Mating system, population genetics, and phylogeography of the devil's garden ant, *Myrmelachista schumanni*, in the Peruvian Amazon

Pierre-Jean G Malé, E. Youngerman, Naomi E Pierce, M. E. Frederickson

Version Post-Print\Accepted Manuscript

Citation (published version) Malé, P.J.G., Youngerman, E., Pierce, N.E. et al. Mating system, population genetics, and phylogeography of the devil's garden ant, *Myrmelachista schumanni*, in the Peruvian Amazon. *Insect. Soc.* 67, 113–125 (2020). <https://doi.org/10.1007/s00040-019-00735-7>

DOI <https://doi.org/10.1007/s00040-019-00735-7>

**Publisher
Terms of
Use**

1. Archived content may only be used for academic research. Any content downloaded for text based experiments should be destroyed when the experiment is complete.
2. Archived content may not be used for purposes that are intended for or directed towards commercial advantage or monetary compensation by means of sale, resale, licence, loan, transfer or any other form of commercial exploitation ("Commercial Purposes").
3. Archived content may not be published verbatim in whole or in part, whether or not this is done for Commercial Purposes, either in print or online.
4. This restriction does not apply to reproducing normal quotations with an appropriate citation. In the case of text-mining, individual words, concepts and quotes up to 100 words per matching sentence may be used, whereas longer paragraphs of text and images cannot (without specific permission from Springer Nature).
5. All use must be fully attributed. Attribution must take the form of a link - using the article DOI - to the published article on the journal's website.
6. All use must ensure that the authors' moral right to the integrity of their work is not compromised.
7. Where content in the document is identified as belonging to a third party, it is the obligation of the user to ensure that any use complies with copyright policies of the owner.
8. Any use of Springer Nature content is at your own risk and Springer Nature accepts no liability arising from such use.

How to cite TSpace items

Always cite the **published version**, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the **author manuscript from TSpace** because you cannot access the published version, then cite the TSpace version **in addition to** the published version using the permanent URI (handle) found on the record page.

**This article was made openly accessible by U of T Faculty.
Please [tell us](#) how this access benefits you. Your story matters.**

**Mating system, population genetics, and phylogeography of the devil's garden ant,
Myrmelachista schumanni, in the Peruvian Amazon**

PJG Malé^{1,2,3*}, E Youngerman¹, NE Pierce², ME Frederickson^{1,3}

¹ Department of Ecology & Evolutionary Biology, University of Toronto, 25 Willcocks Street,
Toronto, ON, M5S 3B2, Canada

² Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street,
Cambridge, MA, 02138, USA

³ Centre for Analysis of Genome Evolution and Function, 25 Willcocks Street, Toronto, ON,
M5S 3B2, Canada

*Corresponding author: pjg.male@gmail.com

Acknowledgements. We thank the Dirección General Forestal y de Fauna Silvestre of the
Ministerio de Agricultura in Peru for permits (Nos. 0299-2011-AG-DGFFS-DGEFFS, 0046-
2014-MINAGRI-DGFFS-DGEFFS). We are grateful to A. Coral for assistance in the field.
This work was supported by FQEB grant RFP-12-06 from the National Philanthropic Trust to
NEP and MEF, NSF SES-0750480 to NEP and the Natural Sciences and Engineering
Research Council of Canada to MEF.

19 **Abstract**

20 Devil's gardens are a remarkable feature of Amazonian rainforests. These clearings result
21 from the cultivation of ant-plants by their symbiotic ant, *Myrmelachista schumanni*. Each
22 devil's garden is inhabited by a single *M. schumanni* colony, often with millions of workers
23 and thousands of queens. Through a combination of field surveys and microsatellite
24 genotyping, we examined *M. schumanni* colony structure, mating system, dispersal, and
25 phylogeography. We discovered that the reproduction of *M. schumanni* is weakly seasonal,
26 exhibits facultative polyandry, and involves split sex ratios potentially leading to sex-biased
27 dispersal. Surprisingly, we only observed very weak clustering of genetic variation, either
28 within or between devil's gardens. We hypothesize that the apparent absence of geographical
29 structure results from the unusually high level of genetic differentiation between colonies.
30 This study adds intriguing observations to the scarce literature about the reproduction and
31 phylogeography of Amazonian ants.

32
33 Keywords: Amazonian rainforests, ant-plant interactions, mutualism, mating,
34 phylogeography, microsatellites

37 **Introduction**

38 In the western Amazon Basin, “devil’s gardens” are orchard-like stands where only one or a
39 few species of myrmecophytic plants grow. The ant *Myrmelachista schumanni* nests in the
40 hollow, swollen stems (termed “domatia”) of these trees, and workers create devil’s gardens
41 by poisoning non-myrmecophytes with formic acid, effectively weeding out other plant
42 species (Frederickson et al. 2005). Devil’s gardens get their name from a local legend that
43 tells that they are cultivated by a mischievous forest spirit, the Chullachaqui. They are a
44 remarkable feature of western Amazonian rainforests because they sometimes grow large,
45 comprising hundreds of ant-plants and spanning a thousand square meters or more
46 (Frederickson and Gordon 2009).

47 Each garden is inhabited by a *M. schumanni* colony that can also grow to as many as 3
48 million workers and 15,000 queens (Frederickson et al. 2005). Little is known about how
49 *M. schumanni* colonies reproduce—how males and gynes mate, disperse, and found new
50 colonies. Other ant species in the genus *Myrmelachista* are also arboreal and restricted to the
51 Neotropics, where they nest in dead twigs or, less frequently, live stems (Longino 2006).
52 *Myrmelachista* nuptial flights have been documented only in the Antillean species
53 *Myrmelachista ramulorum*, which mates year-round (Torres et al. 2001). *Myrmelachista*
54 colonies can be polygynous or monogynous, and are often polydomous, spreading over
55 multiple stems or plants (Kuhn 2014; Longino 2006). Previous work has shown that the
56 devil’s garden ant, *M. schumanni*, nests in anywhere from just one to several hundred
57 myrmecophytic trees (often *Duroia hirsuta*), and that ants move back and forth among trees in
58 a garden, suggesting each garden is maintained by just one *M. schumanni* colony. A
59 demographic study that repeatedly surveyed populations of two myrmecophytic tree species

(*D. hirsuta* and *C. nodosa*) in the Peruvian Amazon found that *M. schumanni* colonized previously unoccupied, spatially isolated trees that were usually less than a year old and bore fewer than five domatia (Frederickson and Gordon 2009). Most small, spatially isolated trees contain just one de-alate *M. schumanni* queen (Frederickson, unpubl. data). Combined, these results suggest that a new colony of *M. schumanni* is founded by a single mated queen, who finds an unoccupied host plant, sheds her wings, and begins making workers. Subsequently, *M. schumanni* workers attack and kill vegetation growing near the base of the tree, creating a clearing in which other myrmecophytes can establish and grow free from competition with other plants. The *M. schumanni* colony then expands to occupy these nearby myrmecophytes, and the colony and garden grow in this manner over many years—potentially for centuries (Frederickson et al. 2005).

As a *M. schumanni* colony grows large, it adds more and more queens, but we do not know whether these queens are related (e.g., the initial queen's daughters or granddaughters, etc.) or whether *M. schumanni* colonies adopt unrelated queens from other colonies that arrive on the wing. Similarly, we do not know whether all queens in a colony have mated, whether they mate with their brothers or other males within a garden, or whether virgin queens (gynes) and males disperse from their natal nest in order to mate. Moreover, little information exists on the seasonal timing of reproduction in *M. schumanni*; in aseasonal tropical forests, some arboreal ants produce alates almost continuously, but with a peak in production during the hottest and wettest months of year (Frederickson 2006). Here, we measured the number of males, gynes, and (de-alate) queens in *M. schumanni* colonies every month for an entire year, and used microsatellite markers to infer the mating system of *M. schumanni*.

To do so, we sampled *M. schumanni* from sites across the Peruvian Amazon, which also allowed us to study how *M. schumanni* genotypes are distributed at a regional scale. Few

studies have explored the phylogeography of Amazonian ants, but studies of other taxa suggest that tectonic movements, sea-level fluctuations, riverine barriers, forest refuges during dry periods, and competitive species interactions may all have contributed to current landscape genetic patterns in the Amazon Basin (Haffer 1997). The scarce literature on South American ants supports a historical role of arches (ancient geological barriers) in promoting allopatric differentiation and isolation by distance among populations (Ahrens et al. 2005; Cristiano et al. 2016; Solomon et al. 2008), but there is little information available for arboreal ants, and no previous research has characterized phylogeographic patterns in *M. schumanni*.

In ants as in other taxa, local patterns of dispersal and gene flow scale up to determine how genotypes are distributed at larger spatial scales. For example, if *M. schumanni* gynes mate with males in their natal nest and remain there to lay eggs, we would expect colonies separated by even small physical distances to be highly genetically differentiated and potentially inbred. At larger spatial scales, restricted gene flow should give rise to isolation by distance, and genetic structure may reflect geographic barriers to ant dispersal, or to the dispersal of their myrmecophytic host trees. However, outside of the range of *M. schumanni*'s main host plant, *D. hirsuta*, this ant species still finds suitable host plants by nesting in *Cordia nodosa* (especially in southern Peru) or melastomes such as *Tococa* spp. or *Clidemia heterophylla* (especially in the foothills of the Andes).

By sampling 43 *M. schumanni* devil's gardens that were widely distributed across much of the Peruvian Amazon and genotyping their workers at 18 microsatellite markers, we characterized the reproductive system, population genetic structure, and phylogeography of the devil's garden ant, *Myrmelachista schumanni*, exploring genetic variation at different spatial scales, ranging from very local (i.e., among trees in a single devil's garden) to regional

(i.e., across the Peruvian Amazon). We also present new data on the reproductive phenology and sex ratio of *M. schumanni* colonies. Thus, this study offers insights into the mating, dispersal, and population structure of *M. schumanni* in the western Amazon.

Materials and methods

Model system

Myrmelachista schumanni Emery (Formicinae) is an arboreal ant species, with workers that attack non-myrmecophytic plants surrounding their host-plant (Frederickson et al. 2005). Depending on the geographic area, the devil's gardens that result from this ant behaviour can be comprised of nearly pure stands or mixtures of up to four myrmecophytic plant species: *Clidemia heterophylla* (Desr.) Gleason, *Cordia nodosa* Lam., *Duroia hirsuta* (Poepp.) K. Schum. and *Tococa guianensis* Aubl (Frederickson and Gordon 2007). Devil's gardens have been described in the Western Amazon (mostly in Peru and Ecuador, but also in Brazil and Colombia), as well as in French Guiana (Salas-Lopez et al. 2016). They are all inhabited by *Myrmelachista* ants, and although it is unclear whether the resident ants are all *M. schumanni* or a complex of different species, *M. schumanni* appears to be the main devil's garden ant species (Edwards et al. 2009).

Sampling and genotyping

In 2004-2005, we collected all domatia from 15-16 *D. hirsuta* trees every month (181 trees in total) at the Allpahuayo Mishana National Reserve in Loreto, Peru. The plants came from at least eight distinct devil's gardens each month; different trees were sometimes collected from the same garden in different months. For each *D. hirsuta* tree, we counted and opened all domatia, and counted all male and female reproductives, including pupae and eclosed adults. It is straightforward to tell which pupae will produce alates because of their size and

developing wings, and their sex can also be determined from their size and shape, particularly of the head. We counted eclosed adults in three categories: alate males, alate females (gynes), and de-alate queens. We did not count workers.

In 2013, we collected and successfully genotyped 50 queens and 597 workers from 43 gardens located in 12 different localities in Peru, using 18 ant-universal microsatellite markers (Butler et al. 2014). Distances between gardens varied from 13 m to 1,108 km (see Table 1 and Fig. 1). We counted the number of trees in each garden and recorded the dominant tree species (either *D. hirsuta* or *C. nodosa*). At least 10 ant workers were collected from one domatium per garden to study the population genetic structure of *M. schumanni* at a country-wide spatial scale. We collected additional workers and queens from two large gardens located 3.46 km apart in Madre Selva and Nuevo Israel, respectively. We used these additional samples to study patterns of relatedness and genetic diversity within gardens. Several trees were sampled for each of these two gardens (see below). Ants were preserved in 96% ethanol. Each garden was geolocated using a GPSMAP® 64s GPS receiver (Garmin Ltd.).

Total DNA was extracted from each ant worker. Each ant worker was incubated overnight at 55°C in 10 µL of Proteinase K solution (Qiagen) and 150 µL of 10% Chelex solution (BioRad). We used 4 µL of the solution obtained to amplify 18 microsatellite markers multiplexed in four different PCR sets using forward primers fluorophore-labeled on the 5' end (Table 2). PCR amplifications were carried out in a 10 µL final volume containing 0.7X QIAGEN Multiplex Master Mix, 0.1-0.025 µM of each primer and 4 µL of genomic DNA. The thermal profile started with initial denaturation at 95°C (15 min), was followed by 40 cycles of denaturation at 94°C (30 sec), annealing at 57°C (1 min 30) and extension at 72°C (1 min 30), and ended with final extension at 72°C for 10 min. Loci were finally sized using an ABI 3730 sequencer (Applied Biosystems) at the Centre for Analysis of Genome

Evolution and Function (CAGEF) with the 600 LIZTM GeneScanTM size standard and GENEIOUS v9.1.6 software.

Assessment of the reproduction mode of *M. schumanni*

We used the data on the number of male and female *M. schumanni* reproductives in *D. hirsuta* trees from Allpahuayo-Mishana to examine reproductive phenology, sex ratio, and how fecundity scales with tree size. Using the *lme4* package in R v 3.5.0, we fit a generalized linear model (GLMM) using a Poisson error distribution with the total number of alates (summed for males and females) as a response variable and the number of domatia (log-transformed) and month of the year (as an ordered categorical variable) as fixed effects. To account for sampling multiple *D. hirsuta* trees per garden, we included garden as a random effect in the model. We also fit a GLMM with the same fixed and random effects to the number of de-alate queens. For both models, Type III Wald chi-square test statistics were calculated using the Anova function from the *car* package. We calculated sex ratio as the proportion of all alates that are female and as the proportion of all reproductives (including queens) that are female for each tree. We tested for split sex ratios using the *diptest* package in R that tests for uni-modality versus multi-modality. For five large devil's gardens, we sampled at least 10 different *D. hirsuta* trees over the course of the year, and thus could also calculate sex ratio at the garden-level.

We inferred the reproduction mode of *M. schumanni* by comparing the genotypes of queens, of the males that inseminated them, and of workers from the same colonies. Two large gardens were sampled for queens and workers. One garden (Garden 30) was in Madre Selva. It comprised 56 *D. hirsuta* trees, and we collected and genotyped 24 queens and 66 workers from five of these trees. The other garden (Garden 45) was in Nuevo Israel. It comprised 270 *D. hirsuta* trees, and we collected and genotyped 26 queens and 87 workers from eight of the trees. The two gardens were 3.46 km apart. Queens were dissected to collect

their spermathecae. Maternal (queen), paternal (male) and worker DNA were extracted separately from heads, spermathecae, and whole bodies, respectively, and microsatellites were then amplified as described above.

We inferred the minimum number of mating events per queen by counting the maximum number of alleles observed for a given microsatellite marker in spermathecae. We then used the program Relatedness 5.0.8 (Queller and Goodnight 1989) run on the QEMU virtual machine emulator (Bellard 2005) to assess mate choice and sex-biased dispersal separately for each garden. Relatedness 5.0.8 accommodates haplodiploid genotype data to calculate relatedness statistics using the method of Queller and Goodnight (Queller and Goodnight 1989). Note that we used genetic information only from singly-mated females and the content of their spermathecae because we cannot infer individual genotypes of males from the pooled genotypic information of the spermatheca contents of multiply mated queens. To assess whether queens and males mate at random within a garden, i.e. irrespective of their relatedness, we determined for each singly-mated female whether they were more or less closely related to their mates than to the overall pool of reproducing males in their garden, and whether males were more or less related to the female they inseminated than to the overall pool of singly-mated females. To address sex-biased dispersal, we compared the relatedness between singly-mated females and their mates to the relatedness among singly mated-females, the relatedness between males and the female they inseminated to the relatedness among males, and the relatedness between workers and males to the relatedness between workers and females. Background allele frequencies were estimated separately for each of these two devil's gardens using data on other devil's gardens at each site (i.e., Madre Selva and Nuevo Israel). We calculated allele frequencies within devil's garden sampled for the large-scale population genetic study using Arlequin 3.5 (Excoffier et al. 2005), and averaged them over all the devil's gardens sampled in Madre Selva and Nuevo Israel separately.

Assessment of fine-scale spatial genetic structure

We studied the spatial structure of genetic diversity at the scale of the two big gardens described above. First, we used Relatedness 5.0.8 to test whether the relatedness among workers, among queens, and among males is greater within gardens than between gardens. Second, we fit an analysis of molecular variance (AMOVA: Excoffier et al. 1992) using Arlequin 3.5 (Excoffier et al. 2005) to characterize the patterns of genetic divergence within gardens. For this analysis, we considered gardens as groups of populations and trees as populations, the underlying question being whether there was genetic substructure among trees within a garden or whether each garden is genetically homogeneous.

Assessment of large-scale spatial genetic structure

We also assessed genetic structure at a country-wide scale for *M. schumanni*. Departure from Hardy-Weinberg equilibrium and F_{IS} inbreeding coefficients were calculated for each sampling location independently, using FSTAT v2.9.3.2 (Goudet 1995).

Isolation by distance (IBD) between gardens was assessed by computing pairwise F_{ST} indexes, and then calculating the significance of the regression of $F_{ST} / (1 - F_{ST})$ on the geographical distance between gardens using SPAGeDi 1.5a (Hardy and Vekemans 2002). We assessed the significance of pairwise F_{ST} through 2,000 permutations of individuals among populations. Assessments were performed separately over a short distance (between 0 and 400km), a long distance (over 400km), and over the whole sampling range.

We analyzed phylogenetic relationships among gardens by constructing neighbour-joining trees with 500 bootstrap iterations using POPTREE (Takezaki et al. 2014) based on the pairwise F_{ST} indexes calculated previously. We used a Discriminant Analysis of Principal Components (DAPC) as a multivariate assessment of the differentiation among the 43 devil's gardens. We also ran a separate DAPC on just the 18 devil's gardens at Madre Selva. The two analyses were calculated with the dapc function in the *adegenet* package in R v2.14.2

software (Jombart 2008) with a total of 34 and 30 principal components, accounting for 89.9% and 92.7% of the total variance.

We also performed Bayesian clustering analyses to infer the spatial structure of the genetic data using STRUCTURE v2.3.4 (Pritchard et al. 2000). Each round of STRUCTURE consisted of 10 runs for each of the K genetic groups tested, with K ranging from 1 to 44. All of the runs were conducted with the uncorrelated allele frequencies model, the LOCPRIOR model and the admixture model for 100,000 iterations after a burn-in period of 10,000 iterations. This burn-in period was previously proven to be sufficient for the likelihood of the runs to stabilize. The number of genetic groups was determined using the ΔK method developed by Evanno *et al.* (2005) which involves assessing the breakpoint in the slope of the distribution of $\ln P(D)$ as a function of K . This method was implemented using the online program STRUCTURE HARVESTER (Earl and vonHoldt 2012). Then, 100 runs were performed as described above, but with K fixed to the previously determined value. To determine the final assignment of individuals to clusters, we summarized the data on the outcomes of these 100 replicates using the CLUMPP program with the greedy algorithm (Jakobsson and Rosenberg 2007).

Results

Assessment of the reproduction mode of *M. schumanni*

Myrmelachista schumanni colonies do not produce reproductives evenly throughout the year. Production of both male and female alates peaked in March-April (Fig. 2), which are also the rainiest months of the year. Overall, few alate pupae were encountered, but peak production of alate pupae preceded peak production of adult alates by about two months; more alate pupae were found in January (females) and February (males) than in other months

(Fig. 2). Unexpectedly, the number of de-alate queens also varied seasonally, reaching a peak in April (Fig. 2). Total production of adult alates (both males and females) was significantly explained by tree size (Wald $X^2_1 = 2513.872$, p -value $\ll 0.001$) and month (Wald $X^2_1 = 5379.01$, p -value $\ll 0.001$).

The sex ratio of alates and alate pupae collected from *D. hirsuta* trees was heavily biased towards males (Fig. 3a). When all reproductives (including de-alate queens) in a tree were considered, the sex ratio was still male-biased, but less so (Fig. 3b). The distribution of sex ratios, measured as the proportion of reproductives that are female, was significantly bimodal when all reproductives were considered (Hartigan's dip test: $D = 0.051$, p -value = 0.039), but not when only alates and alate pupae were considered (Hartigan's dip test: $D = 0.023$, p -value = 0.980), probably because of the small number of trees containing only female alates (Fig. 3a). In other words, at the tree-level, sex ratios are split, with some trees containing mostly males and some trees containing mostly females. Only five gardens (i.e., entire *M. schumanni* colonies) were sufficiently well sampled to calculate sex ratio at the colony-level; all of these were large gardens comprising at least 20 *D. hirsuta* trees. At least several hundred de-alate queens and male alates were found in each of these five gardens; three colonies also produced at least one hundred female alates, but we found only 12 and 23 female alates in the other two colonies (Table 3). Four gardens had sex ratios strongly biased towards males, and one garden had a more even ratio of females to males (Table 3).

From the genetic data, the reproductive mode of *M. schumanni* was found to be facultative polyandry with random mating. All the 50 dissected spermathecae contained sperm; 23 spermathecae contained the sperm of one male, 23 contained the sperm of two males, and four contained the sperm of three males. The departure from a monogynous/monoandrous system was further confirmed by the fact that workers within gardens were less related on average than full-siblings ($r = 0.75$) or even paternal half-siblings ($r = 0.5$), and so were queens (mean

relatedness $\pm$ standard errors in Garden 30: workers: 0.437 ± 0.016 , queens: 0.448 ± 0.024 ; and in Garden 45: workers: 0.425 ± 0.018 , queens: 0.436 ± 0.036). They were however more related than maternal half-siblings ($r = 0.25$). Similarly, males were less related than full-siblings ($r = 0.5$) but were still somewhat related (mean relatedness $\pm$ standard errors in Garden 30: 0.284 ± 0.075 ; and in Garden 45: 0.310 ± 0.032).

In both gardens, singly mated queens were found to be equally related to their mate as to the overall pool of reproducing males within the same garden (mean relatedness differences $\pm$ standard errors; Garden 30: -0.027 ± 0.029 , p -value = 0.176; Garden 45: -0.024 ± 0.037 , p -value = 0.263). Similarly, males were found to be equally related to the female they inseminated as to the overall pool of singly mated females within the same garden (mean relatedness differences $\pm$ standard errors; Garden 30: -0.015 ± 0.020 , p -value = 0.123; Garden 45: -0.004 ± 0.027 , p -value = 0.696).

We found evidence of male-biased dispersal in Garden 30: queens were less related to males than to other queens (mean relatedness differences $\pm$ standard errors; -0.054 ± 0.034 , p -value = 0), males were as related to queens as to other males (mean relatedness differences $\pm$ standard errors; 0.014 ± 0.07 , p -value = 0.607), and workers were less related to males than to queens (mean relatedness differences $\pm$ standard errors; -0.014 ± 0.013 , p -value = 0).

The pattern we observe in Garden 45 was less clear: queens were also less related to males than to other queens (mean relatedness differences $\pm$ standard errors; -0.021 ± 0.032 , p -value = 0.013), but males were less related to queens than to other males (mean relatedness differences $\pm$ standard errors; -0.049 ± 0.024 , p -value = 0), and workers were more related to males than to queens (mean relatedness differences $\pm$ standard errors; 0.024 ± 0.009 , p -value = 0; Fig. 4).

Assessment of fine-scale spatial genetic structure

The two well-sampled gardens were genetically distinct, since workers from one garden

were more related to each other than to workers of the other garden (mean relatedness differences $\pm$ standard errors; 0.248 ± 0.014 , p -value = 0). The same result held true for queens and males (mean relatedness differences $\pm$ standard errors; queens: 0.106 ± 0.012 , p -value = 0; males: 0.116 ± 0.032 , p -value = 0).

AMOVA showed that most of the total genetic variation was distributed within individuals (74.34%) and that a lower percentage (26.95%) corresponded to variation between gardens (Table 4). No variation was significantly explained by the subdivision by trees, indicating the absence of spatial structuring among trees within gardens.

Assessment of large-scale spatial genetic structure

Gardens appeared well differentiated from each other as indicated by high and significant F_{ST} values (mean $\pm$ standard deviation; pairwise F_{ST} : 0.327 ± 0.076 , see Table S1 for all pairwise F_{ST} values). Four gardens, all located in Panguana, showed a significant departure from Hardy-Weinberg equilibrium due to heterozygote excess, as indicated by negative F_{IS} values (Table 1).

No significant IBD was detected, whatever the spatial scale (p -values; from 0 to 400 km: 0.294; over 400km: 0.997; the entire range: 0.743; Fig. 5). Similarly, the population tree reconstructed was very poorly supported (Fig. 6a). Discriminant analyses revealed a weak trend for differentiation among the 43 devil's gardens depending on geography (Fig. 6b). Each devil's garden occupied a small area in the overall space and there was very little overlap between the gardens from central (Neshuya, Quimpichiari, and Panguana) and southern (Las Piedras and Wasai) Peru. There was a much greater overlap with gardens from northern Peru (all other sites). Differentiation was much more pronounced at a small geographic scale, when focusing on the 18 devil's gardens from Madre Selva (Fig. 7).

Two clusters were identified by STRUCTURE as the most plausible population structure, although the DeltaK value calculated using STRUCTURE HARVESTER was low. The

population clustering did not reflect any obvious geographic structure (Fig. S2).

Discussion

Through a combination of field surveys and microsatellite genotyping, we described the colony structure and population genetics of the devil's garden ant, *M. schumanni*. We found that *M. schumanni* reproduces seasonally, exhibits facultative polyandry without assortative mating within gardens, and has split sex ratios potentially leading to sex-biased dispersal. We also found that most genetic variation is distributed within colonies, with high differentiation among colonies that could explain the absence of genetic structure at larger spatial scales.

Like many other tropical ants (Torres et al. 2001), *M. schumanni* reproduction is weakly seasonal, with alates produced throughout the year but peaking in February-April. Seasonality in ant reproduction is thought to result from adaptation to abiotic or biotic constraints, often linked to nest site availability (Kaspari et al. 2001a; Kaspari et al. 2001b; Rathcke and Lacey 1985). In some tropical plant-ants, ant phenology matches the seasonal availability of unoccupied host plants (Frederickson 2006). The same ecological pressure for synchronizing alate release with young *D. hirsuta* emergence could explain the seasonality of *M. schumanni* reproduction, although it could also result from many other factors.

Seasonal change in queen number is a surprising result, as it has been previously documented only in temperate species such as *Brachyponera chinensis*, *Linepithema humile* and *Myrmica limanica* (Elmes and Petal 1990; Gotoh and Ito 2008; Keller et al. 1989). Although an increase in queen number might reflect an increased need for alates or workers to care for developing alates, this is unlikely because the peak in *M. schumanni* queen number occurred in April, after alate production had already peaked (Fig. 2). Alternatively, colonies might adaptively adjust queen numbers to meet a seasonally changing demand for workers to engage in tasks other than brood care; *M. schumanni* workers protect *D. hirsuta* and

C. nodosa against insect herbivores (Frederickson 2005) and actively patrol the forest floor and attack non-myrmecophytic plants that attempt to establish in devil's gardens (Frederickson et al. 2005); insect herbivore abundance, new leaf production, and seedling emergence can all fluctuate seasonally, even in tropical forests (Wolda 1978). Additionally, seasonality in queen number could also be a consequence of the arrival of young queens from elsewhere, or gynes staying in their natal nest following mating, such that the number of queens increases temporarily, but then diminishes over the following weeks as some queens die off.

Weak facultative polyandry instead of consistent multiple matings with numerous males is expected in such a highly polygynous species as *M. schumanni* because of a trade-off between polyandry and polygyny (Hughes et al. 2008). Reproduction in social Hymenoptera varies along a continuum from monogyny to high polygyny, and monoandry to high polyandry. Polygyny and polyandry are thought to have evolved mostly due to selection for genetic diversity, but they both involve costs for queens either in terms of lower relatedness to workers, energy expenditure, or predation exposure (Hughes et al. 2008). Assortative mating (positive or negative) can also influence colony genetic diversity through queen mating choices. However, to the best of our knowledge, it has not been described in hymenopterans.

Our results do not suggest assortative mating in *M. schumanni* within gardens, in the sense that queens and their mates were as related to each other as to the other reproductives in the same devil's garden. Our sampling did not allow us to formally test for assortative mating at the population scale, since we only genotyped queens and males from the two exhaustively sampled gardens and they were in distinct locations. However, the mean relatedness within devil's gardens was very high, even among reproductives, indicating common ancestry. These high values could result from very short dispersal distances and therefore queens mating with males from the same or a genetically (and possibly physically) close devil's garden. Such

behaviour would result in colonies separated by even small physical distances being genetically differentiated, a hypothesis consistent with our large-scale population genetic results.

Nonetheless, we observed patterns of relatedness asymmetry between sexes within colonies. This can result from sex-biased dispersal, a well-documented phenomenon (Sündstrom et al. 2003). Surprisingly, these patterns were inconsistent between colonies: one garden showed signs of male-biased dispersal while another showed an inconclusive pattern contradicting male-biased dispersal. Split sex ratios could potentially be one explanation for these patterns. Conflicts between queens and workers over optimal sex allocation are the main explanation for split sex ratios in ants (Kümmerli and Keller 2009). This conflict is highest with high relatedness asymmetry of workers towards females and males such as the one we uncovered in the present study (Fig. 4). This relatedness asymmetry varies depending on the number of queens per colony and the queen mating frequency. The observed split sex ratio is in line with theoretical predictions according to what we know about *M. schumanni*'s mode of reproduction (high polygyny with monogynous small colonies and facultative polyandry, see Frederickson and Gordon 2009 and this study). Relatedness asymmetry is highest in monogynous small colonies, resulting in incentives for workers to bias alate production towards females. However, in polygynous colonies, workers and queens have similar interests, and sex-ratio is influenced by local resource competition between co-breeding queens. Colonies with a few queens produce more females to recruit new queens, whereas highly polygynous colonies specialize in male production. Consistent with these predictions, we found that *M. schumanni* population sex allocation is male-biased (Fig. 3).

Most of the genetic variation in *M. schumanni* populations was distributed within individuals, and to a lesser extent among devil's gardens. We found no evidence of spatial clustering of related individuals (for example, workers with the same mother or father) on

trees within a garden, although our sampling did not allow us to partition genetic variance among domatia within trees, so we do not know whether a domatium usually contains related brood. *Myrmelachista schumanni* workers readily move between trees in a garden, either along branches that touch between trees or over the ground, leading to the spatial homogenisation of worker genotypes within devil's gardens. Our genetic data support earlier assertions (e.g., Frederickson and Gordon 2009) that each devil's garden is occupied by a single, large *M. schumanni* colony in which multiple, mated queens produce genetically diverse but related offspring that mix freely within a garden, much as in other supercolonial ant species (e.g., *Formica aquilonia*, Schultner et al. 2016).

We failed to uncover any clear geographic structure of genotypes at the country-wide scale. Genetic homogenization through high gene flow between populations cannot explain this pattern. The most distant devil's gardens were over 1,000 km apart, and alates flying over such large distances is unlikely. An alternative explanation is a mismatch between microsatellite mutation rate and the timing of differentiation between populations. A very low mutation rate should lead to a lack of differentiation at microsatellite loci despite differentiation in the rest of the genome. On the other hand, a very high mutation rate should lead to homoplasy between unrelated microsatellite markers, especially among populations that diverged a long time ago. Our results might show some sign of homoplasy between distant populations, since F_{ST} values tend to decrease (although not significantly) over 400km. However, both explanations rely on apparent low divergence between populations, a pattern that does not apply to *M. schumanni* given the very high F_{ST} values observed between devil's gardens (0.323 in average, and up to 0.52). We hypothesize that, on the contrary, the absence of geographical structure is due to a high differentiation rate with little gene flow between devil's gardens. Each garden or small group of geographically close gardens would thus function as an almost isolated population. Discriminant analyses highlighted a few of such

seemingly isolated gardens, and the long branches of the F_{ST} -based neighbour-joining consensus tree supports the hypothesis of an ancient differentiation event between gardens followed by a long period of independent evolution with modest gene flow.

Four gardens in Panguana showed a significant departure from Hardy-Weinberg equilibrium, due to strongly negative F_{IS} values. Significant outbreeding could result from disassortative mating, with ant queens preferentially mating with genetically distant males. Interestingly, these four gardens are in Panguana, at the range limit of *D. hirsuta*, *M. schumanni*'s main host plant. If these gardens are at a contact zone between two genetically differentiated clusters, this could explain negative F_{IS} values, especially given split sex ratios, if each cluster specializes in producing males or females. However, this hypothesis seems to be disproven by the absence of obvious spatial genetic structure in *M. schumanni*.

Although our results about *M. schumanni* reproduction are in line with current knowledge and theoretical predictions about ant mating systems, the observed absence of population genetic structure due to a very high level of differentiation is surprising. The unsuitability of microsatellites for measuring large scale populational events has been suggested as a cause for inflated indices of differentiation due to high mutation rate, genotyping errors or ascertainment bias (Fischer et al. 2017; Hoffman and Amos 2005; Whitlock 2011). Conversely, the low number of markers used in microsatellite studies can result in low resolution, therefore in the overlooking of fine population structure (Bradbury et al. 2015; Jeffries et al. 2016). This study highlighted intriguing patterns, but these results need to be refined through the sampling of more populations in central Peru, as well as the use of other genetic markers such as RAD-Seq generated SNPs (Hodel et al. 2016).

453 **References**

- 454 Ahrens ME, Ross KG, Shoemaker DD (2005) Phylogeographic structure of the fire ant
455 *Solenopsis invicta* in its native South American range: roles of natural barriers and
456 habitat connectivity. *Evolution* 59:1733-1743
- 457 Bellard F QEMU, a fast and portable dynamic translator. In: USENIX Annual Technical
458 Conference, FREENIX Track, 2005. p 46
- 459 Bradbury IR, Hamilton LC, Dempson B, Robertson MJ, Bourret V, Bernatchez L, Verspoor
460 E (2015) Transatlantic secondary contact in Atlantic Salmon, comparing
461 microsatellites, a single nucleotide polymorphism array and restriction-site associated
462 DNA sequencing for the resolution of complex spatial structure. *Mol Ecol* 24:5130-
463 5144
- 464 Butler IA, Siletti K, Oxley PR, Kronauer DJC (2014) Conserved microsatellites in ants
465 enable population genetic and colony pedigree studies across a wide range of species.
466 *PLoS one* 9:e107334
- 467 Cristiano MP, Cardoso DC, Fernandes-Salomão TM, Heinze J (2016) Integrating
468 paleodistribution models and phylogeography in the grass-cutting ant *Acromyrmex*
469 *striatus* (Hymenoptera: Formicidae) in southern lowlands of South America. *PLoS*
470 *one* 11:e0146734
- 471 Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for
472 visualizing STRUCTURE output and implementing the Evanno method. *Conserv*
473 *Genet Resour* 4:359-361 doi:10.1007/s12686-011-9548-7
- 474 Edwards DP, Frederickson ME, Shepard GH, Yu DW (2009) A plant needs ants like a dog
475 needs fleas: *Myrmelachista schumanni* ants gall many tree species to create housing.
476 *Am Nat* 174:734-740
- 477 Elmes GW, Petal J (1990) Queen number as an adaptable trait: evidence from wild
478 populations of two red ant species (Genus *myrmica*). *J Anim Ecol* 59:675-690
479 doi:10.2307/4888
- 480 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using
481 the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611-2620
- 482 Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package
483 for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50
- 484 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from
485 metric distances among DNA haplotypes: application to human mitochondrial DNA
486 restriction data. *Genetics* 131:479-491
- 487 Fischer MC et al. (2017) Estimating genomic diversity and population differentiation – an
488 empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*.
489 *BMC Genomics* 18:69
- 490 Frederickson ME (2005) Ant species confer different partners benefits on two neotropical
491 myrmecophytes. *Oecologia* 143:387-395
- 492 Frederickson ME (2006) The reproductive phenology of an Amazonian ant species reflects
493 the seasonal availability of its nest sites. *Oecologia* 149:418-427
- 494 Frederickson ME, Gordon DM (2007) The devil to pay: a cost of mutualism with
495 *Myrmelachista schumanni* ants in 'devil's gardens' is increased herbivory on *Duroia*
496 *hirsuta* trees. *Proc R Soc B* 274:1117-1123 doi:10.1098/rspb.2006.0415
- 497 Frederickson ME, Gordon DM (2009) The intertwined population biology of two Amazonian
498 myrmecophytes and their symbiotic ants. *Ecology* 90:1595-1607
499 doi:<http://dx.doi.org/10.1890/08-0010.1>
- 500 Frederickson ME, Greene MJ, Gordon DM (2005) 'Devil's gardens' bedevilled by ants.
501 *Nature* 437:495-496

- Gotoh A, Ito F (2008) Gotoh, A., and F. Ito. "Seasonal cycle of colony structure in the Ponerine ant *Pachycondyla chinensis* in western Japan (Hymenoptera, Formicidae). Insectes Soc 55:98-104
- Goudet J (1995) FSTAT Version 1.2.: a computer program to calculate F-statistics. J Hered 86:485-486
- Haffer JR (1997) Alternative models of vertebrate speciation in Amazonia: an overview. Biodivers Conserv 6:451-476
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618-620 doi:10.1046/j.1471-8278.2002.00305.x
- Hodel RGJ et al. (2016) The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. Appl Plant Sci 4:1600025
- Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Mol Ecol 14:599-612
- Hughes WOH, Ratnieks FLW, Oldroyd BP (2008) Multiple paternity or multiple queens: two routes to greater intracolony genetic diversity in the eusocial Hymenoptera. J Evol Biol 21:1090-1095 doi:<https://doi.org/10.1111/j.1420-9101.2008.01532.x>
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806
- Jeffries DL, Copp GH, Handley LL, Olsén KH, Sayer CD, Hänfling B (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. . Mol Ecol 25:2997-3018
- Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405
- Kaspari M, Pickering J, Longino JT, Windsor D (2001a) The phenology of a Neotropical ant assemblage: evidence for continuous and overlapping reproduction. Behav Ecol Sociobiol 50:382-290
- Kaspari M, Pickering J, Windsor D (2001b) The reproductive flight phenology of a neotropical ant assemblage. Ecol Entomol 26:245-257
- Keller L, Passera L, Suzzoni J-P (1989) Queen execution in the Argentine ant, *Iridomyrmex humilis*. Physiological Entomology 14:157-163
- Kuhn KM (2014) Colony founding by the ant *Myrmelachista flavocotea*. Insectes Soc 61:239-245
- Kümmerli R, Keller L (2009) Patterns of split sex ratio in ants have multiple evolutionary causes based on different within-colony conflicts. Biol Lett 5:713-716 doi:10.1098/rsbl.2009.0295
- Longino JT (2006) A taxonomic review of the genus *Myrmelachista* (Hymenoptera: Formicidae) in Costa Rica. Zootaxa 1141:1-54
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 164:1567-1587
- Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43:258-275
- Rathcke B, Lacey EP (1985) Phenological patterns of terrestrial plants. Annu Rev Ecol S 16:179-214
- Salas-Lopez A, Talaga S, Lalagüe H (2016) The discovery of devil's gardens: an ant-plant mutualism in the cloud forests of the Eastern Amazon. J Trop Ecol 32:264-268 doi:<https://doi.org/10.1017/S0266467416000195>
- Schultner E, Saramäki J, Helanterä H (2016) Genetic structure of native ant supercolonies varies in space and time. Mol Ecol 25:6196-6213

- Solomon SE, Bacci Jr M, Martins Jr J, Vinha GG, Mueller UG (2008) Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. PLoS one 3:e2738
- Sündstrom L, Keller L, Chapuisat M (2003) Inbreeding and sex-biased gene flow in the ant *Formica exsecta*. Evolution 57:1552-1561
- Takezaki N, Nei M, Tamura K (2014) POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. Mol Biol Evol 31:1622-1624
- Torres JA, Snelling RR, Canals M (2001) Seasonal and nocturnal periodicities in ant nuptial flights in the tropics (Hymenoptera: Formicidae). Sociobiology 37:601-626
- Whitlock MC (2011) G'_{ST} and D do not replace F_{ST} . Mol Ecol 20:1083-1091
- Wolda H (1978) Seasonal fluctuations in rainfall, food and abundance of tropical insects. J Anim Ecol 47:369-381 doi:10.2307/3789

Figure Legends

Fig. 1. Map of the sampling locations. Samples were collected from 43 devil's gardens in 12 different localities in Peru. Each dot represents a locality. The inset zooms in on the box on the main map.

Fig. 2. Seasonality of alate production. Count of alates produced by devil's gardens every month over one year. Symbols are mean numbers of queens (red squares), gynes (yellow triangles), gyne pupae (blue diamonds), males (green circles), and male pupae (pink upside-down triangles). Whiskers are standard errors.

Fig. 3. Sex-ratio. Number of devil's garden trees as a function of the proportion of females among the alates and alate pupae produced (panel a). Panel b also includes de-alate queens in the calculation of female proportion.

Fig. 4. Relatedness between ants within two gardens. Relatedness between singly-mated queens, the males that inseminated them and the workers as calculated with RELATEDNESS within garden 030 (panel a) and garden 045 (panel b). The x axis represents the Px sets as defined by Queller & Goodnight (1989), i.e. the set of individuals the program summed over. Colours account for the Py sets, i.e. the "partners" of the Px individuals. Dots show mean relatedness and error bars show standard errors of mean relatedness.

Fig. 5. Isolation by distance. Pairwise F_{ST} between devil's gardens as a function of geographic distance between devil's gardens. The line is a simple linear regression with the 95% confidence interval shaded, for visualization purposes only.

Fig. 6. Country-wide genetic clustering. Panel a) Neighbour-joining tree based on pairwise F_{ST} indexes between devil's gardens with 500 bootstrap iterations. Branch tips are labelled with devil's garden codes followed by abbreviations of sampling site names between parentheses. Site names are abbreviated as follows: AM, Allpahuayo-Mishana; GH, Genaro

Herrera; LP, Las Piedras; MS, Madre Selva; Ne, Neshuya; NI, Nuevo Israel; Pa, Panguana; Qu, Quimpichiari; Sa, Sabalillo; SC, Santa Cruz; SU, Santa Ursula; Wa, Wasai. See Table 1 for descriptive data on individual devil's gardens at each location, and Table S1 for exact geographical distances between devil's gardens. Black dots show branch support higher than 50. Panel b) Discriminant Analyses of Principal Components (DAPC) showing the differentiation among the 43 devil's gardens sampled across Peru. The DAPC was created with the *adeigenet* package in R with a total of 34 principal components, accounting for 89.9% of the total variance. Horizontal and vertical axes represent the first and second discriminant functions, respectively, accounting for 19.6% of the variance explained by the full analysis. Each colour represents a different location according to a gradient North-South, from blue to red.

Fig. 7. Madre Selva genetic clustering. Panel a) Map of the locations of the 18 devil's gardens sampled at Madre Selva, with longitude (degrees W) on the x-axis and latitude (degrees S) on the y-axis. Panel b) Discriminant Analyses of Principal Components showing the differentiation among the 18 devil's gardens sample at Madre Selva. The DAPC was created with the *adeigenet* package in R with a total of 30 principal components, accounting for 92.7% of the total variance. Horizontal and vertical axes represent the first and second discriminant functions, respectively, accounting for 32.9% of the variance explained by the full analysis. Each colour is a different devil's garden.

Supplementary material

Table S1. Geographical distance and genetic differentiation indexes between pairs of devil's gardens. Geographical distances are in the upper half-matrix (in km), and F_{ST} calculated using SPAGeDi 1.5a are in the lower half-matrix. All F_{ST} are significant with p -values < 0.001 except for the one indicated with an asterisk.

Fig. S2. Population structure. Results of the clustering analyses conducted with STRUCTURE. Panel (a): Delta K plot according to STRUCTURE HARVESTER. The x axis represents the different values of K tested (from 1 to 44), and the y axis the Delta K value. Panel (b): assignment plot according to CLUMPP. Each vertical bar represents an individual from left to right in the same order as in Table 1, and the y axis represents the percentage of assignment to clusters. Colours show the two clusters.