MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2463–2476 doi: 10.1111/mec.13170

Genetic divergence between the sympatric queen morphs of the ant *Myrmica rubra*

JENNI LEPPÄNEN,* PERTTU SEPPÄ,† KARI VEPSÄLÄINEN* and RIITTA SAVOLAINEN*

**Department of Biosciences, University of Helsinki, P.O. Box 65, Helsinki 00014, Finland, †Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, P.O. Box 65, Helsinki 00014, Finland*

Abstract

Pairs of obligate social parasites and their hosts, where some of the parasites have recently diverged from their host through intraspecific social parasitism, provide intriguing systems for studying the modes and processes of speciation. Such speciation, probably in sympatry, has also been propounded in the ant *Myrmica rubra* and its intraspecific social parasite. In this species, parasitism is associated with queen size dimorphism, and the small microgyne has become a social parasite of the large macrogyne. Here, we investigated the genetic divergence of the host and the parasite queen morphs in 11 localities in southern Finland, using nuclear and mitochondrial markers of queens and workers. We formulated and tested four speciation-related hypotheses that differed in the degree of genetic divergence between the morphs. The queen morphs were genetically distinct from each other with little hybridization. In the nuclear data, when localities were nested within queen morphs in the hierarchical AMOVA, 39% of the genetic variation was explained by the queen morph (standardized $F'_{CT} = 0.63$, uncorrected $F_{CT} = 0.39$), whereas 18% was explained by the locality ($F'_{SC} = 0.39$, $F_{SC} = 0.29$). This result corroborated the *hypothesis of advanced sympatric speciation*. In contrast, the mitochondrial DNA could not settle between the hierarchical levels of locality and queen morph, thus substantiating equally the *hypotheses of incipient* and *advanced sympatric speciation*. Together, our results support the view that the microgynous parasite has genetically diverged from its macrogynous host to the level of a nascent species.

Keywords: hymenoptera, inquilinism, *Myrmica microrubra*, size dimorphism, social parasitism, sympatric speciation

Received 29 July 2014; revision received 18 March 2015; accepted 19 March 2015

Introduction

One of the most debated evolutionary questions during the last decades has focussed on the modes of speciation, particularly on speciation in sympatry (Fitzpatrick *et al.* 2009). Nowadays, it is widely accepted that sympatric speciation occurs (Via 2001), but most examples have been criticized as they rarely fulfil all the criteria for sympatric speciation (reviewed by Fitzpatrick *et al.* 2009). The prevalence of sympatric speciation is unclear but probably rare (Fitzpatrick *et al.* 2009).

Correspondence: Jenni Leppänen, Fax: +358 294157692; E-mail: jenni.leppanen@helsinki.fi

© 2015 John Wiley & Sons Ltd

Promising candidates of sympatric speciation are found in some obligate social parasites of ants (Wasmann 1909; Elmes 1978; Buschinger 1990; Bourke & Franks 1991; Mayr 1993). According to Emery's rule, social parasites are closely related to their host (Le Masne 1956), which has been shown in many obligate social parasites (Huang & Dornhaus 2008; Buschinger 2009; and references therein; Jansen *et al.* 2010). This has prompted the hypothesis that some social parasites have evolved from their hosts sympatrically through intraspecific social parasitism (Elmes 1978; Buschinger 1990; Bourke & Franks 1991). Speciation in sympatry is often, however, difficult to distinguish from the alternative scenario, speciation in allopatry followed by sympatry (Smith *et al.* 2007). Therefore, it would be ideal to

study speciation when it is still an ongoing process and allopatry can be ruled out.

The ant genus *Myrmica* has many social parasites (Radchenko & Elmes 2003) that are closely related to their hosts (Savolainen & Vepsäläinen 2003; Jansen *et al.* 2010). One of these host–parasite pairs is *M. rubra* and its queen morphs, which are easily distinguished from each other by size (Seifert 1993). The microgynes usually produce no or only a few workers (Elmes 1976); such social parasites are called inquilines (Wilson 1971). Because inquilines lack a functional worker caste, they are totally dependent on their macrogynous host. Consequently, in spite of a vast body of work on *M. rubra*'s microgyne, no documented observation on its reproduction is known outside colonies of the host. In parasitized nests, which often form large polydomous colonies (supercolonies; Seppä & Pamilo 1995; Walin *et al.* 2001), breeding queens of both morphs typically are present in large numbers (Elmes 1976; Seppä & Walin 1996). When the microgynous parasite is present in a nest, it produces an excessive amount of sexual offspring (Pearson 1981) and can suppress the growth of sexual offspring of the macrogynes, which supply the workers (Elmes 1976; Pearson 1981). Because each queen morph produces only offspring of its own kind, queen size is presumably genetically regulated (Elmes 1976; Seifert 2010).

The taxonomic status of the microgynous morph has been debated and alternated over the past decades. Seifert (1993) described the microgynous morph as a new species, *M. microrubra*, but later it was synonymized with *M. rubra* (Steiner *et al.* 2006). Nevertheless, the documented genetic differentiation between the queen morphs (Pearson & Child 1980; Pearson 1981; Steiner *et al.* 2006; Vepsäläinen *et al.* 2009; Leppänen *et al.* 2011; Schär & Nash 2014) may owe to ongoing speciation (Savolainen & Vepsäläinen 2003; Seifert 2007, 2010; Vepsäläinen *et al.* 2009). To evaluate the postulated speciation, however, more data on the genetic differentiation between the two morphs are needed.

Here, we studied the genetic composition and spatial structure of *M. rubra* in multiple localities where the microgynous and macrogynous morphs (the parasite and the host) co-occur. By investigating the genetic divergence between the locally co-occurring (syntopic, living side by side) morphs and among the localities of each morph, we tested the following four speciation-related hypotheses that are based on the hierarchical distribution of genetic variation at locality and morph levels. (i) According to the *single-species hypothesis*, the two morphs have not diverged from each other, and the morph level explains poorly the genetic variation in the system. (ii) The *hypothesis of incipient sympatric speciation* states that the two morphs have started to diverge from each other. Both locality and morph levels explain significant amounts of the genetic variation in the system, but the locality level explains a larger proportion than the morph level. (iii) The *hypothesis of advanced sympatric speciation* claims that the two morphs have diverged from each other clearly. Both locality and morph levels explain significant amounts of the genetic variation in the system, but the morph level explains a larger proportion than the locality level. (iv) According to the *two-species hypothesis*, the two morphs have fully diverged from each other and can be considered as separate species. In this case, the locality level explains only a minor part of the genetic variation in the system, depending on the spatial structure and gene flow in the two species in question. We argue that in the second and third hypotheses, speciation is sympatric, as the microgynous morph is dependent on its macrogynous host, and the locality explains a significant amount of genetic variation. In the fourth hypothesis, the speciation mode (allopatry vs. sympatry) is, however, less clear as speciation is already complete.

Materials and methods

Sampling

We collected 80 nest samples of *Myrmica rubra* from 11 localities in southern Finland (Fig. 1, Table 1). We selected nests at least three metres apart from each other and sampled workers and all micro- and macrogynous queens that we found and preserved them in absolute ethyl alcohol. The queens were separated by visual inspection into either queen morph; no intermediate individuals were found. All macrogynes were wingless, presumably established queens, except for one winged gyne from each of three nests. The microgynes included both wingless, presumably old queens, and winged gynes.

For nuclear DNA genetic analyses, we generally used eight macrogynes and eight microgynes from each nest. As eight macrogynes were not available in all nests, we supplemented the data by analysing up to eight workers from these nests. Altogether, our data consisted of 511 microgynes, 136 macrogynes and 500 workers (Table 1).

We used 40 earlier published COI and cyt b mitochondrial DNA sequences (Savolainen & Vepsäläinen 2003; Vepsäläinen *et al.* 2009; Leppänen *et al.* 2011) from eight localities where we also had nuclear data (Table S1, Supporting information). The mitochondrial sequences included 18 microgynes and four macrogynes, but as the Bayesian clustering of nuclear data verified that most workers were produced by the macrogynes, 18 workers were used to represent macrogynes when queens were not available.

© 2015 John Wiley & Sons Ltd

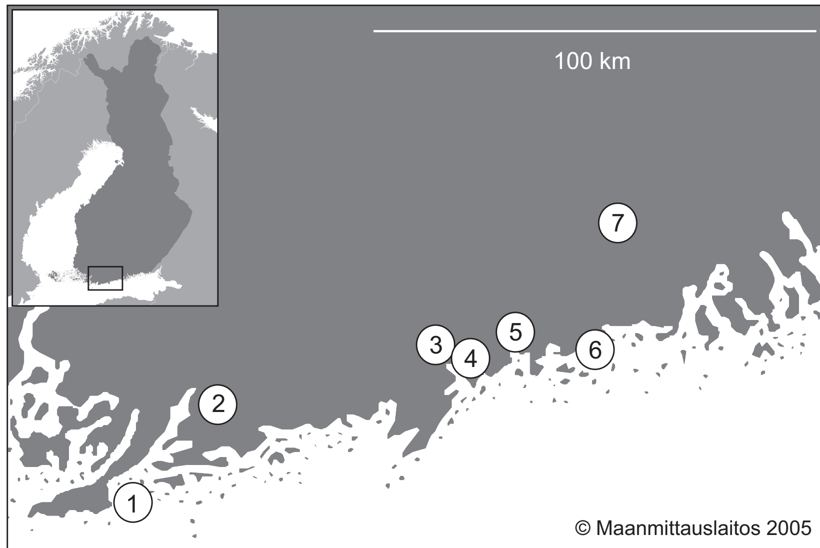

Fig. 1 Sampling areas in southern Finland; the area numbers refer to (1) Tvärminne, localities TK and TN; (2) Karjaa, KA; (3) Vermo, VE; (4) Pikku-Huopalahti, PH; (5) Viikki, VF, VL, VM and VP; (6) Uutela, UU; and (7) Sipoo, SI. For locality information, see Table 1.

Table 1 Locality information. Coordinates are in decimal degrees (N = northern latitude, E = eastern longitude), Nest = number of nests sampled, and Individuals = number of individuals analysed of each morph and caste (m = microgyne, M = macrogyne, w = worker; — = no sample)

		Coordinates			Individuals		
Locality	Area	N	E	Nests	m	M	w
TN	Tvärminne	59.844	23.236	18	119	13	131
TK	Tvärminne	59.842	23.221	5	14	31	9
PH	Pikku-Huopalahti	60.197	24.887	2	15	—	16
UU	Uutela	60.199	25.188	3	17	3	21
VE	Vermo	60.213	24.839	20	103	29	127
VF	Viikki	60.205	25.022	7	50	20	36
VL	Viikki	60.207	24.999	1	8	8	—
VM	Viikki	60.215	25.025	7	55	5	51
VP	Viikki	60.215	24.992	6	48	3	45
KA	Karjaa	60.052	23.641	4	26	3	29
SI	Sipoo	60.340	25.240	7	56	21	35
Total				80	511	136	500

Genotyping

We extracted DNA from ant legs with the Chelex method (Walsh *et al.* 1991; Bio-Rad) and kept the ants as vouchers. We tested primers of 21 microsatellite loci, developed for different *Myrmica* species (Evans 1993; Herbers & Mouser 1998; Henrich *et al.* 2003; Azuma *et al.* 2005; Zeisset *et al.* 2005); of these, eleven loci were polymorphic in *M. rubra* and were analysed in two panels (Table S2, Supporting information). We amplified the microsatellite loci using QIAGEN® Multiplex PCR Kit (QIAGEN Multiplex PCR Handbook 2008). The 9.5 μL PCR reactions contained 0.06–0.2 μM of each primer, 5 μL 2x QIAGEN Multiplex PCR Master Mix, 4 μL RNase-free water and 0.5 μL DNA. We used the following PCR protocol for amplification of the loci: initial denaturation at 95 °C for 15 min, 35 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 90 s, extension at 72 °C for 1 min, and final extension at 60 °C for 30 min. We mixed 1 μL of the 1:300 MilliQ water-diluted PCR product into 9 μL of 1:500 formamide-diluted GS500ROX size standard and resolved the microsatellite loci by Applied Biosystems 3730 DNA Analyzer. We scored the genotypes with the program GENEMAPPER v4.1 (Applied Biosystems).

Genetic data analyses

We analysed the microgynes (from $n = 11$ locations) and macrogynes ($n = 10$ locations) as separate groups in each locality, and as the origin of the workers was not known (which morph produced the worker), we

© 2015 John Wiley & Sons Ltd

analysed also them (n = 10 locations) as a separate group in most analyses (Table 1). Here, for practical reasons, we refer to each group (microgynes, macrogynes and workers) from each locality as separate, syntopic populations. We tested each locus for null alleles and possible genotyping errors in all three groups with individual inbreeding model (IIM) in the program INEST (Chybicki & Burczyk 2009), which is designed for detecting null alleles in populations with high inbreeding (Chybicki & Burczyk 2009; Campagne *et al.* 2012). For most loci, we estimated similar frequencies of null alleles in the data (range of averages 0.07–0.011). In one locus (Msca 7), however, we detected a much higher frequency of null alleles (average ± SD 0.15 ± 0.08), and thus, we excluded Msca7 locus from the analyses.

We calculated the unbiased expected heterozygosity (H_E) (Nei 1987) for each local group for macrogynes, microgynes and workers in ARLEQUIN 3.11 (Excoffier *et al.* 2005). Then, we compared the weighted averages across populations in microgynes and macrogynes for H_E and the number of alleles in each locus with the Wilcoxon signed-rank test in SPSS for Windows 21.0 (SPSS Inc., Chicago). We tested deviations from random mating (i.e. Hardy–Weinberg equilibrium, *HWE*) in each locus and overall in each population with exact tests (Guo & Thompson 1992), and linkage disequilibrium (*LD*) for each locus pair using the log-likelihood ratio statistic (*G*-test). We performed these tests separately for all three groups with GENEPOP on the Web (Raymond & Rousset 1995; Rousset 2008). If significant LD was not constant in a locus across populations, we considered it as a statistical fluke rather than a LD problem with the locus, and we retained all loci in further analyses (see also Discussion).

We identified identical multilocus microsatellite genotypes in two or more individuals with the software GENALEX 6.1 (Peakall & Smouse 2006). We assessed the likelihood for each multilocus genotype to be a result of sexual rather than asexual reproduction in each population (Arnaud-Haond *et al.* 2007) with the software MLGSIM 2.0 (Stenberg *et al.* 2003). MLGSIM uses the population's allele frequencies and the number of heterozygous loci to estimate the likelihood of finding at least as many identical multilocus genotypes as observed in the population (assuming *HWE*), given sexual reproduction (Stenberg *et al.* 2003). We calculated P_{SEX} values, which give the likelihood that individuals share the same multilocus genotype by chance as a result of sexual reproduction, and simulated the corresponding statistically critical P values (1000 Monte Carlo simulations). We restricted the analysis to the microgyne populations, as most identical multilocus genotypes were found in them.

We used several methods to explore the genetic divergence of the host and its microgynous social parasite. First, we tested genic differentiation for each locus and overall for each population pair with the exact *G*-test in GENEPOP on the Web. Second, we illustrated the divergence of the groups with a principal component analysis (PCA). The PCA was based on the covariance matrix of the allele frequencies of nest samples of each female caste and was carried out in SPSS.

Third, we estimated genetic differentiation both within and between the queen morphs, both overall and between all population pairs with *F*-statistics. The traditional F_{ST} values, widely used to estimate differentiation, will remain low even between populations with no alleles in common, when within-population variation approaches unity for highly variable markers (Jost 2008; Bird *et al.* 2011). Thus, we also standardized F_{ST} by dividing it with its theoretical maximum value, to obtain F'_{ST} (Meirmans & Hedrick 2011). We estimated the *F*-statistics in GENODIVE 2.0b24 (Meirmans & Van Tienderen 2004; Meirmans 2012) assuming an infinite allele model, and tested deviations of F_{ST} from zero by 99 999 permutations. We adjusted the P values to multiple tests (treating each set of within-morph and between-morphs pairwise comparisons as one test family) by correcting for false discovery rate under dependency (Benjamini & Yekutieli 2001) in R 2.15.1 (R Core Team 2012) with the function p.adjust.

Jost (2008) criticized the use of F_{ST} as a measure of genetic differentiation, because diversity estimates based on heterozygosity or allele frequency variance do not scale linearly. Thus, we also estimated in GENODIVE Jost's D (D_{EST}), which is expected to yield unbiased estimates of genetic differentiation (Jost 2008). For both pairwise *F*-statistics and D_{EST}, we calculated the test-family means and their 95% confidence intervals in R. We used both *F*-indices and D_{EST} as they provide mutually supporting information, but depending on the circumstances, may have index-specific problems (Jost 2008; Whitlock 2011; Alcala *et al.* 2014). In addition, by reporting multiple estimates, our results are easily comparable to studies that have used only a single estimate.

Fourth, we studied genetic differentiation with hierarchical approaches to choose among the speciation-related hypotheses. We partitioned the genetic diversity in the total queen data (80 nests) and separately in each queen morph including only the nests that included both morphs (40 nests). On the total data in GenoDive, we did two-four-level hierarchical AMOVAS: (i) the two morphs nested within each locality, and (ii) the localities nested within each morph. We performed the AMOVA assuming an infinite allele model and tested with 99 999 permutations deviations from zero. As genetic diversity at the nest level is often high in social insect

© 2015 John Wiley & Sons Ltd

populations, including *M. rubra*, it ought to be controlled for by including it in the hierarchical analysis. Because AMOVA accommodates only four hierarchical levels, we did the AMOVA also separately for each queen morph, now subordinating nests to localities. Then, we did with the total data a hierarchical *F* analysis in HIERFSTAT (Goudet 2005; de Meeûs & Goudet 2007) as above in AMOVA but including nest as an additional level. We tested the *F* values for deviances from zero with 10 000 permutations.

Fifth, we performed AMOVA on published mitochondrial sequences to compare genetic differentiation between the queen morphs in the nuclear and mitochondrial genomes. We did the AMOVA with ARLEQUIN as above, grouping in turn by morph versus locality; the statistical significances of the values were determined with 16 000 permutations.

Finally, we carried out model-based Bayesian clustering analyses to study the number of genetic clusters in each locality and the genetic origin of the workers using the software BAPS 5.4 (Corander & Marttinen 2006; Corander & Tang 2007; Corander *et al.* 2008). BAPS clusters samples with similar allele frequencies but does not minimize *LD* or deviations from *HWE* and should therefore work well for our data (see Results). In the analyses, we grouped the individuals either by nests, or as three separate female groups. First, we performed a mixture analysis in BAPS, which assigns individuals into clusters. Then, we used the clustering of the mixture analysis and performed an admixture analysis to calculate for each individual which proportion of the genome originates from which cluster with 500 iterations, 1000 reference individuals and 100 iterations per reference individual. From the posterior probabilities of the admixture proportions, we calculated the average admixture between the queen morphs in each locality and estimated the share of workers produced by each queen morph. We also studied the genetic origin of the workers in GENODIVE where we estimated hybrid indices (Buerkle 2005) to compare the pooled worker data with those of micro- and macrogynes, respectively. Index values close to one indicate close affinity of the individual worker with the microgynes, and values close to zero, with the macrogynes.

Results

Genetic composition of the study populations

Most microgyne populations shared one frequent allele in all loci; the frequency of the most common allele was 0.84 ± 0.10 (average ± SD), whereas the allele frequencies of the macrogynes and workers were more even (macrogynes 0.50 ± 0.22; workers 0.51 ± 0.21; Table S3, Fig. S1, Supporting information). Consequently, the microgyne populations had significantly fewer alleles and lower expected heterozygosity (H_E) than the syntopic macrogyne populations (Wilcoxon signed-rank tests, $P \leq 0.007$ for both tests; Tables 2, S2 and S3, Fig. S1, Supporting information). The microgynes, macrogynes

Table 2 Estimates of genetic variation in study populations

	N_A			$N_{Private}$		$\%_{HOMO}$			H_E			$N_{CLUSTERS}$	$\%_{IN\ CLUSTERS}$
Loc	m	M	w	m	M+w	m	M	w	m	M	w	m	m
TN	31	37	38	1	9	32	0	5	0.17	0.50	0.55	13	55
TK	17	44	26	1	29	86	0	0	0.07	0.50	0.42	1	86
PH	24	—	28	3	7	40	—	0	0.16	—	0.51	3	60
UU	16	20	31	2	18	12	0	5	0.13	0.45	0.45	4	71
VE	36	42	49	0	13	38	0	2	0.24	0.59	0.58	7	61
VF	22	24	35	5	19	38	0	6	0.16	0.41	0.55	11	92
VL	20	18	—	5	3	13	13	—	0.26	0.31	—	1	25
VM	11	27	30	1	20	98	0	2	0.02	0.44	0.40	1	96
VP	27	25	37	5	17	23	0	0	0.19	0.54	0.48	11	77
KA	16	20	34	2	20	35	0	14	0.19	0.43	0.48	7	81
SI	24	54	42	2	35	32	0	0	0.14	0.60	0.53	6	71
Mean	22	31	35	2.5	17.3	41	1	3	0.16	0.48	0.49	5.9	70
SD	7	12	7	1.8	9.3	27	4	4	0.07	0.16	0.08	4.3	20

For each locality (Loc), total number of alleles (N_A), private alleles ($N_{Private}$), percentage of individuals that were homozygous for all loci ($\%_{HOMO}$), and mean expected heterozygosity (H_E) across loci are given for each female caste sampled (m = microgyne, M = macrogyne, w = worker, — = no sample); for microgynes, also number of clusters of identical multilocus genotypes ($N_{CLUSTERS}$) and percentage of microgynes that belonged to such a cluster of all microgynes sampled ($\%_{IN\ CLUSTERS}$) are given.

© 2015 John Wiley & Sons Ltd

and workers shared 68% of their alleles. The microgynes had less private alleles (not present in the other two groups) than the combined macrogyne and worker data, both as a direct count (5 vs. 21) and relatively to the number of alleles found in each group (9% vs. 30%). The average number of private alleles in syntopic populations was also significantly smaller in the microgynes than in the combined macrogyne and worker data (Wilcoxon signed-rank tests, $P = 0.004$; Tables 2, S2 and S3, Fig. S1, Supporting information).

Many tests showed a significant deviation from Hardy–Weinberg equilibrium (*HWE*) and significant linkage disequilibrium (*LD*) at the significance level $P = 0.05$ (Tables S4 and S5, Supporting information). Linkage disequilibrium together with deviations from *HWE* was concentrated in particular populations (microgynes in the localities TN, VE, VF, VP and KA; macrogynes in TN, TK, VE and SI) rather than in certain loci. Interestingly, these populations had on average higher genetic diversity than randomly mating populations in linkage equilibrium – the average number of alleles in microgynes 2.6 vs. 1.9, and in macrogynes 4.4 vs. 2.2; H_E in microgynes 0.19 vs. 0.13, and in macrogynes 0.55 vs. 0.43 (not tested because of the small number of observations). In addition, these populations that were in *LD* and deviated from *HWE* had significantly more alleles in each locus (Wilcoxon signed-rank test, $P < 0.006$ for both microgynes and macrogynes) and more locus–population combinations having more than three alleles (three is the maximum allele number in a locus in a population colonized by a single, singly mated female; Fisher's exact test, for microgynes $P = 0.0011$, macrogynes $P < 0.0001$) than randomly mating populations with no *LD*. Therefore, the results rather indicate heterogeneity of these population samples (Wahlund effect, Wahlund 1928) than *LD* and true deviation from random mating.

We found 72 genotypes that were identical for multiple loci in two or more individuals. We refer here to such sets of identical multilocus genotypes by the shorthand 'IMG clusters'. These included 72% of the microgynes ($n = 367$), 11% of the workers ($n = 54$) and one macrogyne (Table 2). The microgynes and workers shared 13 IMG clusters and had 45 and 13 unique ones, respectively. One worker shared a cluster with the macrogyne. We found IMG clusters in most nests studied (72 of 80). Usually, the microgyne IMG clusters were found in more than one nest (median 2, range 1–25), but mostly each worker cluster was confined to a single nest (median 1, range 1–4). We found a single IMG cluster in 26 nests. These included eight nests where all microgynes analysed belonged to a single cluster. In 46 nests, we detected members of more than one IMG cluster (median 3, range 2–6). Most clusters (67 of 72) were confined to a single locality, although five were present in six locales.

Many microgynes, including many of the microgyne IMG clusters, were homozygous for most or all ten microsatellite loci studied, signalling possible clonal reproduction; the average proportion of multilocus homozygotes across populations was 41% (Table 2). The P_{SEX} values – which give the likelihood of sexual reproduction – for most identical multilocus genotypes (72%) were, however, high. Because deviations from *HWE* and linkage equilibrium tend to bias P_{SEX} downward, we looked separately at the subset of populations that showed *LD* and deviated from *HWE*. In these, 35% and 49% of the IMG clusters and individuals, respectively, had a statistically significantly small P_{SEX} value. In populations that neither deviated significantly from the Hardy–Weinberg nor linkage equilibrium, on average 22% and 23% of the P_{SEX} values for clusters and individuals, respectively, were small and significant, thus not excluding asexual reproduction. One of these populations (TK) included a single identical multilocus genotype that had a significantly small P_{SEX} value, which biased the averages. Excluding this population decreased the above averages to 7% and 8% of the clusters and individuals, respectively.

Genetic divergence of microgynes and macrogynes

Genic differentiation was always highly significant both within and between the morphs or castes, with syntopic macrogyne and worker samples as exceptions (*G*-tests, Table S6, Supporting information). In the PCA, the first two principal components explained 43% and 8% of the total genetic variation in the microsatellite data. The first principal component separated most microgynes and macrogynes, and most workers and macrogynes sampled from the same nest clustered close to each other (Fig. 2). Occasionally workers, however, were located with or close to the microgynes.

Pairwise estimates of F_{ST}, F'_{ST} and D_{EST} were higher between populations of different queen morphs than between populations of the same morph (Tables 3, S7 and S8, Supporting information). F_{ST} and F'_{ST} estimates were higher between populations of microgynes than macrogynes, but the opposite was true for D_{EST} (Table 3).

In the hierarchical analyses in the AMOVA, when the queen morphs were nested within the localities, the variance component for the among-localities level was strongly negative and explained none of the variation in the nuclear data – the majority of the variation was between the queen morphs (Table 4A, Fig. S2A, Supporting information). This resulted in a negative F_{CT} value among the localities and a highly significant F_{SC}

© 2015 John Wiley & Sons Ltd

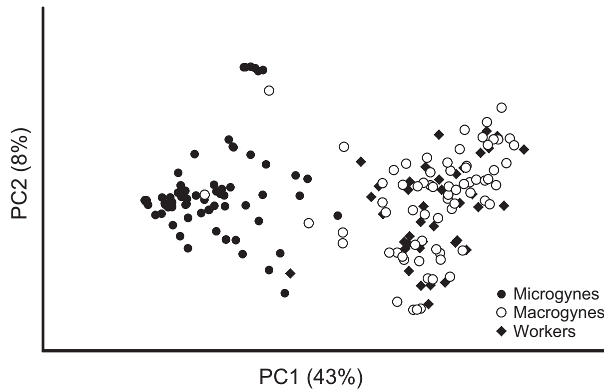

Fig. 2 Ordination with PCA showing the first two components for each nest of microgynes (black circles), macrogynes (white circles) and workers (black squares).

value between the queen morphs. In contrast, when the localities were nested within the queen morphs, the result was more balanced, although the majority of the variation was again between the queen morphs (Table 4B, Fig. S2B, Supporting information). When we did the analysis separately for the queen morphs, now including the nest level, both the locality and the nest level were significant in both morphs (Table 5).

Introducing nests as a hierarchical level to the *F* analyses of the nuclear data did not change the results qualitatively from those obtained with the AMOVA. When the queen morphs were nested within the localities, the *F* index for genetic differentiation among the localities was highly negative (details not shown). In the alternative hierarchy (localities nested within queen morphs), the genetic differentiation was apparently high and statistically significant ($P = 0.0001$) at all levels of interest (between queen morphs $F_{CT} = 0.401$, among localities $F_{SC} = 0.271$ and among nests $F_{ZS} = 0.233$).

Mitochondrial sequences gave contradicting results for the alternative hierarchies. When the queen morphs were nested within the localities, differentiation among the localities was low and not statistically significant, but differentiation between the queen morphs was high and significant (Table 4A, Fig. S2A, Supporting information). In contrast, when the localities were nested within the queen morphs, differentiation between the queen morphs was low and nonsignificant, but among the localities high and statistically significant (Table 4B, Fig. S2B, Supporting information).

In the mixture analysis in BAPS, the microgynes and macrogynes clustered separately in each locality. The number of optimally partitioned clusters was two in all localities ($P = 1$) except in VF, where the workers formed a third cluster ($P = 1$, Fig. 3). The analysis performed by grouping at the nest level resulted in more groups than that based on grouping by queen morph – reflecting the heterogeneity of populations – but the main result persisted: the microgynes and macrogynes belonged to separate genetic clusters.

The admixture analysis implied some hybridization between the queen morphs, with 3% and 2% of microgyne and macrogyne queens being of admixed origin, respectively. Of the workers, 92% clustered with the macrogynes, 5% with the microgynes and 3% had an admixed origin (25 workers in six nests, range 1–10 workers per nest). In most admixed workers, the majority of the genome was of microgyne origin. Buerkle's hybrid index also indicated genetic affinity of some workers with the microgynes, with 25 (5%) having a hybrid index = 1, and another eight (2%) an index >0.9.

Discussion

The traditional biological species concept (Mayr 1942, 2004) defines species by reproductive isolation from other species. This is in contrast to the ecological species concept (Shvarts 1977), which emphasizes that species cannot be defined by reproductive isolation from other species, although they usually do not hybridize with each other. In this view, speciation is an ecologically adaptive process in which reproductive isolation is incidental. Indeed, extensive hybridization of species is well documented (e.g. Shvarts 1977; Arnold 1997) – at least 10% of animal species hybridize and potentially

Table 3 The average pairwise F_{ST}, F'_{ST} and D_{EST} (95% confidence intervals) statistics among the populations with nuclear data (pairwise values are given in Tables S7 and S8)

Level	F_{ST}	F'_{ST}	D_{EST}
Among microgynes	0.402 (0.347–0.456)^{A}	0.462 (0.404–0.520)^{A*}	0.135 (0.114–0.157)^{A}
Among macrogynes	0.225 (0.192–0.257)^{B}	0.396 (0.348–0.443)^{A*}	0.273 (0.238–0.309)^{B}
Between queen morphs	0.595 (0.569–0.620)^{C}	0.784 (0.766–0.803)^{B}	0.506 (0.488–0.524)^{C}
Between local queen morphs	0.594 (0.493–0.696)	0.775 (0.697–0.853)	0.487 (0.397–0.578)

All indices are significantly larger than zero ($P < 0.0001$). Results of Welch's two-sample *t*-tests within F_{ST}, F'_{ST} and D_{EST} summary comparisons: different superscripts A, B and C denote statistically significant differences between the groupings at $P < 0.0001$, for the pair marked with *, $P < 0.10$ (subsample local queen morphs excluded from tests).

© 2015 John Wiley & Sons Ltd

Table 4 Analysis of molecular variance (AMOVA) of the total nuclear DNA microsatellite and mitochondrial DNA sequence data with (A) queen morphs nested within localities and (B) localities nested within queen morphs; population refers to morph within locality; the shares of total variation on each hierarchical level are shown in Fig. S2

	Nuclear DNA				Mitochondrial DNA	
Hierarchical level	% of total	*F* index	SD	*F′* index	% of total	*Φ* index
(A)						
Among localities	−35.4	F_{CT}: −0.354	0.084	−0.789	3.7	Φ_{CT}: 0.037
Between queen morphs within localities	77.5	F_{SC}: 0.572*	0.030	0.778	72.3	Φ_{SC}: 0.751*
Among individuals within populations	11.9	F_{IS}: 0.206*	0.036	—	24.0	Φ_{ST}: 0.760*
Within individuals in the total data	46.0	F_{IT}: 0.540^{NT}	0.035	—	—	—
(B)						
Between queen morphs	39.0	F_{CT}: 0.390*	0.063	0.626	−4.5	Φ_{CT}: −0.045
Among localities within queen morphs	18.0	F_{SC}: 0.294*	0.039	0.385	79.9	Φ_{SC}: 0.765*
Among individuals within populations	8.8	F_{IS}: 0.205*	0.036	—	24.6	Φ_{ST}: 0.754*
Within individuals in the total data	34.3	F_{IT}: 0.657^{NT}	0.031	—	—	—

F, *F′* and *Φ* indices marked with * are significantly larger than zero ($P \leq 0.001$); and ^{NT} not tested.

Table 5 Analysis of molecular variance (AMOVA) with nuclear data for each queen morph, sampled from nests with both queen morphs (n = 40 nests); grouped by locality

	F index		% of total variation	
Hierarchical level	Macrogynes	Microgynes	Macrogynes	Microgynes
Among localities	F_{CT}: 0.144* F'_{CT}: 0.317^{NT}	0.358* 0.438^{NT}	14.36	35.80
Among nests	F_{SC}: 0.172* F'_{SC}: 0.316^{NT}	0.259* 0.300^{NT}	14.70	16.61
Among individuals within nests	F_{IS}: −0.001	0.074*	−0.04	3.53
Within individuals in the total data	F_{IT}: 0.290^{NT}	0.559^{NT}	70.98	44.06

F'_{CT} and F'_{SC} are standardized *F* values. Fixation indices (*F*) marked with *are significantly larger than zero $P \leq 0.01$; and ^{NT} not tested.

experience genetic introgression (Mallet 2005). Species are often incompletely isolated for millions of years (Mallet 2005), during which the species status of a diverging biological entity may be debated and depends on the definition of species (Harrison 1998). Here, we follow the traditional biological species concept and suggest that our study object, the inquiline of *Myrmica rubra*, is a nascent species still genetically diverging from its host and occasionally hybridizing with it.

Next, we will discuss our results to evaluate our hypotheses (formulated in the Introduction) on the speciation of the inquiline of *M. rubra*.

The genetic composition of Myrmica rubra populations

The populations of both the microgyne and macrogyne of *M. rubra* could be divided into two categories. The first one consisted of homogeneous randomly mating populations. The species often forms large polydomous colonies (supercolonies; Seppä & Pamilo 1995; Walin *et al.* 2001), and hence, it is plausible to conclude that each of these populations was sampled from a single polydomous colony. Because a singly mated dispersing female carries with her maximally three alleles, the number of alleles per locus in this category (maximally 3) further suggests that these populations were typically founded by a single colonization event. The other category comprised heterogeneous populations where mating was not random, and these populations probably contained two or more polydomous colonies.

The microgyne populations were characterized by a high frequency of individuals identical in all microsatellite loci studied, such that they often comprised identical multilocus homozygotes. One possible cause for such extreme homogeneity is thelytoky (clonal reproduction), which is rare in ants but has recently been observed in several species (Pearcy *et al.* 2004;

© 2015 John Wiley & Sons Ltd

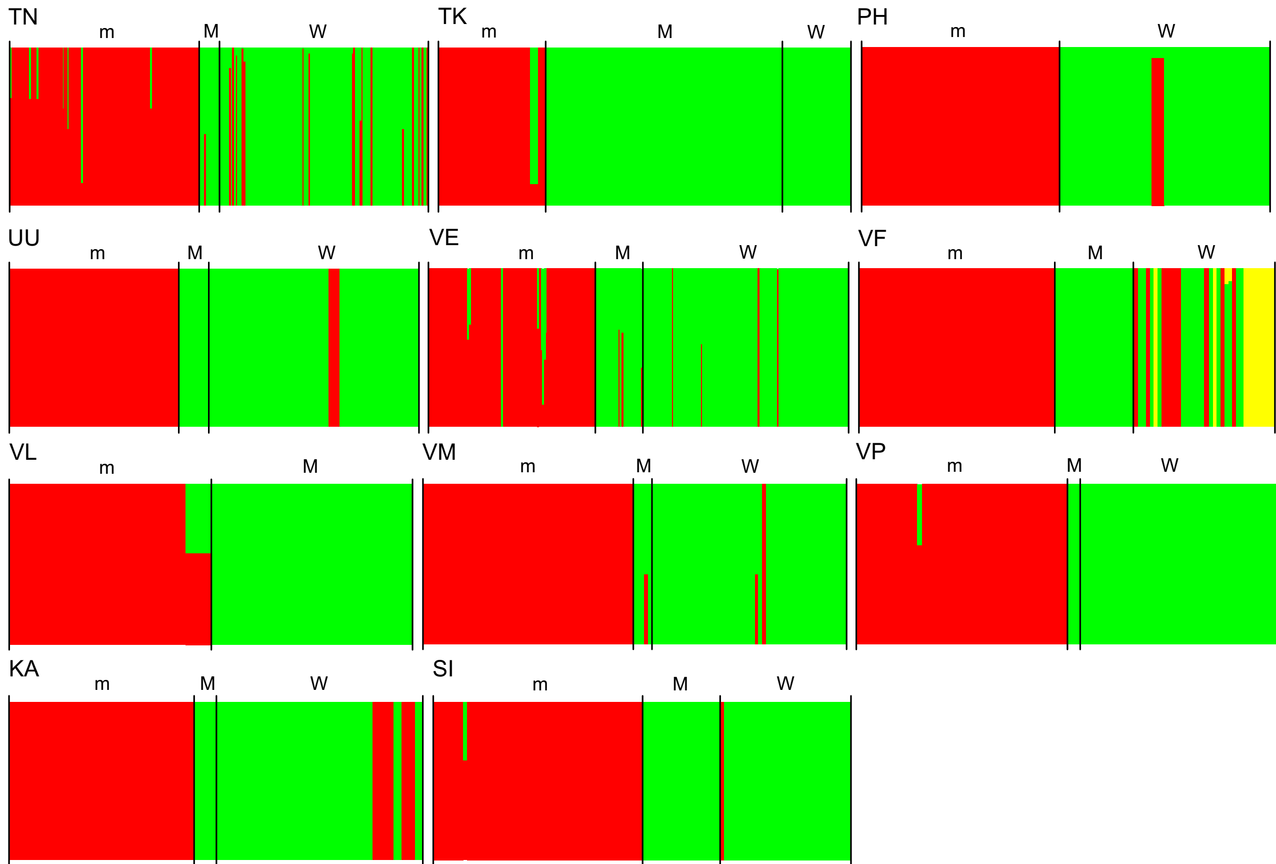

Fig. 3 The admixture partition with BAPS for each locality; significant ($P \leq 0.05$) admixture estimates shown (locality code above left corner of each partition; for locality information, see Table 1). Each bar represents an individual, and clusters of different castes and morphs (m = microgynous parasite, M = macrogynous host and W = worker) are separated with vertical black lines.

Wenseleers & Van Oystaeyen 2011; Rabeling & Kronauer 2013). Thelytoky can quickly cause extreme homozygosity and individuals identical for multiple loci within a population (Wenseleers & Van Oystaeyen 2011). Essentially, thelytoky may contribute to sympatric speciation by promoting barriers to gene flow (Adachi-Hagimori *et al.* 2011; Neumann *et al.* 2011). In our populations, however, sexual reproduction coupled with low genetic variation rather than thelytoky is a more probable explanation for the identical genotypes. A few factors probably contributed to some of the clusters of identical genotypes showing seemingly significant signs of asexual reproduction. First, the test procedure in MLGSIM assumes that populations are randomly mating, but most of our clusters that suggested thelytoky belonged to populations that had an excess of homozygotes and were not in *HWE*, thus breaking the assumptions of the analysis (Stenberg *et al.* 2003). Increasing homozygosity in the populations increases the probability of finding incorrect clones. Second, in ants, nestmates may be closely related (including *M. rubra*, Seppä & Walin 1996), which overestimates the number of (false) clones in the data (Stenberg *et al.* 2003). Third, clusters of identical multilocus genotypes that showed significant signs of asexual reproduction had on average one more locus with rare alleles than clusters that did not show significant signs of asexual reproduction, which could indicate that the analysis is rather sensitive to the occurrence of rare alleles.

The spatial structure of Myrmica rubra populations

Genetic differentiation in the nuclear markers was substantial. This was observed both among populations within each morph, even over distances potentially covered by dispersing alates (winged sexuals), and between the morphs when syntopic. The result is generally robust, and the largest differentiation was always measured between the queen morphs. However, F_{ST} was significantly larger among the microgyne than the macrogyne populations, whereas the standardized F'_{ST} values of the morphs were similar. The dependency of uncorrected F_{ST} on expected heterozygosity (Meirmans & Hedrick 2011) probably resulted in biased estimates

© 2015 John Wiley & Sons Ltd

when comparing the microgyne and macrogyne populations among themselves, owing to the large difference in within-population heterozygosity in the two morphs. Jost's D_{EST}, which is based on the effective number of alleles rather than heterozygosity (Jost 2008), seems thus a better estimate of genetic differentiation in this system than F_{ST}. Indeed, D_{EST} showed that actually the differentiation among the microgyne populations is lower than that among those of the macrogynes. Moreover, similar to the standardized F'_{ST}, D_{EST} signalled most pronounced differentiation among the morphs.

Differences in genetic variation make F_{ST} estimates context-dependent, and comparing results from multiple studies becomes difficult. Like here, Vepsäläinen *et al.* (2009) found significant genetic differentiation in nuclear markers between populations within the morphs and between the morphs when syntopic, and they confirmed strong spatial structuring with PCA ordination and model-based clustering. Steiner *et al.* (2006) also obtained similar results on F_{ST}, but their model-based clustering did not separate the microgyne and macrogyne populations. The most important determinant for the distribution of genetic variation in AMOVA was the queen morph in Vepsäläinen *et al.* (2009), whereas Steiner *et al.* (2006) did not test a model where the localities would have been nested within the queen morphs.

Recurrent founder effects seem to emerge as the major determinant of genetic composition of both microgyne and macrogyne populations in *M. rubra*, although the founder effect seems more pronounced in the microgyne (Vepsäläinen *et al.* 2009). We considered founder effect as the most plausible reason for low genetic diversity in the randomly mating populations and for the clusters of identical multilocus genotypes. The spatial population structure of the macrogyne is strongly influenced by founder effect when new populations are founded, and when they are infested by the microgynous parasite. The microgyne has, however, evolved many peculiar features of the so-called inquiline syndrome (Hölldobler & Wilson 1990), supposedly including restricted dispersal and mating in the nest (Seifert 1993), which ought to further impoverish the gene pool of the microgyne.

Hybridization and parasite worker production

In the admixture analysis, a small number of both microgynous and macrogynous queens and workers were identified as hybrids, but the majority of these individuals were from two localities (6.1% in TN; 4.6% in VE) where populations were in *LD* and not in *HWE*.

Our results on private alleles and admixture analysis gave mixed signals about the direction of the gene flow between the two morphs. The macrogynes and workers had significantly more private alleles than the syntopic microgynes had. This suggests that the microgynous males mate with macrogynous females more often than the other way around; otherwise, more macrogyne alleles ought to end up in the microgyne population. On the other hand, as a nonreproducing caste, workers should first and foremost show the signs of hybridization in their genomes, because workers are probably less susceptible for selection against hybrids than queens are. As the majority of the genomes of admixed workers originated from the gene pool of the microgynes, the macrogynous males must have mated with microgynous females more often than conversely.

Host workers provide the workforce needed for inquiline social parasites, which allows the parasite to allocate to sexual reproduction. Consequently, worker production by inquilines is thought to be selected against and hence expected to disappear quickly when a species becomes parasitic (Nonacs & Tobin 1992). Recently, Schär & Nash (2014), however, showed that the microgynous inquiline of *M. rubra* is able to produce workers in the laboratory. Further, they suggested that the potential to produce workers varies substantially among microgyne lineages — another indication of recent origin of the parasite. Based on an admixture analysis on *M. rubra* microgynes and workers, Vepsäläinen *et al.* (2009) suggested that microgynes sometimes produced workers also in the field. Here, we confirmed this by analysing workers and their putative mothers, microgynous and macrogynous queens, from the same nests. We showed that workers sometimes cluster with the microgynes, leaving little doubt about their progenitor. Overall, worker production by the *M. rubra* microgyne supports the result obtained with phylogenetic analyses (Savolainen & Vepsäläinen 2003; Jansen *et al.* 2010; Leppänen *et al.* 2011) that the microgyne is — relative to the established inquiline species in the genus *Myrmica* (Radchenko & Elmes 2003) — evolutionarily a young social parasite. A relatively young inquiline species, *M. hirsuta* (Jansen *et al.* 2010), also produces in the field occasional workers (Elmes 1994) or worker-like ergatoids (Seifert 2007). Actually, inquiline workers may also have an adaptive role by augmenting sexual production of the inquiline while suppressing that of the host (Sumner *et al.* 2003), which would retard the disappearance of the worker caste.

Is the microgynous parasite speciating?

The various results, strong genetic divergence between the locally co-occurring morphs, PCA ordination, model-based clustering and the hierarchical analyses on the nuclear data grouped by morphs — the morphs

© 2015 John Wiley & Sons Ltd

explaining a significant fraction of the genetic variation — converged towards the same conclusion. They all support the *hypothesis of advanced sympatric speciation*, which postulates that the two morphs have passed the initial phases of divergence. The result is consistent with that of the alternative hierarchy. When strongly diverged morphs are pooled within localities — as if they belonged to a single interbreeding population — the genetic differentiation among the localities may be lower than that within localities. Consequently, as the variance components in the analyses are actually covariances (Yang 1998; Meirmans 2006), a highly negative value at the locality level may ensue.

Sporadic hybridization between the morphs, however, testifies that the microgyne has not yet reached full reproductive isolation from its host, but speciation is still ongoing. The relevance of complete reproductive isolation as a criterion of a 'good' species is, however, disputed as a substantial proportion of both plant and animal species interbreed with other species (Mallet 2005). In *M. rubra*, strong divergence between the morphs at local scale is now emerging as a general result (Pearson & Child 1980; Vepsäläinen *et al.* 2009; this work). However, the strong differentiation among local populations of each morph (Seppä & Pamilo 1995; Nash *et al.* 2008; Vepsäläinen *et al.* 2009, this work), which probably is due to recurrent founder effects, obscures the analytical separation of local, stochastic population processes from divergence during speciation. Each of these processes leads to extreme genetic structure.

In contrast to the results on nuclear markers, in the mitochondrial markers neither grouping by queen morphs nor localities could distinguish between the two hypotheses of ongoing speciation — these results therefore supported equally the *hypotheses of incipient* and *advanced sympatric speciation*. Our study implied, however, that in a smaller geographic scale, the differences between the queen morphs were proportionally more pronounced at the locality level than on a wider geographic scale. Previously, also with mitochondrial data, Leppänen *et al.* (2011) found that on an extensive geographic scale over the Palaearctic region, the local host and parasite were more closely related to each other than to the host or parasite in other localities. Such a pattern is typical of young diverging lineages, where lineage sorting is incomplete and hybrid introgression is still occurring (Funk & Omland 2003). Between the young inquiline species *Mycocepurus castrator* and its host *M. goeldii*, lineage sorting is complete in the mitochondrial (but not in the nuclear) sequences (Rabeling *et al.* 2014). On the contrary, in the inquiline of *M. rubra* lineage sorting is incomplete, which posits it in an earlier phase of speciation than *M. castrator*.

Nevertheless, the microgyne and macrogyne populations appear reproductively isolated to a large extent. If mating of the macrogyne and microgyne is spatially segregated (Buschinger 1997; Seifert 2007), substantial premating reproductive isolation between the morphs already exists. It is not known how frequently the morphs cross-breed when in contact (although experimentally cross-mating is easily achieved; J. Leppänen, P. Seppä, K. Vepsäläinen & R. Savolainen, unpublished), but our genetic results imply that hybridization between the morphs is sporadic and weak at best.

There is good evidence showing that the inquiline of *M. rubra* indeed behaves like a good species; both the parasite and the host produce only progeny of their own kind (e.g. Faber 1969; Elmes 1976; Seifert 2007, 2010). Nevertheless, there are no diagnostic loci in our data that would distinguish the inquiline and its host, which would positively verify the reproductive isolation between the queen morphs of *M. rubra*. The discrepancy becomes comprehensible, if the DNA microsatellite markers are selectively neutral (Avise 2000, 2004; but see Behura & Severson 2015). Namely, particularly in sympatry divergence may be rapid, divergent selection affecting only a small part of the genome, while gene flow may continue over most of the genome (Via 2009). Therefore, gene flow need not be an obstacle to speciation (e.g. Mallet 2005; Forbes *et al.* 2013; Noutsos *et al.* 2014).

In conclusion, the strong and consistent differentiation between the queen morphs of *M. rubra*, with only little hybridization, supports the view that the microgynous morph is speciating sympatrically through intraspecific social parasitism. Obviously, our study system provides an outstanding example of sympatric speciation as the microgynous inquiline lives in obligate contact with its host, from which it is still diverging. For speciation research, this is a serendipitous situation. Usually, sympatric speciation is problematic to distinguish from the allopatric speciation followed by sympatry, because often the genetic lineages under study have already diverged and lineage sorting is complete (Bolnick & Fitzpatrick 2007; Fitzpatrick *et al.* 2009; Mallet *et al.* 2009). Such constraints on the research of speciation, inherent when evaluating historical events of closely related species, can be overcome by studying systems such as *M. rubra*'s inquiline and its host, where the mechanisms and processes of speciation are still detectable.

Acknowledgements

This study was funded by the Academy of Finland (project 210523 to RS), the Ella and Georg Ehrnrooth Foundation, the Finnish Cultural Foundation, the Alfred Kordelin Foundation,

© 2015 John Wiley & Sons Ltd

Societas pro Fauna et Flora Fennica, the Oskar Öflund Foundation, the Otto A. Malm Foundation and the Kuopio Naturalists' Society (to JL). We thank Heini Ali-Kovero, Leena Laaksonen, Minttu Ahjos and Suvi Saarnio for the help with the laboratory work. We obtained valuable comments on the manuscript by Laura Kvist, Serge Aron and anonymous reviewers.

References

Adachi-Hagimori T, Miura K, Abe Y (2011) Gene flow between sexual and asexual strains of parasitic wasps: a possible case of sympatric speciation caused by a parthenogenesis-inducing bacterium. *Journal of Evolutionary Biology*, **24**, 1254–1262.

Alcala N, Goudet J, Vuilleumier S (2014) On the transition of genetic differentiation from isolation to panmixia: what we can learn from G_{ST} and *D*. *Theoretical Population Biology*, **93**, 75–84.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.

Arnold ML (1997) *Natural Hybridization and Evolution*. Oxford University Press, New York City, New York.

Avise JC (2000) *Phylogeography. The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Azuma N, Takahashi J, Kikuchi T *et al*. (2005) Microsatellite loci for *Myrmica kotokui* and their application in some congeneric ant species distributed in northern Japan. *Molecular Ecology Notes*, **5**, 118–120.

Behura SK, Severson DW (2015) Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Research*, **22**, 29–38.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

Bird CE, Karl SA, Smouse PE, Toonen RJ (2011) Detecting and measuring genetic differentiation. *CrustIssues*, **19**, 31–54.

Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology Evolution and Systematics*, **38**, 459–487.

Bourke AFG, Franks NR (1991) Alternative adaptations, sympatric speciation and the evolution of parasitic, inquiline ants. *Biological Journal of the Linnean Society*, **43**, 157–178.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684–687.

Buschinger A (1990) Sympatric speciation and radiative evolution of socially parasitic ants - heretic hypotheses and their factual background. *Zeitschrift Fur Zoologische Systematik Und Evolutionsforschung*, **28**, 241–260.

Buschinger A (1997) Vorkommen der sozialparasitischen ameise *Myrmica microrubra* in Hessen (Hymenoptera, Formicidae). *Hessische Faunistische Briefe*, **16**, 49–57.

Buschinger A (2009) Social parasitism among ants: a review (Hymenoptera: Formicidae). *Myrmecological News*, **12**, 219–235.

Campagne P, Smouse PE, Varouchas G, Silvain JF, Leru B (2012) Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. *Molecular Ecology Resources*, **12**, 975–982.

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, **100**, 106–113.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology*, **15**, 2833–2843.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences*, **205**, 19–31.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

Elmes GW (1976) Some observations on microgyne form of *Myrmica rubra* L. (Hymenoptera, Formicidae). *Insectes Sociaux*, **23**, 3–21.

Elmes GW (1978) Morphometric comparison of three closely related species of *Myrmica* (Formicidae), including a new species from England. *Systematic Entomology*, **3**, 131–145.

Elmes GW (1994) A population of the social parasite *Myrmica hirsuta* Elmes (Hymenoptera, Formicidae) recorded from Jutland, Denmark, with a 1st description of the worker caste. *Insectes Sociaux*, **41**, 437–442.

Evans JD (1993) Parentage analyses in ant colonies using simple sequence repeat loci. *Molecular Ecology*, **2**, 393–397.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Faber W (1969) Beiträge zur Kenntnis sozialparasitischer Ameisen. 2. Aporomyrmex ampeloni nov. gen., nov. spec. (Hym. Formicidae), ein neuer permanenter Sozialparasit bei Plagiolepis vindobonensis Lomnicki aus Österreich. *Pflanzenschutz-Berichte*, **39**, 39–100.

Fitzpatrick BM, Fordyce JA, Gavrilets S (2009) Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, **22**, 2342–2347.

Forbes AA, Kelly PH, Middleton KA, Condon MA (2013) Genetically differentiated races and speciation-with-gene-flow in the sunflower maggot, Strauzia longipennis. *Evolutionary Ecology*, **27**, 1017–1032.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **34**, 397–423.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Harrison RG (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: *Endless Forms: Species and Speciation* (eds Howard DJ, Berlocher SH), pp. 19–31. Oxford University Press, New York City, New York.

Henrich KO, Sander AC, Wolters V, Dauber J (2003) Isolation and characterization of microsatellite loci in the ant *Myrmica scabrinodis*. *Molecular Ecology Notes*, **3**, 304–306.

Herbers JM, Mouser RL (1998) Microsatellite DNA markers reveal details of social structure in forest ants. *Molecular Ecology*, **7**, 299–306.

© 2015 John Wiley & Sons Ltd

Hölldobler B, Wilson EO (1990) *The Ants*. Harvard University Press, Cambridge.

Huang MH, Dornhaus A (2008) A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. *Ecological Entomology*, **33**, 589–596.

Jansen G, Savolainen R, Vepsäläinen K (2010) Phylogeny, divergence-time estimation, biogeography and social parasite-host relationships of the Holarctic ant genus *Myrmica* (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, **56**, 294–304.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Le Masne G (1956) Recherches sur les fourmis parasites. *Plagiolepis grassei* et l'évolution des *Plagiolepis* parasites. *Comptes Rendus des Séances de l'Académie des Sciences, Paris*, **243**, 673–675.

Leppänen J, Vepsäläinen K, Savolainen R (2011) Phylogeography of the ant *Myrmica rubra* and its inquiline social parasite. *Ecology and Evolution*, **1**, 46–62.

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

Mallet J, Meyer A, Nosil P, Feder JL (2009) Space, sympatry and speciation. *Journal of Evolutionary Biology*, **22**, 2332–2341.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York City, New York.

Mayr E (1993) Fifty years of progress in research on species and speciation. *Proceedings of the California Academy of Sciences*, **48**, 131–140.

Mayr E (2004) *What Makes Biology Unique?* Cambridge University Press, Cambridge.

de Meeûs T, Goudet J (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection Genetics and Evolution*, **7**, 731–735.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Meirmans PG (2012) AMOVA-based clustering of population genetic data. *Journal of Heredity*, **103**, 744–750.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Nash DR, Als TD, Maile R, Jones GR, Boomsma JJ (2008) A mosaic of chemical coevolution in a large blue butterfly. *Science*, **319**, 88–90.

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

Neumann P, Hartel S, Kryger P, Crewe RM, Moritz RFA (2011) Reproductive division of labour and thelytoky result in sympatric barriers to gene flow in honeybees (*Apis mellifera* L.). *Journal of Evolutionary Biology*, **24**, 286–294.

Nonacs P, Tobin JE (1992) Selfish larvae: development and the evolution of parasitic behaviour in the Hymenoptera. *Evolution*, **46**, 1605–1620.

Noutsos C, Borevitz JO, Hodges SA (2014) Gene flow between nascent species: geographic, genotypic and phenotypic differentiation within and between *Aquilegia formosa* and *A. pubescens*. *Molecular Ecology*, **23**, 5589–5598.

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

Pearcy M, Aron S, Doums C, Keller L (2004) Conditional use of sex and parthenogenesis for worker and queen production in ants. *Science*, **306**, 1780–1783.

Pearson B (1981) The electrophoretic determination of *Myrmica rubra* microgynes as a social parasite: possible significance in the evolution of ant social parasites. In: *Biosystematics of Social Insects* (eds Howse PE, Clément J-L), pp. 75–84. Academic Press, London.

Pearson B, Child AR (1980) The distribution of an esterase polymorphism in macrogynes and microgynes of *Myrmica rubra* Latreille. *Evolution*, **34**, 105–109.

R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rabeling C, Kronauer DJC (2013) Thelytokous parthenogenesis in eusocial Hymenoptera. *Annual Review of Entomology*, **58**, 273–292.

Rabeling C, Schultz TR, Pierce NE, Bacci M (2014) A social parasite evolved reproductive isolation from its fungus-growing ant host in sympatry. *Current Biology*, **24**, 2047–2052.

Radchenko A, Elmes GW (2003) A taxonomic revision of the socially parasitic *Myrmica* ants (Hymenoptera: Formicidae) of the Palaearctic region. *Annales Zoologici*, **53**, 217–243.

Raymond M, Rousset F (1995) Genepop (Version-1.2) - population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rousset F (2008) Genepop '007: a complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Savolainen R, Vepsäläinen K (2003) Sympatric speciation through intraspecific social parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 7169–7174.

Schär S, Nash DR (2014) Evidence that microgynes of *Myrmica rubra* ants are social parasites that attack old host colonies. *Journal of Evolutionary Biology*, **27**, 2396–2407.

Seifert B (1993) Taxonomic description of *Myrmica microrubra* n. sp. – a social parasitic ant so far known as the microgyne of *Myrmica rubra* (L.). *Abhandlungen und Berichte des Naturkunde Museums Görlitz*, **67**, 9–12.

Seifert B (2007) *Die Ameisen Mittel- und Nordeuropas*. Lutra Verlags- und Vertriebsgesellschaft, Tauer.

Seifert B (2010) Intranidal mating, gyne polymorphism, polygyny, and supercoloniality as factors for sympatric and parapatric speciation in ants. *Ecological Entomology*, **35**, 33–40.

Seppä P, Pamilo P (1995) Gene flow and population viscosity in *Myrmica* ants. *Heredity*, **74**, 200–209.

Seppä P, Walin L (1996) Sociogenetic organization of the red ant *Myrmica rubra*. *Behavioral Ecology and Sociobiology*, **38**, 207–217.

Shvarts SS (1977) *The Evolutionary Ecology of Animals*. Studies in Soviet Science, Consultant Bureau, New York City, New York (translated from Russian, original 1969).

Smith JA, Tierney SM, Park YC, Fuller S, Schwarz MP (2007) Origins of social parasitism: the importance of divergence ages in phylogenetic studies. *Molecular Phylogenetics and Evolution*, **43**, 1131–1137.

Steiner FM, Schlick-Steiner BC, Konrad H *et al*. (2006) No sympatric speciation here: multiple data sources show that the

© 2015 John Wiley & Sons Ltd

ant *Myrmica microrubra* is not a separate species but an alternate reproductive morph of *Myrmica rubra*. *Journal of Evolutionary Biology*, **19**, 777–787.
Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. *Molecular Ecology Notes*, **3**, 329–331.
Sumner S, Nash DR, Boomsma JJ (2003) The adaptive significance of inquiline parasite workers. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 1315–1322.
Vepsäläinen K, Ebsen JR, Savolainen R, Boomsma JJ (2009) Genetic differentiation between the ant *Myrmica rubra* and its microgynous social parasite. *Insectes Sociaux*, **56**, 425–437.
Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, **16**, 381–390.
Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.
Wahlund S (1928) Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas*, **11**, 65–106.
Walin L, Seppä P, Sundström L (2001) Reproductive allocation within a polygyne, polydomous colony of the ant *Myrmica rubra*. *Ecological Entomology*, **26**, 537–546.
Walsh PS, Metzger DA, Higuchi R (1991) Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **10**, 506–513.
Wasmann E (1909) Über den ursprung des sozialen parasitismus, der sklaverei und der myrmekophilie bei den ameisen. *Biologische Centralblatt*, **29**, 683–703.
Wenseleers T, Van Oystaeyen A (2011) Unusual modes of reproduction in social insects: shedding light on the evolutionary paradox of sex. *BioEssays*, **33**, 927–937.
Whitlock MC (2011) G'_{ST} and D do not replace F_{ST}. *Molecular Ecology*, **20**, 1083–1091.
Wilson EO (1971) *The Insect Societies*. Harvard University Press, Cambridge.
Yang RC (1998) Estimating hierarchical *F*-statistics. *Evolution*, **52**, 950–956.
Zeisset I, Ebsen JR, Boomsma JJ (2005) Dinucleotide microsatellite DNA loci from the ant *Myrmica scabrinodis*. *Molecular Ecology Notes*, **5**, 163–164.

R.S. and K.V. conceived the original idea; J.L., R.S. and P.S. designed the study; R.S., K.V. and J.L. collected the samples; J.L. performed the laboratory work; J.L., P.S. and K.V. conducted the data analyses; and J.L., P.S., K.V. and R.S. wrote the manuscript.

Data accessibility

Sampling locations: Table 1 and Fig. 1; characteristics and references of the microsatellite loci: Table S2 (Supporting Information); GenBank accession nos and references of mitochondrial DNA sequences: Table S1 (Supporting Information); and microsatellite genotypes and mitochondrial DNA sequences: Dryad doi:10.5061/dryad.07nj3.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Mitochondrial DNA sequences used in AMOVA.

Table S2 Characteristics of the microsatellite loci used in the study for the two primer panels; locus name and reference (in superscript), allele size range in base pairs (bp), number of alleles (N_A) and private alleles ($N_{Private}$) in microgynes (m), combined macrogyne and worker data (M+w), macrogynes (M) and workers (w).

Table S3 Allele frequencies of each locus for each caste and morph sampled from each locality (detailed locality information in Table 1).

Table S4 Probabilities for deviation from Hardy–Weinberg equilibrium (*HWE*) in exact test for each locus in each caste and morph sampled from each locality (detailed locality information in Table 1).

Table S5 Proportions of *LD* tests (*G* tests with $P < 0.05$, not corrected for the number of tests) in each locus for each caste and morph from each locality (detailed locality information in Table 1).

Table S6 Genic differentiation between *Myrmica rubra* castes and morphs (M = macrogynes; m = microgynes; w = workers) sampled from each locality (LOC, detailed locality information in Table 1).

Table S7 Genetic differentiation among populations (first two letters refer to locality of the population, the third one to queen morph: m = microgyne, M = macrogyne; detailed locality information in Table 1); below the diagonal pairwise F_{ST} estimates and above the diagonal pairwise D_{EST} estimates.

Table S8 Genetic differentiation among populations (first two letters refer to locality of the population, the third one to queen morph: m = microgyne, M = macrogyne; detailed locality information in Table 1); above the diagonal pairwise F'_{ST} estimates.

Fig. S1 Allele frequencies of the 10 microsatellite loci of the three female groups (details of loci in Tables S2 and S3).

Fig. S2 Scaled proportions of nuclear DNA and mitochondrial DNA variation explained by hierarchical levels in AMOVA, with (A) queen morphs nested within localities and (B) localities nested within queen morphs; see details in Table 4.

© 2015 John Wiley & Sons Ltd