Fukushima Hotaka Chiba Nara Tokushima
Cluster1
Cluster2

***K*=3**
1.00
0.75
0.50
0.25
0.00
Yubari Oshamanbe Fukushima Hotaka Chiba Nara Tokushima
Cluster1
Cluster2
Cluster3

K=4
1.00
0.75
0.50
0.25
0.00
Yubari Oshamanbe Fukushima Hotaka Chiba Nara Tokushima
Cluster1
Cluster2
Cluster3
Cluster4

(b)

P. napi groups

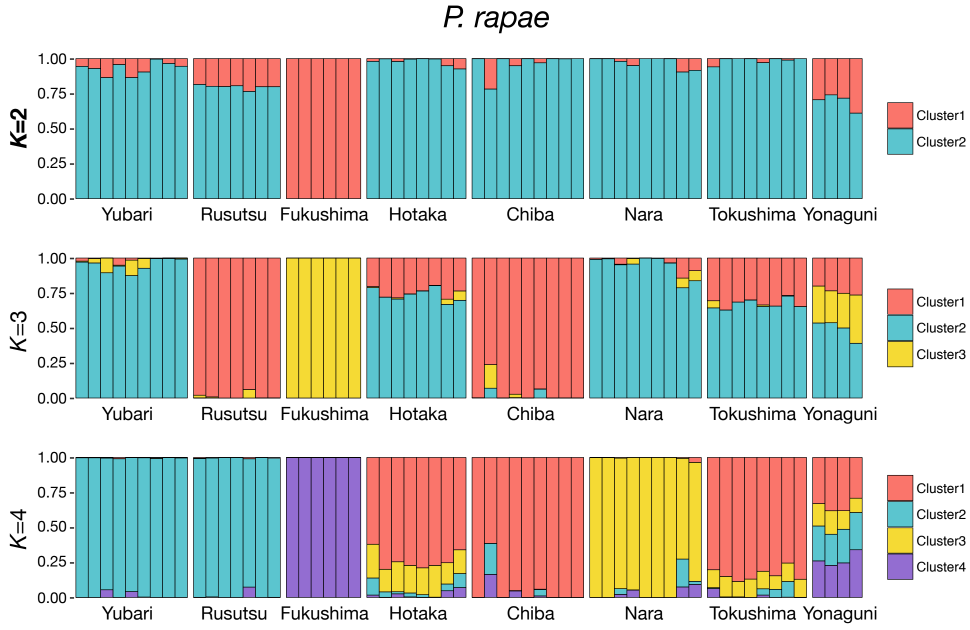
***K*=2**
1.00
0.75
0.50
0.25
0.00
Yubari Rusutsu Fukushima Hotaka Kanagawa Nara Miyazaki
Cluster1
Cluster2

K=3
1.00
0.75
0.50
0.25
0.00
Yubari Rusutsu Fukushima Hotaka Kanagawa Nara Miyazaki
Cluster1
Cluster2
Cluster3

K=4
1.00
0.75
0.50
0.25
0.00
Yubari Rusutsu Fukushima Hotaka Kanagawa Nara Miyazaki
Cluster1
Cluster2
Cluster3
Cluster4

ND5 identification
P. dulcinea *P. napi nesis*

(c)



(d)

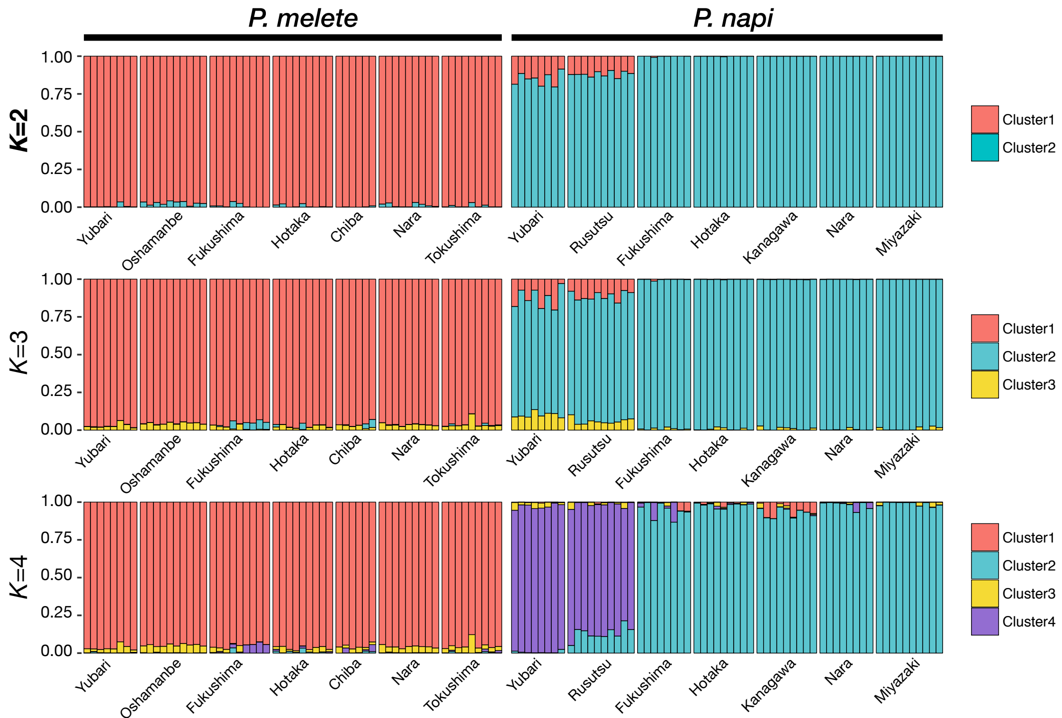


Fig. 4-9 STRUCTURE analysis results of (a) *P. melete*, (b) *P. napi*, (c) *P. rapae*, (d) *P. melete* & *P. napi*. The optimal clustering number *K* is in bold. Northern populations come to left in these figures.

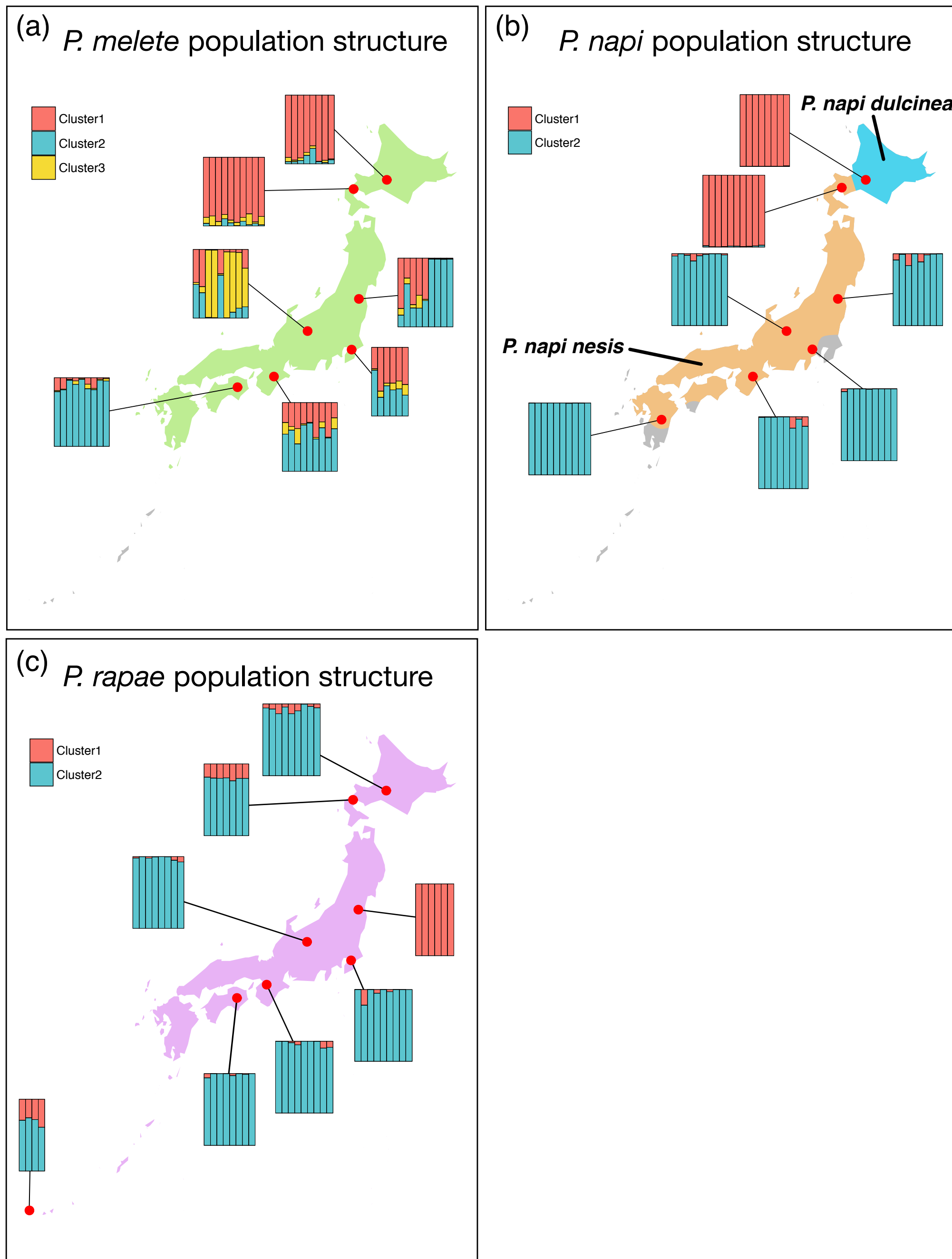


Fig. 4-10 STRUCTURE analysis results of optimal clustering number are shown for (a) *P. melete*, (b) *P. napi*, and (c) *P. rapae* with habitat range of each species.

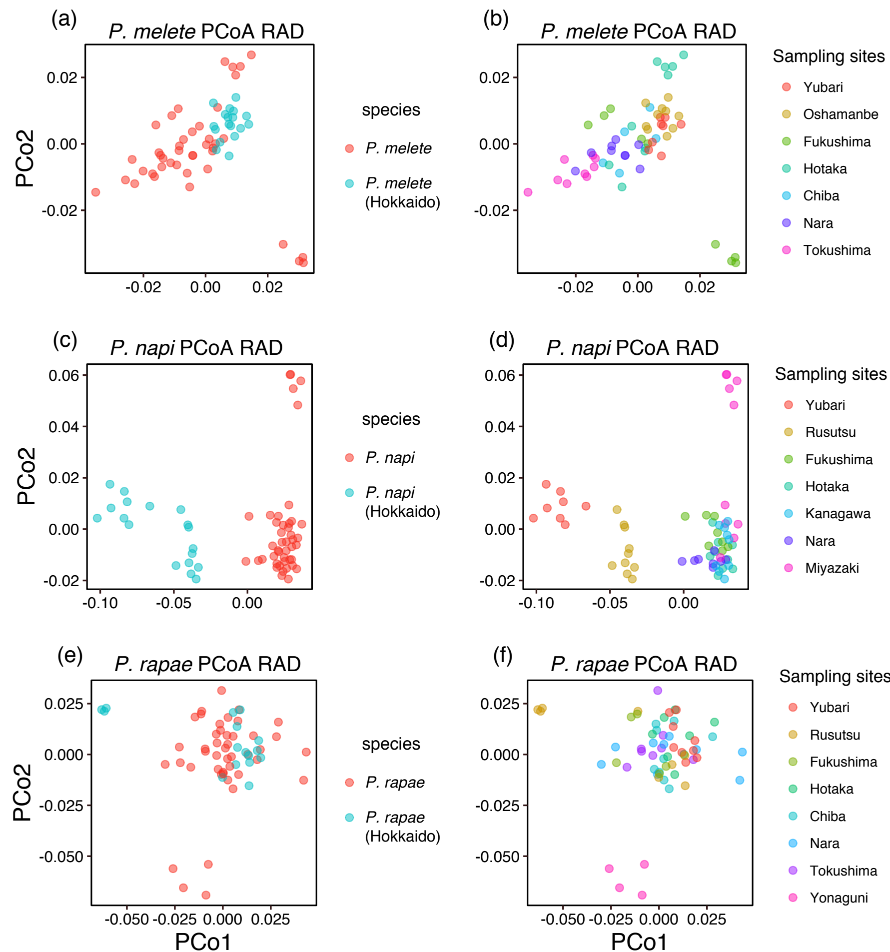


Fig. 4-11 PCoA on SNPs from RAD-seq. (a) Biplot of *P. melete* colored by Hokkaido and Honshu populations difference. (b) Biplot of *P. melete* colored by each population. (c) Biplot of *P. napi* colored by Hokkaido and Honshu populations difference. (d) Biplot of *P. napi* colored by each population. (e) Biplot of *P. rapae* colored by Hokkaido and Honshu populations difference. (f) Biplot of *P. rapae* colored by each population.

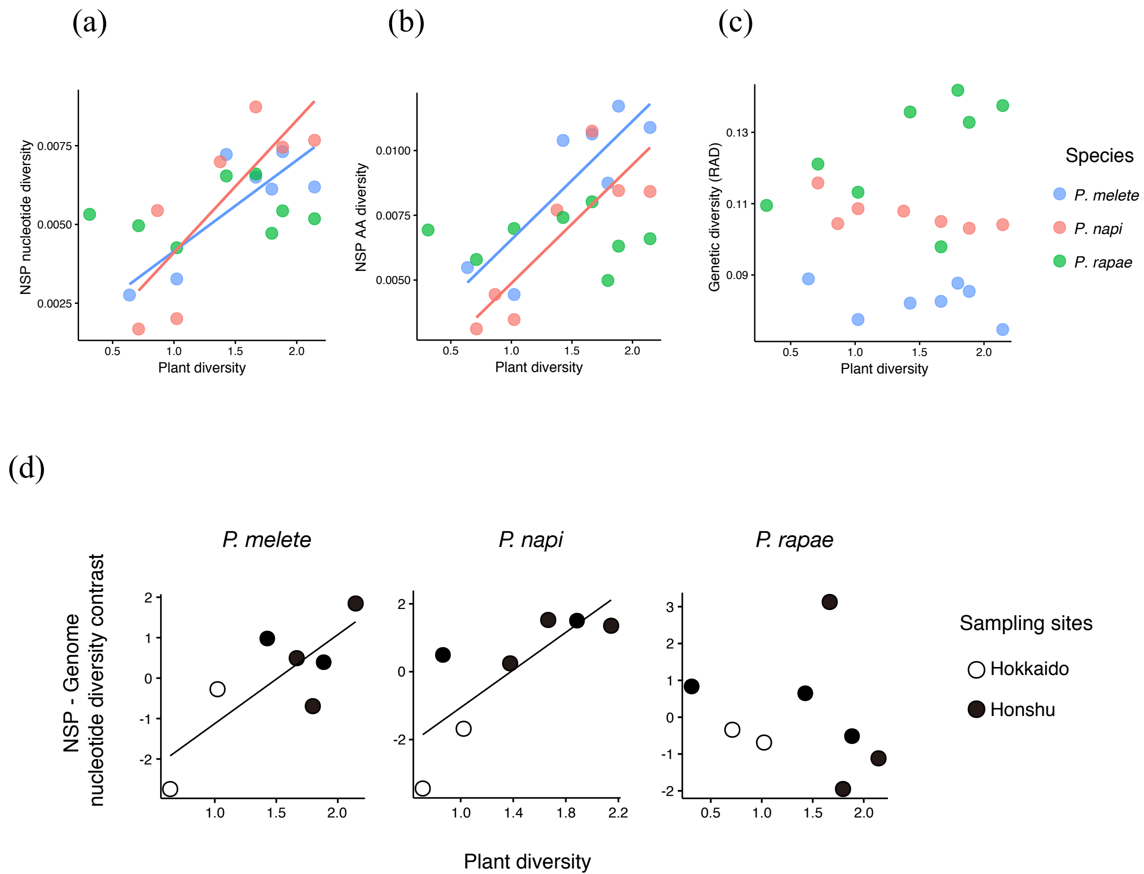


Fig. 4-12 Observed relationship between plant diversity and (a) NSP nucleotide diversity, *P. melete*; *P* = 0.0234, *P. napi*; *P* = 0.0282 *P. rapae*; *P* = 0.627, (b) NSP amino acid diversity, *P. melete*; *P* = 0.0170, *P. napi*; *P* = 0.0166 *P. rapae*; *P* = 0.850. (c)No correlation found in Genome associated genetic diversity from RAD-seq with plant diversity. (d) Genetic diversity contrast between NSP and Genome wide SNP is also correlated with plant diversity. *P. melete*; *P* = 0.0346, *P. napi*; *P* = 0.0328 *P. rapae*; *P* = 0.624. Hokkaido has relatively lower contrast score than Honshu area in *P. melete* and *P. napi*.

Table 4-1 The sampling sites in this study and acquired numbers of individuals of each *Pieris* butterfly species. Observed plant diversity indices (Shannon) in each sampling sites are also shown.

Sampling sites	*P. melete*	*P. napi*	*P. rapae*	lat.	long.	Plant diversity index
Yubari (Hokkaido)	20	28	31	43.0	142.1	1.0224
Rusutsu (Hokkaido)	-	43	21	42.7	140.8	0.7115
Oshamanbe (Hokkaido)	24	-	-	42.6	140.3	0.6374
Fukushima	47	24	12	37.8	140.6	1.6662
Nagano	35	77	39	36.3	137.8	2.1451
Chiba	112	-	34	35.5	140.2	1.7968
Kanagawa	44	14	-	35.5	139.2	1.3765
Nara	51	33	20	34.3	136.0	1.8850
Tokushima	150	-	44	33.9	134.2	1.4266
Miyazaki	-	34	-	32.7	131.3	0.8650
Yonaguni (Okinawa)	-	-	8	24.5	123.0	0.3125
Total	483	253	209			

Table 4-2 Sequenced NSP and its genetic statistics. NSPπ shows nucleotide diversity and significant Tajima's *D* values are in bold. Sequences: number of sequences, Number of sites: NSP length, S sites: number of segregating sites, NSP π: nucleotide diversity, AA diversity: amino acid diversity.

Species	Sampling sites	Sequences	Number of sites	S sites	NSP π	AA diversity	Tajima's *D*	*P* value
P. melete	**Yubari (Hokkaido)**	10	1861	29	0.0033	0.0044	**-1.936**	**0.006**
	Oshamanbe (Hokkaido)	10	1861	25	0.0028	0.0055	**-1.983**	**0.001**
	Fukushima	10	1861	37	0.0065	0.0106	-0.342	0.391
	Nagano	10	1861	32	0.0062	0.0109	0.113	0.587
	Chiba	10	1861	37	0.0061	0.0087	-0.605	0.289
	Nara	10	1858-1861	47	0.0073	0.0117	-0.872	0.213
	Tokushima	10	1858-1861	39	0.0072	0.0104	-0.105	0.496
P. napi	**Yubari (Hokkaido)**	10	1861	18	0.002	0.0035	**-1.913**	**0.008**
	Rusutsu (Hokkaido)	10	1861	14	0.0017	0.0031	**-1.68**	**0.036**
	Fukushima	10	1861	63	0.0087	0.0107	-1.31	0.098
	Nagano	10	1861	48	0.0077	0.0084	-0.758	0.232
	Kanagawa	10	1861	43	0.007	0.0077	-0.683	0.278
	Nara	10	1861	48	0.0075	0.0085	-0.873	0.207
	Miyazaki	10	1861	35	0.0054	0.0044	-0.857	0.206
P. rapae	**Yubari (Hokkaido)**	10	1870	35	0.0043	0.0070	**-1.733**	**0.017**
	Rusutsu (Hokkaido)	9	1870	35	0.005	0.0058	-1.432	0.083
	Fukushima	9	1870	39	0.0066	0.0080	-0.731	0.252
	Nagano	10	1870	42	0.0052	0.0066	**-1.704**	**0.037**
	Chiba	8	1870	29	0.0047	0.0050	-1.135	0.152
	Nara	10	1870	37	0.0054	0.0063	-1.099	0.15
	Tokushima	10	1870	40	0.0065	0.0074	-0.674	0.271
	Yonaguni (Okinawa)	8	1870	31	0.0053	0.0069	-0.916	0.228

Table 4-3 PCR-RFLP genotyping of *P. melete* NSP (n = 10 for each population). The observed genotype frequency and allele frequency are shown. Expected genotype frequency is also listed as Exp.

Sampling sites	AA	AB	BB	pA	pB	Exp. AA	Exp. AB	Exp. BB
Yubari (Hokkaido)	0	0	1	0	1	0	0	1
Oshamanbe (Hokkaido)	0	0	1	0	1	0	0	1
Fukushima	0.2	0.6	0.2	0.5	0.5	0.25	0.5	0.25
Nagano	0.2	0.6	0.2	0.5	0.5	0.25	0.5	0.25
Chiba	0.1	0.3	0.6	0.25	0.75	0.0625	0.375	0.5625
Nara	0.6	0.4	0	0.8	0.2	0.64	0.32	0.04
Tokushima	0.6	0.3	0.2	0.75	0.25	0.5625	0.375	0.0625

Table 4-4 Genetic statistics from RAD sequencing of three *Pieris* butterfly species from each population. Ind./loci: individual number per loci, RAD: genetic diversity from RAD-seq, Obs. Het: observed heterozygosity, Exp. Het: expected heterozygosity.

Species	population	Sample size	Private alleles	Ind. / loci	Obs. Het	Exp. Het	RAD π	Fis
P. melete	Yubari (Hokkaido)	8	380	7.106	0.059	0.077	0.083	0.071
	Oshamanbe (Hokkaido)	10	535	9.164	0.066	0.086	0.091	0.082
	Fukushima	9	357	8.122	0.065	0.082	0.087	0.068
	Nagano	9	312	8.251	0.066	0.076	0.081	0.042
	Chiba	6	352	5.519	0.075	0.091	0.100	0.064
	Nara	9	402	8.175	0.066	0.087	0.092	0.080
	Tokushima	9	422	8.082	0.066	0.083	0.089	0.068
P. napi	Yubari (Hokkaido)	8	491	7.036	0.078	0.113	0.122	0.125
	Rusutsu (Hokkaido)	10	374	8.982	0.084	0.107	0.113	0.093
	Fukushima	8	217	7.114	0.081	0.111	0.120	0.102
	Nagano	9	182	7.976	0.081	0.107	0.115	0.094
	Kanagawa	9	177	8.249	0.084	0.109	0.116	0.089
	Nara	8	162	7.239	0.088	0.105	0.113	0.064
	Miyazaki	10	219	9.138	0.078	0.102	0.108	0.086
P. rapae	Yubari (Hokkaido)	9	161	8.166	0.106	0.114	0.121	0.042
	Rusutsu (Hokkaido)	7	136	6.557	0.085	0.119	0.129	0.109
	Fukushima	7	87	5.516	0.084	0.103	0.114	0.070
	Nagano	8	253	7.241	0.098	0.142	0.153	0.148
	Chiba	9	298	8.026	0.095	0.139	0.149	0.153
	Nara	9	253	8.093	0.101	0.134	0.143	0.115
	Tokushima	8	264	7.153	0.096	0.142	0.153	0.151
	Yonaguni (Okinawa)	4	180	3.552	0.078	0.104	0.121	0.086

Table 4-5 Measured F_{st} values from RAD-seq analysis for two closely related *Pieris* species (*P. melete* and *P. napi*)

		P. melete						*P. napi*						
		Osha	Fuku	Naga	Chib	Nara	Toku	Yuba	Rusu	Fuku	Naga	Kana	Nara	Miya
P. melete	Yubari (Hokkaido)	0.002	0.010	0.018	0.006	0.007	0.016	0.159	0.157	0.208	0.229	0.218	0.235	0.212
	Oshamanbe (Hokkaido)		0.008	0.013	0.007	0.007	0.015	0.156	0.160	0.206	0.216	0.222	0.227	0.219
	Fukushima			0.015	0.006	0.010	0.015	0.187	0.179	0.217	0.236	0.240	0.257	0.236
	Nagano				0.013	0.013	0.018	0.167	0.186	0.221	0.253	0.255	0.264	0.249
	Chiba					0.006	0.013	0.152	0.164	0.207	0.235	0.233	0.237	0.238
	Nara						0.008	0.169	0.164	0.221	0.235	0.235	0.244	0.227
	Tokushima							0.137	0.149	0.188	0.212	0.200	0.212	0.209
P. napi	Yubari (Hokkaido)								0.021	0.057	0.072	0.079	0.070	0.082
	Rusutsu (Hokkaido)									0.041	0.039	0.049	0.039	0.055
	Fukushima										0.007	0.011	0.017	0.022
	Nagano											0.010	0.016	0.017
	Kanagawa												0.020	0.030
	Nara													0.031
	Miyazaki													

General Discussion

Glucosinolate diversification in Brassicales plants and its ecological role

In this thesis, I focused on the underlying mechanisms and evolutionary consequences of arms-race between brassicales plant defense and pierid butterfly adaptation traits. Brassicales have a variety of glucosinolates (GLSs) as a defense against herbivores, and these GLSs were diversified associated with whole genome duplication events in brassicales plants (Hofberger et al. 2013; Edger et al. 2015; Halkier & Gershenzon 2006). This suggests whole genome duplication enabled brassicales plants to have genetic flexibility for acquiring new types of GLSs. However, the ecological factors which drove this diversification was still hard to be explained, even some of the GLSs are known to have differentiated functions which can act as a defense against different guilds of enemies (herbivores or fungi) (Hopkins et al. 2009; Kos et al. 2012; Weigel 2012; Abdalsamee & Müller 2012; Müller et al. 2010; Beekwilder et al. 2008). In Chapter 1, I found Brassicaceae, which possess highest variety of GLSs, has several multiple defense strategies (defense syndrome) which was associated with a certain types of GLS profiles (Agrawal & Fishbein 2006; Silva & Batalha 2010; Travers-Martin & Müller 2008). For example, dense trichome can co-occur with short chain aliphatic GLSs but long chain aliphatic GLSs can be found with high nutrition without trichomes. This result suggests GLSs also have evolved as a part of plant defense and factors which ecologically drove the diversification of GLS can not be separated from the concept of multiple defense strategies in plants.

Plant defense that affect Pieris butterfly species-specific host utilization

Assuming brassicales and pierid butterflies coevolution occurred mediated by GLS diversification, a part of species-specific host utilization of pierid butterflies should be explained by differential capability of pierid butterflies against different types of GLSs (Althoff et al. 2014). Differential host utilization among pierid butterflies were reported in some cases (Chew 1980; Ohsaki & Sato 1994), however, relationships between host use and GLS profiles of the host plants have not been well understood. In Chapter 2, I observed differential host utilization patterns among four *Pieris* butterfly species in larval performance level, and compared them with a number of plant defense traits including non-GLS defenses as it was stressed in previous Chapter. I found their host use difference could be correlated only with GLS profile of plants but not with their non-GLS defenses. Although this observation was from only a restricted number of detectable GLS, our result suggests even *Pieris* butterfly, which is one of Brassicaceae plant specialist, is not fully capable of a set of GLSs and their GLS detoxification ability would be rather species-specific. This implies *Pieris* might evolve their detoxification mechanisms through arms-race with GLS diversification, and differentiate the mechanisms as species-specific as evolutionary consequences. This result produced a correlation type support which can guide the previous research, although further researches about molecular mechanisms of GLS detoxification in *Pieris* are needed to confirm this.

Pieris butterflies appears to handle two different detoxification genes for dealing with wider range of GLSs in their host plants.

According to the previous chapter, I showed plants use several types of defense and form multiple defenses to defend themselves from a variety of herbivores. This pointed

out we need to consider GLS as a part of plant multiple defense. Species specific host utilizations of *Pieris* butterflies, on the other hand, were not affected by non-GLS defenses but GLS profiles of Brassicaceae plants. This suggests resistance abilities of *Pieris* butterflies against GLS are not identical among species, although they share the same detoxification mechanism, namely NSP (Wheat et al. 2007; Wittstock et al. 2004; Fischer et al. 2008). These result support the coevolutionary hypothesis between brassicales and pierid butterflies, however, it is still difficult to mention interspecific difference of detoxification ability against GLSs because of the lack of information about molecular detoxification mechanisms of NSP(Heidel-Fischer & Vogel 2015; Fischer et al. 2008). In Chapter 3, I focused on NSP gene family and its regulation patterns of *Pieris* butterflies against different types of plants with completely different GLS profiles. I found not only NSP but also its sister gene called MA would act as a part of GLS detoxification or digestion in *Pieris* butterflies (Fischer et al. 2008). In addition, I observed that NSP was strongly down regulated in larvae which fed on a certain type of Brassicaceae plants, however, the low expression NSP was complemented by higher expression of MA and inverse regulation pattern was observed in different plant species. Therefore, I would stress the function of MA would be different from that of NSP in point of its substrate specificity. Interestingly, this pattern of compliment expressions of NSP and MA were rather universal in all of the four *Pieris* butterflies. These results clearly showed that the function of NSP related gene can be dramatically changed with its minor sequence change and this would strongly affect the arms-race between brassicales defense and pierid adaptation. Although I could not observe any species specific regulation patterns among *Pieris* butterflies, more focused researches on functional differences of NSP and MA among pierid species would give

us insights to understand counter adaptation history in pierid and its evolutionary consequences on their speciation.

Microevolution of NSP

In the last chapter, I focused on microevolutionary consequences of NSPs in three Japanese *Pieris* butterflies in the wild. I found that NSP nucleotide diversity was significantly decreased in populations which are exposed to lower host plant community diversity. Interestingly, this trend was only found in two native Brassicaceae weed feeding *Pieris* species and not observed in crop-feeder *Pieris rapae*. It seems *P. rapae* was not affected by local Brassicaceae plant community, since they are hardly rely on Brassica crops in each populations. The exciting results that NSP diversity can correlate with local host plant diversity suggested that the key innovation traits in *Pieris* can respond to local host plant community and can have different microevolutionary consequences. This would be strong evidence which support arms-race between GLS and NSP can be occurred in the wild.

In this thesis, I revealed that GLS are involved in multiple defense strategies of Brassicaceae plants, however, *Pieris* butterflies only respond to the GLS profiles of plants in larval performance level. Differential interspecific capability of *Pieris* butterflies against GLS was also observed and this supported the idea of coevolutionary relationship between brassicales and pierid was mediated by GLS diversification. Additional molecular work on expression levels of NSP gene family revealed that not only NSP but its sister gene MA also would have important role in GLS detoxification in *Pieris* butterflies and both of the genes would differentiate for their specific functions. Microevolutionary dynamics which was observed in the wild population of three

Japanese butterflies was also supported that GLS and NSP arms-race can be happen in the wild. These support that arms-race between brassicales and pierid butterflies can be generated by GLS diversification in the wild, however, GLS have diversified not only for pierid butterflies. Pierid butterflies, on the other hand, would have evolved their adaptation traits responding to GLS diversification. Especially in genus *Pieris*, they appear to utilize two different genes to deal with diversified GLS possessed by their host plants. This strongly indicates we need to include MA as well to understand the adaptation traits evolution in pierid butterflies in this arms-race system. Although detoxification enzyme functional differentiation in this gene family and its interspecific difference should be confirmed, my results produces important insights for understanding the mechanisms underlies in the arms-race of brassicales and pierid butterflies.

Acknowledgement

First of all, I would like to appreciate Dr. Masashi Murakami (Chiba university). He has always kindly helped me for doing my work for a long time and supported me in general. This work could not be accomplished without his help. I also would like to thank Dr. Heiko Vogel, Dr. Hanna-Heidel Fischer and the member of Entomology department in Max Planck Institute for Chemical Ecology in Jena. They helped me for doing major part of my work in Chapter 3 and 4.

I also would like to stress that members of Community ecology laboratory (Chiba university) critically contributed to my work in general. Especially, Natsumi Tsuzuki dedicated a lot for a part of experimental work in Chapter 3 and Ai Sato greatly helped me for doing essential fieldworks and laboratory works included in Chapter 4. I would like to express my special thanks to them.

Finally, I would like to thank my family and my great friends for supporting me. I am sure I could not complete this thesis without their kind help.

A part of this study was supported by Grants-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science

References

Abdalsamee, M.K. & Müller, C., 2012. Effects of indole glucosinolates on performance and sequestration by the sawfly Athalia rosae and consequences of feeding on the plant defense system. *Journal of chemical ecology*, 38(11), pp.1366–75. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23053922.

Agrawal, A. a & Fishbein, M., 2006. Plant defense syndromes. *Ecology*, 87(7 Suppl), pp.S132-49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16922309.

Althoff, D.M., Segraves, K. a & Johnson, M.T.J., 2014. Testing for coevolutionary diversification: linking pattern with process. *Trends in ecology & evolution*, 29(2), pp.82–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24314843.

Beekwilder, J. et al., 2008. The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. *PloS one*, 3(4), p.e2068. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2323576&tool=pmcentrez&rendertype=abstract.

Chew, F.S., 1980. Foodplant Preferences of Pieris Caterpillars (Lepidoptera). *Oecologia*, pp.347–353.

Edger, P.P. et al., 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, p.201503926. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1503926112.

Fischer, H.M. et al., 2008. Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular biology and evolution*, 25(5), pp.809–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18296701.

Halkier, B.A. & Gershenzon, J., 2006. Biology and biochemistry of glucosinolates. *Annual review of plant biology*, 57, pp.303–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16669764.

Heidel-Fischer, H.M. & Vogel, H., 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, pp.8–14. Available at: http://dx.doi.org/10.1016/j.cois.2015.02.004.

Hofberger, J. a. et al., 2013. Whole Genome and Tandem Duplicate Retention Facilitated Glucosinolate Pathway Diversification in the Mustard Family. *Genome Biology and Evolution*, 5(11), pp.2155–2173. Available at: http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt162.

Hopkins, R.J., van Dam, N.M. & van Loon, J.J. a, 2009. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual review of entomology*, 54, pp.57–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18811249.

Kos, M. et al., 2012. Effects of glucosinolates on a generalist and specialist leaf-chewing herbivore and an associated parasitoid. *Phytochemistry*, 77, pp.162–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22281379.

Müller, R. et al., 2010. Differential effects of indole and aliphatic glucosinolates on lepidopteran herbivores. *Journal of chemical ecology*, 36(8), pp.905–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20617455.

Ohsaki, N. & Sato, Y., 1994. Food plant choice of Pieris butterflies as a trade-off between parasitoid avoidance and quality of plants. *Ecology*, 75(1), pp.59–68.

Silva, D.M. & Batalha, M.A., 2010. Defense syndromes against herbivory in a cerrado plant community. *Plant Ecology*, 212(2), pp.181–193. Available at: http://link.springer.com/10.1007/s11258-010-9813-y.

Travers-Martin, N. & Müller, C., 2008. Matching plant defence syndromes with performance and preference of a specialist herbivore. *Functional Ecology*, 22(6), pp.1033–1043. Available at: http://doi.wiley.com/10.1111/j.1365-2435.2008.01487.x.

Weigel, D., 2012. Natural variation in Arabidopsis: from molecular genetics to ecological genomics. *Plant physiology*, 158(1), pp.2–22. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3252104&tool=pmcentrez&rendertype=abstract.

Wheat, C.W. et al., 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(51), pp.20427–31. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2154447&tool=pmcentrez&rendertype=abstract.

Wittstock, U. et al., 2004. Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), pp.4859–64. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=387339&tool=pmcentrez&rendertype=abstract.