doi:10.1111/j.1558-5646.2010.01170.x

ORIGIN AND EVOLUTION OF THE DEPENDENT LINEAGES IN THE GENETIC CASTE DETERMINATION SYSTEM OF *POGONOMYRMEX* ANTS

Anu Sirviö,[1,2] Pekka Pamilo,[1,3,4] Robert A. Johnson,[5,6] Robert E. Page, Jr.,[5,7] and Jürgen Gadau[5,8]

[1]*Department of Biology, University of Oulu 90014, Linnanmaa P.O. Box 3000, Finland*

[2]*E-mail: anu.sirvio@oulu.fi*

[3]*Department of Biosciences, University of Helsinki 00014, P.O. Box 65, Finland*

[4]*E-mail: pekka.pamilo@helsinki.fi*

[5]*School of Life Sciences, 1711 South Rural Road, State University of Arizona, Tempe, Arizona 85287*

[6]*E-mail: robert.johnson4@asu.edu*

[7]*E-mail: robert.page@asu.edu*

[8]*E-mail: jgadau@asu.edu*

Received January 27, 2010
Accepted September 30, 2010

Hybridizing harvester ants of the *Pogonomyrmex barbatus/rugosus* complex have an exceptional genetic caste determination (GCD) mechanism. We combined computer simulations, population genomics, and linkage mapping using >1000 nuclear AFLP markers and a partial mtDNA sequence to explore the genetic architecture and origin of the dependent lineages. Our samples included two pairs of hybridizing lineages, and the mitochondrial and nuclear data showed contradicting affinities between them. Clustering of individual genotypes based on nuclear markers indicated some exceptions to the general GCD system, that is, interlineage hybrid gynes as well as some pure-line workers. A genetic linkage map of *P. rugosus* showed one of the highest recombination rates ever measured in insects (14.0 cM/Mb), supporting the view that social insects are characterized by high recombination rates. The population data had 165 markers in which sibling pairs showed a significant genetic difference depending on the caste. The differences were scattered in the genome; 13 linkage groups had loci with F_{ST} > 0.9 between the hybridizing lineages J1 and J2.The mapping results and the population data indicate that the dependent lineages have been initially formed through hybridization at different points in time but the role of introgression has been insignificant in their later evolution.

KEY WORDS: AFLP, genetic differentiation, hybridization, linkage map, mtDNA, recombination.

Division of individuals into reproductive (queen) and nonreproductive (worker) castes is a fundamental property of ants and other social insects. The developmental and evolutionary processes involved in caste determination are mostly unknown (Schwander et al. 2010). Caste is usually considered to be based on environmental factors (environmental caste determination, ECD) but some ant species have been shown to have genetic caste determination (GCD) (*Solenopsis xyloni* and *S. geminata*; Hung and Vinson 1977; Helms Cahan and Vinson 2003, *Pogonomyrmex*; Helms Cahan et al. 2002; Julian et al. 2002; and Volny and Gordon 2002) or to rely on a combination of sexual and parthenogenetic reproduction (*Vollenhoevia*; Ohkawara et al. 2006, *Wasmannia*

© 2010 The Author(s). *Evolution* © 2010 The Society for the Study of Evolution.

auropunctata; Fournier et al. 2005). GCD is usually detected as co-segregation of genetic markers/patrilines and castes within colonies.

In *Pogonomyrmex* harvester ants with GCD, queens are derived from mating within a genetic lineage, whereas workers are derived from mating between two lineages and are thus hybrids (Helms Cahan and Keller 2003; Anderson et al. 2006a,b and Schwander et al. 2007a,b). The mtDNA phylogeny suggests that at least the maternal genealogies of such hybridizing lineages have not mixed recently (Helms Cahan and Keller 2003; Anderson et al. 2006a). Although the *Pogonomyrmex* GCD system is arguably the best understood system of GCD, we know very little about the proximate genetic architecture underlying caste determination in the *P. barbatus/P. rugosus* species complex. The genotype of an individual (homozygote vs. heterozygote) could affect larval growth rates, leading to different developmental trajectories of larvae (Schwander et al. 2007b). Such an effect could depend on many loci and on interaction between nuclear–nuclear and/or nuclear–cytoplasmic genes. Alternatively, there may be a simple switch encoded by a single gene or by a few genes that act as major regulatory genes similar to genes in the sex determination pathway. GCD can impact many biological and social traits such as survival/mortality of individuals and colonies, and colony-level sex ratios (Schwander et al. 2006, 2007b; Anderson et al. 2009). It can also cause conflicts within societies as certain genotypes gain a fitness advantage by becoming queens. The *Pogonomyrmex* lineages with GCD are related to and evidently derived from *P. barbatus* and *P. rugosus*, which are characterized by ECD even though the female caste in *P. rugosus* seems to be influenced by the genetic interaction between the parents (Schwander and Keller 2008) and by maternal effects (Schwander et al. 2008a). It has been suggested that hybridization between these species was at some time involved in generating the GCD lineages. However, it is still unclear whether hybridization was the causal mechanism or a secondary event that led to the introgression of GCD into new populations and species after it had evolved within one species (Anderson et al. 2006a; Schwander et al. 2007a).

Pogonomyrmex rugosus and *P. barbatus* are closely related species in the *P. barbatus* group (Parker and Rissing 2002) that live in hot, arid deserts in the southwestern United States and Mexico and occur sympatrically in some areas, for example, southeastern Arizona (Johnson 1998, 2000). Since the first morphological and molecular studies on sympatric populations of the two species, signs of hybrid lineages in the form of exceptional color morphs and deviations from Hardy–Weinberg equilibrium have been detected (Julian et al. 2002; Volny and Gordon 2002). Further studies indicated that there are at least two pairs, perhaps four or more pairs, of dependent lineages that morphologically resemble either *P. rugosus* or *P. barbatus* (Helms Cahan and Keller 2003; Anderson et al. 2006a; Schwander et al. 2007a) but are reproductively isolated from them. Mitochondrial DNA sequences, together with a few nuclear markers, indicated that the nuclear genomes of the dependent lineages are mosaics of *P. barbatus* and *P. rugosus*, suggesting that the lineages are of a hybrid origin (Helms Cahan and Keller 2003). The lineage pairs are dependent on each other and hence they cannot exist as pure lineage populations. Mating between individuals belonging to two different lines that are morphologically either of the *P. rugosus* type (e.g., lines H1/H2) or the *P. barbatus* type (e.g., lines J1/J2) is needed to produce workers and thus ensure successful colony founding. In contrast, within-lineage mating is necessary for production of virgin queens. Although historical gene flow is evident, no present gene flow has been reported between lineages, suggesting genetic isolation of the lineages from each other and from the two parental species (Helms Cahan and Keller 2003, but see Anderson et al. 2006a).

There are still many unanswered questions about the GCD system of *Pogonomyrmex*. These concern the origin of GCD and the first hybrid lineages, whether these two events are causally linked or independent, the maintenance of GCD, the origin of new dependent lineages (Schwander et al. 2007a,b), and the number and expression pattern of the putative caste-determining genes. Three genetic models have been presented to explain the origin of the GCD system in *Pogonomyrmex*:

(1) The single-factor model (Volny and Gordon 2002; Anderson et al. 2006a) assumes one major gene that determines the female caste by zygosity. Homozygotes will develop into queens and heterozygotes into workers. Anderson et al. (2006a) suggested, on the basis of mtDNA divergence patterns, that GCD evolved first within *P. barbatus* and spread to other phylogenetic lineages by introgression.

(2) The nuclear epistatic model of Helms Cahan and Keller (2003) suggests a different genetic background and a different origin of the interacting dependent lineages, for example, J1/J2. According to this model, dependent lineages were formed by hybridization of two species (genotypes *aabb* and *AABB* at the loci A and B). Inbreeding among the hybrid offspring (*AaBb* females) could later result in lines with the genotypes *aaBB* and *AAbb*. Assuming that the double heterozygotes *AaBb* develop into workers and the homozygotes *aaBB* and *AAbb* into queens, the genotypes behave as two dependent lineages. The model assumes that the double heterozygotes do not generally reproduce but that they are (or at least have earlier been) bipotential in the sense that the early hybrids must have been reproductive to produce the new genetic lines. If a dependent line hybridizes with the parental species, e.g., *aabb* × *Ab*, and the *Aabb* offspring are reproductive, gene flow (both nuclear and mitochondrial) may exist between the dependent line

and the parental species unless specific mechanisms have evolved to prevent this.

(3) The model with cyto-nuclear epistasis (Linksvayer et al. 2006) differs from the above epistatic model by assuming that one of the involved factors is located in the mitochondrial genome (types *m* and *M*). The parental types *aam* and *AAM* should be modified in such a way that the dependent lines $a'a'm'$ and $A'A'M'$ are separate from the parental species and the hybrids $A'a'm'$ and $A'a'M'$ develop into workers. When one of the interacting factors is mitochondrial, the mitochondrial haplotype will stay associated with a given dependent lineage.

The origin and history of the dependent lineages have so far been mainly inferred from the phylogenetic tree of mitochondrial haplotypes, whereas the history of the nuclear genomes has been evaluated with very few genetic markers (Anderson et al. 2006a; Schwander et al. 2007a). Our present study focuses on the origin and evolution of the hybrid lineages by using a large nuclear marker dataset combined with mitochondrial sequences. We used a set of 751 variable AFLP markers that are scattered throughout the genome. This dataset allowed a more detailed inference of the genetic composition of the GCD lineages and a reevaluation of the origin of hybrid lineages. We also constructed a linkage map based on an ECD *P. rugosus* family using the same AFLP markers. The linkage study had two aims. First, we estimated the genome-wide recombination rate in *P. rugosus* to check whether it confirms the pattern of high recombination rates detected in other highly social insects (Sirviö et al. 2006). Second, the linkage map allowed us to evaluate whether the genomes of the hybridizing lineages are porous in the sense that introgression has been possible in some parts of their genomes but restricted in others. Such porosity, resulting from historical hybridization, has been detected in the genomes of related species even when no recent hybrids or individuals with hybrid phenotypes are known (Kane et al. 2009).

Material and Methods

SAMPLES

The linkage map was based on 92 haploid males that were collected from one ECD *P. rugosus* colony (RA Johnson #3111) located 1.1 miles southwest of the junction of Interstate 10 and McCartney Road, Pinal County, Arizona. Population samples were collected from the same populations and sometimes the same colonies that were analyzed by Anderson et al. (2006a) and in which GCD has been documented. Our samples came from populations 2, 6, 8, and 9 in Arizona and 10 and 11 in New Mexico (population numbers according to Fig. 1 in Anderson et al. 2006a). We analyzed one worker and one alate queen from each of 44 colonies. In addition, workers and alate queens from one *P. rugosus* hybrid line colony (RA Johnson #483) were used to prescreen for suitable primer pairs that show consistent marker differences between the two castes. Vouchers are deposited in the collection of Robert A. Johnson, Tempe, Arizona.

GENOTYPING

The AFLP Core Reagent Kit of Invitrogen Life Technologies (Carlsbad, CA) was used according to the manufacturer's protocol for the haploid males and nine diploid individuals (five workers and four queens from colony 483). However, in our study we used only one-half of the proposed reaction volumes. We applied 15 ng/μl of genomic DNA to the enzyme restriction-adaptor ligation step. In total, 39 different combinations of selective EcoRI- and MseI- primers were used to search for polymorphic markers for mapping purposes and to screen for caste-specific markers. Samples were amplified and run in acrylamide gels as described by Sirviö et al. (2006). Segregating markers for males were scored as 1 = presence allele or 0 = absence allele. Fragment length polymorphisms were scored by marking 1 = longer allele and 0 = shorter allele. Bands that were not clear were rerun or marked as missing. Films were scored twice. Mapmaker version 3.0 (Lander et al. 1987; Lincoln et al. 1992) was used for building linkage groups for the variable markers from haploid *P. rugosus* males according to the phase unknown procedure described by Sirviö et al. (2006). Recombination fractions were translated in centiMorgans using Kosambi (1944) mapping function. Marker linkage groups were formed with the criteria LOD 3.0 and theta 0.35.

Prescreening of individuals from colony 483 resulted in 17 selective primer pairs that produced markers with systematic caste specific differences. All population samples were prepared using the AFLP Core Kit for small plant genome (Applied Biosystems, Carlsbad, CA) according to the manufacturer's protocol and they were run on an ABI 3730 sequencer. AFLP data were analyzed by GeneMapper version 3.7 and the validity of all peaks was evaluated by manual inspection. Again, a presence phenotype was scored as 1 and absence as 0. AFLP differences among individuals were calculated by transforming the 0/1 data into a dinucleotide A/G data and by using the sequence analysis programs in Mega 3.1 (Kumar et al. 2004). Principal coordinate analysis (Anderson 2003) was used to cluster individuals on the basis of the pairwise differences shown by the nuclear AFLP markers. Marker differences and identities between each pair of lines were scored in each linkage group and possible clustering of differences or identities in the linkage groups was tested by 1000 permutations and using the goodness-of-fit criteria.

Population samples (one worker and one queen per colony) were sequenced with the Ready Reaction Kit of Applied Biosystems for mitochondrial *cox1*-gene using primers LCO and HCO resulting in ~650 bp sequences congruent with sequences used

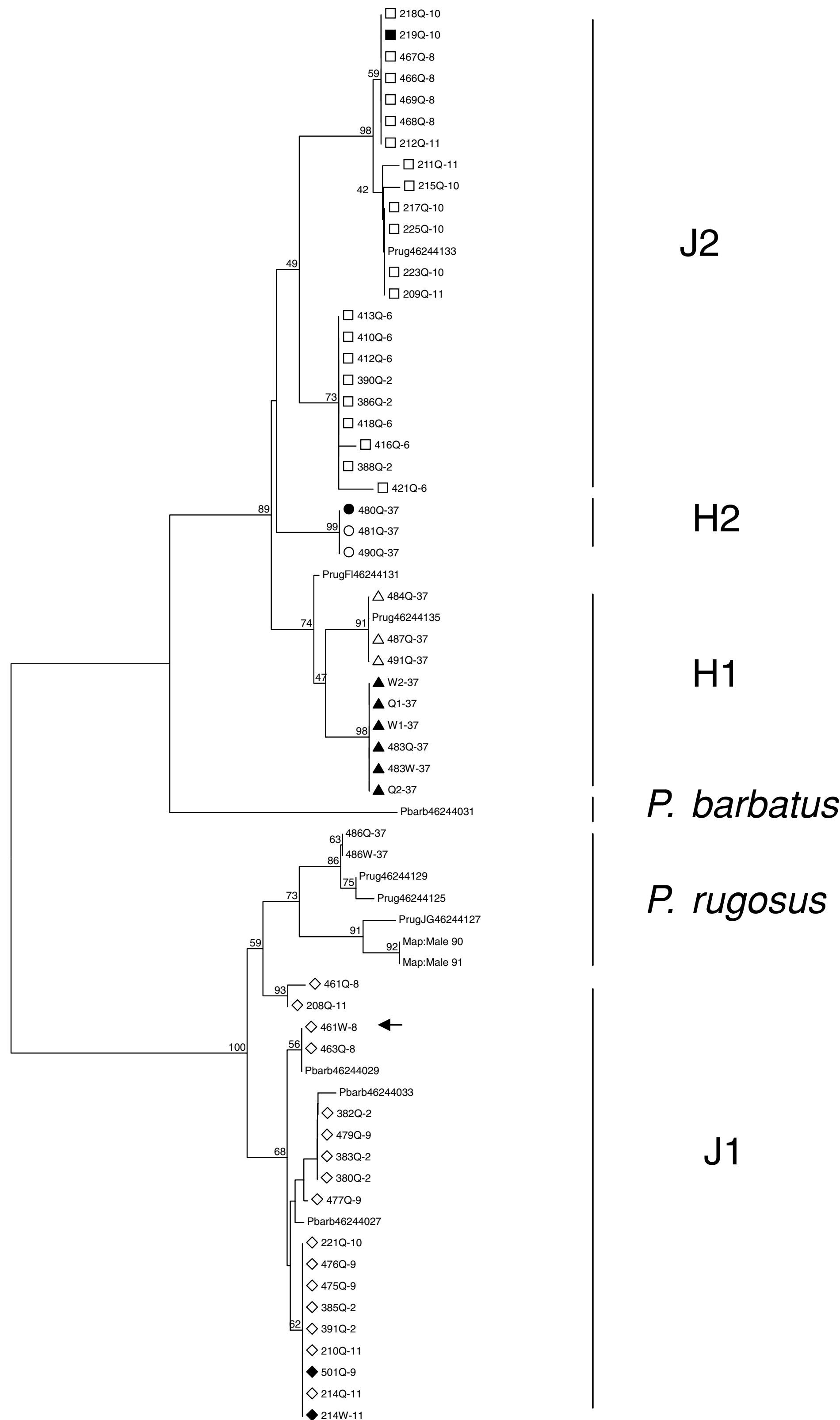

Figure 1. Neighbor-joining tree of the mitochondrial *cox1* sequences (bootstrap tested with 1000 replicates). Distances were calculated according to Kimura's two-parameter method (sites with missing data or alignment gaps were deleted). Support values below 20% are not shown. Solid symbols indicate individuals whose nuclear genotypes depart from that expected on the basis of their caste. The individuals are marked with the colony number, caste (W or Q) and the population number. Sequences from Genbank are marked with the accession number.

by Anderson et al. (2006a). Products were run in a ABI 3730 capillary sequencer and the sequences were aligned and edited with Sequencher version 4.7. Mega version 3.1 was used to estimate nucleotide differences and to build neighbor-joining trees. The trees were tested using an interior branch test (1000 replicates).

SIMULATIONS

The origin of the observed genetic lineages was evaluated by comparing the observed genetic differences to those obtained by simulating various evolutionary scenarios. In the simulations we constructed an ancestral type by giving a value of 0 or 1 at 1500 loci (corresponding to the order of magnitude of our AFLP data). We let this ancestral type (time point t_0) evolve in two lines (one leading to *P. barbatus* and the other to *P. rugosus*). At the time point t_1, two new lines (J1 and J2) were formed as mixtures of the two parental species (Fig. 6B). Later at point t_2, two new lines (H1 and H2) were formed by one of the following three schemes: as a mixture between *P. rugosus* and *P. barbatus*, between J1 and J2, or between *P. rugosus* and the line J2. The alternative of having J2 as one contributing parent is attractive for the reason that the mitochondrial haplotypes of H1, H2, and J2 share a common ancestor. The time periods from t_0 to t_1, from t_1 to t_2, and from t_2 to t_3 (present) were assumed to be proportional to 3:2:2 on the basis of the mtDNA phylogeny. The three schemes thus differed by having different sources for the H lines. Mutations were generated stochastically and hybrids were formed as alternative mixtures of the parental lines. Simulations were run for 300 times, and we evaluated the mixing proportions by checking how well the final genotypic differences matched with the real AFLP data.

Results

MITOCHONDRIAL MARKERS

We sequenced the mitochondrial gene *coxI* (555 bp) from all individuals so that we could assign each individual to a lineage or species based on its mitotype. We also included sequences of the putative parental species *P. barbatus* and *P. rugosus* from GenBank. The neighbor-joining tree based on pairwise nucleotide differences (Kimura's two-parameter model) clustered the mtDNA sequences in groups that corresponded to the lines J1, J2, H1, H2, *P. rugosus* and *P. barbatus* (Fig. 1) according to Anderson et al. (2006a). The only disagreement was that sequences from colonies 208 and 461 clustered loosely with *P. rugosus* but the colonies could otherwise be considered as J1. The sequences from GenBank originated apparently from samples that had only been identified morphologically as our results suggest that some of them represented dependent lineages (Fig. 1). Individuals from colony 486 (population 37) grouped with *P. rugosus* and they are considered to represent the parental species *P. rugosus* with ECD.

The alate queens and workers from the same colony had an identical haplotype with one exception; the worker and the gyne from colony 461 (population 8) had different haplotypes, both belonging to lineage J1 but differing by seven nucleotides. This could be interpreted at least in three ways. The mother of these two individuals might have been heteroplasmic, the colony had multiple queens, or the worker had been brood-raided from a neighboring colony.

The majority of our colonies were morphologically similar to *P. barbatus* and, as expected, they had mitochondrial types belonging to either the J1 (16 colonies) or J2 lineage (22 colonies). The haplotypes of J2 showed some geographical differentiation with one major haplotype clade in populations 2 and 6, and another in the populations 8, 10, and 11 (Fig. 1). All colonies belonging to the H lineages were from a single population, four of type H1 and three of H2. The lineage H2 was represented by a single haplotype in our samples, whereas H1 had two haplotypes differing by five nucleotide substitutions. Nucleotide diversity (mtDNA) within a lineage was highest in J2 (0.008 or on average 4.4 differences), being almost twice as high as in J1 and H1 (Table 1). When we add the sequence data from previous studies (Helms Cahan and Keller 2003; Schwander et al. 2007a), the comparable sequence becomes shorter (302 bp instead of our 555 bp) and the level of nucleotide diversity among the lineages ranges from 0.006 to 0.010 (Table 1).

The greatest sequence difference was between J1 and J2 (corrected average pairwise difference, $n = 35.7$ nucleotides) whereas differences between J2, H2, and H1 were smaller (7–8 differences). The average distance calculated according to Kimura's two-parameter method between the mitochondrial lineages in our data (555-bp-long sequence) ranged from 0.018 between J2 and H2 to 0.067 and 0.076 between J1 and H2 and J1 and J2, respectively (Table 2). The result was similar when we included shorter sequences from previous studies (data not shown).

Sequence divergence between the two major clusters of the mtDNA phylogeny, one with *P. rugosus* and J1, the other with *P. barbatus,* J2, H1, and H2, was on average $d = 0.075$. This is of the same magnitude as the divergence between *P. rugosus* and *P. barbatus* obtained earlier (Helms Cahan and Keller 2003; Anderson et al. 2006a). Removing the sequence variation within the parental species resulted in a net divergence of about 0.072 between the two main clusters. The net divergence between J1 and *P. rugosus* was 0.02 and between *P. barbatus* and the cluster of J2, H1, and H2 on average 0.049 (0.048–0.050). The divergence among the lines J2, H1, and H2 was 0.022.

As a large number of short sequences (302 bp, Table 1) of the *coxI* region exist from earlier studies (Helms Cahan and Keller 2003; Schwander et al. 2007a), we combined them with our data to make a comprehensive phylogeny (Fig. S1). The main findings remain the same when using this extended dataset.

Table 1. Genetic diversity within the lineages given as nucleotide diversity (π) for the mitochondrial sequences (*cox1*) and as the proportion (*p*) of differences for the AFLP data. The number of individuals is given within parentheses. Nucleotide diversities are estimated from our own data (555 bp) and from a combined dataset (302 bp) that also includes the data of Schwander et al. (2007a) and Helms Cahan and Keller (2003).

Lineage	mtDNA (555 bp)		mtDNA (302 bp)		AFLP differences
	π	haplotypes	π	haplotypes	*p*
J1	0.004 (16)	6	0.010 (23)	8	0.060 (14)
J2	0.008 (22)	7	0.007 (28)	7	0.061 (21)
H1	0.005 (4)	2	0.006 (7)	3	0.066 (3)
H2	0 (3)	1	0.009 (6)	3	0.050 (2)
Prug	0.011 (5)		0.013 (20)	11	0.032 (2)
Pbar	na	na	0.012 (9)	4	na

NUCLEAR MARKERS

Seventeen selective AFLP primer pairs produced 1147 scorable markers, of which 751 showed variation within or among lineages in our population samples.

Clustering of the gynes based on the nuclear loci agreed well with that presented earlier by Schwander et al. (2007a), separating J1, H1, and *P. rugosus* in one cluster and J2 and H2 in the other (Fig. 2). However, because the dependent lineages may have hybrid background, a distance-based tree does not necessarily well reveal the historical relationships between the lineages. For this reason we used principal co-ordinate analysis (PCA) to cluster individuals on the basis of their pairwise differences. PCA classified the individuals into seven main clusters (Fig. 3) and the first two axes explained 80% of the variation. Four of the groups were mainly formed by gynes representing the maternal lineages J1, J2, H1, and H2. Two clusters were located between the two pairs of dependent lineages and consisted mainly of workers. The seventh group was formed by *P. rugosus* from nest 486, including both the gyne and the worker from that nest. Some exceptional individuals were found in all studied lineages, including the gynes 219Q, 501Q, 480Q, 483Q, and the worker 214W. All the exceptional gynes were clearly of hybrid origin and clustered with workers, that is, some larvae from interlineage crosses are able to develop into gynes. The mitochondrial sequence of 219Q and 501Q belonged to the lineages J2 and J1 and 480Q to H2 (Fig. 1).

The gynes and workers from the nest 483 belonged to the maternal lineage H1 (there were several specimens that were studied because they were used in prescreening for selective primer pairs in PCR). However, nuclear DNA placed the gyne 483Q into the proximity of workers that were hybrids between the dependent lineages J1 and J2 (Fig. 3). The worker 483W shared the maternal lineage with 483Q but clustered with the other workers that were H1/H2 hybrids (Figs. 1 and 3). Finally, the worker 214W represented the J1 mitochondrial lineage, but in contrast to other workers, its nuclear genotype placed it among queens of the J1 lineage.

The average number of marker differences between individuals within a lineage (i.e., counting only gynes that clustered in the corresponding groups in PCA) varied from 58 in H2 to 75 in H1 (Table 3). Many pairwise divergence estimates calculated from the AFLP data were within a narrow range of values (0.29–0.34), but the divergence among *P. rugosus* (nest 486), J1, and H1 (0.17–0.20) and between H2 and J2 (0.18) was clearly smaller than the others. The highest pairwise differences were between J1 and J2 (0.34 or on average 394 differences).

Spatial genetic differentiation (excluding the hybrid individuals) could be examined in lines J1 and J2 that had samples from five populations each (Table 4). The populations 8, 9, and 11 are very similar to each other (no J2 from population 9). Interestingly, population 10 was similar to these populations in line J2 but not in J1. Population 2 is geographically most distant (~500 km) from the other populations and is also genetically quite distinct. Finally, population 6 is geographically distant (300–500 km) to other populations and it turns out to be also genetically isolated from all the other J2 populations. Our results show that 9% (four out of 44) of the dependent line gynes were most likely first-generation hybrids, suggesting that gene flow between the lines might take place if these females would successfully

Table 2. Genetic distances between the lineages calculated as means of individual pairwise distances. Above diagonal: nucleotide differences (Kimura's two-parameter distance) from the mtDNA sequences (514 bp). Below diagonal: average *p*-distance for the 1163 AFLP markers. The apparent hybrid queens were not used in AFLP analysis. AFLP data was not available for *P. barbatus*.

	J1	J2	H1	H2	Prug	Pbar
J1	–	0.076	0.068	0.067	0.020	0.079
J2	0.344	–	0.022	0.018	0.083	0.047
H1	0.206	0.308	–	0.016	0.078	0.049
H2	0.328	0.183	0.290	–	0.077	0.049
Prug	0.166	0.322	0.191	0.311	–	0.088

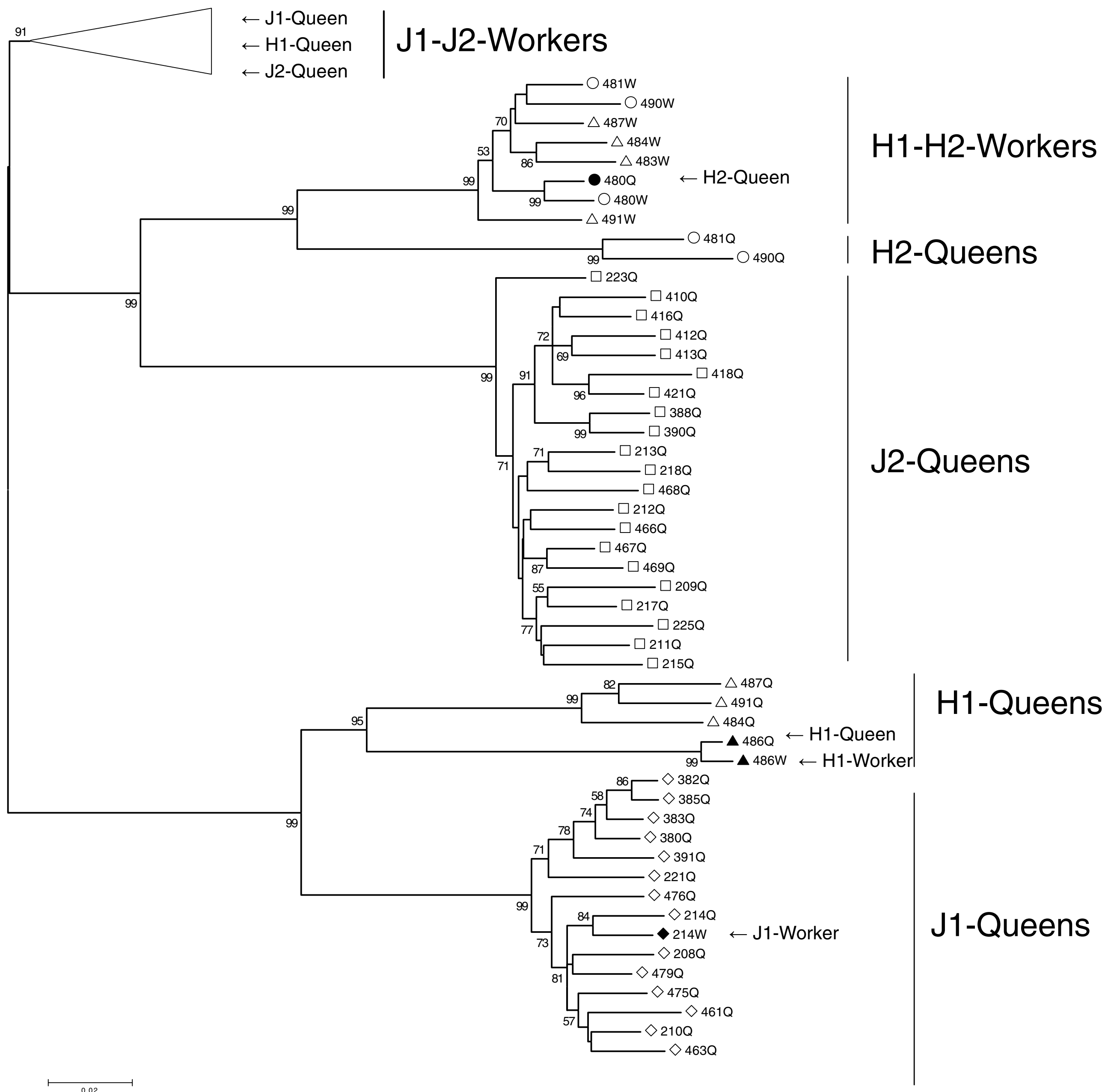

Figure 2. Neighbor-joining tree of the AFLP distances (*p* distance) with 1000 replicates of interior branch testing. Individuals whose clustering departs from that expected on the basis of their caste are marked with solid symbols. Support values below 50 are not shown.

found new colonies. Yet, the genetic divergence between J1 and J2 was not different when comparing either sympatric or allopatric nests.

LINKAGE MAP AND RECOMBINATION (*P. RUGOSUS*)

We generated a linkage map based on haploid males from a single *P. rugosus* (ECD) colony using the same AFLP markers that were used for the population study. To confirm that our mapping population is indeed *P. rugosus* (ECD), we sequenced the mitochondrial *cox1* of two males and compared them to the *cox1* sequences in our tree. The NJ-tree of mtDNA (Fig. 1) clustered these males in the same branch with nest 486 and the *P. rugosus* (ECD) sequences retrieved from GenBank.

By using 39 selective primer pairs, we obtained 258 polymorphic AFLP markers for the *P. rugosus* males produced by a single queen. Of these, 215 markers were mapped into 37 linkage groups (Fig. 4). Linkage groups 2, 3, 4, 5, 6, 7, 10, 11, 12, and 14 were formed by combining two or three smaller groups together with the Near-command in Mapmaker. The map is not saturated as the haploid chromosome number for *P. rugosus* is 16. The fraction of markers that showed length polymorphism was ~22%. The physical size of the genome was estimated by flow

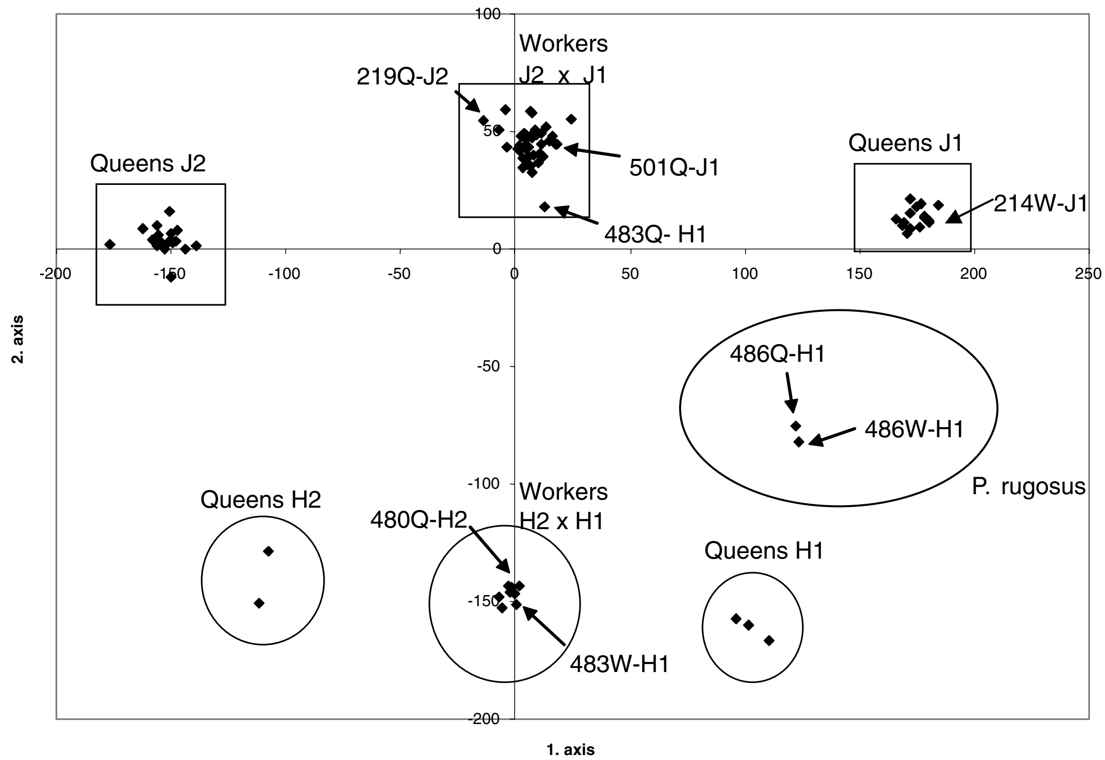

Figure 3. **Principal coordinate analysis on the nuclear AFLP distances between individuals. The circles delineate individuals belonging to the same lineage and caste. The individuals whose clustering departs from the expected are marked separately.**

cytometry to be 255 Mb (see Sirviö et al. 2006). After error detection option of the Mapmaker, recombinational size of the genome was estimated to be 2823 cM. Finally, we added 735 cM (21× 35 cM) to cover additional gaps between linkage groups and obtained the total estimate of 3558 cM for the mapping size of the genome. This leads to a recombination rate estimate of 71.7 kb/cM or 14.0 cM/Mb for *P. rugosus* (ECD).

Table 3. **Mean pairwise lineage differences for AFLP data. N_G is the number of gynes and N_P the number of AFLP markers that were polymorphic within a lineage. The diagonal of the differentiation matrix is the mean number of pairwise differences within a line (P_X, in bold). Above diagonal are the mean number of pairwise differences between population (P_{XY}) and below diagonal are corrected average pairwise differences between populations ($P_{XY} - (P_X + P_Y)/2$). The value within *P. rugosus* is from the comparison between worker and queen from the nest 486.**

	N_G	N_P	Pairwise AFLP differences				
			J1	J2	H1	H2	Prug
J1	14	181	***68***	394	234	377	192
J2	21	203	325	***70***	351	211	372
H1	3	110	162	279	***75***	333	218
H2	2	58	314	147	267	***58***	362
Prug	2	37	139	318	162	314	***37***

GENOMIC DISTRIBUTION OF MARKERS WITH CASTE OR LINEAGE DIFFERENCES

Possible association of AFLP markers with caste and lineage was estimated in both family data and population data. The rationale behind these analyses is that the level of linkage disequilibrium between putative genetic caste-determining factors and AFLP markers could vary if there has been gene flow between the lineages, and differences between the lineages could reflect the contribution of the two parental genomes (*P. barbatus* and *P. rugosus*) in the formation of the dependent lines.

AFLP markers showed consistent differences between both J1 and J2 and H1 and H2, and 165 markers were statistically significantly differentiated. Of these, 19 markers were among those we mapped and they were placed in 15 linkage groups (square marks in Fig. 4). We also estimated the allele frequencies among the gynes of the lineages J1 and J2, in which we had most samples, by assuming Hardy–Weinberg frequencies. Fifteen loci had $F_{ST} > 0.9$ (calculated as $1 - H_S/H_T$) and these were located in 13 different linkage groups. Even though the data violated the assumptions underlying the use of Hardy–Weinberg frequencies, the result shows that genetic differences between the lineages are scattered within the genome.

We also tested whether the marker differences or identities between the lineages (*P. rugosus*, J1, J2, H1, and H2) were clustered. Such clustering could indicate that the lines have received

Table 4. AFLP-differences between populations within (A) the line J1 and (B) the line J2. *N* is the number of gynes, *P* is the detected number of polymorphic loci within the population. The differences between the populations are the mean numbers of AFLP differences in pairwise comparisons (above diagonal) and net pairwise differences (below diagonal). The diagonal gives the mean differences in pairwise comparisons within the population.

(A)	N	P	Differences between J1 populations				
			Pop2	Pop8	Pop9	Pop10	Pop11
Pop2	5	98	**42**	76	75	66	70
Pop8	2	57	26	**57**	61	62	54
Pop9	3	79	31	9	**47**	56	46
Pop10	1	na	22	5	9	**na**	52
Pop11	2	44	27	3	0	8	**44**
(B)	N	P	Differences between J2 populations				
			Pop2	Pop6	Pop8	Pop10	Pop11
Pop2	2	36	**36**	62	77	80	81
Pop6	6	122	16	**57**	72	79	79
Pop8	4	94	32	17	**54**	61	62
Pop10	5	137	29	18	2	**65**	63
Pop11	4	115	33	20	4	0	**61**

different/identical segments from the parental genomes. Forty-six markers that showed some variation among the lineages were among those we mapped. Marker differences and identities between each pair of lines were scored in each linkage group. The permutation test (1000 permutations) showed that clustering did not depart from random (*P*-values from 0.15 to 0.80 for different pairs of lineages).

Discussion

RECOMBINATION

The rate of recombination within the parental species *P. rugosus* is extremely high. Our genetic map of *P. rugosus* (Fig. 4) is only the second published map for ants (Sirviö et al 2006). The physical size 255 Mb (Sirviö et al. 2006) is close to the values obtained for three other *Pogonomyrmex* species (249.5–282.9 Mb) and below the mean of the subfamily Formicinae, or of all ants (Tsutsui et al. 2008). The map was not saturated and the number of linkage groups exceeded the haploid chromosome number. The estimated recombination rate of 14 cM/Mb is the second-highest reported in insects, or any animal, and very close to the value in the honey bee (19 cM/Mb, Beye et al. 2006). As the map was not saturated, we added 735 cM to the genome size to cover the gaps. Even if this 735 cM is not added, the recombination rate (11.1 cM/Mb) would still be higher than reported for any other insect except the honey bee. The current recombination rate estimate for *Pogonomyrmex* is more than twice that estimated earlier for the ant *Acromyrmex echinatior* (6.2 cM/Mb, Sirviö et al. 2006) which is also a high value for an insect. Even though the ants do not provide phylogenetically independent datapoints, the two genera are not closely related and their respective lines may have separated some 80 million years ago (Brady et al. 2006; Moreau et al. 2006). We can therefore conclude that a high recombination rate characterizes eusocial insects. The proximate and ultimate factors behind this trend are unclear. It is possible that the recombination rate has increased because it has positive effects and has been selected for (e.g., due to an increased intracolonial genotypic variability), or that the effects are neutral or even slightly harmful, but the recombination rate has drifted up stochastically. Because the effective population sizes of advanced social insects have probably been small, slightly harmful traits can spread in their populations as if neutral (see Beye et al. 2006; Sirviö et al. 2006 and Wilfert et al. 2007 for further discussion).

INTROGRESSION AND CASTE DETERMINATION

Our study covered a significant proportion of the nuclear genome and indicated that a major fraction of the nuclear genome showed fixed differences between the dependent lineage pairs. These widespread genetic differences between lineages explain why earlier studies have readily detected caste differences using less than a handful of nuclear markers (Helms Cahan and Keller 2003; Anderson et al. 2006a,b; Schwander et al. 2007a).

If gene flow would take place between a pair of dependent lineages, linkage disequilibrium between a putative caste determination locus and a neutral unlinked locus would break down because (a) heterozygous females occasionally develop into queens, or (b) heterozygous workers produce males. We examined the breakdown of linkage disequilibrium (D) assuming that it starts with an initial maximal value of D_0. In case (a), a female heterozygous at the caste-determining locus develops into a queen at a small probability, and the fraction of hybrids among all the queens is x. Assuming that the two dependent lines are equally frequent and mating takes place at random, the rate of change of linkage disequilibrium is approximately $\Delta D = -{}^2/_3 xD$ for small x, and the level of linkage disequilibrium at time t will reduce to

$$D_t/D_0 = {}^1/_4\, e^{-2xt/(3+2x)}. \quad (1)$$

In the case (b), workers (which are always heterozygotes at the caste-determining locus) produce a small fraction x of all males in the population. In this case, $\Delta D = -{}^1/_3 xD$ for small x and

$$D_t/D_0 = {}^1/_4\, e^{-xt/3}. \quad (2)$$

The two models show that even a small frequency of introgression between two lines would lead to a substantial decay

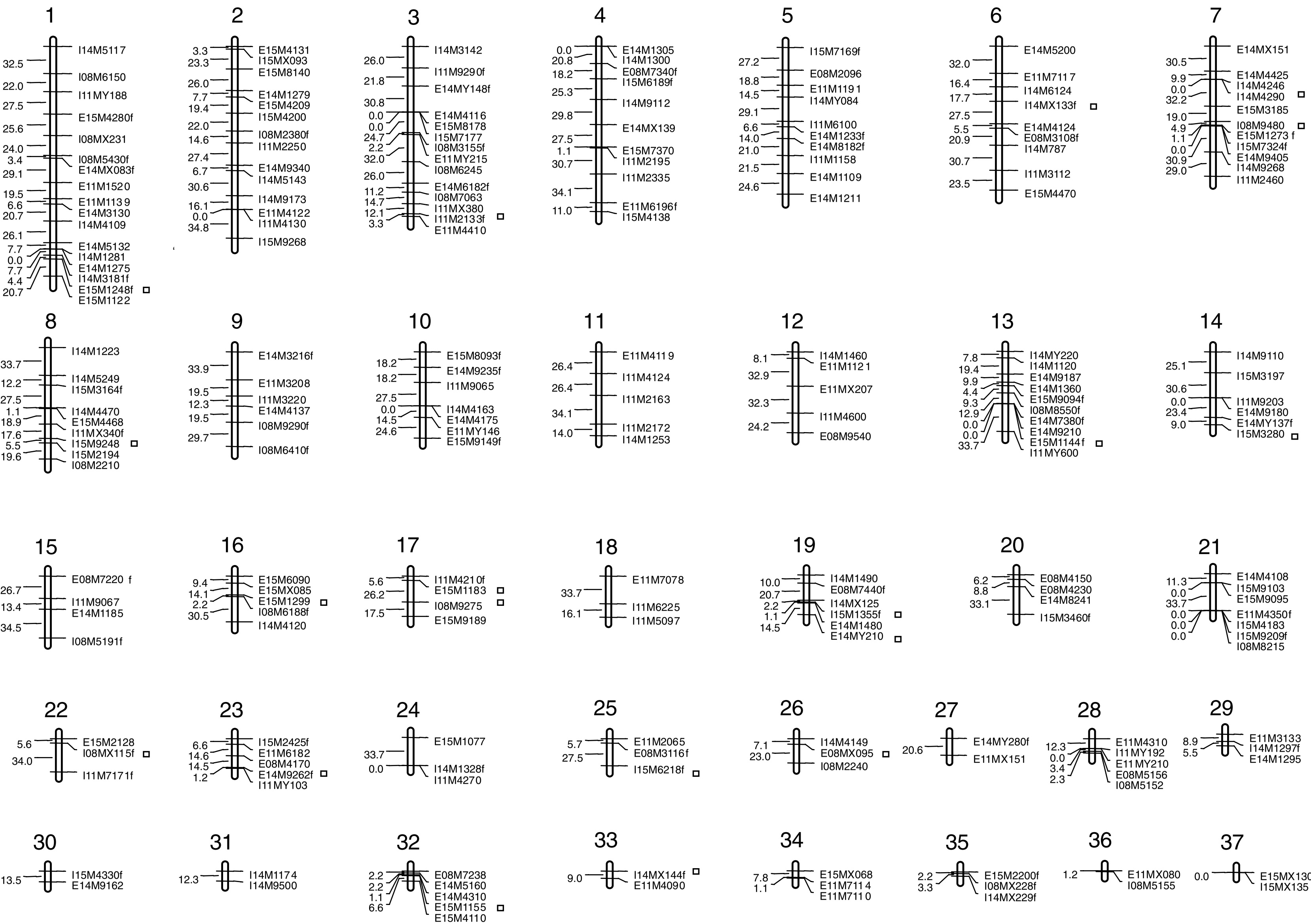

Figure 4. Marker linkage groups constructed for haploid *P. rugosus* males produced by single queen. The name of fragment length polymorphism marker ends with the letter f. Location of candidate GCD markers are labeled with squares.

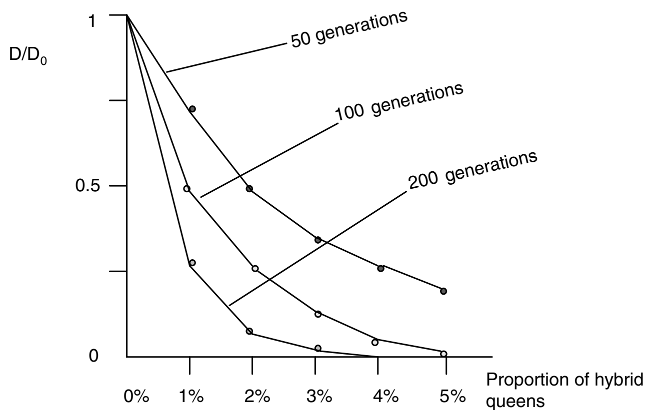

Figure 5. Break-down on linkage disequilibrium (eq. 1) as the proportion of hybrid queens out of all the queens increases. The curves represent the situation after 50, 100, and 200 generations. The curves for worker-produced males (eq. 2) are very similar.

of linkage disequilibrium between unlinked genes given enough time (Fig. 5). Mapping the *Pogonomyrmex* genome showed that the loci that showed significant differences between the lineages were scattered over many different linkage groups. This indicates that there has been no recent gene flow or recombination between the different genetic lineages. For that reason, any difference between the complementary dependent genomes appears as a systematic difference between the worker and queen castes within a lineage (Schwander et al. 2007a,b).

Nine percent of the gynes in the GCD colonies were hybrids (4 out of 44). This is similar to the fraction detected earlier, 7.6% by Schwander et al. (2007a). These observations suggest that the determination of caste involves a stochastic (or environmental) component and females with a similar genetic background can develop as either workers or gynes. Alternatively, there may be rare genetic variation influencing the caste determination within a lineage, and the unusual genotypes reflect genetic segregation within crosses. Such "errors" in caste determination can go in both directions as we also found a worker that apparently resulted from mating within a line (line J1). Even though Schwander et al. (2007b) found hybrid females among the newborn gynes, they did not find old reproductive hybrid queens. This can also be indirectly concluded from the tight clustering of castes in the PCA (Fig. 3). If individuals would have different degrees of hybrid history, we would expect a greater scatter of points in the PCA figure. Our results therefore support the view that there has been no recent introgression between the lineages. Male production by workers would also lead to recombination between the dependent lines if the hybrid workers would produce gametes. Worker reproduction is not uncommon in ants (Crozier and Pamilo 1996) but Suni et al. (2007) found no worker-produced males in *P. barbatus*, and *P. badius* is the only *Pogonomyrmex* species known to have reproductive workers (Smith et al. 2007).

Workers resulting from intralineage mating and queens from interlineage crossing have been detected earlier in these taxa (Helms Cahan and Keller 2003; Helms Cahan et al. 2004; Anderson et al. 2006a,b; Schwander et al. 2007b) but interlineage queens between J1 and J2 have been considered particularly rare (Schwander et al. 2007b). This has been interpreted as a loss of phenotypic plasticity in the putatively old lines. Our samples had hybrid queens within each of the four lineages. In addition, we detected one J1 worker that is of pure-lineage ancestry. The fitness of the hybrid queens has been suggested to be low because hybrid mother queens were not found in natural colonies (Helms Cahan and Keller 2003) and because their fertility has been shown to be low in laboratory colonies (Schwander et al. 2007b). The latter study reported, however, a small fraction of daughter queens that could have arisen from backcrossing in natural conditions. The frequency of such backcrossed queens was 0.5%, a value which would, in the long run, lead to a complete mixing of the genomes (see Fig. 5). This raises questions concerning the gene flow between the lineages. If introgression takes place, it should be seen in the nuclear genetic distances, in the association between the mitochondrial haplotype and the lineage, and in the association between putative caste-determining loci and other nuclear markers. It is known that the genomes of hybridizing species can be porous in the sense that introgression takes place in some parts of the genome but is strongly restricted in others. Even species that have no phenotypically apparent hybrids and no recent hybrids can show historical introgression in many parts of their genomes (Kane et al. 2009). Generally, introgression is restricted in genomic areas associated with strongly selected loci and in areas with low recombination rates. In *Pogonomyrmex*-dependent lineages, the genes affecting caste determination act as barriers to gene flow if interlineage hybrids develop as workers. If the dependent lineages have a high recombination rate, as observed in *P. rugosus*, linkage disequilibria decay rapidly and introgression could take place in those genomic areas that are not associated with caste determination or other strongly selected loci. Our results showed large nuclear distances between the pairs of dependent lineages in many linkage groups. This clearly points to the lack of recent gene flow. If hybrid queens can establish colonies, as shown by Schwander et al. (2007b), it seems likely that selection among their offspring is strong and leads to hybrid break-down, thus effectively preventing introgression.

DIFFERENTIATION OF THE DEPENDENT LINEAGES

Hymenopteran females can produce nonhybrid haploid sons even if they have mated with a male from another species. In several cases of hybridization in ants, diploid hybrid offspring develop into workers and the reproductive success of the queen relies on the production of haploid sons (Seifert 1999; Umphrey 2006). Despite obvious disadvantages of hybridization, viable and fertile

hybrid females that have fitness advantages over either parental species can sometimes be generated.

The lineages in the *Pogonomyrmex* hybrid zone have been initially divided into two pairs of dependent lines, J1/J2 and H1/H2 that can be distinguished on the basis of both their mitochondrial haplotypes and segregation of nuclear markers (Helms Cahan and Keller 2003). Recently, Schwander et al. (2007a) suggested that the pair H1 and H2 can be further divided into three pairs of dependent lines (the pairs F, G, and H), which are recognized on the basis of nuclear microsatellites but are only partially supported by mitochondrial haplotypes. All the H1 and H2 queens in our data formed two tight clusters on the basis of their nuclear genotypes (Figs. 2 and 3) suggesting no clear divergence within the groups even though the mtDNA data included haplotypes identical to lineages F1, G1, and H1 (as presented by Schwander et al. 2007a and Helms Cahan and Keller 2003). Furthermore, our H2 haplotypes clustered with four F2 and one G2 haplotypes (Schwander et al. 2007a). There are two possible explanations for the apparent discrepancy between the nuclear differences among these lineages in the microsatellite genotypes (Schwander et al. 2007a) and the lack of such differences in the present AFLP data. First, it is possible that our restricted samples included only H lineages and the apparent clustering of mtDNA haplotypes results from incomplete lineage sorting. Second, it is likely that the formation of new lineage pairs has included population bottlenecks that have led to allele frequency changes and elevated genetic distances at the microsatellite loci, whereas the large number of AFLP markers better reflects the nuclear sequence differences between the lines.

The origin of the dependent lineages is controversial as the mitochondrial and nuclear genomes show different clustering (Fig. 6A). Furthermore, the J line females are morphologically like *P. barbatus* and the H line females like *P. rugosus*, resemblances that depart from the overall nuclear affinities of the lines (Fig. 6A). Anderson et al. (2006a) suggested that GCD evolved first within the species *P. barbatus* and spread to other lineages through introgression, whereas Helms Cahan and Keller (2003) and Helms Cahan et al. (2006) suggested that the dependent lines and GCD originated through hybridization. Helms Cahan and Keller (2003) further claimed that the dependent lines are mosaics of *P. rugosus* and *P. barbatus*, and Schwander et al. (2007a) estimated that the line J1 has received 75% of its genome from *P. rugosus* whereas J2 has received 99% from *P. barbatus*. These estimates were based on current allele frequencies at nine microsatellite loci. As the current allele frequencies may not be a good representation of those at the time of lineage formation, we reestimate the genome composition of the different lineages and species on the basis of 1147 AFLP markers.

When simulating the evolution of hypothetical multilocus genotypes (Fig. 6B), the best fit to the actual AFLP distance matrix is obtained by assuming that the dependent lines are almost pure representatives of the parental species. The best result was based on the assumptions that J1 started as pure *P. rugosus* and J2 as pure *P. barbatus*, and that H1 received 80–90% of its genes from *P. rugosus* and H2 85–95% from J2. The estimates differ somewhat from those of Schwander et al. (2007a) in suggesting that the nuclear genomes of both J1 and J2 would mainly represent one parental species only. If there has been later gene flow between the dependent lineages or between them and the two parental species, the proportions calculated above will be affected. Nevertheless, the genetic distances based on AFLP show that within a pair of dependent lines, one partner is originally close to *P. rugosus* and the other one close to *P. barbatus*.

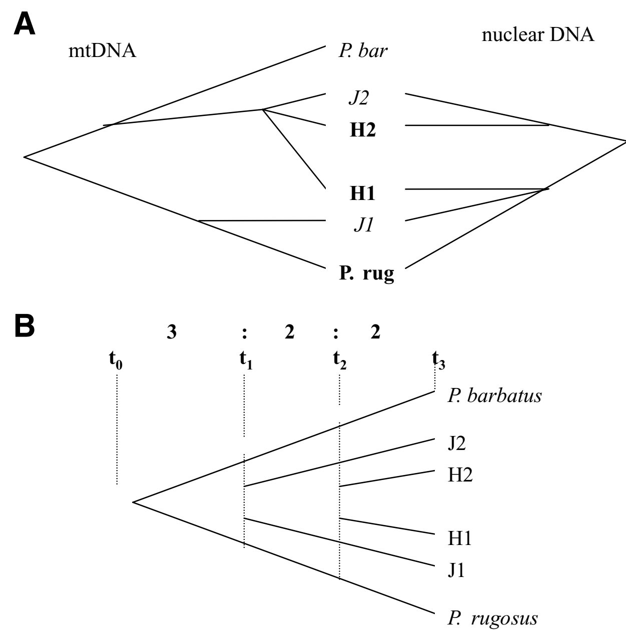

Figure 6. (A) A schematic view of the mtDNA phylogeny and nuclear genetic affinities (AFLP) of the lineages. The females in the lineages shown in bold have morphology of *P. rugosus*, those in italics of *P. barbatus*. (B) The model used in simulations

The results give support to the conclusion that the dependent lines have hybrid history, but they do not give strong (or any) support to the hypothesis that the dependent lines J1 and J2 initially originated through hybridization. The mitochondrial haplotypes tend to suggest that there has been later introgression between some of the lines. Evidence for that is provided by the sequence similarity of the mitochondrial haplotypes in *P. rugosus* and J1, and between J2 and the H-lineages. Furthermore, the females of both H1 and H2 are morphologically of the *P. rugosus* type and those of J1 and J2 of the *P. barbatus* type, even though such similarities are not evident in the nuclear AFLP distances. It thus seems clear that the dependent lineages themselves have some hybrid origin, and the workers resulting from a mating between two dependent lineages are hybrid offspring of hybrid parents. The AFLP results indicate that the level of introgression in the

dependent lines is relatively weak, 80–100% of the nuclear genes coming from one of the parental species.

Sequence divergence between the two major clusters of the mtDNA phylogeny, one with *P. rugosus* and J1, the other with *P. barbatus,* J2, H1, and H2, was on average $d = 0.075$. This is of the same magnitude as the divergence between *P. rugosus* and *P. barbatus*. When we removed the sequence variation within the parental species, the net divergence between the two main clusters was 0.072. Based on our distance estimates and assuming a rate of divergence of 2% per million years for the mitochondrial sequences (DeSalle et al. 1987) *P. rugosus* and *P. barbatus* diverged about 3.6 million years ago, J1 has been derived from the mitochondrial haplotype of *P. rugosus*, and the others from *P. barbatus*. Using the same rate estimate, the dependent lines would have diverged from the parental species about 0.5 and 2 million years ago, respectively. The divergence among the lines J2, H1, and H2 was 0.022. As the diversity within these lines was smaller than within the parental species, the net divergence is probably almost the same, suggesting that the lineages diverged about 1 million years ago (see also Schwander et al. 2008b). These, of course, are rough estimates liable to errors of sampling and of calibration of the substitution rate, but they may well reflect the relative order of divergence.

POPULATION DIFFERENCES

For the lineages J1 and J2, we sampled different subpopulations. The populations 2 and 6 are geographically the most distant to the other J2 populations and accordingly they form a clade of their own in the mitochondrial phylogeny. Our nuclear genotypes support this relationship (Table 4). The mtDNA haplotypes of J1 also have a tendency to form clusters in the mitochondrial tree according to the geographic location. Population 2 is nuclearly the most divergent from the others. The observed geographical differentiation of populations could simply reflect the biology of *Pogonomyrmex* species where high levels of genetic substructure and restricted dispersal, possibly due to patchy habitat, has been reported in *P. rugosus*/*P. barbatus* and another related species, *P. badius* (Strehl and Gadau 2004; Schwander et al. 2007a). A patchy habitat distribution has been suggested as a prerequisite for the independent evolution of hybridizing populations (Buerkle et al. 2000), and it could promote the development of lineages with a GCD system because it could automatically restrict the gene flow from parental species.

Even though distant populations shared some mtDNA haplotypes, the nuclear markers showed that remote populations were genetically well differentiated from each other. Any introgression between the lineages would lead to a genetic resemblance of sympatric populations of dependent lineages. However, this was not the case. The pattern of geographical differentiation thus agrees with the above conclusions of no, or very little, recent introgression between the dependent lineages.

WHICH MODEL OF GCD IS SUPPORTED BY OUR RESULTS?

As described above (Table 5), three different models have been presented for the GCD system in *Pogonomyrmex*. It is noteworthy that all models explain the maintenance of the system rather than its origin and the formation of new lineage pairs. Yet, these aspects are intimately connected. Any valid genetic model must therefore be able to explain the following points: (1) all lineages are at least to some degree of hybrid origin and their genomes are a mixture of the two species *P. rugosus* and *P. barbatus*, (2) each

Table 5. Overview of the three presented models relating to hybrid lineages.

Feature	One-factor model	Epistatic nuclear model	Cytonuclear model
Nature of GCD	Homozygotes develop into queens and heterozygotes into workers	Reciprocally homozygous individuals (aaBB, AAbb) develop into queens and double heterozygotes into workers	Heterozygotes (A′a′m′, A′a′M′) develop into workers and mtDNA-homozygote combinations into queens (a′a′m′, A′A′M′)
Origin of GCD	By intraspecific polymorphism	Epistatic interactions in hybrid offspring	Epistatic interaction in hybrid offspring and new incompatibility mutations
Spreading of GCD	Selective advantage for genetically determined queens	Inbreeding; double heterozygotes must initially reproduce	
Separation from parental species	Not specified	Not specified	Incompatibility mutations in mtDNA
Association of mtDNA haplotype and nuclear genotype	Breaks down with gene flow and recombination between lines	Breaks down with gene flow and recombination between lines	Remains

lineage is associated with a specific mitotype, (3) the mitotypes of lineages J2, H1, and H2 share a most recent common ancestor with *P. barbatus*, whereas J1's mitotype shares a recent common ancestor with *P. rugosus*, (4) the nuclear genomes of J1 and H1 are more similar to *P. rugosus*, whereas the nuclear genomes of H2 and J2 are more related with the genome of *P. barbatus*. We next discuss the presented models in light of our results.

All three models include implicit assumptions that have not been clearly delineated, for example reinforcement of incompatibilities that prevent back crossing with the parental populations. Also, all models explain the maintenance of the dependent lines, assuming no recombination takes place between them. However, all models assume that at an early stage of the dependent lineage system formation, the hybrids must have reproduced, and the models remain unclear regarding how the dependent lines are kept separate from the parental species. Finally, the models are not very clear on how new pairs of dependent lineages are formed.

One of the main differences between the models concerns the association of lineages and specific mitotypes. The nuclear models of Anderson et al. (2006a) and Helms Cahan et al. (2006) were criticized by Linksvayer et al. (2006) on the basis that the association between the nuclear caste-determining loci and the mitochondrial haplotype would easily break down by recombination if there would be gene flow between the lineages. Linksvayer et al. (2006) proposed that the association could be easily explained if mitochondrial factors would be directly involved in caste determination, whereas the nuclear models produce such an association as a byproduct of the reduction of gene flow between lineages. The model of Linksvayer et al. (2006) would have been supported if we would have detected nuclear gene flow between the lineages. In fact, Anderson et al. (2006a) suggested that there has been gene flow between dependent lineages as the level of genetic differentiation varied among three nuclear markers they used. However, such a result could also arise from historical differences in allele frequencies, and the demonstration of gene flow requires significant covariation of allele frequencies in different geographical populations. Our results using multilocus nuclear genotypes show that introgression between the lineages has played an insignificant role in the long-term evolution of the *Pogonomyrmex* hybrid lineages, thus removing the major criticism presented by Linksvayer et al. (2006) toward a solely nuclear gene model. Therefore, it is not possible to reject the two nuclear hypotheses (Table 5). However, the discrepancy between the genetic and morphological differentiation indicates that historical gene flow must have occurred at some point and the resulting nuclear differences (or similarities) may have been affected by strong selection in the hybrids. Nevertheless we should note that even if the separation of the lines J1 and J2 would have originally included a mitochondrial factor, it is hard to explain how that system was transferred to the lines H1 and H2 that share a relatively recent mitochondrial ancestor. Clearly, that would require new mitochondrial mutations affecting the caste-determination system.

The two proposed nuclear models differ from each other regarding the number of loci involved, i.e., whether several or a single genetic factor underlie GCD, and whether the caste determination system evolved first within one population/species and then spread to other populations via hybridization or was itself the result of hybridization (Anderson et al. 2006a; Helms Cahan et al. 2006). As indicated in Table 5, both models require additional assumptions to explain the formation of new lineage pairs and the separation from the parental species. Both models also assume an early hybridization event during the evolution of GCD; the two-locus model to explain the origin of the caste determination system itself and the one-locus model to explain its transfer to a new lineage. Our AFLP data, together with the simulation model, show that all lineages show some hybrid features but that either repeated back-crossing or strong selection has shaped the nuclear genomes in such a way that they are not clearly identified as hybrids.

The one-locus model by Anderson et al. (2006a) also assumed that the system first evolved in one population within the *P. barbatus* ECD species, giving rise to the J2 lineage. The most relevant rationales for this model were that the genetic distance between J2 and its mitochondrial parental species *P. barbatus* is three times higher than that between the most recently formed J1 and its mitochondrial parental *P. rugosus,* and the oldest GCD lineage J2 mtDNA also has the highest level of nucleotide diversity. The present results combined with earlier mtDNA studies (Table 1) show that it is not clearly evident that J2 is genetically more diverse and thus older than the others. Even though we did not have *P. barbatus* in our AFLP analysis, the observed nuclear data nicely fit a model that has the lines J1 and J2 forming at the same time. Even though this evidence is not very strong, we conclude that the support for J2 being the oldest line and the source of GCD is still unconfirmed. Transfer of mtDNA between species through hybridization without any clear sign of marked nuclear gene flow is known from several other species including ants (Goropashnaya et al. 2004).

The two-locus model has been criticized because it assumes that initially the double heterozygotes also became gynes, whereas currently such double heterozygotes develop into workers (Anderson et al. 2006a). The model thus requires that later enforcement of the caste-determination system must have taken place. Even though the criticism is valid, here we have pointed out that similar problems exist with the other models and can thus not be used as strong evidence.

One critical question addresses the origin of new dependent pairs. It seems possible that the lineage pairs F1–F2, G1–G2, and H1–H2 may represent, at least partly, geographically separate populations that have differentiated through accumulation of

genetic changes (e.g., by drift and bottlenecks) over time. Could such an explanation also explain the origin of H1–H2 from J1–J2? If the system of GCD first evolved in the J lines, a new pair could have been formed in an isolated allopatric population. The nuclear resemblance between H1 and J1, as well as between H2 and J2, suggests that H1/H2 or any of the other new lineages, if they really exist as independent entities, are derived from the original J1/J2 system, otherwise H1 and H2 should cluster together. The mitochondrial affinities could thus indicate an occasion of later introgression (although not regular gene flow) between the lines. That would be a parsimonious way to explain the origin of new pairs of dependent lines. The comparison of our simulation model with the obtained AFLP data, however, lends greater support to a model in which the new lines would have been formed anew by hybridization involving at least one of the initial parental species. Even though we had a very large nuclear dataset covering a significant proportion of the ant genome, the precise history of the lineages may be difficult to infer because of repeated, although rare, hybridization events that erased the historical information captured in an regularly evolving genome.

A combination of functional genomics and detailed comparative genomics analyses involving all lineages and parental species is necessary to solve the genetic basis, the mystery of the origin and further spread of GCD in *Pogonomyrmex*. One step toward this direction is the upcoming genome and transcriptome sequencing of *Pogonomyrmex barbatus* (Smith et al. 2010).

ACKNOWLEDGMENTS

We thank our co-workers in Davis and Oulu for advice and help, especially R. Jokela for valuable work in the laboratory. Financial support was provided by the Academy of Finland (122210 to PP), by the Science Foundation of the University of Oulu, Finnish Cultural Foundation and the Finnish Graduate School of Population Genetics (to AS).

LITERATURE CITED

Anderson, K. E., J. Gadau, B. M. Mott, R. A. Johnson, A. Altamirano, C. Strehl, and J. H. Fewell. 2006a. Distribution and evolution of genetic caste determination in *Pogonomyrmex* seed-harvester ants. Ecology 87:2171–2184.

Anderson, K. E., B. Hölldobler, H. Fewell, B. M. Mott, and J. Gadau. 2006b. Population-wide lineage frequencies predict genetic load in the seed-harvester ant *Pogonomyrmex*. Proc. Natl. Acad. Sci. USA 103:13433–13438.

Anderson, K. E., C. R. Smith, T. A. Linksvayer, B. M. Mott, J. Gadau, and J. H. Fewell. 2009. Modeling the maintenance of a dependent lineage system: the influence of positive frequency-dependent selection on sex ratio. Evolution 63:2142–2152.

Anderson, M. J. 2003. PCO: a FORTRAN computer program for principal coordinate analysis. Department of Statistics, Univ. of Auckland, New Zealand.

Beye, M., I. Gattermeier, M. Hasselmann, T. Gempe, M. Schioett, J. F. Baines, D. Schlipalius, F. Mougel, C. Emore, O. Rueppell, et al. 2006. Exceptionally high levels of recombination across the honey bee genome. Genome Res. 16:1339–1344.

Brady, S. G., T. R. Schultz, B. L. Fisher, and P. S. Ward. 2006. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc. Natl. Acad. Sci. USA 103:18172–18177.

Buerkle, C. A., J. M. Morris, M. A. Asmussen, and L. H. Rieseberg. 2000. The likelihood of homoploid hybrid speciation. Heredity 84:441–451.

Crozier, R. H., and P. Pamilo. 1996. Evolution of social insect colonies: sex allocation and kin selection. Oxford Univ. Press, Oxford, U.K.

DeSalle, R., T. Freedman, E. M. Prager, and A. C. Wilson 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian Drosophila. J. Mol. Evol. 26:157–164.

Fournier, D., A. Estoup, F. Orivel, J. Foucad, H. Jourdan, J. Le Breton, and L. Keller. 2005. Clonal reproduction by males and females in the little fire ant. Nature 435:1230–1234.

Goropashnaya, A., V. B. Fedorov, and P. Pamilo. 2004. Recent speciation in the *Formica rufa* group ants (Hymenoptera, Formicidae): inference from mitochondrial DNA phylogeny. Mol. Phylogent. Evol. 32:198–206.

Helms Cahan, S., and L. Keller. 2003. Complex hybrid origin of genetic caste determination in harvester ants. Nature 424:306–309.

Helms Cahan, S., and S. B. Vinson. 2003. Reproductive division of labor between hybrid and non-hybrid offspring in a fire ant hybrid zone. Evolution 57:1562–1570.

Helms Cahan, S., J. D. Parker, S. W. Rissing, R. A. Johnson, T. S. Polony, M. D. Weiser, and D. R. Smith. 2002. Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. Proc. R. Soc. Lond. B 269:1871–1877.

Helms Cahan, S., G. E. Julian, S. W. Rissing, T. Schwander, J. D. Parker, and L. Keller. 2004. Loss of phenotypic plasticity generates genotype-caste association in harvester ants. Curr. Biol. 14:2277–2282.

Helms Cahan, S., E. J. Glennis, T. Schwander, and L. Keller. 2006. Reproductive isolation between *Pogonomyrmex rugosus* and two lineages with genetic caste determination. Ecology 87:2160–2170.

Hung, A. C. F., and S. B. Vinson 1977. Interspecific hybridization and caste specificity of protein in fire ant. Science 196:1458–1460.

Johnson, R. A. 1998. Foundress survival and brood production in the desert seed-harvester ants *Pogonomyrmex rugosus* and *P. barbatus* (Hymenoptera, Formicidae). Ins. Soc. 45:255–266.

———. 2000. Habitat segregation based on soil texture and body size in the seed-harvester ants *Pogonomyrmex rugosus* and *P. barbatus*. Ecol. Entomol. 25:403–412.

Julian, G. E., J. H. Fewell, J. Gadau, R. A. Johnson, and D. Larrabee. 2002. Genetic determination of the queen caste in an ant hybrid zone. Proc. Natl. Acad. Sci. USA 99:8157–8160.

Kane, N. C., M. G. King, M. S. Barker, A. Raduski, S. Karrenberg, Y. Yatabe, S. J. Knapp, and L. H. Riesenberg. 2009. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent Helianthus species. Evolution 63:2061–2075.

Kosambi, D. D. 1944. The estimation of map distances from recombination values. Ann. Eugen. 12:172–175.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Lander, E. S., P. Green, J. Abrahamson, A. Baelow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Lincoln, S. E., M. Daly, and E. S. Lander. 1992. Constructing genetic maps with MAPMAKER/EXP 3.0. Whithehead Institute Technical Report, 3rd edn. Cambridge, MA.

Linksvayer, T. A., M. J. Wade, and D. Gordon. 2006. Genetic caste determination in harvester ants: possible origin and maintenance by cyto-nuclear epistasis. Ecology 87:2185–2193.

Moreau, C. S., C. D. Bell, R. Vila, B. Archibald, and N. E. Pierce. 2006. Phylogeny of the ants: diversification in the age of angiosperms. Science 312:101–104.

Ohkawara, K., M. Nakayama, A. Satoh, A. Trindl, and J. Heinze. 2006. Clonal reproduction and genetic caste differences in a queen-polymorphic ant, *Vollenhovia emeryi*. Biol. Lett. 2:359–363.

Parker, J. D., and S. W. Rissing. 2002. Molecular evidence for the origin of workerless social parasites in the ant genus *Pogonomyrmex*. Evolution 56:2017–2028.

Schwander, T., and L. Keller. 2008. Genetic compatibility affects queen and worker caste determination. Science 322:552.

Schwander, T., S. Helms Cahan, and L. Keller 2006. Genetic caste determination in *Pogonomyrmex* harvester ants imposes costs during colony founding. J. Evol. Biol. 19:402–409.

———. 2007a. Characterization and distribution of *Pogonomyrmex* harvester ant lineages with genetic caste determination. Mol. Ecol. 16:367–387.

Schwander, T., L. Keller, and S. Helms Cahan. 2007b. Two alternate mechanisms contribute to the persistence of interdependent lineages in *Pogonomyrmex* harvester ants. Mol. Ecol. 16:3533–3542.

Schwander, T., J. Y. Humbert, C. S. Brent, S. Helms Cahan, L. Chapuis, E. Renai, and L. Keller 2008a. Maternal effect on female caste determination in a social insect. Curr. Biol. 18:265–269.

Schwander, T., S. S. Suni, S. Helms Cahan, and L. Keller. 2008b. Mechanisms of reproductive isolation between an ant species of hybrid origin and one of its parents. Evolution 62:1635–1643.

Schwander, T., N. Lo, M. Beekman, B. P. Oldroyd, and L. Keller. 2010. Nature versus nurture in social insect caste differentiation. Trends Ecol. Evol. 25:275–282.

Seifert, B. 1999. Interspecific hybridisations in natural populations of ants by example of a regional fauna (Hymenoptera, Formicidae). Insect Soc. 46:45–42.

Sirviö, A., J. Gadau, O. Rueppell, D. Lamatsch, J. J. Boomsma, P. Pamilo, and R. E. Jr. Page. 2006. High recombination frequency creates genotypic diversity in colonies of the leaf-cutter ant *Acromyrmex echinatior*. J. Evol. Biol. 19:1475–1485.

Smith, C. R., C. Schoenick, K. E. Anderson, J. Gadau, and A. V. Suarez. 2007. Potential and realized reproduction by different worker castes in queen-less and queen-right colonies of *Pogonomyrmex badius*. Insect Soc. 54:260–267.

Smith, C. D., C. R. Smith, U. Mueller, and J. Gadau. 2010. Ant genomics: strength and diversity in numbers. Mol. Ecol. 19:31–35.

Strehl, C. P., and J. Gadau. 2004. Cladistic analysis of paleo-island populations of the Florida harvester ant (Hymenoptera:Formicidae) based upon divergence of mitochondrial DNA sequences. Florida Entomol. 87:576–581.

Suni, S. S., C. Gignoux, and D. M. Gordon. 2007. Male parentage in dependent-lineage populations of the harvester ant *Pogonomyrmex barbatus*. Mol. Ecol. 16:5149–5155.

Tsutsui, N. D., A. V. Suarez, J. C. Spagna, and J. S. Johnston. 2008. The evolution of genome size in ants. BMC Evol. Biol. 8:64.

Umphrey, G. J. 2006. Sperm parasitism in ants: selection for interspecific mating and hybridization. Ecology 87:2148–2159.

Volny, V. P., and D. M. Gordon. 2002. Genetic basis for queen-worker dimorphism in a social insect. Proc. Natl. Acad. Sci. USA 99:6108–6111.

Wilfert, L., J. Gadau, and P. Schmid-Hempel. 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98:189–197.

Associate Editor: J. Strassmann

Supporting Information

The following supporting information is available for this article:

Figure S1. In comprehensive neighbor-joining phylogeny, mitochondrial *cox1* sequences (302 bp) for eight *Pogonomyrmex* lineages (J1, J2, H1, H2, F1, F2, G1, and G2) are shown.

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

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