MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4681–4694 doi: 10.1111/j.1365-294X.2012.05727.x

Speciation history of three closely related oak gall wasps, *Andricus mukaigawae*, *A. kashiwaphilus*, and *A. pseudoflos* (Hymenoptera: Cynipidae) inferred from nuclear and mitochondrial DNA sequences

NAKATADA WACHI,* YOSHIHISA ABE,† NOBUYUKI INOMATA,‡§ ALFRED EDWARD SZMIDT‡ and HIDENORI TACHIDA‡

*Graduate School of Systems Life Sciences, Kyushu University, 6-10-1 Hakozaki, Fukuoka, 812-8581, Japan, †Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, 744 Motooka, Fukuoka, 819-0395, Japan, ‡Department of Biology, Faculty of Sciences, Kyushu University, 6-10-1 Hakozaki, Fukuoka, 812-8581, Japan, §Department of Environmental Science, International College of Arts and Sciences, Fukuoka Women's University, 1-1-1 Kasumigaoka, Fukuoka, 813-8529, Japan

Abstract

The *Andricus mukaigawae* complex of oak gall wasps is composed of cyclically parthenogenetic species: *A. mukaigawae* and *Andricus kashiwaphilus*, and a parthenogenetic species, *Andricus pseudoflos*. The component species differ in life history, host plant, karyotype, and asexual gall shape, although little difference is found in the external morphology of asexual adults. To understand the speciation history of this species complex, DNA sequences of one mitochondrial region and nine nuclear gene regions were investigated. The genetic relationship among the species suggested that a loss of sex occurred after host shift. Unexpectedly, two or three distinct groups in the parthenogenetic species, *A. pseudoflos*, were revealed by both mitochondrial and nuclear DNA data. Gene flow in nuclear genes from the species not infected by *Wolbachia* (*A. kashiwaphilus*) to the species infected by it (*A. mukaigawae*) was suggested by a method based on coalescent simulations. On the other hand, gene flow in mitochondrial genes was suggested to be in the opposite direction. These findings indicate possible involvement of *Wolbachia* infection in the speciation process of the *A. mukaigawae* complex.

Keywords: gall-inducing insects, species complex, DNA polymorphism, population structure, *Wolbachia*

Received 26 March 2012; revision received 4 June 2012; accepted 18 June 2012

Introduction

The oak gall wasps (Cynipini) comprise around 1000 species that make galls on oaks (*Quercus*) and other related species of Fagaceae (Liljeblad & Ronquist 1998). Most species of Cynipini show cyclical parthenogenesis, i.e., alternation of sexual and asexual generations (Stone *et al.* 2002). Cyclical parthenogenesis is known to occur in seven groups of animals: Rotifera, Cladocera, Digenea, Aphidoidea, Cecidomyiidae, Micromalthidae, and Cynipidae (Hebert 1987; Suomalainen *et al.* 1987) but an obligate alternation of a sexual and an asexual generations is a characteristic of only cynipid wasps (Suomalainen *et al.* 1987). Within Cynipini, the *Andricus mukaigawae* complex provides a good opportunity to study differences in host plants, reproductive modes and their relationship with genetic divergence. The species in the complex differ little in adult morphology in the asexual generation and induce galls of asexual generation in the same location (bud) on their host plants, but they differ in chromosome number, female host preference, larvae ability for the gall induction, shape of the asexual gall, reproductive mode,

Correspondence: Nakatada Wachi, The Kyushu University Museum, 6-10-1 Hakozaki, Fukuoka 812-8581, Japan, Fax: +81 92 642 4252; E-mail: nwachi@kyudai.jp

and *Wolbachia* infection status (Abe 1986, 1988, 1991, 1998, 2007; Abe & Miura 2002). The cyclically parthenogenetic species *A. mukaigawae* (Mukaigawa) ($2n = 12$) induces bur-shaped galls of asexual generation on *Quercus aliena*, *Quercus mongolica*, and *Quercus serrata*. The other cyclically parthenogenetic species *Andricus kashiwaphilus* Abe ($2n = 10$) induces flower-shaped galls of asexual generation on *Quercus dentata*, and the obligatory parthenogenetic species *Andricus pseudoflos* (Monzen) (also $2n = 10$) produces flower-shaped galls on *Q. dentata* (Table 1). In the present study, we focus on two features of the complex, *Wolbachia*-infection and inclusion of an asexual species, revealed in the previous studies. Below, we explain these two features in more detail.

First, the complex seems to have experienced speciation without shifts in the host plant organ [such shifts are rare in the diversification of oak gall wasps (Cook *et al.* 2002)]. *Wolbachia* may have played an important role in the splitting of the *Wolbachia*-infected *A. mukaigawae* and *Wolbachia*-free *A. kashiwaphilus* (Abe & Miura 2002). The phenotypic effect of infection with *Wolbachia* in *A. mukaigawae* remains unknown, although its possible role in reproductive isolation between *A. mukaigawae* and *A. kashiwaphilus* was suggested (Abe & Miura 2002). However, most insects infected by *Wolbachia* show cytoplasmic incompatibility (CI) (Werren 1997; Stouthamer *et al.* 1999). Because CI reduces gene flow between populations, infection with *Wolbachia* has been suggested to enhance speciation by acting in combination with other isolation barriers (Werren 1998; Bordenstein 2003). Theoretical (Flor *et al.* 2007; Telschow *et al.* 2007) and empirical studies (Haine *et al.* 2006; Jaenike *et al.* 2006) showed that *Wolbachia* infection in some populations of various insect species can be stable and can even facilitate speciation.

In *Culex* mosquitos, *Drosophila* fruitflies, *Tetranychus* mites, *Diabrotica* beetles and *Tribolium* beetles (Bordenstein 2003 and references therein) and *Leptopilina heterotoma* (Thompson) (Cynipoidea: Figitidae) (Vavre *et al.* 2000), CI-inducing *Wolbachia* kills their offspring in the cross between a *Wolbachia*-infected male and an uninfected female, but in the reciprocal cross between an uninfected male and an infected female, CI does not occur allowing the survival of their offspring. This is called unidirectional CI. If *Wolbachia* infection causes unidirectional CI, then gene flow will also be unidirectional. So, if *Wolbachia* infection caused unidirectional CI in *A. mukaigawae* and *A. kashiwaphilus* as previously suggested (Abe & Miura 2002), we should observe unidirectional gene flow: from the *Wolbachia*-free species to the infected species at nuclear genes, but in the opposite direction at mitochondrial genes, as shown by Jaenike *et al.* (2006). We briefly explain here how this occurs under unidirectional CI. In this situation, four types of crosses can occur, depending on the sex of the migrant and the population where the cross occurs (see Fig. 1): (i) between an infected female resident and an uninfected male migrant in the infected population, (ii) between an uninfected female migrant and an infected male resident in the infected population, (iii) between an uninfected female resident and an infected male migrant in the uninfected population, and (iv) between an infected female migrant and an uninfected male resident in the uninfected population. Only crosses 1 and 4

Table 1 Key features of the species in the *Andricus mukaigawae* complex used in this study (Abe 1988; Abe & Miura 2002; Abe 2007)

	Life cycle, reproductive mode	Gall shape of asexual generation	Chromosome number ($2n$)	Host plant	Female host preference*	Larval performance†	*Wolbachia* infection
Andricus mukaigawae	Bivoltine Alternation of sexual and asexual generations	Bur	12	*Q. serrata* *Q. mongolica* *Quercus aliena* –	*Q. serrata* *Q. mongolica* –‡ –	*Q. serrata* *Q. mongolica* –‡ *Q. dentata*	+
Andricus kashiwaphilus	Bivoltine Alternation of sexual and asexual generations	Flower	10	*Q. dentata*	*Q. dentata*	*Q. dentata*	–
Andricus pseudoflos	Univoltine Only asexual generation	Flower	10	*Q. dentata*	na	na	–

na, not analyzed (Abe 1988).

*Host plant species that the adult female wasp showed ovipositional preference.

†Host plant species that the sexual galls were induced by the no-choice experiments.

‡The female host preference and the larval performance of *A. mukaigawae* to *Q. aliena* were not examined (Abe 1988).

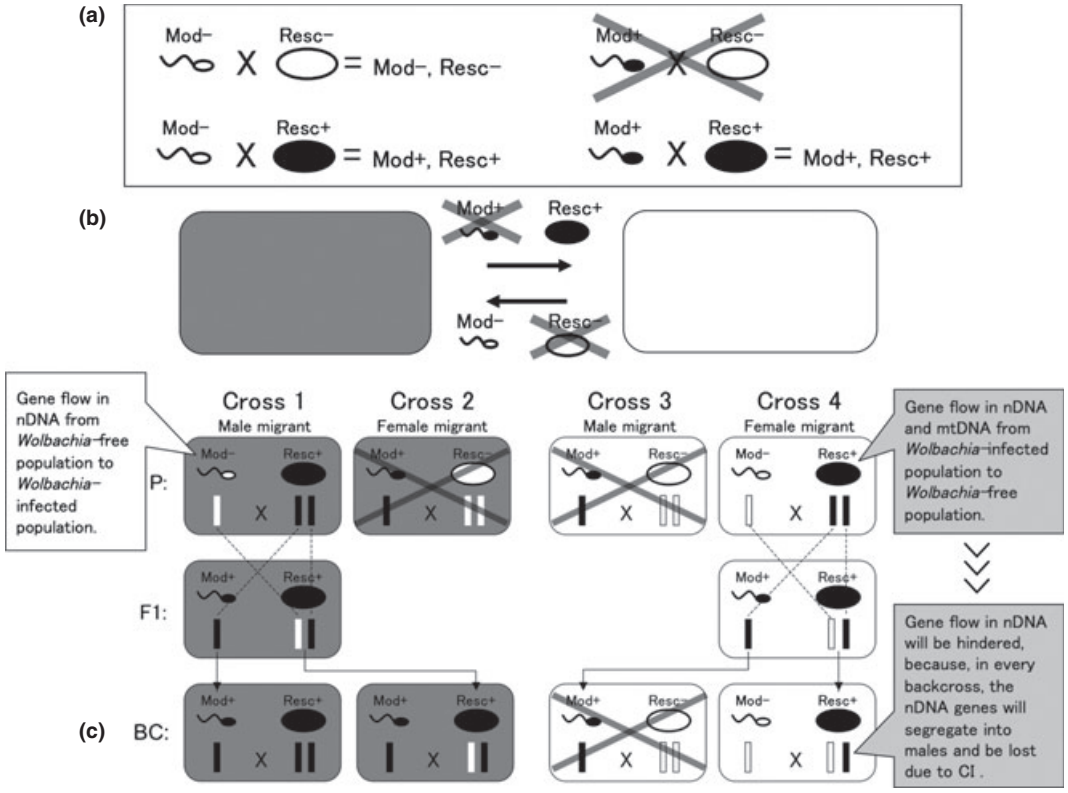

Fig. 1 The consequences of the cross involving *Wolbachia* infection. (a) Cytoplasmic incompatibility (CI) is assumed following Werren (1998). The phenotypic effects of *Wolbachia* infection are represented by Resc+ for female, and Mod+ for male. (b) Contact between *Wolbachia*-uninfected (open) and infected (filled) populations. (c) The crosses between Resc+ female and Mod− male (1 and 4) do not result in CI, while the reciprocal crosses between Resc− female and Mod+ male (2 and 3) result in CI. Diploid: female, haploid: male. P: parental generation, F1: hybrid generation, and BC: backcross generation.

do not cause CI and produce offspring. Cross 1 results in gene flow of nuclear DNA (nDNA) from the uninfected population to the infected population because migrants are male. Cross 4 results in gene flow of both nDNA and mitochondrial DNA (mtDNA) from the infected population to the uninfected population in the first generation, because migrants are female. However, the migrant nDNA cannot persist in the uninfected population. This is because in every backcross, half of the nuclear genes of the initial hybrids will segregate into males and be lost due to CI (Jaenike *et al.* 2006). Consequently, all such genes will be lost in the subsequent backcrosses to the uninfected population (Jaenike *et al.* 2006). Thus, if a *Wolbachia*-infected population and an uninfected population exchange migrants, unidirectional and opposite directions of gene flow involving nDNA and mtDNA occur.

Second, the *A. mukaigawae* complex includes an asexual species, *A. pseudoflos*. A study of this species can help us to understand origin of asexuality in oak gall wasps. Among Cynipini, an asexual life cycle also occurs in *Plagiotrochus suberi* Weld, *Andricus quadrilineatus* Hartig, and *Dryocosmus kuriphilus* Yasumatsu (Abe 2007). For these asexual species, the origin of asexuality has not been investigated using molecular genetic data, although population genetic approaches have been applied to infer oviposition behavior for the apparent asexuality in four oak gall wasp species (Atkinson *et al.* 2002). The number of asexual species in Cynipini is very small compared to those in rotifers (Bdelloida), seed shrimps (Ostracoda, Darwinulidae), and mites (some of Oribatida), which are considered as ancient asexual taxa (Welch & Meselson 2000; Schön & Martens 2003; Maraun *et al.* 2004; Schaefer *et al.* 2006). Rarity of obligate parthenogenetic species has been explained by recent origins of asexuality possibly due to a loss of the sexual generation (Stone *et al.* 2002) as in some other groups of cyclically parthenogenetic animals (Hebert 1987). However, recent origin of asexuality in oak gall wasps has not been tested using molecular genetic data. Specifically, we do not know whether the origin of asexuality precedes the host shift or not and whether it has a single origin or not.

In the present study, the pattern of nucleotide variation in a mitochondrial region and nine nuclear regions in the three species belonging to the *A. mukaigawae* complex

were investigated to understand the aforementioned features of the complex. The main objectives of our present study were to answer the following questions:

1 Which change occurred first, host shift or change of reproductive mode? If the loss of sex occurred on *Q. dentata* after host shift as suggested by Abe (2007), genetic divergence between the species on the same host plant species (*A. kashiwaphilus* and *A. pseudoflos*) will be smaller than that between the species on different host plant species (*A. kashiwaphilus* and *A. mukaigawe*, or *A. pseudoflos* and *A. mukaigawae*).
2 Does the asexuality of *A. pseudoflos* have a single origin?
3 Are the directions of gene flow in mtDNA and nDNA between *A. mukaigawae* and *A. kashiwaphilus* as expected from the role of *Wolbachia* infection in speciation?

Materials and methods

Sampling

A total of 51 asexual galls induced by *A. mukaigawae*, *A. kashiwaphilus*, and *A. pseudoflos* were collected from ten sites: three in Hokkaido, five in Honshu, and one in Kyushu, Japan; and one in Hongcheong, Korea (Fig. 2 and Table S1, Supporting information). After collection, the galls were kept in the laboratory until adult gall wasps emerged. Total DNA was extracted from the abdomen of adult wasps that emerged from the galls, using the GenElute Mammalian Genomic DNA kit (SIGMA) following manufacturer's instructions. The DNA was suspended in 250 μl TE buffer.

Fig. 2 Sampling points in Japan and Korea. Filled square: *Andricus mukaigawae*, filled circle: *Andricus kashiwaphilus*, open circle: *Andricus pseudoflos*.

Primer design

For the design of primers in the mitochondrial DNA, the mitochondrial genome of *Apis mellifera* L. (Crozier & Crozier 1993) was used as a reference. The region from *ATPase6* to *COX3* of the mitochondrial genome was amplified using primers A6-N4552RC-M and C3-N-5460 (Table S2, Supporting information). Then, a new primer L1F was designed based on the obtained sequences. PCR was performed using Ex-Taq DNA Polymerase (TaKaRa) in 25 μL of reaction mixture containing 1 μL DNA template. The cycling profile was 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 47 °C for 60 s, and 72 °C for 60 s, then 72 °C for 10 min. A primer NAD4R1 was designed based on the *NAD4* sequence of *Vanhornia eucnemidarum* Crawford (Vanhorniidae, GenBank accession number: NC_008323) (Table S3, Supporting information) by aligning the sequences of the *NAD4* region of the other hymenopteran species (Crozier & Crozier 1993; Castro *et al.* 2006; Oliveira *et al.* 2008; Dowton *et al.* 2009; Wei *et al.* 2010).

Partial regions of putative nuclear genes homologous to *Ant_SesB*, *RpL37*, and *RpS18* were amplified using the primer sets developed for hymenopteran species (Lohse *et al.* 2010, 2011). Six additional primer sets were designed based on published sequences, and subsequently refined using the obtained sequences (Tables S2 and S3, Supporting information; Folmer *et al.* 1994; Simon *et al.* 1994; Cook *et al.* 2002; Rokas *et al.* 2002; Sharanowski *et al.* 2010). As expected, two of the six PCR products (*RpL27A* and *LWRh*) had homology to nuclear gene regions. However, the *COX-like* region had homology to the mitochondrial *COX1* gene. Nevertheless, it was regarded as a nuclear gene because some individuals were heterozygotes and there were stop codons and deletions compared to the mitochondrial *COX1* gene sequences of other Hymenopteran species, including *Andricus curvator* Hartig (GenBank accession numbers: AF395177, DQ012621, FJ154897, and FJ478177). The remaining sequences were regarded as from anonymous nuclear loci (Locus 2, Locus 4, Locus 5; Table S3, Supporting information), because they could not be annotated on the basis of sequence homology. Furthermore, the groups detected by the Structure program (see later) were in Hardy–Weinberg equilibrium (HWE) (data not shown).

PCR amplification and sequencing

Nine nDNA regions and one mDNA region including four genes (partial *COX3*, *NAD3*, *NAD5*, partial *NAD4* and tRNA) were amplified and sequenced using primers shown in Table S3. For the mtDNA region, PCR was performed using LA-Taq DNA Polymerase (TaKa-

Ra) in 30 μL of reaction mixture containing 3 μL of the DNA template. The cycling profile was 94 °C for 1 min followed by 30 cycles of 98 °C for 10 s, 60 °C for 5 min, then a final extension at 72 °C for 10 min.

For the nuclear genes, PCR was performed using Ex-Taq DNA Polymerase (TaKaRa). The cycling profile was at 95 °C for 2 min followed by 35 cycles of 95 °C for 15 s, 40 s at annealing temperature depending on the genes, and 72 °C for 1 min (except for the case of *RpL27A*, for which the extension time per cycle was 5 min), then a final extension at 72 °C for 10 min. The annealing temperature was 55 °C for *COX-like* and *RpL27A*, and 60 °C for *Ant_SesB*, *Locus 2*, *Locus 4*, *Locus 5* and *LWRh*. For *RpL37* and *RpS18*, a touchdown PCR was performed following Lohse *et al.* (2010).

The PCR products were checked by agarose gel electrophoresis before polyethyleneglycol (20% PEG 6000, 2.5 M NaCl) purification. The purified PCR products were used as templates for direct sequencing, from both directions. The DNA sequencing reaction was performed using an ABI BigDye Terminator Cycle version 3.1 sequencing kit (Applied Biosystems) following manufacturer's instructions and analyzed on an ABI PRISM 3100 or ABI 3730 Genetic Analyzer (Applied Biosystems). Whenever necessary, internal sequencing primers were designed at about every 500 bp for both strands (Table S3, Supporting information). When sequences had two or more heterozygous sites, the PCR products were first cloned into the pGEM-T Easy Vector System (Promega) following manufacturer's instructions after purification using the Wizard SV Gel and PCR Clean-Up System (Promega). Then, six or more clones were sequenced to determine individual alleles in each sample. The sequence of each allele was determined by at least two independent clones that had identical sequence.

Data analysis

All sequences were assembled using SEQMAN packaged in Lasergene 7.1 (DNASTAR). The obtained sequences were aligned by CLUSTALX ver. 2 (Larkin *et al.* 2007) and the alignment was manually corrected using SE-AL ver. 2.0a11 (Rambaut 1996). All indels found in the aligned sequences were excluded from the following analyses. In the mtDNA region and in the *RpL27A* nuclear gene, discontinuous sequence data was obtained because repeated sequences interrupted sequencing. For both regions, the obtained partial sequences were aligned and concatenated after removing ambiguous or unalignable sites flanking the repeat sequences, and used in the following analyses.

Standard measures of DNA polymorphism, including number of segregating sites (S), nucleotide diversity (π) (Nei & Li 1979), scaled mutation rate (θ_W) (Watterson 1975), number of haplotypes (h), and haplotype diversity (hd) (Nei 1987) were estimated. To estimate minimum number of recombination events (R_M) (Hudson & Kaplan 1985), the four-gamete test was performed. Neutrality test statistics, D (Tajima 1989) and D^*, F^* (Fu & Li 1993) were also estimated. Tests of means of Tajima's D were conducted *via* 10 000 coalescent simulations. All these analyses were carried out using the following programs: DNASP ver. 5.10.01 (Librado & Rozas 2009), SITES, and HKA (developed by Jody Hey, available at http://genfaculty.rutgers.edu/hey/software). Deviations from HWE were examined using software GENALEX ver. 6.3 (Peakall & Smouse 2006).

To infer the population structure using nDNA data, STRUCTURE software ver. 2.3 (Pritchard *et al.* 2000; Falush *et al.* 2003), which assigns individuals to one of a given number (K) of populations, was used. The method assumes that each population is at HWE and linkage equilibrium. These assumptions may be inappropriate for the individuals of *A. pseudoflos*, because they are known to be parthenogenetic. The program was run with a burn-in of 100 000 iterations followed by a further run of 100 000 iterations. Admixture and correlated allele frequency models for a given K were simulated. The run was repeated 20 times for each K and the log likelihood, ln $P[D]$, of the data was estimated by taking the averages over the replicates. These averages were also used for the calculation of ΔK (Evanno *et al.* 2005) for each K value. The values of ln $P[D]$ and ΔK were obtained by Structure Harvester ver. 0.6.1 (Earl 2012). CLUMPP (Jakobsson & Rosenberg 2007) and DISTRUCT programs (Rosenberg 2004) were used to align and visualize the results of multiple runs.

The hierarchical Bayesian model implemented in *BEAST ver. 1.7 (Drummond *et al.* in press) was used to construct trees for the seven populations that were suggested by STRUCTURE. Two types of data sets: nDNA +mtDNA and nDNA, were analyzed. Sequences of the *LWRh* region were monomorphic and therefore were excluded from this analysis. A strict clock model was assumed and constant size coalescent was chosen as the coalescent prior. The analysis was repeated twice and run each time for 40 000 000 generations (sampling at every 1000 steps) after 10 000 000 burn-in steps. Convergence was assessed by examining the likelihood plots through time using TRACER ver. 1.5 software (Rambaut & Drummond 2007). For mtDNA data neighbor-joining (NJ) trees (Saitou & Nei 1987) were constructed using MEGA 5 beta#4 (Tamura *et al.* 2011) based on Kimura's two-parameter distance (Kimura 1980). The root of the tree was placed using the mid-point rooting. For the mtDNA NJ tree, bootstrapping was performed with 1000 replicates and Bayesian posterior probabilities of each major clade were obtained using the *BEAST analy-

sis. Haplotype networks were constructed using TCS program ver. 1.21 (Clement *et al.* 2000).

A Bayesian method of species delimitation implemented in BPP ver. 2.1 software (Rannala & Yang 2003; Yang & Rannala 2010) was used to conduct multilocus, coalescent population delimitation. The longest non-recombining nDNA regions were used in this analysis (Table S4, Supporting information). To estimate posterior distribution for the seven-population delimitation models, the population tree estimated using *BEAST was used as a guide tree. The combinations of gamma priors (G) for the population size parameters (θ) and the age of the root (τ) were set at $\theta \sim G(2, 1000)$, $\tau \sim G(2, 200)$. The Dirichlet priors (Yang & Rannala 2010: eq. 2) were assigned to the other divergence time parameters. The algorithm 0 was chosen with the fine-tune parameter 5. Each reversible-jump Markov chain Monte Carlo analysis was run for 100 000 generations (sampling every five steps) following a burn-in of 50 000 generations. This analysis was run twice to check consistency between runs.

The isolation with migration analytical model implemented in the program IMA2 (Hey 2010a,b) was used to estimate the following six demographic parameters for a pair of populations: population sizes of the two populations and of the ancestral population, time since divergence, and migration rates in both directions. The longest non-recombining nDNA regions were used in this analysis (Table S4, Supporting information) because the underlining model of IMA2 assumes that no recombination occurs in the analyzed gene segment. Inheritance scalars of 0.75 and 0.25 were assigned to nine nDNA genes and one mtDNA region, respectively, because Hymenoptera is considered to be haplodiploid, i.e., female: $2n$; male: n. The HKY substitution model was chosen for each locus (Hasegawa *et al.* 1985). The analysis was carried out only for the *A. kashiwaphilus* and *A. mukaigawae* pair. *Andricus pseudoflos* was not included in this analysis because of its asexuality. Three types of datasets: nDNA + mtDNA, nDNA, and mtDNA, were separately analyzed. The prior distributions for the parameters were chosen following software manual assuming uniform distributions with minimums at zero. Maximum values for the population size parameter were set at 20, 10, and 1600 for nDNA + mtDNA, nDNA, and mtDNA, respectively, which were chosen after preliminary runs, following the software manual. Maximum migration prior values were set at 2 for the nDNA + mtDNA and nDNA datasets, and 1 for the mtDNA dataset. Maximum times of population splitting were set at 10 for the nDNA + mtDNA and nDNA datasets, and 1000 for the mtDNA dataset. Four independent simulations were conducted to check for convergence (after a burn-in of 1 000 000, 25 000 genealogies were saved at every 100 steps) with 40 chains with a heating scheme option command, -hfg -ha0.975 -hb0.75. A total of 100 000 genealogies of each dataset were used to calculate marginal distributions of the posterior probability. Log-likelihood ratio tests were also conducted using those genealogies to statistically evaluate estimates of the migration rates. If a negative value was obtained for the log-likelihood ratio statistics, it was treated as 0.0001. This is because theoretically the value is positive and distributed approximately as a chi-squared (Hey & Nielsen 2007).

Although the IMA estimation is not sensitive to population structure with regard to discriminating gene flow from ancestral polymorphism (Strasburg & Rieseberg 2010), estimates of effective population sizes and divergence time will be distorted in unpredictable ways (Becquet & Przeworski 2009; Strasburg & Rieseberg 2010). Therefore, the scaled estimates of population sizes or time were not converted to units of real individuals or years, respectively.

Results

DNA polymorphism and test of neutrality

Six individuals in the MN population of *A. pseudoflos* from Hokkaido had a single heterozygous site at one of the mitochondrial loci. Because there was only one such site, the sequences were manually divided into two sequences. The obtained alleles were treated as pseudohaplotypes, and were included in the following analyses. The presence of such pseudohaplotypes can be explained by: the presence of a nuclear pseudogene, heteroplasmy or intramitochondrial duplication (discussion in Data S1, Supporting information).

Estimates of nucleotide diversity in the three species are shown in Table 2. In the mtDNA region (3271 bp), the highest estimate of nucleotide diversity was found in *A. kashiwaphilus*. Nucleotide diversity of *A. pseudoflos* would be inflated because of the presence of pseudohaplotypes. In the nDNA regions (total 11 635 bp from 321 to 1955 bp), the highest nucleotide diversity was found in *A. mukaigawae* (average = 0.00469). The levels of nucleotide diversity of *A. kashiwaphilus* (average = 0.000828) and *A. pseudoflos* (average = 0.000664) were similar. Six of the nine nuclear loci in *A. pseudoflos* were monomorphic, while two loci were monomorphic in *A. mukaigawae* and one locus was monomorphic in *A. kashiwaphilus*.

Tajima's D for the mtDNA region was significantly positive in *A. kashiwaphilus* and *A. pseudoflos*, but not significant in *A. mukaigawae* (Table 2). Tajima's D values for the nDNA genes were not significant except for *RpS18*. Thus, the standard neutral model could not be rejected at these loci. In contrast, Tajima's D was significantly positive for the *RpS18* region in *A. kashiwaphilus* and *A. pseudoflos* (Table 2), suggesting action of

Table 2 Estimates of nucleotide variation

Species	Locus		*n*	*S*	*K*	π	θ_W	*h*	*hd*	R_M	Tajima's *D*	Fu & Li's *D**	Fu & Li's *F**
Andricus	mtDNA		7	12	4.286	0.00131	0.0015	5	0.857	–	−0.681	−0.904	−0.938
mukaigawae	nDNA	*Ant_SesB*	14	5	2.297	0.00134	0.00092	3	0.604	0	1.595	0.505	0.903
		RpL37	14	4	0.956	0.00298	0.00392	4	0.67	0	−0.79	−1.414	−1.426
		RpS18	14	3	1.56	0.00147	0.00089	4	0.714	0	2.002	1.07	1.492
		Locus 2	14	0	0	0	0	1	0	0	–	–	–
		Locus 4	14	22	5.275	0.00365	0.00479	5	0.505	1	−1	−1.169	−1.288
		COXlike	14	2	0.67	0.00042	0.00039	3	0.582	0	0.179	−0.446	−0.324
		Locus 5	14	19	6.538	0.00401	0.00366	10	0.923	2	0.393	0.585	0.611
		RpL27A	14	23	5.527	0.0283	0.0037	8	0.824	1	−0.995	−0.852	−1.022
		LWRh	14	0	0	0	0	1	0	–	–	–	–
		Total/mean		78	2.536	0.00469	0.00203				0.198		
Andricus	mtDNA		22	140	64.069	0.01959	0.01174	11	0.922	–	2.712**	1.689*	2.351*
kashiwaphilus	nDNA	*Ant_SesB*	44	6	0.951	0.00055	0.0008	5	0.634	0	−0.811	1.185	0.658
		RpL37	44	1	0.0089	0.00028	0.00072	2	0.089	0	−0.853	0.555	0.177
		RpS18	44	13	5.834	0.00548	0.00281	4	0.675	0	2.919*	1.015	1.939
		Locus 2	44	4	0.525	0.00038	0.00066	5	0.393	0	−1.001	−1.145	−1.284
		Locus 4	44	2	0.091	0.00006	0.00032	3	0.09	0	−1.478	−2.473	−2.530
		COXlike	44	5	0.542	0.00034	0.00072	5	0.419	0	−1.316	−0.776	−1.101
		Locus 5	44	3	0.266	0.00016	0.00042	3	0.173	0	−1.307	0.905	0.285
		RpL27A	44	3	0.416	0.00021	0.00035	4	0.388	0	−0.843	−0.378	−0.601
		LWRh	44	0	0	0	0	1	0	–	–	–	–
		Total/mean		37	0.959	0.000829	0.000756				−0.586		
Andricus	mtDNA		28	37	15.376	0.0047	0.00291	6	0.802	–	2.290*	1.758*	2.271*
pseudoflos	nDNA	*Ant_SesB*	44	2	0.584	0.00034	0.00027	3	0.545	0	0.496	0.761	0.792
		RpL37	44	0	0	0	0	1	0	–	–	–	–
		RpS18	44	12	5.666	0.00533	0.00259	3	0.664	0	3.189**	1.476	2.402
		Locus 2	44	1	0.089	0.00006	0.00016	2	0.089	0	−0.853	0.555	0.177
		Locus 4	44	0	0	0	0	1	0	–	–	–	–
		COXlike	44	0	0	0	0	1	0	–	–	–	–
		Locus 5	44	0	0	0	0	1	0	–	–	–	–
		RpL27A	44	1	0.495	0.00025	0.00012	2	0.495	0	1.601	0.555	0.986
		LWRh	44	0	0	0	0	1	0	–	–	–	–
		Total/mean		16	0.759	0.000664	0.000349				1.108*		

S, total number of segregating sites; π, average number of pairwise nucleotide differences per site; θ_W, average Watterson's parameter, *h*, number of haplotypes; *hd*, haplotype diversity.

*$0.01 < P < 0.05$, **$0.001 < P < 0.01$, ***$P < 0.001$ (for nDNA regions, the star indicate *P* value after Bonferroni correction).

balancing selection, gene duplication, or the presence of population structure (discussion in Data S2 and Figs S1 and S2, Supporting information). Similar results were obtained for Fu and Li's D^* and F^*. Fay and Wu's H (Fay & Wu 2000) was also calculated for the nDNA genes of *A. kashiwaphilus* and *A. pseudoflos* with *A. mukaigawae* as an outgroup. In both species, no locus showed significant result and there was no clear trend in the values (Table S5, Supporting information).

Phylogenetic relationship

To determine the order of events (host shift and change to asexuality) and origins of asexuality, we reconstructed the history of the species based on sequences of mtDNA and nDNA. The NJ tree of the mitochondrial region is shown in Fig. 3. It contained two groups: one group consisted of IS and MN populations of *A. kashiwaphilus* from Hokkaido and *A. pseudoflos*, and the other group consisted of KJ and KR populations of *A. kashiwaphilus* from Kyushu and Korea and *A. mukaigawae*. Therefore, *A. kashiwaphilus* is paraphyletic for the mitochondrial DNA.

For the nuclear genes, first all data sets were analyzed (*A. mukaigawae*, *A. kashiwaphilus*, and *A. pseudoflos*) using the STRUCTURE program. The posterior probability for K increased until $K = 4$, then plateaued between $K = 4$ and 12 and started to decrease afterwards, while Evanno's ΔK showed a peak at $K = 2$ (Fig. 4, see also Fig. S1, Supporting information). The assignment of individuals when $K = 4$ suggested that *A. kashiwaphilus* and *A. pseudoflos* could not be separated. The MN and IS populations of *A. kashiwaphilus* from Hokkaido, and the MN, HR, and SK populations of *A. pseudoflos* from Hokkaido and Honshu formed one group, while the KJ population of *A. kashiwaphilus* from Kyushu and the TN population of *A. pseudoflos* from

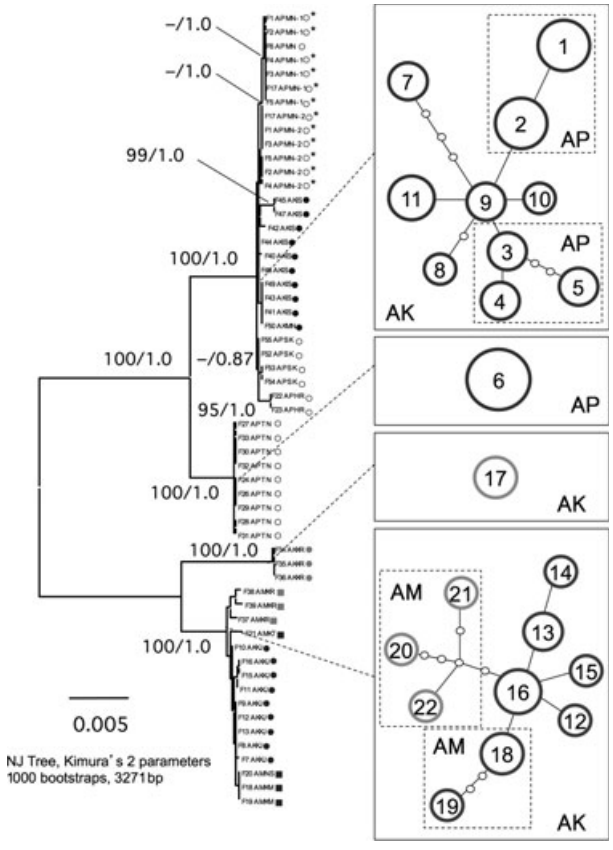

Fig. 3 Neighbor-joining tree based on Kimura's two parameters distance calculated from the mitochondrial region (3271 bp). Numbers associated with branches indicate (bootstrap values with 1000 replicates)/(posterior probabilities in Bayesian analyses). Filled square: *Andricus mukaigawae*, filled circle: *Andricus kashiwaphilus*, open circle: *Andricus pseudoflos*, asterisk: pseudohaplotype. Figure on the right shows the maximum-parsimonious networks of haplotypes. Haplotypes are indicated by numbers (1–22). AM: *A. mukaigawae*, AK: *A. kashiwaphilus*, and AP: *A. pseudoflos*.

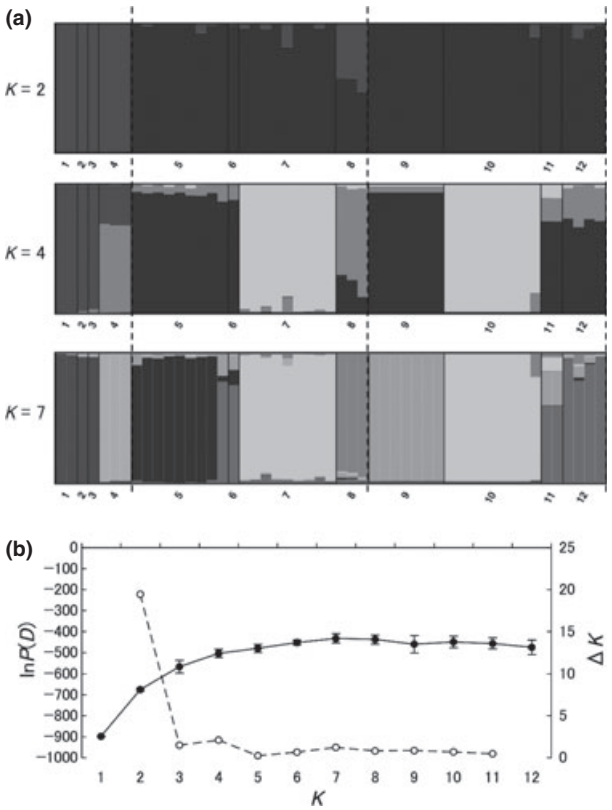

Fig. 4 (a) Genetic relationships among the 12 sampled populations using Structure based on nDNA data. Admixture and correlated allele frequency models for given K were simulated. 1–4: *Andricus mukaigawae* (1: KM, 2: KT, 3: NS, and 4: KR), 5–8: *Andricus kashiwaphilus* (5: IS, 6: MN, 7: KJ, and 8: KR), and 9–12: *Andricus pseudoflos* (9: MN, 10: TN, 11: HR, and 12: SK). (b) Plots of mean ln $P(D)$ with standard deviation (filled circle) and ΔK (open circle) based on 20 replicates for each values of K.

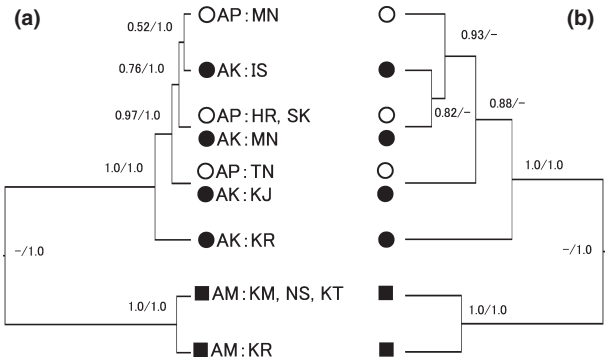

Fig. 5 Population trees estimated by *BEAST for the seven populations identified by the Structure analysis, numbers associated with branches indicate (posterior probabilities obtained from the *BEAST analyses)/(speciation probabilities obtained from the BPP analyses). AM: *Andricus mukaigawae* (filled square), AK: *Andricus kahiwaphilus* (filled circle), and AP: *Andricus pseudoflos* (open circle). (a) using nDNA data, (b) using nDNA + mtDNA data.

Hokkaido formed another group. The highest posterior probability was observed when $K = 7$. When $K = 7$, two individuals from the MN and IS populations of *A. kashiwaphilus* were assigned to the group consisting of the HR, and SK populations of *A. pseudoflos*, and all individuals from the KJ population of *A. kashiwaphilus* were assigned to the other group consisting of the TN population of *A. pseudoflos*. With $K = 7$, the three groups found within *A. pseudoflos* were the same as those in the mitochondrial genealogy. Analysis using the dataset without *RpS18* region, which showed significant positive Tajima's *D*, still showed assignment of individuals similar to that with the full dataset (data not shown).

Based on the result of the STRUCTURE analysis when $K = 7$, one individual of the MN population of *A. kashiwaphilus* and all individuals of the HR and SK populations of *A. pseudoflos* were assigned to the same group. Furthermore, individuals of the KJ population of *A. kashiwaphilus* and individuals of the TN populations of *A. pseudoflos* were assigned to a separate group. The population trees estimated for the seven populations in *BEAST analysis are shown in Fig. 5. In contrast to the mitochondrial genealogy (Fig. 3), the population trees suggest monophyly of host plant association (*A. kashiwaphilus* and *A. pseudoflos* formed a monophyletic group separated from *A. mukaigawae*). In obligate parthenogenetic *A. pseudoflos*, there is strong support for three groups (MN, TN, and HR + SK populations) in the trees obtained using the nDNA + mtDNA dataset (Fig. 5b), while there is moderate support for those groups in the trees based on the nDNA dataset (Fig. 5a). Analysis using BPP supported the seven-population model with high posterior probabilities (Fig. 5a).

Estimates of population parameters

To determine the direction of gene flow between *A. mukaigawae* and *A. kashiwaphilus*, we applied IMA2 to the sequence data. The maximum likelihood estimates (MLE) for the migration rate parameters between *A. kashiwaphilus* and *A. mukaigawae* were significantly positive in both directions when the mtDNA + nDNA dataset was analyzed (Fig. 6 and Table S6, Supporting information). With the mtDNA dataset, only the MLE for the forward migration rate from *A. mukaigawae* to *A. kashiwaphilus* was positive and the 95% credibility interval did not include zero. On the other hand, with the nDNA dataset, only the MLE for the forward migration rate from *A. kashiwaphilus* to *A. mukaigawae* was positive and the 95% credibility interval did not include zero (Fig. 6 and Table S6, Supporting information). Exclusion of the data of *RpS18*, which showed significantly positive Tajima's *D* values, did not change the aforementioned result (data not shown). For the migration rates that showed positive estimates, the likelihood ratio tests rejected zero-migration-rate model (Table 3). Estimates for the size of ancestral population, and the

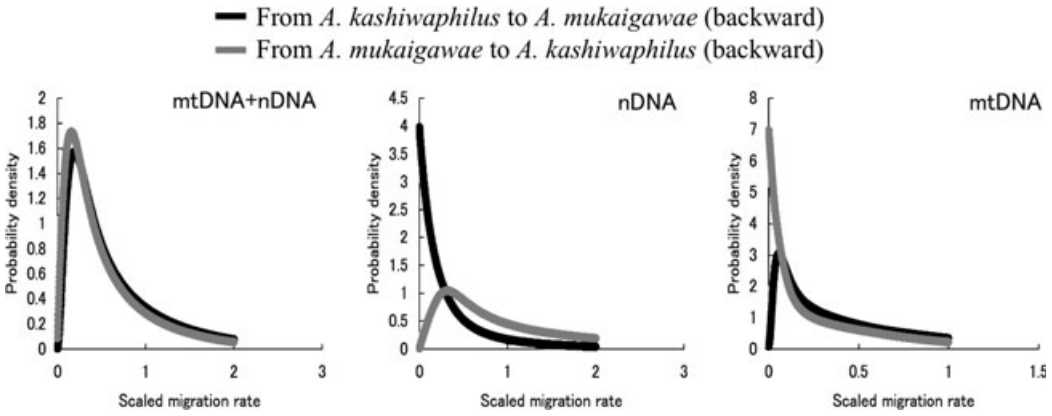

Fig. 6 Marginal posterior distributions for scaled migration rate (backward in time) between *Andricus kashiwaphilus* and *Andricus mukaigawae*.

Table 3 Log-likelihood ratio statistics for the zero migration model, the equal migration model and the equal population size model

Datasets	Model		2LLR	*df*	*P*
mtDNA + nDNA	Full model	θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0}$	–	–	
	Nested models	θ_0, θ_1, θ_2, $m_{0>1} = 0$, $m_{1>0}$	35.54***	1	2.50E−09
		θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0} = 0$	5.4*	1	0.0201
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0}$	3.563	1	0.0591
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0} = 0$	39.49***	2	2.66E−09
nDNA	Full model	θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0}$	–	–	
	Nested models	θ_0, θ_1, θ_2, $m_{0>1} = 0$, $m_{1>0}$	0.0001 (−0.9031)	1	0.992
		θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0} = 0$	7.789**	1	0.00526
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0}$	0.387	1	0.534
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0} = 0$	9.108*	2	0.0105
mtDNA	Full model	θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0}$	–	–	
	Nested models	θ_0, θ_1, θ_2, $m_{0>1} = 0$, $m_{1>0}$	13.37***	1	0.000256
		θ_0, θ_1, θ_2, $m_{0>1}$, $m_{1>0} = 0$	0	1	1
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0}$	1.981	1	0.159
		θ_0, θ_1, θ_2, $m_{0>1} = m_{1>0} = 0$	15.49***	2	0.000433

If nested model is true, 2LLR asymptotically will follow a chi-squared distoribution with degrees of freedom.
Species 0: *Andricus kashiwaphilus* and species 1: *Andricus mukaigawae*.
θ_n: combined parameter $4N_eu$ of population 0, population 1, and population 2 (=ancestral population), where u is the mutation rate per site per generation.
$m_{x>y}$: combined parameter $M_{x>y}/u$, the migration rate per gene per generation from population x to population y in the coalescent (backward in time). In time moving forward, it means migration from population y to population x.
* $0.01 < P < 0.05$, ** $0.001 < P < 0.01$, *** $P < 0.001$.

divergence time did not seem to be reliable, because the peaks of their posterior density distributions were not distinct, and the curves of the distributions were not declining at the upper limit (data not shown).

Discussion

The order of diversification in the Andricus mukaigawae complex

Genetic divergence between the species on the same host plant species (*A. kashiwaphilus* and *A. pseudoflos*) was smaller than that between the species on different host plant species (*A. kashiwaphilus* and *A. mukaigawae*, or *A. pseudoflos* and *A. mukaigawae*) at the nDNA genes (Fig. 5). This result suggests that the loss of sex occurred after host shift to *Q. dentata* (host plant for *A. kashiwaphilus* and *A. pseudoflos*). Although *A. kashiwaphilus* is paraphyletic for the mtDNA (Fig. 3), the occurrence of the loss of sex on *Q. dentata* is still supported. Therefore, host shift occurred first and then loss of sexuality followed in this species complex.

The origin of asexuality in the Andricus mukaigawae complex

The asexual species, *A. pseudoflos* seemed to consist of two or three distinct groups, as suggested by the mtDNA and nDNA data. If these groups originated independently from the cyclically parthenogenetic species (*A. kashiwaphilus*), asexuality has evolved at least twice in this species complex. Polyphyly of parthenogenesis in a species was also suggested in species of Homoptera, Coleoptera, Lepidoptera, Hymenoptera, Orthoptera and Phasmatoptera (Simon *et al.* 2003 and references therein), although types of parthenogenesis and modes of origin differ among those species.

Direction of gene flow and speciation scenario for the Andricus mukaigawae complex

Unidirectional but opposite directions of gene flow in nDNA and mtDNA between *A. mukaigawae* and *A. kashiwaphilus* were detected in this study. If there is a third unknown population or the ancestral population was structured, estimates of the presence and the timing of gene flow may not be reliable (Becquet & Przeworski 2009; Strasburg & Rieseberg 2010). However, if these estimates were reliable, then they could be explained by a scenario of diversification involving *Wolbachia* infection (Introduction and Fig. 1) with two assumptions: (i) the phenotypic effect of *Wolbachia* infection in *A. mukaigawae* is CI and, (ii) there was a loss of *Wolbachia* infection in *A. kashiwaphilus*. Under this scenario, individuals on *Q. dentata* (host plant of *A. kashiwaphilus*) are *Wolbachia*-free and those on other species of *Quercus* (host plant of *A. mukaigawae*) are infected.

The observation that *Wolbachia*-free *A. kashiwaphilus* has the mtDNA type associated with infected *A. mukaigawae* is apparently inconsistent with this scenario because females inherit both mtDNA and *Wolbachia* from their mothers. Nevertheless, the observation that not all individuals of *A. mukaigawae* were *Wolbachia*-infected (only 15 of 17 individuals were found to be infected; Abe & Miura 2002) suggested that *Wolbachia* transmission was imperfect. If the transmission is imperfect, some of the offspring of infected females will not be infected but have the mtDNA type coming from the infected population. The migration of, or the production of such uninfected individuals can explain the observation of uninfected individuals of *A. kashiwaphilus* that have the mtDNA type associated with infected *A. mukaigawae*. Theoretically, when the prevalence of infection is low, loss of infection can result from imperfect transmission, with the lack of reproductive advantage of infected females due to their compatiblity with infected males (Turelli 1994). Empirically, in mushroom-feeding *Drosophila*, the occurrence of uninfected *Drosophila subquinaria* individuals carrying mtDNA haplotypes of infected *Drosophila recens* has been suggested to be due to loss of infection (Jaenike *et al.* 2006).

Alternatively, unidirectional gene flow can be explained by the female host preference and the assumption of dominance of larval performance as predicted in *Juglans* and *Lyonia* races of the leaf-mining moth, *Acrocercops transecta* (Ohshima & Yoshizawa 2010). This prediction, however, can only explain unidirectional gene flow in nDNA from the recessive type to the dominant type (Ohshima & Yoshizawa 2010). Because female host preference would restrict introgression of maternally inherited genes (mtDNA) between host races (Ohshima & Yoshizawa 2010), no gene flow in mtDNA will be expected. Furthermore, even if there is dominance of larval performance, the gene flow in nDNA would appear to be bidirectional, possibly because of backcross matings (Ohshima & Yoshizawa 2010). Therefore, unidirectional and opposite direction of gene flow in nDNA and mtDNA seems to be difficult to explain with only female host preference and dominance of larval performance.

Conclusion

In conclusion, the genetic relationship among the cyclically parthenogenetic species and the obligate parthenogenetic species suggested that the loss of sex occurred after the host shift in this species complex. This study also revealed unidirectional and opposite directions of gene flow for nDNA and mtDNA between the two cyclically parthenogenetic species. *Wolbachia* infection in one of the two populations ancestral to the two species can explain this result if *Wolbachia* infection causes CI. If this is the case, *Wolbachia* infection may have enhanced reproductive isolation of the two incipient species and promoted host shift to a closely related tree species, which has rarely occurred in Cynipini. Similarly, some studies on the fig-pollinating wasp suggested that *Wolbachia* has a role on fig wasp speciation, because cryptic pollinator species on the same host plant species have different *Wolbachia* infections (Haine *et al.* 2006; Sun *et al.* 2011). Furthermore, unidirectional CI might have affected the host specialization (female preference and larval performance) of *Wolbachia*-free *A. kashiwaphilus*. The theoretical (Telschow *et al.* 2007) and empirical studies (Jaenike *et al.* 2006) suggested that *Wolbachia* infection in one of the host populations can maintain divergence of the hosts of *Wolbachia* by selection for premating isolation (reinforcement). So *Wolbachia* infection might have an important role not only in the direction of gene flow but also in the host specialization.

Acknowledgements

We would thank Dr. Kosuke M. Teshima (Kyushu University) for suggestions, discussions and for allowing us to use his equipment. We deeply appreciate Professor Jong Kuk Kim (Kangwon National University) for preparing the specimens of *A. kashiwaphilus* and *A. mukaigawae* from Korea and Professor Graham N. Stone, Dr. James A. Nicholls, and Dr. Konrad R. Lohse (University of Edinburgh) for providing information about the primers. This study was partly supported by the Kyushu University Foundation, Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (19380037, 22370083 and 21405011), and Global COE Program (Center of Excellence for Asian Conservation Ecology as a Basis of Human-Nature Mutualism), Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Abe Y (1986) Taxonomic status of the *Andricus mukaigawae* complex and its speciation with geographic parthenogenesis (Hymenoptera: Cynipidae). *Applied Entomology and Zoology*, **21**, 436–447.

Abe Y (1988) Two host races in *Andricus mukaigawae* (Mukaigawa) (Hymenoptera: Cynipidae). *Applied Entomology and Zoology*, **23**, 381–387.

Abe Y (1991) Host race formation in the gall wasp *Andricus mukaigawae*. *Entomologia Experimentalis et Applicata*, **58**, 15–20.

Abe Y (1998) Karyotype differences and speciation in the gall wasp *Andricus mukaigawae* (s. lat.) (Hymenoptera: Cynipidae), with description of the new species *A. kashiwaphilus*. *Entomologica scandinavica*, **29**, 131–135.

Abe Y (2007) Parallelism in secondary loss of sex from a heterogonic life cycle on different host plants in the *Andricus mukaigawae* complex (Hymenoptera: Cynipidae), with taxonomic notes. *Journal of Natural History*, **41**, 473–480.

Abe Y, Miura K (2002) Does *Wolbachia* induce unisexuality in oak gall wasps (Hymenoptera: Cynipidae)? *Annals of the Entomological Society of America*, **95**, 583–586.

Atkinson RJ, McVean GAT, Stone GN (2002) Use of population genetic data to infer oviposition behaviour: species-specific patterns in four oak gallwasps (Hymenoptera: Cynipidae). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **269**, 383–390.

Becquet C, Przeworski M (2009) Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–2562.

Bordenstein SR (2003) Symbiosis and the origin of species. In: Insect Symbiosis (eds Bourtzis K, Miller T), pp. 283–304. CRC Press, New York.

Castro LR, Ruberu K, Dowton M (2006) Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera). *Genome*, **49**, 752–766.

Clement M, Posada D, Crandall K (2000) tcs: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Cook JM, Rokas A, Pagel M, Stone GN (2002) Evolutionary shifts between host oak sections and host-plant organs in *Andricus* gallwasps. *Evolution*, **56**, 1821–1830.

Crozier RH, Crozier YC (1993) The mitochondrial genome of the Honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133**, 97–117.

Dowton M, Cameron SL, Dowavic JI, Austin AD, Whiting MF (2009) Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, **26**, 1607–1617.

Drummond AJ, Suchard MA, Xie D, Rambaut A (in press) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*.

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

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fay JC, Wu CI (2000) Hitchhiking positive Darwinian selection. *Genetics*, **155**, 1405–1413.

Flor M, Hammerstein P, Telschow A (2007) Dynamics and stability of *Wolbachia*-induced unidirectional cytoplasmic incompatibility in parapatric host populations. *Journal of Evolutionary Biology*, **20**, 696–706.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Haine E, Martin J, Cook J (2006) Deep mtDNA divergences indicate cryptic species in a fig-pollinating wasp. *BMC Evolutionary Biology*, **6**, 83.

Hasegawa M, Kishino H, Yano T (1985) Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hebert PDN (1987) Genotypic characteristics of cyclic parthenogens and their obligately asexual derivatives. In: The Evolution of Sex and its Consequences (ed. Stearns SC), pp. 175–195. Birkhauser, Basel.

Hey J (2010a) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J (2010b) The divergence of Chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution*, **27**, 921–933.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Jaenike J, Dyer KA, Cornish C, Minhas MS (2006) Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biology*, **4**, e325.

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

Kimura M (1980) A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

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

Liljeblad J, Ronquist F (1998) A phylogenetic analysis of higher-level gall wasp relationships. *Systematic Entomology*, **23**, 229–252.

Lohse K, Sharanowski B, Stone GN (2010) Quantifying the Pleistocene history of the oak gall parasitoid *Cecidostiba fungosa* using twenty intron loci. *Evolution*, **64**, 2664–2681.

Lohse K, Sharanowski B, Blaxter M, Nicholls JA, Stone GN (2011) Developing EPIC markers for chalcidoid Hymenoptera from EST and genomic data. *Molecular Ecology Resources*, **11**, 521–529.

Maraun M, Heethoff M, Schneider K *et al.* (2004) Molecular phylogeny of oribatid mites (Oribatida, Acari): evidence for multiple radiations of parthenogenetic lineages. *Experimental and Applied Acarology*, **33**, 183–201.

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

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.

Ohshima I, Yoshizawa K (2010) Differential introgression causes genealogical discordance in host races of *Acrocercops transecta* (Insecta: Lepidoptera). *Molecular Ecology*, **19**, 2106–2119.

Oliveira DC, Raychoudhury R, Lavrov DV, Werren JH (2008) Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, **25**, 2167–2180.

Peakall R, Smouse PE (2006) GenAlEx6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

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

Rambaut A (1996) SE-AL: Sequence Alignment Editor. http://tree.bio.ed.ac.uk/software/seal/ (last accessed 5 April 2009).

Rambaut A, Drummond AJ (2007) Tracer v1.4. http://beast.bio.ed.ac.uk/Tracer/ (last accessed 6 May 2012).

Rannala B, Yang Z (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, **164**, 1645–1656.

Rokas A, Nylander JAA, Ronquist F, Stone GN (2002) A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. *Molecular Phylogenetics and Evolution*, **22**, 206–219.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Schaefer I, Domes K, Heethoff M *et al.* (2006) No evidence for the 'Meselson effect' in parthenogenetic oribatid mites (Oribatida, Acari). *Journal of Evolutionary Biology*, **19**, 184–193.

Schön I, Martens K (2003) No slave to sex. *Proceedings of the Royal Society of London Series B*, **270**, 827–833.

Sharanowski BJ, Robbertse B, Walker J *et al.* (2010) Expressed sequence tags reveal Proctotrupomorpha (minus Chalcidoidea) as sister to Aculeata (Hymenoptera: Insecta). *Molecular Phylogenetics and Evolution*, **57**, 101–112.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Simon JC, Delmotte F, Rispe C, Crease T (2003) Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, **79**, 151–163.

Stone GN, Schönrogge K, Atkinson RJ, Bellido D, Pujade-Villar J (2002) The population biology of oak gall wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology*, **47**, 633–668.

Stouthamer R, Breeuwer JAJ, Hurst GDD (1999) *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology*, **53**, 71–102.

Strasburg JL, Rieseberg LH (2010) How robust are 'Isolation with Migration' analyses to violations of the IM model? A simulation study *Molecular Biology and Evolution*, **27**, 297–310.

Sun X-J, Xiao J-H, Cook JM, Feng G, Huang D-W (2011) Comparisons of host mitochondrial, nuclear and endosymbiont bacterial genes reveal cryptic fig wasp species and the effects of *Wolbachia* on host mtDNA evolution and diversity. *BMC Evolutionary Biology*, **11**, 86.

Suomalainen E, Saura A, Lokki J (1987) Cytology and Evolution in Parthenogenesis. CRC Press, Boca Raton, Florida.

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

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Telschow A, Flor M, Kobayashi Y, Hammerstein P, Werren JH (2007) *Wolbachia*-induced unidirectional cytoplasmic incompatibility and speciation: mainland-island model. *PLoS One*, **2**, e701.

Turelli M (1994) Evolution of incompatibility-inducing microbes and their hosts. *Evolution*, **48**, 1500–1513.

Vavre F, Fleury F, Varaldi J, Fouillet P, Boulétreau M (2000) Evidence for female mortality in *Wolbachia*-mediated cytoplasmic incompatibility in haplodiploid insects: epidemiologic and evolutionary consequences. *Evolution*, **54**, 191–200.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Wei SJ, Tang P, Zheng LH, Shi M, Chen XX (2010) The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A+T content and a long intergenic spacer between atp8 and atp6. *Molecular Biology Reports*, **37**, 1931–1942.

Welch DBM, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211–1215.

Werren JH (1997) Biology of *Wolbachia*. *Annual Review in Entomology*, **42**, 587–609.

Werren JH (1998) *Wolbachia* and speciation. In: Endless Forms: Species and Speciation (eds Howard D, Berlocher S), pp. 245–260. Oxford University Press, Oxford.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**, 9264–9269.

This study is part of N.W.'s doctoral research. He is interested in the mechanisms and processes of genetic differentiation and speciation in insects. Y.A. is interested in systematics and ecology of insects, especially Cynipoidea (Hymenoptera) and Agromyzidae (Diptera). N.I. is interested in understanding the maintenance mechanisms of naturally occurring genetic variation within and between species. He is working on Drosophila and tropical forest trees. A.E.S. is currently studying population genetics of endangered tropical trees. H.T. is interested in population genetics and molecular evolution.

Data accessibility

DNA sequences: Genbank accessions AB731763–AB731898.

Collection information, including individual sampling locations and accession numbers: Table S1 (Supporting information).

Alignment data and IMA2/BPP input files are available through the Dryad repository: doi: 10.5061/dryad.k41qk.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Figure 4 in colour.

Fig. S2 Means of Tajima's *D* calculated from 1000 simulations based on the parameters of *Andricus kashiwaphilus* in Table S6 (nuc).

Fig. S3 Means of Fay and Wu's *H* calculated from 1000 simulations based on the same parameters as Fig. S1 (Supporting information).

Table S1 Collection information for the samples used.

Table S2 List of PCR primers for primer developments.

Table S3 List of PCR primers and sequencing primers used in this study.

Table S4 Length of regions and estimation of recombination of pooled sequences.

Table S5 Test of neutrality, Fay and Wu's *H*, using the sequence of *A. mukaigawae* as an outgroup.

Table S6 Maximum-likelihood estimates and 95% highest posterior densities of demographic parameters.

Data S1 Multiple mitochondria-like sequences in *Andricus pseudoflos*.

Data S2 Significantly positive Tajima's *D* at *mtDNA* and *RpS18*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.