Received: 3 October 2021 | Revised: 1 March 2022 | Accepted: 23 March 2022

DOI: 10.1111/mec.16447

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

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

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, Jena, Germany

[2]Community Ecology Lab, Faculty of Science, Chiba University, Chiba, Japan

[3]Research Administration Office, Kyoto University, Kyoto, Japan

[4]Faculty of Agriculture, Ryukoku University, Shiga, Japan

[5]Institute for Advanced Biosciences, Keio University, Yamagata, Japan

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

Correspondence
Yu Okamura, Department of Insect Symbiosis, Max Planck Institute for Chemical Ecology, Jena, Germany.
Email: yokamura@ice.mpg.de

Funding information
Japan Society for the Promotion of Science, Grant/Award Number: 15J00320 and 202060676; Max-Planck-Gesellschaft

Handling Editor: Jeremy Yoder

Abstract

Herbivorous insects have evolved counteradaptations to overcome the chemical defences 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 particular, it is not known whether and how host plant diversity shapes diversity in insect counteradaptations. 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 paralogue major allergen (MA) in their gut to overcome the highly diversified glucosinolate–myrosinase defence 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. Both *P. melete* and *P. napi* 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 defence.

KEYWORDS
arms race, host plant adaptation, insects, selection

1 | 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 defences of their host plants. Available evidence suggests that specialist herbivores acquired key innovations that enabled them to colonize their host plants by circumventing host plant chemical defences (Berenbaum et al., 1996; Ratzka et al., 2002; Wheat et al., 2007; Wittstock et al., 2004). Defences 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), and

several studies indicate that this variation is non-neutral (Benderoth et al., 2006, 2009; Bidart-Bouzat & Kliebenstein, 2008; Kroymann et al., 2003). This raises the question of whether there is similar variation in insect counteradaptations. However, in the field 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 defence (Wheat et al., 2007; Wittstock et al., 2004; Wittstock & Halkier, 2002). GLSs are hydrolysed 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 defence 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). Previous genetic and molecular analyses in *Anthocharis cardamines*, another Pierid species feeding on Brassicaceae, indicated that major allergen (MA), an ancient NSP paralogue, has NSP-like activity (Edger et al., 2015). Of note, *MA* transcript levels in the gut of *Pieris* larvae are upregulated in response to certain types of GLS (Okamura, Sato, Tsuzuki, Sawada, et al., 2019). Together, *NSP* and *MA* constitute the *NSP*-like gene family, a key innovation that enabled Pierid butterflies to colonize their host plants from the order Brassicales (Edger et al., 2015; Fischer et al., 2008; Wittstock et al., 2004).

GLSs are a highly diverse group of plant secondary metabolites, with about 135 different structures identified in the Brassicales (Blažević et al., 2020). GLS profiles vary both quantitatively and qualitatively (Agerbirk & Olsen, 2012; Fahey et al., 2001; Olsen et al., 2016) and *Pieris* species typically use several genera within the Brassicaceae as their hosts (Friberg & Wiklund, 2019; Ohsaki, 1979; Ohsaki & Sato, 1994; Okamura, Sato, Tsuzuki, Sawada, et al., 2019; Okamura et al., 2016). A previous study with *Pieris rapae*, the cabbage white butterfly, found no evidence for local adaptation of *NSP* genes in Europe or the USA (Heidel-Fischer et al., 2010). However, their study did not address patterns of host plant use of *P. rapae*. This species depends mainly on cruciferous crop plants and is a considerable pest on *Brassica* (Grishin et al., 2016; Ryan et al., 2019). Hence, *P. rapae* may not be suitable to detect microevolutionary patterns that are associated with different host usage in the field. Instead, it is necessary to study insect species that depend on wild host plants to possibly detect microevolutionary patterns in the genes that shape the insect's counteradaptations against their host plants' defences.

In this study, we focus on three *Pieris* species, *Pieris melete*, *Pieris napi* and *P. rapae*, that co-occur across most of Japan to compare the microevolutionary dynamics of NSP and MA among butterfly species with different host usage patterns. *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 contrast, *P. rapae* uses mostly cultivated Brassicaceae. In Japan, plant communities vary substantially along a north–south cline (Kubota et al., 2015). Because this holds true also for the latitudinal distribution of wild Brassicaceae, we expected that the plant community cline should affect host usage patterns of *P. melete* and *P. napi* more strongly than of *P. rapae* feeding on cultivated plants.

We sampled local populations of these three *Pieris* species across Japan and acquired data on 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 et al., 2007). Using those approaches we expected to identify microevolutionary patterns of NSP or MA potentially correlating with local host plant usage among *Pieris* species strongly associated with wild Brassicaceae plants. However, we also expected weaker or no such correlation in crop-dependent *P. rapae* due to a much more uniform host usage across Japan. In addition, we also tested whether *NSP* and *MA* showed any differences in molecular evolutionary patterns along a host plant community cline.

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. In *P. melete*, two *NSP* alleles were maintained by balancing selection across Japan, and allele frequency was correlated with local host plant diversity. *P. napi*, on the other hand, 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. Together, our study indicates that microevolution of counteradaptive traits has an important role for herbivores to adapt to heterogeneous chemical defences in the host plants that they encounter in the field.

2 | MATERIALS AND METHODS

2.1 | Selection of sampling sites

To determine suitable sampling sites that cover different host plant communities, we estimated the diversity of Brassicaceae across Japan with MAXENT version 3.4.1 (Phillips et al., 2004, 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 11 sampling sites across Japan to cover a gradient of local host plant diversities for each butterfly species (Table S2).

2.2 | 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 *Pieris 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 the 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 reached the 3rd instar and then dissected the larvae to separate guts from the rest of the bodies. We processed larvae larger than 3rd instar immediately. We stored guts in RNAlater (Qiagen) and body rests at −80°C. We excluded larvae that contained parasitoid wasps from further analyses.

2.3 | RFLP-based identification of *Pieris* butterfly species

To determine the identity of the sampled larvae, we used restriction fragment length polymorphisms (RFLPs) in the mitochondrial *ND5* gene (GenBank accession nos.: LC090587–LC090590). *P. melete ND5* has a *Hin*cII site, and *P. napi ND5* has a *Hin*fI and a *Hin*dIII site, whereas *P. rapae ND5* lacks these restriction sites. We amplified *ND5* directly from larval bodies, using MIGHTYAMP DNA POLYMERASE Version 3 (TaKaRa) and *ND5* universal primers (Table S3), followed by restriction and separation on 2% agarose gels. To verify that our species assignments worked correctly, we used the same RFLP technique for adult males previously classified as *P. melete* ($N = 24$), *P. napi* ($N = 16$) or *P. rapae* ($N = 24$) based on differences in the shape of the androconium. *ND5*-based identification matched androconium-based identification for all 64 individuals.

2.4 | *NSP* and *MA* amplification, cloning and sequencing

We randomly picked 10 larvae per species and sampling site to sequence *NSP* and *MA* genes. In total, we used 70 larvae (10 larvae per site from seven sampling sites) of *P. melete* and *P. napi*, respectively, and 78 larvae of *P. rapae* (10 larvae per site from seven sampling sites, plus eight larvae from Yonaguni Island). We extracted RNA from dissected larval gut using the RNeasy Mini Kit (Qiagen), synthesized cDNA with the ReverTra Ace qPCR RT Master Mix (Toyobo) and used TaKaRa Ex Taq (TaKaRa) for amplification of *NSP* and *MA* genes. We designed PCR (polymerase chain reaction) primers for *NSP* and *MA* genes based on available RNA-seq data (accession nos.: ERX2829492–ERX2829499). We gel-purified PCR products with the GEL/PCR Purification Mini Kit (Favorgen), cloned them using the 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).

2.5 | Network analysis of *NSP* and *MA*

We trimmed and aligned *NSP* and *MA* reads with MEGA6 (Tamura et al., 2013) and MAFFT (Katoh & Standley, 2013). For each species, we excluded singleton single nucleotide polymosphisms (SNPs), assuming that they were PCR errors. To analyse NSP and MA divergence across Japan in all three *Pieris* species, we performed NeighborNet network analyses with SPLITSTREE 4.15.1 (Huson & Bryant, 2006).

2.6 | PCR-RFLP and RT-qPCR of *P. melete NSP* and *MA* genes

Since our NeighborNet network analyses showed that there were two diverged *NSP* variants in *P. melete* across Japan, we tested whether these variants were different alleles or different gene copies. First, we used PCR-RFLP to distinguish two major *NSP* variants in *P. melete* and observed whether those appeared as different alleles among individual samples. We amplified *NSP* exon 1 from genomic DNA (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 between the two diverged *NSP* variants, and separation on 3% TBE agarose gels. Second, we performed real-time quantitative PCR (RT-qPCR) from gDNA to determine whether *NSP* and *MA* gene copy numbers differed among *P. melete* samples. For primer design, we used PRIMER3PLUS (Untergasser et al., 2007) with a product size of 70–180 bp, a T_m of 59–61°C, a GC content of 40%–60% and a maximum polybase of 3. For RT-qPCR, we used the CFX Connect Real-Time PCR Detection System (Bio-Rad) using TB Green Premix Ex Taq II (Tli RnaseH Plus; TaKaRa) with *Ef1a* as a reference gene. We used the ddCq method (Pfaffl, 2001) and analysis of variance (ANOVA) to compare relative gene copy numbers among genotypes.

2.7 | RAD sequencing

We extracted gDNA for RAD-seq from the same 218 individuals that we had used for sequencing of *NSP* and *MA* genes, using the Maxwell 16 LEV Plant DNA Kit (Promega), with *Eco*R1 as the restriction enzyme for generating RAD-seq libraries. We ran all 218 samples in a single lane on a HiSeq2500 (Illumina SE, 50 bp) and trimmed reads with TRIMMOMATIC (Bolger et al., 2014), using ILLUMINACLIP:2:10:10, TRAILING:20, SLIDINGWINDOW:4:15 and MINLEN:30. We excluded samples with fewer than 500,000 reads from further analysis. We called SNPs with STACKS version 1.48 (Catchen, 2013). For

USTACKS, we set $n = 3$ and $M = 3$, and for CSTACKS, we used the $n = 3$ option. We performed this analysis for each species independently. To acquire values for genetic diversity (π) and genetic distance (F_{ST}) of each population, we used "populations" with parameter p set as sampled population number and $r = .70$. Since we suspected that *P. melete* and *P. napi* could hybridize, we also called SNPs by setting $p = 1$ and $r = .85$ in multiple species scales without involving population information for assessing the genetic structure of the three species.

2.8 | Population and evolutionary genetic analyses

For population structure analyses we used SNPs called across the three species and performed sparse non-negative matrix factorization (SNMF) with the "snmf" function (K = 1–10, repetition = 20, iterations = 200) implemented in the R package "LEA" (Frichot & François, 2015; Frichot et al., 2014). We used the cross-entropy criterion to determine the number of ancestral populations (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 replicates. We performed these analyses on the entire coding sequence as well as on individual exons. To calculate genome-wide Tajima's D at the species level, we used VCFTOOLS (Danecek et al., 2011). We performed the same random sampling of individuals and selected 30 random SNPs, which corresponded to the average number of SNPs in *NSP* within populations. We used the poly-div_sfs.pl script to calculate the allele frequency of *NSP* and *MA* for each species. At the population level, we used DNASP version 6 (Rozas et al., 2017) to compute π and F_{ST} of *NSP* and *MA*, and ARLEQUIN version 3.5 (Excoffier & Lischer, 2010) to compute Tajima's D. We estimated F_{ST}-like amino acid divergence in NSP and MA between populations as follows; F_{ST}–AA = {Dij − mean (D within population)}/Dij, where Dij corresponds to the 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 et al., 2008) to conduct McDonald–Kreitman tests (McDonald & Kreitman, 1991).

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

3 | RESULTS

3.1 | Field sampling

3.1.1 | 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 (Figure 1a). We chose 11 sites along this predicted cline of Brassicaceae diversity: Yubari, Rusutsu and Oshamanbe (all on the island of Hokkaido), Fukushima, Nagano, Chiba, Kanagawa and 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; Figure 1a). In total, we collected data on 4777 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 ($r = .701$, $p = .016$). Diversity was highest in central Japan and declined towards the north and the south (Figure 1b).

3.1.2 | Sampling of *Pieris* larvae in the field

In total, we collected 945 larvae from 11 sampling sites (Table S4). Based on PCR-RFLP using restriction site polymorphisms in the mitochondrial *ND5* region, we identified 483 larvae of *Pieris melete*, 253 of *P. napi* and 209 of *P. rapae*.

3.2 | Population structure of three *Pieris* species in Japan

We randomly chose 10 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 version 1.48 called 2490 SNPs for *P. melete*, 1376 for *P. napi* and 1410 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).

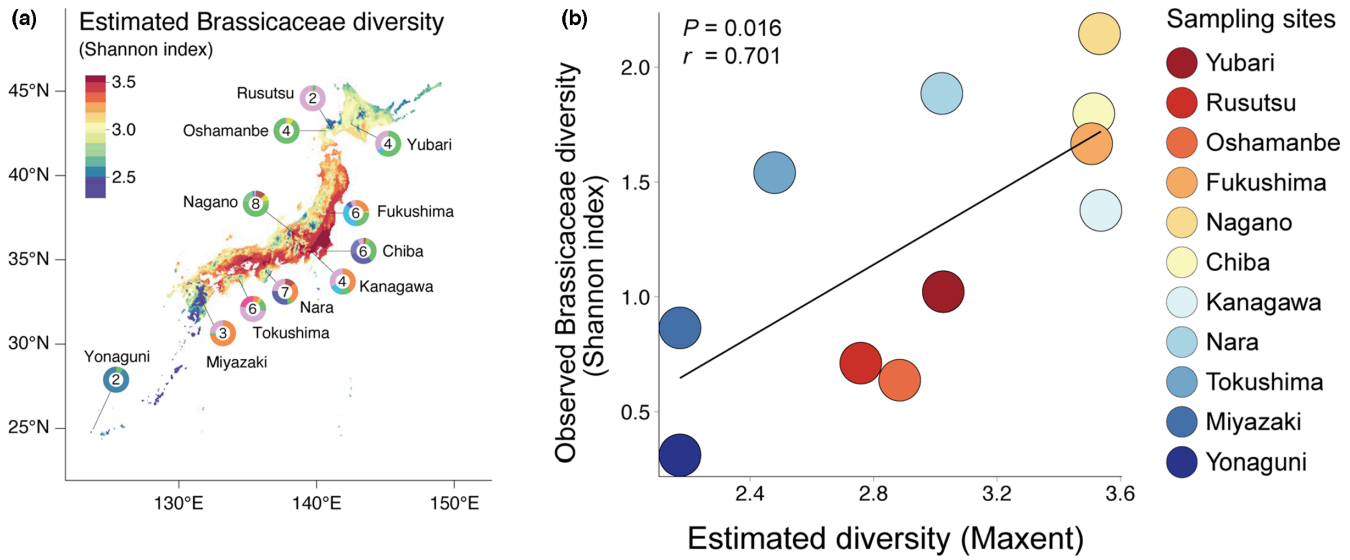

FIGURE 1 (a) MAXENT-based diversity in Brassicaceae across Japan. The colour code indicates the Shannon index for estimated diversity. Pie charts show the number and composition of observed Brassicaceae genera at each sampling site. (b) Pearson correlation between estimated and observed diversity of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

Genome-wide data did not suggest any apparent population genetic structure of *P. melete* or *P. rapae* across Japan (Figures 2 and 3d). In contrast, *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 (Figure 2) but not for recent hybridization events among these closely related butterfly species.

3.3 | Diversity of *NSP* and *MA*

3.3.1 | *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 (Tables 1b and S5). PCR products contained near full-length coding sequences (CDS) of *NSP* (1860 bp of a total CDS of 1896 bp) and *MA* (1869 bp of 1899 bp).

Network analysis identified two main clusters of *NSP* variants in both *P. melete* and *P. napi* (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 3b), with one of the two *P. napi MA* variants present only in Yubari on Hokkaido (Figure 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 an RT-qPCR-based quantification of gene copy number. The *P. melete NSP* variants differed by an 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 (Figure S1). Furthermore, gene copy numbers determined by RT-qPCR (Figure 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.

3.4 | NSP and MA diversity and plant community diversity

3.4.1 | 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 11 different plant species, *P. napi* on 14 and *P. rapae* on nine (Table S4; Figure S3). We found 36% of all *P. melete* larvae on *Rorippa indica*, 19% on *Cardamine leucantha*,

TABLE 1 Population genetic parameters (a) from genome-wide RAD-seq, (b) for *NSP* and *MA* in three *Pieris* species

Species	Population	Private alleles	Individuals	Observed heterozygosity	π
(a)					
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

		Segregating sites (S)		Nucleotide diversity (π)		Amino acid diversity		Tajima's *D*	
Species	Sampling sites	*NSP*	*MA*	*NSP*	*MA*	*NSP*	*MA*	*NSP*	*MA*
(b)									
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

TABLE 1 (Continued)

Species	Sampling sites	Segregating sites (S)		Nucleotide diversity (π)		Amino acid diversity		Tajima's *D*	
		NSP	*MA*	*NSP*	*MA*	*NSP*	*MA*	*NSP*	*MA*
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

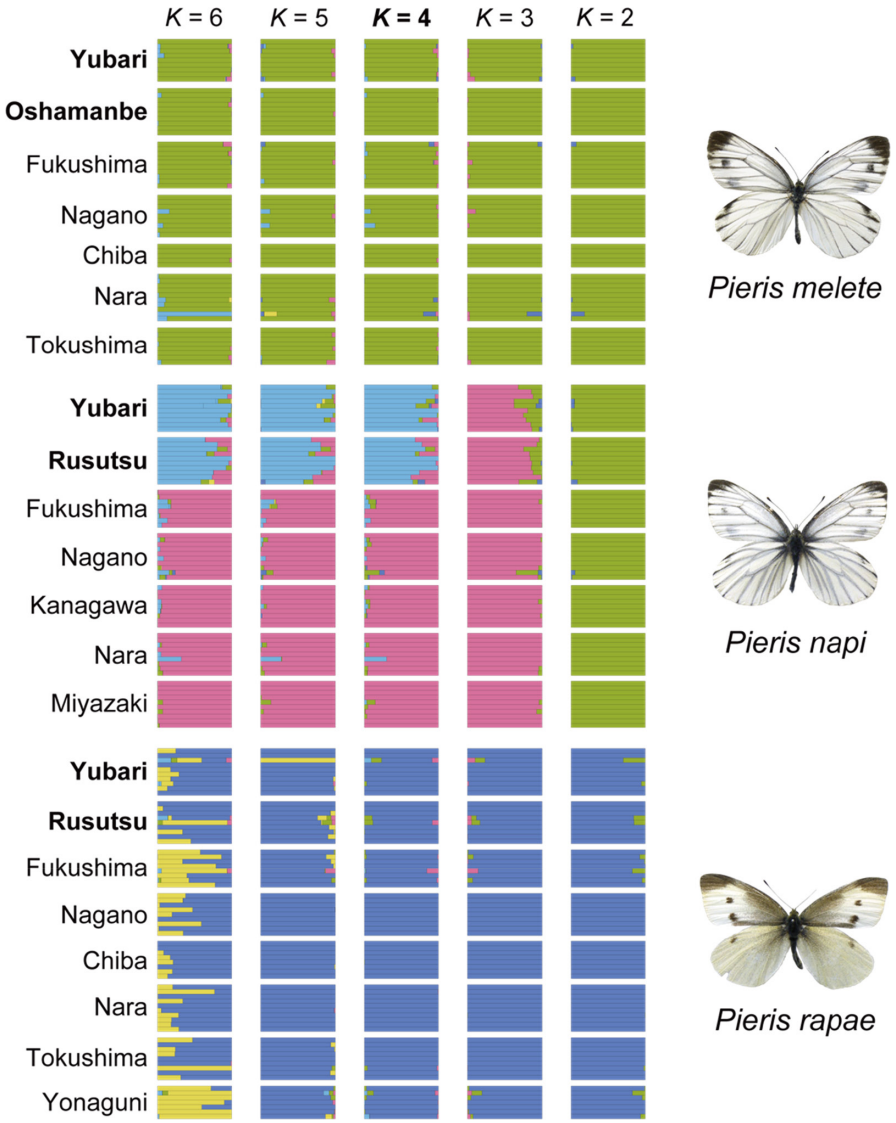

FIGURE 2 Inferred population genetic structures of three *Pieris* spp. across Japan for different clustering numbers (*K*), based on RAD-seq data. The cluster number most highly supported by the cross-entropy criterion (*K* = 4) is shown in bold. Coloured horizontal lines correspond to individual samples,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

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%), *A. hirsuta* (12%) and *Barbarea orthoceras* (11%). Finally, we found *P. rapae* larvae most often on feral plants of *Brassica napus* (33%), followed by *R. indica* and *R. sylvestris* (both: 15%), *Cardamine kiusiana* (12%), *Nasturtium*

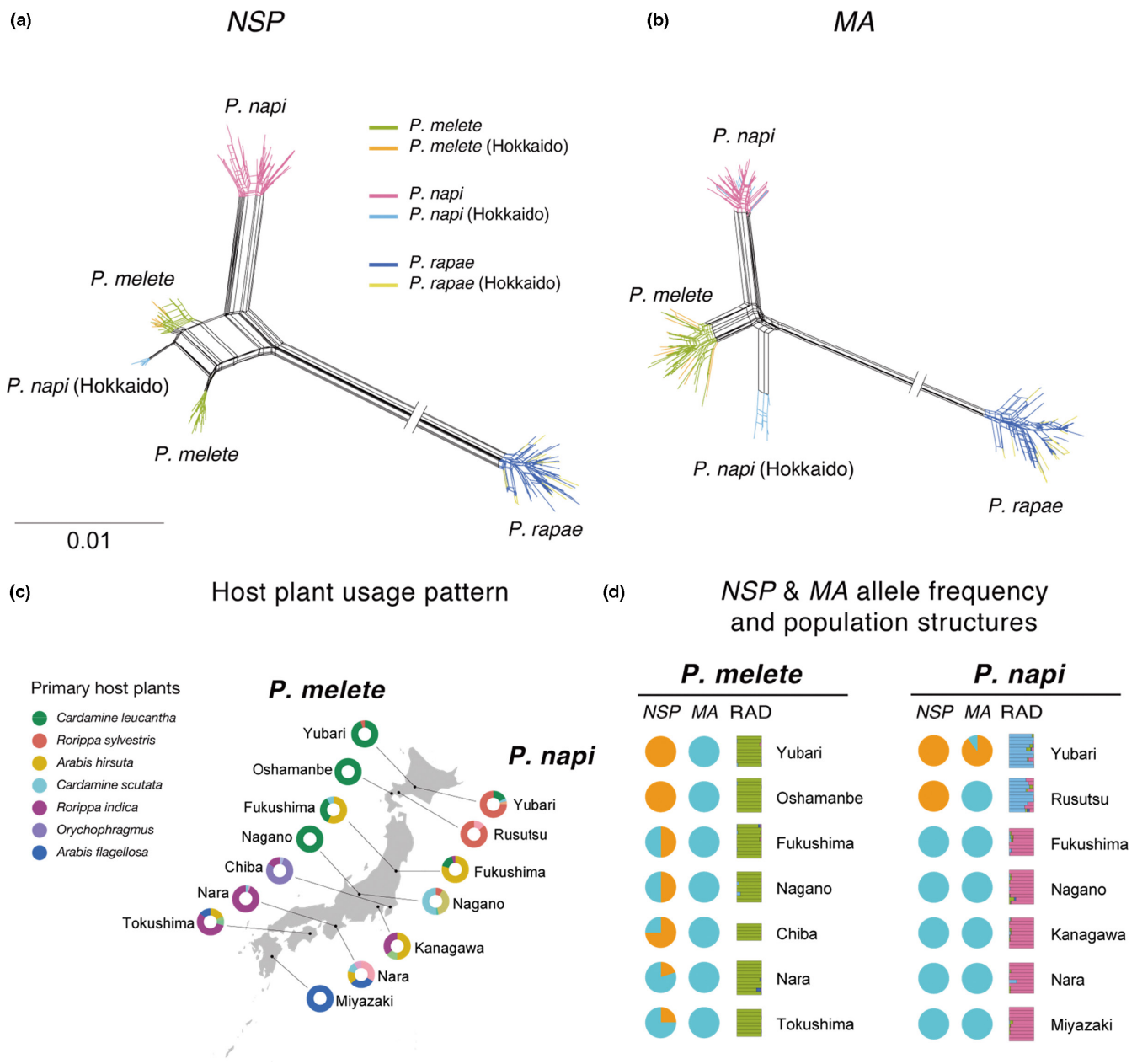

FIGURE 3 NeighborNet networks of (a) *NSP*s and (b) *MA*s from three *Pieris* species across Japan, with different colours for species and for Hokkaido populations. Scale bar indicates number of substitutions per site. Note that *P. melete* and *P. napi* both have two major *NSP* alleles. Note further that *P. napi* has two major *MA* alleles. (c) Observed host plant usage patterns of *P. melete* and *P. napi* at sampling sites. (d) Allele frequencies of major *NSP* and *MA* alleles in comparison with genome-wide genetic structure of *P. melete* and *P. napi* populations at each sampling site, ordered from north to south

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.

Together, *Pieris* species appeared to use a more diversified set of host plants in central Japan than in northern and southern Japan (Figure 3c; Table S4). Therefore, we compared diversity in detoxification genes with the diversity of Brassicaceae across Japan (Figure 4a). Local host plant diversity correlated significantly with NSP amino acid diversity for both *P. melete* ($r = .83$, $p = .022$) and *P. napi* ($r = .88$, $p = .010$), but not *P. rapae* ($r = .021$, $p = .961$). By contrast, local host plant diversity did not correlate with MA amino acid diversity or genome-wide diversity for any of the three species (Figure 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 populations for both *P. melete* ($r = .443$, $p = .039$) and *P. napi* ($r = .436$, $p = .009$) but not *P. rapae*. Likewise, results for MA remained nonsignificant (Table 2).

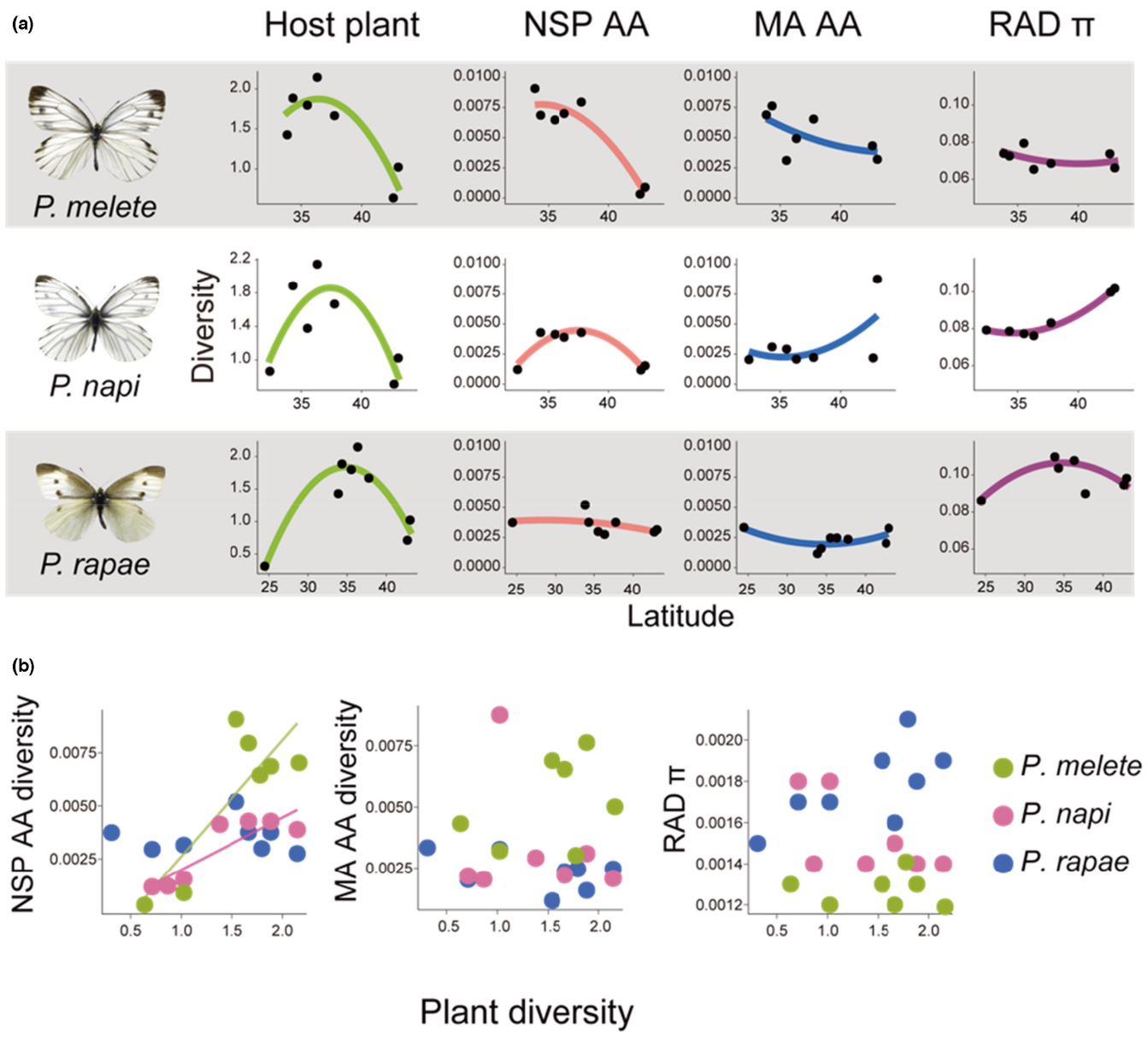

FIGURE 4 (a) Host plant diversity (Shannon index), NSP amino acid diversity (NSP AA), MA amino acid diversity (MA AA) and genome-wide diversity (RAD π) within three *Pieris* species at sampling sites ordered according to latitude. Black dots correspond to sampling sites, with coloured lines showing regression curves. (b) Correlations between NSP AA, MA AA or RAD π with host plant diversity in three *Pieris* species. Note the significant positive correlation ($p < .05$) between plant diversity and NSP AA diversity in *P. melete* and *P. napi*, but not in *P. rapae*

TABLE 2 Partial mantel tests for NSP and MA in three *Pieris* species

Comparison	Species	*r*	*p*-value
Partial mantel test			
NSP AA F_{ST}, plant dissimilarity|RAD F_{ST}	***P. melete***	**.444**	**.039**
	P. napi	**.436**	**.010**
	P. rapae	−.047	.594
MA AA F_{ST}, plant dissimilarity|RAD F_{ST}	*P. melete*	.1405	.258
	P. napi	−.318	.867
	P. rapae	−.183	.877

Note: NSP AA F_{ST} and MA AA F_{ST}: F_{ST}-like divergences based on amino acid sequences. Plant dissimilarity: Bray–Curtis plant community dissimilarity. Statistically significant results are in bold ($p \leq .05$).

3.5 | Statistical tests for selection

3.5.1 | Different evolutionary regime between *NSP* and *MA*

To identify patterns of potential selection, we calculated Tajima's *D* for *NSP* (D_{NSP}), *MA* (D_{MA}) and at the genome-wide level (D_{RAD}). D_{RAD} was negative in all three species, compatible with overall population expansion (Figure 5a). In contrast, D_{NSP} was positive and substantially higher than D_{RAD} in *P. melete* and *P. napi*. Also, D_{MA} was higher than D_{RAD} in *P. rapae*. Furthermore, D_{NSP} varied substantially among exons in *P. melete* and moderately in *P. napi* but not in *P. rapae*, while D_{MA} displayed a more uniform pattern across exons of all three species (Figure 5b). In *P. melete*, exon 1 had a particularly high positive *D*. Similarly, *P. napi* exon 3 displayed an elevated *D*. Close inspection of these exons revealed that both contained several linked, nonsynonymous SNPs at intermediate frequency (Figure 5c), suggesting that *NSP* was under balancing selection in *P. melete*.

In addition, we compared the ratio of nonsynonymous to synonymous nucleotide diversity (π_a/π_s) of *NSP* and *MA* of the three *Pieris* spp. *MA* had a lower π_a/π_s ratio compared to *NSP* in all three species (Figure 5d), suggesting that *MA* was more evolutionarily conserved than *NSP* across *Pieris* spp.

3.5.2 | Divergent selection of *NSP* alleles in *P. napi*

We found that *P. melete* and *P. napi NSP*s, as well as *P. napi MA*s, were all grouped into two clearly distinct clusters (Figure 3a). Therefore, we tested for potential divergent selection between major alleles, using a population-based diversity test (McDonald & Kreitman, 1991). Results for *P. melete NSP* and *P. napi MA* were nonsignificant (Table 3), but *P. napi NSP* showed a strong signal of divergent selection (NI = 0.15, $\chi^2 = 10.6$, $p = .001$), caused by an elevated level of nonsynonymous vs. synonymous substitutions between populations from Hokkaido and populations outside of Hokkaido. To identify where mutations had occurred, we compared *NSP*s of *P. napi* and *P. melete*, which allowed us to infer the direction of evolutionary change. To our surprise, all nonsynonymous substitutions that differentiated the Hokkaido populations of *P. napi* from those of central and southern Japan had occurred outside of Hokkaido, while each two synonymous substitutions had occurred in the lineages leading to the Hokkaido and non-Hokkaido populations. Furthermore, 10 nonsynonymous and 24 synonymous polymorphisms were specific for populations from central and southern Japan, and only two nonsynonymous polymorphisms were specific for populations from Hokkaido (Figure 6). These patterns of polymorphism indicate that the lineage leading to the extant *P. napi* non-Hokkaido populations experienced strong directional selection on NSP function, followed by stabilizing selection under relaxed constraints, whereas the lineage leading to the extant *P. napi* Hokkaido populations experienced only stabilizing selection.

4 | DISCUSSION

Several insect counteradaptations against host plant defences have been characterized at the molecular level (Berenbaum et al., 1996; Gloss et al., 2014; Heidel-Fischer et al., 2019; Li et al., 2003; Ratzka et al., 2002; Wittstock et al., 2004) but only few have been investigated for genetic variation in response to host plant usage. Bono et al. (2008) found some evidence for adaptive evolution of two cytochrome P450 genes from the fly *Drosophila mettleri* that were putatively involved in the detoxification of alkaloids from two cactus species. Heidel-Fischer et al. (2010) investigated genetic variation in the *NSP* gene of three European and one USA population of *Pieris rapae* and found high amounts of amino acid diversity but little evidence for local adaptation. Here, we investigated two genes that underlie the counteradaptation of Pierid butterflies against the GLS–myrosinase defence system of their host plants for potential microevolutionary dynamics in response to host plant usage along a latitudinal cline in Japan. In addition to the crop-dependent *P. rapae*, we included two *Pieris* species that feed on wild Brassicaceae, *P. melete* and *P. napi*, to test whether different patterns of host plant use (wild plant-dependent vs. crop-dependent) would result in differences in the observed microevolutionary patterns of NSP and/or MA. As in the previous study of Heidel-Fischer et al. (2010), we found no evidence for adaptation of *P. rapae NSP*-like genes to local communities of wild Brassicaceae. In contrast, both *P. melete* and *P. napi* showed a significant and positive correlation between the diversity of the *NSP* (but not the *MA*) gene product and host plant diversity (Figure 4). In addition, both species harboured two distinct *NSP* alleles, which shaped the observed correlations in a species-specific manner.

P. melete did not show any indication of population substructure (Figure 2). In this species, NSP diversity correlated with host plant diversity, which varied along a latitudinal gradient (Figure 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 (Figure 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,

FIGURE 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 *Pieris 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 colour coding. Exon 1 of *P. melete* and exon 3 of *P. napi NSP*s accumulate positions with intermediate allele frequencies. (d) Nonsynonymous (π_a) and synonymous nucleotide diversity (π_s) at the species level and ratio of nonsynonymous to synonymous diversity (π_a/π_s) of *NSP* and *MA* in three *Pieris* species

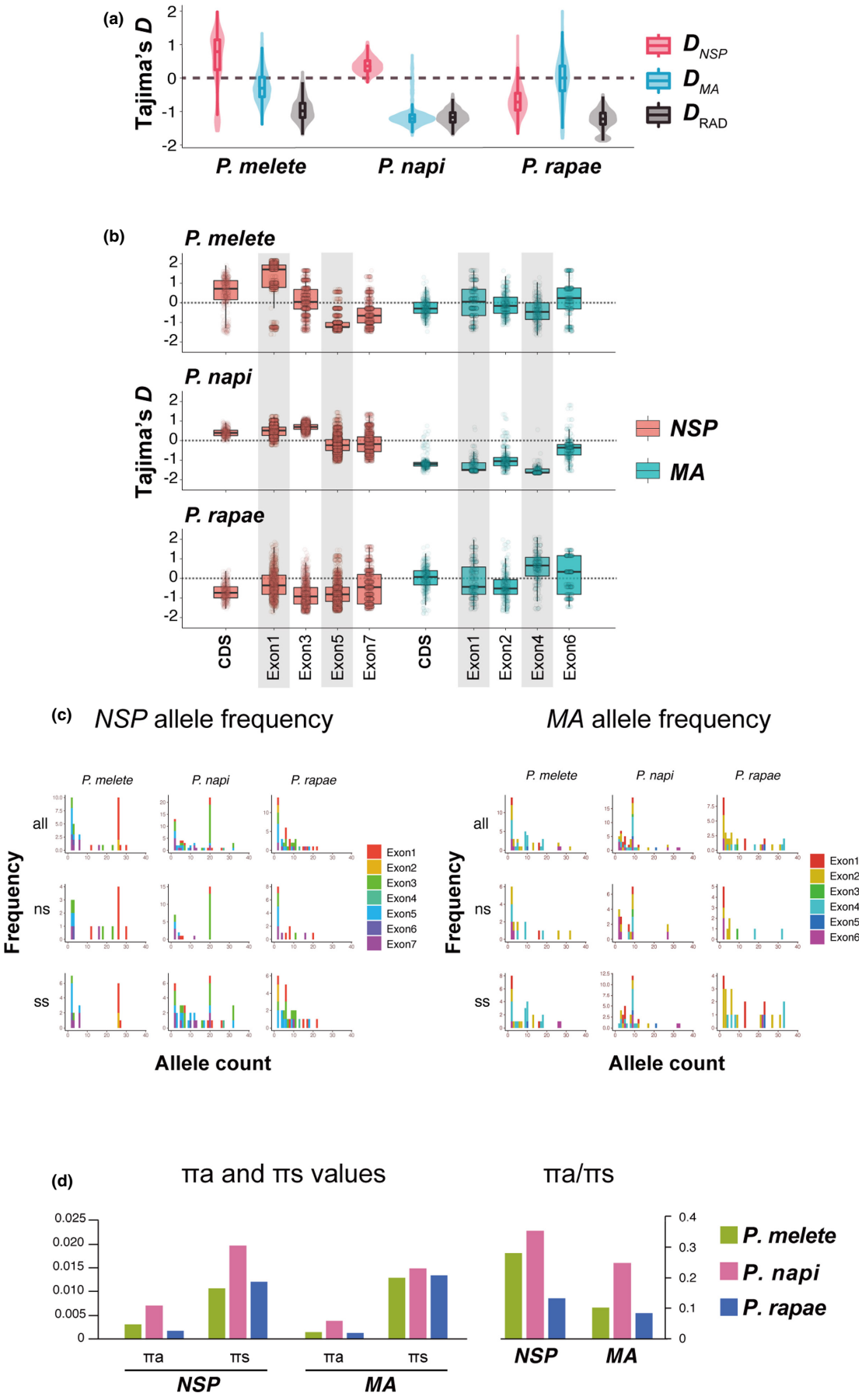
(a)
Tajima's D
2
1
0
-1
-2
P. melete
P. napi
P. rapae
D_{NSP}
D_{MA}
D_{RAD}
(b)
P. melete
P. napi
P. rapae
Tajima's D
NSP
MA
CDS
Exon1
Exon3
Exon5
Exon7
CDS
Exon1
Exon2
Exon4
Exon6
(c)
NSP allele frequency
MA allele frequency
P. melete
P. napi
P. rapae
all
ns
ss
Frequency
Exon1
Exon2
Exon3
Exon4
Exon5
Exon6
Exon7
Allele count
(d)
πa and πs values
πa/πs
0.025
0.020
0.015
0.010
0.005
0
πa
πs
πa
πs
NSP
MA
0.4
0.3
0.2
0.1
0
NSP
MA
P. melete
P. napi
P. rapae

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

		Polymorphisms		Substitutions				
Comparison	**Gene**	**syn**	**nsyn**	**syn**	**nsyn**	**NI**	χ^2	***p***
P. melete	*NSP*	20	13	6.05	4	0.982	0	.981
P. napi	*NSP*	29	21	4.02	19.16	0.152	10.568	**.001**
	MA	22	17	8.1	13.07	0.521	1.807	.178

Abbreviations: NI, Neutrality Index; nsyn, nonsynonymous; syn, synonymous.
Statistically significant results are in bold ($p \leq .05$).

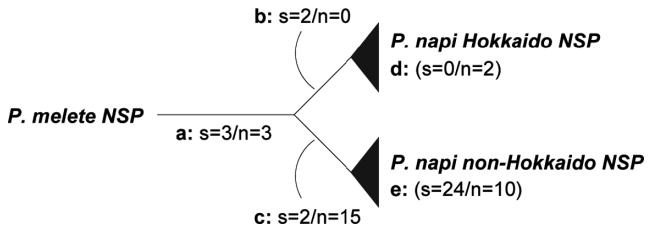

FIGURE 6 Divergent selection on *NSP* in populations of *Pieris napi*. Synonymous (s) and nonsynonymous (*n*) mutations in *NSP*s on branches of a simplified molecular phylogeny consisting of two *P. napi* populations and *Pieris melete*. Polymorphisms segregating in each population are shown in parentheses. A very high number of nonsynonymous mutations on the non-Hokkaido branch of *P. napi* (branch c) indicates positive selection in the lineage leading to the non-Hokkaido populations of *P. napi*. In comparison, *NSP* of the Hokkaido populations appears to be highly conserved

in particular in the first exon (Figure 5b). This region indeed harboured an elevated number of linked synonymous and nonsynonymous SNPs at intermediate frequency (Figure 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 defence (Carley et al., 2021; Kroymann et al., 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 defences. Of course, functional data would be helpful to substantiate this idea.

Hokkaido populations of *P. napi* had a major *NSP* allele that was distinct from alleles found outside of Hokkaido (Figure 3a). This difference mirrors genome-wide differentiation between both populations (Figure 2). Based on the analysis of the mitochondrial *ND5* gene, a previous 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 (Figure 2). Furthermore, *MA* alleles showed similar divergence patterns as shown in *ND5*-based genotyping (Figure 3d). However, despite gene flow and in contrast to *MA* alleles, both Hokkaido populations had a fixed *NSP* allele (Figure 3d). Furthermore, *NSPs* from Hokkaido and from central and southern Japan differed by more nonsynonymous substitutions than expected (Table 3), indicative of diversifying selection. However, all amino acid changes that distinguished Hokkaido and non-Hokkaido populations occurred in the non-Hokkaido lineage (Figure 6), indicating that strong directional selection occurred 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 et al., 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*.

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 and 3; Figures 4 and 5). Instead, the comparison of nonsynonymous version synonymous diversity suggested that *MA*s were more strongly conserved than *NSP*s in all three *Pieris* species (Figure 5d). This is consistent with a previous study (Okamura, Sato, Tsuzuki, Murakami, et al., 2019), which found evidence for positive selection acting on *NSP* but not *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 homologues play different roles in GLS detoxification, with the more conserved MAs detoxifying more common GLSs, and NSP acting on rarer GLSs.

Despite the ecological importance of herbivores' counteradaptations in the field, the underlying microevolutionary dynamics have remained largely unknown. Here, we tested the microevolutionary

patterns of two key genes, *NSP* and *MA*, in *Pieris* butterflies that have different host usage patterns across Japan. We found that the evolution of *NSP*s varied among the three species, which differ in their ecology, host plant use and population dynamics. On the other hand, *MA*s did not show any particular pattern that correlated with host plant diversity. The different evolutionary dynamics of NSP and MA indicates that functional differentiation could be important for adaptation to chemically heterogeneous host plants in the field. If the observed sequence differences could be linked to functional differentiation, it would advance our understanding of the chemical arms-race between Pierid butterflies and their host plants of the Brassicaceae taking place in the field.

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 Fellowship 15J00320 and JSPS Overseas Research Fellowship 202060676 to Y.O., and Max-Planck-Gesellschaft. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

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.

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. The host plant data, as well as NSP and MA sequences, have been deposited at the Edmond server: https://edmond.mpdl.mpg.de/imeji/collection/3Ne_ZYG2Vs7H8pAa

ORCID

Yu Okamura https://orcid.org/0000-0001-6765-4998
Atsushi J. Nagano https://orcid.org/0000-0001-7891-5049
Heiko Vogel https://orcid.org/0000-0001-9821-7731

REFERENCES

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

Benderoth, M., Pfalz, M., & Kroymann, J. (2009). Methylthioalkylmalate synthases: Genetics, ecology and evolution. *Phytochemistry Reviews*, *8*(1), 255–268. https://doi.org/10.1007/s11101-008-9097-1

Benderoth, M., Textor, S., Windsor, A. J., Mitchell-Olds, T., Gershenzon, J., & Kroymann, J. (2006). Positive selection driving diversification in plant secondary metabolism. *Proceedings of the National Academy of Sciences*, *103*(24), 9118–9123. https://doi.org/10.1073/pnas.0601738103

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. https://doi.org/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. https://doi.org/10.1046/j.1420-9101.2003.00531.x

Bidart-Bouzat, M. G., & Kliebenstein, D. J. (2008). Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in *Arabidopsis thaliana*. *Journal of Chemical Ecology*, *34*(8), 1026–1037. https://doi.org/10.1007/s10886-008-9498-z

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*, 112100. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.1365-294X.2008.03823.x

Carley, L. N., Mojica, J. P., Wang, B., Chen, C.-Y., Lin, Y.-P., Prasad, K. V. S. K., Chan, E., Hsu, C.-W., Keith, R., Nuñez, C. L., Olson-Manning, C. F., Rushworth, C. A., Wagner, M. R., Wang, J., Yeh, P.-M., Reichelt, M., Ghattas, K., Gershenzon, J., Lee, C.-R., & Mitchell-Olds, T. (2021). Ecological factors influence balancing selection on leaf chemical profiles of a wildflower. *Nature Ecology & Evolution*, *5*, 1135–1144. https://doi.org/10.1038/s41559-021-01486-0

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., & Bernatchez, L. (2007). Clinal variation in MHC diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution*, *61*(9), 2154–2164. https://doi.org/10.1111/j.1558-5646.2007.00178.x

Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., Heckel, D. G., Der, J. P., Wafula, E. K., Tang, M., Hofberger, J. A., Smithson, A., Hall, J. C., Blanchette, M., Bureau, T. E., Wright, S. I., dePamphilis, C. W., Eric Schranz, M., Barker, M. S., … Wheat, C. W. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 8362–8366. https://doi.org/10.1073/pnas.1503926112

Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: A website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, *36*, 157–162. https://doi.org/10.1093/nar/gkn337

Excoffier, L., & Lischer, H. E. L. 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), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fahey, J. W., Zalcmann, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, *56*(1), 5–51. https://doi.org/10.1016/S0031-9422(00)00316-2

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), 4302–4315. https://doi.org/10.1002/joc.5086

Fischer, H. M., Wheat, C. W., Heckel, D. G., & Vogel, H. (2008). Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular Biology and Evolution*, *25*(5), 809–820. https://doi.org/10.1093/molbev/msn014

Friberg, M., & Wiklund, C. (2019). Host preference variation cannot explain microhabitat differentiation among sympatric *Pieris napi* and *Pieris rapae* butterflies. *Ecological Entomology*, *44*(4), 571–576. https://doi.org/10.1111/een.12728

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/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. https://doi.org/10.1073/pnas.0904106106

Gloss, A. D., Vassão, D. G., Hailey, A. L., Nelson Dittrich, A. C., Schramm, K., Reichelt, M., Rast, T. J., Weichsel, A., Cravens, M. G., Gershenzon, J., Montfort, W. R., & 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. https://doi.org/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. https://doi.org/10.12688/f1000research.9765.1

Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology*, *57*, 303–333. https://doi.org/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. https://doi.org/10.1146/annurev.ecolsys.37.091305.110132

Heidel-Fischer, H. M., Kirsch, R., Reichelt, M., Ahn, S.-J., Wielsch, N., Baxter, S. W., Heckel, D. G., Vogel, H., & Kroymann, J. (2019). An insect counteradaptation against host plant defenses evolved through concerted neofunctionalization. *Molecular Biology and Evolution*, *36*(5), 930–941. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1104/pp.126.2.811

Kroymann, J., Donnerhacke, S., Schnabelrauch, D., & Mitchell-Olds, T. (2003). Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proceedings of the National Academy of Sciences*, *100*(Supplement 2), 14587–14592. https://doi.org/10.1073/pnas.1734046100

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), 639–648. https://doi.org/10.1111/ecog.00981

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*(24), 14593–14598. https://doi.org/10.1073/pnas.1934643100

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, *353*, 412–414. https://doi.org/10.1038/351652a0

Mitchell-Olds, T., & Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature*, *441*(7096), 947–952. https://doi.org/10.1038/nature04878

Ohsaki, N. (1979). Comparative population studies of three *Pieris* butterflies, *P. rapae*, *P. melete* and *P. napi*, living in the same area—I. Ecological requirements for habitat resources in the adults. *Researches on Population Ecology*, *20*(2), 278–296. https://doi.org/10.1007/BF02512633

Ohsaki, N., Ohata, M., Sato, Y., & Rausher, M. D. (2020). Host plant choice determined by reproductive interference between closely related butterflies. *The American Naturalist*, *196*(4), 512–523. https://doi.org/10.1086/710211

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), 59–68. https://doi.org/10.2307/1939382

Okamura, Y., Sato, A., Tsuzuki, N., Murakami, M., Heidel-Fischer, H., & Vogel, H. (2019). Molecular signatures of selection associated with host plant differences in *Pieris* butterflies. *Molecular Ecology*, *28*(22), 4958–4970. https://doi.org/10.1111/mec.15268

Okamura, Y. U., Sato, A. I., Tsuzuki, N., Sawada, Y., Hirai, M. Y., Heidel-Fischer, H., Reichelt, M., Murakami, M., & Vogel, H. (2019). Differential regulation of host plant adaptive genes in *Pieris* butterflies exposed to a range of glucosinolate profiles in their host plants. *Scientific Reports*, *9*(1), 7256. https://doi.org/10.1038/s41598-019-43703-8

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

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H. (2017). *Community ecology package*, *package 'vegan*.' ISBN 0-387-95457-0

Olsen, C. E., Huang, X. C., Hansen, C. I. C., Cipollini, D., Ørgaard, M., Matthes, A., Geu-Flores, F., Koch, M. A., & 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. https://doi.org/10.1016/j.phytochem.2016.09.013

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), 2002–2007. https://doi.org/10.1093/nar/29.9.e45

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*(7), 887–893. https://doi.org/10.1111/ecog.03049

Phillips, S., Dudík, M., & Schapire, R. (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 655–662. doi: https://doi.org/10.1145/1015330.1015412

Prasad, K. V. S. K., Song, B.-H., Olson-Manning, C., Anderson, J. T., Lee, C.-R., Schranz, M. E., Windsor, A. J., Clauss, M. J., Manzaneda, A. J., Naqvi, I., Reichelt, M., Gershenzon, J., Rupasinghe, S. G., Schuler, M. A., & Mitchell-Olds, T. (2012). A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science*, *337*(6098), 1081–1084. https://doi.org/10.1126/science.1221636

Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T., & Kroymann, J. (2002). Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11223–11228. https://doi.org/10.1073/pnas.172112899

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Ryan, S. F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dincă, V., & Shoemaker, D. (2019). Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. *Proceedings of the National Academy of Sciences*, *116*(40), 20015–20024. https://doi.org/10.1073/pnas.1907492116

Shinkawa, T., Yagi, T., Fujii, H., & Kondoh, K. (2003). Molecular phylogeny and divergence of *Pieris* group. *Butterfly DNA Research Society News Letter*, *10*, 14–17.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology*, *35*(4), 627–632. https://doi.org/10.2307/2413122

Sokal, R. R. (1979). Testing statistical significance of geographic variation patterns. *Systematic Zoology*, *28*(2), 227. https://doi.org/10.2307/2412528

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(Suppl. 2), 71–74. https://doi.org/10.1093/nar/gkm306

Wheat, C. W., Vogel, H., Wittstock, U., Braby, M. F., Underwood, D., & Mitchell-Olds, T. (2007). The genetic basis of a plant-insect coevolutionary key innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20427–20431. https://doi.org/10.1073/pnas.0706229104

Windsor, A. J., Reichelt, M., Figuth, A., Svatoš, A., Kroymann, J., Kliebenstein, D. J., Gershenzon, J., & Mitchell-Olds, T. (2005). Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochemistry*, *66*(11), 1321–1333. https://doi.org/10.1016/j.phytochem.2005.04.016

Wittstock, U., Agerbirk, N., Stauber, E. J., Olsen, C. E., Hippler, M., Mitchell-Olds, T., Gershenzon, J., & Vogel, H. (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), 4859–4864. https://doi.org/10.1073/pnas.0308007101

Wittstock, U., & Halkier, B. A. (2002). Glucosinolate research in the *Arabidopsis* era. *Trends in Plant Science*, *7*(6), 263–270. https://doi.org/10.1016/S1360-1385(02)02273-2

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Okamura, Y., Sato, A., Kawaguchi, L., Nagano, A. J., Murakami, M., Vogel, H., & Kroymann, J. (2022). Microevolution of *Pieris* butterfly genes involved in host plant adaptation along a host plant community cline. *Molecular Ecology*, *31*, 3083–3097. https://doi.org/10.1111/mec.16447