Title:

Microevolution of *Pieris* butterfly genes involved in host-plant adaptation along a host-plant community cline

Short running title: Microevolution of host-plant adaptation genes

Yu Okamura[1,2], Ai Sato[2], Lina Kawaguchi[3], Atsushi J. Nagano[4,5], Masashi Murakami[2], Heiko Vogel[1], Juergen Kroymann[6]

1 Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, Hans-Knöll-Str. 8, Jena, 07745, Germany

2 Community Ecology Lab, Faculty of Science, Chiba University, Chiba, 263-8522, Japan

3 Research Administration Office, Kyoto University, Kyoto, 606-8501, Japan

4 Faculty of Agriculture, Ryukoku University, Shiga, 520-2194, Japan

5 Institute for Advanced Biosciences, Keio University, Yamagata, 997-0017, Japan

6 Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, Orsay, 91405, France

Corresponding author:

Yu Okamura

yokamura@ice.mpg.de

Key words: Insects, host-plant adaptation, arms race, selection

Abstract

Herbivorous insects have evolved counteradaptations to overcome the chemical defenses of their host plants. Several of these counteradaptations have been elucidated at the molecular level, in particular for insects specialized on cruciferous host plants. While the importance of these counteradaptations for host plant colonization is well established, little is known about their microevolutionary dynamics in the field. In this study, we examine patterns of host plant use and insect counteradaptation in three *Pieris* butterfly species across Japan. The larvae of these butterflies express nitrile-specifier protein (NSP) and its paralog major allergen (MA) in their gut to overcome the highly diversified glucosinolate-myrosinase defense system of their cruciferous host plants. *Pieris napi* and *Pieris melete* colonize wild Brassicaceae whereas *Pieris rapae* typically uses cultivated *Brassica* as a host, regardless of the local composition of wild crucifers. As expected, NSP and MA diversity was independent of the local composition of wild Brassicaceae in *P. rapae*. In contrast, NSP diversity correlated with local host plant diversity in both species that preferred wild Brassicaceae. *P. melete* and *P. napi* both revealed two distinct major *NSP* alleles, which shaped diversity among local populations, albeit with different evolutionary trajectories. In comparison, *MA* showed no indication for local adaptation. Altogether, *MA* appeared to be evolutionary more conserved than *NSP*, suggesting that both genes play different roles in diverting host plant chemical defense.

Introduction

Herbivorous insects encounter heterogeneous plant communities in the field and often use a subset of those plants as hosts. While feeding, herbivores are exposed to the chemical defenses of their host plants. Defenses typically vary between plant species and populations (Futuyma & Agrawal, 2009; Kliebenstein et al., 2001; Mitchell-Olds & Schmitt, 2006; Prasad et al., 2012; Windsor et al., 2005). Available evidence suggests that specialist herbivores acquired key innovations that enabled them to colonize their host plants by circumventing host plant chemical defenses (Berenbaum, Favret, & Schuler, 1996; Ratzka, Vogel, Kliebenstein, Mitchell-Olds, & Kroymann, 2002; Wheat et al., 2007; Wittstock et al., 2004). The microevolutionary dynamics of these herbivore counteradaptations is largely unknown, despite their ecological and evolutionary importance.

Pieris butterflies use plants from the family Brassicaceae as their hosts. These plants rely on the glucosinolate (GLS)-myrosinase system as their main chemical defense (Wheat et al., 2007; Wittstock et al., 2004; Wittstock & Halkier, 2002). GLSs are hydrolyzed by plant myrosinase enzymes when plant tissue is macerated by insect herbivory, and the main hydrolysis products, isothiocyanates, are highly toxic to most herbivores (Halkier & Gershenzon, 2006; Wittstock & Halkier, 2002). *Pieris* butterfly larvae overcome this defense system by expressing nitrile specifier proteins (NSPs) in their gut. NSPs affect the outcome of GLS breakdown to form nitriles rather than toxic isothiocyanates by an as yet unknown mechanism (Wittstock et al., 2004). Recent genetic and molecular analyses in *Anthocharis cardamines*, another Pierid species feeding on Brassicaceae, indicated that major allergen (MA), an ancient NSP paralog, has NSP-like activity (Edger

69 et al., 2015). Noteworthy, *MA* transcript levels in the gut of *Pieris* larvae are upregulated
70 in response to certain types of GLS (Okamura, Sato, Tsuzuki, Sawada, et al., 2019).
71 Together, *NSP* and *MA* constitute the *NSP*-like gene family, a key innovation that enabled
72 Pierid butterflies to colonize their host plants from the order Brassicales (Edger et al.,
73 2015; Fischer, Wheat, Heckel, & Vogel, 2008; Wittstock et al., 2004).
74
75 GLSs are a highly diverse group of plant secondary metabolites, with about 135 different
76 structures identified in the Brassicales (Blažević et al., 2020). GLS profiles vary both
77 quantitatively and qualitatively (Agerbirk & Olsen, 2012; Fahey, Zalcmann, & Talalay,
78 2001; Olsen et al., 2016) and *Pieris* species typically use several genera within the
79 Brassicaceae as their hosts (Friberg & Wiklund, 2019; Ohsaki & Sato, 1994; Ohsaki,
80 1979; Okamura, Sato, Tsuzuki, Sawada, et al., 2019; Okamura, Sawada, Hirai, &
81 Murakami, 2016). Therefore, each species encounters a range of different GLS profiles
82 in the field, a scenario that should result in local adaptation to host plant communities.
83 Surprisingly, a previous study with *Pieris rapae*, the cabbage white butterfly, found no
84 evidence for local adaptation of *NSP* genes in Europe or in the U.S. (Heidel-Fischer,
85 Vogel, Heckel, & Wheat, 2010). However, this study did not address patterns of host-
86 plant use. Furthermore, *P. rapae*, a considerable pest of *Brassica*, depends mainly on
87 cultivated plants (Grishin et al., 2016; Ryan et al., 2019) and not on local populations of
88 wild Brassicaceae. Hence, it seems important to test species that depend on wild plants in
89 order to detect potential microevolutionary patterns in the genes that underlie insect
90 counteradaptations against their host plants' defenses.
91

In this study, we focus on three *Pieris* species, *Pieris melete*, *Pieris napi* and *P. rapae* that co-occur across most of Japan. *P. melete* and *P. napi* use wild Brassicaceae as host plants, including the genera *Arabis*, *Cardamine* and *Rorippa*, but their larvae are rarely observed on cultivated *Brassica* (Ohsaki & Sato, 1994; Okamura, Sato, Tsuzuki, Sawada, et al., 2019). In Japan, plant communities vary substantially along a north-south cline (Kubota, Shiono, & Kusumoto, 2015). Because this holds also true for the latitudinal distribution of wild Brassicaceae, we expected that the plant community cline should affect host-plant use in *P. melete* and *P. napi*, which depend on wild Brassicaceae, but not *P. rapae*, which feeds on cruciferous crops. We sampled local populations of these three species across Japan and acquired data about host plant communities at the sampling sites. We sequenced *NSP* and *MA* and used restriction-site associated DNA sequencing (RAD-seq) data for comparison, to distinguish between the potential impact of selection and neutral genetic processes (Bernatchez & Landry, 2003; Dionne, Miller, Dodson, Caron, & Bernatchez, 2007). As expected, we did not find any evidence for local adaptation of crop-dependent *P. rapae*. In contrast, we found a clear correlation between local host plant diversity and NSP, but not MA, amino-acid sequence diversity for both *Pieris* species that preferred wild Brassicaceae. Both *P. melete* and *P. napi* possessed two major *NSP* alleles, which caused the observed correlation with host plant diversity, albeit with completely different distribution patterns among Japanese populations. *P. napi* consisted of two distinct populations using very different sets of host plants, with a different *NSP* allele fixed in each population and evidence for directional selection of one allele in the past. In *P. melete*, on the other hand, two *NSP* alleles were maintained by balancing selection across Japan and allele frequency was correlated with local host plant diversity. Altogether, our study indicates that microevolution of counteradaptive traits

116 has an important role for herbivores to adapt to heterogeneous chemical defenses in the

117 host plants that they encounter in the field.

118

Materials and Methods

Selection of sampling sites

To determine suitable sampling sites, we estimated the diversity of Brassicaceae across Japan with Maxent ver. 3.4.1 (Phillips, Dudík, & Schapire, 2004; Phillips, Anderson, Dudík, Schapire, & Blair, 2017). We used 11,325 individual location data of 44 genera of the Brassicaceae in the Global Biodiversity Information Facility (GBIF.org; https://doi.org/10.15468/dl.a2nqtv) and S-Net (http://science-net.kahaku.go.jp), and 19 climate variables for Japan at 2.5 arc-minute resolution from WorldClim (Fick & Hijmans, 2017) to infer additional potential locations of genera from the Brassicaceae (Table S1). We calculated the Shannon index for genera-based diversity of Brassicaceae using the R package "vegan" (Oksanen et al., 2017). Based on this estimation, we chose eleven sampling sites across Japan for each of the three *Pieris* spp., *P. melete*, *P. napi* and *P. rapae* (Table S2).

Sampling of Brassicaceae and *Pieris* spp.

We collected *Pieris* larvae from April to August 2017. At each sampling site, we used the transect method to search for Brassicaceae along a > 2-km sampling path. We noted the number of plants per species, and collected larvae or eggs of *P. melete*, *P. napi* and *P. rapae* from their host plants. We disregarded two Brassicaceae genera, *Capsella* and *Erysimum*, because these are unsuitable hosts for those *Pieris* species (Okamura et al., 2016). Similarly, we did not take into account cultivated Brassicaceae, *i.e.*, cabbage and kale, growing in crop fields near sampling sites. We calculated the host plant diversity using Shannon diversity index for each sampling site, based on the number of species and plants per species. We provided eggs or small larvae with their host plants until they

143 reached the 3rd instar and then dissected the larvae to separate guts from the rest of the
144 bodies. We processed larvae larger than 3rd instar immediately. We stored guts in
145 RNAlater (QIAGEN) and body rests at -80°C. We excluded larvae that contained
146 parasitoid wasps from further analyses.
147
148 **RFLP-based identification of *Pieris* butterfly species**
149 We used restriction fragment length polymorphisms (RFLPs) in the mitochondrial *ND5*
150 gene (GenBank accession numbers: LC090587- LC090590) to determine species identity
151 of larvae. *P. melete ND5* has a *Hin*cII site, and *P. napi ND5* has a *Hin*fI and a *Hin*dIII site,
152 whereas *P. rapae ND5* lacks these restriction sites. We amplified *ND5* directly from larval
153 bodies, using MightyAmp™ DNA Polymerase Ver.3 (TaKaRa) and *ND5* universal
154 primers (Table S3), followed by restriction and separation on 2% agarose gels. To verify
155 that our species assignments worked correctly, we used the same RFLP technique for
156 adult males previously classified as *P. melete* (N = 24), *P. napi* (N = 16) or *P. rapae* (N
157 = 24) based on differences in the shape of the androconium. *ND5*-based identification
158 matched androconium-based identification for all 64 individuals.
159
160 ***NSP* and *MA* amplification, cloning and sequencing**
161 We randomly picked ten larvae per species and sampling site to sequence *NSP* and *MA*
162 genes. In total, we used 70 larvae (ten larvae per site from seven sampling sites) of *P.*
163 *melete* and *P. napi,* respectively, and 78 larvae of *P. rapae* (ten larvae per site from seven
164 sampling sites, plus eight larvae from Yonaguni Island). We extracted RNA from
165 dissected larval gut using the RNeasy Mini Kit (QIAGEN), synthesized cDNA with the
166 ReverTra Ace qPCR RT Master Mix (TOYOBO) and used TaKaRa Ex Taq (TaKaRa)

for amplification of *NSP* and *MA* genes. We gel-purified PCR products with the GEL/PCR Purification Mini Kit (FAVORGEN). We designed PCR primers for *NSP* and *MA* genes based on available RNA-seq data (Accession numbers: ERX2829492-ERX2829499) to confirm that primer-binding sites were conserved and to avoid a potential allelic bias in amplification (Table S3). We cloned it with Mighty TA-cloning kit (TaKaRa), purified plasmids with NucleoSpin Plasmid EasyPure (TaKaRa), and sequenced one plasmid for each *NSP* and *MA* gene from each of the 218 larval samples, using an ABI 3730xl DNA Analyzer (Applied Biosystems).

Network analysis of *NSP* and *MA*, and PCR-RFLP of *P. melete NSP* gene variants

We trimmed and aligned *NSP* and *MA* reads with MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) and Mafft (Katoh & Standley, 2013). For each species, we excluded singleton SNPs, assuming that they were PCR errors. We performed NeighborNet network analyses with SplitsTree 4.15.1 (Huson & Bryant, 2006). We used PCR-RFLP to distinguish two major *NSP* variants in *P. melete* to determine whether these variants were different gene copies or alleles. We amplified *NSP* exon 1 from gDNA with EmeraldAmp MAX PCR Master Mix (TaKaRa), using primers specific for exon 1 (Table S3), followed by a *Hin*cII digest to target a fixed SNP, and separation on 3% TBE agarose gels.

RT-qPCR of *P. melete NSP* and *MA* genes

To determine whether *NSP* and *MA* gene copy numbers differed among *P. melete* genotypes, we conducted Real Time quantitative PCR (RT-qPCR) from gDNA using *Ef1a* as a reference gene. For primer design, we used Primer3Plus (Untergasser et al.,

191 2007) with a product size of 70 to180 bp, a Tm of 59 to 61°C, a GC content of 40 to 60%
192 and a maximum polybase of 3. For RT-qPCR, we used the CFX Connect Real-Time PCR
193 Detection System (Bio-Rad) using TB Green Premix Ex Taq II (Tli RnaseH Plus)
194 (TaKaRa). We used the ddCq method (Pfaffl, 2001) and ANOVA to compare relative
195 gene copy numbers among genotypes.
196
197 **RAD sequencing**
198 We extracted gDNA for RAD-seq from the same 218 individuals that we had used for
199 sequencing of *NSP* and *MA* genes, using the Maxwell 16 LEV Plant DNA Kit (Promega),
200 with *Eco*R1 as the restriction enzyme for generating RAD-seq libraries. We ran all 218
201 samples in a single lane on a HiSeq2500 (Illumina SE 50 bp) and trimmed reads with
202 trimmomatic (Bolger, Lohse, & Usadel, 2014), using ILLUMINACLIP:2:10:10,
203 TRAILING:20, SLIDINGWINDOW:4:15, and MINLEN:30. We excluded samples with
204 less than 500,000 reads from further analysis. We called SNPs with stacks ver. 1.48
205 (Catchen, 2013). For ustacks, we set n = 3 and M = 3, and for cstacks, we used the n = 3
206 option. We performed this analysis for each species independently. To acquire values for
207 genetic diversity (π) and genetic distance (F_{st}) of each population, we used "populations"
208 with parameter p set as sampled population number and r = 0.70. Since we suspected that
209 *P. melete* and *P. napi* could hybridize, we also called SNPs by setting p = 1 and r = 0.85
210 in multiple species scales without involving population information for assessing genetic
211 structure of the three species.
212
213 **Population and evolutionary genetic analyses**

For analyses of potential population structure we used RAD-seq data in multiple species scales and performed sparse non-negative matrix factorization (SNMF) with the "snmf" function (K = 1 to 10, repetition = 20, iterations = 200) implemented in the R package "LEA" (Frichot & François, 2015; Frichot, Mathieu, Trouillon, Bouchard, & François, 2014). We used cross-entropy criterion to determine the number of ancestral population (K). We calculated species-wide distributions of Tajima's D (Tajima, 1989) of *NSP* and *MA* using the poly-div_sfs.pl script (https://github.com/santiagosnchez/poly-div_sfs) iterating random sampling of one individual per population and calculation of D for 300 times. We performed these analyses on both, entire coding sequence and individual exons. For calculation of genome-wide Tajima's D at the species level, we used VCFtools (Danecek et al., 2011). We did the same random sampling coupling with selecting 30 random SNPs corresponding to the average number of SNPs in *NSP* within populations. We used poly-div_sfs.pl script to calculate allele frequency of *NSP* and *MA* for each species. At the population level, we used DNAsp ver.6 (Rozas et al., 2017) for calculation of π and F_{st} of *NSP* and *MA*, and Arlequin ver. 3.5 (Excoffier & Lischer, 2010) for Tajima's D. We estimated F_{st}-like amino acid divergence in NSP and MA between populations as follows; Fst_AA = {Dij – mean (D within population)} / Dij, where Dij shows mean p-distance in amino acid sequences between population i and j. We used the online tool at mkt.uab.es/mkt/MKT.asp (Egea, Casillas, & Barbadilla, 2008) to conduct McDonald–Kreitman tests (McDonald & Kreitman, 1991).

Comparison of *NSP* and *MA* with host-plant diversity

We used linear regression to test for each *Pieris* species whether diversity of *NSP* or *MA* correlated with host-plant diversity across sampling sites, using measured host plant

238 diversity as the explanatory variable and NSP or MA amino acid diversity, or genome-
239 wide genetic diversity estimated from RAD-seq data as response variables. Furthermore,
240 we performed a partial Mantel test (Mantel, 1967; Smouse, Long, & Sokal, 1986; Sokal,
241 1979) to test whether NSP or MA amino acid sequence divergence between *Pieris*
242 populations correlated with host-plant community dissimilarities, using the "vegan"
243 package in R (Oksanen et al., 2017). We used the partial Mantel test to compare F_{st}-like
244 amino acid sequence distances of NSP or MA with the Bray-Curtis index for host plant
245 dissimilarity, after controlling for genetic differences between populations (Fst) using
246 RAD-seq data. We permuted rows and columns of the first dissimilarity matrix 100,000
247 times to evaluate the reliability of the partial Mantel test results.
248

Results

1. Field Sampling

Diversity of Brassicaceae in the field reflects Maxent-based predictions

The Maxent presence-probability estimation predicted that diversity in Brassicaceae communities should be higher in central Japan than in northern and southern regions (Fig. 1a). We chose eleven sites along this predicted cline of Brassicaceae diversity: Yubari, Rusutsu and Oshamanbe (all: on the island of Hokkaido), Fukushima, Nagano, Chiba, Kanagawa, Nara (all: on the island of Honshu), Tokushima (on the island of Shikoku), Miyazaki (on the island of Kyushu), and the island Yonaguni (from north to south; Fig. 1a). In total, we collected data on 4,777 individual plants across 14 genera and 25 species of the Brassicaceae (Table S2). We used these data to estimate the Shannon index for the diversity of Brassicaceae communities at each site. As expected, field-observed diversity in Brassicaceae communities correlated positively with Maxent predictions ($r = 0.701$, $P = 0.016$), was highest in central Japan, and declined towards the north and the south (Fig. 1b).

Sampling of *Pieris* larvae in the field

In total, we collected 945 larvae from eleven sampling sites (Table S4). Based on PCR-RFLP using restriction site polymorphisms in the mitochondrial *ND5* region, we identified 483 larvae of *P. melete*, 253 of *P. napi* and 209 of *P. rapae*.

2. Population structure of three *Pieris* species in Japan

We randomly chose ten individual larvae per population and species for RAD-seq. On average, we obtained 1,062,550 RAD-seq reads per sample. After removal of samples with low read counts (< 500,000 reads), 183 samples remained for population genetic analyses. Stacks ver.1.48 called 2,490 SNPs for *P. melete*, 1,376 for *P. napi*, and 1,410 for *P. rapae.* Genome-wide diversity was higher in *P. rapae* and *P. napi* (both: $\pi = 0.055$) than in *P. melete* ($\pi = 0.042$) (Table 1a). Individual populations of *P. melete* displayed relatively uniform diversity among sampling sites. *P. rapae* populations from central Japan tended to have slightly higher diversity than northern or southern populations. Finally, Hokkaido populations of *P. napi* had higher diversity than other populations in the remainder of Japan (Table 1a)

Genome-wide data did not suggest any apparent population genetic structure of *P. melete* or *P. rapae* across Japan, in contrast to *P. napi* (Fig. 2, 3d). In *P. napi*, populations from Hokkaido were distinct from populations in other parts of Japan. In addition, we found evidence for some, very limited gene flow between species (Fig. 2) but not for recent hybridization events among these closely related butterfly species.

3. Diversity of *NSP* and *MA*

NSP and *MA* sequences are diverse but single copy in *P. melete* and *P. napi*

We cloned and sequenced *NSP* and *MA* cDNAs from the same larvae that we used for RAD-seq. PCR products contained near full-length coding sequences (CDS) of *NSP* (1,860 bp of a total CDS of 1,896 bp) and *MA* (1,869 bp of 1,899 bp).

Network analysis identified two main clusters of *NSP* variants in both, *P. melete* and *P. napi* (Fig. 3a). In contrast, all *P. rapae NSP* sequences grouped together, separate from *P. melete* and *P. napi NSP*s. Hokkaido populations of *P. melete* were fixed for one of the

main *NSP* variants (Figs. 3a, d), while all other *P. melete* populations possessed both variants, albeit with an overall decline in the frequency of the Hokkaido variant towards the south (Fig. 3d). One of the *P. napi NSP* variants was only found in populations from Hokkaido, while the other variant was fixed in central and southern populations (Fig. 3a, d). Network analysis of *MA* showed one single sequence cluster for each of *P. melete* and *P. rapae,* while *P. napi* sequences fell into two distinct clusters (Fig. 3b), with one of the two *P. napi MA* variants present only in Yubari on Hokkaido (Figs. 3b, d).

To test whether the different *NSP* variants of *P. melete* corresponded to different alleles or two different gene copies, we conducted two experiments, a discrimination of variants with PCR-RFLP and a RT-qPCR-based quantification of gene copy number. The *P. melete NSP* variants differed by a SNP causing a *Hin*cII restriction site polymorphism. After PCR, restriction digest, and agarose gel electrophoresis, we expected to obtain one undigested and one digested PCR product across all individuals if *NSP* variants corresponded to two distinct gene copies. In contrast, we expected to obtain undigested and digested PCR products in various combinations in the case that *NSP* variants corresponded to segregating alleles of the same gene. Indeed, we found samples with only undigested or with only digested PCR products, as well as samples with a mix of both, indicating that *NSP* variants corresponded to different alleles of the same gene (Fig. S1). Furthermore, gene copy numbers determined by RT-qPCR (Fig. S2) from eight randomly chosen samples per genotype were compatible with the same number of *NSP* copies in all *P. melete* genotypes, confirming that *NSP* variants were indeed alleles of the same gene.

4. NSP and MA diversity and plant community diversity

NSP but not MA amino acid diversity correlates positively with host-plant diversity in *P. melete* and *P. napi*

We found larvae of *P. melete* feeding on eleven, *P. napi* on 14, and *P. rapae* on nine different plant species (Table S4). We found 36% of all *P. melete* larvae on *Rorippa indica*, 19% on *Cardamine leucantha*, 18% on *Orychophragmus violaceus*, and 13% on *Arabis hirsuta* (Table S4). *P. napi* larvae were mostly present on *Rorippa sylvestris* (23%), *Arabis flagellosa* (17%), *Cardamine occulta* (17%), *Arabis hirsuta* (12%) and *Barbarea orthocera* (11%). Finally, we found *P. rapae* larvae most often on feral plants of *Brassica napus* (33%), followed by *Rorippa indica* and *Rorippa sylvestris* (both: 15%), *Cardamine kiusiana* (12%), *Nasturtium officinale* and *Rorippa palustris* (both: 10%). However, we also saw a substantial number of *P. rapae* larvae and eggs in crop fields of cabbage, broccoli or kale near sampling sites but we did not have permission to sample from these fields.

Altogether, *Pieris* species appeared to use a more diversified set of host plants in central Japan than in northern and southern Japan (Fig. 3c, Table S4). Therefore, we compared diversity in detoxification genes with the diversity of Brassicaceae at sampling sites (Fig. 4a). Local host plant diversity correlated significantly with NSP amino acid diversity for both *P. melete* ($r = 0.83$, $P = 0.022$) and *P. napi* ($r = 0.88$, $P = 0.010$), but not *P. rapae* ($r = 0.021$, $P = 0.961$). By contrast, local host plant diversity did not correlate with MA amino acid diversity or genome-wide diversity for any of the three species (Fig. 4b).

To confirm the relationship between host plant and NSP sequence, we conducted a partial Mantel test (Mantel, 1967; Smouse et al., 1986; Sokal, 1979). Again, we found positive correlations between host plant dissimilarities and NSP sequence divergence between

342 populations for both, *P. melete* ($r = 0.443$, $P = 0.039$) and *P. napi* ($r = 0.436$, $P = 0.009$)
343 but not *P. rapae*. Likewise, results for MA remained non-significant (Table 2).
344

345 ## 5. Statistical tests for selection

346 **Different evolutionary regime between *NSP* and *MA***

347 To identify patterns of potential selection, we calculated Tajima's *D* for *NSP* (D_{NSP}), *MA*
348 (D_{MA}), and at the genome-wide level (D_{RAD}). D_{RAD} was negative in all three species,
349 compatible with overall population expansion (Fig. 5a). In contrast, D_{NSP} was positive
350 and substantially higher than D_{RAD} in *P. melete* and *P. napi*. Also, D_{MA} was higher than
351 D_{RAD} in *P. rapae*. Furthermore, D_{NSP} varied substantially among exons in *P. melete* and
352 moderately in *P. napi* but not in *P. rapae*, while D_{MA} displayed a more uniform pattern
353 across exons of all three species (Fig. 5b). In *P. melete*, exon 1 had a particularly high
354 positive *D*. Similarly, *P. napi* exon 3 displayed an elevated *D*. Close inspection of these
355 exons revealed that both contained several linked, nonsynonymous SNPs at intermediate
356 frequency (Fig. 5c), suggesting that *NSP* was under balancing selection in *P. melete*.
357 In addition, we compared the ratio of non-synonymous to synonymous nucleotide
358 diversity (π_a/π_s) of *NSP* and *MA* of the three *Pieris* spp. *MA* had a lower π_a/π_s ratio
359 compared to *NSP* in all three species (Fig. 5d), suggesting that *MA* was more evolutionary
360 conserved than *NSP* across *Pieris* spp.
361

362 **Divergent selection of *NSP* alleles in *P. napi***

363 We found that *P. melete* and *P. napi NSP*s, as well as *P. napi MA*s, were all grouped into
364 two clearly distinct clusters (Fig. 3a). Therefore, we tested for potential divergent
365 selection between major alleles, using a population-based diversity test (McDonald &

366 Kreitman, 1991). Results for *P. melete NSP* and *P. napi MA* were non-significant
367 (Table 3), but *P. napi NSP* showed a strong signal of divergent selection (NI = 0.15, X^2 =
368 10.6, P = 0.001), caused by an elevated level of non-synonymous *versus* synonymous
369 substitutions between populations from Hokkaido and populations outside of Hokkaido.
370 In order to identify where mutations had occurred, we compared *NSP*s of *P. napi* and *P.*
371 *melete*, which allowed us to infer the direction of evolutionary change. To our surprise,
372 all non-synonymous substitutions that differentiated the Hokkaido populations of *P. napi*
373 from those of central and southern Japan had occurred outside of Hokkaido, while each
374 two synonymous substitutions had occurred in the lineages leading to the Hokkaido and
375 non-Hokkaido populations. Furthermore, ten non-synonymous and 24 synonymous
376 polymorphisms were specific for populations from central and southern Japan, and only
377 two non-synonymous polymorphisms were specific for populations from Hokkaido (Fig.
378 6). These patterns of polymorphism indicate that the lineage leading to the extant *P. napi*
379 non-Hokkaido populations experienced strong directional selection on NSP function,
380 followed by stabilizing selection under relaxed constraints, whereas the lineage leading
381 to the extant *P. napi* Hokkaido populations experienced only stabilizing selection.
382
383

384 ## Discussion

385 Several insect counteradaptations against host plant defenses have been characterized at
386 the molecular level (Berenbaum et al., 1996; Gloss et al., 2014; Heidel-Fischer et al.,
387 2019; Li, Schuler, & Berenbaum, 2003; Ratzka et al., 2002; Wittstock et al., 2004) but
388 only few have been investigated for genetic variation in response to host plant usage.
389 Bono et al. (2008) found some evidence for adaptive evolution of two cytochrome P450
390 genes from the fly *Drosophila mettleri* that were putatively involved in the detoxification
391 of alkaloids from two cactus species. Heidel-Fischer et al. (2010) investigated genetic
392 variation in the *NSP* gene of three European and one US American population of *P. rapae*
393 and found high amounts of amino acid diversity but little evidence for local adaptation.
394 Here, we investigated two genes that underlie the counteradaptation of Pierid butterflies
395 against the GLS-myrosinase defense system of their host plants for potential
396 microevolutionary dynamics in response to host plant usage along a latitudinal cline in
397 Japan. In addition to the crop-dependent *P. rapae*, we included two species that feed on
398 wild Brassicaceae, *P. melete* and *P. napi*. As in the previous study of Heidel-Fischer et
399 al. (2010), we found no evidence for adaptation of *P. rapae NSP*-like genes to local
400 communities of wild Brassicaceae. In contrast, both *P. melete* and *P. napi* showed a
401 significant and positive correlation between the diversity of the *NSP* (but not the *MA*)
402 gene product and host plant diversity (Fig. 4). Both species harbored two distinct *NSP*
403 alleles, which shaped the observed correlations in a species-specific manner.
404
405 Hokkaido populations of *P. napi* had a major *NSP* allele that was distinct from alleles
406 found outside of Hokkaido (Fig. 3a). This difference mirrors genome-wide differentiation
407 between both populations (Fig. 2), which suggests two independent colonization events

by *P. napi*, one in the north and one in the south of Japan. Differences in the number of acquired polymorphisms in both groups of populations suggest that colonization in the south happened earlier than colonization in the north. Based on the analysis of the mitochondrial *ND5* gene, a recent study concluded that *P. napi* populations on the island of Hokkaido fell into a (south-) western and a (north-) eastern group of populations (Shinkawa et al. 2003), with a hybridization zone in between. Indeed, both Hokkaido populations had elevated genome-wide diversity (Table 1a), and both populations (and in particular Rusutsu) showed signs of gene flow from the south (Fig. 2). Furthermore, *MA* alleles prevalent in the south had apparently replaced the original *MA* alleles in Rusutsu and partially also those in Yubari (Fig. 3d). However, despite gene flow and in contrast to the replacement of *MA* alleles, both Hokkaido populations retained their original *NSP* allele (Fig. 3d). Furthermore, *NSPs* from Hokkaido and from central and southern Japan differed by more non-synonymous substitutions than expected (Table 3), indicative of diversifying selection. However, all amino acid changes that distinguished Hokkaido and non-Hokkaido populations occurred in non-Hokkaido lineage (Fig. 6), indicating that strong directional selection happened in this group, potentially as an adaptive response to novel sets of glucosinolate profiles after colonization. In contrast, both Hokkaido populations displayed very low *NSP* nucleotide diversity, suggesting strong conservation of NSP function. This finding is somewhat surprising because *Rorippa sylvestris*, the preferred host plant species of *P. napi* on the island of Hokkaido (Table S4), was introduced only about 50 years ago (Ohsaki, Ohata, Sato, & Rausher, 2020). It would therefore be helpful to characterize the functional properties of *NSP* gene products from both metapopulations to better understand the contrasting patterns of selection on *NSP* of Japanese *P. napi*.

In contrast to *P. napi*, *P. melete* did not show any indication of population substructure (Fig. 2). In this species, NSP diversity correlated with host plant diversity, which varied along a latitudinal gradient (Fig. 4). Two distinct NSP alleles were present in areas of high host plant diversity, *i.e.*, in central Japan, whereas areas of low host plant diversity were biased towards one of the two alleles, up to fixation of one allele in the Hokkaido populations (Fig. 3d). Although we did not find evidence for diversifying selection, Tajima's *D* was substantially higher for *NSP* than for *MA* or genome-wide sequence, in particular in the first exon (Fig. 5b). This region indeed harbored an elevated number of linked synonymous and non-synonymous SNPs at intermediate frequency (Fig. 5c), suggesting balancing selection acting on this exon across Japan. Balancing selection is an important mechanism to maintain polymorphisms in natural populations, helping organisms to adapt to environments varying in space and time (Hedrick, 2006). Considering the heterogeneous chemical interactions between herbivores and their host plants in the field, it is not surprising to see such polymorphisms maintained in genes involved in these interactions within natural populations. Indeed, there is evidence on the plant side for balancing selection on genes causing variation in chemical defense (Carley et al., 2021; Kroymann, Donnerhacke, Schnabelrauch, & Mitchell-Olds, 2003; Mitchell-Olds & Schmitt, 2006). By contrast, such patterns have previously not been described for counteradaptive traits of herbivorous insects, mainly due to a lack of information about the genetic bases of such traits. Therefore, *NSP* in Japanese *P. melete* may represent the first example for balancing selection acting on an insect counteradaptation against host plant defenses. Of course, functional data would be helpful to substantiate this idea.

For MA we did not find a clear evolutionary trend. *P. napi* had two distinct MA alleles but these did not show any indication for divergent or balancing selection (Tables 2, 3, Figs. 4, 5). Instead, the comparison of non-synonymous *versus* synonymous diversity suggested that *MA*s were more strongly conserved than *NSP*s in all three *Pieris* species (Fig. 5d). This is consistent with a previous study (Okamura, Sato, Tsuzuki, Murakami, et al., 2019), which found evidence for positive selection only for *NSP* but not for *MA* among different *Pieris* species. Although the functional differences between *NSP* and *MA* are still unknown, both genes displayed distinct expression patterns in response to different GLSs (Okamura, Sato, Tsuzuki, Sawada, et al., 2019). This could indicate that these ancient homologs play different roles in GLS detoxification, with the more evolutionary conserved MAs detoxifying more common GLSs, and NSP acting on rarer GLSs.

Remarkably, the same genes show substantial differences in polymorphism patterns and evolutionary trajectories in three different *Pieris* species that populate Japan. This finding can be partially (but not entirely) explained by different host plant usage. The observed different evolutionary dynamics between *NSP* and its paralog *MA* shows that gene duplication and potential functional differentiation appears to be important for adaptation to chemically heterogeneous host plants in the field. Demographic history, and in particular timing and number of colonization events may also play a major role. Therefore, it would be interesting to investigate the same genes, *NSP* and *MA*, in Pierids collected along a latitudinal cline from the eastern coasts of the Asian continent for comparison.

Acknowledgements

We thank Domenica Schnabelrauch for her help in sequencing. We are grateful to the members of the community ecology laboratory of Chiba University and Ryo Maeda for their help with field work. We also thank Hanna Heidel-Fischer for comments and suggestions. We thank Emily Wheeler, Boston, for editorial assistance. This work was supported by JSPS KAKENHI Grant-in-Aid for JSPS Fellowships 15J00320 and JSPS Overseas Research Fellowships 202060676 for YO, and Max-Planck-Gesellschaft.

References

Agerbirk, N., & Olsen, C. E. (2012). Glucosinolate structures in evolution. *Phytochemistry*, *77*, 16–45. doi: 10.1016/j.phytochem.2012.02.005

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*, 139–155. doi: 10.1086/285907

Bernatchez, L., & Landry, C. (2003). MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? *Journal of Evolutionary Biology*, *16*(3), 363–377. doi: 10.1046/j.1420-9101.2003.00531.x

Blažević, I., Montaut, S., Burčul, F., Olsen, C. E., Burow, M., Rollin, P., & Agerbirk, N. (2020). Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. *Phytochemistry*, *169*(July 2019), 112100. doi: 10.1016/j.phytochem.2019.112100

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bono, J. M., Matzkin, L. M., Castrezana, S., & Markow, T. A. (2008). Molecular evolution and population genetics of two *Drosophila mettleri* cytochrome P450 genes involved in host plant utilization. *Molecular Ecology*, *17*(13), 3211–3221. doi: 10.1111/j.1365-294X.2008.03823.x

Carley, L. N., Mojica, J. P., Wang, B., Chen, C., Lin, Y., Prasad, K. V. S. K., … Mitchell-Olds, T. (2021). Ecological factors influence balancing selection on leaf chemical profiles of a wildflower. *Nature Ecology & Evolution*, 1–10. doi: 10.1038/s41559-021-01486-0

Catchen, J. M. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354.Stacks

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

516 Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., & Bernatchez, L. (2007). Clinal
517 variation in MHC diversity with temperature: Evidence for the role of host-
518 pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, *61*(9),
519 2154–2164. doi: 10.1111/j.1558-5646.2007.00178.x
520 Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., …
521 Wheat, C. W. (2015). The butterfly plant arms-race escalated by gene and genome
522 duplications. *Proceedings of the National Academy of Sciences of the United States*
523 *of America*, *112*, 8362–8366. doi: 10.1073/pnas.1503926112
524 Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald-
525 Kreitman test: A website to detect selection by comparing different classes of
526 DNA sites. *Nucleic Acids Research*, *36*(Web Server issue), 157–162. doi:
527 10.1093/nar/gkn337
528 Excoffier, L., & Lischer, H. E. L. L. (2010). Arlequin suite ver 3.5: A new series of
529 programs to perform population genetics analyses under Linux and Windows.
530 *Molecular Ecology Resources*, *10*(3), 564–567. doi: 10.1111/j.1755-
531 0998.2010.02847.x
532 Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and
533 distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*,
534 *56*(1), 5–51.
535 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate
536 surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–
537 4315. doi: 10.1002/joc.5086
538 Fischer, H. M., Wheat, C. W., Heckel, D. G., & Vogel, H. (2008). Evolutionary origins
539 of a novel host plant detoxification gene in butterflies. *Molecular Biology and*
540 *Evolution*, *25*(5), 809–820. doi: 10.1093/molbev/msn014
541 Friberg, M., & Wiklund, C. (2019). Host preference variation cannot explain
542 microhabitat differentiation among sympatric *Pieris napi* and *Pieris rapae*
543 butterflies. *Ecological Entomology*, *44*(4), 571–576. doi: 10.1111/een.12728
544 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological
545 association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. doi:
546 10.1111/2041-210X.12382
547 Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and
548 efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983.
549 doi: 10.1534/genetics.113.160572

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(43), 18054–18061.

Gloss, A. D., Vassão, D. G., Hailey, A. L., Nelson Dittrich, A. C., Schramm, K., Reichelt, M., … Whiteman, N. K. (2014). Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the Drosophilidae. *Molecular Biology and Evolution*, *31*(9), 2441–2456. doi: 10.1093/molbev/msu201

Grishin, N. V., Shen, J., Cong, Q., Kinch, L. N., Borek, D., & Otwinowski, Z. (2016). Complete genome of *Pieris rapae*, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. *F1000Research*, *5*(5), 2631. doi: 10.12688/f1000research.9765.1

Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology*, *57*, 303–333. doi: 10.1146/annurev.arplant.57.032905.105228

Hedrick, P. W. (2006). Genetic polymorphism in heterogeneous environments: The age of Genomics. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 67–93. doi: 10.1146/annurev.ecolsys.37.091305.110132

Heidel-Fischer, H. M., Kirsch, R., Reichelt, M., Ahn, S.-J., Wielsch, N., Baxter, S. W., … Kroymann, J. (2019). An Insect Counteradaptation against Host Plant Defenses Evolved through Concerted Neofunctionalization. *Molecular Biology and Evolution*, *36*(5), 930–941. doi: 10.1093/molbev/msz019

Heidel-Fischer, H. M., Vogel, H., Heckel, D. G., & Wheat, C. W. (2010). Microevolutionary dynamics of a macroevolutionary key innovation in a Lepidopteran herbivore. *BMC Evolutionary Biology*, *10*, 60. doi: 10.1186/1471-2148-10-60

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. doi: 10.1093/molbev/msj030

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. doi: 10.1093/molbev/mst010

Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., & Mitchell-Olds, T. (2001). Genetic control of natural variation in arabidopsis glucosinolate accumulation. *Plant Physiology*, *126*(2), 811–825. doi: 10.1104/pp.126.2.811

585 Kroymann, J., Donnerhacke, S., Schnabelrauch, D., & Mitchell-Olds, T. (2003).
586 Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus.
587 *Proceedings of the National Academy of Sciences*, *100*(Supplement 2), 14587–
588 14592. doi: 10.1073/pnas.1734046100
589 Kubota, Y., Shiono, T., & Kusumoto, B. (2015). Role of climate and geohistorical
590 factors in driving plant richness patterns and endemicity on the east Asian
591 continental islands. *Ecography*, *38*(6), 639–648. doi: 10.1111/ecog.00981
592 Li, W., Schuler, M. A., & Berenbaum, M. R. (2003). Diversification of furanocoumarin-
593 metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity
594 and substrate encounter rate. *Proceedings of the National Academy of Sciences of*
595 *the United States of America*, *100*(24), 14593–14598. doi:
596 10.1073/pnas.1934643100
597 Mantel, N. (1967). The detection of disease clustering and a generalized regression
598 approach. *Cancer Research*, *27*(February), 209–220.
599 McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus
600 in *Drosophila*. *Nature*, *353*, 412–414.
601 Mitchell-Olds, T., & Schmitt, J. (2006). Genetic mechanisms and evolutionary
602 significance of natural variation in *Arabidopsis*. *Nature*, *441*(7096), 947–952. doi:
603 10.1038/nature04878
604 Ohsaki, N., & Sato, Y. (1994). Food plant choice of *Pieris* butterflies as a trade-off
605 between parasitoid avoidance and quality of plants. *Ecology*, *75*(1), 59–68. doi:
606 10.2307/1939382
607 Ohsaki, N. (1979). Comparative population studies of three *Pieris* butterflies, *P. rapae*,
608 *P. melete* and *P. napi*, living in the same area—I. Ecological requirements for
609 habitat resources in the adults. *Researches on Population Ecology*, *20*(2), 278–296.
610 doi: 10.1007/BF02512633
611 Ohsaki, N., Ohata, M., Sato, Y., & Rausher, M. D. (2020). Host plant choice
612 determined by reproductive interference between closely related butterflies. *The*
613 *American Naturalist*, *196*(4), 512–523. doi: 10.1086/710211
614 Okamura, Y., Sato, A., Tsuzuki, N., Murakami, M., Heidel-Fischer, H., & Vogel, H.
615 (2019). Molecular signatures of selection associated with host plant differences in
616 *Pieris* butterflies. *Molecular Ecology*, *28*(22), 4958–4970. doi: 10.1111/mec.15268
617 Okamura, Y., Sato, A., Tsuzuki, N., Sawada, Y., Hirai, M. Y., Heidel-Fischer, H., …
618 Vogel, H. (2019). Differential regulation of host plant adaptive genes in Pieris
619 butterflies exposed to a range of glucosinolate profiles in their host plants.
620 *Scientific Reports*, *9*(1), 7256. doi: 10.1038/s41598-019-43703-8

621 Okamura, Y., Sawada, Y., Hirai, M. Y., & Murakami, M. (2016). Effects of different
622 secondary metabolite profiles in plant defense syndromes on specialist and
623 generalist herbivores. *Entomological Science*, *19*, 97–103. doi: 10.1111/ens.12172
624 Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn,
625 D., … Wagner, H. (2017). *Community Ecology Package, Package ' vegan .'* doi:
626 ISBN 0-387-95457-0
627 Olsen, C. E., Huang, X. C., Hansen, C. I. C., Cipollini, D., Ørgaard, M., Matthes, A., …
628 Agerbirk, N. (2016). Glucosinolate diversity within a phylogenetic framework of
629 the tribe Cardamineae (Brassicaceae) unraveled with HPLC-MS/MS and NMR-
630 based analytical distinction of 70 desulfoglucosinolates. *Phytochemistry*, *132*, 33–
631 56. doi: 10.1016/j.phytochem.2016.09.013
632 Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time
633 RT-PCR. *Nucleic Acids Research*, *29*(9), 2002–2007. doi: 10.1093/nar/29.9.e45
634 Phillips, S., Dudík, M., & Schapire, R. (2004). A maximum entropy approach to species
635 distribution modeling. *Proceedings of the Twenty-First International Conference*
636 *on Machine Learning*, 655–662. doi: 10.1145/1015330.1015412
637 Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017).
638 Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7),
639 887–893. doi: 10.1111/ecog.03049
640 Prasad, K. V. S. K., Song, B.-H. H., Olson-Manning, C., Anderson, J. T., Lee, C.-R. R.,
641 Schranz, M. E., … Mitchell-Olds, T. (2012). A gain-of-function polymorphism
642 controlling complex traits and fitness in nature. *Science*, *337*(6098), 1081–1084.
643 doi: 10.1126/science.1221636
644 Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T., & Kroymann, J. (2002).
645 Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences*
646 *of the United States of America*, *99*(17), 11223–11228. doi:
647 10.1073/pnas.172112899
648 Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P.,
649 Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence
650 polymorphism analysis of large data sets. *Molecular Biology and Evolution*,
651 *34*(12), 3299–3302. doi: 10.1093/molbev/msx248
652 Ryan, S. F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dincă, V., … Shoemaker,
653 D. (2019). Global invasion history of the agricultural pest butterfly *Pieris rapae*
654 revealed with genomics and citizen science . *Proceedings of the National Academy*
655 *of Sciences*, *116*(40), 20015–20024. doi: 10.1073/pnas.1907492116

656 Shinkawa, T., Yagi, T., Fujii H., Kondoh, K. (2003). Molecular phylogeny and
657 divergence of *Pieris* group. *Butterfly DNA research society News letter*, 10, 14-17.
658 Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation
659 extensions of the mantel test of matrix correspondence. *Systematic Zoology*, *35*(4),
660 627–632. doi: 10.2307/2413122
661 Sokal, R. R. (1979). Testing statistical significance of geographic variation patterns.
662 *Systematic Zoology*, *28*(2), 227. doi: 10.2307/2412528
663 Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by
664 DNA polymorphism. *Genetics*, *123*(3), 585–595. doi: 10.1093/genetics/123.3.585
665 Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6:
666 Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and*
667 *Evolution*, *30*(12), 2725–2729. doi: 10.1093/molbev/mst197
668 Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M.
669 (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids*
670 *Research*, *35*(SUPPL.2), 71–74. doi: 10.1093/nar/gkm306
671 Wheat, C. W., Vogel, H., Wittstock, U., Braby, M. F., Underwood, D., & Mitchell-
672 Olds, T. (2007). The genetic basis of a plant-insect coevolutionary key innovation.
673 *Proceedings of the National Academy of Sciences of the United States of America*,
674 *104*(51), 20427–20431. doi: 10.1073/pnas.0706229104
675 Windsor, A. J., Reichelt, M., Figuth, A., Svatos, A., Kroymann, J., Kliebenstein, D. J.,
676 … Mitchell-Olds, T. (2005). Geographic and evolutionary diversification of
677 glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae).
678 *Phytochemistry*, *66*(11), 1321–1333. doi: 10.1016/j.phytochem.2005.04.016
679 Wittstock, U., Agerbirk, N., Stauber, E. J., Olsen, C. E., Hippler, M., Mitchell-Olds, T.,
680 … Vogel, H. (2004). Successful herbivore attack due to metabolic diversion of a
681 plant chemical defense. *Proceedings of the National Academy of Sciences of the*
682 *United States of America*, *101*(14), 4859–4864. doi: 10.1073/pnas.0308007101
683 Wittstock, U., & Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis* era.
684 *Trends in Plant Science*, *7*(6), 263–270.
685
686

Data Availability Statement

The RAD-seq short read data have been deposited in the EBI short read archive (SRA) with the following project URL: http://www.ebi.ac.uk/ena/data/view/PRJEB47778.

Author Contributions

Y.O., A.S., L.K., and A.J.N. carried out the laboratory work. Y.O., M.M., H.V., and J.K. conceived, designed and coordinated the study. Y.O., H.V. and J.K. wrote the manuscript. All authors, drafted parts of the manuscript, gave approval for publication and agree to be accountable for the content.

Tables and Figures

Table 1

(a) Population genetic parameters from genome-wide RAD-seq in three *Pieris* species

Species	Population	private alleles	Individuals	Obs Het	π
P. melete	ALL		57.6	0.0029	0.0042
	Yubari (Hokkaido)	188	8.27	0.0009	0.0012
	Oshamanbe (Hokkaido)	273	9.33	0.0011	0.0013
	Fukushima	181	9.12	0.0010	0.0012
	Nagano	160	8.52	0.0010	0.0012
	Chiba	195	6.20	0.0011	0.0014
	Nara	202	8.81	0.0010	0.0013
	Tokushima	201	8.30	0.0010	0.0013
P. napi	ALL		59.6	0.0034	0.0055
	Yubari (Hokkaido)	269	8.82	0.0013	0.0018
	Rusutsu (Hokkaido)	153	9.18	0.0015	0.0018
	Fukushima	72	8.06	0.0011	0.0015
	Nagano	57	8.94	0.0011	0.0014
	Kanagawa	56	8.59	0.0012	0.0014
	Nara	38	8.61	0.0013	0.0014
	Miyazaki	84	9.39	0.0012	0.0014
P. rapae	ALL		63.7	0.0039	0.0055
	Yubari (Hokkaido)	38	9.55	0.0017	0.0017
	Rusutsu (Hokkaido)	51	8.22	0.0013	0.0017
	Fukushima	36	7.44	0.0013	0.0016
	Nagano	74	8.64	0.0015	0.0019
	Chiba	121	8.19	0.0014	0.0021
	Nara	71	9.45	0.0016	0.0018
	Tokushima	80	8.52	0.0015	0.0019
	Yonaguni	83	6.34	0.0012	0.0015

(b) Population genetic parameters for *NSP* and *MA* in three *Pieris* species

Species	Sampling sites	Segregating sites (S)		nucleotide diversity (π)		amino acid diversity		Tajima's *D*	
		NSP	*MA*	*NSP*	*MA*	*NSP*	*MA*	*NSP*	*MA*
P. melete	ALL	34	44	0.0047	0.0048	0.0070	0.0056	0.5611	-0.2495
	Yubari (Hokkaido)	5	21	0.0007	0.0035	0.0009	0.0032	-1.0353	-0.5370
	Oshamanbe (Hokkaido)	4	23	0.0005	0.0045	0.0003	0.0043	-1.2447	0.1832
	Fukushima	23	20	0.0052	0.0044	0.0080	0.0065	0.7376	0.8698
	Nagano	18	24	0.0047	0.0041	0.0070	0.0049	1.7720	-0.4326
	Chiba	22	16	0.0045	0.0039	0.0065	0.0031	0.4060	0.9582
	Nara	24	31	0.0049	0.0059	0.0069	0.0076	0.3140	0.0867
	Tokushima	27	29	0.0059	0.0052	0.0091	0.0069	0.7735	-0.2318
P. napi	ALL	66	56	0.0098	0.0061	0.0156	0.0087	0.3700	-1.1531
	Yubari (Hokkaido)	5	39	0.0006	0.0056	0.0015	0.0088	-1.3882	-1.2381
	Rusutsu (Hokkaido)	2	11	0.0004	0.0023	0.0012	0.0022	0.1203	0.4869
	Fukushima	27	17	0.0052	0.0029	0.0043	0.0022	-0.1066	-0.3511

	Nagano	29	14	0.0056	0.0021	0.0039	0.0021	0.1320	-0.9809
	Kanagawa	27	15	0.0053	0.0029	0.0041	0.0029	0.1615	0.2393
	Nara	26	14	0.0051	0.0028	0.0043	0.0031	0.1779	0.3970
	Miyazaki	19	10	0.0037	0.0021	0.0012	0.0020	0.1825	0.5833
P. rapae	ALL	43	34	0.0039	0.0038	0.0037	0.0025	-0.7042	-0.0747
	Yubari (Hokkaido)	22	18	0.0029	0.0040	0.0032	0.0033	-1.4854	0.8915
	Rusutsu (Hokkaido)	21	20	0.0033	0.0042	0.0030	0.0020	-1.0311	0.3914
	Fukushima	23	18	0.0047	0.0033	0.0038	0.0024	0.1521	-0.0542
	Nagano	21	15	0.0029	0.0032	0.0028	0.0025	-1.2568	0.6447
	Chiba	18	16	0.0033	0.0038	0.0030	0.0025	-0.6758	1.0538
	Nara	26	14	0.0043	0.0027	0.0038	0.0016	-0.6641	0.2119
	Tokushima	28	22	0.0053	0.0041	0.0052	0.0012	-0.2214	-0.0399
	Yonaguni	18	22	0.0036	0.0046	0.0037	0.0033	-0.4853	0.0754

703

704 **Table 2 Partial Mantel tests for NSP and MA in three *Pieris* species**

Comparison	Species	*r*	*P*-value
Partial Mantel Test			
NSP AA Fst, Plant dissimilarity | RAD Fst	***P. melete***	**0.444**	**0.039**
	P. napi	**0.436**	**0.010**
	P. rapae	-0.047	0.594
MA AA Fst, Plant dissimilarity | RAD Fst	*P. melete*	0.1405	0.258
	P. napi	-0.318	0.867
	P. rapae	-0.183	0.877

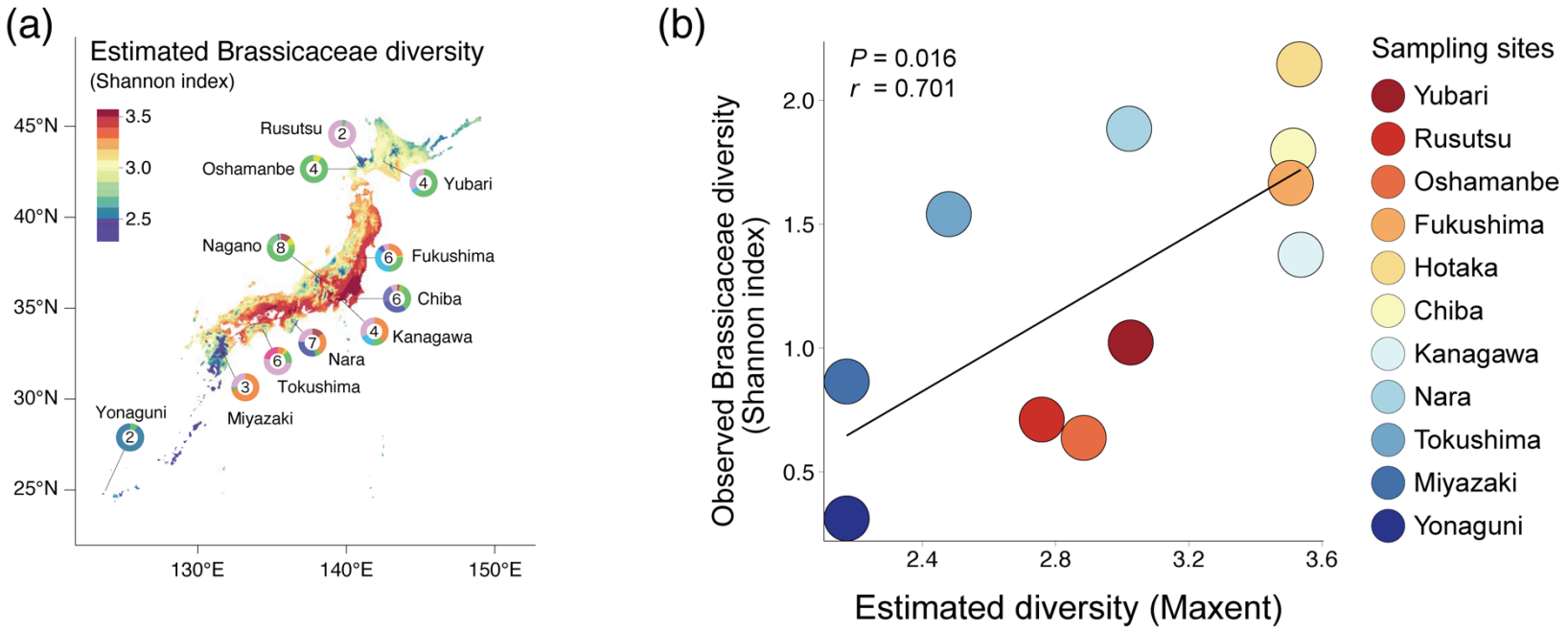
705 NSP AA Fst and MA AA Fst: Fst-like divergences based on amino acid sequences. Plant dissimilarity: Bray-Curtis plant community
706 dissimilarity. Statistically significant results are in bold ($P \leq 0.05$).
707

Table 3 Population-based tests for diversifying selection (McDonald–Kreitman)

Comparison	Gene	Polymorphisms		Substitutions		NI	X^2	*P*
		syn	nsyn	syn	nsyn			
P. melete	*NSP*	20	13	6.05	4	0.982	0	0.981
P. napi	*NSP*	29	21	4.02	19.16	0.152	10.568	**0.001**
	MA	22	17	8.1	13.07	0.521	1.807	0.178

syn: synonymous, nsyn: nonsynonymous, NI: Neutrality Index

713 **Fig. 1**

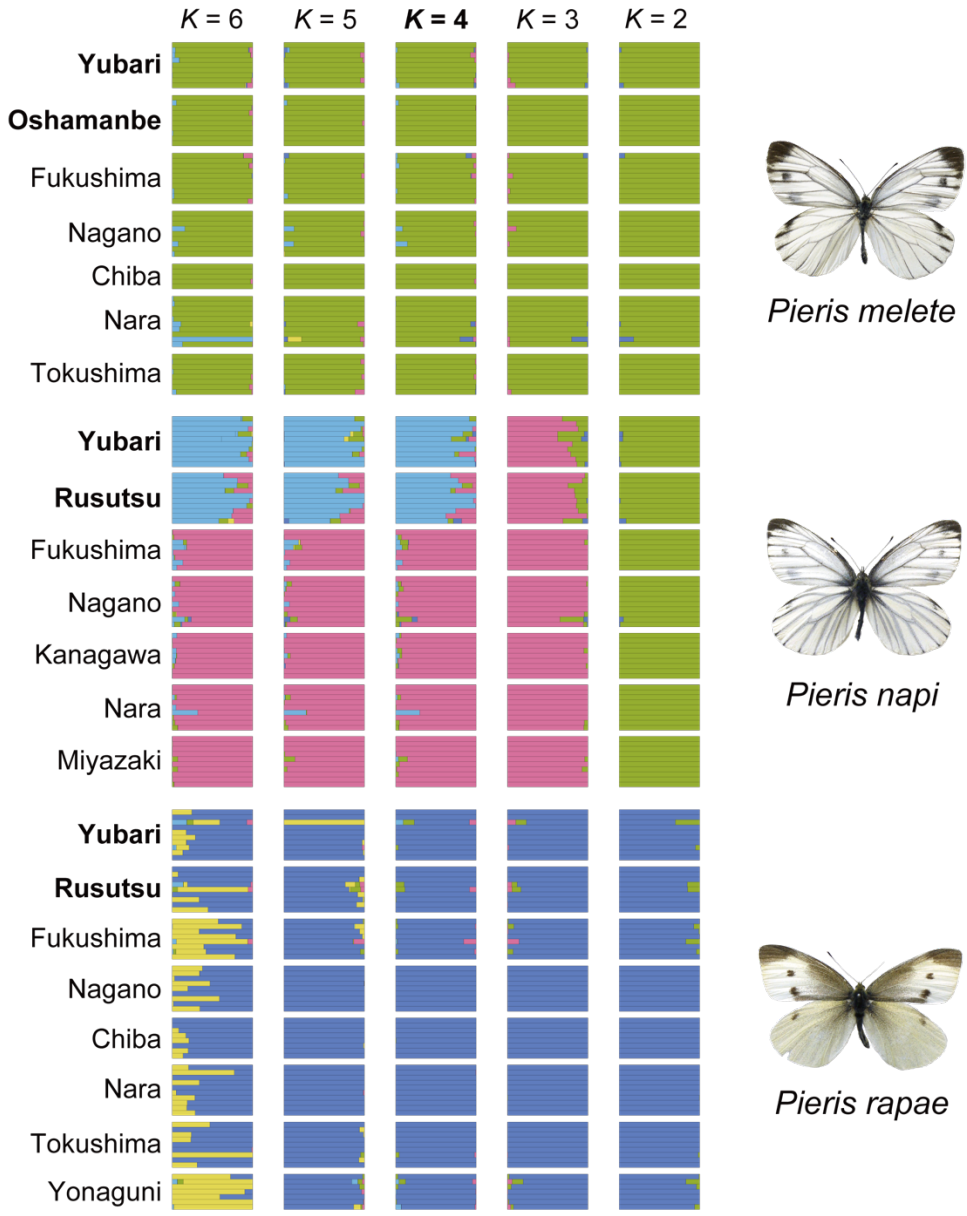


714

715 **Fig. 1** (a) Maxent-based diversity in Brassicaceae across Japan. The color code indicates the Shannon index for estimated diversity. Pie
716 charts show number and composition of observed Brassicaceae genera at each sampling site. (b) Pearson correlation between estimated
717 and observed diversity of Brassicaceae at each sampling site, with the north-south order of sampling sites depicted to the right of the
718 panel. Diversity in the Brassicaceae is higher in central Japan than in northern or southern Japan.

719

720 **Fig. 2**

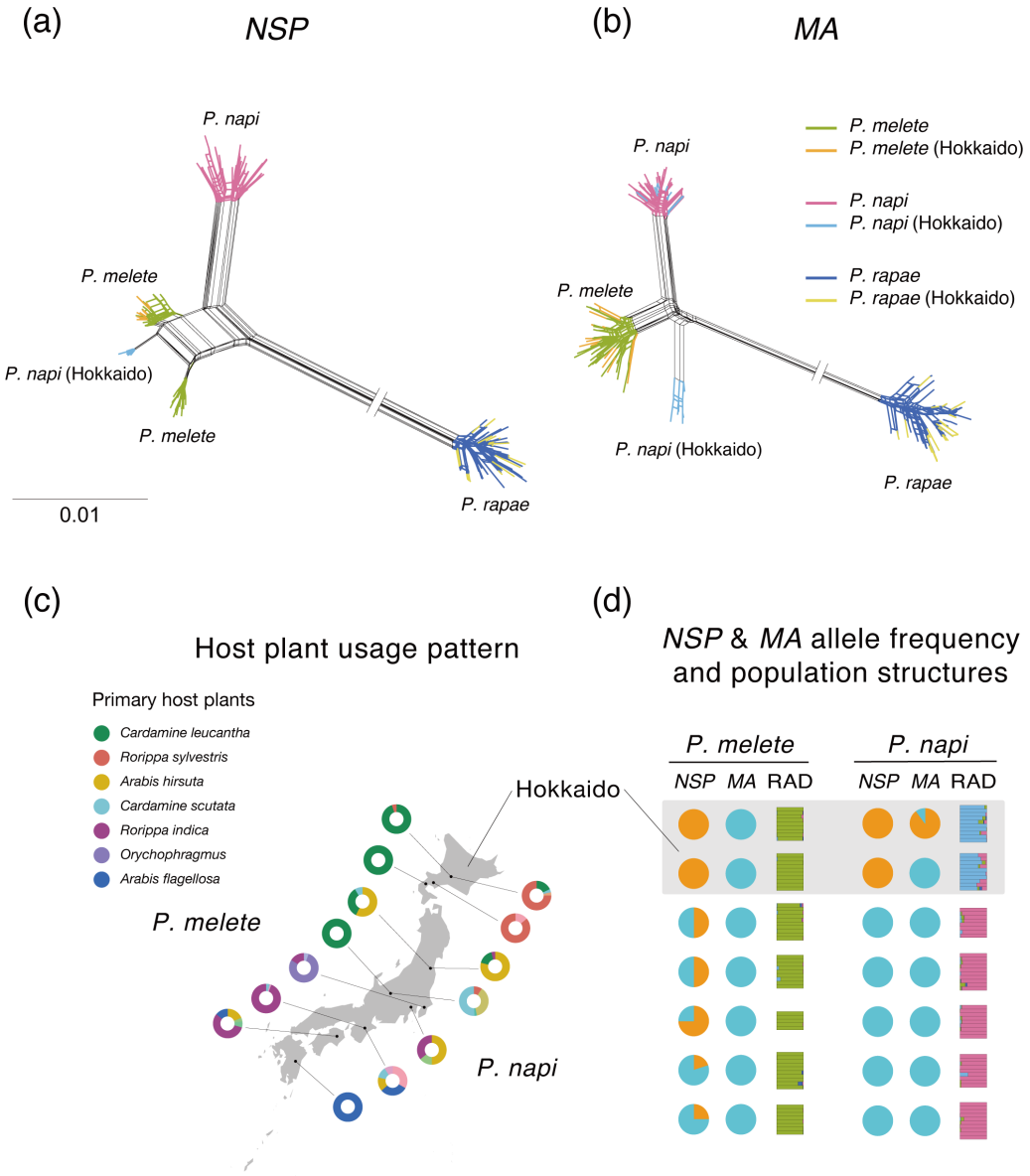


721

722 **Fig. 2** Inferred population genetic structures of three *Pieris* spp. across Japan for
723 different clustering numbers (*K*), based on RAD-seq data. The cluster number most
724 highly supported by the cross-entropy criterion (*K* = 4) is shown in bold. Colored
725 horizontal lines correspond to individual samples, and blocks correspond to a
726 population at a particular sampling site, ordered from north to south with Hokkaido sites
727 in bold.

728

729 **Fig. 3**



730

731 **Fig. 3** NeighborNet networks of (a) *NSP*s and (b) *MA*s from three *Pieris* species across
732 Japan, with different colors for species and for Hokkaido populations. Note that *P.*
733 *melete* and *P. napi* both have two major *NSP* alleles. Note further that *P. napi* has two
734 major *MA* alleles. (c) Observed host plant usage patterns of *P. melete* and *P. napi* at
735 sampling sites. (d) Allele frequencies of major *NSP* and *MA* alleles in comparison with
736 genome-wide genetic structure of *P. melete* and *P. napi* populations at each sampling
737 site, ordered from north to south.

Fig. 4

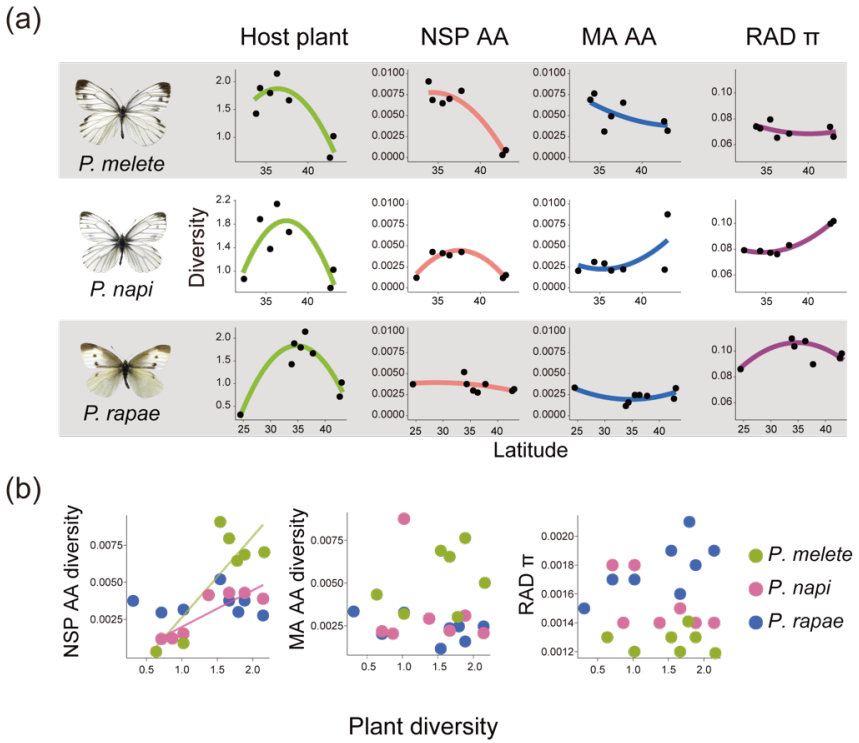
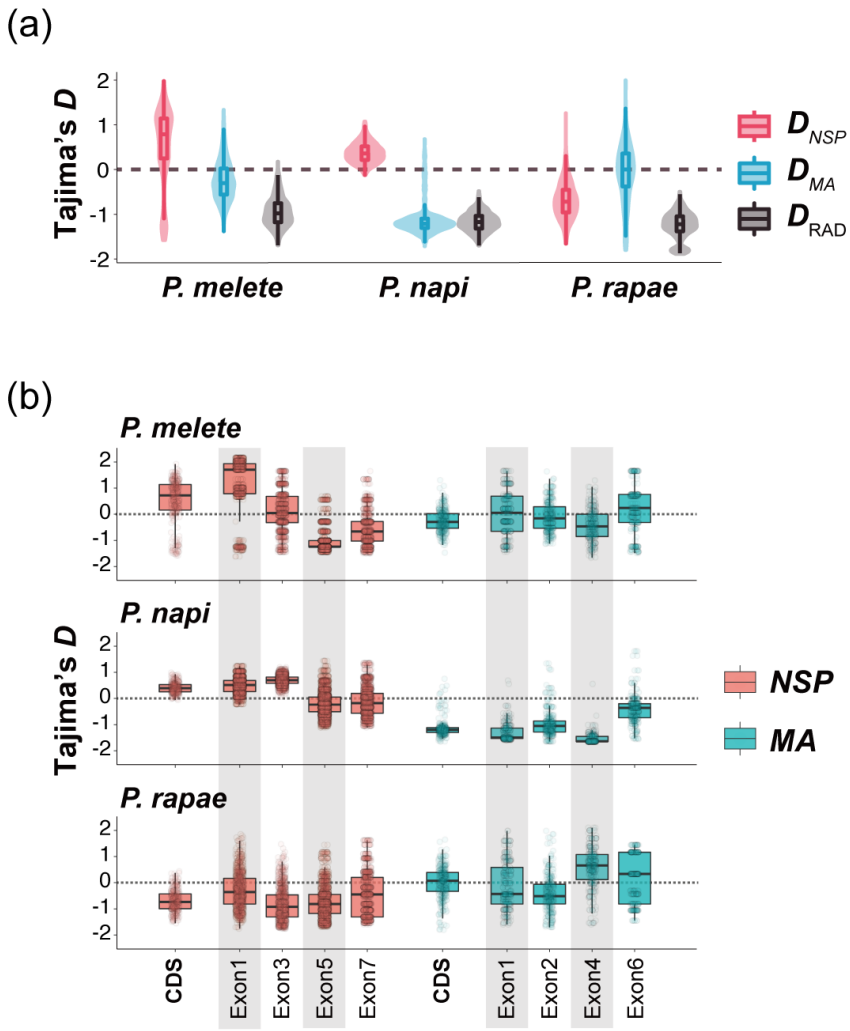


Fig. 4 (a) Host plant diversity (Shannon index), NSP amino acid diversity (NSP AA), MA amino acid diversity (MA AA), and genome-wide diversity (RAD π) within three *Pieris* species at sampling sites ordered according to latitude. Black dots correspond to sampling sites, with colored lines showing regression curves. (b) Correlations between NSP AA, MA AA, or RAD π with host plant diversity in three *Pieris* species. Note the significant positive correlation ($P < 0.05$) between plant diversity and NSP AA diversity in *P. melete* and *P. napi*, but not in *P. rapae*.

746 **Fig. 5**



747

748

749

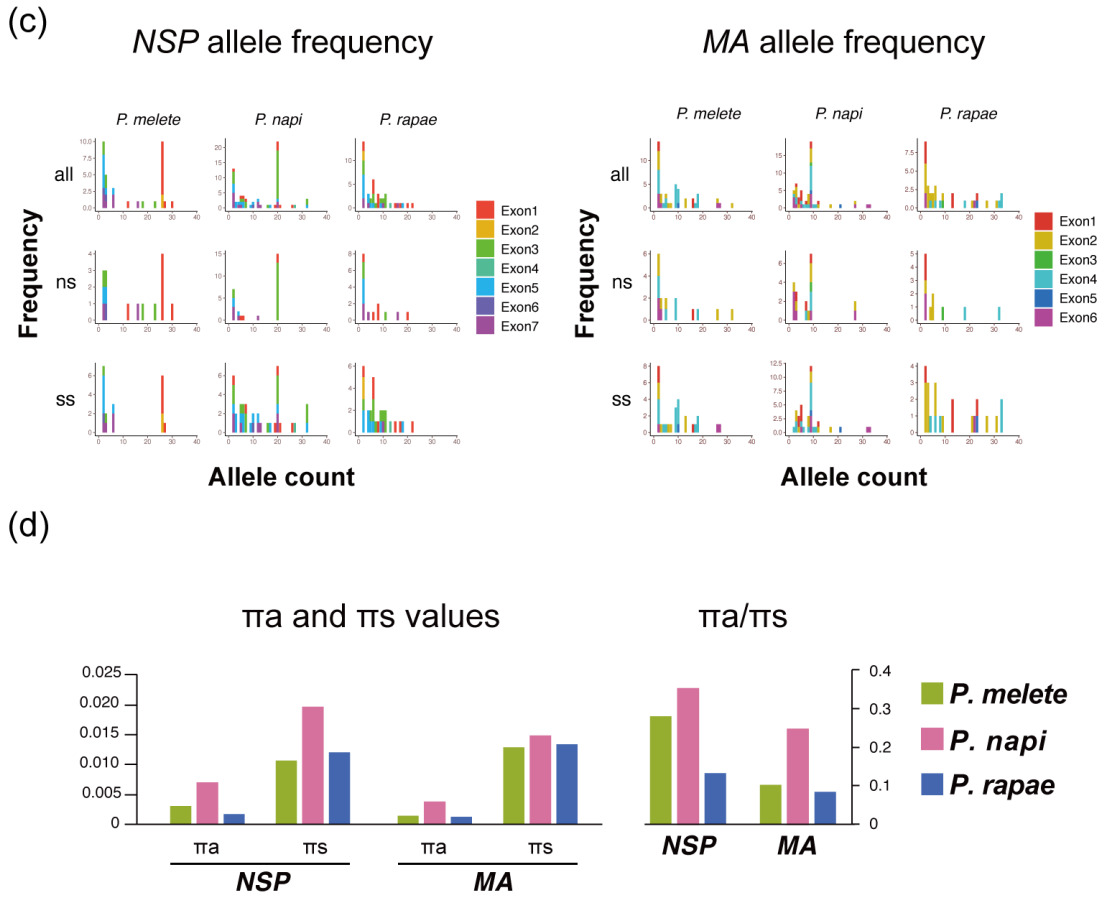
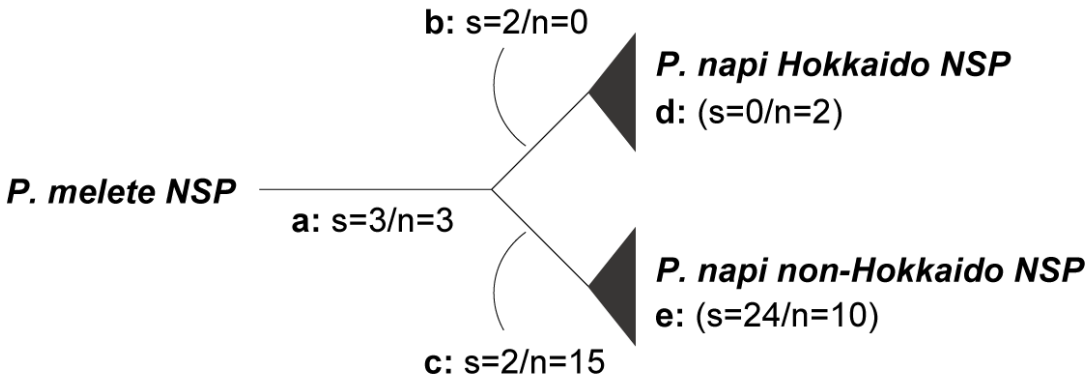


Fig. 5 (a) Species-wide distributions of Tajima's *D* for *NSP*, *MA*, and genome-wide data (RAD). Note high values of *D* for *NSP* in *P. melete* and *P. napi.* (b) Exon-based Tajima's *D* distribution for *NSP* and *MA*. Only exons > 200 bp are shown. (c) Allele frequency spectra of *NSP* and *MA* based on exon color coding. Exon 1 of *P. melete* and exon 3 of *P. napi NSP*s accumulate positions with intermediate allele frequencies. (d) Non-synonymous (π_a) and synonymous nucleotide diversity (π_s) at the species level and ratio of non-synonymous to synonymous diversity (π_a/π_s) of *NSP* and *MA* in three *Pieris* species.

761



762

763

764 **Fig. 6 Divergent selection on *NSP* in populations of *Pieris napi*.**

765 Synonymous (s) and non-synonymous (n) mutations in *NSP*s on branches of a
766 simplified molecular phylogeny consisting of two *Pieris napi* populations and *Pieris*
767 *melete*. Polymorphisms segregating in each population are shown in brackets. A very
768 high number of non-synonymous mutations on the non-Hokkaido branch of *Pieris napi*
769 (branch c) indicates positive selection in the lineage leading to the non-Hokkaido
770 populations of *P. napi*. In comparison, *NSP* of the Hokkaido populations appears to be
771 highly conserved.

772

773 **Supplementary Tables and Figures**

774 **Table S1 19 climate variables used for Maxent**

Variables	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation ×100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

775

776

777 **Table S2 Sampling sites and observed number of Brassicaceae**

Sampling Sites	Yubari	Rusutsu	Oshamanbe	Fukushima	Nagano	Chiba	Kanagawa	Nara	Tokushima	Miyazaki	Yonaguni
latitude	43.0	42.7	42.6	37.8	36.3	35.5	35.5	34.3	33.9	32.4	24.5
longitude	142.1	140.8	140.3	140.6	137.8	140.2	139.2	136.0	134.1	131.2	123.0
Plant diversity	1.0224	0.7115	0.6374	1.6662	2.1451	1.7968	1.3765	1.8850	1.5404	0.8650	0.3125
Plant Species											
Arabidopsis thaliana	0	0	0	0	60	20	0	0	0	0	0
Arabidopsis halleri	0	0	0	0	0	0	0	83	0	0	0
Arabidopsis kamchatica	0	0	0	0	46	0	0	0	0	0	0
Arabis flagellosa	0	0	0	0	0	0	0	150	0	245	0
Arabis hirsuta	0	0	0	78	0	0	72	12	12	0	0
Barbarea orthoceras	6	0	15	7	72	0	0	0	4	0	0
Brassica napus	0	0	0	0	51	45	0	10	3	0	0
Cardamine impatiens	0	0	0	0	12	4	7	0	5	0	0
Cardamine kiusiana	0	0	0	0	0	0	0	0	14	0	0
Cardamine leucantha	438	20	130	42	140	0	0	0	0	0	0
Cardamine regeliana	0	0	0	0	81	125	0	0	0	0	0
Cardamine occulta	30	0	12	45	210	60	18	31	0	13	25
Draba nemorosa	0	0	0	0	140	0	0	0	0	0	0
Eutrema japonicum	0	0	0	130	15	0	30	0	0	0	0
Eutrema tenue	47	0	0	0	0	0	0	0	0	0	0
Lepidium virginicum	0	0	0	0	0	0	0	10	0	0	240
Nasturtium officinale	0	0	0	23	0	200	0	174	60	0	0
Orychophragmus violaceus	0	0	0	0	0	140	0	0	0	0	0
Raphanus sativus	0	0	0	0	8	0	0	0	0	0	0
Rorippa dubia	0	0	0	0	0	0	0	62	0	58	0
Rorippa indica	0	0	2	25	0	40	58	75	70	27	0
Rorippa palustris	0	101	0	0	21	10	0	10	0	0	0
Rorippa sylvestris	262	317	0	0	0	0	0	0	0	0	0
Turritis glabra	0	0	0	0	0	0	0	0	30	0	0

778

779 **Table S3 Primers used in this study**

Primers	Description	Sequence
Primers for cloning and RFLP		
M13_F	M13 Forward	GTTTTCCCAGTCACGAC
M13_R	M13 Reverse	CAGGAAACAGCTATGAC
ND5_F	*ND5* Forward	CCTGTTTCTGCTTTAGTTCA
ND5_R	*ND5* Reverse	AATATDAGGTATAAATCATAT
NSP_RFLP_F	*NSP* Forward primer for RFLP (NSPmn_F)	ATGAAAGCTGTTGTAGTCTTATTAGC
NSP_RFLP_R	*NSP* Reverse primer for RFLP	GTCTTGTACTTCGGACTCCTTTT
Primers for sequencing		
NSPmn_F	*NSP* Forward for *P. melete* and *P. napi*	ATGAAAGCTGTTGTAGTCTTATTAGC
NSPmn_R	*NSP* Reverse for *P. melete* and *P. napi*	CTGTCCGTAAAGAGCAGGTAC
NSPmn_int1	*NSP* internal1 for *P. melete* and *P. napi*	GCTTAGATGCCTTGTCAAAGACT
NSPmn_int2	*NSP* internal2 for *P. melete* and *P. napi*	AATAGCGTGGTCGTTCTTAGC
NSPr_F	*NSP* Forward for *P. rapae*	ATGAAAGGTGTTGTAGTCTTCTTAG
NSPr_R	*NSP* Reverse for *P. rapae*	TTACTGTCCGTAAAGGGCA
NSPr_int1	*NSP* internal1 for *P. rapae*	CTCTGGAAGAACGAAGCATT
NSPr_int2	*NSP* internal2 for *P. rapae*	AACTCGGCTAGTCCTGCTTTC-
MAmn_F	*MA* Forward for *P. melete* and *P. napi*	ATGAAGACAACAATAGTGCTCCTAAG
MAr_F	*MA* Forward for *P. rapae*	ATGAAGACAACAATAGTGCTCCTC
MAmnr_R	*MA* Reverse for all the three species	TTATTGTCCCCAGAGGGTTG
MAmn_int1	*MA* internal1 for *P. melete* and *P. napi*	CCTGTTTGAGGAATTACTTCCA
MAmn_int2	*MA* internal2 for *P. melete* and *P. napi*	ATAGGCGTGTTTTGTCCGAAA
MAr_int1	*MA* internal1 for *P. rapae*	CTTCCAGCATTCTTCGGAC
MAr_int1	*MA* internal2 for *P. rapae*	TCTGATTGCACGATGATGTCC
Primers for RT-qPCR		
NSPm_qPCR_F	Forward qPCR primer for *P. melete NSP*	AATTGGCGGCTTTATACACG
NSPm_qPCR_R	Reverse qPCR primer for *P. melete NSP*	TTCTTTCCTTCGGCACTTGT
MAm_qPCR_F	Forward qPCR primer for *P. melete MA*	TGTTGCTAACGCACTGGAAG
MAm_qPCR_R	Reverse qPCR primer for *P. melete MA*	CCCTCCAACGCAGTAATGAT
Ef1am_qPCR_F	Forward qPCR primer for *P. melete Ef1a*	AGGAATTGCGTCGTGGTTAC
Ef1am_qPCR_R	Reverse qPCR primer for *P. melete Ef1a*	GCAAGCAATGTGAGCTGTGT

780

781

Table S4 Host plant usage patterns of three *Pieris* species at different sampling sites

Columns show 'observed number of larvae' and 'checked number of plants'. Hyphens (-) indicate absence of a plant species at the sampling site.

P. melete

Plant sp./Sampling Sites	Yubari	Rusutsu	Oshamanbe	Fukushima	Nagano	Chiba	Kanagawa	Nara	Tokushima	Miyazaki	Yonaguni	**SUM**
Arabidopsis halleri	-	-	-	-	-	-	-	0/83	-	-	-	0/83
Arabidopsis kamchatica	-	-	-	-	0/46	-	-	-	-	-	-	0/46
Arabidopsis thaliana	-	-	-	-	0/60	0/20	-	-	-	-	-	0/80
Arabis flagellosa	-	-	-	-	-	-	-	0/150	-	0/245	-	0/395
Arabis hirsuta	-	-	-	27/78	-	-	10/72	0/12	28/12	-	-	65/174
Barbarea orthoceras	0/6	-	0/15	0/7	0/72	-	-	-	0/4	-	-	0/104
Brassica napus	-	-	-	-	0/51	0/45	-	0/10	0/3	-	-	0/109
Cardamine impatiens	-	-	-	-	0/12	0/4	9/7	-	6/5	-	-	15/28
Cardamine kiusiana	-	-	-	-	-	-	-	-	10/14	-	-	10/14
Cardamine leucantha	19/438	0/20	24/130	16/42	35/140	-	-	-	-	-	-	94/770
Cardamine occulta	0/30	-	0/12	4/45	0/210	0/60	0/18	2/31	-	0/13	0/25	6/444
Cardamine regeliana	-	-	-	-	0/81	0/125	-	-	-	-	-	0/206
Draba nemorosa	-	-	-	-	0/140	-	-	-	-	-	-	0/140
Eutrema japonicum	-	-	-	0/130	0/15	-	0/30	-	-	-	-	0/175
Eutrema tenue	0/47	-	-	-	-	-	-	-	-	-	-	0/47
Lepidium virginicum	-	-	-	-	-	-	-	0/10	-	-	0/240	0/250
Nasturtium officinale	-	-	-	0/23	-	6/200	-	0/174	0/6	-	-	6/403
Orychophragmus violaceus	-	-	-	-	-	88/140	-	-	-	-	-	88/140
Raphanus sativus	-	-	-	-	0/8	-	-	-	-	-	-	0/8
Rorippa dubia	-	-	-	-	-	-	-	1/62	-	0/58	-	1/120
Rorippa indica	-	-	0/2	0/25	-	18/40	25/58	48/75	85/70	0/27	-	176/297
Rorippa palustris	-	0/101	-	-	0/21	0/10	-	0/10	-	-	-	0/142
Rorippa sylvestris	1/262	0/317	-	-	-	-	-	-	-	-	-	1/579
Turritis glabra	-	-	-	-	-	-	-	-	21/30	-	-	21/30

SUM	20/783	0/438	24/159	47/350	35/856	112/644	44/185	51/617	150/144	0/343	0/265	483/4784

P. napi

Plant sp./Sampling Sites	Yubari	Rusutsu	Oshamanbe	Fukushima	Nagano	Chiba	Kanagawa	Nara	Tokushima	Miyazaki	Yonaguni	**SUM**
Arabidopsis halleri	-	-	-	-	-	-	-	11/83	-	-	-	11/83
Arabidopsis kamchatica	-	-	-	-	7/46	-	-	-	-	-	-	7/46
Arabidopsis thaliana	-	-	-	-	0/60	0/20	-	-	-	-	-	0/80
Arabis flagellosa	-	-	-	-	-	-	-	10/150	-	34/245	-	44/395
Arabis hirsuta	-	-	-	19/78	-	-	7/72	5/12	0/12	-	-	31/174
Barbarea orthoceras	0/6	-	0/15	0/7	29/72	-	-	-	0/4	-	-	29/104
Brassica napus	-	-	-	-	0/51	0/45	-	0/10	0/3	-	-	0/109
Cardamine impatiens	-	-	-	-	0/12	0/4	2/7	-	0/5	-	-	2/28
Cardamine kiusiana	-	-	-	-	-	-	-	-	0/14	-	-	0/14
Cardamine leucantha	5/438	0/20	0/130	4/42	0/140	-	-	-	-	-	-	9/770
Cardamine occulta	1/30	-	0/12	0/45	39/210	0/60	0/18	4/31	-	0/13	0/25	44/444
Cardamine regeliana	-	-	-	-	2/81	0/125	-	-	-	-	-	2/206
Draba nemorosa	-	-	-	-	0/140	-	-	-	-	-	-	0/140
Eutrema japonicum	-	-	-	0/130	0/15	-	0/30	-	-	-	-	0/175
Eutrema tenue	1/47	-	-	-	-	-	-	-	-	-	-	1/47
Lepidium virginicum	-	-	-	-	-	-	-	0/10	-	-	0/240	0/250
Nasturtium officinale	-	-	-	0/23	-	0/200	-	0/174	0/6	-	-	0/403
Orychophragmus violaceus	-	-	-	-	-	0/140	-	-	-	-	-	0/140
Raphanus sativus	-	-	-	-	0/8	-	-	-	-	-	-	0/8
Rorippa dubia	-	-	-	-	-	-	-	3/62	-	0/58	-	3/120
Rorippa indica	-	-	0/2	1/25	-	0/40	5/58	0/75	0/70	0/27	-	6/297
Rorippa palustris	-	6/101	-	-	0/21	0/10	-	0/10	-	-	-	6/142
Rorippa sylvestris	21/262	37/317	-	-	-	-	-	-	-	-	-	58/579
Turritis glabra	-	-	-	-	-	-	-	-	0/30	-	-	0/30
SUM	28/783	43/438	0/159	24/350	77/856	0/644	14/185	33/617	0/144	34/343	0/265	253/4784

P. rapae

Plant sp./Sampling Sites	Yubari	Rusutsu	Oshamanbe	Fukushima	Nagano	Chiba	Kanagawa	Nara	Tokushima	Miyazaki	Yonaguni	**SUM**
Arabidopsis halleri	-	-	-	-	-	-	-	0/83	-	-	-	0 /83
Arabidopsis kamchatica	-	-	-	-	0/46	-	-	-	-	-	-	0 /46
Arabidopsis thaliana	-	-	-	-	0/60	0/20	-	-	-	-	-	0 /80
Arabis flagellosa	-	-	-	-	-	-	-	0/150	-	0/245	-	0 /395
Arabis hirsuta	-	-	-	0/78	-	-	0/72	0/12	0/12	-	-	0 /174
Barbarea orthoceras	2/6	-	0/15	0/7	0/72	-	-	-	0/4	-	-	2 /104
Brassica napus	-	-	-	-	35/51	34/45	-	0/10	0/3	-	-	69 /109
Cardamine impatiens	-	-	-	-	0/12	0/4	0/7	-	0/5	-	-	0 /28
Cardamine kiusiana	-	-	-	-	-	-	-	-	25/14	-	-	25 /14
Cardamine leucantha	0/438	0/20	0/130	0/42	0/140	-	-	-	-	-	-	0 /770
Cardamine occulta	0/30	-	0/12	0/45	0/210	0/60	0/18	0/31	-	0/13	0/25	0 /444
Cardamine regeliana	-	-	-	-	0/81	0/125	-	-	-	-	-	0 /206
Draba nemorosa	-	-	-	-	0/140	-	-	-	-	-	-	0 /140
Eutrema japonicum	-	-	-	0/130	0/15	-	0/30	-	-	-	-	0 /175
Eutrema tenue	0/47	-	-	-	-	-	-	-	-	-	-	0 /47
Lepidium virginicum	-	-	-	-	-	-	-	0/10	-	-	8/240	8 /250
Nasturtium officinale	-	-	-	1/23	-	0/200	-	13/174	6/6	-	-	20 /403
Orychophragmus violaceus	-	-	-	-	-	0/140	-	-	-	-	-	0 /140
Raphanus sativus	-	-	-	-	4/8	-	-	-	-	-	-	4 /8
Rorippa dubia	-	-	-	-	-	-	-	0/62	-	0/58	-	0 /120
Rorippa indica	-	-	0/2	11/25	-	0/40	0/58	7/75	13/70	0/27	-	31 /297
Rorippa palustris	-	20/101	-	-	0/21	0/10	-	0/10	-	-	-	20 /142
Rorippa sylvestris	29/262	1/317	-	-	-	-	-	-	-	-	-	30 /579
Turritis glabra	-	-	-	-	-	-	-	-	0/30	-	-	0 /30
SUM	31 /783	21 /438	0 /159	12 /350	39 /856	34 /644	0 /185	20 /617	44 /144	0 /343	8 /265	209 /4784

Table S5 Statistics for *NSP* and *MA*

(a) *NSP*

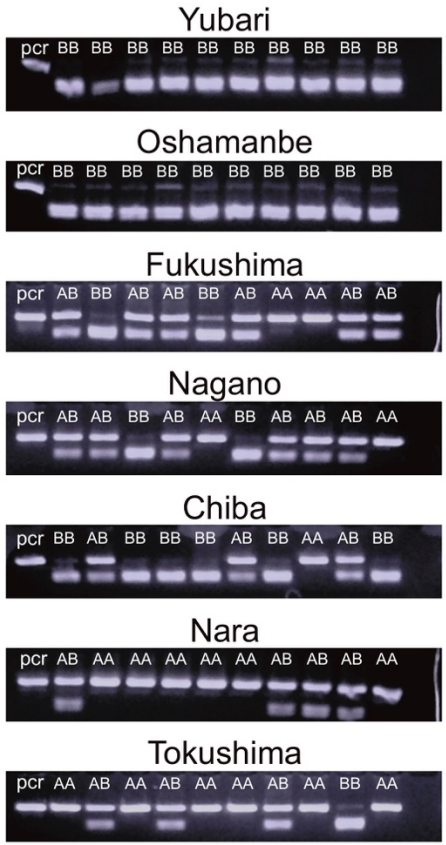
Species	Sampling sites	Sequences	Number of sites	S sites	*NSP* π	AA diversity	Tajima's	*P* value
P. melete	ALL	69	1860	34	0.0047	0.0070	0.561	-
	Yubari (Hokkaido)	10	1860	5	0.0007	0.0009	-1.035	0.188
	Oshamanbe	10	1860	4	0.0005	0.0003	-1.245	0.116
	Fukushima	9	1860	23	0.0052	0.0080	0.738	0.803
	Nagano	10	1860	18	0.0047	0.0070	1.772	0.980
	Chiba	10	1860	22	0.0045	0.0065	0.406	0.685
	Nara	10	1857-1860	24	0.0049	0.0069	0.314	0.662
	Tokushima	10	1857-1860	27	0.0059	0.0091	0.773	0.815
P. napi	ALL	69	1860	66	0.0098	0.0156	0.370	-
	Yubari (Hokkaido)	10	1860	5	0.0006	0.0015	-1.388	0.094
	Rusutsu (Hokkaido)	10	1860	2	0.0004	0.0012	0.120	0.686
	Fukushima	9	1860	27	0.0052	0.0043	-0.107	0.483
	Nagano	10	1860	29	0.0056	0.0039	0.132	0.584
	Kanagawa	10	1860	27	0.0053	0.0041	0.162	0.610
	Nara	10	1860	26	0.0051	0.0043	0.178	0.614
	Miyazaki	10	1860	19	0.0037	0.0012	0.182	0.615
P. rapae	ALL	74	1869	43	0.0039	0.0037	-0.704	-
	Yubari (Hokkaido)	10	1869	22	0.0029	0.0032	-1.485	0.062
	Rusutsu (Hokkaido)	9	1869	21	0.0033	0.0030	-1.031	0.164
	Fukushima	9	1869	23	0.0047	0.0038	0.152	0.587
	Nagano	10	1869	21	0.0029	0.0028	-1.257	0.105
	Chiba	8	1869	18	0.0033	0.0030	-0.676	0.281
	Nara	10	1869	26	0.0043	0.0038	-0.664	0.265
	Tokushima	10	1869	28	0.0053	0.0052	-0.221	0.518
	Yonaguni (Okinawa)	8	1869	18	0.0036	0.0037	-0.485	0.427

(b) *MA*

Species	Sampling sites	Sequences	Number of sites	S sites	*MA* π	AA diversity	Tajima's *D*	*P* value
P. melete	ALL	68	1896	44	0.0048	0.0056	-0.2495	-
	Yubari (Hokkaido)	10	1896	21	0.0035	0.0032	-0.5370	0.314
	Oshamanbe	10	1896	23	0.0045	0.0043	0.1832	0.617
	Fukushima	10	1896	20	0.0044	0.0065	0.8698	0.840
	Nagano	10	1896	24	0.0041	0.0049	-0.4326	0.348
	Chiba	8	1896	16	0.0039	0.0031	0.9582	0.859
	Nara	10	1896	31	0.0059	0.0076	0.0867	0.579
	Tokushima	10	1896	29	0.0052	0.0069	-0.2318	0.437
P. napi	ALL	70	1896	56	0.0061	0.0087	-1.1531	-
	Yubari (Hokkaido)	10	1896	39	0.0056	0.0088	-1.2381	0.134
	Rusutsu	10	1896	11	0.0023	0.0022	0.4869	0.711
	Fukushima	10	1896	17	0.0029	0.0022	-0.3511	0.389
	Nagano	10	1896	14	0.0021	0.0021	-0.9809	0.176
	Kanagawa	10	1896	15	0.0029	0.0029	0.2393	0.632
	Nara	10	1896	14	0.0028	0.0031	0.3970	0.687
	Miyazaki	10	1896	10	0.0021	0.0020	0.5833	0.747
P. rapae	ALL	76	1896	34	0.0038	0.0025	-0.0747	-
	Yubari (Hokkaido)	10	1896	18	0.0040	0.0033	0.8915	0.844
	Rusutsu	9	1896	20	0.0042	0.0020	0.3914	0.687
	Fukushima	10	1896	18	0.0033	0.0024	-0.0542	0.508
	Nagano	10	1896	15	0.0032	0.0025	0.6447	0.772
	Chiba	9	1896	16	0.0038	0.0025	1.0538	0.881
	Nara	10	1896	14	0.0027	0.0016	0.2119	0.625
	Tokushima	10	1896	22	0.0041	0.0012	-0.0399	0.513
	Yonaguni	8	1896	22	0.0046	0.0033	0.0754	0.548

Fig. S1(a) PCR-RFLP of *NSP* exon 1 from seven populations of *P. melete*. "pcr" refers to an undigested PCR product, corresponding to 'type A'. Note that 'type B' bands in the agarose gels consist of two restriction fragments of similar size. (b) Exon 1 sequences of type A and type B *NSP* with *Hin*cII restriction site in type B. SNPs distinguishing types A and B of *NSP* are shown in red.

(a)



(b)

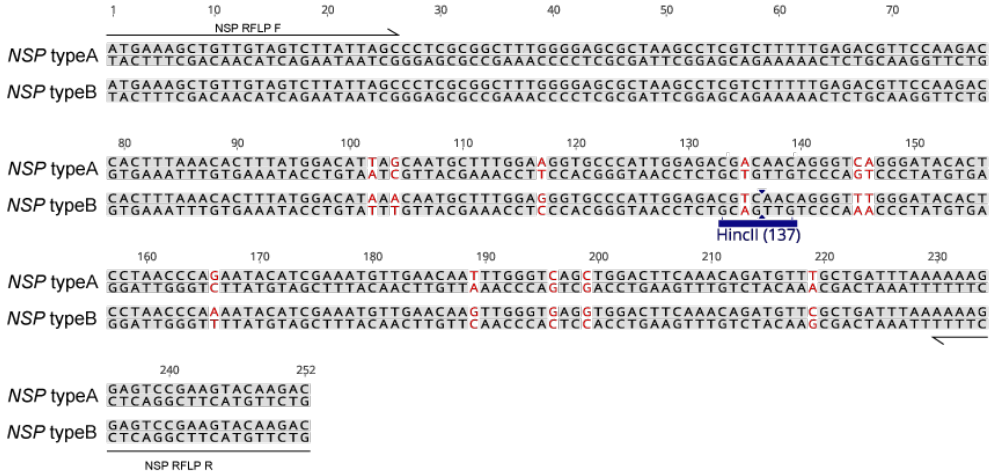
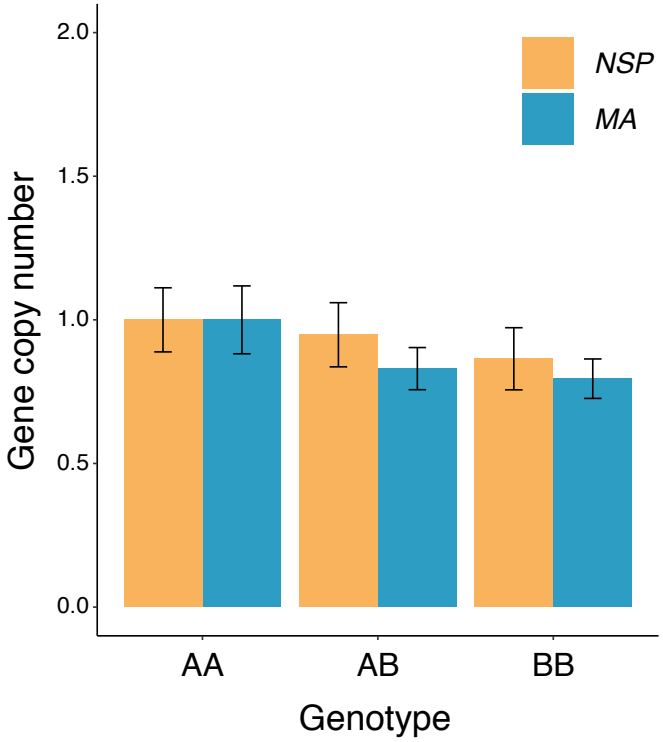
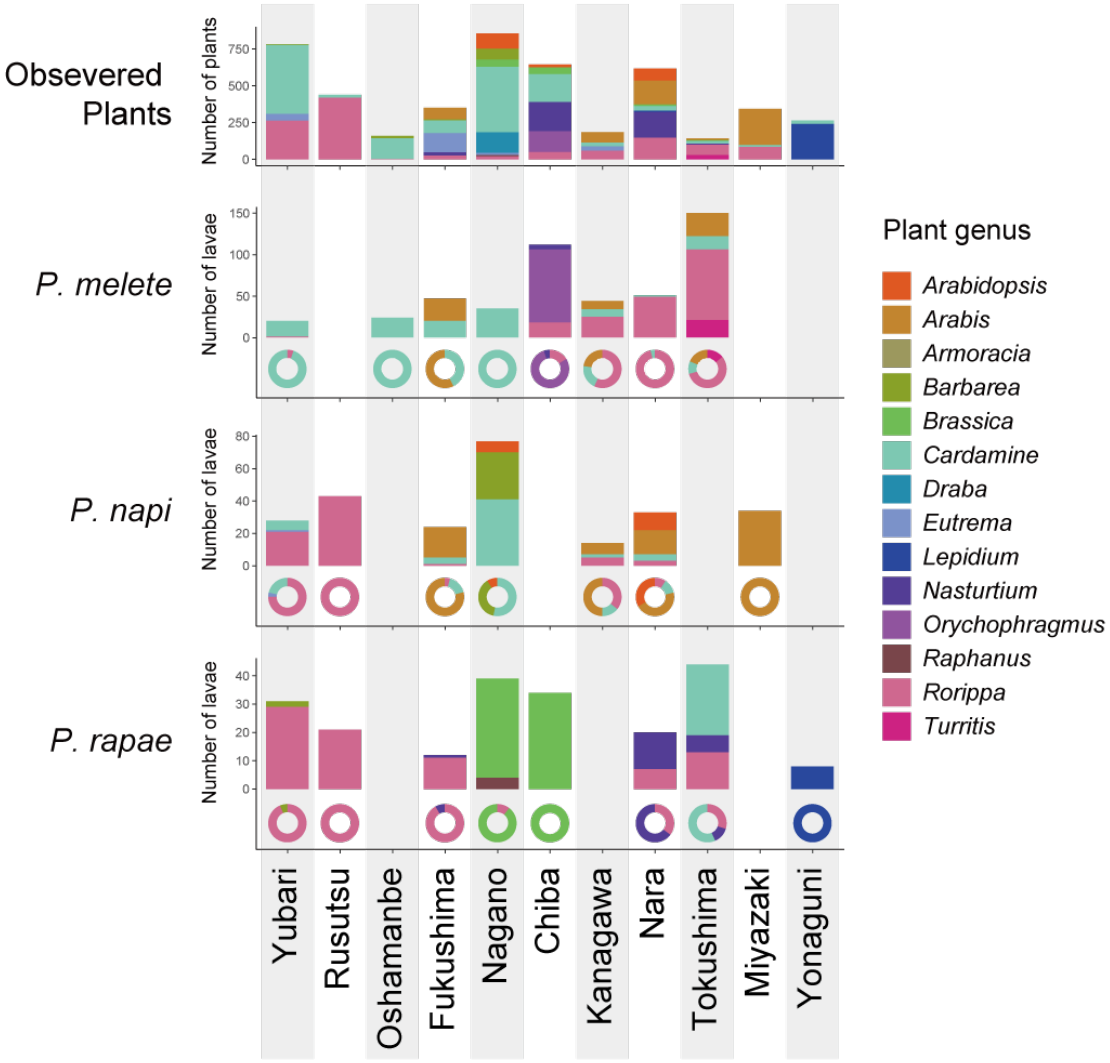


Fig. S2 *NSP* and *MA* gene copy numbers in *P. melete*. Genotypes (AA, AB and BB) are the same as in Fig. S1. RT-qPCR was done with N = 8 replicates per genotype and *Ef1α* as a reference gene. Data were standardized with mean of genotype AA set as 1. Bars shows means (± SE). Genotypes do not differ in *NSP* (yellow bars) or *MA* (blue bars) gene copy numbers.



800 **Fig. S3** Patterns of host plant use of *Pieris* species in the field. The plant species data
801 were summarized to genus level for visualization. The pie charts show composition of
802 plant genus and each bar show number of individuals (either plants or larvae).



803
804