Differential introgression causes genealogical discordance in host races of *Acrocercops transecta* (Insecta: Lepidoptera)

Issei Ohshima (I.O.)[1,2§] and Kazunori Yoshizawa (K.Y.)[1]

[1]Systematic Entomology, Department of Ecology and Systematics, Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

[2]Division of Evolutionary Biology, National Institute for Basic Biology, Okazaki 444-8585, Japan

Keywords: ecological speciation, gene flow, host adaptation, hybridization, mitochondrial gene, Z-linked gene.

[§]Correspondence: Issei Ohshima

Present Address: Division of Evolutionary Biology, National Institute for Basic Biology, Okazaki 444-8585, Japan

Fax: +81-564-55-7549

Email addresses:

IO: issei@nibb.ac.jp

KY: psocid@res.agr.hokudai.ac.jp

Running title: Genealogical discordance in host races

Abstract

Recently diverged populations often exhibit incomplete reproductive isolation, with a low level of gene flow continuing between populations. Previous studies have shown that, even under a low level of gene flow, genetic divergence between populations can proceed at the loci governing local adaptation and reproductive isolation but not at other neutral loci. A leaf-mining moth, *Acrocercops transecta*, consists of *Juglans*- and *Lyonia*-associated host races. The two host races differ in host preferences of ovipositing females and in larval adaptation to host plants but mate readily in the laboratory, producing fertile hybrids. The *Juglans* and *Lyonia* races are often sympatric in the wild, implying that gene introgression could occur in nature between the two host races. We tested this hypothesis by combining phylogenetic analyses with coalescent simulations, focusing on mitochondrial genes (COI and ND5) and the nuclear *Tpi*, *Per* and *Ldh* genes located on the Z-chromosome. The mitochondrial genes clearly distinguished the *Lyonia* race from the *Juglnas* race, whereas the *Tpi*, *Per* and *Ldh* genealogies did not reflect the two host races. Coalescent simulations indicated gene flow at the three Z-linked genes in both directions, whereas there was no introgression in the mitochondrial genes. The lack of introgression in mitochondrial genes suggests that female host preference is the primary force leading to the bifurcation of maternally inherited loci. Thus, the results show that a low level of gene flow coupled with the inflexible female host preference differentiates histories of divergence between maternally and biparentally inherited genes in this host race system.

Introduction

Shifts to a novel environment play an important role in adaptive radiation for both plants and animals (Ramsey *et al*. 2008; Malenke *et al*. 2009; Singer & McBride 2009). Ecological differences between environments cause divergent selection that may yield phenotypic divergence between populations (Nosil & Crespi 2006; Yamashiro *et al*. 2008; Schluter 2009; Monteiro & Nogueira 2009). In the context of gene flow, divergent selection results in genetic differentiation at the loci responsible for adaptation to each environment, because immigrants and hybrids with ecologically maladaptive genetic combinations are likely to be eliminated. Genomic regions surrounding such loci will also hitchhike toward fixation due to tight linkage (Maynard Smith & Haigh 1974; Charlesworth *et al*. 1997; Fay & Wu 2000). However, the remaining genomic regions may be homogenized between populations even under a low level of gene flow (Barton 2000; Clarke *et al*. 1996). Multilocus or genome-wide analyses of populations with different ecological requirements would elucidate the role of selection and gene flow in genomic differentiation and speciation.

Significant heterogeneity in gene genealogies among genomic regions has been reported in *Drosophila* (Machado & Hey 2003); the corn borer moth, *Ostrinia nubilalis* (Dopman *et al*. 2005); *Gryllus* crickets (Maroja *et al*. 2009); *Heliconius* butterflies (Beltrán *et al*. 2002; Bull *et al*. 2006; Kronforst *et al*. 2006); *Papilio* butterflies (Putnam *et al*. 2007); *Dioryctria* moths (Roe & Sperling 2007); and the European rabbit, *Oryctolagus cuniculus* (Geraldes *et al*. 2006). Other genome-wide studies using amplified fragment length polymorphism (AFLP) markers demonstrated that only a small fraction of genomes (< 5%) showed high genetic differentiation, as a result of

linkage with loci involved in local adaptation and reproductive isolation (Wilding *et al*. 2001; Campbell & Bernatchez 2004; Bonin *et al*. 2006; Savolainen *et al*. 2006; Egan *et al*. 2008; Nosil *et al*. 2008; see Nosil *et al*. 2009 for a review). These studies based on closely related species or races indicate that patterns of genetic divergence and of introgression vary greatly among genomic regions.

Host races in phytophagous insects are defined as host-associated populations with little gene flow between them, and they provide useful models for studying the process of genetic divergence (Diehl & Bush 1984; Drès & Mallet 2002). These host races are ecologically and genetically differentiated but are sympatric (not geographically isolated) (Diehl & Bush 1984; Drès & Mallet 2002). In this paper, we compare the patterns of genetic introgression among different genes across host races in a leaf-mining moth, *Acrocercops transecta* (Insecta: Lepidoptera: Gracillariidae), which consists of putative host races that are associated with either *Juglans ailanthifolia* (Juglandaceae) or *Lyonia ovalifolia* (Ericaceae). A previous mtDNA-based phylogeny clearly separated the *Lyonia*-associated race from the *Juglans*-associated race throughout Japan and indicated that the *Lyonia* race evolved once from the *Juglans* race (Ohshima 2008). The *Juglans* and *Lyonia* races differ clearly in host preferences of ovipositing females and in larval adaptation to host plants (Ohshima 2008). Despite the differences in host adaptation, the two host races mate readily under laboratory conditions without detectable intrinsic incompatibilities (Ohshima 2008). The *Juglans* and *Lyonia* races are often sympatric in the wild, i.e., the two populations co-occur within the range of normal dispersal for the adult moths. The high mating compatibilities, coupled with the sympatric distribution, imply that gene introgression could occur in nature between the two host races.

In general, maternally inherited mitochondrial or chloroplast genes are more frequently subject to introgression than biparentally or paternally inherited components (e.g., Bachtrog *et al*. 2006; Forister *et al*. 2008; Gompert *et al*. 2008; for a review see Chan & Levin 2005). However, the strong preference of females to their respective host plant should restrict introgression of maternally inherited genes between host races. This implies that mitochondrial DNA can differentiate between the host races, even in the presence of gene flow at the genomic level.

In *A. transecta*, larval performance is mainly determined by a single autosomal locus, and *Juglans* feeding is completely dominant over *Lyonia*-feeding (Ohshima 2008). Thus, F_1 hybrids between the two host races cannot develop into adulthood when they are fed on *Lyonia* as a host plant because of their inability to digest *Lyonia* tissues, whereas on *Juglans* hybrids grow to adulthood as well as the *Juglans* race does. Because the preference of ovipositing females is not affected by the genotype of mating partners, *Lyonia* females exclusively deposit F_1 eggs on *Lyonia* leaves after they mate with *Juglans* males, although their offspring are inviable there. Thus, it can be predicted that gene flow from the *Juglans* race to the *Lyonia* race should be completely hindered across the entire genome. However, because *Juglans* females deposit eggs on *Juglans* leaves when mating with *Lyonia* males, and because F_1 larvae exhibit high viability on *Juglans*, gene flow from the *Lyonia* race to the *Juglans* race is possible via dispersal of *Lyonia* males to the *Juglans* race.

In order to test these hypotheses of differential introgression between mitochondrial and nuclear genes and of asymmetrical gene flow, we assess the pattern of genetic divergence and introgression across the two host races in *A. transecta* by combining phylogenetic analyses with coalescent simulations. Gene genealogies and

migration rates across the two host races were compared between mitochondrial COI + ND5 and the nuclear *Tpi*, *Period* and *Ldh* genes that are located on the Z chromosome in lepidopterans (Dopman *et al*. 2005; Bull *et al*. 2006). Since host adaptation is determined by autosomal genes in the two host races (Ohshima 2008), Z-linked genes are neutral in respect to host adaptation. Therefore, Z-linked genes could be homogenized even under a low degree of gene flow. Female moths have a single copy of *Tpi*, *Period* and *Ldh* genes that are inherited from their paternal parent because females are the heterogametic sex in Lepidoptera (Fig. 1). Thus, phylogenetic analyses using only females enable us to examine the possibility of genealogical discordance between maternally inherited and Z-linked genes (Fig. 1).

Several studies have assessed genetic divergence using either mtDNA for host races (Brunner *et al*. 2004; Diegisser *et al*. 2004; Stireman *et al*. 2005; Diegisser *et al*. 2006; Ohshima & Yoshizawa 2006) or Z-linked genes for closely related lepidopteran taxa (Jiggins *et al*. 2001; Nason *et al*. 2002; Dopman *et al*. 2005; Bull *et al*. 2006; Kronforst *et al*. 2006; Narita *et al*. 2006). However, few studies have utilized both mitochondrial and Z-linked genes to assess genetic divergence of closely related phytophagous-insect taxa. Furthermore, in phytophagous insects, difficulties crossing and rearing continuously have hampered the assessment of hybrid fitness, so that it has been difficult to predict the pattern of gene flow between host races. We have used the two putative host races of *A. transecta* for which the genetic basis of larval adaptation to host plants has been clarified. This information about the genetic basis enables a prediction for the pattern of gene flow between the putative host races. Thus, the putative host race system in *A. transecta* is a suitable model for testing the hypotheses of differential introgression between mitochondrial and Z-linked genes and of

asymmetrical gene flow. The phylogenetic and coalescent analyses in the present study corroborated the hypothesis of differential introgression due to females' fixed and distinct host preferences. However, despite no viability of F1 hybrid larvae on *Lyonia*, we detected introgression from the *Juglans* race to the *Lyonia* race as well as the opposite direction, suggesting that backcross hybridization between F1 hybrids and the *Lyonia* race contribute to gene flow in this direction.

Materials and methods

Collection

Moths were collected from the following two localities; Mt. Aoba, Sendai-city (38˚15'N, 140˚49'E), northern Honshu, Japan, and Niimi-city (34˚59'N, 133˚25'E), western Honshu, Japan. Note that samples collected from Mt. Aoba were different from those used in the previous mtDNA-phylogeny work (Ohshima 2008). At both localities, the two host plants, *Juglans ailanthifolia* and *Lyonia ovalifolia*, grow side by side and at some sites their canopies overlapped. Larvae were collected where *Juglans* and *Lyonia* trees occurred within 10 meters from each other; at this distance, adult moths are probably able to move between trees. Larvae were collected together with the mined leaves from the host plants, and the larvae were reared in the cut leaves in the laboratory following the method described by Ohshima (2005). Eclosed adults were sexed and stored in 99.5% ethanol at -20 ˚C until DNA extraction. Only female samples were used for analyses, and we refer to samples collected from *Juglans ailanthifolia* as *Juglans* moths and to those from *Lyonia ovalifolia* as *Lyonia* moths. For the Sendai

population, a total of 78 samples were prepared for *Juglans* moths and 68 for *Lyonia* moths; for the Niimi population, there were 44 for *Juglans* moths and 40 for *Lyonia* moths. As an outgroup, we chose *A. leucophaea*, which is a sister species of *A. transecta*, but is divergent from the two putative host races (Ohshima & Yoshizawa 2006).

Sequence determination

Total DNA was extracted from the head and thorax of each sample following the method described by Boom *et al*. (1990). The abdomen of each sample was resuspended in 99.5% ethanol and stored at -20 ˚C in the Department of Evolutionary Biology, National Institute for Basic Biology, Japan, kept as a voucher. For molecular phylogenetic inference, we used partial sequences of two mitochondrial genes, cytochrome oxidase subunit I (COI) and NADH dehydrogenase subunit 5 (ND5), and introns of three house-keeping nuclear genes, triose-phosphate isomerase (*Tpi*), *Period* and lactate dehydrogenase (*Ldh*) that are located on the Z chromosome in lepidopterans (e.g., *Ostrinia*, Dopman *et al*. 2004, 2005; *Heliconius*, Jiggins *et al*. 2005; *Bombyx*, Niu *et al*. 2005). Polymerase chain reaction (PCR) was performed following the procedures described in Ohshima and Yoshizawa (2006) for COI and ND5 genes. Partial sequences of the *Tpi*, *Period* and *Ldh* genes were amplified using the primers listed in Table S1, Supporting Information. The PCR reaction cycle was 94 ˚C for 2 min followed by 40 cycles of 94 ˚C for 45 sec, 55 ˚C for 45 sec, 72 ˚C for 78 sec, and 72 ˚C for 5 min for *Tpi* (but 94 ˚C for 3 min followed by 40 cycles of 94 ˚C for 30 sec, 59 ˚C for 30 sec, 72 ˚C for 45 sec, and 72 ˚C for 3 min for the primer set with Tpi_IR1), 94 ˚C for 1 min

followed by 40 cycles of 94 ˚C for 30 sec, 42 ˚C for 30 sec, 65 ˚C for 45 sec, and finally 65 ˚C for 3 min for *Period*, and 94 ˚C for 2 min followed by 40 cycles of 94 ˚C for 30 sec, 57 ˚C for 45 sec, 72 ˚C for 2 min, and finally 72 ˚C for 5 min for *Ldh*. Although all the samples (78 *Juglans* moths and 68 *Lyonia* moths for the Sendai population, and 44 *Juglans* moths and 40 *Lyonia* moths for the Niimi population) were used for the PCR reactions for *Ldh*, *Period* and *Tpi*, we could not obtain amplified fragments from some samples (Table S1, Supporting Information). For the PCR reactions for the COI and ND5, we used moths whose partial sequences of the *Tpi*, *Period* and/or *Ldh* were amplified, and we consequently obtained the fragments from 56 *Juglans* and from 60 *Lyonia* moths for the Sendai samples, while using all moths for the Niimi samples. We could not amplify the *Period* and *Ldh* fragments for one of the outgroup samples (IO-087), so only one outgroup (IO-095) was included in the *Period* and *Ldh* dataset.

The amplified products were purified using a PCR Purification Kit (Qiagen) or EconoSpin (GeneDesign). Because lepidopteran females are heterogametic in sex chromosomes and because each of the *Tpi*, *Period* and *Ldh* loci can be regarded as a single-copy nuclear gene (Naisbit *et al*. 2002), *A. transecta* females carry only one allele for the Z-linked loci (Fig. 1). Amplicons could thus be sequenced without cloning. The PCR products were sequenced on an ABI3130 sequencer (Applied Biosystems) using the BigDye sequencing kit (ver. 3.1) or the CEQ2000 DNA Analysis System (Beckman Coulter). The alignments of COI and ND5 were straightforward and required no gaps, and were aligned manually. Sequences of *Period* and *Tpi* were aligned using ClustalX 1.5a (Thompson *et al*. 1997) with Gap-open costs at 20 and Gap-extension costs at 1, and then optimized by eye. The obtained sequences have been deposited in the GenBank database (GU809539 - GU809738 [COI], GU809339 -

GU809538 [ND5], GU809843 - GU809977 [*Tpi*], GU809739 - GU809842 [*Period*], GU809242 - GU809338 [*Ldh*]).

Detection of recombination

Since the *Ldh*, *Period* and *Tpi* sequences contained various insertions and deletions (indels), we excluded gaps from the raw datasets (no indels were observed in any mitochondrial sequences). We also conducted phylogenetic analyses using the datasets, including gaps for *Tpi*, *Period* and *Ldh* by a maximum likelihood method. The results (not shown) were identical to those using the datasets excluding gaps; therefore, we used the datasets excluding gaps for the present study. Recombinations were inferred using the algorithms implemented in the IMgc software package (Woerner *et al*. 2007), and we detected the longest nonrecombining region for each locus. We applied the same methods for the mitochondrial genes but did not detect any evidence of recombination events. NEXUS files of the aligned sequences are available from the URL [http://insect3.agr.hokudai.ac.jp/~issei/data/] or by request to the first author.

Molecular population genetics

We used DnaSP 5.10.00 (Librado & Rozas 2009) for basic polymorphism analyses. Numbers of haplotypes (h), total numbers of polymorphic sites (S), average pairwise diversity per nucleotide, π (Tajima 1983) and divergence (D_{xy}) between the *Juglans* and *Lyonia* races based on π (Nei 1987) were estimated. In addition, in order to test for departures from neutrality, Tajima's D (Tajima 1989) was calculated using DnaSP.

Phylogenetic analysis

On each dataset, maximum likelihood (ML) and parsimony (MP) analyses were conducted in PAUP* 4.0b10 PPC (Swofford, 2002) and Bayesian MCMC analyses were conducted in MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). Parameters for the ML analysis were chosen on the basis of the Akaike Information Criterion (AIC) (Akaike 1974; Sakamoto *et al*. 1986) as implemented in Modeltest 3.06 (Posada & Crandall 1998; Posada & Buckley 2004) (Table S2, Supporting Information). For the *Tpi*, *Period* and *Ldh* genes, in order to remove effects of recombination, we also conducted the ML analyses using the longest nonrecombining regions, which were inferred by IMgc. Parameters for these longest nonrecombining region are listed in Table S3, Supporting Information. The ML tree of each dataset was searched with TBR branch swapping using a neighbor-joining tree as a starting point. For the MP analyses, all characters were equally weighted. MP trees were searched with 100 random addition replications using TBR branch swapping. To assess confidence limits in clades for ML and MP analyses, nonparametric bootstrap tests (Felsenstein, 1985) were performed using 100 replicates with TBR branch swapping. However, due to computational difficulties under the ML methods, bootstrap supports for the mitochondrial dataset of the Sendai population and the longest nonrecombining regions of the Z-linked-gene datasets were calculated using 1000 replicates with NNI rearrangements by PhyML 3.0 (Guindon & Gascuel 2003). For the Bayesian analyses, we performed two independent runs for each dataset, and each run consisted of four chains for 5,000,000 generations, and a tree was sampled every 1,000 generations. The

first 1,000 trees were excluded as a burn-in period, and we composed a 50% majority consensus tree of the remaining trees to estimate posterior probabilities of branches in the tree. Substitution models for Bayesian analyses were estimated using MrModeltest 2.3 (Nylander, 2004) based on the AIC criterion.

Statistical test for historical rates of introgression

The isolation with migration analytic (IMa) model implemented in the program IMa (Hey & Nielsen 2004, 2007) was adopted to estimate introgression between the two putative host races. This method uses DNA sequences from a pair of populations to infer six demographic parameters (population sizes of both extant populations as well as the ancestor population, time since divergence, and per gene migration rates in both directions). Since the IMa analysis cannot accommodate regions that exhibit evidence of recombination in DNA alignments (Hey & Nielsen 2004), we used the longest nonrecombining region of the *Tpi*, *Period* and *Ldh* datasets inferred by IMgc. Because the *Tpi*, *Period* and *Ldh* loci are located on the same chromosome (Z chromosome), this combined treatment still violates the requirement that such loci should be segregated independently. However, these loci could be considered to be substantially unlinked due to recombination when a long time scale is assumed (Hey & Nielsen 2006). The IMa analysis is applicable to datasets from multiple loci (Hey & Nielsen 2004, 2007), so we analyzed the combined dataset of *Tpi*, *Period*, *Ldh* and mitochondrial genes and the dataset of the *Tpi*, *Period* and *Ldh* genes, as well as each dataset of the respective genes. Thus, we analyzed six datasets: (1) mitochondrial COI+ND5 (referring to mt), (2) *Tpi*, (3) *Period*, (4) *Ldh*, (5) *Tpi* + *Period* + *Ldh* (referring to Z-linked) and (6) mt +

Tpi + *Period* + *Ldh* (referring to mt + Z-linked). Inheritance scalars, which are per locus effective population sizes relative to those for autosomal loci, were set at 0.25 for mitochondrial COI + ND5 and 0.75 for Z-linked *Tpi*, *Period* and *Ldh*. For each of the six parameters in each analysis, we recorded the marginal probability density over the course of the simulations. We assumed the following prior distribution for each of the six parameters: for population sizes of both extant and ancestor populations, 20; for per gene migration rates in both directions, and for time since divergence, 10. The peak of the resulting distribution of the marginal probability density was taken as the maximum-likelihood estimates (MLE) of the parameter (Nielsen & Wakeley 2001). Since the lower limit of the bin was 0.0050 for the migration parameter, we interpreted the MLE at the lower limit as being zero (Nielsen & Wakeley 2001). The HKY model was chosen for each locus because it is the most complex model currently available in IMa. For each dataset, we ran five independent simulations. Individual simulations were run for 5,000,000 steps following 1,000,000 burn-in steps with five Metropolis-coupled chains (Geyer 1991) with heating scheme commands, -ft -g1 0.05 -g2 2, but for the Z-linked + mt dataset each simulation was run for 10,000,000 steps following 2,000,000 burn-in steps. The analysis was considered to have converged upon a stationary distribution if the independent runs showed effective sample size (ESS) values above 100 (Kuhner & Smith 2007), and all runs satisfied this criterion. In order to test statistical differences of gene-migration-rate values between the two directions as well as differences between the values and zero, we conducted log-likelihood ratio tests with nested models (L mode option in IMa) for the two combined datasets (Z-linked and mt + Z-linked).

Although the IMa model is robust to population structure with regard to

discriminating introgression from persistent ancestral polymorphism, estimates of effective population sizes and divergence time will be distorted in unpredictable ways (Whitlock & Barton 1997; Wakeley 2000). Therefore, we make no attempt to convert the scaled estimates of population sizes or time to units of real individuals or years, respectively (Won & Hey 2005). Although we did not obtain PCR products for more than 50% of most of the Z-linked markers, coalescent simulations are robust to sampling bias (Felsenstein 2004) and we successfully amplified all PCR products for mitochondrial genes, thus low PCR yields in Z-linked markers should not hamper the conclusion about higher gene flow in nuclear genes.

Results

Molecular population genetics

Polymorphism analyses for the two putative host races are summarized in Table S4, Supporting Information. In general, the *Juglans* race has more haplotypes than the *Lyonia* race, except for *Ldh* in both populations and *Tpi* in the Niimi population. Also, the *Juglans* race showed more nucleotide variation in all loci, indicating larger effective population sizes of the race. Except for *Ldh*, the average nucleotide difference per site between the putative host races (D_{xy}) was substantially greater than the average difference per site (π) for at least one of the two races. However, *Ldh* showed similar D_{xy} values as π values in both races and in both populations. The results of Tajima's *D* (TD) demonstrated a tendency toward negative values (15 of 16 values were negative), implying the recent population expansion of *A. transecta*. However, none of the 16

values of TD within the putative host races were significantly deviated from the neutral expectation of zero, satisfying the assumption of neutrality for coalescent simulations.

Phylogenetic analysis

Gene genealogies estimated by the ML, MP and Bayesian analyses were highly concordant for all loci so that only ML trees are shown (Figs. 2, 3). For mitochondrial datasets of both Sendai and Niimi populations, the *Lyonia* race composed a monophyletic group with strong support and was clearly distinguished from the *Juglans* race. However, for each dataset of the Z-linked genes, the two putative host races were not clearly distinguished. One or more *Juglans* moths were clustered with the *Lyonia* moths with strong support in both Sendai and Niimi populations (Figs. 2B-D, 3B-D). Phylogenetic analyses of the longest nonrecombining region of *Tpi*, *Period* and *Ldh* each yielded almost identical results to those using the datasets with effects of recombination (Figs. S1, S2, Supporting Information). In order to confirm the results of shared haplotypes between the two putative host races, total DNA was extracted from the voucher specimens of the samples (the abdomen of each moth). Analyses of both mitochondrial and Z-linked genes using new extractions resulted in identical gene sequences.

Statistical test for historical rates of introgression

When each of the six datasets (mt, *Tpi*, *Period*, *Ldh*, Z-linked, and mt + Z-linked) was analyzed by IMa, the marginal posterior probability for each of the six parameters

converged on a similar distribution after five independent runs. For the mitochondrial genes datasets, the MLE for the migration-rate parameters (gene-flow rates per locus per generation per neutral mutation rate) was at the lower limit (0.0050) in both directions after the five runs in both of the populations (Fig. 4; Tables S5, S6, Supporting Information). In contrast, the MLE for migration parameters of the *Tpi* datasets clearly differed between the two directions in both the populations. The mean value of MLE for the rate of migration from the *Lyonia* race to the *Juglans* race was above zero; however, the MLE for migration in the opposite direction was at the lower limit 0.0050 in all five runs in both populations (Fig. 4; Tables S5, S6, Supporting Information). For both *Period* and *Ldh*, the MLE values for migration were positive in both directions, and the MLE values for *m*1 (migration from the *Lyonia* race to the *Juglans* race) were higher than those for *m*2 (migration from the *Juglans* race to the *Lyonia* race), except for *Period* in the Niimi population (Fig. 4; Tables S5, S6, Supporting Information). The combined dataset of the three Z-linked genes (Z-linked) also yielded similar results in both of the populations; both migration rates were positive, and *m*1 was higher than *m*2 (Fig. 4; Tables S5, S6, Supporting Information). Interestingly, this trend was also recovered when the combined dataset of mitochondrial and Z-linked genes (mt + Z-linked) was analyzed (Fig. 4; Tables S5, S6, Supporting Information).

The log-likelihood ratio tests with nested models did not reject the zero-migration-rate model (*m*1 or *m*2 = 0) in both directions for mitochondrial datasets in both of the populations (Table 1). These results suggest that introgression has not occurred in the mitochondrial genes in either direction in recent time periods. In contrast, for both the Z-linked and mt + Z-linked datasets, the zero-migration-rate model was rejected in both directions except for *m*2 of the Z-linked dataset in the Niimi population (Table 1).

However, although all MLEs for $m1$ were higher than $m2$ for both the Z-linked and mt + Z-linked datasets, the identical-migration-rate model was not rejected. Therefore, the results lead to the conclusion that, in the Z-linked genes, gene flow has occurred in both directions.

Interestingly, when CJ45 and 90 were removed from the dataset of Z-linked genes in the Sendai population, the MLE for $m1$ was at zero (0.0050, the lower limit) in all five runs (but not at zero for $m2$), although positive migration rates were recovered when only CJ90 was deprived. However, both $m1$ and $m2$ were not at zero when NJ2, 7, 35, 36, 37, 38 and 41 were removed from the Z-linked dataset. Estimates of effective population sizes varied between datasets. However, the results of all datasets indicated that the effective population size of the *Juglans* race was larger than that of the *Lyonia* race (Fig. S3, Supporting Information; Tables S5, S6, Supporting Information). This is consistent with the result that the *Juglans* race has more haplotypes than the *Lyonia* race (Table S4, Supporting Information).

Discussion

Alleles unique to a population or species are sometimes shared by some members of a different population or species because of the persistence of ancestral polymorphism or introgression between them. In cases of such phenomena, the extent of introgression and of ancestral polymorphism differs among genes, so that gene trees for these populations or species are sometimes incongruent depending upon the genetic region examined (Ting *et al*. 2000; Beltrán *et al*. 2002; Shaw 2002; Machado & Hey 2003; Dopman *et al*. 2005; Bull *et al*. 2006; Kronforst *et al*. 2006). Hey & Nielsen (2004)

estimated gene migration rates between *Drosophila pseudoobscura* and *D. persimilis* based on fourteen nuclear loci and mitochondrial COI + ND4, and they showed that five nuclear loci and mitochondrial genes introgressed rather freely between species, whereas the remaining loci showed strong evidence of isolation. However, it is usually difficult to discriminate introgression from persistent ancestral polymorphism (Hey *et al*. 2004; Mallet 2005).

The present phylogenetic analyses demonstrated discordant genealogies between the mtDNA and the Z-linked nuclear genes in *A. transecta* (Figs. 2, 3). Under the ancestral-polymorphism scenario, introgression between races is limited in the genomes, and the phylogeny will be constructed based on those genomic regions where frequent extinctions of gene genealogies resulted in the loss of polymorphism in one race (i.e., lineage sorting). Assuming that the sexes are equal in number, the effective population size for mitochondrial genes is one-third of that for Z-linked loci (Birky Jr. *et al*. 1983). Therefore, mitochondrial genes should be more sensitive to the effects of lineage sorting than the Z-linked *Tpi*, *Period* and *Ldh* genes. This possibility, coupled with the phylogenetic position of CJ45 in the Sendai population in the *Tpi* trees (Fig. 2B) and the positions of NJ7, 35, 36 and 41 in the Niimi population in the *Period* trees (Fig. 3C), suggests that the present phylogenetic discordance may be partially due to ancestral polymorphism in the Z-linked genes. However, the ML analysis based on *Tpi* and *Period* showed that the CJ90 haplotype was embedded in the *Lyonia* clade and clustered with 18 *Lyonia* moths in the Sendai population (Fig. 2B,C). Similarly, in the ML tree of *Tpi* in the Niimi population NJ38 was embedded in the *Lyonia* race (Fig. 3B). These results are inconsistent with the ancestral-polymorphism scenario.

Alternatively, the genealogical discordance can be explained if introgression of

mitochondrial genes is hindered, whereas gene exchange is not restricted in Z-linked genes. IMa analyses detected greater introgression in the three Z-linked genes than in the mitochondrial genes in both the Sendai and Niimi populations (Fig. 4; Tables S5, S6, Supporting Information), strongly supporting the differential introgression scenario between the Z-linked and mitochondrial genes. When assuming a zero migration rate between populations in IMa, all shared polymorphisms should be the result of ancestral polymorphisms. IMa significantly rejected the zero-migration-rate model for Z-linked datasets except for *m*2 (from the *Juglans* race to the *Lyonia* race) in the Niimi population; thus these results indicate that the ancestral polymorphism scenario is not sufficient for explaining all the presently shared polymorphisms. Furthermore, migration estimates for the Z-linked dataset in the Sendai population were at zero when both the CJ45 and 90 were removed from the dataset. This implies that CJ45 is a putative hybrid as well as CJ90. Thus, the present discordance is most likely to be interpreted as the result of completely restricted introgression in mitochondrial genes and a low level of gene flow in Z-linked genes. Also, the presence of 1-2 putative hybrids in 78 individuals is compatible with host race status (> 1% hybridization, Drès & Mallet [2002]).

In sympatric or parapatric populations, restricted introgression in a genetic marker often is because of a positive association between that genetic marker and the genomic regions responsible for local adaptation or reproductive isolation (Machado & Hey 2003; Colosimo *et al*. 2005; Dopman *et al*. 2005). In the present system, restricted introgression in mitochondrial genes is best explained by the genetically based host preference of females. Females' strict host preference (Ohshima 2008) probably hinders introgression in maternally inherited genes between the two host races. In contrast, Z-

linked genes are not included in host adaptation because genes for host adaptations are not linked to sex chromosomes (Ohshima 2008). Haldane's rule might also lead to restricted introgression in mitochondrial genes because females are the heterogametic sex in Lepidoptera. However, the fact that both sexes of hybrids do not suffer from reduced fecundity (Ohshima unpubl. data) or intrinsic incompatibility (Ohshima 2008) negates the possiblity of Haldane's rule in *A. transecta*.

In other phytophagous insect systems, Putnam *et al*. (2007) estimated divergence times of *Papilio glaucus* and *P. canadensis* based on mitochondrial COI+COII and five separate Z-linked genes, but they did not find significant differences between the estimated divergence times, regardless of whether Z-linked genes or mitochondrial genes were used. Between *P. glaucus* and *P. canadensis*, host preferences of ovipositing females are not differentiated (Hagen 1990), and maternally inherited genes can introgress across the two species as well as neutral nuclear genes. In contrast, Smith *et al*. (2002) revealed that the mitochondrial gene tree of *Eurosta solidaginis* demonstrated the monophyly of the two host races, whereas the nuclear allozyme patterns did not distinguish the two host races completely (Waring 1990) which agrees with the results of our study. Peccoud *et al*. (2009) also demonstrated that sequences of maternally inherited bacterial endosymbiont genes from 11 host-associated biotypes in *Acyrthosiphon pisum* sort mostly into distinct matrilines despite their very recent diversification (3,600 – 9,500 years ago). These studies, as well as our results, suggest that restricted introgression in maternally inherited genes is widespread among closely related phytophagous insect taxa with inflexible host associations. Putnum *et al*. (2007) also showed that there was considerable variation in estimated divergence times depending upon the particular Z-linked genes used. Thus, future studies based on other

Z-linked and autosomal genes will help to elucidate the extent of gene flow across the genomes in *A. transecta*.

In contrast to the expectation, our coalescent results indicate that gene flow has occurred in both directions between the two host races. Although previous studies revealed that F_1 hybrid larvae in *A. transecta* cannot survived on *Lyonia* without exception (Ohshima 2008), half of offspring of the backcross with the *Lyonia* race can survive on *Lyonia* (Ohshima unpubl. data) because the gene for larval performance is autosomal and the *Lyonia* feeding allele is recessive (Ohshima 2008). Given recombinations between Z chromosomes in F_1 hybrid males, all the viable offspring possess *Juglans*-race-type Z-linked loci, and, even if there are no recombinations between Z chromosomes, half of the viable backcross offspring possess *Juglans*-race-type Z-linked alleles. Therefore, there would be an opportunity for a breakdown between the larval performance and the marker Z-linked genes, resulting in gene flow from the *Juglans* race to the *Lyonia* race through backcross matings between F_1 hybrid males and *Lyonia* females.

A difference in population sizes between adjacent populations often causes asymmetric gene flow; asymmetrical gene flow can occur from large to small populations (Nosil 2004; Palstra *et al*. 2007). In the present case, the estimated effective population size of the *Lyonia* race was smaller than that of the *Juglans* race (Tables S5, S6, Supporting Information), indicating that gene flow occurred from a small to a large population as well as in the opposite direction. This implies that in *A. transecta* the differential fitness of F_1 hybrids on the two host plants is also affecting the direction of gene flow. Kronforst *et al*. (2006) also reported that ecological phenotypes in F_1 hybrids determining the direction of gene flow between mimic butterflies

Heliconius cydno and *H. pachinus*.

The present analyses detected only a few putative hybrid individuals, despite high mating and genomic compatibility between the two host races. This suggests that gene flow between the two host races is largely hindered by other isolating barriers. Small-sized phytophagous insects usually mate on or near the host plant (Bush 1975; Berlocher & Feder 2002; Matsubayashi *et al*. 2010). For this reason, differences in host use may directly affect the choice of mating sites (Craig *et al*. 1993; Feder *et al*. 1994; Via 1999; Emelianov *et al*. 2003; Hirai *et al*. 2006; Matsubayashi & Katakura 2009), resulting in pre-mating isolation between host races. In *A. transecta*, Ohshima (2010) revealed that *Lyonia* females significantly prefer *Lyonia* as mating sites and that in the *Juglans* race the proportion of successful matings was significantly reduced when the host was absent. Therefore, these results, coupled with the present results of the molecular analyses, suggest that the combined effects of the mating-site preference of *Lyonia* females and the mating propensity of the *Juglans* race contribute to reproductive isolation between the two host races.

In conclusion, our study clearly demonstrates that maternally and biparentally inherited genes have different histories of divergence in this host-race system, and they suggest that distinct female host preference leads to the bifurcation of maternally inherited loci even in the presence of gene flow. If this is generally true in other host races or closely related species with different host associations, phylogenetic reconstruction based only on mitochondrial genes must be used with great caution. The estimation of genealogies from multiple nuclear loci coupled with genome-wide methods (e.g., using AFLP; Storz 2005) will help to elucidate the genetic architecture of host specialization. In addition, although the present analyses have used just two

sympatric populations, analyses including allopatric populations (i.e., samples from a locality where only the *Juglans* race is distributed) as well as other sympatric populations will reveal the detailed genomic architecture of the two host races.

Acknowledgements

We thank T. Kumata for many insights into the evolution of *Acrocercops*, S. Akimoto and P. Nosil for discussions and comments on the draft manuscript, J. Peccoud, M. L. Niemiller and one anonymous reviewer for valuable comments on the manuscript, and M. Ôhara and Y. Saito for allowing the use of their laboratories. I.O. thanks R. Deguchi, T. Fukushi, N. Goto, Y. Hirabuki, K. Mizota and A. Munakata for allowing him to use their laboratory at Miyagi University of Education for rearing, Y. Ohmizu for laboratory assistance, as well as the NIBB Center for Analytical Instruments for DNA sequencing. Part of this work was carried out using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by the Microsoft Corporation. This study was supported by Research Fellowships of the Japan Society for the Promotion of Science (JSPS) for Young Scientists (16-9250 and 20-5555), the Hokkaido University Clark Foundation, the Saito Gratitude Foundation to I.O. and a Grant-in Aid from JSPS (18770058) to K.Y.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Bachtrog D, Thornton K, Clark A, Andolfatto P (2006) Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution*, **60**, 292-302.

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

Beltrán M, Jiggins CD, Bull V *et al*. (2002) Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. *Molecular Biology and Evolution*, **19**, 2176-2190.

Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomol*ogy, **47**, 773-815.

Birky Jr CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, **103**, 513-527.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution*, **23**, 773-783.

Boom R, Sol CJ, Salimans MM *et al*. (1990) Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, **28**, 495-503.

Brunner PC, Chatzivassiliou EK, Katis NI, Frey JE (2004) Host-associated genetic differentiation in Thrips tabaci (Insecta; Thysanoptera), as determined from mtDNA sequence data. *Heredity*, **93**, 364-370.

Bull V, Beltrán M, Jiggins CD *et al*. (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biology*, **4**,11.

Bush GL (1975) Sympatric speciation in phytophagous parasitic insects. In:

Evolutionary Strategies of Parasitic Insects and Mites, (ed Price PW), pp. 187-206. Plenum Press, New York.

Campbell D, Bernatchez L (2004) Genetic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, **21**, 945-956.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720-729.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, **70**, 155-174.

Clarke B, Johnson MS, Murray J (1996) Clines in the genetic distance between two species of island land snails: how 'molecular leakage' can mislead us about speciation. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, **351**, 773-784.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal Jr G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, **307**, 1928-1933.

Craig TP, Itami JK, Abrahamson WG, Horner JD (1993) Behavioral evidence for host-race formation in *Eurosta solidaginis*. *Evolution*, **47**, 1696-1710.

Diegisser T, Johannesen J, Lehr C, Seitz A (2004) Genetic and morphological differentiation in *Tephritis bardanae* (Diptera: Tephritidae): evidence for host-race formation. *Journal of Evolutionary Biology*, **17**, 83-93.

Diegisser T, Seitz A, Johannesen J (2006) Phylogeographic patterns of host-race

evolution in *Tephritis conura* (Diptera: Tephritidae). *Molecular Ecology*, **15**, 681-694.

Diehl SR, Bush GL (1984) An evolutionary and applied perspective of insect biotypes. *Annual Review of Entomology*, **29**, 471-504.

Dopman EB, Bogdanowicz SM, Harrison RG (2004) Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). *Genetics*, **167**, 301-309.

Dopman EB, Pérez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14706-14711.

Drès M, Mallet J (2002) Host-races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society London B: Biological Science*, **357**, 471-492.

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

Emelianov I, Simpson F, Narang P, Mallet J (2003) Host choice promote reproductive isolation between host races of the larch budmoth *Zeiraphera diniana*. *Journal of Evolutionary Biology*, **16**, 208-218.

Fay J, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.

Feder JL, Opp S, Wlazlo B *et al*. (1994) Host fidelity is an effective pre-mating barrier

between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 7990-7994.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783-791.

Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associations, Massachustts.

Forister ML, Nice CC, Fordyce JA, Gompert Z, Shapiro AM (2008) Considering evolutionary processes in the use of single-locus genetic data for conservation, with examples from the Lepidoptera. *Journal of Insect Conservation*, **12**, 37-51.

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

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **173**, 919-933.

Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: Computing science and statistics: proceedings of the 23rd symposium on the interface, (ed Keramidas EM), pp. 156-163. Interface Foundation of North America, Seattle.

Gompert Z, Forister ML, Fordyce JA, Nice CC (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, **17**, 5231-5244.

Hagen RH (1990) Population structure and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). *Evolution*, **44**, 1914-1930.

Hey J, Nielsen R (2004) Multilocus methods for estimating population size, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747-760.

Hey J, Nielsen R (2006) *IM Documentation*. Available from http://lifesci.rutgers.edu/%7Eheylab/HeylabSoftware.htm.

Hey J, Won Y-J, Sivasundar A, Nielsen R, Markert JA (2004) Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology*, **13**, 909-919.

Hirai Y, Kobayashi H, Koizumi T, Katakura H (2006) Field-cage experiments on host fidelity in a pair of sympatric phytophagous ladybird beetles. *Entomologia Experimentalis et Applicata*, **118**, 129-135.

Jiggins CD, Linares M, Naisbit RE *et al*. (2001) Sex-linked hybrid sterility in a butterfly. *Evolution*, **55**, 1631-1638.

Jiggins CD, Mavarez J, Beltrán M *et al*. (2005) A genetic linkage map of the mimetic butterfly *Heliconius melpomene*. *Genetics*, **171**, 557-570.

Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, **60**, 1254-1268.

Kuhner MK, Smith LP (2007) Comparing likelihood and bayesian coalescent estimation of population parameters. *Genetics*, **175**, 155-165.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.

Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic Drosophila species group. Proceedings of the Royal Society of London B: Biological Science, **270**, 1193-1202.

Malenke JR, Johmson KP, Clayton DH (2009) Host specialization differentiates cryptic species of feather-feeding lice. *Evolution*, **63**, 1427-1438.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229-237.

Maroja LS, Andres JA, Harrison RG (2009) Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*, **63**, 2999-3015.

Matsubayashi KW, Katakura H (2009) Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles. *Evolution*, **63**, 2563-2580.

Matsubayashi KW, Ohshima I, Nosil P (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, **134**, 1-27.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. *Genetical Research*, **23**, 23-35.

Monteiro LR, Nogueira MR (2009) Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. *Evolution*, in press.

Naisbit RE, Jiggins CD, Linares M, Salazar C, Mallet J (2002) Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*, **161**, 1517-1526.

Narita S, Nomura M, Kato Y, Fukatsu T (2006) Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Molecular Ecology*, **15**, 1095-1108.

Nason JD, Heard SB, Williams FR (2002) Host-associated genetic differentiation in the goldenrod elliptical-gall moth, *Gnorimoschema gallaesolidaginis* (Lepidoptera: Gelechiidae). *Evolution*, **56**, 1475-1488.

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

NY.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a markov chain monte carlo approach. *Genetics*, **158**, 885-896.

Niu B-L, Lu S-L, Weng H-B, Shen W-F, Meng Z-Q (2005) Localization of the *Tpi* gene on Z chromosome of the silkworm, *Bombyx mori* by real-time quantitative PCR. *Acta Entomologia Sinica*, **48**, 622-626 (in Chinese with English summary).

Nosil P (2004) Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society of London B*, **103**, 9090-9095.

Nosil P, Crespi BJ (2006) Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 9090-9095.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differetiation between walking-stick ecotypes: 'isolation by adapation' and multiple roles for divergent selection. *Evolution*, in press.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18,** 375-402.

Nylander JAA (2004) *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Ohshima I (2005) Techniques for continuous rearing and assessing host preference of a multivoltine leaf-mining moth, *Acrocercops transecta* (Lepidoptera: Gracillariidae). *Entomological Science*, **8**, 227-228.

Ohshima I (2008) Host race formation in the leaf-mining moth *Acrocercops transecta* (Lepidoptera: Gracillariidae). *Biological Journal of the Linnean Society*, **93**,

135-145.

Ohshima I (2010) Host-associated pre-mating reproductive isolation between host races of *Acrocercops transecta*: mating site preferences and effect of host presence on mating. *Ecological Entomology*, in press.

Ohshima I, Yoshizawa K (2006) Multiple host shifts between distantly related plants, Juglandaceae and Ericaceae, in the leaf-mining moth *Acrocercops leucophaea* complex (Lepidoptera: Gracillariidae). *Molecular Phylogenetics and Evolution*, **38**, 231-240.

Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology*, **16**, 4504-4522.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793-808.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Putnam AS, Scriber JM, Andolfatto P (2007) Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution*, **61**, 912-927.

Ramsey J, Robertson A, Husband B (2008) Rapid adaptive divergence in new world *Achillea*, an autopolyploid complex of ecological races. *Evolution*, **62**, 639-653.

Roe AD, Sperling FAH (2007) Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. *Molecular Ecology*, **16**,

3617-3633.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

Sakamoto Y, Ishiguro M, Kitagawa G (1986) *Akaike information criterion statistics*. Springer, New York.

Savolainen V, Anstett M-C, Lexer C *et al*. (2006) Sympatric speciation in palms on an oceanic island. *Nature*, **441**, 210-213.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737-741.

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16122-16127.

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, **17,** 1246-1247.

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Systematic Biology*, **51,** 492-508.

Singer MC, McBride CS (2009) Multitrait, host-associated divergence among sets of butterfly ppopulations: implications for reproductive isolation and ecological speciation. *Evolution*, in press.

Smith PT, Krager K, Cronin JT, Kambhampati S (2002) Mitochondrial DNA variation among host races of *Eurosta solidaginis* Fitch (Diptera: Tephritidae). *Molecular Phylogenetics and Evolution*, **25**, 372-376.

Stireman III JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in

phytophagous insects: general phenomenon or isolated exceptions? evidence from a goldenrod-insect community. *Evolution*, **59**, 2573-2587.

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

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4, Sinauer Assoc., Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*,**105**, 437-460.

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

Ting C-I, Tsaur S-C, Wu C-I (2000) The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. Proceedings of the National Academy of Sciences of the United States of America, 97, 5313-5316.

Thompson JD, Gibson TJ, Plewniak F *et al*. (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876-4882.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446-1457.

Wakeley J (2000) The effects of subdivision on the genetic divergence of populations and species. *Evolution*, **54**, 1092-1101.

Waring GL, Abrahamson WG, Howard DJ (1990) Genetic differentiation among host-associated populations of the gallmaker *Eurosta solidaginis* (Diptera: Tephritidae). *Evolution*, **44**, 1648-1655.

Whitlock CM, Barton NH (1997) The effective size of a subdivided population.

Genetics, **146**, 427-441.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, **14**, 611-619.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851-1853.

Won Y-J, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297-307.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14,** 851-865.

Yamashiro T, Yamashiro A, Yokoyama J, Maki M (2008) Morphological aspects and phylogenetic analyses of pollination systems in the Tylophora-Vincetoxicum complex (Apocynaceae-Asclepiadoideae). *Biological Journal of the Linnean Society*, **93,** 325-341.

Figures

Figure 1. Inheritance of mitochondrial and Z-linked genes in hybrid females. In Lepidoptera, females are heterogametic in sex chromosomes and have a single Z chromosome. When hybridization between a female of the *Juglans* race and a male of the *Lyonia* race occurs, all hybrid females possess the Z chromosome inherited from the paternal parent, the *Lyonia* male. All hybrids inherit mtDNA from the maternal parent, the *Juglans* female. Thus, analyses using only females enable us to detect genealogical discordance due to hybridization between the host races.

Figure 2. The maximum likelihood gene trees for the Sendai population. (A) mitochondrial COI + ND5 (-ln = 1465.40303), (B) Z-linked *Tpi* (2574.56073), (C) Z-linked *Period* (807.60084) and (D) Z-linked *Ldh* (1382.09037). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from *Juglans*; blue: from *Lyonia*). Nodes with 50% or greater bootstrap support are labeled. Key individuals in the *Juglans* race are labeled by the orange color.

Figure 3. The maximum likelihood gene trees for the Niimi population. (A) mitochondrial COI + ND5 (-ln = 1462.97356), (B) Z-linked *Tpi* (1620.87698), (C) Z-linked *Period* (829.96110) and (D) Z-linked *Ldh* (1635.86014). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from *Juglans*; blue: from *Lyonia*). Nodes with 50% or greater bootstrap support are labeled. Key individuals in the *Juglans* race are labeled by the orange color.

Figure 4. IMa analysis of migration rate estimates between the two sympatric host races: (A) the Sendai population, and (B) the Niimi population. The marginal posterior density distributions for the migration rate parameters (gene flow rates per gene copy per generation per neutral mutation rate) between host races are shown. Bidirectional introgression was estimated for each of six datasets. Posterior-density curves are means of the five independent simulations for each data set.

Table

Table 1. Log-likelihood ratio statistics for the zero-migration-rate ($m1$ or $m2 = 0$) and identical-migration-rate ($m1 = m2$) models.

Supplemental Figures

Figure S1. The maximum likelihood gene trees for the Sendai population using the nonrecombining region of each gene: (A) Z-linked *Tpi* (-ln = 827.973796), (B) Z-linked *Period* (511.024649) and (C) Z-linked *Ldh* (486.481697). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from *Juglans*, blue; from *Lyonia*). Nodes with 50% or greater bootstrap support are labeled.

Figure S2. The maximum likelihood gene trees for the Niimi population using the nonrecombining region of each gene: (A) Z-linked *Tpi* (-ln = 683.857079), (B) Z-linked

Period (471.350554) and (C) Z-linked *Ldh* (501.208436). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from *Juglans*; blue: from *Lyonia*). Nodes with 50% or greater bootstrap support are labeled.

Figure S3. IMa analysis of effective population size estimates for the two host races in the two populations: (A) the Sendai population, and (B) the Niimi population. Posterior-density curves are means of the five independent simulations for the mt + Z-linked dataset.

Supplemental Tables

Table S1. Primers for the Z-linked loci and numbers of individuals from which amplified fragments were successively obtained for each locus.

Table S2. Parameters for the ML analyses inferred by Modeltest 3.7.

Table S3. Parameters inferred by Modeltest 3.7 for the ML analyses with datasets of the nonrecombining region.

Table S4. Basic polymorphism analyses using datasets without indels.

Table S5. Parameter estimates from the IMa analyses for the Sendai population. Means and standard deviations (SD, in parentheses) of the maximum likelihood

estimates (MLE)† and of the 90% highest posterior density (HPD) intervals§ of the IMa model parameters for each dataset are shown.

Table S6. Parameter estimates from the IMa analyses for the Niimi population. Means and standard deviations (SD, in parentheses) of the maximum likelihood estimates (MLE)† and of the 90% highest posterior density (HPD) intervals§ of the IMa model parameters for each dataset are shown.

Table 1

Population	Dataset	Model					
		$m1 = 0$		$m2 = 0$		$m1 = m2$	
		2LLR (df = 1)	*P*	2LLR (df = 1)	*P*	2LLR (df = 1)	*P*
Sendai	mt	0.0049	0.9442	0.0007	0.9789	0.0002	0.9887
	Z-linked	8.8445	0.0029	15.7149	< 0.0001	0.0278	0.8676
	mt + Z-linked	39.1262	< 0.0001	8.3686	0.0038	0.5239	0.4692
Niimi	mt	0.0027	0.9586	0.0005	0.9822	0.0008	0.9774
	Z-linked	9.3729	0.0022	2.3032	0.1291	0.0135	0.9075
	mt + Z-linked	7.7173	0.0055	5.5106	0.0189	3.7795	0.0519

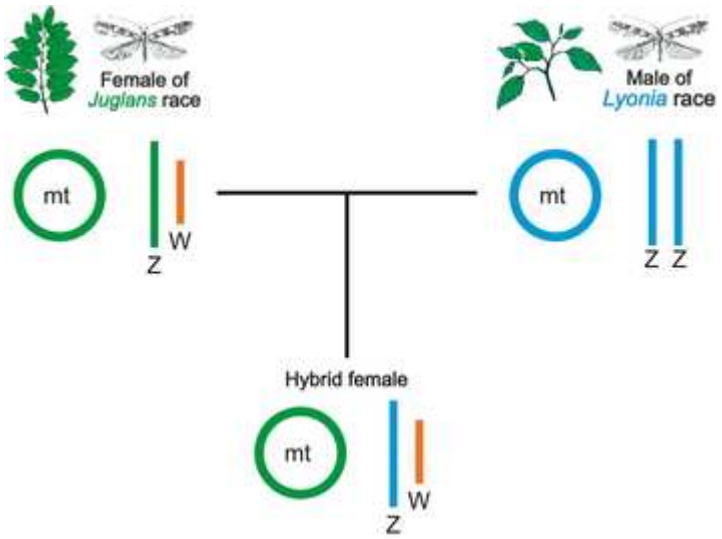

Figure 1. Inheritance of mitochondrial and Z-linked genes in hybrid females. In Lepidoptera, females are heterogametic in sex chromosomes and have a single Z chromosome. When hybridization between a female of the Juglans race and a male of the Lyonia race occurs, all hybrid females possess the Z chromosome inherited from the paternal parent, the Lyonia male. All hybrids inherit mtDNA from the maternal parent, the Juglans female. Thus, analyses using only females enable us to detect genealogical discordance due to hybridization between the host races.
15x11mm (600 x 600 DPI)

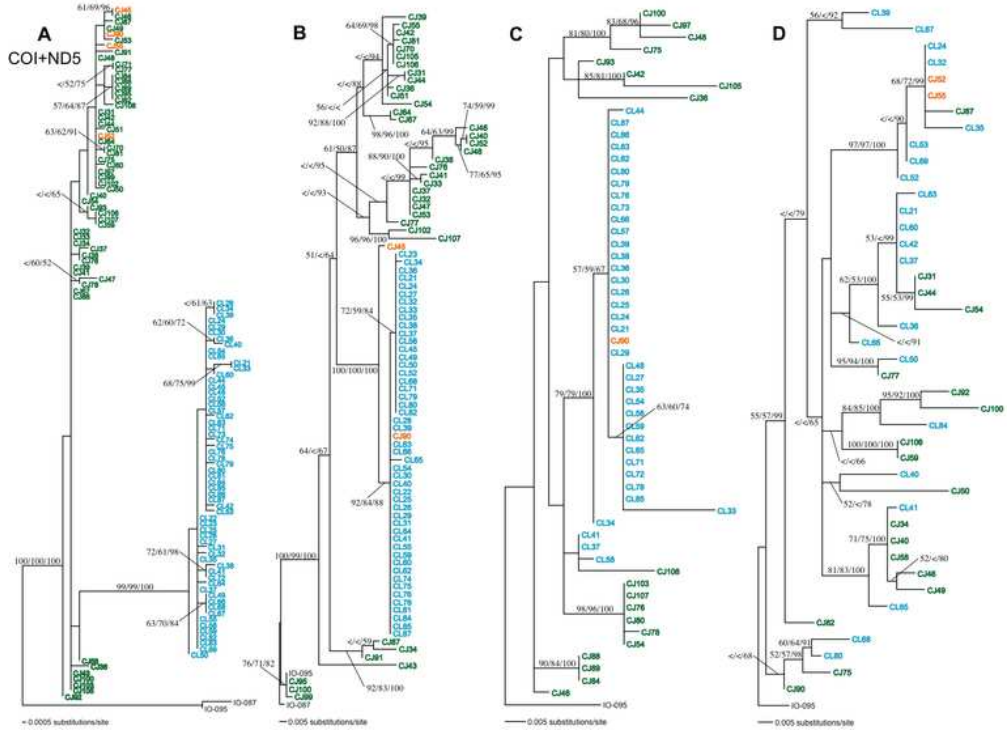

Figure 2. The maximum likelihood gene trees for the Sendai population. (A) mitochondrial COI + ND5 (-ln = 1465.40303), (B) Z-linked Tpi (2574.56073), (C) Z-linked Period (807.60084) and (D) Z-linked Ldh (1382.09037). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from Juglans; blue: from Lyonia). Nodes with 50% or greater bootstrap support are labeled. Key individuals in the Juglans race are labeled by the orange color.
32x24mm (600 x 600 DPI)

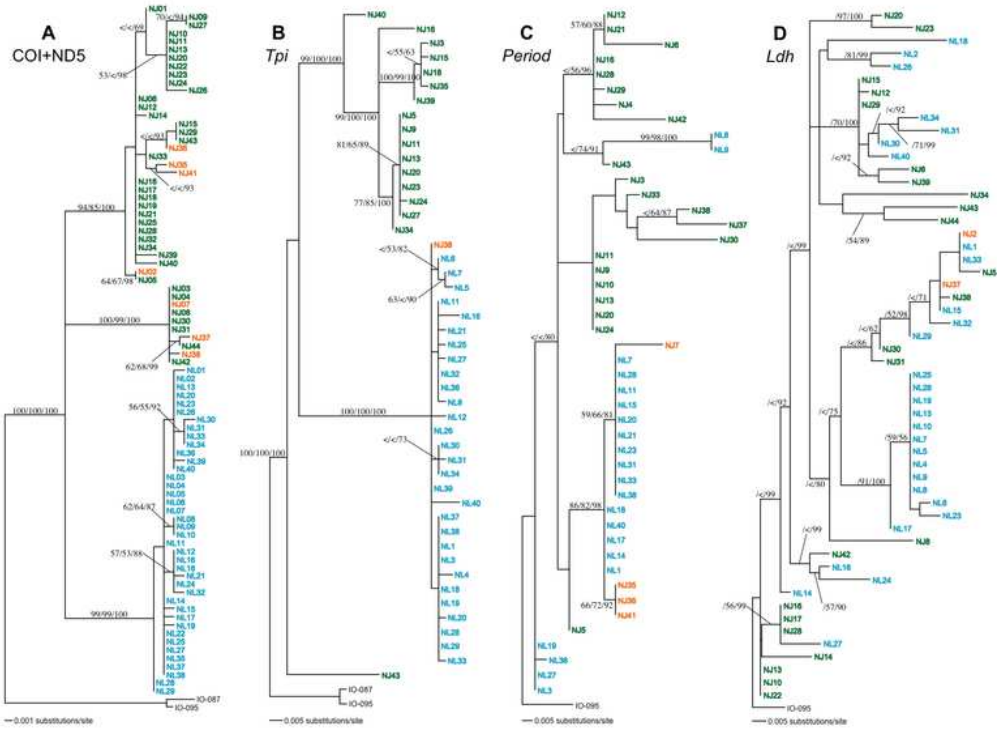

Figure 3. The maximum likelihood gene trees for the Niimi population. (A) mitochondrial COI + ND5 (-ln = 1462.97356), (B) Z-linked Tpi (1620.87698), (C) Z-linked Period (829.96110) and (D) Z-linked Ldh (1635.86014). Terminal tree branches are voucher numbers that are labeled with a host-specific color (green: from Juglans; blue: from Lyonia). Nodes with 50% or greater bootstrap support are labeled. Key individuals in the Juglans race are labeled by the orange color.
32x24mm (600 x 600 DPI)

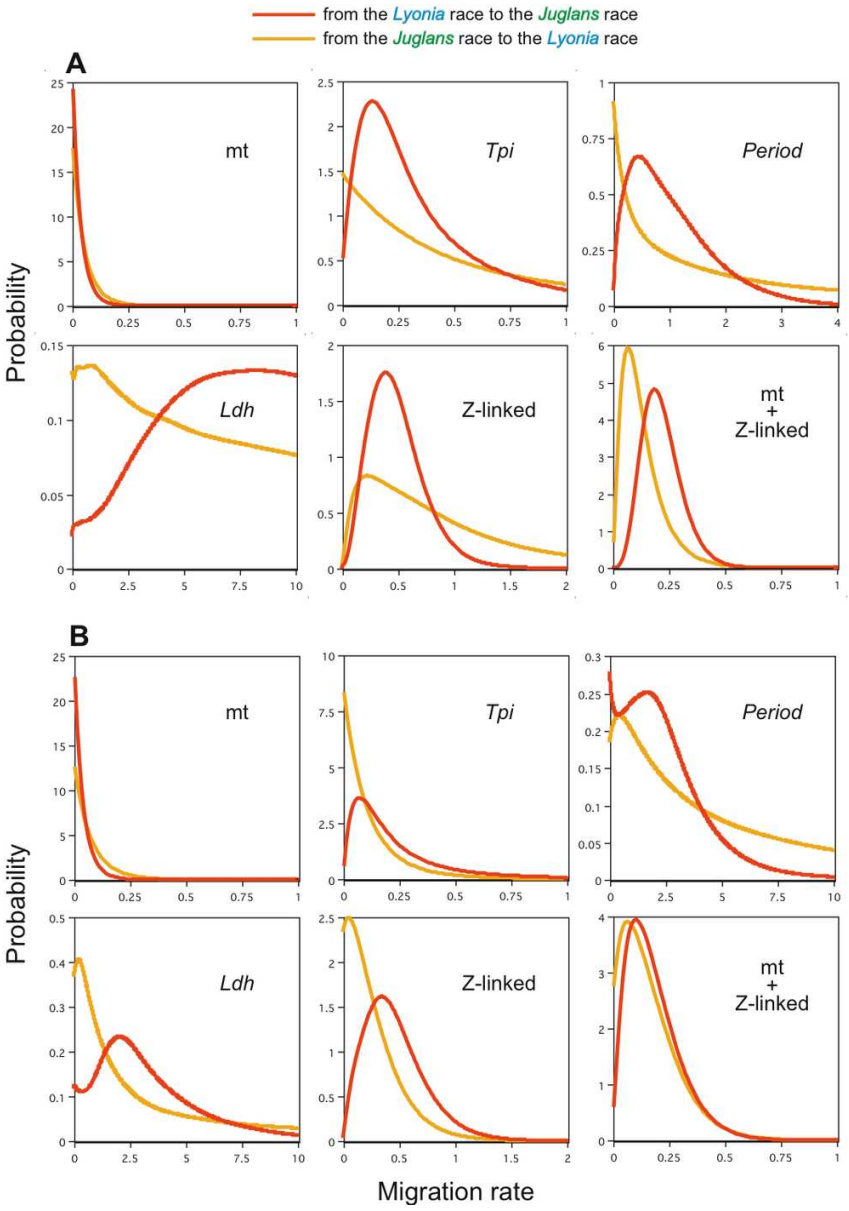

Figure 4. IMa analysis of migration rate estimates between the two sympatric host races: (A) the Sendai population, and (B) the Niimi population. The marginal posterior density distributions for the migration rate parameters (gene flow rates per gene copy per generation per neutral mutation rate) between host races are shown. Bidirectional introgression was estimated for each of six datasets. Posterior-density curves are means of the five independent simulations for each data set.
42x59mm (600 x 600 DPI)