Research Article

Received: 28 August 2018 Revised: 15 February 2019 Accepted article published: 29 March 2019 Published online in Wiley Online Library: 30 April 2019

(wileyonlinelibrary.com) DOI 10.1002/ps.5427

Genome-wide analyses of single nucleotide polymorphisms reveal the consequences of traditional mass-rearing on genetic variation in *Aphytis melinus* (Hymenoptera: Aphelinidae): the danger of putting all eggs in one basket

Marco Gebiola,[a*,†] Jeffrey W Streicher,[b†] Paul F Rugman-Jones,[a] Joseph G Morse[a] and Richard Stouthamer[a]

Abstract

BACKGROUND: *Aphytis melinus* DeBach (Hymenoptera: Aphelinidae) is a highly effective biocontrol agent of the California red scale *Aonidiella aurantii* (Maskell) (Hemiptera: Diaspididae). It is commercially reared and used for augmentative releases within integrated pest management programs. However, mass rearing of biocontrol agents can result in population bottlenecks and high levels of inbreeding and/or adaptation to the factitious rearing conditions. Although these factors can all negatively impact field performance of biocontrol agents, few empirical studies have examined the genetic consequences of mass rearing. We used double-digest RAD sequencing (ddRADseq) to investigate the effect of traditional mass rearing on genetic variation among insectary colonies of *A. melinus* relative to wild populations in native (Pakistan) and introduced (California) ranges.

RESULTS: Analyses of up to 9700 single nucleotide polymorphisms (SNPs) revealed that insectary populations had less genomic variation than introduced populations. This was evidenced by fewer private alleles, reduced heterozygosity, and greater missing data in the insectary populations. Further, California insectaries formed a distinct genomic cluster relative to the other samples, a surprising result given that the insectary colonies were putatively established at different times and from different source populations. These differences were evident across most data sets also after we filtered out contaminant DNA from the most common host species (*Aspidiotus nerii* Bouché and *A. aurantii*).

CONCLUSION: We hypothesize that this pattern would only result if: (i) directional selection for 'captive' phenotypes produces convergent patterns of genomic variation across insectaries; or (ii) the California insectary colonies were all founded from a unifying source population and/or that the insectaries regularly exchange 'genetic' stocks. We show that RADseq is an effective method to investigate the effects of mass rearing on genetics of biocontrol agents.

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

Keywords: *Aonidiella aurantii*; *Aspidiotus nerii*; ddRADseq; parasitoids; population genetics

1 INTRODUCTION

The use of commercially raised beneficial insects (e.g. sterile conspecific competitors and natural enemies) to control arthropod pests is widespread. However, mass rearing of these insects has historically focused more on the quantity of insects produced than on their quality[1–3]. Today, there are quality control guidelines for some mass-reared insects, but performance testing is rare, especially under field conditions.[4] A frequent practice for improving the quality of mass-reared beneficial insects is the periodic introduction of conspecifics from wild populations.[5] This practice is intended to mitigate the impact of inbreeding that results from the population size reduction so that mass-reared beneficial insects can be more effective when they are released into the field.[4,6] Minimizing domestication may be critical for the success of biocontrol programs because long-term captivity can lead to selection for phenotypes that are adapted to captivity but not to field conditions.[7–11]

Aphytis melinus DeBach (Hymenoptera: Aphelinidae) is a wasp native to India and Pakistan, where it was first collected and

* *Correspondence to: M Gebiola, Department of Entomology, 417 Entomology Bldg, University of California, Riverside, CA 92521, USA. E-mail: marco.gebiola@gmail.com*

† *Joint first authors.*

a *Department of Entomology, University of California, Riverside, CA, USA*

b *Department of Life Sciences, The Natural History Museum, London, UK*

described in the late 1950s.[12] At about the same time it was being described, the species was introduced to Southern California to control the California red scale *Aonidiella aurantii* (Maskell) (Hemiptera: Diaspididae). The control exerted by *A. melinus* was so effective in Southern California that insectaries readily began rearing and releasing it, adapting methods developed previously for *Aphytis lingnanensis* Compere.[13] In the most important area for California citrus production, the San Joaquin Valley, *A. melinus* now represents a critical, commercially reared natural enemy, used in augmentative releases within biologically-based insect pest management (IPM) programs by ~10% of growers (after reaching a peak of ~30% of growers in the mid-1990s).[14,15] Recently, Vasquez and Morse[16] evaluated temporal variation in the quality of *A. melinus* produced by five commercial insectaries in California, concluding that the wasps produced by each of the insectaries were likely to be of lower quality (i.e. less fit) at certain times of the year, as a result of reduced adult survival times, reduction in the size of the female wasps, and shifts in the sex ratio towards the production of higher numbers of males. Vasquez and Morse[16] speculated that the reduction in quality may correlate with decreasing levels of genetic variation in the insectary colonies as a result of inbreeding and/or adaptation to the mass-rearing conditions, but presented no empirical evidence to support this hypothesis.

Given the findings of Vasquez and Morse[16], we were primarily interested in testing the hypothesis that insectary-reared colonies of *A. melinus* have lower and/or divergent levels of genetic variation when compared with native field populations and established introduced field populations. Therefore, in this study we estimated levels of genetic variation among mass-reared 'populations' of *A. melinus* from commercial insectaries in California and Spain, and compared that to similar estimates in two highly inbred laboratory colonies held at UC Riverside (UCR), and wild populations from the native (Pakistan) and introduced (California) range. Our assessment of genetic variation was based on thousands of genome-wide single nucleotide polymorphisms (SNPs), characterized using a variant of restriction site-associated DNA sequencing (RADseq) called double-digest RADseq (ddRADseq)[17]. We also tested for the impact of filtering out host DNA by sequencing comparable ddRADseq libraries from two host species, the natural host *A. aurantii* and the host used for mass rearing, *Aspidiotus nerii* Bouché (Hemiptera: Diaspididae).

2 METHODS

2.1 Sampling of *Aphytis melinus*

Organismal size (< 1 mm) and budgetary constraints required that genetic characterization of each population was based on a pooled sample (i.e. multiple specimens) rather than multilocus genotypes for each individual. Samples of *A. melinus* were obtained between 2004 and 2005 from the same five California insectaries sampled by Vasquez and Morse[16], identified herein by the same anonymous A–E code. We also obtained a sample of a mass-reared colony of *A. melinus* produced by an anonymous Spanish insectary (November 2006). The insectary populations sampled were all reared on oleander scale, *A. nerii*, following long-standing industry protocols.[13] Established, introduced field populations of *A. melinus* were sampled in 2004 and 2005 from localities spanning four California counties: Redlands in San Bernardino Co. (November 2005), Porterville in Tulare Co. (October 2004), Edison in Kern Co. (October 2004), and Newman in Stanislaus Co. (September 2005), and a linear distance of ~600 km. Field samples were collected by bringing citrus fruit naturally infested with California red scale, *A. aurantii*, into the laboratory and holding it in isolation until wasps emerged.[18] In addition, between 2007 and 2008, small numbers of *A. melinus* were obtained from each of seven localities in Northern Pakistan (Islamabad, Peshawar, Rawalpindi, Taxila, Bhalwal, Kot Momin and Lahore). Given the constraints of organismal size, we chose to combine these to create one 'artificial' population that would both provide a fair estimate of genetic variation in the native range of *A. melinus*, and (potentially) recreate the variation originally released into California by DeBach in 1957.[12] Finally, two captive colonies held in the quarantine facility at UCR that were thought to be highly inbred were also sampled in 2008. The first was a colony (hereafter referred to as 'DeBach colony') established from the material collected by DeBach in the late 1950s in India and Pakistan (i.e. the genetic material that was originally introduced to California), and the second colony (hereafter referred to as 'China colony') was one established from wasps collected in Ming Ho, China, in 1990, but never subsequently released in California. These two colonies were exclusively inbred for about 600 and 230 generations, respectively. In total, we estimated genetic variation across 16 population samples.

2.2 Genomic sampling

For all but one of the populations sampled, DNA was extracted from pools of 50 diploid females, using the DNeasy Blood & Tissue Kit (Qiagen Inc., Valencia, CA, USA). The remaining sample, from the Spanish insectary, contained only 49 females, and so two haploid males were included, thereby resulting in the sampling of an equivalent ploidal number ($n = 100$) across all populations. This number was thought to be enough to avoid sampling error that could arise from small pools of individuals.[19,20] Importantly, species of *Aphytis* are haplodiploid, that is, the males are produced from unfertilized eggs, hence no male specific variation should arise to create bias in the Spanish insectary pool. Reduced genomic DNA (gDNA) libraries were created using the original ddRADseq protocol[17] with the following modifications. Approximately 200 ng of gDNA was digested with the restriction enzymes *Sph*I and *Mlu*CI (New England Biolabs, Ipswich, MA, USA) for 3 h at 37 °C. The digested DNA was purified using AMPure XP beads with a volume of beads 1.5× the volume of the restriction digest, and subsequently eluted in 40 μL of TE buffer. Concentration of the purified DNA was estimated using a Nanodrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA), and for each sample ~110 ng of digested DNA was ligated to one of four indexed P1 adapters (containing the index ACGGT, ATTAC, TCGAT, or TGGAA), and a universal P2 adapter (for complete oligo sequences see Peterson *et al.*[17]). After successful ligation, samples were combined into one of five composite reactions, each containing no more than one population sample labeled with each P1 adapter index. The five composite reactions were again purified with AMPure XP beads. The Illumina® flowcell annealing sequence, a second multiplexing index, and a sequencing primer annealing region were added to the DNA fragments with two rounds of polymerase chain reaction (PCR) using Q5® High-Fidelity DNA Polymerase (New England BioLabs), the universal PCR1 primer, and five PCR2 primers (PCR2_idx_1 to 5) (again, for complete oligo sequences see Peterson *et al.*[17]). This step adds a second index to each sample within a composite reaction, such that each of the original population samples can subsequently be identified by the combination of the two indexes. Amplicons were size selected using an automated BluePippin (Sage Science Inc., Beverly, MA, USA), targeting the range between 200 and 500 bp. After size selection, the composite

libraries were purified using AMPure XP beads, and the concentration of each was determined using a Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA). Equal amounts of DNA from each of the five composite libraries were combined, and the 16 originating sample libraries were sequenced on a single Illumina® MiSeq run at the University of California, Davis' Genome Center, using the 250 bp paired-end Illumina® kit. This produces two sets of sequences, hereafter referred to as 'read 1' and 'read 2' (*sensu* Davey *et al.*[21]) that read each fragment from the 5′- and 3′-ends, respectively.

The sequenced libraries (consisting of 260 bp reads 1 and 2) were demultiplexed and cleaned using the *process_radtags* program of the software pipeline Stacks v. 1.35[22] and quality-checked using FastQC (Babraham Bioinformatics, Cambridge, UK). Using the FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), all sequences were end-trimmed, resulting in two non-overlapping reads of 100 bp for read 1 and 90 bp for read 2. We observed far fewer data in the read 2 libraries, most likely as a result of large amounts of P1 adapter contamination that was removed prior to analyses. Demultiplexing, quality filtering and trimming resulted in a final data set that contained 9 305 668 read 1 sequences and 4 895 447 read 2 sequences.

2.3 Data assembly in Stacks

The Stacks programs *ustacks*, *cstacks* and *sstacks* were used to build *de novo* stacks for each population, to build a catalog of stacks from all populations, and to identify loci within stacks, respectively. A 'stack' is a set of identical sequences in the terminology of this pipeline; several of these stacks that have similar (but not identical) sequence identities are then merged to form putative loci. We allowed a maximum difference of four nucleotides to align all loci (default settings). Because errant population structure could result from low data coverage,[20] we attempted to identify SNPs at four different stack depths across treatments (6×, 10×, 15× and 20×), allowing a maximum of 25% missing populations per SNP. The Stacks program populations was then used to output SNPs for downstream analysis. Reads 1 and 2 were analyzed independent of each other to identify SNPs, and the respective paired-end reads were subsequently concatenated using a basic text editor, following their output from populations.

Current methods for analyzing pooled genetic 'libraries' have several limitations, such as requiring a reference genome for mapping short-length reads (e.g. PoPoolation[23]). Furthermore, prominent analytical tools, like the program Stacks,[22] are designed for diploid organisms, making highly polymorphic or low frequency alleles difficult to detect in pooled samples. Given these limitations for analyzing pooled data and the lack of a reference genome for any *Aphytis* species, it was necessary to modify certain interpretations from the standard Stacks pipeline in order to effectively analyze our RADseq data. Specifically, we developed two different approaches (described in detail below) that offer insight into the genomic variation across and within our pooled samples.

2.4 Analytical approach 1: using fixed sites within pools to infer relatedness

We used Stacks to identify those SNPs that were fixed within pools, but variable across pools. We used this approach to avoid violating assumptions that each individual RADseq library belongs to a diploid individual. This SNP data set (hereafter 'fixed approach') was then used to assess how: (i) variation was partitioned across the data set; and (ii) different samples were related to one another. Specifically, we used a stack depth of 10× and performed multivariate analyses.

To explore the distribution of variation across the fixed approach data set, we performed a principal components analysis in *adegenet* 1.3–9.2.[24] We replaced missing data using the *na.replace* command. To more objectively identify the total number of clusters, we also used *adegenet* to identify the most likely number of genomic clusters in our data via k-means clustering and Bayesian Information Criterion (BIC) scores. After identifying the most likely range of number of clusters, we used the program STRUCTURE[25] to approximate the membership of each putative cluster. All STRUCTURE runs were done using the default settings, i.e. the admixture model and correlated allele frequencies, letting the software infer the lambda parameter, with 100 000 runs in the data collection phase following a burn-in period of 10 000 runs, for a K cluster ranging from 2 to 4, with 100 iterations per each K.

2.5 Analytical approach 2: approximating genetic diversity within pools

The previous 'fixed' approach excludes all data that are polymorphic within a pooled population sample (a large proportion of the total observed variation). Thus, to gain insight into variation within populations, we used a second data analysis approach, hereafter referred to as the 'variable approach'. This method used the measure of 'heterozygosity' generated by Stacks as a proxy for within-pool genetic diversity. Although this approach cannot identify low-frequency alleles, we expect that there will be more sites with two high-frequency alleles in pools of individuals (i.e. populations) that have higher genetic diversity (i.e. higher heterozygosity). We term this measure 'partial heterozygosity' because although it is not equivalent to measures of heterozygosity made for a single individual, it should provide relevant insight into patterns of within-population genetic variation. We compared levels of partial heterozygosity between insectary and field (native and introduced) populations.

2.6 Using paired-end reads to support data quality

After trimming the read 1 and read 2 sequences, the respective reads were typically non-overlapping and ~100–300 bp apart from one another on the genome (based on our BluePippin selected fragment size; see above). Despite not overlapping, both sets of reads can be surveyed for SNPs by using the recognition site of the restriction enzyme as a reference point for alignment (in our case *Mlu*CI for read 1 and *Sph*I for read 2). These sites are effectively linked because of their close proximity to one another on the genome, but because they are non-overlapping they may each contain unique SNP data. We utilized this to test the integrity of our research pipeline, by comparing variance structure across our samples between read 1 and read 2. Similar variance structure between the two reads will provide validation for our pipeline. Specifically, we compared estimates of (i) the number of private alleles (fixed and variable approaches); (ii) partial heterozygosity (variable approach); and (iii) cluster assignment using STRUCTURE, a stack depth of 6× and K setting ranging from 2 to 4 (variable approach).

2.7 Summarizing allelic content to test for evidence of hybridization and shared ancestry

The Californian insectaries evaluated differed with regards to the practice of introducing conspecific wasps from field populations to augment genetic variation in their colonies, with protocols

ranging from complete replacement of the colony each year, to complete disregard for the practice.[16] Given the surprising finding that the Californian insectary and the field populations from which they were derived formed tight genetic clusters separate from each other (see Results), we looked for evidence of common ancestry and panmixia. Therefore, we summarized the number of segregating characters between the insectaries and introduced populations. First, we summarized the number of segregating sites between these populations using the concatenated SNPs data set from our fixed approach. We counted the number of sites that (i) were missing data in insectary pools but not introduced pools; (ii) were missing data in introduced pools but not insectary pools; (iii) were segregating between insectary and introduced pools; and (iv) were shared between insectary and introduced pools. The first two categories were relevant because these site patterns are consistent with the allelic drop out (i.e. null alleles) expected to occur in samples that have shared ancestry.[26] Second, we examined the distribution of missing data between these groups using a modified version of the four sample Patterson's D-statistic.[27] The D-statistic test was originally designed to identify putative introgression, but we modified it to address patterns of SNP diversity (including missing data as a possible allelic state) between insectary and introduced field populations. Based on the findings of our STRUCTURE analyses (see Results) we hypothesized that insectaries would share more nucleotide characters with each other, relative to the field populations, than would be expected by chance. Thus, we arranged genotypes in the order W_1, I_1, I_2, W_2 (where I_1 and I_2 are different insectary populations and W_1 and W_2 are different field populations), and then counted the number of SNPs with the pattern NXXN and XNXN, where X is one character state and N is the other character state. If the numbers of SNP sites with the pattern NXXN were greater than those with the pattern XNXN, this would imply that the two insectaries have more shared character states than either has with the field populations. If the number of NXXN and XNXN sites were similar, then it would imply that character states were randomly distributed across the sampled insectary and field populations. D-statistics reflect these patterns and can be significant with positive or negative values. Significantly positive values provide evidence that there are more NXXN sites than would be expected by chance, and significantly negative values provide evidence that there are more NXNX sites than would be expected by chance. To count the number of sites with each pattern, we used the R package *evobiR* 1.0 (available at: http://cran.r-project.org/web/packages/evobiR/index.html).

2.8 Testing for host and symbiont contamination

Based on STRUCTURE analyses, higher levels of missing data, and D-statistics, our results suggested genomic similarities among insectary populations. One possible explanation for this pattern could be the shared host species of the insectary populations. Whereas its natural host is *Aonidiella aurantii*, *Aphytis melinus* is reared in insectaries on a substitute host, the oleander scale *Aspidiotus nerii*.[13] Furthermore, adults of *A. melinus* are known to perform host-feeding.[28,29] Thus, the presence of DNA of the natural (California red scale) and insectary (oleander scale) host in the guts of field and insectary populations, respectively, could potentially explain genetic clustering of these two groups. To test this we created RADseq libraries for California red scale and oleander scale according to methods described by Fouet *et al.*[30] The main difference between this protocol and the one employed for *A. melinus* was that the restriction endonuclease *Nla*III (with a four-base recognition site) instead of *Sph*I (with a six-base recognition site) was used in conjunction with *Mlu*CI for the initial digestion of genomic DNA. However, the recognition site *Nla*III is the same as that of *Sph*I, but with a base removed from either end, hence any RAD site cleaved by *Sph*I is also cleaved by *Nla*III. Host libraries were also subjected only to single-end 101 bp reads. The California red scale library was based on a pooled extraction of 10 adult females collected from the biocontrol grove (Field 21) at UC Riverside, whereas that of oleander scale was prepared using a pool of ~500 immature 'crawlers' obtained from one of the California insectaries.

After quality filtering and removal of adapter contamination, host reads were exported as 'fasta' files using Stacks and concatenated into a 'RADnome' using a custom script. Each of the 16 *A. melinus* population reads were then mapped to the hosts' RADnome via the 'Map to Reference' function in Geneious 8.0.5 (BioMatters Ltd., Auckland, New Zealand) using default settings. Reads of *A. melinus* that matched host reads were removed using the 'save list of unused read' option in Geneious, which gave us a list of reads for each *A. melinus* population that did not match the hosts' RADnome. We repeated this procedure by keeping the two host 'RADnomes' separate. Following host sequence removal, we reran a subset of the analyses using both filtered read 1 and read 2 data as described above. This included estimating the number of private alleles and partial heterozygosity with the variable approach at 6× and 10× stack depth for the unfiltered data set, as well as rerunning the K = 4 structure analysis using identical settings of the fixed approach at 10× stack depth. If host contamination was a significant contributor to the patterns we observed in the unfiltered data sets, we expected to see the distinction between the insectary and field populations receive less support or vanish altogether.

Because *A. melinus* is known to be infected with *Wolbachia*[18], using the filtering procedure just described, we also checked all reads against the genome of *Wolbachia* infecting *Trichogramma pretiosum*,[31] the closest hymenopteran taxon for which a *Wolbachia* genome is currently available. We also checked against other common bacterial contaminants (e.g. *Helicobacter*) using Bowtie2[32].

2.9 Comparing allele frequencies between insectary and introduced populations in California

To compare patterns of genetic variation in insectary and field populations, respectively, we generated within-group genetic diversity values (population module of Stacks) obtained by analyzing the 953 (read1) and 107 (read 2) loci shared between Californian insectaries and field populations (using the Stacks variable approach). For each locus, we summarized the frequency of nucleotide character states for insectary and field populations separately. We then plotted these frequencies against each other and performed a regression analysis to assess the level of correlation. Our expectations were that if insectaries have lost substantial genetic diversity during laboratory domestication we would not see a significant correlation in nucleotide frequencies between insectary and field populations and, in general, would observe more fixed character states (i.e. frequencies of 0 or 1) in insectary populations.

3 RESULTS

3.1 SNP identification

Using the fixed approach and a stack depth of 10× we identified a total of 722 SNPs (661 read 1; 61 read 2) across the 16 population

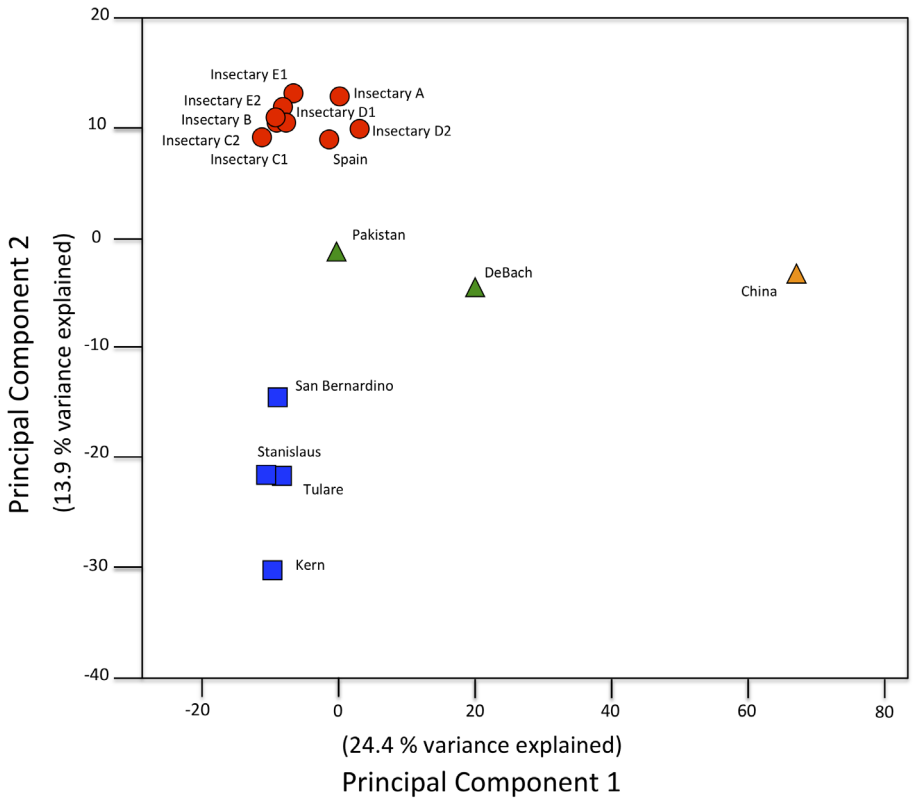

Figure 1. First two axes from principle components analysis of 722 single nucleotide polymorphisms (SNPs) among pools of *Aphytis melinus*. These SNPs were identified using the fixed approach with a stack depth of 10× (concatenated read 1 and read 2). Insectaries are indicated by red circles, introduced populations sampled in California are by blue squares, and all other samples are by green or orange triangles.

samples of *A. melinus*. Using the variable approach and a stack depth of 6× and 10×, we identified a total of 9700 SNPs (7113 read 1; 2587 read 2) and 5921 SNPs (5411 read1; 510 read2), respectively. Increasing the stack depth to 15× and 20× limited the number of SNPs to the point where subsequent analyses were uninformative (data not shown).

3.2 Genomic differentiation based on the fixed approach

The principal components analysis revealed several interesting clustering patterns (Fig. 1). The most explanatory axis (24.4%) showed clear differences between the two highly inbred colonies (DeBach and China), and between each of these and the remaining population samples. By contrast, the second axis (13.9%) revealed clear differentiation between the insectary and introduced field populations. Together, the two most explanatory axes explained 38.3% of the variance and resulted in the insectary samples (including that from Spain) clustering tightly. BIC scores from k-means clustering suggested that four clusters best describe variation within our data set (Fig. 2A). The STRUCTURE analysis using a k-value of 4 revealed that all insectary population samples formed a distinct genomic cluster, as did the introduced Californian field populations (Fig. 2B).

3.3 Partial heterozygosity based on the variable approach

Indicative of a reduction in genetic variation, we found that Californian insectary populations had far lower levels of partial heterozygosity and fewer private alleles than their field counterparts (Fig. 3A,B, Table S1). This was also the case following the filtering of contaminating host DNA (see below). The DeBach and China colonies had the lower levels of partial heterozygosity (Table S1), which is consistent with their high level of inbreeding. However, the China colony had the highest number of private alleles, which is strident with its history of nearly 20 years of inbreeding. A phylogenetic analyses (see Appendix S1 for details) placed this population as an outgroup and at the tip of a very long branch. Furthermore, the 28S-D2 rDNA of a specimen from this colony was sequenced by Gillespie *et al.*[33] and included by Dao *et al.*[34] in a phylogenetic analysis of *Aphytis*, where it clustered with two other specimens (whose pupae resembled those of another *Aphytis* species) on a distinct clade than the typical *A. melinus*[34]. This suggests that this population represents a different species, which would be consistent with its levels of genetic variation and clustering pattern (see below).

3.4 Comparisons of paired-end data variation

We found that for both analytical approaches (fixed and variable), SNPs identified from read 1 and read 2 produced almost identical patterns across the population samples in our study (Fig. 4). Specifically, across populations, we found strong positive relationships between the number of private alleles (both fixed and variable approaches) and partial heterozygosity (variable approach) whether considering read 1 or read 2 data. Furthermore, clustering patterns were nearly identical for all STRUCTURE analyses (K = 2, 3 or 4) conducted using read 1, read 2 and concatenated data sets from the variable approach data set (Fig. 5).

3.5 Allelic variation across Californian insectary and field populations

The summary of allelic patterns (across SNPs from the 6× data set) between insectary and Californian field populations revealed

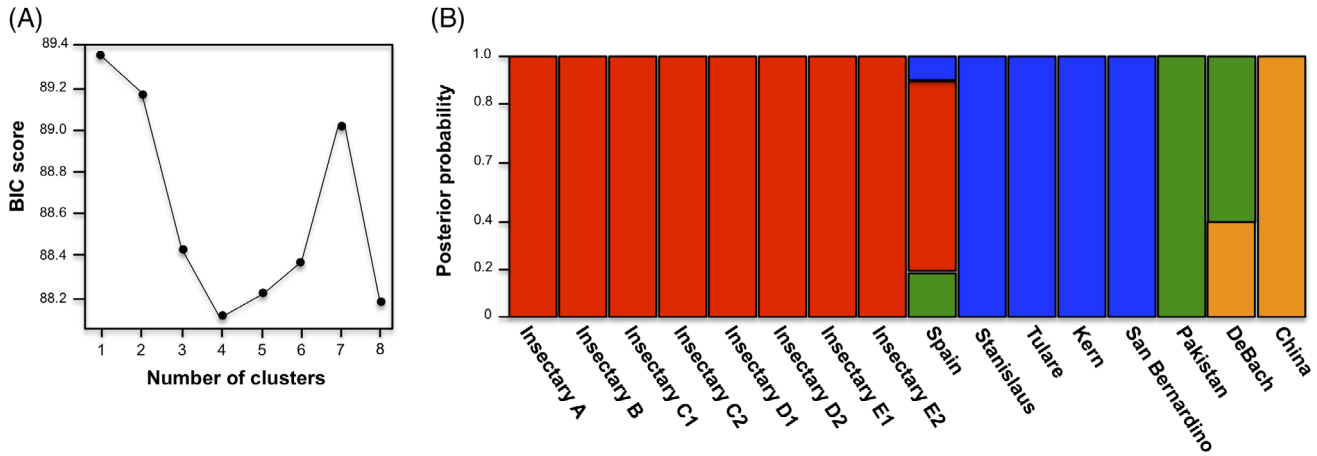

Figure 2. Based on k-means analysis (which uses Bayesian Information Criterion [BIC] scores to rank different clustering schemes) of sites fixed within pools of individuals (but variable across pools), the number of distinct genomic clusters in the *A. melinus* data set is likely 4 (A). STRUCTURE plot at K = 4 clustering level made using 722 single nucleotide polymorphisms (SNPs) obtained using the fixed approach for 16 pools of individuals with a stack depth of 10× (B).

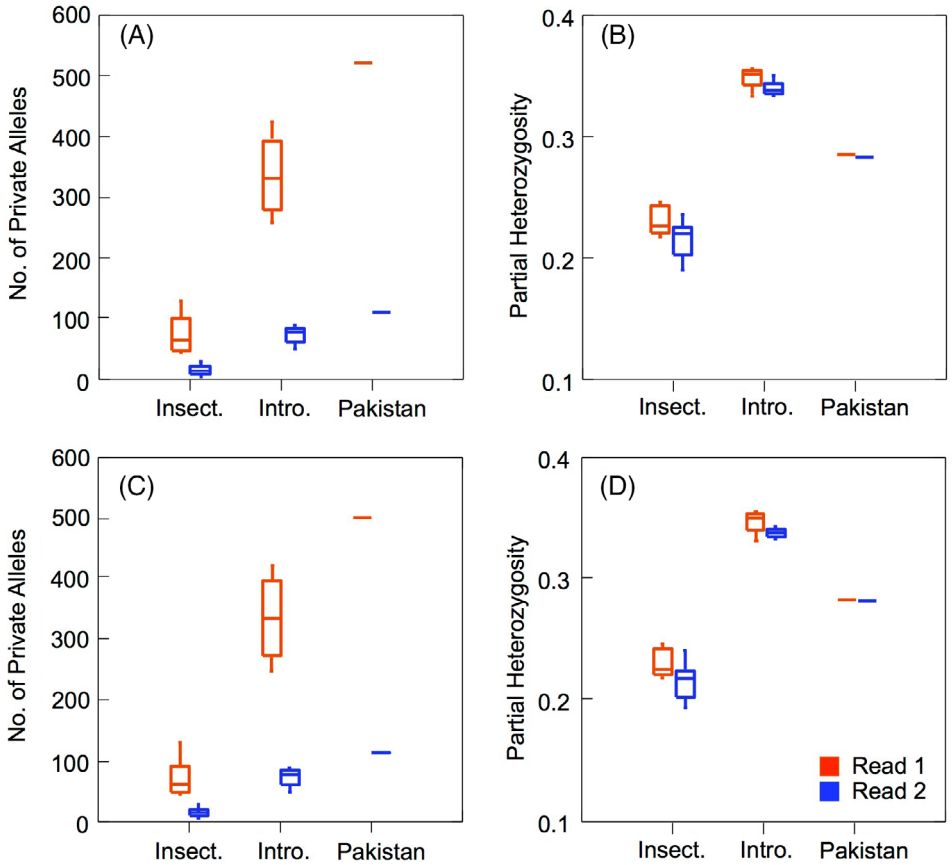

Figure 3. Number of private alleles and partial heterozygosity as inferred from read 1 and read 2 data analyzed separately for insectary and introduced pools before (A, B) and after (C, D) removal of scale host DNA using the variable approach with a stack depth of 6×. Pakistan samples are included in all comparisons for reference.

that, in terms of total numbers, the majority of alleles were shared (Fig. 6A). Following this category (but almost fourfold less common) were segregating sites that differentiated insectary and field populations (Fig. 6A). All four-sample comparisons of insectary colonies with field populations yielded significant positive D-statistics (Table 1) indicating that the insectaries were much more likely to share alleles at a given locus with another insectary population than with a field population. To better understand these results, we summarized the number of unique genotypic states including all relevant nucleotide and missing data pairs. This summary revealed that cases where missing data were observed in insectary population samples and fixed alleles were observed in field populations substantially outnumbered sites with other patterns (Fig. 6B).

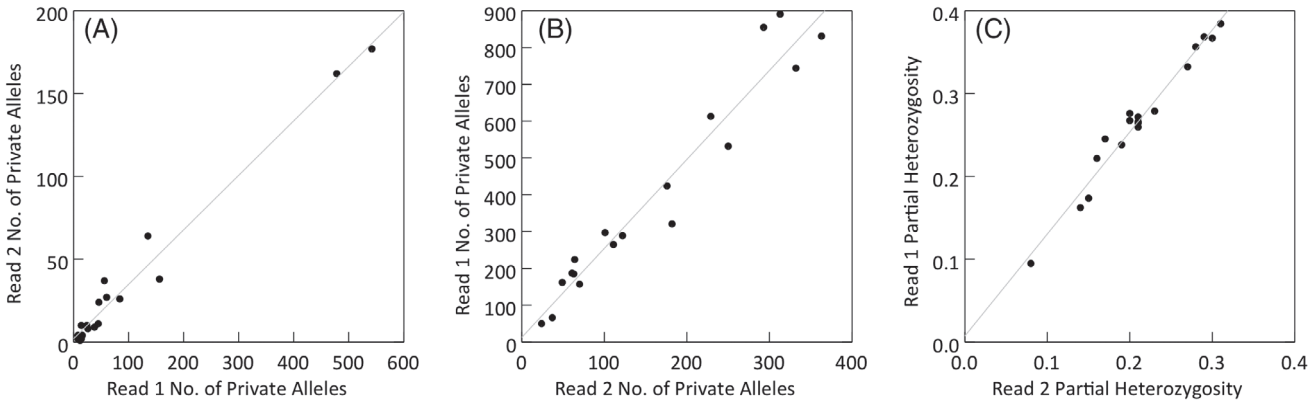

Figure 4. Linear relationships between single nucleotide polymorphism (SNP) parameters identified from read 1 and read 2 data sets. (A) Number of private alleles from fixed approach analysis. (B) Number of private alleles from variable approach analysis. (C) Partial heterozygosity from variable approach analysis.

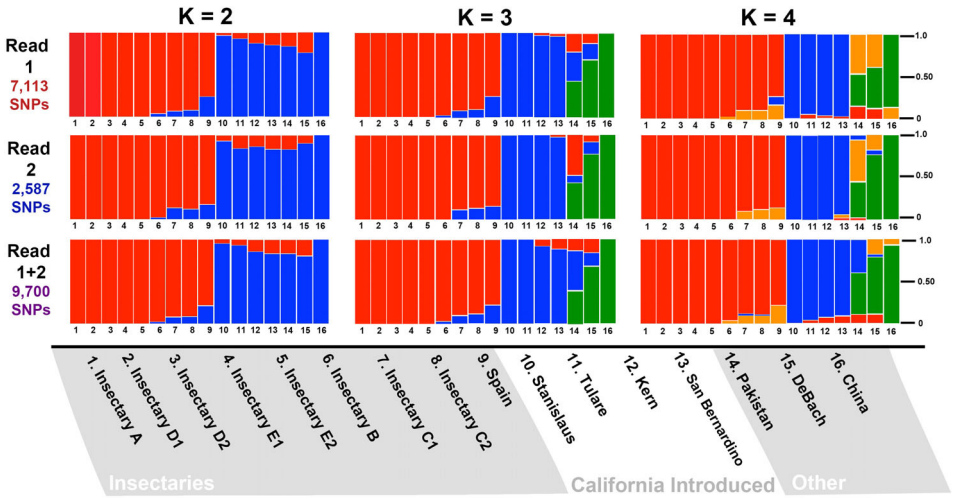

Figure 5. STRUCTURE analyses (K = 2–4) demonstrating the similar results of read 1, read 2, and concatenated data sets from the variable approach analysis when using a stack depth of 6×. Pools of individuals are labeled 1–16. This analysis was conducted before host filtering.

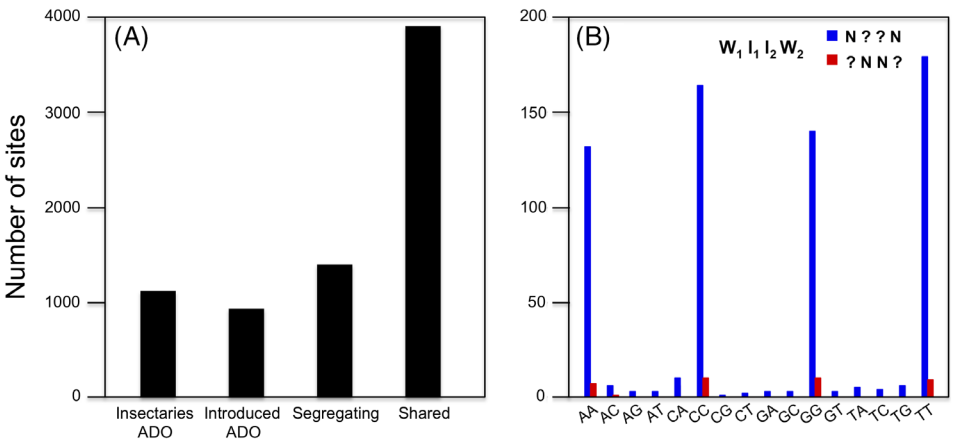

Figure 6. Site-specific patterns of allelic drop out (ADO), segregating and shared alleles between insectary and introduced pools of *Aphytis melinus* demonstrating that (A) the most common allelic pattern is the shared category and that there are over 1000 segregating sites between the groups. (B) A summary of ADO patterns (i.e. missing data) between two insectaries and two introduced pools demonstrating that the most common pattern of missing data is allelic dropout in the insectaries. This analysis was conducted before host filtering.

Table 1. Four-sample tests for random single nucleotide polymorphism patterns

Test	NXXN sites	NXNX sites	D (p)
1. W1, I1, I2, W2	680	449	0.20 (0.000)
2. W1, I2, I3, W2	838	301	0.47 (0.000)
3. W1, I3, I4, W2	842	343	0.42 (0.000)
4. W1, I4, I5, W2	665	421	0.22 (0.000)
5. W1, I5, I6, W2	759	413	0.30 (0.000)
6. W1, I6, I7, W2	1101	199	0.69 (0.000)
7. W1, I7, I8, W2	792	349	0.36 (0.000)
8. W1, I1, I2, W4	632	440	0.18 (0.000)
9. W1, I2, I3, W4	774	300	0.44 (0.000)
10. W1, I3, I4, W4	797	352	0.39 (0.000)
11. W1, I4, I5, W4	601	389	0.21 (0.000)
12. W1, I5, I6, W4	725	423	0.26 (0.000)
13. W1, I6, I7, W4	1039	206	0.67 (0.000)
14. W1, I7, I8, W4	763	356	0.36 (0.000)
15. W2, I1, I2, W3	694	251	0.47 (0.000)
16. W2, I2, I3, W3	675	453	0.20 (0.000)
17. W2, I3, I4, W3	816	192	0.62 (0.000)
18. W2, I4, I5, W3	701	297	0.40 (0.000)
19. W2, I5, I6, W3	665	283	0.40 (0.000)
20. W2, I6, I7, W3	853	472	0.29 (0.000)
21. W2, I7, I8, W3	799	231	0.55 (0.000)
22. W3, I1, I2, W4	980	254	0.59 (0.000)
23. W3, I2, I3, W4	1037	409	0.43 (0.000)
24. W3, I3, I4, W4	820	274	0.50 (0.000)
25. W3, I4, I5, W4	958	287	0.54 (0.000)
26. W3, I5, I6, W4	1128	190	0.71 (0.000)
27. W3, I6, I7, W4	907	433	0.35 (0.000)
28. W3, I7, I8, W4	913	257	0.56 (0.000)
29. W1, I2, I5, W4	769	212	0.57 (0.000)
30. W1, I4, I8, W4	797	237	0.54 (0.000)
31. W1, I3, I7, W4	753	254	0.50 (0.000)

3.6 Genetic structuring after genome filtering

The matches between the *A. melinus* and the hosts' concatenated 'RADnome' ranged between 4.9% and 6.5% for read 1 and between 0.3% and 0.9% for read 2. When the filtering was done by keeping the two host 'RADnomes' separate, it turned out that nearly all matching reads belonged to oleander scale (Table 2, Fig. 7A). However, as expected almost all detected red scale DNA was in the four field populations from California (Fig. 7B). Filtering for *Wolbachia* resulted in 0.19% of matching reads at most, whereas we did not find any read matching *Helicobacter*. Filtering resulted in an almost undetectable reduction in private alleles (Fig. 3C) and partial heterozygosity (Fig. 3D) across samples (Table S1). Similarly, structure-inferred clustering assignment patterns among the samples were not altered by the filtering process (Fig. 3C,D and Fig. S2). Thus, even though our pooled samples contained host and symbiont DNA, neither of these explained the patterns of variation we discovered in the unfiltered data sets. Interestingly, the higher coverage (10×) read 2 filtered data set had different patterns of heterozygosity than the other filtered data sets; however, this was largely due to one of the most incomplete insectary samples (two orders of magnitude lower read numbers) having a high level of heterozygosity (data not shown).

3.7 Results of allele frequency comparisons in Californian wasps

We found a positive but weak correlative relationship between allele frequencies of California insectary and field populations (read 1, $r^2 = 0.19$; read 2, $r^2 = 0.28$; Fig. S1). For read 1 and read 2, respectively, the average partial allele frequency in the insectary group was 0.43 and 0.52, whereas this value was 0.47 and 0.54 in the field group. Thus, the insectary populations had fewer unique loci than field populations (also see allelic drop out results; Fig. 6), and on average lower allele frequencies than conspecific field samples – consistent with our expectations of laboratory-induced reduction of genetic variation.

4 DISCUSSION

In this study, we used genome-wide sampling to study the effects of mass-rearing on genetic variation in a widely used biocontrol agent, *Aphytis melinus*. We found several patterns that are putatively related to domestication, including: (i) genomic similarity between all commercial insectary samples included in the study; (ii) lower numbers of private alleles in insectary populations compared with introduced and native field populations; and (iii) lower partial heterozygosity in the insectary populations relative to the field populations from which they were derived. Below we discuss our findings in the context of previous work and outline recommendations for future management of these valuable biocontrol agents.

4.1 Why do the California insectaries form a distinct genomic cluster?

One of the more salient findings of our study is that all of the commercial insectaries we sampled formed a genomic cluster (Figs 1, 2 and 6). This is a noteworthy finding because the colonies were likely established at different times, from different sources, and are reported to be augmented (or even re-established) with wasps from field populations to varying degrees.[16] In terms of raw data, this finding is supported by the large number of segregating sites observed between the insectary samples and their closest relative in our study, the Californian field populations (Fig. 6). This separation of insectary and field populations was also supported following the removal of contaminating DNA from the different armored scale hosts and common bacterial symbionts/contaminants (Fig. S2).

There are at least two, not necessarily mutually exclusive, explanations for these patterns. First, the insectaries may be genetically similar simply due to common descent. In other words, each insectary colony may have been established using a common ancestral stock, and/or the insectaries may regularly exchange wasps to bolster colony size (and with the belief that they are also boosting genetic variation within their stock). This seems to be a common practice. In a survey of 40 North American insectaries producing the green lacewing *Chrysoperla carnea* (Stephens) (Neuroptera: Chrysopidae), Henry and Wells[35] discovered that only four insectaries grew their own insects. Second, the similarities we observed could relate to strong, parallel selection for traits (genotypes) that infer higher fitness in such an artificial setting, with an excessive supply of hosts and no predation risks (e.g. development time and body size). Future investigation to test for patterns consistent with these explanations is warranted.

A further potential contributing factor is genomic coverage. Despite the apparent appropriateness of RADseq for analyzing population genomics systems like *A. melinus*, several studies have

Table 2. Number of *Aphytis melinus* reads matching reads of the natural (California red scale, *Aonidiella aurantii*) and laboratory (oleander scale, *Aspidiotus nerii*) host

Populations	Reads before host filtering	Reads after host filtering	Reads matching oleander scale*	Reads matching red scale
CA_Kern	456 312	431 007	24 911	400
CA_SBernardino	789 158	738 221	49 907	1053
CA_Stanislaus	847 671	796 119	50 537	1026
CA_Tulare	682 420	643 441	38 418	559
China	620 204	589 918	32 466	8
DeBach	757 280	716 571	12 532	2
Insectary_A	556 538	524 069	36 826	10
Insectary_B	231 801	219 267	43 578	1
Insectary_C1	616 706	579 872	25 689	2
Insectary_C2	793 904	750 325	16 270	0
Insectary_D1	460 768	435 080	29 262	4
Insectary_D2	280 331	264 061	32 907	5
Insectary_E1	541 613	512 347	40 709	0
Insectary_E2	591 432	558 522	20 398	2
Pakistan	692 955	655 197	37 754	5
Spain	386 575	366 177	30 155	130

*See comments in the main text.

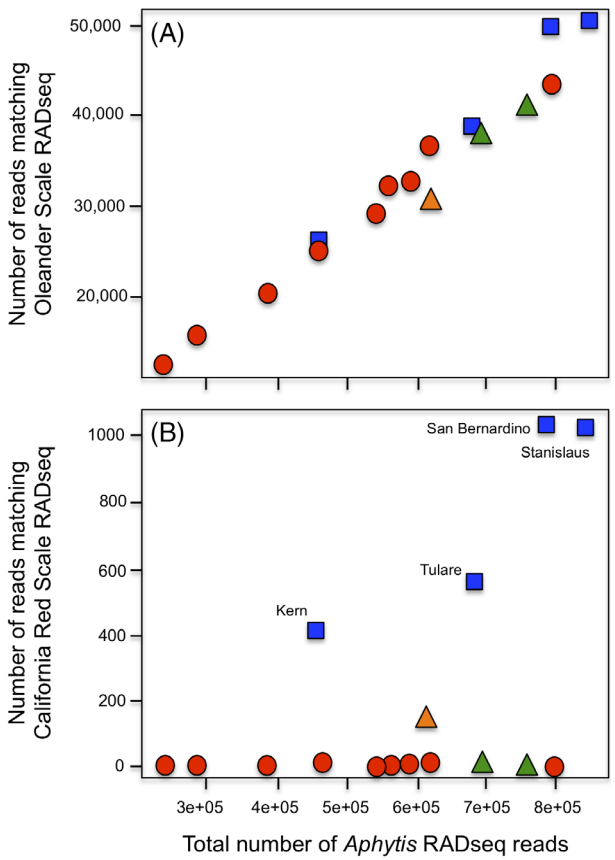

Figure 7. Read-filtering to test for host contamination from oleander scale (A; *Aspidiotus nerii*) and California red scale (B; *Aonidiella aurantii*). Number of *A. melinus* reads for each pooled sample is on the x-axis and the number of RADseq reads matching scale insect 'RADnomes' is on the *y*-axis. See text for more detail.

urged caution when using this methodology,[20,26] particularly when using RADseq protocols to identify divergent SNPs from pools of individuals. Anderson *et al.*[20] pointed out that unexpected population structure observed in RADseq analyses of pooled individuals could simply be an artifact of data coverage. To address this concern, we attempted to increase the stack depth to 15× and 20× in our fixed and variable approaches. However, we found that these criteria limited the number of SNPs to the point where subsequent analyses were uninformative (data not shown). That said, given that we observed similar clustering patterns when using a stack depth of 6× and 10× in our STRUCTURE analyses (variable versus fixed approaches; Figs 2 and 5), we do not think that the clustering of insectaries is solely an artifact of data coverage.

4.2 The distribution of missing data

Another important finding of our study was that commercial insectary population samples had more allelic drop out relative to the field population samples than would be expected by chance. We found that the most common SNP pattern between insectary and natural populations was missing data in insectaries where natural populations had fixed SNPs (Fig. 7). This pattern is curious because in the case of bottlenecks (as we might expect to occur in mass-reared colonies), the effective population size is decreased and therefore genetic diversity is decreased. Thus, under this scenario, we should have observed fewer loci with less missing data.[26] One factor that could explain the apparent high levels of allelic drop out in the insectary populations, in loci that were fixed in field populations (and vice versa), is linked to differences in the armored scale hosts on which the field and insectary populations developed. Field populations developed on *Aonidiella aurantii*, whereas insectary populations were reared on *Aspidiotus nerii*. Thus, host-feeding by *A. melinus* individuals may have resulted in contamination of our initial gDNA libraries with host DNA. This explanation is at least partially supported by the results of our host/symbiont filtering analyses in which we found that *A. aurantii* contamination was only present in the California field populations

(Fig. 7B). Given the nearly prefect linear relationship between the number of *A. melinus* and matched oleander scale reads (Fig. 7), we suspect that although the sample of 'crawlers' used to create the oleander scale library was confirmed under a microscope to be free from intact *A. melinus* specimens (or obvious body parts), it may have been contaminated with *A. melinus* DNA. Although the number of contaminating reads was relatively low (between 400 and 1000 reads), this could easily explain the N??N sites we observed, which were represented by only a few hundred sites per nucleotide motif (Fig. 6B). In light of this, we recommend that future RADseq investigations on parasitoids include a host DNA-filtering step to remove any potential for misleading patterns of missing data.

5 CONCLUSION

Aphytis melinus is currently reared by four commercial insectaries in California and sold to citrus growers as a supplement to endemic predators and parasitoids important in biological control of California red scale. Vasquez and Morse[16] discussed the importance of insectaries producing female-biased sex ratios of this wasp and attempting to maximize female wasp size, which is correlated with improved longevity and fecundity. The present study, however, suggests that despite some insectaries attempting to periodically restart or reinvigorate insectary colonies using field-collected wasps, these attempts have failed to maintain genetic diversity similar to that observed in field populations, and/or if such diversity was initially obtained, it was lost over time as colonies were maintained in culture. The solution may not be trivial because field-collected *A. melinus* colonies are initially difficult to rear (JGM, unpublished data) and selection that leads to prolific propagation under insectary conditions likely selects at the same time for reduced genetic diversity. We suggest that additional research is needed to develop strategies whereby insectary producers can simultaneously maintain genetically diverse parasitoids populations and rapidly rear large numbers of wasps for sale.

ACKNOWLEDGEMENTS

We thank Drs Cassandra Vasquez and Inamullah Khan for contributing *A. melinus* collections that were used for parts of this research, and Len Nunney for commenting on an earlier draft of the manuscript. We also thank the California Citrus Research Board for funding, in part, which lead to aspects of this publication.

DATA ACCESSIBILITY

The sequencing reads are archived at the NCBI Sequence Read Archive, BioProject PRJNA312209, accession number SRP070197.

SUPPORTING INFORMATION

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

REFERENCES

1 Mackauer M, Genetic aspects of insect production. *Entomophaga* **17**:27–48 (1972).
2 Mackauer M, Genetic problems in the production of biological control agents. *Annu Rev Entomol* **21**:369–385 (1976).
3 Chambers DL, Quality-control in mass rearing. *Annu Rev Entomol* **22**:289–308 (1977).
4 Nunney L, Managing small populations: a population genetic perspective, in *Quality Control and Production of Biological Control Agents: Theory and Testing Procedures*, ed. by Van Lenteren JC. CABI Publishing, Wallingford, UK, pp. 73–87 (2003).
5 Gilchrist AS and Meats AW, An evaluation of outcrossing to improve mass-reared strains of the Queensland fruit fly *Bactrocera tryoni*. *Int J Trop Insect Sci* **34**:S35–S44 (2014).
6 Nunney L, in *Captive Rearing for Field Release: A Population Genetic Perspective in Encyclopedia of Pest Management*, ed. by Pimentel D. Taylor & Francis, London, pp. 1–4 (2006).
7 Miyatake T, Genetic changes of life history and behavioral traits during mass-rearing in the melon fly, *Bactrocera cucurbitae* (Diptera : Tephritidae). *Res Popul Ecol* **40**:301–310 (1998).
8 Miyatake T, Quantitative genetic aspects of the quality control of mass-reared insects: the case of the melon fly (*Bactrocera cucurbitae*). *Formosan Entomol* **26**:307–318 (2006).
9 Kuriwada T, Kumano N, Shiromoto K and Haraguchi D, Effect of mass rearing on life history traits and inbreeding depression in the sweetpotato weevil (Coleoptera: Brentidae). *J Econ Entomol* **103**:1144–1148 (2010).
10 Francuski L, Djurakic M, Ludoški J, Hurtado P, Pérez-Bañón C, Ståhls G *et al.*, Shift in phenotypic variation coupled with rapid loss of genetic diversity in captive populations of *Eristalis tenax* (Diptera: Syrphidae): Consequences for rearing and potential commercial use. *J Econ Entomol* **107**:821–832 (2014).
11 Gilchrist AS, Cameron EC, Sved JA and Meats AW, Genetic consequences of domestication and mass rearing of pest fruit fly *Bactrocera tryoni* (Diptera: Tephritidae). *J Econ Entomol* **105**:1051–1056 (2012).
12 DeBach P, New species and strains of *Aphytis* (Hymenoptera: Aphelinidae) parasitic on the California red scale, *Aonidiella aurantii* (Mask.), in the Orient. *Ann Entomol Soc Am* **52**:354–362 (1959).
13 DeBach P and White EB, Commercial mass culture of the California red scale parasite *Aphytis lingnanensis*. *Calif Agric Exp Stn Bul* **770**:4–58 (1960).
14 Rill S, Grafton-Cardwell EE and Morse JG, Effects of two insect growth regulators and a neonicotinoid on various life stages of *Aphytis melinus* (Hymenoptera: Aphelinidae). *BioControl* **53**:579–587 (2008).
15 Morse JG and Grafton-Cardwell EE, Evolution of biologically-based integrated pest management in California citrus. *Univ Calif Plant Prot* **Q17**:1–11 (2012).
16 Vasquez C and Morse JG, Fitness components of *Aphytis melinus* (Hymenoptera: Aphelinidae) reared in five California insectaries. *Environ Entomol* **41**:51–58 (2012).
17 Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE, Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* **7**:e37135 (2012).
18 Vasquez CJ, Stouthamer R, Jeong G and Morse JG, Discovery of a CI-inducing *Wolbachia* and its associated fitness costs in the biological control agent *Aphytis melinus* DeBach (Hymenoptera: Aphelinidae). *Biol Control* **58**:192–198 (2011).
19 Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A *et al.*, Estimation of population allele frequencies from next-generation sequencing data: pool- versus individual-based genotyping. *Mol Ecol* **22**:3766–3779 (2013).
20 Anderson EC, Skaug HJ and Barshis DJ, Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Mol Ecol* **23**:502–512 (2014).
21 Davey JW, Cézard T, Fuentes-Utrilla P, Eland C, Gharbi K and Blaxter ML, Special features of RAD sequencing data: implications for genotyping. *Mol Ecol* **22**:3151–3164 (2013).
22 Catchen J, Hohenlohe P, Bassham S, Amores A and Cresko W, Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**:3124–3140 (2013).
23 Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A *et al.*, PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**:e15925 (2011).
24 Jombart T, adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**:1403–1405 (2008).
25 Pritchard JK, Stephens M and Donnelly P, Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959 (2000).

26 Arnold B, Corbett-Detig RB, Hartl D and Bomblies K, RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol* **22**:3179–3190 (2013).
27 Durand EY, Patterson N, Reich D and Slatkin M, Testing for ancient admixture between closely related populations. *Mol Biol Evol* **28**:2239–2252 (2011).
28 Collier TR, Host feeding, egg maturation, resorption, and longevity in the parasitoid *Aphytis melinus* (Hymenoptera: Aphelinidae). *Ann Entomol Soc Am* **88**:206–214 (1995).
29 Heimpel GE, Rosenheim JA and Kattari D, Adult feeding and lifetime reproductive success in the parasitoid *Aphytis melinus*. *Entomol Exp Appl* **83**:305–315 (1997).
30 Fouet C, Kamdem C, Gamez S and White BJ, Extensive genetic diversity among populations of the malaria mosquito *Anopheles moucheti* revealed by population genomics. *Infect Genet Evol* **48**:27–33 (2017).
31 Lindsey ARI, Werren JH, Richards S and Stouthamer R, Comparative genomics of a parthenogenesis-inducing *Wolbachia* symbiont. *G3 (Bethesda)* **6**:2113–2123 (2016).
32 Langmead B and Salzberg SL, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359 (2012).
33 Gillespie J, Munro J, Heraty J, Yoder M, Owen A and Carmichael A, A secondary structural model of the 28S rRNA expansion segments D2 and D3 for chalcidoid wasps (Hymenoptera: Chalcidoidea). *Mol Biol Evol* **22**:1593–1608 (2005).
34 Dao HT, Beattie GAC, Spooner-Hart R, Riegler M and Holford P, Primary parasitoids of red scale (*Aonidiella aurantii*) in Australia and a review of their introductions from Asia. *Insect Sci* **24**:150–168 (2015).
35 Henry CS and Wells MM, Can what we don't know about lacewing systematics hurt us? *Am Entomol* **53**:42–47 (2007).

