Conserved Microsatellites in Ants Enable Population Genetic and Colony Pedigree Studies across a Wide Range of Species

Ian A. Butler*, Kimberly Siletti*, Peter R. Oxley, Daniel J. C. Kronauer

Laboratory of Insect Social Evolution, The Rockefeller University, New York, New York, United States of America

Abstract

Broadly applicable polymorphic genetic markers are essential tools for population genetics, and different types of markers have been developed for this purpose. Microsatellites have been employed as particularly polymorphic markers for over 20 years. However, PCR primers for microsatellite loci are often not useful outside the species for which they were designed. This implies that a new set of loci has to be identified and primers developed for every new study species. To overcome this constraint, we identified 45 conserved microsatellite loci based on the eight currently available ant genomes and designed primers for PCR amplification. Among these loci, we chose 24 for in-depth study in six species covering six different ant subfamilies. On average, 11.16 of these 24 loci were polymorphic and in Hardy-Weinberg equilibrium in any given species. The average number of alleles for these polymorphic loci within single populations of the different species was 4.59. This set of genetic markers will thus be useful for population genetic and colony pedigree studies across a wide range of ant species, supplementing the markers available for previously studied species and greatly facilitating the study of the many ant species lacking genetic markers. Our study shows that it is possible to develop microsatellite loci that are both conserved over a broad range of taxa, yet polymorphic within species. This should encourage researchers to develop similar tools for other large taxonomic groups.

Introduction

Microsatellites, also called short tandem repeats (STRs) or simple sequence repeats (SSRs), are sequential repeats of 1 to 6 base pair motifs that have been used as genetic markers for more than 20 years [1,2,3]. Often found in noncoding regions, they are common in the genomes of eukaryotes [4–6]. An important feature of these sequences is their high degree of length polymorphism within populations of single species, which has been attributed to DNA polymerase slippage during replication [7,8]. This can result in a large number of alleles per locus that differ from one another in the number of repeats, making them distinguishable by size alone. This high degree of polymorphism and the ease of genotyping make them particularly suitable for studies in population genetics and pedigree analyses [9,10]. For example, microsatellites have been used to measure population differentiation and hybridization [11,12], to investigate ploidy levels [13,14], and to reconstruct parentage and pedigrees in wild and domestic populations [15,16]. Microsatellites are comparatively cheap to genotype and can be used with low concentrations of DNA. Furthermore, they typically have more alleles per locus than single nucleotide polymorphisms (SNPs) and thus provide more information per locus [17]. Although they often have a high degree of polymorphism within species, some microsatellite loci can be conserved across species that diverged 100 million years ago or more [18–24].

More recently, next generation sequencing (NGS) techniques have risen in popularity, mainly because of the large number of marker loci they can generate at relatively low per locus cost. For example, restriction site-associated DNA (RAD) tags can generate thousands of markers and have proven instrumental for measuring gene flow between populations [25], as well as for reconstructing shallow phylogenies [26]. However, the data generated from these techniques can be complex and difficult to analyze. There are techniques to reduce the complexity of DNA libraries such as double digest RADseq (ddRAD) [27], 2b-RAD [28], or genotyping by sequencing (GBS) [29], but these still require expensive NGS platforms. On the other hand, for many studies a smaller number of markers is sufficient, and markers such as microsatellites can be more attractive.

Despite their utility, a significant impediment to the use of microsatellites is the cost and effort associated with identifying a set of loci and developing PCR primers. Although the same loci can sometimes be useful for studying closely related species, loci that are polymorphic in one species are often not informative in another, and primers quickly lose affinity as species become more divergent. This usually requires new microsatellite loci to be characterized for each studied species. Depending on the research...
question, studies typically require a set of five to ten or more independent microsatellite loci. Paying a commercial service to develop these markers can be costly, and developing markers independently can be labor intensive and time consuming. Nevertheless, the utility of microsatellites in determining pedigree structures, relatedness and mating systems makes them particularly useful for social insect research because they can be used to address important questions related to inclusive fitness theory, including social organization (e.g. [30]), worker caste determination (e.g. [31]), and the evolution of supercolonies (e.g. [32]). Of the social insects, ants are a particularly speciose and ecologically diverse group being intensively studied. Current estimates place the ant family Formicidae at 115 to 158 million years of age [33–35], and close to 13,000 species have been described, according to the Hymenoptera Name Server (v. 1.5, accessed 14 April 2014). Eight ant genomes are currently available representing most major ant clades, allowing highly conserved regions to be identified over most of the family. To help overcome the constraints of narrowly applicable primers and to make microsatellites broadly available as population genetic markers, we aimed to develop a set of microsatellite markers that would be conserved across a wide range of species, yet polymorphic within species.

Results

To design a set of broadly applicable microsatellite primers we searched the eight currently available ant genomes for conserved microsatellite motifs with conserved flanking regions. The eight available ant genomes are from the red harvester ant Paraponera clavata (subfamily Myrmicinae) [36], Jerdon’s jumping ant Harpegnathos saltator (subfamily Ponerinae), the Florida carpenter ant Camponotus floridanus (subfamily Formicinae) [37], the leaf-cutting ants Atta cephalotes (subfamily Myrmicinae) and Acromyrmex echinatior (subfamily Myrmicinae) [38], the Argentine ant Linepithema humile (subfamily Dolichoderinae) [39], the red imported fire ant Solenopsis invicta (subfamily Myrmicinae) [40], and the clonal raider ant Cerapachys biroi (subfamily Myrmicinae) [41].

Figure 1. Phylogeny of the ants, showing the phylogenetic distribution of the species used in this study. The size of each triangle is proportional to the number of species in each group, and the approximate number of species is given in parentheses next to the group name. Boxes next to species names indicate whether that species’ genome was used to design (green) or test (purple) the PCR primers. Figure adapted with permission from Libbrecht et al. 2013 [73]. doi:10.1371/journal.pone.0107334.g001
Table 1. PCR amplification success across six ant species for the 21 microsatellite loci that were only tested with unlabeled PCR primers.

| Locus | Primer sequence (5'-3') | P. clavata | S. pentadentata | D. molestus | L. nearcticus | E. ruidum | S. invicta |
|-------|-------------------------|------------|----------------|-------------|---------------|-----------|-----------|
| Ant21 | F-TTCTGGGACGACGGTTGTT   | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-CCATCGGACTTCACCTCG     |            |                |             |               |           |           |
| Ant608| F-AGCAGATCTAGTGCTCTGGG   | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-ATGGACGGAGATAAGACGGA   |            |                |             |               |           |           |
| Ant1049| F-GAGGATCGGAGGGGTTGCGGA   | Yes        | Yes            | No          | Yes           | Yes       | Yes       |
|       | R-CGGAGCGCTTGCTGTTGAT    |            |                |             |               |           |           |
| Ant1052| F-GCGACCTTCTGAGCGATCTC   | Yes        | Yes            | No          | Yes           | Yes       | Yes       |
|       | R-CCTTTTAGTACAGACGCGCG   |            |                |             |               |           |           |
| Ant1387| F-ATAGGTGCGACATACGGTGT   | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-CAGGCGGACTCCCTCTCTCC   |            |                |             |               |           |           |
| Ant1732| F-ATGATAAGCAGATGAGTCC    | Yes        | No             | No          | Yes           | Yes       | Yes       |
|       | R-GCCACGCTCCGGGCCTCCTAT  |            |                |             |               |           |           |
| Ant2409| F-ATAGGCGCTACATGCGATTT   | Yes        | Yes            | No          | Yes           | Yes       | Yes       |
|       | R-GCTGATCTCTTCTGAGCCGCAC |            |                |             |               |           |           |
| Ant3362| F-CCGCAATCACTCTCGTGCC    | Yes        | Yes            | No          | Yes           | Yes       | Yes       |
|       | R-GTCTACAAAGCTGGGATAGGA  |            |                |             |               |           |           |
| Ant3395| F-CGCGGCGGAGCGATCCAGCCA  | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-CGGAGCACTTGGTACACGTTA  |            |                |             |               |           |           |
| Ant3411| F-GCGAGGCGGACGAGCTACCCCC | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-TGACGAGAACAAACTGCGGGC  |            |                |             |               |           |           |
| Ant3452| F-TTGAGTGCAGCGACTGCGAGA   | No         | No             | No          | Yes           | Yes       | Yes       |
|       | R-ATGCCAGCAAATCTGCGGCC   |            |                |             |               |           |           |
| Ant3505| F-TTACCGGACATCTGTGTTGG   | Yes        | No             | Yes         | Yes           | Yes       | Yes       |
|       | R-TACACACAGAGCAGAATTCT   |            |                |             |               |           |           |
| Ant3541| F-TGCACAAGCTGCTGGAGT     | No         | No             | No          | No            | No        | No        |
|       | R-TCACTGCTGCCGCGYGCAT    |            |                |             |               |           |           |
| Ant4709| F-AAGCGGTAAGGAGTATGAGGA  | No         | No             | MP          | Yes           | Yes       | Yes       |
|       | R-AAGCGGAGGAGATAGTGAGAG  |            |                |             |               |           |           |
| Ant5033| F-TTCCCTCTCCCTTGCACACC   | Yes        | No             | Yes         | Yes           | Yes       | No        |
|       | R-GGGGAGAAAGGAGGTGGGAGG  |            |                |             |               |           |           |
| Ant5204| F-GCCGAAATTCTGCACTTCCT   | Yes        | Yes            | Yes         | Yes           | Yes       | Yes       |
|       | R-GGGGAGAAAGGAGGTGGGAGG  |            |                |             |               |           |           |
| Ant5544| F-GGGGTGCGGAGCGACGTGCTG  | Yes        | Yes            | Yes         | Yes           | MP        | No        |
|       | R-CATGAGATGCTGAGGCAACA   |            |                |             |               |           |           |
| Ant5656| F-TTGGACGCGGACGCTGCGT    | Yes        | Yes            | No          | Yes           | Yes       | Yes       |
|       | R-AGCCAGCACTCCTGCGCAGCT  |            |                |             |               |           |           |
| Ant10290| F-CAGCTTCTCGGCCTCCGCGG  | No         | No             | No          | No            | No        | No        |
|       | R-ACGCGGCGCTTCGGGCTCGG   |            |                |             |               |           |           |
| Ant10427| F-AATTCACTTACGGCGCTCTAA | Yes        | No             | No          | Yes           | No        | Yes       |
|       | R-ATCCAGAGCATCTCGGAGTTC  |            |                |             |               |           |           |
| Ant11610| F-GGTAATGGCGGGGCGGTCAA   | No         | No             | No          | No            | No        | No        |
|       | R-GCCTGAAGATGGGTCCGATTGC |            |                |             |               |           |           |

See Table 2 for details on the remaining 24 loci that were also tested using labeled PCR primers. "Yes" indicates clear amplification of a single product. "No" indicates no amplification of any product. "MP" indicates that there were multiple products from which the desired product could not be determined.

doi:10.1371/journal.pone.0107334.t001
(subfamily Dorylinae) [42]. The available genomes represent five of the 21 recognized extant ant subfamilies, allowing us to select primer sequences that are conserved in a wide range of species across the ants (Figure 1). We identified 176 potential microsatellite loci with conserved flanking regions across all eight genomes, and among those selected 45 that had a repeat motif in most or all of the available genomes (Table S1 in File S1). To demonstrate their usefulness in species other than those with available genomes, we tested these primers for amplification in six species from six different subfamilies, only one of which was also used for primer design (Solenopsis invicta, subfamily Myrmicinae) (Figure 1). The other five species in which the markers were tested were the bullet ant Paraponera clavata (subfamily Paraponerinae), the army ants Simopelta pentadentata (subfamily Ponerinae) and Dorylus molestus (subfamily Dorylinae), Lasius nearcticus (subfamily Formicinae), and Ectatomma ruidum (subfamily Ectatomminae). The success of PCR amplification varied by locus and species (Tables 1 

From those 45 loci, we selected 24 that amplified well in all or most of the six species tested and also had at least ten consecutive repeats of their motif in the genomes of more than one of the species with available genome sequences (Table S1 in File S1). We genotyped those 24 loci across all six species using fluorescently labeled primers (Applied Biosystems). PCR amplification was successful for all 24 loci in L. nearcticus and D. molestus, for 23 loci in S. invicta, for 22 loci in P. clavata and E. ruidum, and for 21 loci in S. pentadentata (Table 2, Figure 2). To determine which of the microsatellite loci were polymorphic in any given species, we genotyped ten individuals from ten different colonies from the same population of each species for each locus. On average, 12.83 (±6.15 SD) of the 24 loci were polymorphic in a given species, and 11.16 (±5.27 SD) were polymorphic and in Hardy-Weinberg equilibrium (Table 2, Figure 2). Across those polymorphic loci in Hardy-Weinberg equilibrium, the average number of alleles per locus per species was 4.39 (±2.41 SD). The average observed heterozygosity was 0.534 (±0.22 SD), and the average expected heterozygosity was 0.61 (±0.22 SD). Most of the loci were monomorphic for multiple species. However, in all cases the monomorphic allele at a given locus was different for each species. We found no statistical linkage disequilibrium (at p<0.00003 after Bonferroni correction) between any pair of loci in any species, but this is likely due to small sample sizes and reduced power due to the large number of tests performed. In fact, in all eight genomes there are scaffolds containing multiple loci, i.e. these loci occur on the same chromosome and are therefore physically linked (Table S2 in File S1).

Discussion

To reduce the time and cost associated with developing microsatellite primers for a large number of different species, we designed a set of 43 primer pairs for potential use in a broad range of ant species spanning many millions of years of evolution. We tested 24 of these primer pairs in detail across six distantly related ant species from six different subfamilies. The number of useful polymorphic loci ranged from 5 to 20 for the six species we tested, although those loci were not always the same across species. Although we found no statistical linkage between any loci, some loci were located on the same scaffold in the genome assemblies of the reference species, and the location of the loci in the reference genomes should be considered when selecting primers from this set (Tables S1 & S2 in File S1). In assessing the utility of these markers in other species, it may be initially beneficial to test the entire set using inexpensive unlabeled primers. Then fluorescently labeled primers can be used for genotyping only those loci that amplify and yield clean PCR products. To further reduce costs, the primers described here could be used as unlabeled locus-specific primers in combination with universal labeled-tail primers [43].

Microsatellites have been an important tool for studies in population genetics for more than 20 years [1–3]. They are excellent markers for many types of studies including pedigree analyses and mating system studies, but their applicability has previously been limited by the narrow range of taxa in which each locus can be used. Researchers usually develop sets of primers specifically for their study species or a group of closely related species, and ants are no exception in this respect (e.g. [44–50]). For example, we found 32 publications of microsatellite primer notes for ants in the journal Molecular Ecology Resources, a leading outlet for the publication of population genetic markers. These primer notes represented 31 species and 26 genera. Looking only at those studies that described more than ten polymorphic loci per species, the number of alleles per locus ranged from 2 to 21.

![Figure 2. Overview results of genotyping 24 microsatellite loci for six different ant species.](image-url)
Table 2. Characteristics of 24 microsatellite loci tested in six different ant species.

| Locus | Primer sequence (5'-3') | Paraponera clavata | Sinopelta pentadentata | Dorylus molestus | Lasius nearcticus | Ectatomma ruidum | Solenopsis invicta |
|-------|-------------------------|---------------------|------------------------|------------------|-------------------|------------------|-------------------|
|       |                         | n | A | Size range (bp) | H0 | HE | Deviates from HWE | n | A | Size range (bp) | H0 | HE | Deviates from HWE | n | A | Size range (bp) | H0 | HE | Deviates from HWE | n | A | Size range (bp) | H0 | HE | Deviates from HWE | n | A | Size range (bp) | H0 | HE | Deviates from HWE |
| Ant20 | F-AGGTCTTAGCA-GGTAACTTTG R-CCTCGTCGACGAGGACG | 10 | 1 | 137 | 0 | 0 | no | 10 | 3 | 171–177 | 0.3 | 0.54 | no | 10 | 1 | 150 | 0 | 0 | no | 10 | 1 | 74 | 0 | 0 | no | 10 | 1 | 153 | 0 | 0 | no |
| Ant575 | F-TCAAGGTTCGAC-ACATGTGCC R-TCAAGCTGATT-TGTCAAGGTCG | 10 | 4 | 370–379 | 0.9 | 0.63 | no | 10 | 11 | 334–375 | 0.4 | 0.96 | yes | 10 | 2 | 230–250 | 0.2 | 0.19 | no | 10 | 4 | 209–234 | 0.3 | 0.37 | no | 10 | 1 | 248 | 0 | 0 | no | 10 | 3 | 218–239 | 0.7 | 0.63 | no |
| Ant859 | F-TACGCGGAGA-AACGTTGCTG R-GCTGATCAGAA-CTTCGATGAAC | 10 | 5 | 180–206 | 0.7 | 0.77 | no | 10 | 3 | 180–184 | 0.3 | 0.54 | no | 10 | 2 | 197–199 | 0.2 | 0.19 | no | 9 | 11 | 175–204 | 1 | 0.94 | no | 10 | 1 | 158 | 0 | 0 | no | 10 | 1 | 191 | 0 | 0 | no |
| Ant1343 | F-TCCGTCCCTCC-TCTTGCATT R-GRGGGCCTGCTG-CAATTTGCT | 10 | 4 | 229–235 | 0.6 | 0.53 | no | 10 | 1 | 186 | 0 | 0 | no | 10 | 4 | 263–269 | 0.6 | 0.76 | no | 10 | 3 | 206–211 | 0.4 | 0.58 | no | 10 | 4 | 252–272 | 0.9 | 0.71 | no |
| Ant1368 | F-ACTACCCCAATGACGCCAG C-TATGCTGAGGTGTTAGT | 10 | 1 | 251 | 0 | 0 | no | 7 | 6 | 266–313 | 0.14 | 0.93 | yes | 10 | 8 | 299–322 | 0.9 | 0.85 | no | 10 | 5 | 278–309 | 0.6 | 0.62 | no | 10 | 1 | 269 | 0 | 0 | no | 10 | 1 | 280 | 0 | 0 | no |
| Ant2341 | F-RAACAGCAAGG TGGCCGAGG | no amplification | no amplification | no amplification | 10 | 5 | 345–359 | 0.7 | 0.76 | no | 10 | 4 | 256–267 | 0.2 | 0.55 | no | 10 | 2 | 212–215 | 0.4 | 0.51 | no | 10 | 1 | 184 | 0 | 0 | no | 10 | 2 | 245–251 | 0.3 | 0.27 | no |
| Ant2794 | F-TGGTGTCGGGTGTGGCTGACG R-GATGCTGCA-CTTCAGGACTC | 10 | 3 | 241–251 | 0.5 | 0.42 | no | 9 | 9 | 280–336 | 0.67 | 0.90 | no | 10 | 5 | 246–270 | 0.4 | 0.77 | no | 10 | 10 | 240–268 | 0.1 | 0.89 | no | 10 | 9 | 218 | 0 | 0 | no | 9 | 1 | 258 | 0 | 0 | no |
| Ant2956 | F-GGGGAGTC-GGTAATCTCCTC R-TGGGTGCCTGCA-GTAAATGTGT | no amplification | no amplification | no amplification | 10 | 7 | 314–336 | 0.3 | 0.92 | yes | 10 | 9 | 352–390 | 0.4 | 0.9 | no | no amplification | 10 | 5 | 349–365 | 0.1 | 0.81 | yes | 10 | 1 | 337 | 0 | 0 | no |
| Ant3648 | F-CTTCCTGTCGCC-TGGGATCCTC R-TAAACCCATG-CCCTCTGTG | 9 | 1 | 337 | 0 | 0 | no | 10 | 10 | 368–410 | 0.5 | 0.94 | yes | 10 | 7 | 376–421 | 0.5 | 0.83 | no | 10 | 3 | 332–343 | 0.6 | 0.57 | no | 10 | 4 | 393–401 | 0.3 | 0.67 | no | 10 | 1 | 337 | 0 | 0 | no |
| Locus   | Primer sequence (5’-3’) | Paraponera clavata | Sinopelta pentadentata | Dorylus molestus | Lasius nearcticus | Ectatomma ruidum | Solenopsis invicta |
|---------|------------------------|---------------------|------------------------|------------------|-------------------|------------------|-----------------|
|         | n A Size range (bp) Ho H_e Deviates from HWE | n A Size range (bp) Ho H_e Deviates from HWE | n A Size range (bp) Ho H_e Deviates from HWE | n A Size range (bp) Ho H_e Deviates from HWE | n A Size range (bp) Ho H_e Deviates from HWE | n A Size range (bp) Ho H_e Deviates from HWE |
| Ant3653 | F-AGCAGAGACC- AATCAAGCGGA R-GGCAATTATC- GGACCAGGTT | 10 1 273 0 0 no | 10 9 238– 254 0.8 0.85 no | 10 3 255– 259 0.6 0.62 no | 10 9 261– 319 0.4 0.9 yes | 10 4 357– 363 0.4 0.74 no | 10 2 254– 256 0.4 0.33 no |
| Ant3993 | F-TGATCCGCTC- TTAAATTTAG- ATGGGA R-AGCTTTCCG- ARGCATTAAC- ATTTTCCT | 8 7 368– 387 0.88 0.88 no | 10 5 368– 379 0.2 0.81 yes | 10 2 311– 317 0.5 0.48 no | 10 7 379– 419 0.7 0.77 no | 8 1 454 0 0 no | 10 3 375– 363 0.7 0.47 no |
| Ant4155 | F-AGAATCTCT- TGAGCCTCGGAGG R-AGCTCTCTG- CTGTGAGG | 10 1 162 0 0 no | 8 3 206– 211 0.38 0.64 no | 10 4 176– 195 0.8 0.61 no | 10 1 170 0 0 no | 10 1 158 0 0 no | 10 2 200– 203 0.2 0.19 no |
| Ant5035 | F-AGGATAATTGTT- TCCGCGTTTATGGG R-GCTACTCTCGY- AGGGTTAAGG | 10 2 340– 342 0.4 0.33 no | 10 9 412– 442 0.3 0.94 yes | 10 6 365– 384 0.8 0.77 no | 10 8 284– 341 0.8 0.9 no | 10 1 331 0 0 no | 10 1 311 0 0 no |
| Ant7249 | F-AAGTGTCAAG- GCCGCTGATGAG R-AGGATTGTTGAGGT | 10 1 425 0 0 no | 10 7 320– 359 0.4 0.86 yes | 10 6 369– 398 0.5 0.68 no | 10 5 345– 368 0.6 0.74 no | 10 1 325 0 0 no | 10 1 358 0 0 no |
| Ant7680 | F-TCCGGAGAGAG- CTGATTCTCC R-TAGGACAAAC- TGGAGACCAC | 10 1 306 0 0 no | 10 9 332– 386 0.56 0.97 yes | 10 6 310– 328 0.6 0.74 no | 10 1 257 0 0 no | 10 1 219 0 0 no | 10 1 264 0 0 no |
| Ant8424 | F-TCATATGGCAG- ATGATGAAAATTCCT R-GGCGATTACA- AATGAGGCC | 10 2 262– 265 0.2 0.19 no | 10 8 894– 318 0.5 0.82 no | 10 3 232– 238 0.5 0.48 no | 10 4 193– 240 0.4 0.36 no | 10 2 266– 275 0.4 0.44 no | 10 3 235– 259 0.4 0.35 no |
| Ant8498 | F-GATGCGAGAGAGA- GCCAGAGC R- TGTTGGGAAATTCTC | 10 2 214– 218 0.4 0.51 no | 10 1 181 0 0 no | 10 1 147 0 0 no | 10 1 145 0 0 no | 10 1 172 0 0 no | 10 1 201 0 0 no |
| Locus sequence (5'-3') | Paraponera clavata | Sphinctopelta pentadentata | Dorylus molestus | Lasius nearcticus | Ectatomma ruidum | Solenopsis invicta |
|------------------------|--------------------|---------------------------|-----------------|------------------|----------------|-----------------|
|                        | $n$ | $A$ | Size range (bp) | $H_O$ | $H_E$ | Deviates from HWE | $n$ | Size range (bp) | $H_O$ | $H_E$ | Deviates from HWE | $n$ | Size range (bp) | $H_O$ | $H_E$ | Deviates from HWE | $n$ | Size range (bp) | $H_O$ | $H_E$ | Deviates from HWE |
| Ant9181 F-TGCCACTTACG-CTGGAGCACAC-R-CAAATGCAGCC-CAAGAGAAA | 10 | 1 | 280 | 0 | 0 | no | no amplification | 10 | 4 | 355–371 | 0.3 | 0.62 no | no | no amplification | no amplification |
| Ant9218 F-GACCTACCTT-GCCTTCGTTAA-R-CCTATGTTAACATGCAGG | 10 | 1 | 335 | 0 | 0 | no | 5 | 6 | 500–564 | 0.6 | 0.93 no | no | no amplification | no amplification |
| Ant10878 F-CGGTGGTTATATCAAGTCATTGCAGCAGG | 10 | 1 | 358 | 0 | 0 | no | 10 | 7 | 358–377 | 0.6 | 0.88 no | no | no amplification | no amplification |
| Ant11315 F-AGCGTGGTACAAGCTGGCGAAGGCTTTAGACCTAA | 10 | 1 | 358 | 0 | 0 | no | 10 | 1 | 380 | 0 | 0 | no | 10 | 1 | 317 | 0 | 0 | no | 10 | 1 | 355 | 0 | 0 | no | 10 | 1 | 343 | 0 | 0 | no |
| Ant1400 F-CAGCACTTGTGGCGCGGATACCTTTAATGAAATTCACAC | 10 | 1 | 258 | 0 | 0 | no | 10 | 9 | 259–336 | 0.7 | 0.85 no | no | no amplification | no amplification |
| Ant1893 F-CAGCACTTGGGTAAGAGCCCGACGCTTAGG | 10 | 9 | 375–392 | 0.89 | 0.9 no | 10 | 5 | 377–412 | 0.2 | 0.82 yes | no | no amplification | no amplification |
| Ant12220 F-AAGAGAGGCGG-GCGCTTCTCAAGAGG | 10 | 1 | 378 | 0 | 0 | no | no amplification | 10 | 3 | 274–280 | 0.3 | 0.28 no | no | no amplification | no amplification |

$n$ is the number of individuals successfully genotyped for each locus, $A$ is the number of alleles, $H_O$ is observed heterozygosity, $H_E$ is expected heterozygosity, and the last column for each species indicates whether that locus deviates from Hardy-Weinberg equilibrium in that species.

doi:10.1371/journal.pone.0107334.t002
Table 3. Overview of number of alleles and expected and observed heterozygosity in eight studies of species-specific microsatellite primers in ants.

| Species               | Number of loci | Mean A | Hr range | Mean Hr | Hr range | Mean Ho | Ho range |
|-----------------------|----------------|--------|----------|---------|----------|---------|----------|
| Allomerus octoarticulatus | 15             | 7.03   | 2–21     | 0.65    | 0.20–0.30| 0.20    | 0.05–0.80|
| Oecophylla smaragdina   | 13             | 5.00   | 2–14     | 0.58    | 0.15–0.89| 0.58    | 0.05–1.00|
| Petalomyrmex phylax      | 14             | 7.43   | 2–15     | 0.68    | 0.05–0.93| 0.68    | 0.05–1.00|
| Formica exsecta          | 12             | 6.48   | 2–14     | 0.72    | 0.05–0.97| 0.72    | 0.05–1.00|
| Wasmannia auropunctata   | 12             | 9.91   | 4–19     | 0.81    | 0.02–0.97| 0.81    | 0.02–0.97|
| Lasius austriacus        | 11             | 9.91   | 4–19     | 0.78    | 0.01–0.97| 0.78    | 0.01–0.97|

Number of loci is the number of polymorphic loci described in that study. Mean A is the average number of alleles per locus, Hr range is the range of allele numbers in each study, mean Hr and mean Ho are the average expected and observed heterozygosity respectively, and Ho range is the range of expected heterozygosity.

Materials and Methods

Specimen collection

All specimens of *Ectatomma raidum* and *Paraponera clavata* were collected at the Organization for Tropical Studies field station in La Selva, Costa Rica. *Simopelta pentadentata* specimens were collected in Monteverde, Costa Rica. *Dorylus molestus* specimens were collected in Kakamega Forest, Kenya. *Lasius*

### Specimen collection

All specimens of *Ectatomma raidum* and *Paraponera clavata* were collected at the Organization for Tropical Studies field station in La Selva, Costa Rica. *Simopelta pentadentata* specimens were collected in Monteverde, Costa Rica. *Dorylus molestus* specimens were collected in Kakamega Forest, Kenya. *Lasius*
nearticus specimens were collected at the Rockefeller University Center for Field Research in Millbrook, New York, USA, and specimens of Solenopsis invicta were collected in Tallahassee, Florida, USA.

Collection permits were acquired for all samples where necessary. A permit for specimens from Kakamega National Park, Kenya was granted by the National Council for Science and Technology (permit number NCST/RCD/12B/012/37B). A permit for specimens from Costa Rica was granted by Ministerio de Ambiente, Energía y Telecomunicaciones (permit number 192-2012-SINAC). Permits were not required for specimens collected in the United States. No protected species were sampled.

Bioinformatics

Seven available ant genomes were downloaded from Ant Genomes Portal (hymenopteragenome.org/ant_genome), and our lab has recently published the C. biroi genome [42]. The genome versions for each species were A. cephalotes v1.0, A. echinatior v2.0, C. floridanus v3.3, C. biroi v2.0, H. saltator v3.3, L. humile v1.0, P. barbatus v3.0, S. invicta v1.0. Microsatellites in the C. biroi genome were located using Tandem Repeats Finder (TRF; v. 4.04) [68], which utilizes Smith-Waterman style local alignment. Tandem repeats are reported only if they exceed a minimum alignment score, specified as 50 (Minscore = 50). Alignment mismatches were assigned a weight of five (Mismatch = 5). Additionally, the size of the repeat pattern was limited to five bases (Maxperiod = 5). The microsatellite indices returned were used to generate a masked BLAST query for each microsatellite, extended to include 200-bp flanking regions. The query sequence was used to search all eight sequenced ant genomes, including A. cephalotes v2.0, C. floridanus v3.3, A. echinatior v2.0, H. saltator v3.3, L. humile v1.0, P. barbatus v3.0, S. invicta v1.0. Microsatellites in the C. biroi genome were aligned using MUSCLE [70]. To confirm that these conserved flanking regions indeed contained microsatellite sequences, TRF was used to search for microsatellites in all database genomes at the indices returned by BLAST for each hit (settings as stated above). Primer3 software (v. 2.3.4; http://primer3.sourceforge.net/releases.php) [71] generated primers from the consensus sequence in each flanking region. A maximum of four unknown bases were allowed in any primer set (PRIMER_MAX_NS_ACCEPTED = 4). All unspecified parameters used the default or recommended settings. Custom Python scripts were used to parse TRF and Primer3 outputs, prepare files for BLAST and Primer3, and filter the BLAST results. These scripts are available upon request from the corresponding author. Initially, 176 loci were identified across all genomes with the described bioinformatics pipeline, from which we chose 45 loci for further study. These 45 loci were chosen subjectively based on the number of perfect repeats in different species and the presence of a microsatellite motif in as many ant genomes as possible.

DNA extraction, PCR amplification and genotyping

DNA was extracted by first homogenizing the tissue in a Qiagen TissueLyser II and then heating the sample at 96 °C for 15 minutes in 200 μl of 10% Chelex in TE solution. The samples were then centrifuged at 9100 rpm for three minutes, and the supernatant containing the DNA was removed and used as the template for PCR amplification.

The PCR cocktail (10 μl total volume) for all reactions contained 1 μl PCR Gold Buffer (10x), 0.5 μl MgCl₂ (25 mM), 0.5 μl dNTPs (10 mM total, 2.5 mM each), 0.1 μl of each forward and reverse primer (10 μM), 0.1 μl AmpliTaq Gold (5 U/μl), 1 μl DNA template and 6.7 μl H₂O. PCR reactions were run on an Eppendorf Mastercycler Pro S under the following conditions: 10 min at 95°C followed by 40 cycles of 15 s at 94°C, 30 s at 55°C and 30 s at 72°C, and a final extension of 10 min at 72°C. PCR products were sent to a commercial facility (Genewiz, Inc.) for genotyping. Analysis of chromatograms was performed using PeakScanner (Applied Biosystems). Calculations of observed and expected heterozygosity, as well as tests for linkage disequilibrium and deviations from Hardy-Weinberg equilibrium were performed using F-STAT (v2.9.3.2) [72].

Supporting Information

File S1 Contains Tables S1 and S2 described below. Table S1. Details of microsatellite loci in eight ant genomes. Numbers in parentheses behind the size of the targeted fragment indicate that there are a number of unknown ("N") bases inserted into the available genome sequence. In some cases, these can be larger stretches of "N" bases in the published genome assembly. This number is included in the total size of the targeted fragment. In most cases, this implies that the given size of the targeted fragment is probably imprecise. The column "Size of targeted fragment in base pairs" thus gives "Total base pairs (number of N bases among the total base pairs)". Some loci have multiple motifs listed. All motifs are in the same region and are included in the size of the targeted fragment. Table S2. Physical linkage of microsatellite loci. X indicates where two loci are on the same scaffold in that species. Order of species left to right in every box is P. barbatus, H. saltator, At. cephalotes, C. boro, L. humile, Ca. floridanus, S. invicta, Ac. echinatior.

Acknowledgments

We thank Sean O'Donnell for supplying samples of S. pennadentata, Ingrid Fetter-Prumeda and Sean McKenzie for collecting samples of S. invicta, Aniek Ivens for specimens of L. neasciens, Marcell Peters for assistance in collecting D. molestus, Christoph von Beeren for assistance in collecting P. clausata and E. ruidum, and Romain Libbrecht for assistance in creating figures.

Author Contributions

Conceived and designed the experiments: IAB PRO DJCK. Performed the experiments: IAB KS. Analyzed the data: IAB KS DJCK. Wrote the paper: IAB DJCK. Wrote the custom code used in the analysis: KS PRO. Commented on the manuscript and approved the final version: IAB KS PRO DJCK.
7. Leclercy S, Rivals E, Jarne P (2010) DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach. Genome Biol Evol 2: 325–335.

8. Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in rukyarygenomes. BioEssays 28: 1069–1070.

9. Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11: 424–429.

10. Selkoe KA, Tooman RJ (2006) Microsatellites for ecologists: a practical guide to microsatellite evaluation. Microsatellite markers. Trends Ecol Evol 9: 615–629.

11. Hansson B, Tarka M, Dawson DA, Horsburgh GJ (2011) Hybridization but no evidence for backcrossing and introgression in a sympatric population of great reed warblers and flamboyant reed warblers. PLoS One 7: e31667.

12. Kronauer DJC, Peters MK, Schöning C, Boomsma JJ (2011) Hybridization in African swamp-riding army ants. Front Ecol Environ 9: 20.

13. Jungman L, Vigna BBZ, Boldrini KR, Sousa ACB, do Valle CB, et al. (2010) Mosaic genetic diversity and population structure analysis of the tropical pasture grass Brachiaria humidicola based on microsatellites, cytogenetics, morphological traits and geographical origin. Genome 53: 689–709.

14. Mishlya T, Takada M, Tashkima H, Nakano M, Tabata R, et al. (2014) Molecular identification of species and ploidy of Carassius fishes in Lake Biwa, using mtDNA and microsatellite multiplex PCRs. Ichthyol Res 61: 169–175.

15. Wang J, Scribner KT (2014) Parentage and sibship inference from markers in. PLoS Genet 7: e1002007.

16. Nygaard S, Zangh G, Schmit M, Li C, Wurm Y, et al. (2011) The genome of the leafcutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. PLoS Genet 7: e1002007.

17. Gaërke C, Ytournel F, Bed’hom B, Gut I, Lathrop M, et al. (2011) Comparison of the genetic diversity and population structure of the terrestrial pastaure grass *Brachytrum lanosum* based on microsatellites, cytogenetics, morphological traits and geographical origin. Genome 53: 689–709.

18. Seppä P, Johansson H, Gyllenstrand N, Pålsson S, Pamilo P (2012) Mosaic microsatellite markers in the ant *Atta laevigata*. BMC Res Notes 6: 328.

19. Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, et al. (1991) The use of heterologous PCR primer pairs in closely related species. Genomics 7: 223–234.

20. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple Genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19377.

21. Buschiazzo E, Gemmell NJ (2009) Evolution and phylogenetic significance of whitefish species pairs assessed by RAD sequencing. Evolution 67: 2483–2497.

22. Omland KS, Waples RS, Gaggiotti OE, Schierup MH (2005) Analysis of the genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution 67: 2483–2497.

23. Gyllenstrand N, Gertsch PJ, Pamilo P (2002) Polymorphic microsatellite DNA markers in the ant *Atta laevigata*. Conserv Genet 3: 415–421.

24. Vaccari S, Sbirrazzuoli LL, King TJ, Matrick JS, Georges M, et al. (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10: 634–660.

25. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple Genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19377.

26. Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution 67: 2483–2497.

27. Fournier D, Foucaud J, Loiseau A, Cros-Arteil S, Jourdan H, et al. (2005) PCR primers for the army ant *Lasius flavius* (Hymenoptera: Formicidae). Mol Ecol Notes 7: 328.

28. Wang S, Meyer E, McKay JK, Mata MV (2012) 2b-RAD: a simple and flexible approach to identify conserved microsatellite loci and develop primer sets of high utility. Mol Ecol Resour 10: 654–660.

29. Wang S, McKay JK, Mata MV (2012) 2b-RAD: a simple and flexible approach to identify conserved microsatellite loci and develop primer sets of high utility. Mol Ecol Resour 10: 654–660.

30. Suefuji M, Trindl J, Heinze J (2011) Characterization of 13 microsatellite markers for the threatened snake-making ant *Myrmecocystus muelleri* (Formicidae: Myrmicinae). Conserv Genet Resour 3: 229–231.

31. Qian Z-Q, Ceccarelli FS, Guo J, Schulze H, Schlick-Steiner BC, et al. (2011) Characterization of polymorphic microsatellites in the giant bulldog ant, *Myrmecia brevior* and the jumper ant, *M. pilosula*. J Insect Sci 11: 1–8.

32. Ascunce MS, Boomsma AJ, Shoemaker DJ (2009) Characterization of 24 microsatellite markers in the fire ant *Solenopsis invicta* genosys (Hymenoptera: Formicidae). Mol Ecol Resour 9: 1476–1479.

33. Gyllenstrand N, Gertsch PJ, Pamilo P (2002) Polymorphic microsatellite DNA markers in the ant *Atta laevigata*. Conserv Genet 3: 415–421.

34. Deboni GDG, Labrador MS, Jouve M, Sequeiros T (2009) Isolation and characterization of microsatellite markers from the acacia-ant *Crematogaster selvaarum*. Mol Ecol Resour 9: 1476–1479.

35. Maudet C, Luikart G, Taberlet P (2001) Development of microsatellite markers in the leaf-cutter ant *Atta laevigata*. Curr Biol 21: 5880–5891.

36. Deboni GDG, Labrador MS, Jouve M, Sequeiros T (2009) Isolation and characterization of microsatellite markers from the acacia-ant *Crematogaster selvaarum*. Mol Ecol Resour 9: 1476–1479.

37. Rubin BE, Makarevich CA, Tsunba LA, Stenzel I, Bogdanowicz SM, et al. (2009) Isolation and characterization of microsatellite markers from the acacia-ant *Crematogaster selvaarum*. Mol Ecol Resour 9: 1476–1479.

38. Fraze F, Sannini G, Natali C, Chelazzi G, Cotti C (2009) Characterization of polymorphic microsatellite loci in the acacia-ant *Crematogaster selvaarum*. Mol Ecol Resour 9: 1476–1479.

39. Bruschke C, Bed’hom B, Gut I, Lathrop M, et al. (2011) Comparison of the structure of chicken populations. Curr Genet 43: 415–421.

40. Urushibara A, Kajihara K, Sasaki M (2004) Microsatellite loci for the ant *Oecophylla smaragdina*. Mol Ecol Notes 4: 608–610.

41. Wurm Y, Jiang J, Ribas-Gromuz O, Corona M, Nygaard S, et al. (2011) The genome of the fire ant *Solenopsis invicta*. Proc Natl Acad Sci USA 108: 5670–5674.

42. Oxley PR, Lee J, Fetter-Pruneda I, McKenzie SK, Li C, et al. (2014) The evolution of the colonial ruler and Cephalotes thorax. Curr Biol 24: 431–438.

43. Forstner J, Fournier D, Foucaud J, Loiseau A, Cros-Arteil S, et al. (2005) PCR primers for the leaf-cutter ant *Atta laevigata*. Curr Biol 15: 811–814.

44. Tranuel E, Rousset F, Duret L, Leplae R, Gouyon PH, et al. (2001) Phylogenet Evol 10: 168–177.

45. Kronauer DJC, Boomsma JJ, Pierce NE (2011) Nine novel microsatellite loci for the ant *Atta laevigata*. BMJ Res Notes 6: 328.

46. Kleinert J, Stahl T, Weigmann M, Haller K, Habermehl T, et al. (2006) High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. BMC Genomics 14: 176.

47. Qian Z-Q, Ceccarelli FS, Guo J, Schluck-Steiner BC, et al. (2011) Characterization of polymorphic microsatellites in the giant bulldog ant, *Myrmecia brevior* and the jumper ant, *M. pilosula*. J Insect Sci 11: 1–8.

48. Ascunce MS, Boomsma AJ, Shoemaker DJ (2009) Characterization of 24 microsatellite markers in the fire ant *Solenopsis invicta* genosys (Hymenoptera: Formicidae). Mol Ecol Resour 9: 1476–1479.

49. Deboni GDG, Labrador MS, Jouve M, Sequeiros T (2009) Isolation and characterization of microsatellite markers from the acacia-ant *Crematogaster selvaarum*. Mol Ecol Resour 9: 1476–1479.
67. Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP (2008) Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biol 6: 6.
68. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580.
69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

70. Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
71. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al. (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40: e115.
72. Goudet J (1995) FSTAT Version 1.2: a computer program to calculate F-statistics. J Hered 86: 485–486.
73. Libbrecht R, Oxley PR, Kronauer DJC, Keller L (2013) Ant genomics sheds light on the molecular regulation of social organization. Genome Biol 14: 212.