

## Quick Guidelines to Reviewing your Proof PDF below

1. Please mark your corrections in the proof PDF OR clearly describe corrections in a text file.

**DO NOT:** implement desired changes in the PDF or supply newly edited Word files, TEX files, or XML files.

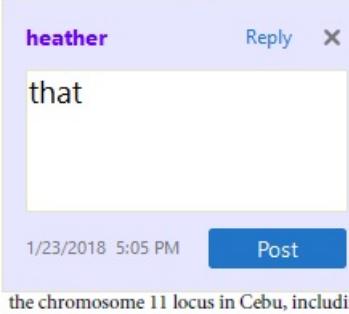
2. When marking the PDF, clearly show or describe requested changes. We must be able to read every correction in order to add changes to our master copy of the proof (see examples below).

### Example of correctly showing text deletion

Rhesus macaques are an ideal host species to study the behavioral bases of ~~dyadic and/or community-wide~~ bacterial sharing. They are biologically, socially and cognitively analogous to human societies (Cobb, 1976; Suomi, 2011). In nature, they live in large (approximate range: 20–150 individuals), multi-male-multi-female social groups, in which individuals

### Example of correctly showing text replacement

mice. Significance of *SIDT2* for TRIG in LCL was also replicated in GLGC at  $P = 1.1 \times 10^{-102}$  (Table 5). The similar effects and significance of *SIDT2* in knockout mice studies, the Cebu cohort analysis, and GLGC PrediXcan results indicate the increased expression levels. Therefore, variation in the regulation of *TRIG* levels. This analysis showed that *BACE1* has the largest effect, followed by *CEP164* contributing smaller effects. *BACE1* and *CEP164* are located on chromosome 11q23.3 locus in Asian populations. Our study have been previously associated with *TRIG* at the 11q23.3 locus (Rasmussen-Torvik et al., 2012). In our study, we found additional genes at the chromosome 11 locus in Cebu, including genes with predictive SNPs located hundreds



2. Please check the entire article and **review all critical information carefully**. Examples (this list is not comprehensive):

- Title
- Author names & affiliations
- Names and email addresses of corresponding author(s)
- Figures
- Tables

3. Please answer all queries listed at the end of the article.

### NOTES

We ask that you return proofs in **3 business days** but please do not rush. Take all the time that you need to review this thoroughly and submit all your corrections in a complete draft. ***We only accept one set of corrections.***

Any substantive changes introduced at this stage such as the removal or addition of new text, revising Supplemental Information files, and changing table or figure values **requires Academic Editor approval**. This may cause delays in publication.

Please only  
ANNOTATE  
the proof. Do  
not edit the  
PDF. If multiple  
authors will  
review this  
PDF, please  
return one file  
containing all  
corrections.

Submitted 25 April 2018  
Accepted 10 January 2019  
Published TBD

Corresponding author  
Matthew K. Lau,  
matthewklau@fas.harvard.edu

Academic editor  
Nigel Andrew

Additional Information and  
Declarations can be found on  
page 14

DOI 10.7717/peerj.6447

© Copyright  
2019 Lau et al.

Distributed under  
Creative Commons CC-BY 4.0

## OPEN ACCESS

# Draft *Aphaenogaster* genomes expand our view of ant genome size variation across climate gradients

Matthew K. Lau<sup>1</sup>, Aaron M. Ellison<sup>1</sup>, Andrew Nguyen<sup>2,3</sup>, Clint Penick<sup>4</sup>, Bernice DeMarco<sup>5</sup>, Nicholas J. Gotelli<sup>3</sup>, Nathan J. Sanders<sup>6</sup>, Robert R. Dunn<sup>7</sup> and Sara Helms Cahan<sup>3</sup>

[Q1]

<sup>1</sup> Harvard Forest, Harvard University, Petersham, MA, USA

<sup>2</sup> Department of Entomology and Nematology, University of Florida, Gainesville, FL, USA

<sup>3</sup> Department of Biology, University of Vermont, Burlington, VT, USA

<sup>4</sup> The Biomimicry Center, Arizona State University, Tempe, AZ, USA

<sup>5</sup> Smithsonian Institution, Washington, DC, USA

<sup>6</sup> Environmental Program, Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA

<sup>7</sup> Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA

## ABSTRACT

Given the abundance, broad distribution, and diversity of roles that ants play in many ecosystems, they are an ideal group to serve as ecosystem indicators of climatic change. At present, only a few whole-genome sequences of ants are available (19 of >16,000 species), mostly from tropical and sub-tropical species. To address this limited sampling, we sequenced genomes of temperate-latitude species from the genus *Aphaenogaster*, a genus with important seed dispersers. In total, we sampled seven colonies of six species: *Aphaenogaster ashmeadi*, *Aphaenogaster floridana*, *Aphaenogaster fulva*, *Aphaenogaster miamiana*, *Aphaenogaster picea*, and *Aphaenogaster rudis*. The geographic ranges of these species collectively span eastern North America from southern Florida to southern Canada, which encompasses a latitudinal gradient in which many climatic variables are changing rapidly. For the six genomes, we assembled an average of 271,039 contigs into 47,337 scaffolds. The *Aphaenogaster* genomes displayed high levels of completeness with 96.1% and 97.6% of Hymenoptera BUSCOs completely represented, relative to currently sequenced ant genomes which ranged from 88.2% to 98.5%. Additionally, the mean genome size was 370.5 Mb, ranging from 310.3 to 429.7, which is comparable to that of other sequenced ant genomes (212.8–396.0 Mb) and flow cytometry estimates (210.7–690.4 Mb). In an analysis of currently sequenced ant genomes and the new *Aphaenogaster* sequences, we found that after controlling for both spatial autocorrelation and phylogenetics ant genome size was marginally correlated with sample site climate similarity. Of all examined climate variables, minimum temperature, and annual precipitation had the strongest correlations with genome size, with ants from locations with colder minimum temperatures and higher levels of precipitation having larger genomes. These results suggest that climate extremes could be a selective force acting on ant genomes and point to the need for more extensive sequencing of ant genomes.

---

**Subjects** Ecology, Genomics

**Keywords** Ants, Genomics, Ecology, Climate change, Evolution, Adaptation

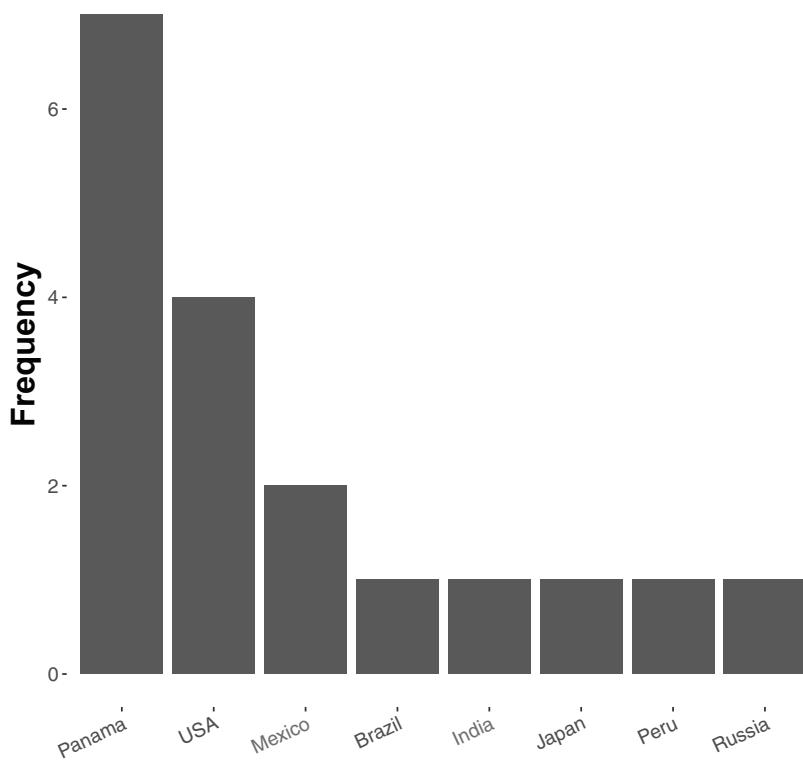
## INTRODUCTION

Understanding how terrestrial ecosystems will respond to ongoing shifts in climatic variables, such as temperature and precipitation, will improve our ability to manage communities and mitigate impacts of climatic change. The mean global temperature is currently on track to meet or exceed that predicted by the most extreme forecasting models ([Brown & Caldeira, 2017](#)). Climatic change is also pushing local conditions outside the boundaries of historic ranges, potentially leading to combinations of species or entire ecosystems that have no contemporary analogs ([Burrows et al., 2014](#)). As climate-driven impacts on evolutionary responses are likely to occur over contemporary time scales, with the potential for ecological and evolutionary dynamics to affect communities and ecosystem processes ([Rowntree, Shuker & Preziosi, 2011](#); [Des Roches et al., 2017](#)), there is a need for a comprehensive study of the genetic basis of species' responses to climate ([Parmesan, 2006](#)).

The biodiversity of most terrestrial systems is great enough to be intractable to study in its entirety. To deal with this, researchers often study “indicator” species whose responses to environmental change are broadly representative of a much wider range of taxa ([Siddig et al., 2016](#)). Ants (Formicidae), in particular, are widely used as indicator taxa ([Agosti et al., 2000](#)) as they play key roles in community dynamics and ecosystem processes, including key interactions, such as seed dispersal and the movement of soil via colony construction ([Del Toro, Ribbons & Pelini, 2012](#)). Ants also are responsive to changes in temperature and other climatic variables via individual responses, changes in colony structure and community assembly ([Kaspari et al., 2015](#); [Spicer et al., 2017](#); [Diamond et al., 2017](#); [Diamond & Chick, 2018](#)).

Multiple studies support the perspective that a more complete knowledge of ant genetics will increase our understanding of ant responses to environmental change ([Diamond et al., 2012](#); [Nygaard & Wurm, 2015](#); [Stanton-Geddes et al., 2016](#); [Boomsma et al., 2017](#); [Penick et al., 2017](#)). Studies of ant genomes have shed light on the evolution and social organization of ants ([Libbrecht et al., 2013](#)). One promising avenue is the possibility of genome size as an adaptive trait in ants. Recent observational studies have reported biogeographic patterns in genome size in arthropod taxa, for example, Crustacea ([Hultgren et al., 2018](#)), and patterns in insect genomes suggest that climate may constrain genome size with cold temperatures possibly selecting for larger genome sizes ([Mousseau, 1997](#); [Petrov, 2001](#); [Alfsnes, Leinaas & Hessen, 2017](#)). Specific to ants, previous research into genome size variation using flow cytometry found that ants have small genomes relative to other insect taxa and that their genomes display large variation across subfamilies with patterns indicative of both gradual and rapid evolution in genome size ([Tsutsui et al., 2008](#)).

At present relatively few ant species have been sequenced—20 in total, of which 19 are currently available in the NCBI Genome Database (accessed 18 December 2018,

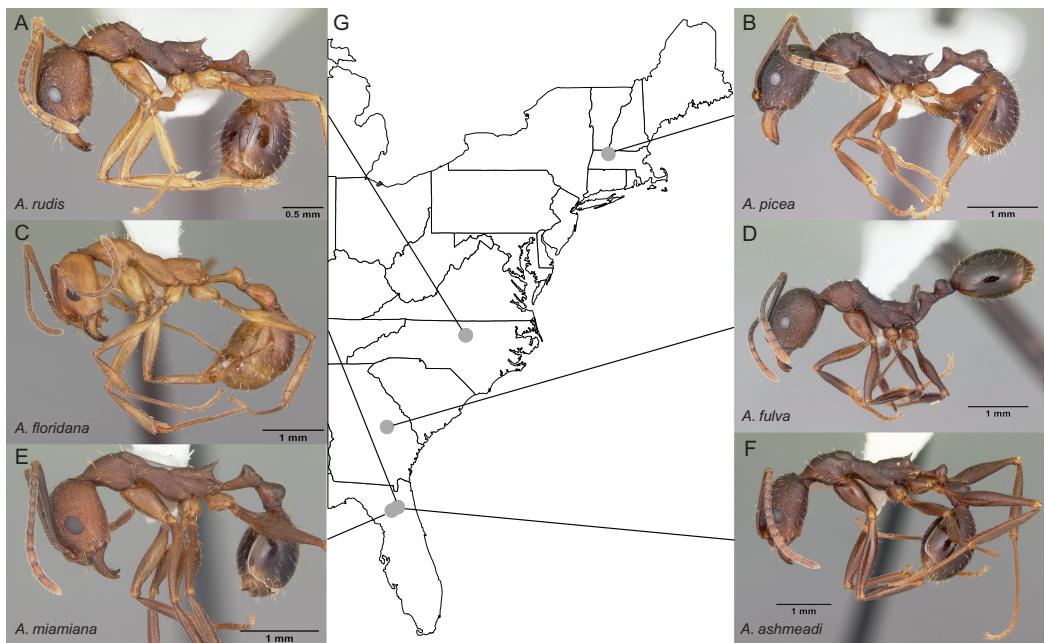


**Figure 1** Frequency distribution of currently sequenced ant genomes by geographic location. Number of ant (Formicidae) whole-genome sequences available in NCBI by country (accessed August 2018).

[Full-size](#) DOI: 10.7717/peerj.6447/fig-1

see Table S1). Of these, most are from tropical and subtropical assemblages (Fig. 1), and all but five represent unique genera (the exceptions being two species of *Atta* and three of *Trachymyrmex*). No species of *Aphaenogaster*, which are abundant ants that play key roles in the dispersal of understory plant species in North America and temperate Asia, have yet been sequenced. Previous studies have also shown that *Aphaenogaster* species' ecological and physiological responses to climatic change appear to depend both on species identity and on the geographic region in which climatic change occurs (Warren & Chick, 2013; Stanton-Geddes et al., 2016).

To increase the number of genomes of temperate-zone ant species, we sequenced the genomes of *Aphaenogaster* species. We conducted whole genome sequencing for six species: *Aphaenogaster ashmeadi*, *Aphaenogaster floridana*, *Aphaenogaster fulva*, *Aphaenogaster miamiana*, *Aphaenogaster picea*, and *Aphaenogaster rudis*. These species were collected from across a broad biogeographic gradient spanning 10° of longitude and 12° of latitude. We also conducted an initial exploration of biogeographic patterns in ant genome sequences, focusing on genome size. To do this we analyzed the newly collected *Aphaenogaster* sequences together with all publicly available ant whole genome sequences. We present the newly sequenced *Aphaenogaster* genomes and investigate biogeographic (i.e., location and climate) related patterns of ant genomes using all currently sequenced ant genomes.



**Figure 2** Sample location map with overlaid photos of sequenced *Aphaenogaster* species. We sampled seven colonies representing six species of *Aphaenogaster*, including (A) *A. rudis*, (B) *A. picea*, (C) *A. floridana*, (D) *A. fulva*, (E) *A. miamiana*, and (F) *A. ashmeadi* from (G) sampling locations across eastern North America (see Table 1). All photos by April Noble (available from <http://www.antweb.org>).

Full-size DOI: 10.7717/peerj.6447/fig-2

## MATERIALS AND METHODS

### Sampling and whole-genome sequencing and assembly

Entire colonies of the six *Aphaenogaster* species were collected by A. Nguyen and C. Penick from field sites in eastern North America (Fig. 2). Ants were identified to species and voucher specimens have been deposited at the Museum of Comparative Zoology, Harvard University. Individuals from each colony were isolated from nest material and debris, weighed, placed in 50 ml Falcon centrifuge tubes, and immediately flash frozen in a  $-80^{\circ}\text{C}$  freezer. Colony weights were: 794 mg (*Aphaenogaster ashmeadi*), 652 mg (*Aphaenogaster floridana*), 520 mg (*Aphaenogaster fulva*), 749 mg (*Aphaenogaster picea*), 862 mg (*Aphaenogaster miamiana*), 280 mg (*Aphaenogaster rudis* 1), and 236 mg (*Aphaenogaster rudis* 2).

Whole colony DNA was used to have sufficient concentrations for sequencing. DNA was then extracted from each colony using methods developed previously for genomic sequencing of whole colonies of colonial mosquitos (*Anopheles* spp.) (Neafsey *et al.*, 2010) and sequenced using an Illumina HiSeq 2500 at the Broad Institute (Cambridge, MA, USA). A combination of fragment and jump sequences were used to generate higher quality, long sequence reads.

Raw sequences were processed to remove chimeric and contaminant sequences, screened for contaminants by BLAST searches (using *blastn*) to identify sequences with likely matches to non-target species (primarily *Wolbachia* and *Mycoplasma*), and assembled using ALLPATHS-LG (version r48559) (Gnerre *et al.*, 2011).

**Table 1** Climate variables for colony sample sites. Climate are 30 year normal values (1970–2000) for minimum temperature of the coldest month ( $T_{min}$ ), maximum temperature of the warmest month ( $T_{max}$ ), and total precipitation (Precip) from the WorldClim database accessed on 08 August 2018.

|                                | Lat   | Lon    | $T_{min}$ (C) | $T_{max}$ (C) | Precip (mm) |
|--------------------------------|-------|--------|---------------|---------------|-------------|
| <i>Aphaenogaster ashmeadi</i>  | 29.79 | -82.03 | 5.80          | 32.70         | 1,314       |
| <i>Aphaenogaster floridana</i> | 29.79 | -82.03 | 5.80          | 32.70         | 1,314       |
| <i>Aphaenogaster fulva</i>     | 32.69 | -82.51 | 1.30          | 33.30         | 1,155       |
| <i>Aphaenogaster miamiana</i>  | 29.66 | -82.30 | 5.90          | 32.80         | 1,322       |
| <i>Aphaenogaster picea</i>     | 42.60 | -72.58 | -12.40        | 28.30         | 1,122       |
| <i>Aphaenogaster rudis1</i>    | 36.02 | -78.98 | -2.70         | 31.50         | 1,164       |
| <i>Aphaenogaster rudis2</i>    | 36.02 | -78.98 | -2.70         | 31.50         | 1,164       |

Additional assembly processing using PILON (version 1.13) (Walker *et al.*, 2014) was applied to reduce base-call errors and gaps in coverage. On average, across all seven genomes, PILON reduced coverage gaps by 3.1% or 3.9 Mb. GAEMR (<http://www.broadinstitute.org/software/gaemr/>) software produced summary statistics of the final assembled genomes. Once assembled, repeat regions in the *Aphaenogaster* genomes were detected and masked using Repeatmasker (version 4.0.5 Institute for Systems Biology). Genome completeness was assessed for all ant genome genome assemblies using BUSCO version 3.0.2 (Waterhouse *et al.*, 2018) with the Hymenoptera BUSCO lineage (ODB9 from February 13, 2017) with the honey bee starting genome and HMMER version 3.1b1, BLAST+ version 2.2.31 and AUGUSTUS version 3.0.3 dependencies. For each genome we ran BUSCO on the final contaminant removed, repeat-masked *Aphaenogaster* assemblies and the most recent published assembly for the NCBI published genomes.

### Analysis of genomes along climate gradients

After masking repeat regions, we applied MASH distance (Ondov *et al.*, 2016) to measure pairwise dissimilarity of genomic sequences. The MASH method extends a data compression and dimensionality-reduction algorithm to generate estimates of sequence similarity with low computational overhead. Briefly, the pairs of genomic sequences were pre-processed into sets of k-mers of size 21 with the size of the non-redundant hashes retained set to 1,000. These settings have been demonstrated to provide good representation of genomic similarity with minimal computational costs (Ondov *et al.*, 2016). These sets were then used to estimate the Jaccard similarity coefficient (the ratio of shared k-mers to total k-mers) of subsampled k-mer pairs of genomes. This unbiased estimate of the Jaccard similarity ( $J$ ) was then used to calculate the dissimilarity of the two genomes ( $D$ ) as  $D = 1 - J$ . All Jaccard similarity estimates had  $p$ -values less than  $10^{-14}$ , which is below the recommended  $10^{-3}$  probability of observing values of  $J$  due to chance.

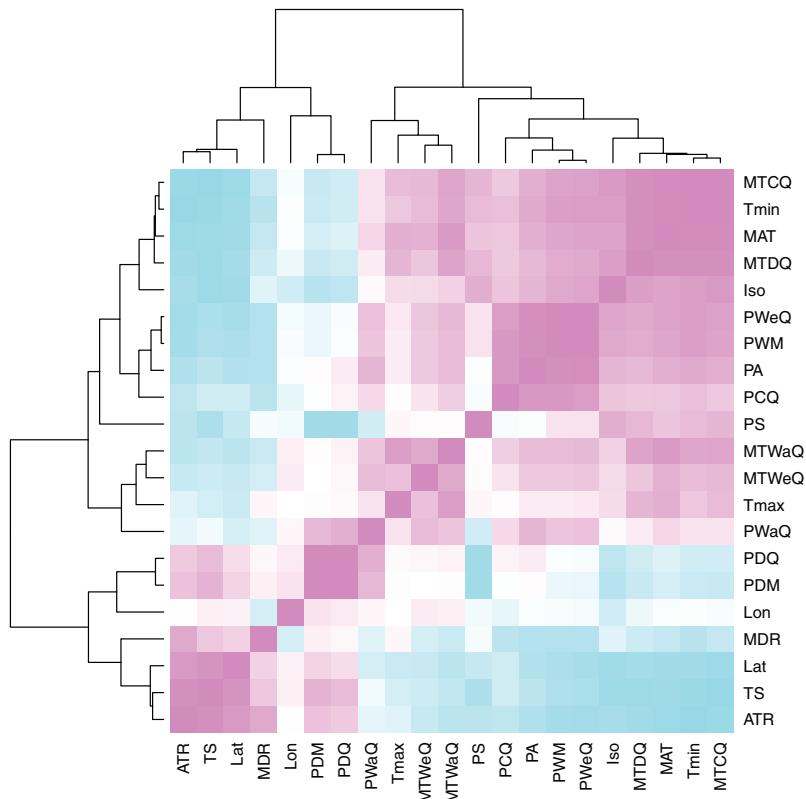
We used multivariate correlation analyses to examine biogeographic patterns of ant genomes. Mantel tests of multivariate correlation of distance matrices were used to examine correlations among ant genomes and climate variables. Specifically, we used

directional ( $H_0$ : Mantel  $r \leq 0$ ) partial mantel tests, which calculate the correlation between two distance matrices while controlling for the covariance with other matrices (Goslee & Urban, 2007). First, we examined the correlations between genomic similarity (MASH distance), whole-genome size similarity (Euclidean distance of assembly size in total base pairs) and climate variables (also using Euclidean distance). Via partial Mantel tests, we were able to isolate the correlation between genome size and climate by controlling for spatial autocorrelation and potential phylogenetic patterns by including geodesic and MASH distances as terms.

We obtained previously sequenced ant whole genome and climate data from a publicly available databases. Whole genome sequences for ants were obtained from the NCBI Genome database (accessed August 2018, see Table S1). Climatic variables for each sampling location was obtained from the WorldClim database (version 2.0) at a 2.5 arc minute spatial resolution from the years 1970 to 2000 (Fick & Hijmans, 2017). Although used in the previous analyses of ant genomes, two species, (*Wasmannia auropunctata* and *Monomorium pharaonis*), which did not have published location information, were excluded from biogeographic analyses.

Using a permutational multivariate analysis of variance (PerMANOVA) procedure, we parsed the individual variables that were correlated with both genome size and MASH similarity. PerMANOVA is a flexible multivariate analog of ANOVA that permits the use of a wider set of similarity metrics to be used for the response matrix (Anderson, 2001), such as the MASH distance. We ran a total of 10,000 permutations of the original distance matrices for each statistical permutation procedure. We chose a subset of all possible climate variables available via WorldClim for this analysis. A visual inspection of the sampled climate variable correlations indicated that the primary climate variables, mean annual temperature (MAT), minimum temperature of the coldest month (Tmin), maximum temperature of the hottest month (Tmax), annual precipitation (PA) and precipitation seasonality (PS), represented the majority of climate variation (Fig. 3). Based on this, we only included these variables, along with latitude and longitude coordinates, as factors in the PerMANOVAs.

It is important to note that we are using assembly length as an indicator of genome size. As genome size estimates are generally used to set assembly size targets for whole genome sequencing efforts (Hare & Johnston, 2011), we expect there to be a high degree of correlation between assembly size and genome size. Also, as a test of the potential relationship between assembly size and true genome size, we examined the correlation between the average assembly sizes of ant genera that overlapped with flow cytometry estimates of those published in Tsutsui et al. (2008). We applied leave-one-out outlier detection (Cook & Weisberg, 1982) to least squares regression analysis examining the predictive potential of assembly length and flow cytometry size estimates at the genus level and found one strong outlier (Fig. S1), *Dinoponera quadriceps*, whose assembly size was less than half the flow cytometry estimate (259.7 vs. 554.7 Mb). After removing this species from the regression analysis we found a strong, significant relationship between assembly length and genome size ( $R^2 = 0.76$ ,  $F_{1,5} = 15.99$ ,  $p$ -value = 0.010), which supports the use of assembly length as an indicator of genome size. Based on this, we did not include *D. quadriceps* in any of the analyses examining patterns of assembly length.



**Figure 3 Heatmap of climate variable intercorrelations.** Heatmap of Pearson correlations among climate variables. Cells in the heatmap are colored by the correlation between the two variables that intersect at that location ranging from blue =  $-1$  to white =  $0$  to pink =  $1$ . The variables are arrayed by hierarchical clustering of the correlations, as shown by the dendograms on the top and left side. For variable descriptions see Table 2.

Full-size DOI: 10.7717/peerj.6447/fig-3

## Data, computation and statistics

The raw and assembled genome sequences are currently archived at Harvard Forest (Petersham, MA, USA) and in NCBI's genome database (Genome Accessions [NJKR00000000–NJRQ00000000](#) and BioSample Accessions [SAMN06892346–SAMN06892352](#)). Genomic distance (MASH) computations were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. All analyses were conducted in [R Core Team \(2017\)](#). Analytical scripts for the project have been versioned and archived ([DOI: 10.5281/zenodo.1341982](#)) and are available online at <https://zenodo.org/record/1341982>. We used the *vegan* ([Oksanen et al., 2016](#)) and *ecodist* ([Goslee & Urban, 2007](#)) packages in R for multivariate analyses.

## RESULTS

### Genome quality and composition

DNA extractions yielded substantial amounts of high quality DNA with concentrations and quality scores ranging from  $3.45\text{--}5.39$  to  $4.05\text{--}4.27 \text{ ng}\mu\text{L}^{-1}$ , respectively. All genome assemblies displayed good coverage, with an average of 70% across all genomes (Table 3). Across all species, the length of the shortest contig at 50% of the genome (i.e., N50)

**Table 2** WorldClim variables, abbreviations, and numbers for the climate variables used in the analysis of size and MASH similarity of ant genomes.

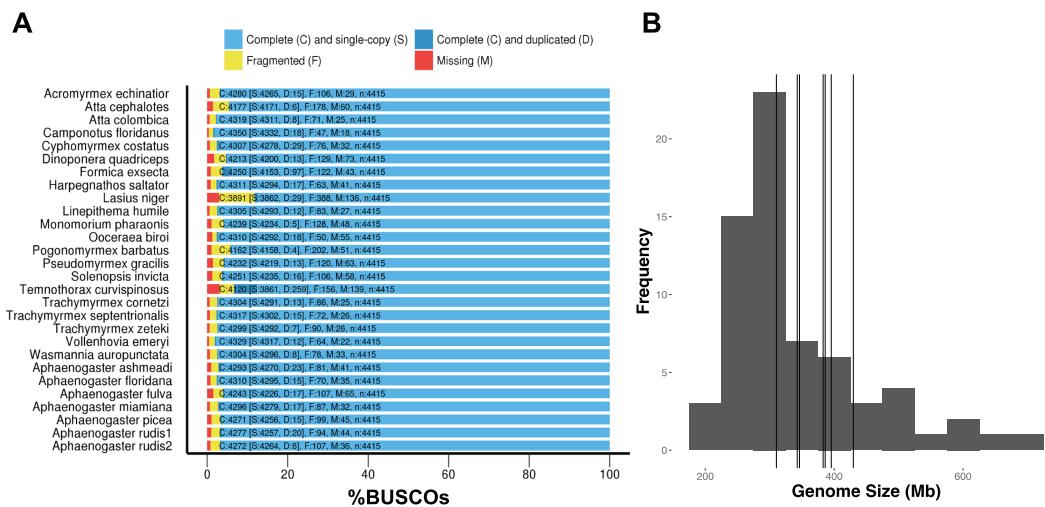
| WorldClim variable                             | BIO number |
|--|------------|
| Annual mean temperature (MAT)                  | BIO1       |
| Mean diurnal range (MDR)                       | BIO2       |
| Isothermality (Iso)                            | BIO3       |
| Temperature seasonality (TS)                   | BIO4       |
| Max temperature of warmest month ( $T_{max}$ ) | BIO5       |
| Min temperature of coldest month ( $T_{min}$ ) | BIO6       |
| Temperature annual range (ATR)                 | BIO7       |
| Mean temperature of wettest quarter (MTWeQ)    | BIO8       |
| Mean temperature of driest quarter (MTDQ)      | BIO9       |
| Mean temperature of warmest quarter (MTWaQ)    | BIO10      |
| Mean temperature of coldest quarter (MTCQ)     | BIO11      |
| Annual precipitation (PA)                      | BIO12      |
| Precipitation of wettest month (PWM)           | BIO13      |
| Precipitation of driest month (PDM)            | BIO14      |
| Precipitation seasonality (PS)                 | BIO15      |
| Precipitation of wettest quarter (PWeQ)        | BIO16      |
| Precipitation of driest quarter (PDQ)          | BIO17      |
| Precipitation of warmest quarter (PWaQ)        | BIO18      |
| Precipitation of coldest quarter (PCQ)         | BIO19      |

was 18,864 bases; average assembly GC content was 38.18%; and average genome size was 370.45 Mb. Using GAEMR's BLAST feature to conduct a search of the contigs against the NCBI's nucleotide sequence database, we discovered that 38.98% and 22.04% of the top hits were "ant" and *Aphaenogaster*, respectively. Using the Hymenoptera BUSCO set, the *Aphaenogaster* genomes displayed a high degree of completeness. Between 96.1% and 97.6% with an average of  $97.0\% \pm 0.2$  (S.E.) of Hymenoptera BUSCOs were completely represented. The *Aphaenogaster* genomes were among the more complete genomes, relative to the other sequenced ant genomes from NCBI, which ranged from 88.2% to 98.5% complete Hymenoptera BUSCOs with a an average of  $96.3\% \pm 0.5$  (S.E.) (Fig. 4A). Also, the sizes of the *Aphaenogaster* genomes were within the range of other ant genomes based on size from both flow cytometry (Tsutsui *et al.*, 2008) and the previously sequenced ant genomes available in NCBI (Fig. 4B).

Using the MASH genomic distances, we observed patterns of genomic similarity that were in line with expectations of ant relatedness. Sequences formed groups that corresponded with subfamilies (Fig. 5). *Aphaenogaster* clustered with other genera from the Myrmicinae and, in general, subfamily level clustering tended to follow previously observed patterns of subfamily relatedness (Bolton, 2006; Moreau, 2006; Ward, 2014). The *Aphaenogaster* sequences formed a single cluster containing only *Aphaenogaster* species and displayed intra-generic levels of genomic variance comparable to other genera (e.g., *Trachymyrmex* spp.). The separation of the two *Aphaenogaster rudis* species was initially surprising, as these two samples were collected at the same site

**Table 3** Sequencing statistics for the genomes of the sequenced colonies of *Aphaenogaster*.

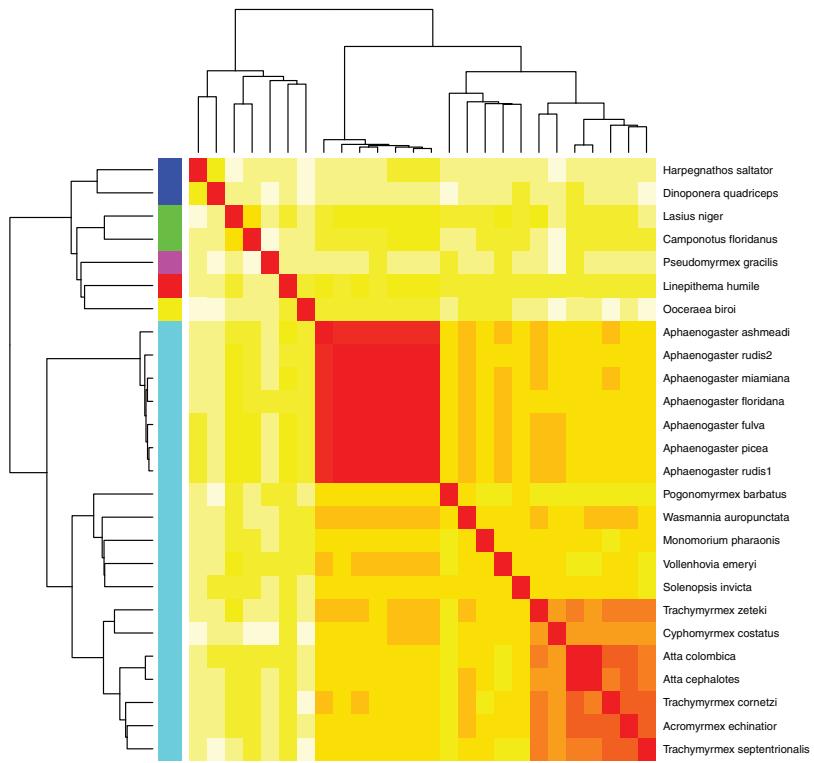
|                            | <i>A. ashmeadi</i> | <i>A. floridana</i> | <i>A. fulva</i> | <i>A. miamiana</i> | <i>A. picea</i> | <i>A. rudis1</i> | <i>A. rudis2</i> |
|----------------------------|--------------------|---------------------|-----------------|--------------------|-----------------|------------------|------------------|
| Total scaffold length (Mb) | 310.33             | 382.86              | 346.13          | 342.64             | 386.04          | 395.41           | 429.70           |
| Coverage (%)               | 81.46              | 71.88               | 70.70           | 77.40              | 67.47           | 66.49            | 65.59            |
| Scaffold N50 (bp)          | 336,807.00         | 439,114.00          | 255,328.00      | 351,517.00         | 322,984.00      | 300,103.00       | 269,776.00       |
| Scaffolds                  | 5,087.00           | 6,422.00            | 7,031.00        | 6,920.00           | 6,808.00        | 7,404.00         | 7,665.00         |
| Max gap (bp)               | 13,070.00          | 15,108.00           | 12,104.00       | 11,453.00          | 14,952.00       | 18,586.00        | 24,564.00        |
| Captured gaps              | 26,350.00          | 30,858.00           | 32,881.00       | 28,801.00          | 36,417.00       | 34,062.00        | 34,313.00        |
| Total gap length (Mb)      | 57.69              | 107.89              | 101.40          | 77.64              | 125.15          | 131.71           | 148.75           |
| Total contig length (Mb)   | 252.64             | 274.96              | 244.73          | 265.00             | 260.90          | 263.70           | 280.95           |
| Contig N50 (bp)            | 21,677.00          | 23,448.00           | 15,753.00       | 20,738.00          | 15,440.00       | 15,622.00        | 18,941.00        |
| Contigs                    | 31,437.00          | 37,280.00           | 39,912.00       | 35,721.00          | 43,225.00       | 41,466.00        | 41,978.00        |
| Assembly GC (%)            | 38.27              | 38.03               | 38.39           | 38.21              | 38.32           | 38.25            | 37.88            |
| Contaminants (%)           | 0.30               | 0.24                | 0.02            | 0.26               | 1.14            | 1.25             | 0.61             |



**Figure 4** Frequency distribution of sequenced assembly and flow cytometry estimated genome sizes with overlaid *Aphaenogaster* genome sizes. The *Aphaenogaster* genomes displayed a high degree of completeness and were within the size range of previously published ant genomes. (A) Stacked bar plot of the BUSCO metrics including the number and percentage of complete Hymenoptera (ODB9) BUSCOs (C), also including both single (S) and duplicated (D) copies, as well as fragmented (F) and missing (M) out of 4,415 BUSCOs for the *Aphaenogaster* and previously published ant genomes available from NCBI (accessed August 2018). (B) Frequency distribution of previously published genome size estimates using flow cytometry from [Tsutsui et al. \(2008\)](#) and those available in the NCBI Genomes database. Vertical lines identify the sizes of the *Aphaenogaster* assemblies (see Table 3).

Full-size DOI: 10.7717/peerj.6447/fig-4

(Duke Forest, NC, USA) and were identified as the same species based on their morphological characteristics ([Ellison, 2012](#); [DeMarco & Cognato, 2016](#)). However, two recent studies of targeted gene regions have demonstrated the polyphyletic nature of *Aphaenogaster rudis*. One study of the evolution of the subfamily Myrmicinae observed that the genus as a whole could be split into at least four different lineages ([Ward et al., 2015](#)). Another, more detailed study of the genus in North America found that



**Figure 5 Heatmap of MASH genomic distances for all sequenced ant genomes.** Heatmap of the MASH genomic distances of the *Aphaenogaster* species that we sampled together with other ant species in NCBIs. Heat colors shown in the central matrix range from high (white = 1) through moderate (orange = 0.5) to low (red = 0) genomic distance; the diagonal is entirely red because it illustrates the distance of each sequence to itself. The cladograms on the left and top show hierarchical clustering of the genomes. Colors shown to the left of the matrix indicate ant subfamilies: *Ponerinae* (dark blue), *Formicinae* (green), *Pseudomyrmecinae* (pink), *Dolichoderinae* (red), *Dorylinae* (yellow), *Myrmicinae* (light blue).

Full-size DOI: [10.7717/peerj.6447/fig-5](https://doi.org/10.7717/peerj.6447/fig-5)

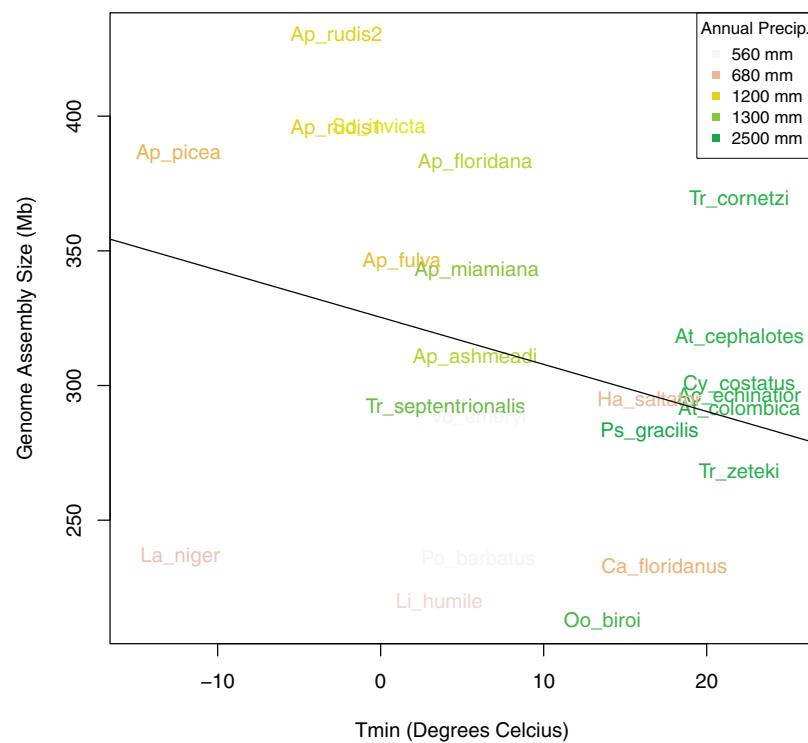
multiple individuals of *Aphaenogaster rudis* separated out into distinct groupings, each with other species, specifically, individuals of *Aphaenogaster rudis* from North Carolina (USA) were observed to form distinct clusters with individuals of *Aphaenogaster carolinensis*, *Aphaenogaster miamiana*, *Aphaenogaster lamellidens* and *Aphaenogaster texana* (DeMarco & Cognato, 2016).

### Biogeographic patterns of ant genomes

After controlling for both spatial autocorrelation and potential phylogenetic patterns, we found a marginally significant, positive correlation between ant assembly size similarity and climate similarity (Mantel  $R = 0.12$ ,  $p$ -value = 0.068). Although assembly size similarity and MASH genome similarity were not significantly correlated ( $p$ -value = 0.126), we included MASH as a covariate in addition to geodesic distance because previous research indicated that genome size is associated with phylogenetic relatedness (Alfsnes, Leinaas & Hessen, 2017). We found that different spatial and climatic variables were associated with the size similarity of ant genomes. Longitude but not latitude was a significant predictor of assembly size similarity (Table 4). Temperature of

**Table 4** PerMANOVA pseudo-F table for the analysis of the relationship between climate variables and ant assembly size similarity.

|                                 | df | SS        | MS        | Pseudo-F | R <sup>2</sup> | p-value |
|---------------------------------|----|-----------|-----------|----------|----------------|---------|
| <b>Assembly size similarity</b> |    |           |           |          |                |         |
| Lat                             | 1  | 2,428.30  | 2,428.30  | 1.70     | 0.03           | 0.2093  |
| Lon                             | 1  | 9,528.44  | 9,528.44  | 6.67     | 0.11           | 0.0165  |
| MAT                             | 1  | 147.67    | 147.67    | 0.10     | 0.00           | 0.7,531 |
| Tmin                            | 1  | 18,869.62 | 18,869.62 | 13.22    | 0.22           | 0.0029  |
| Tmax                            | 1  | 8,994.04  | 8,994.04  | 6.30     | 0.11           | 0.0230  |
| PA                              | 1  | 22,424.69 | 22,424.69 | 15.71    | 0.26           | 0.0011  |
| PS                              | 1  | 1,062.82  | 1,062.82  | 0.74     | 0.01           | 0.4132  |
| Residuals                       | 15 | 21,414.46 | 1,427.63  |          | 0.25           |         |
| Total                           | 22 | 84,870.04 |           |          | 1.00           |         |

**Figure 6** Bivariate plot of the correlation between sampling location minimum temperature of the coldest month (Tmin), annual precipitation (PA), and genome size. Bivariate plot showing the relationship between ant assembly size and minimum temperature of the coldest month (Tmin) with Annual Precipitation (PA) overlaid as colors. Ants from locations with lower minimum temperatures and annual precipitation tended to have larger genomes.

Full-size DOI: 10.7717/peerj.6447/fig-6

the coldest (*Tmin*) and hottest (*Tmax*) month and total PA, were significant predictors, but neither MAT nor PS were significant predictors of assembly size similarity. Overall, *Tmin* and PA had the strongest relationships with assembly size similarity with both exhibiting generally negative relationships with size (Fig. 6). When the newly sequenced *Aphaenogaster* genomes were excluded from the analysis, only PA

was a significant predictor of assembly size similarity ( $R^2 = 0.31$ ,  $p$ -value = 0.020 and see Table S2).

## DISCUSSION

We have produced seven draft whole-genome sequences of six species of ants in the genus *Aphaenogaster*. The addition of the *Aphaenogaster* sequences increases the breadth of global ant genomic sampling, as these are the first whole-genomes from a previously un-sequenced genus, adding to the sequences of the diverse “formicoid” clade, which contains 90% of all extant ant species (Ward, 2014). Our genomic sequences were comparable in quality to other ant and insect genomes and the patterns of genomic similarity were in line with expectations based on current ant systematics. With the addition of the new *Aphaenogaster* sequences, our initial biogeographic analysis revealed that ant genomes from more similar climates have more similarly sized genomes with minimum temperatures having the strongest correlation with genome size, which is consistent with the hypothesis that climate has been a force shaping ant genome size.

Although correlative, our genome analysis results are consistent with the hypothesis that ants from regions with more similar climates tend to have similar sized genomes. Previous studies have observed physiological and ecological responses of ants to climate gradients and shifting temperatures (Warren & Chick, 2013; Stanton-Geddes et al., 2016; Diamond et al., 2016, 2017; Nguyen et al., 2017; Helms Cahan et al., 2017; Penick et al., 2017) that could act as agents of selection or as environmental filters. For example, Warren & Chick (2013) found that cold, but not warm, temperatures limited shifts in the distributions of *Aphaenogaster picea* and *Aphaenogaster rudis*. Diamond et al., (2016) reported that the rate of colonization and occupancy of nests by *Aphaenogaster* species in a 5-year experimental warming study (Pelini et al., 2014) declined with temperature in the warm, southern study site (Duke Forest, NC, USA) but not in the cooler, northern study site (Harvard Forest, MA, USA). In addition to the direct impacts of climate, some studies support the importance for the indirect effects of climate via biotic interactions. For example, the distribution of the species *Atta texana* is limited by the cold-tolerance of its fungal symbiont, cultivars of the genus *Attamyces* (Mueller et al., 2011). The evolution of the ant-fungus relationship has led to reductions in some ant species ranges by cold temperatures.

Although we found support for increasing genome size with colder temperatures and PA, we are cautious to offer possible mechanisms for these patterns. In addition to these variables themselves being correlated, in general arthropod genome size appears to be influenced by a complex array of selection pressures. This is evidenced by the recent study by Alfsnes, Leinaas & Hessen (2017), which found that genome size patterns varied greatly among major arthropod taxa with high potential for different mechanisms affecting genome size. For example, insects displayed clear phylogenetic correlations with genome size while genome size patterns in crustaceans were nearly independent of phylogeny but strongly related to biogeographic gradients (e.g., decreasing genome size with increasing maximum observed latitude). In addition, Hultgren et al. (2018) found evidence for increasing genome size with latitude in crustaceans but not

[Q3]

decapods, adding another example of the potential complexity of genome size as an adaptive trait.

There is the potential for both direct and indirect selection for increased genome size in colder conditions. *Hessen et al. (2010)* has proposed that there is a complex set of relationships among genome size, developmental rate, cell size, and body size and that selection can act at different points in this causal network. One possible direct pathway is that increased expression of gene products that deal with cold stress could lead to larger genomes via whole genome or individual gene duplications *Dufresne & Jeffery (2011)*. Previous work with *Aphaenogaster* supports this hypothesis, as *Stanton-Geddes et al. (2016)* found that exposure to extreme cold induced expression of genes in the cold-climate *Aphaenogaster picea* more so than in *Aphaenogaster carolinensis* (a more southern, warm climate species). An example of a possible indirect pathway is that cold could select for increased body size to deal with heat-loss in cold conditions (*Brown et al., 2004*), which could lead to larger genome sizes. There is evidence of cold selecting for greater body size (i.e., Bergmann's Rule) in ants (*Heinze et al., 2003*; *Bernadou et al., 2016*), and some have hypothesized that increased body size could lead indirectly to increased genome size via increased cell size (*Ryan Gregory, 2005*); however, the most recent, broad analysis of genome size in ants (that we are aware of) did not find support for a relationship between ant genome size and body size after controlling for phylogenetic patterns (*Tsutsui et al., 2008*). Therefore, assuming that our analysis adequately controlled for phylogenetics, indirect selection on genome size via body size is not a likely explanation for our observed relationship between genome size and temperature.

It is important to keep in mind that the climate related genomic patterns observed in this study should be considered an initial view of possible biogeographic patterns in ant genomes. As the addition of these sequences had a marked impact on the statistical results of the climate analysis (see *Table S2*), we expect that further sequencing work will continue to enhance our understanding of the ecological genomics of ants. Also, these findings should be tested with additional sequencing efforts, as we could not control for several potentially important intercorrelated variables. Factors such as sampling bias and sequencing methodology (e.g., 454 vs Illumina) also varied among sequencing efforts, which could have contributed to some of the observed correlations with climate. We did not attempt to control for these factors statistically due to the limitations of the current ant genome sample size. Future work should methodologically and/or statistically control for such sources of variation in ant genomes as more sequences become available to elucidate clearer patterns and resolve underlying ecological and evolutionary mechanisms.

## CONCLUSION

Although we have increased the total number of sequenced ant genomes by over 30%, the total number of ant sequences analyzed here is still a relatively small sample ( $n = 26$ ) of the estimated  $>16,000$  ant species and subspecies ([www.antweb.org](http://www.antweb.org), accessed 20 August 2018). Efforts such as the global ant genomics alliance (*Boomsma et al., 2017*),

which aims to greatly increase the number of ant species sequenced from across the world, will provide additional resources for ecological genomics studies. Further work investigating the variation in genomic content and mapping of target coding regions from previous physiological ([Nguyen et al., 2017](#)), biochemical ([Helms Cahan et al., 2017](#)), and transcriptomic ([Stanton-Geddes et al., 2016](#)) studies of *Aphaenogaster* and other ant species will inform predictions of how these species, and the ecosystems that they inhabit, may respond to ongoing climatic change. For instance, determining the genomic factors underlying the temperature response of ant assemblages to climatic gradients ([Warren & Chick, 2013](#); [Diamond et al., 2016, 2017](#)) could provide useful insights into the response of these important organisms to non-analog ecosystem states and idiosyncratic community responses ([Bewick et al., 2014](#)). In addition, as species distribution models have been significantly improved by the inclusion of genetic information ([Ikeda et al., 2016](#)), an ecological genetics approach that couples ant genomic and ecologically relevant data will provide a useful window into the response of many terrestrial ecosystems to a changing climate.

## ACKNOWLEDGEMENTS

Thank you to the genome sequencing team at the Broad Institute (particularly, James Bochicchio, Sarah Young, Terrance Shay and Caroline Cusick) and to Manisha Patel at Harvard Forest's Torrey Lab for assistance in obtaining and properly storing ant colonies.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by a US National Science Foundation Dimensions of Biodiversity grant (DEB 11-36646) to Nathan J. Sanders, Robert R. Dunn, Aaron M. Ellison, Nicholas J. Gotelli and Sara Helms Cahan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosure

The following grant information was disclosed by the authors:  
US National Science Foundation Dimensions of Biodiversity grant: DEB 11-36646.

### Competing Interests

Aaron  Ellison is an Academic Editor for PeerJ.

### Author Contributions

- Matthew K. Lau conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Aaron M. Ellison conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

- Andrew Nguyen conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Clint Penick performed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Bernice DeMarco performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Nicholas J. Gotelli authored or reviewed drafts of the paper, approved the final draft.
- Nathan J. Sanders conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Robert R. Dunn conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Sara Helms Cahan conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The assembled *Aphaenogaster* whole genome sequences have been deposited in NCBI's Genomes Database with BioSample accessions ([https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample\\_bioproject&from\\_uid=6894596](https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample_bioproject&from_uid=6894596)) SAMN06894590–SAMN06894596 and Genome Accessions [NJRQ00000000](https://www.ncbi.nlm.nih.gov/nuccore/NJRQ00000000)–[NJRK00000000](https://www.ncbi.nlm.nih.gov/nuccore/NJRK00000000).

## Data Availability

The following information was supplied regarding data availability:

The raw sequences for the *Aphaenogaster* genomes are being deposited in NCBI's Sequence Read Archive (SRA).

Analytical code is available at <https://github.com/HarvardForest/apGenomes>

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.6447#supplemental-information>.

## REFERENCES

- Agosti D, Majer JD, Alonso LE, Schultz TR. 2000. *Standard methods for measuring and monitoring biodiversity*. Vol. 233. Smithsonian Institution Press.
- Alfsnes K, Leinaas HP, Hessen DO. 2017. Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecology and Evolution* 7(15):5939–5947 DOI [10.1002/ece3.3163](https://doi.org/10.1002/ece3.3163).
- Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26(1):32–46.
- Bernadou A, Römermann C, Gratiashvili N, Heinze J. 2016. Body size but not colony size increases with altitude in the holarctic ant, *Leptothorax acervorum*. *Ecological Entomology* 41(6):733–736 DOI [10.1111/een.12338](https://doi.org/10.1111/een.12338).

All supplemental information will be made available for download exactly as they were supplied. This link to the SI will only work when the article is published.

- Bewick S, Stuble KL, Lessard J-P, Dunn RR, Adler FR, Sanders NJ.** 2014. Predicting future coexistence in a North American ant community. *Ecology and Evolution* **4**(10):1804–1819 DOI [10.1002/ece3.1048](https://doi.org/10.1002/ece3.1048).
- Bolton B.** 2006. *Bolton's catalogue of ants of the world, 1758–2005*. Harvard University Press.
- Boomsma JJ, Brady SG, Dunn RR, Gadau J, Heinze J, Keller L, Moreau CS, Sanders NJ, Schrader L, Schultz TR, Sundström L, Ward PS, Wcislo WT, Zhang G.** 2017. The global ant genomics alliance (GAGA). *Myrmecological News* **25**:61–66.
- Brown PT, Caldeira K.** 2017. Greater future global warming inferred from Earth's recent energy budget. *Nature* **552**(7683):45–50 DOI [10.1038/nature24672](https://doi.org/10.1038/nature24672).
- Brown JH, Gillooly JF, Allen AP, Savage VM, West GB.** 2004. Toward a metabolic theory of ecology. *Ecology* **85**(7):1771–1789 DOI [10.1890/03-9000](https://doi.org/10.1890/03-9000).
- Burrows MT, Schoeman DS, Richardson AJ, Molinos JG, Hoffmann A, Buckley LB, Moore PJ, Brown CJ, Bruno JF, Duarte CM, Halpern BS, Hoegh-Guldberg O, Kappel CV, Kiessling W, O'Connor MI, Pandolfi JM, Parmesan C, Sydeman WJ, Ferrier S, Williams KJ, Poloczanska ES.** 2014. Geographical limits to species-range shifts are suggested by climate velocity. *Nature* **507**(7493):492–495 DOI [10.1038/nature12976](https://doi.org/10.1038/nature12976).
- Cook RD, Weisberg S.** 1982. *Residuals and influence in regression*. Chapman and Hall.
- Del Toro I, Ribbons RR, Pelini SL.** 2012. The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). *Myrmecological News* **17**:133–146.
- DeMarco BB, Cognato AI.** 2016. A multiple-gene phylogeny reveals polyphyly among eastern North American *Aphaenogaster* species (Hymenoptera: Formicidae). *Zoologica Scripta* **45**(5):512–520 DOI [10.1111/zsc.12168](https://doi.org/10.1111/zsc.12168).
- Des Roches S, Post DM, Turley NE, Bailey JK, Hendry AP, Kinnison MT, Schweitzer JA, Palkovacs EP.** 2017. The ecological importance of intraspecific variation. *Nature Ecology & Evolution* **2**:57–64.
- Diamond SE, Chick LD.** 2018. Thermal specialist ant species have restricted, equatorial geographic ranges: implications for climate change vulnerability and risk of extinction. *Ecography* **41**(9):1507–1509 DOI [10.1111/ecog.03264](https://doi.org/10.1111/ecog.03264).
- Diamond SE, Chick L, Penick CA, Nichols LM, Cahan SH, Dunn RR, Ellison AM, Sanders NJ, Gotelli NJ.** 2017. Heat tolerance predicts the importance of species interaction effects as the climate changes. *Integrative and Comparative Biology* **57**(1):112–120 DOI [10.1093/icb/icx008](https://doi.org/10.1093/icb/icx008).
- Diamond SE, Nichols LM, Pelini SL, Penick CA, Barber GW, Cahan SH, Dunn RR, Ellison AM, Sanders NJ, Gotelli NJ.** 2016. Climatic warming destabilizes forest ant communities. *Science Advances* **2**(10):e1600842 DOI [10.1126/sciadv.1600842](https://doi.org/10.1126/sciadv.1600842).
- Diamond SE, Sorger DM, Hulcr J, Pelini SL, Toro ID, Hirsch C, Oberg E, Dunn RR.** 2012. Who likes it hot? A global analysis of the climatic, ecological, and evolutionary determinants of warming tolerance in ants. *Global Change Biology* **18**(2):448–456 DOI [10.1111/j.1365-2486.2011.02542.x](https://doi.org/10.1111/j.1365-2486.2011.02542.x).
- Dufresne F, Jeffery N.** 2011. A guided tour of large genome size in animals: what we know and where we are heading. *Chromosome Research* **19**(7):925–938 DOI [10.1007/s10577-011-9248-x](https://doi.org/10.1007/s10577-011-9248-x).
- Ellison AM.** 2012. *A field guide to the ants of New England*. Yale University Press.
- Fick SE, Hijmans RJ.** 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12):4302–4315 DOI [10.1002/joc.5086](https://doi.org/10.1002/joc.5086).

- Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB.** 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **108**(4):1513–1518 DOI [10.1073/pnas.1017351108](https://doi.org/10.1073/pnas.1017351108).
- Goslee SC, Urban DL.** 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**(7):1–19 DOI [10.18637/jss.v022.i07](https://doi.org/10.18637/jss.v022.i07).
- Hare EE, Johnston JS.** 2011. Genome size determination using flow cytometry of propidium iodide-stained nuclei. *Methods in Molecular Biology* **772**(5):3–12 DOI [10.1007/978-1-61779-228-1\\_1](https://doi.org/10.1007/978-1-61779-228-1_1).
- Heinze J, Foitzik S, Fischer B, Wanke T, Kipyatkov VE.** 2003. The significance of latitudinal variation in body size in a holarctic ant, *Leptothorax acervorum*. *Ecography* **26**(3):349–355 DOI [10.1034/j.1600-0587.2003.03478.x](https://doi.org/10.1034/j.1600-0587.2003.03478.x).
- Helms Cahan S, Nguyen AD, Stanton-Geddes J, Penick CA, Hernández Y, DeMarco BB, Gotelli NJ.** 2017. Modulation of the heat shock response is associated with acclimation to novel temperatures but not adaptation to climatic variation in the ants *Aphaenogaster picea* and *A. rudis*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **204**:113–120 DOI [10.1016/j.cbpa.2016.11.017](https://doi.org/10.1016/j.cbpa.2016.11.017).
- Hessen DO, Jayasingh PD, Neiman M, Weider LJ.** 2010. Genome streamlining and the elemental costs of growth. *Trends in Ecology & Evolution* **25**(2):75–80 DOI [10.1016/j.tree.2009.08.004](https://doi.org/10.1016/j.tree.2009.08.004).
- Hultgren KM, Jeffery NW, Moran A, Gregory TR.** 2018. Latitudinal variation in genome size in crustaceans. *Biological Journal of the Linnean Society* **123**(2):348–359 DOI [10.1093/biolinnean/blx153](https://doi.org/10.1093/biolinnean/blx153).
- Ikeda DH, Max TL, Allan GJ, Lau MK, Shuster SM, Whitham TG.** 2016. Genetically informed ecological niche models improve climate change predictions. *Global Change Biology* **23**(1):164–176 DOI [10.1111/gcb.13470](https://doi.org/10.1111/gcb.13470).
- Kaspari M, Clay NA, Lucas JA, Revzen S, Kay AD, Yanoviak SP.** 2015. Thermal adaptation and phosphorus shape thermal performance in an assemblage of rainforest ants. *Ecology* **97**(4):1038–1047 DOI [10.1890/15-1225.1](https://doi.org/10.1890/15-1225.1).
- Libbrecht R, Oxley PR, Kronauer DJC, Keller L.** 2013. Ant genomics sheds light on the molecular regulation of social organization. *Genome Biology* **14**(7):212 DOI [10.1186/gb-2013-14-7-212](https://doi.org/10.1186/gb-2013-14-7-212).
- Moreau CS.** 2006. Phylogeny of the ants: diversification in the age of angiosperms. *Science* **101**(281):1249–1481. [Q8]
- Mousseau TA.** 1997. Ectotherms follow the converse to Bergmann's rule. *Evolution* **51**(2):630 DOI [10.2307/2411138](https://doi.org/10.2307/2411138).
- Mueller UG, Mikheyev AS, Hong E, Sen R, Warren DL, Solomon SE, Ishak HD, Cooper M, Miller JL, Shaffer KA, Juenger TE.** 2011. Evolution of cold-tolerant fungal symbionts permits winter fungiculture by leafcutter ants at the northern frontier of a tropical ant-fungus symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **108**(10):4053–4056 DOI [10.1073/pnas.1015806108](https://doi.org/10.1073/pnas.1015806108).
- Neafsey DE, Lawiczak MKN, Park DJ, Redmond SN, Coulibaly MB, Traoré SF, Sagnon N, Costantini C, Johnson C, Wiegand RC, Collins FH, Lander ES, Wirth DF, Kafatos FC, Besansky NJ, Christophides GK, Muskavitch MAT.** 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* **330**(6003):514–517 DOI [10.1126/science.1193036](https://doi.org/10.1126/science.1193036).

- Nguyen AD, DeNovellis K, Resendez S, Pustilnik JD, Gotelli NJ, Parker JD, Cahan SH.** 2017. Effects of desiccation and starvation on thermal tolerance and the heat-shock response in forest ants. *Journal of Comparative Physiology B* **187**(8):1107–1116 DOI [10.1007/s00360-017-1101-x](https://doi.org/10.1007/s00360-017-1101-x).
- Nygaard S, Wurm Y.** 2015. Ant genomics (Hymenoptera: Formicidae): challenges to overcome and opportunities to seize. *Myrmecological News* **21**:59–72.
- Oksanen J, Blanchet F, Kindt R, Legendre P, O'Hara R.** 2016. Vegan: community ecology package. [Q9]
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM.** 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology* **17**(1):132 DOI [10.1186/s13059-016-0997-x](https://doi.org/10.1186/s13059-016-0997-x).
- Parmesan C.** 2006. Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics* **37**(1):637–669 DOI [10.1146/annurev.ecolsys.37.091305.110100](https://doi.org/10.1146/annurev.ecolsys.37.091305.110100).
- Pelini SL, Diamond SE, Nichols LM, Stuble KL, Ellison AM, Sanders NJ, Dunn RR, Gotelli NJ.** 2014. Geographic differences in effects of experimental warming on ant species diversity and community composition. *Ecosphere* **5**(10):art125 DOI [10.1890/es14-00143.1](https://doi.org/10.1890/es14-00143.1).
- Penick CA, Diamond SE, Sanders NJ, Dunn RR.** 2017. Beyond thermal limits: comprehensive metrics of performance identify key axes of thermal adaptation in ants. *Functional Ecology* **31**(5):1091–1100 DOI [10.1111/1365-2435.12818](https://doi.org/10.1111/1365-2435.12818).
- Petrov DA.** 2001. Evolution of genome size: new approaches to an old problem. *Trends in Genetics* **17**(1):23–28 DOI [10.1016/s0168-9525\(00\)02157-0](https://doi.org/10.1016/s0168-9525(00)02157-0).
- R Core Team.** 2017. *R: a language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. Available at <http://www.R-project.org/>.
- Rowntree JK, Shuker DM, Preziosi RF.** 2011. Forward from the crossroads of ecology and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **366**(1569):1322–1328.
- Ryan Gregory T.** 2005. Genome size evolution in animals. *Evolution of the Genome* 3–8 DOI [10.1016/b978-012301463-4/50003-6](https://doi.org/10.1016/b978-012301463-4/50003-6). [Q10]
- Siddig AA, Ellison AM, Ochs A, Villar-Leeman C, Lau MK.** 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in ecological indicators. *Ecological Indicators* **60**:223–230 DOI [10.1016/j.ecolind.2015.06.036](https://doi.org/10.1016/j.ecolind.2015.06.036).
- Spicer ME, Stark AY, Adams BJ, Kneale R, Kaspari M, Yanoviak SP.** 2017. Thermal constraints on foraging of tropical canopy ants. *Oecologia* **183**(4):1007–1017 DOI [10.1007/s00442-017-3825-4](https://doi.org/10.1007/s00442-017-3825-4).
- Stanton-Geddes J, Nguyen A, Chick L, Vincent J, Vangala M, Dunn RR, Ellison AM, Sanders NJ, Gotelli NJ, Helms Cahan S.** 2016. Thermal reactionomes reveal divergent responses to thermal extremes in warm and cool-climate ant species. *BMC Genomics* **17**(1):171–186 DOI [10.1186/s12864-016-2466-z](https://doi.org/10.1186/s12864-016-2466-z).
- Tsutsui ND, Suarez AV, Spagna JC, Johnston JS, Gregory T, Evans J, Gundersen-Rindal D, Gardner T, Gregory T, Wilson E, Hölldobler B, Wilson E, Li J, Heinz K, Johnston J, Ross L, Beani L, Hughes D, Kathirithamby J, Geraci N, Johnston J, Robinson J, Wikle S, Hill C, Gregory T, Bennett M, Leitch I, SanMiguel P, Gaut B, Tikhonov A, Nakajima Y, Bennetzen J, Kazazian H, Kidwell M, Comeron J, Ustinova J, Achmann R, Cremer S, Mayer F, Hancock J, Hancock J, Toth G, Gaspari Z, Jurka J, Redon R, Ishikawa S, Fitch K, Feuk L, Perry G, Andrews T, Fiegler H, Shapero M, Carson A, Chen W, Cho E, Dallaire S, Freeman J, Gonzalez J, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald J, Marshall C,**

Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville M, Tchinda J, Valsesia A, Woodwork C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad D, Estivill X, Tyler-Smith C, Carter N, Aburatani H, Lee C, Jones K, Scherer S, Hurles M, Gregory T, Gregory T, Petrov D, Lozovskaya E, Hartl D, Devos K, Brown J, Bennetzen J, Bennetzen J, Ma J, Devos K, Ma J, Bennetzen J, Oliver M, Petrov D, Ackerly D, Falkowski P, Schofield O, Gregory T, Hebert P, Kolasa J, Finston T, Hebert P, Foottit R, Ferrari J, Rai K, Ellegren H, Vandenbussche R, Longmire J, Baker R, Organ C, Shedlock A, Meade A, Pagel M, Edwards S, Hughes A, Hughes M, Reinhold K, Gregory T, Pittendrigh B, Clark J, Johnston J, Lee S, Romero-Severson J, Dasch G, Gregory T, Weinstock G, Robinson G, Gibbs R, Worley K, Evans J, Maleszka R, Robertson H, Weaver D, Beye M, Bork P, Elsik C, Hartfelder K, Hunt G, Zdobnov E, Amdam G, Bitondi M, Collins A, Cristino A, Latorff H, Lobo C, Moritz R, Nunes F, Page R, Simoes Z, Wheeler D, Carninci P, Fukuda S, Hayashizaki Y, Kai C, Kawai J, Sakazume N, Sasaki D, Tagami M, Albert S, Baggerman G, Beggs K, Bloch G, Cazzamali G, Cohen M, Drapeau M, Eisenhardt D, Emore C, Ewing M, Fahrbach S, Foret S, Grimmelikhuijzen C, Hauser F, Hummon A, Huybrechts J, Jones A, Kadowaki T, Kaplan N, Kucharski R, Leboulle G, Linial M, Littleton J, Mercer A, Richmond T, Rodriguez-Zas S, Rubin E, Sattelle D, Schlipalius D, Schoofs L, Shemesh Y, Sweedler J, Velarde R, Verleyen P, et al. 2008. The evolution of genome size in ants. *BMC Evolutionary Biology* 8(1):64.

Walker BJ, Abeel T, Shea T, Priest M, Abouellil A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 9(11):e112963 DOI [10.1371/journal.pone.0112963](https://doi.org/10.1371/journal.pone.0112963).

Ward PS. 2014. The phylogeny and evolution of ants. *Annual Review of Ecology, Evolution, and Systematics* 45(1):23–43 DOI [10.1146/annurev-ecolsys-120213-091824](https://doi.org/10.1146/annurev-ecolsys-120213-091824).

Ward PS, Brady SG, Fisher BL, Schultz TR. 2015. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). *Systematic Entomology* 40(1):61–81 DOI [10.1111/syen.12090](https://doi.org/10.1111/syen.12090).

Warren RJ, Chick L. 2013. Upward ant distribution shift corresponds with minimum, not maximum, temperature tolerance. *Global Change Biology* 19(7):2082–2088 DOI [10.1111/gcb.12169](https://doi.org/10.1111/gcb.12169).

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* 35(3):543–548 DOI [10.1093/molbev/msx319](https://doi.org/10.1093/molbev/msx319).

# Author Query Form

**Journal:** PeerJ

**Article ID:** peerj-6447

**Article Title:** Draft *Aphaenogaster* genomes expand our view of ant genome size variation across climate gradients

|                   |   |
|-------------------|---|
| <b>Q1</b> (p.1)   | Please check all author names, initials, and affiliations to confirm they are accurate.   |
| <b>Q2</b> (p.9)   | On page 10, line 200, “Duke Forest” the city is missing for the manufacturer. Please provide.   |
| <b>Q3</b> (p.12)  | On page 16, line 244, “Harvard Forest” the city is missing for the manufacturer. Please provide.  |
| <b>Q4</b> (p.15)  | Reference Agosti et al. (2000) is incomplete. Please provide the relevant missing information: location of publisher.   |
| <b>Q5</b> (p.16)  | Reference Bolton (2006) is incomplete. Please provide the relevant missing information: location of publisher.  |
| <b>Q6</b> (p.16)  | Reference Cook & Weisberg (1982) is incomplete. Please provide the relevant missing information: location of publisher.   |
| <b>Q7</b> (p.16)  | Reference Ellison (2012) is incomplete. Please provide the relevant missing information: location of publisher.   |
| <b>Q8</b> (p.17)  | The reference was originally supplied with the incorrect [Journal title]. We have corrected this so that it matches the publication. Please check to verify that this is correct. |
| <b>Q9</b> (p.18)  | Reference Oksanen et al. (2016) is incomplete. Please provide the relevant missing information: website.  |
| <b>Q10</b> (p.18) | Reference Ryan Gregory (2005) is incomplete. Please provide the relevant missing information: volume.   |