

Article

Population Genomics of Pooled Samples: Unveiling Symbiont Intrapopulation Diversity and Host–Symbiont Coevolution

Alix E. Matthews ^{1,2,*} , Than J. Boves ², Katie L. Percy ^{3,4}, Wendy M. Schelsky ^{5,6}  and Asela J. Wijeratne ² 

¹ College of Sciences and Mathematics and Molecular Biosciences Program, Arkansas State University, Jonesboro, AR 72401, USA

² Department of Biological Sciences, Arkansas State University, Jonesboro, AR 72401, USA; tboves@astate.edu (T.J.B.); awijeratne@astate.edu (A.J.W.)

³ Audubon Delta, National Audubon Society, Baton Rouge, LA 70808, USA; katie.percy@usda.gov

⁴ United States Department of Agriculture, Natural Resources Conservation Service, Addis, LA 70710, USA,

⁵ Department of Evolution, Ecology, and Behavior, School of Integrative Biology, University of Illinois, Urbana-Champaign, Champaign, IL 61801, USA; schelsky@illinois.edu

⁶ Prairie Research Institute, Illinois Natural History Survey, University of Illinois, Urbana-Champaign, Champaign, IL 61820, USA

* Correspondence: alix.matthews@smail.astate.edu

Abstract: Microscopic symbionts represent crucial links in biological communities. However, they present technical challenges in high-throughput sequencing (HTS) studies due to their small size and minimal high-quality DNA yields, hindering our understanding of host–symbiont coevolution at microevolutionary and macroevolutionary scales. One approach to overcome those barriers is to pool multiple individuals from the same intrapopulation (i.e., individual host) and sequence them together (Pool-Seq), but individual-level information is then compromised. To simultaneously address both issues (i.e., minimal DNA yields and loss of individual-level information), we implemented a strategic Pool-Seq approach to assess variation in sequencing performance and categorize genetic diversity (single nucleotide polymorphisms (SNPs)) at both the individual-level and intrapopulation-level for microscopic feather mites. To do so, we collected feathers harboring mites (Proctophyllodidae: *Amerodectes protonotaria*) from four individual Prothonotary Warblers (Parulidae: *Protonotaria citrea*). From each of the four hosts (i.e., four mite intrapopulations), we conducted whole-genome sequencing on three extraction pools consisting of different numbers of mites (1 mite, 5 mites, and 20 mites). We found that samples containing pools of multiple mites had more sequencing reads map to the feather mite reference genome than did the samples containing only a single mite. Mite intrapopulations were primarily genetically structured by their associated individual hosts (not pool size) and the majority of SNPs were shared by all pools within an intrapopulation. Together, these results suggest that the patterns observed are driven by evolutionary processes occurring at the intrapopulation level and are not technical signals due to pool size. In total, despite the challenges presented by microscopic symbionts in HTS studies, this work highlights the value of both individual-level and intrapopulation-level sequencing toward our understanding of host–symbiont coevolution at multiple evolutionary scales.

Keywords: Astigmata; diversity; ectosymbionts; feather mites; high-throughput sequencing; low-input DNA



Citation: Matthews, A.E.; Boves, T.J.; Percy, K.L.; Schelsky, W.M.; Wijeratne, A.J. Population Genomics of Pooled Samples: Unveiling Symbiont Intrapopulation Diversity and Host–Symbiont Coevolution. *Life* **2023**, *13*, 2054. <https://doi.org/10.3390/life13102054>

Academic Editors: Koichiro Tamura and Jorge Doña

Received: 29 June 2023

Revised: 30 September 2023

Accepted: 10 October 2023

Published: 14 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Symbiotic relationships—broadly defined as two different species living in close physical or physiological association with one another—are among the most common interactions in the natural world [1]. These intimate interactions play a fundamental role in an individual's health [2], in biological community structure [3], and in the entire functioning of ecosystems [4]. Consequently, symbionts (i.e., organisms that form parasitic,

mutualistic, or commensal interactions with hosts [5]) have strong effects on biodiversity, and symbionts themselves comprise the majority of species on Earth [6,7]. Despite their ubiquity and importance in the natural world, basic biological knowledge about many symbionts is still lacking, including their ecology and coevolutionary history with hosts.

Technical challenges associated with studying symbionts add to these knowledge gaps and can consequently hinder our understanding of symbionts and their relationships with hosts at multiple evolutionary scales. One major limitation is that there are few taxonomic and morphological experts for most groups of symbionts, so basic research on species identification, description, and biodiversity is subjected to a severe bottleneck [8,9]. Fortunately, the field of molecular genetics has helped overcome some of the taxonomic expertise limitations and has aided in rapid species discovery and augmented species identification [10,11]. In particular, the application of whole-genome high-throughput sequencing (HTS) has transformed our understanding of host–symbiont macroevolutionary processes and patterns. For example, coevolutionary studies applying whole-genome HTS have unveiled major historical host-switching events between distantly related taxa, which has revolutionized our understanding of symbiont diversification and geographic distribution [12,13]. The application of HTS has also allowed us to explore genomic underpinnings of the variation (or lack thereof) in symbiont host specificity [14,15], providing us fine-scale insight into symbiont–host ecology and coevolution. However, genome-wide HTS studies for microscopic symbionts are often limited by the minimal amount of high-quality DNA that can be extracted for downstream use [16]. The lack of high-quality reference genomes for non-model organisms (as is the case with many undescribed symbionts [17]) further complicates genome-wide investigations. These challenges have not only limited our understanding of macroevolutionary patterns, but also the microevolutionary processes that drive those patterns, such as dispersal and gene flow.

One approach to surpass the barrier of low DNA yield/quality for microscopic symbionts in a HTS context is to pool multiple individuals into a single DNA extraction prior to sequencing (i.e., pooled sequencing or Pool-Seq). This approach can be strategically implemented based on budget, goals of the study, and biology of the system [18]. In the case of symbionts, individuals from the same population on an individual host (i.e., intrapopulation or deme) can be pooled and sequenced together. Genomic information gleaned from HTS at the intrapopulation level can provide insight not only on processes occurring on a much finer scale, such as natural selection and genetic drift, but also about the role of life history and ecology in those processes from both the symbiont and host perspectives [19]. However, valuable information about an individual's genetic contribution to the intrapopulation is lost with Pool-Seq [20]. To address this issue, a strategic pooled-sequencing approach can be applied to explore both individual-level and intrapopulation-level diversity simultaneously, such that individuals from the same intrapopulation can be sequenced in different sized pools (e.g., 1–50 individuals from a single intrapopulation). Variations of this approach have been applied to assess the accuracy and cost-effectiveness of individual sequencing versus pooled sequencing in both model (e.g., *Drosophila melanogaster* [21]; *Apis mellifera* [22]) and non-model (e.g., pine processionary moth, *Thaumetopoea pityocampa* [23]) species. These studies have established that pooled sequencing is a comparable approach to individual sequencing and can accurately (and cost-effectively) capture the genetic diversity of a population. However, this type of pooled HTS approach has not been applied in a host–symbiont context, where it may provide valuable insight regarding the coevolution of symbioses at both the individual and intrapopulation levels.

Vane-dwelling feather mites (Acari: Acariformes: Astigmata: Analgoidea, Pterolichoidea) represent an ideal system with which to implement both individual-level and intrapopulation-level sequencing. Feather mites are ubiquitous, obligate symbionts of birds [24,25] with variable intrapopulation sizes [26–28]. Although they are dispersal-limited and typically vertically transmitted [29,30], macroevolutionary (i.e., species level) studies have demonstrated that they often do not perfectly coevolve with their hosts. Instead, there is evidence for clade-limited host switching and a wide range of host specificity,

which could be associated with the differential life histories of mites [31] or hosts [32–34]. Microevolutionary studies (i.e., intrapopulation level) are rare and have been restricted to barcoding markers [35,36], limiting our understanding of their intrapopulation genetic diversity to a small number of conserved loci instead of genome-wide. This is at least partly due to their minute size (0.3–0.7 mm in length [37]), which makes obtaining sufficient amounts of high-quality DNA for whole-genome HTS endeavors challenging, particularly for individual mites. As a result, individual-level and intrapopulation-level feather mite diversity and the importance of that diversity with respect to coevolution with their hosts remain largely underexplored.

In this study, we investigated the individual and intrapopulation genetic diversity of a symbiotic feather mite species, *Amerodectes protonotaria* Hernandez 2018 (Astigmata: Proctophyllodidae), through a pooled HTS approach. *Amerodectes protonotaria* is a host-specialist mite found thus far only on *Protonotaria citrea* Boddaert 1783 (Prothonotary Warbler; Parulidae) [28,33,38,39] and is the only feather mite species with a fully sequenced and assembled genome [40]. We sequenced DNA from individual mites as well as various sized pools of mites from the same intrapopulation and then compared the (1) sequencing and read-mapping performance on low-input samples obtained from the different amounts of tissue (i.e., different numbers of individual mites) and (2) genetic diversity across and within datasets. We then discussed the insights gleaned from these different datasets with respect to basic biology, ecology, and (co)evolution of small symbiotic organisms such as feather mites, as well as the advantages and disadvantages of applying these techniques in (co)evolutionary studies.

2. Materials and Methods

2.1. Field Collections

We collected *Amerodectes protonotaria* from four adult male Prothonotary Warblers in Illinois (Big Cypress Access Nature Preserve; $n = 2$) and Louisiana (BREC's Bluebonnet Swamp Nature Center; $n = 2$), USA. All birds were captured during the breeding season with mist nets using audio playback lures. We collected one to two tail feathers harboring mites and placed them in individually labeled containers. Mites were later isolated from feathers under a stereomicroscope and preserved in 95% ethyl alcohol at $-20\text{ }^{\circ}\text{C}$ until DNA extraction. Birds were captured and ethically handled under the necessary and appropriate federal and local permits (see Acknowledgements). Field collection data are listed in Table S1.

2.2. DNA Extraction, Library Preparation, and Sequencing

We conducted three genomic DNA (gDNA) isolations of mites per individual host (i.e., intrapopulation of mites) for a total of 12 isolations. From each intrapopulation, we randomly selected mites to be added to different sized gDNA isolation pools. Specifically, from each of the four intrapopulations, we isolated gDNA from: (1) a single mite (one female); (2) five mites pooled together (three females and two males); and (3) 20 mites pooled together (15 females and five males). Although lower than the recommended pool size of 40 individuals [18], we wanted to determine the lowest ranges of input tissue needed for whole-genome sequencing experiments (e.g., microscopic symbionts). Generally, around 0.5 ng of gDNA can be obtained from a single (female) *A. protonotaria* extraction (A.E.M., personal observations). We used the QIAamp DNA Micro Kit (QIAGEN, Germantown, MD, USA) tissue sample protocol, with slight modifications that have been previously described [33,38]. Briefly, our modifications included the following: using a sterile mounting needle to pierce a minute hole in the mite exoskeletons (for the samples containing one and five mites) or using a small sterile pestle to crush mites (for the samples containing 20 mites) while in lysis buffer; instead of vortexing samples, we gently pipetted and/or inverted them; for elution, we preheated the elution buffer and extended both the incubation and centrifugation steps.

Library construction and sequencing were carried out at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. The shotgun libraries were constructed from up to 5 ng of DNA after sonication with a Covaris ME220 sonicator (Covaris Inc., Woburn, MA, USA) to an average fragment size of 400 bp using the UltraLow-Input DNA Library Construction kit (Tecan, Morrisville, NC, USA). The dual-indexed libraries were amplified with 8 cycles of PCR and run on a Fragment Analyzer (Agilent, Santa Clara, CA, USA). Libraries were pooled in equimolar concentration and further quantitated via qPCR on a BioRad CFX Connect Real-Time System (Bio-Rad Laboratories, Hercules, CA, USA). The libraries were loaded on an Illumina NovaSeq 6000 SP lane and sequenced for 151 cycles from both ends of the fragments (i.e., paired-end sequencing). The fastq read files were generated and demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina, San Diego, CA, USA). Raw reads were deposited to the NCBI SRA database (Table S1).

2.3. Sequence Quality Control and Species Validation

Demultiplexed reads were trimmed of Illumina sequencing adapters first using cutadapt version 3.0 [41] with a Phred score quality threshold of 30. Using BBduk (<https://sourceforge.net/projects/bbmap/>, accessed on 6 September 2021), we further trimmed the 3' end of resulting reads to 140 bases and removed duplicate reads. The resulting reads were then analyzed using FastQC version 0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, accessed on 6 September 2021).

To confirm that the mites isolated from feathers were indeed *Amerodectes protonotaria*, we used the automated Target Restricted Assembly Method (aTRAM) version 2.3.4 [42] to target and assemble the cytochrome c oxidase subunit 1 (*cox1*) gene from our quality-controlled reads. We first conducted a BLASTn on the resulting *cox1* sequences to confirm that samples had high sequence similarity to *A. protonotaria*. All but one sample (PROW953_R1; 1 mite sample from PROW953) had high sequence similarity with *A. protonotaria* (>99% identity over 1190 bp); PROW953_R1 had high sequence similarity to multiple bacteria of the family Burkholderiaceae. For this sample, we mapped reads to the top three bacterial matches using BWA MEM [43]. We extracted the unmapped (i.e., non-bacterial) reads for use in aTRAM; after removing these bacterial reads, the top *cox1* BLASTn result was *A. protonotaria* for this sample (>99% identity over 1190 bp). We then aligned the *cox1* sequences from all 12 samples sequenced herein along with previously curated *cox1* sequences of *A. protonotaria* and other closely related (other *Amerodectes* spp. and a *Tyrannidectes* sp., from the same subfamily as *Amerodectes* (Pterodectinae)) and distantly related (*Proctophyllodes* spp., from a different subfamily in Proctophyllodidae (Proctophyllodinae) and a *Trouessartia* sp., from a different family of feather mites (Trouessartiidae)) feather mite lineages collected from parulid warblers [33,38]. We aligned sequences in MAFFT version 7.503 [44] and estimated a maximum likelihood (ML) phylogeny in IQ-Tree version 1.6.12 [45] and ModelFinder [46] with 1000 ultrafast bootstrap replicates as measures of branch support. We confirmed species identity if samples fell clearly within the *A. protonotaria* clade.

2.4. Read Mapping and Variant Detection

High-quality reads were mapped to the *A. protonotaria* draft reference genome (PROW981_R1; the single-mite sample from PROW981 in this study [40]) using BWA MEM and sorted using SAMtools version 1.13 [47]. As a read-mapping comparison, we also mapped high-quality reads for all samples to three Burkholderiaceae genomes, which were highly abundant in one sample (see previous section). We marked PCR duplicates using picard tools version 2.17.10 within the Genome Analysis Toolkit (GATK) version 4.2.6.1 [48]. We applied several hard filters to the resulting reads. We first removed low-quality mapped reads by skipping alignments with a MAPQ score < 20 in SAMtools version 1.15.1. We then calculated the average read depth for each species (~15X) using SAMtools

to determine an appropriate maximum read depth (50) for generating mpileup files using BCFtools version 1.15.1 [47].

We conducted joint genotyping to call SNPs on the resulting mpileup files using BCFtools. We excluded low-genotyping-quality sites (<30) and sites that were different from the reference but equal to each other. We included a minor allele frequency threshold of 0.05 and excluded sites that had depth <10 using BCFtools. We excluded sites that were not present in 100% of individuals (i.e., we allowed 0% missing data) using vcftools version 0.1.15 [49]. Although linkage is difficult to estimate for Pool-Seq data given the loss of haplotype information [18], we removed potentially linked SNPs above two contrasting r^2 thresholds (0.01 and 0.99) with sliding windows of 100 SNPs using a step size of 50 in PLINK version 1.90b5.2 [50] as a comparative exercise to explore the effect on our results. Many other analytical frameworks exist for inferring different population genetic statistics from Pool-Seq data [51–53] with variable benchmarking results [54,55], which alludes to the difficulty of untangling Pool-Seq data [18,23,56]. Our strategy described above is highly conservative, utilizes well-maintained public programs, and allows us to more directly compare population structure [57]. We visually assessed the impact of the different linkage thresholds through principal components analyses (PCA) using PLINK. PCAs were visualized in R version 3.6.3 [58] using tidyverse version 3.1 [59], ggplot2 version 3.3.5 [60], and ggpubr version 0.4.0 [61]. We used sambamba version 0.8.2 [62] to calculate the read depth at each SNP site for both linkage thresholds. We also explored how the total number of SNPs varied within and between samples under the different linkage filtering parameters. Lastly, UpSet plots [63] were created in ggupsetr version 0.3.0 [64] to visualize how many SNPs were shared within and between samples and to identify which SNPs were shared.

As a preliminary analysis to explore the potential function of the genes in which SNPs are located, we conducted a BLASTx on FlyBase [65] against the well-annotated genome database of *Drosophila melanogaster* (Insecta: Drosophilidae). We used the UpSet plot matrix interactions to identify SNP sites that were common among all samples ($n = 6$) under both linkage thresholds. Our BLASTx queries were regions of our reference genome that consisted of 100 bp upstream and 100 bp downstream of each SNP site (201 bp total). We investigated the top hit polypeptides based on the lowest bit scores and E-values. From these BLAST hits and associated coordinates, we manually inspected the alignments to determine whether the SNP was located within a protein domain. We also identified the associated gene(s) and their Gene Ontology (GO) Annotations on FlyBase.

3. Results

3.1. Sequencing Results and Phylogenetic Analysis

All samples were successfully sequenced. All samples, regardless of pool size (e.g., 1, 5, or 20 mites), produced a relatively even number of reads both before ($15,172,758 \pm 515,109$; average $\pm$ standard error (SE)) and after ($9,915,847 \pm 311,391$) quality control measures were implemented (e.g., excluding reads with Phred quality <30 and duplicated reads). This read evenness was also prevalent within each inrapopulation both before and after quality control (Table S1). The best fit model of nucleotide substitution for the *cox1* ML phylogenetic analysis was TIM3+F+I+G4 according to AICc. As expected, all mites were identified as *Amerodectes protonotaria* (Figure S1).

3.2. Read Mapping and Variant Detection

Initial read mapping (i.e., before removing low-quality mapped reads) varied across samples using the feather mite reference genome and bacterial genomes (Figure 1). Using the feather mite reference genome, read mapping varied from 41.8% to 96.5% (average of 70.7% with SE of 6.5%). For three of the four one-mite (individually sequenced) samples, less than 60% of reads mapped the feather mite reference; the average across all four one-mite samples was $61.9\% \pm 10.9\%$. Average read mapping to the feather mite genome increased to $71.5 \pm 11.5\%$ with the five-mite pools and increased further to $78.5 \pm 12.8\%$

with the 20-mite pools. With an average of about 7 million reads mapped per sample, and given the genome size (59.6 Mb), this amounts to an average sequencing depth of 33X. Conversely, using the Burkholderiaceae genomes as a reference, read mapping varied from 2.6% to 56.6% across all samples (average: $28.6 \pm 6.4\%$). Average read mapping to these bacteria generally decreased as pool sizes increased, from $39.4 \pm 11.7\%$ (1-mite samples) to $28.9 \pm 12\%$ (5-mite pools) to $17.4 \pm 10\%$ (20-mite pools). Only reads that passed the read mapping quality filter (i.e., mapping score > 20 to feather mite genome) were used for downstream SNP calling.

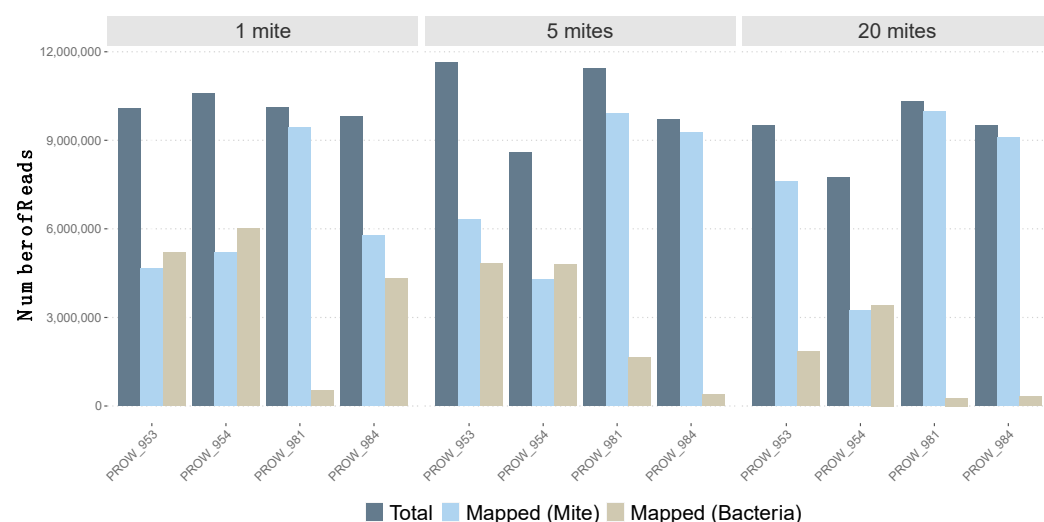


Figure 1. Number of total reads (after quality control, in dark blue) for each sample versus the number of mapped reads to both the feather mite genome reference (light blue) and three Burkholderiaceae bacterial genomes (tan).

A total of 3,338,797 candidate genome-wide SNPs were initially identified prior to any filtering. After removing low-quality sites and sites different from the reference but identical to one another, a total of 2,735,601 candidate SNPs remained. A total of 2,706,989 SNPs remained after setting the MAF threshold to 0.05 and depth to 10X. Before removing missing genotype data, 41,108 SNPs remained with a linkage threshold of 0.01, whereas 250,142 SNPs remained with a linkage threshold of 0.99. The number of SNPs (and variation in the number of SNPs) differed both within and across infrapopulations prior to removing missing genotype data. Generally, the pattern followed expectations of the fewest SNPs in the one-mite samples, an intermediate number of SNPs in the five-mite samples, and the greatest number of SNPs in the 20-mite samples (Figure 2).

After filtering missing genotype data, only 737 SNPs remained. Implementing a conservative linkage threshold of 0.01 retained 235 SNPs. The variation in the number of SNPs per sample followed a similar pattern in the filtered dataset as the non-filtered dataset (Figure 3A). In other words, the number of SNPs increased with the number of individuals in the pool. While additional SNPs were retained using the more lenient linkage threshold of 0.99 ($n = 439$), the pattern of variation across samples was nearly identical to the dataset using a 0.01 linkage threshold (Figure 3B).

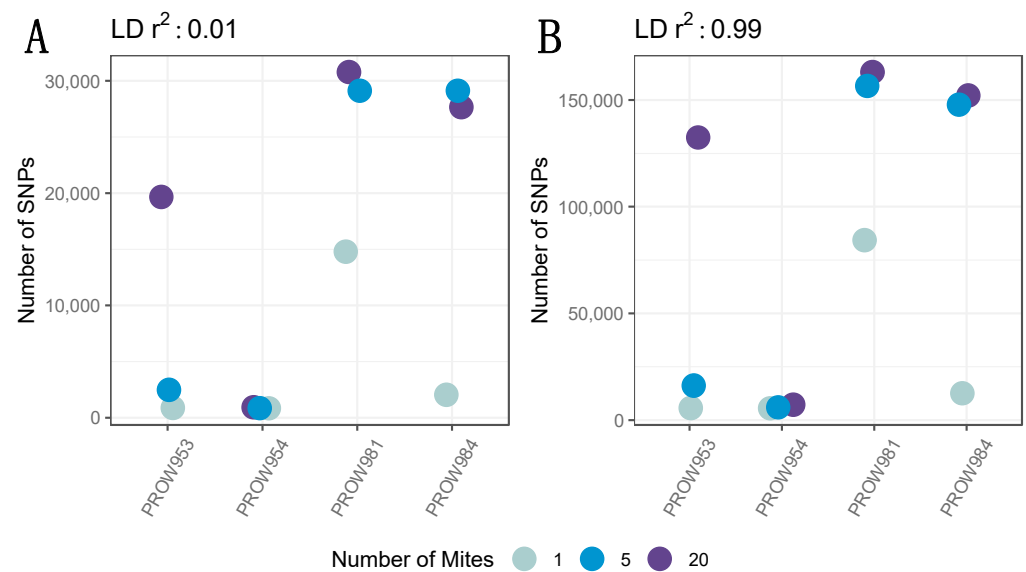


Figure 2. The number of SNPs across intrapopulations (labeled as the individual hosts) in which genotyping rate was not considered (i.e., all SNPs were included) with a linkage threshold of (A) 0.01 and (B) 0.99. Colors represent the number of mites in each pool and points are jittered around each individual host.

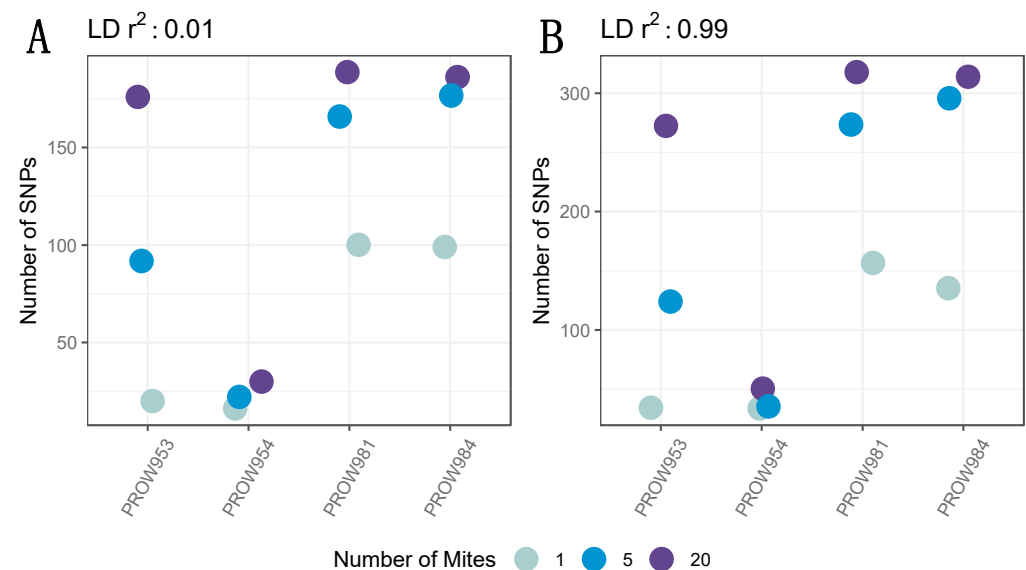


Figure 3. The number of SNPs across intrapopulations (labeled as the individual hosts) after filtering missing genotyping data with a linkage threshold of (A) 0.01 and (B) 0.99. Colors represent the number of mites in each pool and points are jittered around each individual host.

SNP coverage varied within and across samples (Figure 4, rightmost panels). Average SNP coverage for the filtered dataset (i.e., no missing genotype data) varied between linkage thresholds, but the median SNP coverage values were similar (Figure 4, leftmost panels). Specifically, the SNPs retained using a more conserved linkage threshold of 0.01 had an average coverage of 59X (median = 38X; Figure 4A), whereas the SNPs retained using a linkage threshold of 0.99 had an average coverage of 96X (median = 40X; Figure 4B).

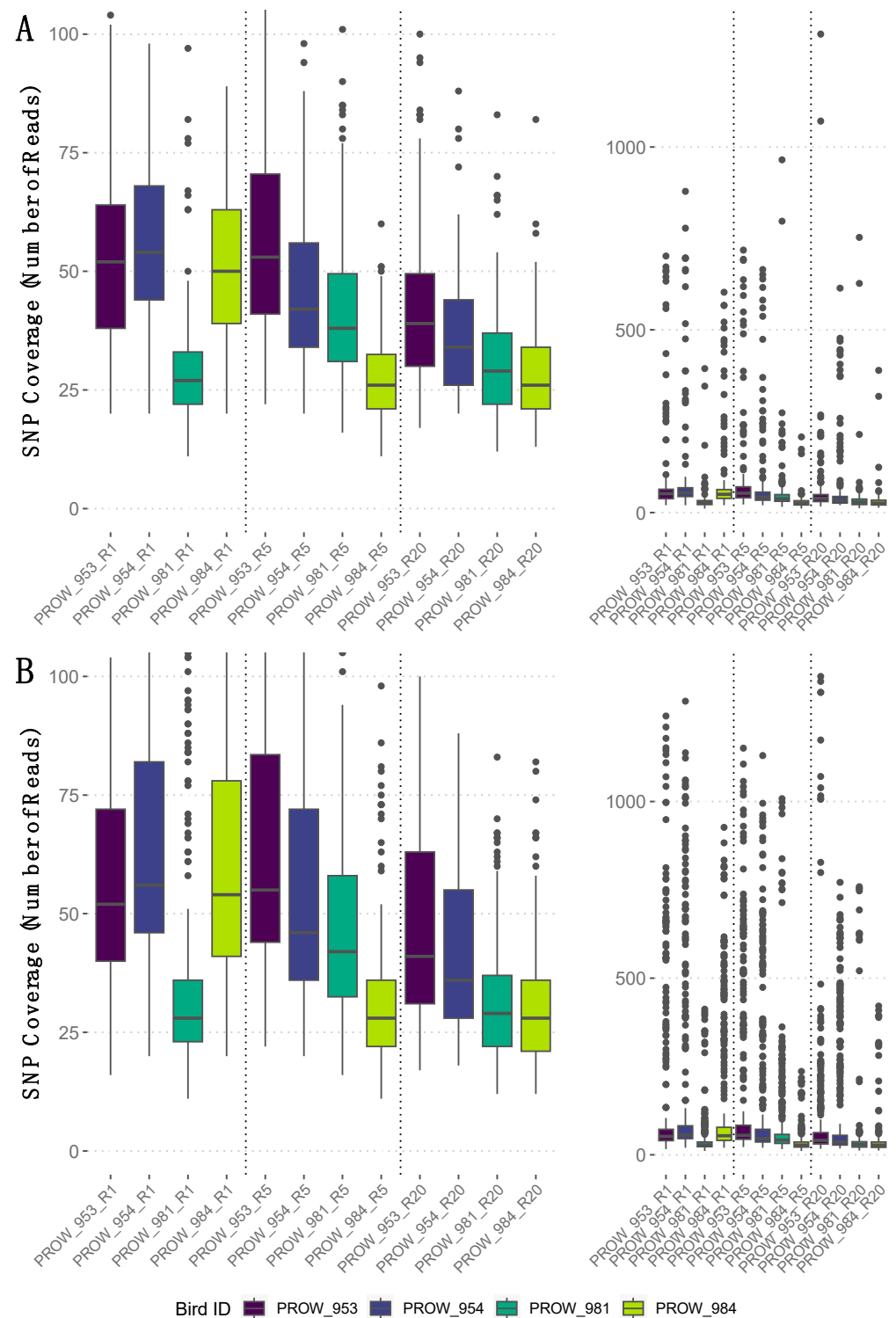


Figure 4. The number of reads that mapped to each SNP (i.e., SNP coverage) for the filtered dataset (i.e., no missing genotype data) per sample with a linkage threshold of (A) 0.01 and (B) 0.99. Rightmost panels show the full range of SNP coverage values per sample, whereas leftmost panels are enlarged and truncated versions of the rightmost panels to highlight the median and interquartile ranges of SNP coverage values per sample. Samples are ordered by the number of mites represented in each pool and divided by a dotted vertical line. Specifically, leftmost samples (i.e., those ending in “R1”) indicate one-mite samples, the middle four (“R5”) indicate five-mite samples, and the rightmost four (“R20”) indicate 20-mite samples. Darker colors represent hosts from Illinois and lighter colors represent hosts from Louisiana.

3.3. Intrapopulation Genetic Structure and Diversity

Both before (Figure 5) and after (Figure 6) removing missing genotype data, principal components analyses indicated that data were structured primarily by the individual bird from which the mites were collected (i.e., samples were primarily structured by intrapopulation). No obvious structure was seen by the number of mites in the pool or the location of collection. For the dataset in which missing data were retained, PC1 accounted for over 20% of the variation using both linkage thresholds, whereas PC2 accounted for 21.6% and 14.6% of the variation using linkage thresholds of 0.01 and 0.99, respectively (Figure 5). For the dataset in which missing genotype data were excluded, PC1 accounted for over 40% of the variation in the data and PC2 accounted for around 12% using both linkage thresholds (Figure 6).

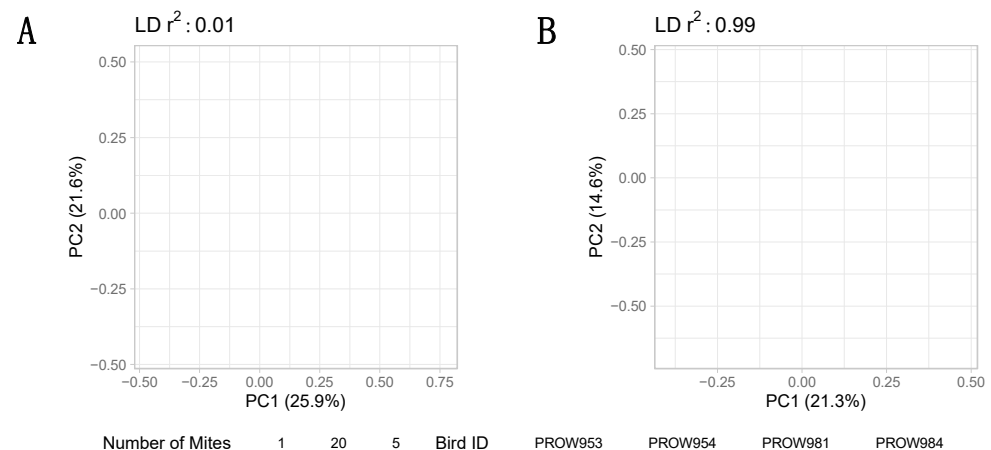


Figure 5. Principal components analysis (PCA) plots of *Amerodectes protonotaria* mites based on genome-wide SNPs prior to removing missing data with linkage thresholds of (A) 0.01 and (B) 0.99. Different pool sizes are represented by different shapes (circle = 1 mite; square = 5 mites; triangle = 20 mites) and colors represent individual hosts. Darker colors represent hosts from Illinois and lighter colors represent hosts from Louisiana.

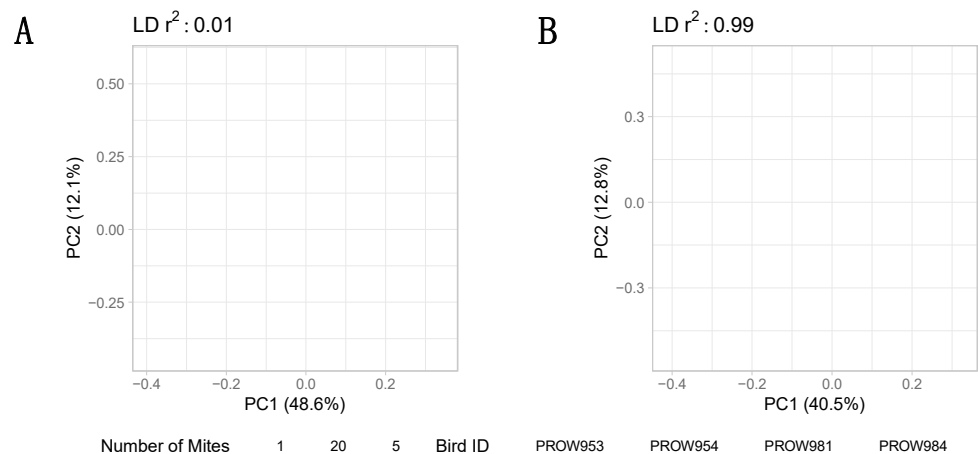


Figure 6. Principal components analysis (PCA) of *Amerodectes protonotaria* mites based on genome-wide SNPs after removing missing data with linkage thresholds of (A) 0.01 and (B) 0.99. Different pool sizes are represented by different shapes (circle = 1 mite; square = 5 mites; triangle = 20 mites) and different colors represent individual hosts. Darker colors represent hosts from Illinois and lighter colors represent hosts from Louisiana.

UpSet plots were used to illustrate how many and which SNPs were shared within and across intrapopulations. When using the more conservative linkage threshold of 0.01, the greatest number of the shared SNPs are shared at all scales for three out of the four hosts

(one mite, five mites, and twenty mites; Figure 7B–D). For the remaining host (PROW953), most of the SNPs are unique to the 20-mite sample (Figure 7A). Results using a linkage threshold of 0.99 were slightly different; more SNPs were shared within only the five- and 20-mite samples as opposed to all three scales, except for host PROW954 (Figure 8). There were only six SNPs that were common to all four samples under both linkage thresholds.

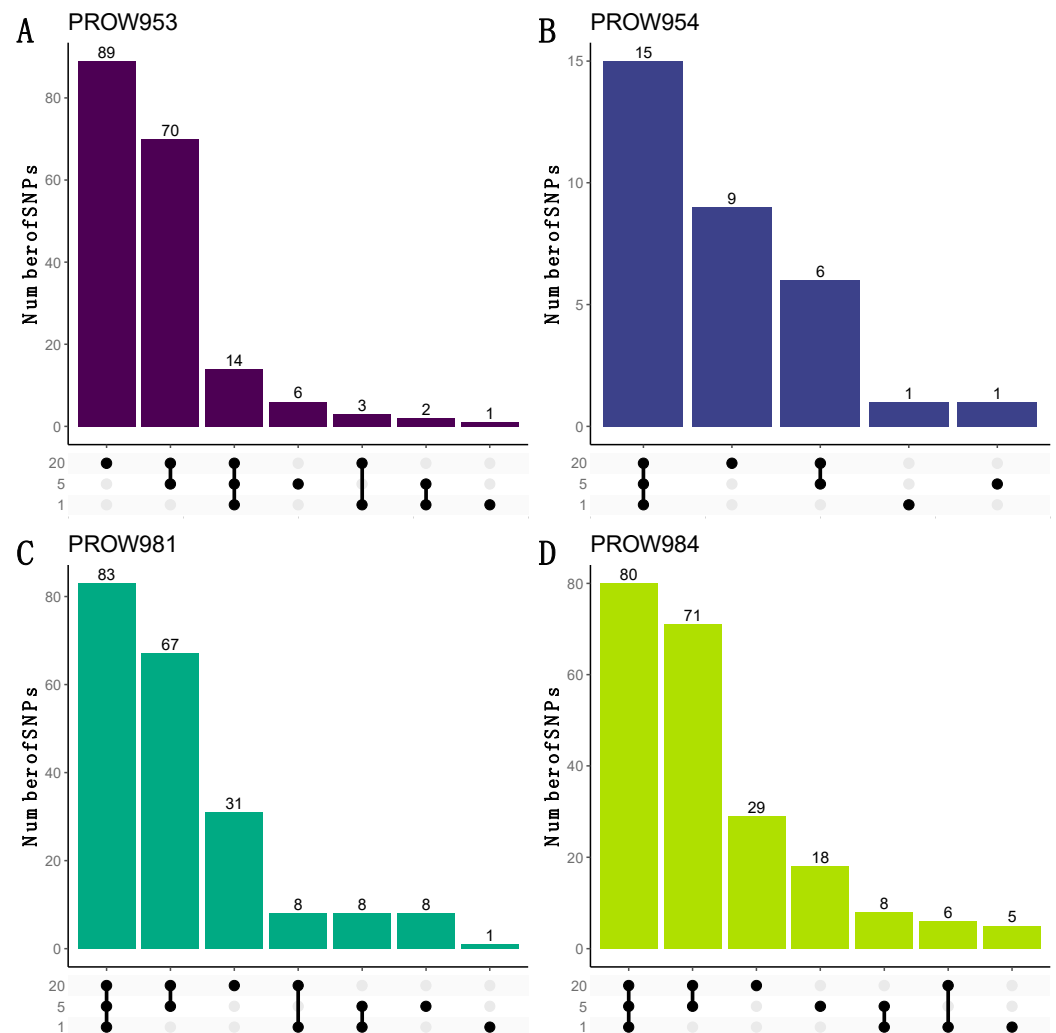


Figure 7. UpSet plots to visualize the number of shared and unique SNPs within and across sample scales from the dataset with a linkage threshold of 0.01. Rows represent different numbers of mites in each sample (20 mites, 5 mites, and 1 mite). Each column represents sets of SNPs that are either shared by samples (denoted by filled black dots connected with lines) or unique to a sample (denoted by black dots with no connections). Values on top of columns represent the size of the set. Each panel represents a different host individual, with the unique identifier as the panel title. Panels (A,B) were samples collected in Illinois and panels (C,D) were samples collected in Louisiana.

We conducted BLASTx analyses of regions flanking the six SNPs that were common to all samples under both linkage thresholds to explore the potential function of the genes in which SNPs are located. We found that five of the six SNPs had BLASTx hits on the FlyBase database, and three of the six SNP sites had E-value scores less than 1 (Table 1). For those three SNP sites with E values less than 1, we identified the associated gene(s) and GO terms (Table 2). No SNPs were located within a protein domain.

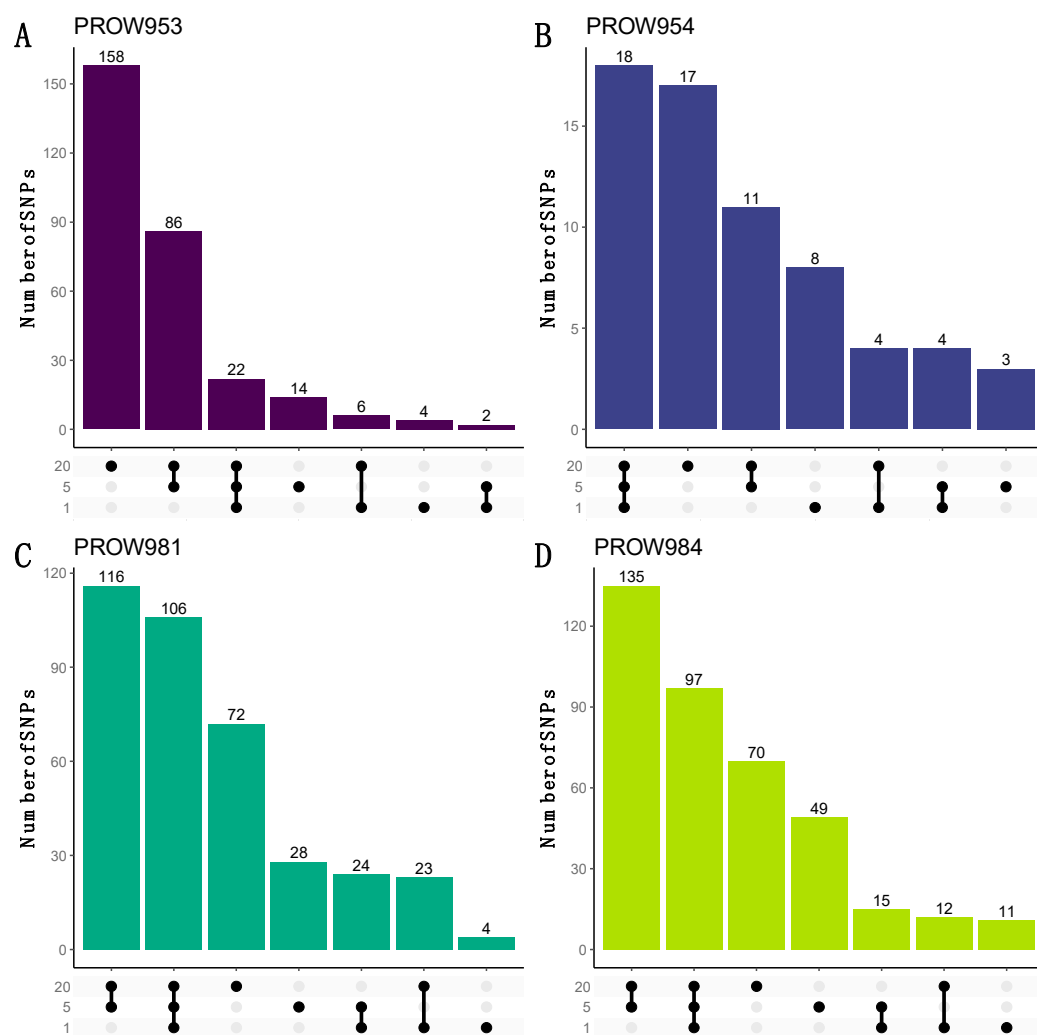


Figure 8. UpSet plots to visualize the number of shared and unique SNPs within and across sample scales from the dataset with a linkage threshold of 0.99. Rows represent different numbers of mites in each sample (20 mites, 5 mites, and 1 mite). Each column represents sets of SNPs that are either shared by samples (denoted by filled black dots connected with lines) or unique to a sample (denoted by black dots with no connections). Values on top of columns represent the size of the set. Each panel represents a different host individual, with the unique identifier as the panel title. Panels (A,B) were samples collected in Illinois and panels (C,D) were samples collected in Louisiana.

Table 1. FlyBase BLASTx results of 100 bp regions flanking the six SNPs (listed as contig:position) common to all samples under both linkage thresholds.

SNP	Polypeptide	Score	E Value	FlyBase ID (Polypeptide)	Length (aa)
866:1415	GABA-B-R2-PC	31.5722	0.173236	FBpp0308298	1224
	GABA-B-R2-PB	31.5722	0.173236	FBpp0083566	1221
	GABA-B-R2-PA	31.5722	0.173236	FBpp0083565	1220
1716:115	mamo-PC	25.7942	9.50577	FBpp0311886	799
1885:6843	CG32102-PA	30.4166	0.385929	FBpp0075710	88
2045:5306	Mp-PS	26.5646	5.57281	FBpp0308935	804
	Mp-PR	26.5646	5.57281	FBpp0301686	1039
2173:2309	NA	NA	NA	NA	NA
2301:4976	kl-5-PC	33.8834	0.0349058	FBpp0312367	4559

Table 2. Gene Ontology (GO) annotations for the three SNP (listed as contig:position) regions with E values less than 1 from the FlyBase BLASTx results.

SNP	Associated Gene	Annotation Symbol	FlyBase ID (Gene)	Molecular				Biological				Cellular							
866:1415	metabotropic GABA-B receptor subtype 2	CG6706	FBgn0027575	X	X	X	X	X	X	X	X	X	X	X	X				
1885:6843	CG32102	CG32102	FBgn0052102								X								
2301:4976	male fertility factor kl5	CG45786	FBgn0267433		X	X		X			X	X		X	X				
				receptor	small molecule binding	other molecular function	cell organization/biogenesis	transport/localization	reproduction	behavior	response to stimulus	signaling	other biological process	cytoskeleton	membrane	cell periphery	cell projection	macromolecular complex	other cellular component

4. Discussion

Microscopic symbionts present challenges in HTS studies. Nonetheless, we have demonstrated through pooled HTS that valuable information can be gleaned from sequencing individuals and pools of individuals from the same infrapopulation. Although both scenarios (individual sequencing and pooled sequencing) have advantages and disadvantages, our results demonstrate special technical and biological considerations to consider during the planning stages of HTS genomic studies of symbionts. Our data indicated that the downstream filtering and interpretation of data can also influence our understanding of coevolution at different evolutionary scales. Our results highlight both evolutionary and ecological insight into symbionts and their coevolutionary relationships with hosts that may have otherwise been overlooked by individual or pooled sequencing alone.

All samples, regardless of the number of individuals included in the pool, yielded a relatively even number of reads before and after implementing quality control measures (Table S1). This indicates that even an ultralow amount of tissue from an individual sample (e.g., a ~450 µm in length feather mite with ~0.5 ng of gDNA; A.E.M., personal observations) can be successfully sequenced in whole-genome shotgun sequencing projects using similar DNA extraction and library preparation protocols. This information can be important for future HTS studies on microscopic symbionts as well as the coevolutionary relationships they have with their hosts. However, the number of reads that mapped to the feather mite genome varied considerably within and across individual hosts (Figure 1). For three of the four single-mite samples, only half of the reads ($51.4 \pm 0.04\%$ (SE)) mapped to the feather mite genome. This suggests that researchers who intend to conduct whole-genome HTS of individual microscopic symbionts, such as feather mites, may expect to discard approximately half of the raw reads at the mapping stage for most samples. However, in this study, we did not surface-sterilize mites prior to DNA extractions [66] to remove mite ectosymbionts (e.g., bacteria). Doing so may increase the number of feather mite reads while reducing the number of non-mite reads sequenced (see below). This knowledge may impact decisions on the type of sequencing technology to use or the number of samples that can be sequenced.

It may be expected that the relative abundance of bacteria, whether endosymbiotic (e.g., found in the gut), ectosymbiotic, or environmental, will scale proportionally with the number of mites in a pool. Contrary to this expectation, the number of target (i.e., feather mite) reads increased with pool size and off-target (i.e., discarded) reads generally decreased with pool size (Figure 1). The source of the discarded reads in this study primarily

belonged to bacteria of the family Burkholderiaceae. A similar negative relationship between pool size and the relative abundance of bacteria has been observed in a multiplex PCR experiment of spider mites [67]. It is possible that the ultralow-input samples are at an inherently higher risk of contamination during wet lab procedures (e.g., DNA extraction, sonication, library preparation, sequencing; [68]). It is equally possible that bacterial symbionts vary in concentration among individual mites. In any case, for microscopic organisms, it is important to account for reads that do not map to the target genome to distinguish biological (e.g., horizontal gene transfer [69]) from non-biological (e.g., true contaminants [70]) signals. There are several well-established strategies and tools to determine target from off-target DNA for HTS data [71–74].

The SNP filtering criteria can be a critical component to identify genomic variants and interpret results [75]. Therefore, we explored datasets generated using different filters. Our data indicated that the number of SNPs differed between datasets (e.g., including or excluding missing data, contrasting linkage thresholds). Despite these differences, the overall patterns between and within samples remained nearly identical (Figures 2 and 3). This result demonstrates that the overall patterns are robust to different filtering criteria. Namely, the number of SNPs generally increases with the number of individuals in the pool, regardless of how the data are filtered. One infrapopulation, PROW954, did not deviate from this pattern, but we found much fewer SNPs within this infrapopulation (across all pool sizes) than the others. It is unclear why the genetic diversity of the mites from this infrapopulation is so uniquely low but highlights the variation of evolutionary processes occurring at the infrapopulation-level in host–symbiont systems. We also found that enough SNPs remained at a high enough coverage in the 5- and 20-mite samples after filtering (Figure 4) to conduct downstream population genetic analyses, suggesting that genetic diversity is proportional to the pool size and possibly infrapopulation size, although this was not directly tested here.

These results are consistent with previous feather mite studies, which demonstrated that mitochondrial DNA genetic diversity within infrapopulations can be significant [35] and that genetic diversity increases with infrapopulation sizes [27]. Whether or not there is an asymptote of diversity within an infrapopulation (i.e., how much genetic diversity can one more individual contribute to the infrapopulation?) requires more replication and potentially larger pool sizes. This insight can help us fill basic biological knowledge gaps of mite sexual reproduction behaviors (e.g., inbreeding [76]) and transmission (e.g., immigration [77]). Sequencing male and female mites separately (individually or pooled) may also help to answer sex-specific genomic questions of many types of avian mites, such as sex-based variation in genome size and ploidy [78–80]. These results may also have important implications for coevolutionary studies of hosts and symbionts. For example, sequencing only one individual symbiont from an infrapopulation may underestimate the genetic diversity of the symbionts on an individual host (depending on the life history of the symbiont [81]). This underestimation may then lead to a misinterpreted coevolutionary signal. However, with carefully designed sampling and a large enough sample size, these risks can be minimized [82].

Although the number of SNPs varied with pool size within infrapopulations, we found that infrapopulations were primarily structured by the individual host more so than pool size (Figure 6). This pattern, like the patterns seen with the number of SNPs, was also consistent when missing genotype data were included (Figure 5). These results indicate that the mites collected from an infrapopulation (i.e., a single feather from a single bird) are more closely related to each other than to another infrapopulation, similar to what was observed in Virrueta Herrera et al.'s work [82], in which two lice per freshwater seal host were individually sequenced. These results indicate that infrapopulation-level evolutionary processes are driving these general patterns of structure rather than technical choices like pool size. UpSet plots also suggest a similar pattern such that most SNPs identified within an infrapopulation are shared by all three pool sizes (Figure 7B–D). In other words, most SNPs are not unique to specific pool sizes, which further supports a biological signal of

intrapopulation relatedness. This also indicates that it is possible to detect SNPs of high confidence (i.e., SNPs that were identified from one-mite samples) in pools of 5 or 20 individuals. Thus, we can be highly confident in the SNPs identified from pooled samples, especially those shared with the one-mite samples. With a more relaxed linkage threshold (Figure 8), we observed the pattern shift for samples collected in Louisiana. Most SNPs were shared only between the five- and 20-mite samples (i.e., not all three pool sizes). This is not largely unexpected because pooled samples potentially contribute redundant information in clustered genomic regions compared to individual samples [83]; thus, they are more likely to have linked SNPs. Despite these differences, these patterns still suggest that most SNPs are not unique to specific pool sizes. Ideally, each sample will contain an equal number of individuals so that each pool contains an equal amount of DNA and an equal representation of individuals [18]. However, our results suggest that enough similarity exists within intrapopulations (Figure 7) and enough diversity exists between intrapopulations (Figure 6) that if an equal number of individuals cannot be obtained for every sample (for example, due to difficulty of collection, rarity of the species, or variation in intrapopulation sizes), signals of (intra)population differentiation can still be detected.

It is possible that some SNPs are located on genes that are important for mite fitness (e.g., survival and reproduction), which enables mites to quickly adapt to, persist on, and maintain a symbiotic relationship with avian hosts. We preliminarily tested if our pooled data could be used for such analyses. Two of the three FlyBase BLASTx hits with E-values less than 1 (Table 1) were associated with specific genes annotated for *Drosophila melanogaster* (Table 2). First, according to FlyBase, metabotropic GABA-B-R2 enables G protein-coupled GABA-B receptor activity and has several biological functions, such as a response to mechanical stimuli and olfactory behaviors. One mechanical pressure that feather mites experience is related to their need for permanent attachment to flight feathers; they must remain attached without being dislodged, even while the bird is in flight. Adaptive behavioral responses to changing turbulence and other dynamic conditions would certainly increase their chances of survival on hosts. Feather mites also encounter a large number of food resources on feathers [84] and may eat a diverse array of those available resources [85]; the ability to sense and adapt to novel food resources may also increase survival. Second, according to FlyBase, the male fertility factor kl5 is involved in sperm motility, which could significantly impact mite reproduction or sexual selection. Future investigations on these genomic regions could help clarify the important role these SNPs may play in mite fitness. Alternatively, exploring variants in genomic regions that are known to be involved in coevolutionary processes (e.g., [86–88]) could be valuable to our understanding of how host–symbiont interactions are maintained or change over time. However, these specific regions are not yet known for this system.

5. Conclusions

Our feather mite population genomics study highlights the advantages and disadvantages of applying HTS for individual-level sequencing versus intrapopulation-level pooled sequencing of microscopic symbionts in coevolutionary studies. Trade-offs exist with respect to individual haplotype information capture (individual sequencing) or loss (pooled sequencing) as well as an increase (pooled sequencing) or decrease (individual sequencing) in read mapping. Pooled sequencing is more likely to capture “hidden” genetic variants that exist in an intrapopulation compared with individual sequencing, but because intrapopulation diversity seems to be structured by the individual host regardless of pool size in this system, a loss of this information at the individual level may not impact biological inferences. Based on our results, a balanced strategy of sequencing multiple individuals per intrapopulation may be ideal. This approach would confidently capture haplotype information while not overlooking potentially informative intrapopulation genetic diversity and would ultimately improve our understanding of host–symbiont coevolution at multiple evolutionary scales.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life13102054/s1>, Table S1: Sample information.; Figure S1: Maximum likelihood (ML) phylogeny estimated from the cytochrome c oxidase subunit 1 (cox1) barcoding gene using a curated dataset of parulid *Amerodectes*, *Tyrannidectes*, *Proctophyllodes*, and *Trouessartia* feather mite lineages. *Amerodectes protonotaria* samples are highlighted in yellow with the 12 samples from the present work ending in “pw” at the branch tips. Values at nodes indicate ML ultrafast bootstrap (BS) support values for each mite lineage; nodes with >99 BS values are indicated by asterisks (*). The scale bar represents nucleotide substitutions per site.

Author Contributions: Conceptualization, A.E.M., T.J.B. and A.J.W.; funding acquisition, A.E.M. and T.J.B.; sample collection and coordination, K.L.P. and W.M.S.; methodology and analyses, A.E.M. and A.J.W.; writing—original draft preparation, A.E.M.; writing—review and editing, T.J.B., K.L.P., W.M.S. and A.J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the Arkansas State University College of Sciences and Mathematics provided to T.J.B. and the Students United in Preserving, Exploring, and Researching Biodiversity Program (National Science Foundation; DUE-1564954).

Institutional Review Board Statement: Samples were obtained under appropriate guidelines by the Institutional Animal Care and Use Committee (IACUC; #63-8363 and #19-039) and permitted by the United States Geological Survey Bird Banding Laboratory (#23805 and #23931), the U.S. Fish and Wildlife Service (migratory bird scientific collecting permits #MB30850D and #MB59069B-0), the Illinois Department of Natural Resources (scientific permits #NH21.5777 and #W21.6355), the Illinois Nature Preserves Commission (permit 2020), Cypress Creek National Wildlife Refuge (special use permit #2021-001R), and BREC (scientific research permit CONS-2017-04).

Data Availability Statement: The data presented in this study are available in Table S1 or are available upon request from the corresponding author. All bioinformatic pipelines and analytical scripts are available on GitHub (https://github.com/alixmatthews/protonotaria_pool_experiment) (accessed on 6 October 2023).

Acknowledgments: We thank the Guest Editors and two anonymous reviewers for constructive feedback that improved the manuscript. We thank those who helped to coordinate and/or collect samples for this project, especially Mark E. Hauber, Sara E. Lipshutz, and those in the Prothonotary Warbler Working Group; Jeff Pummill, David Chaffin, and Pawel Wolinski at the AHPCC (University of Arkansas) for their computational assistance; and Alvaro Hernandez and Chris Wright at the Roy J. Carver Biotechnology Center at the University of Illinois for their assistance in library preparation and sequencing of these ultra-low-input samples. All ethical guidelines for the use of wild birds in research were followed [89].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Larsen, B.B.; Miller, E.C.; Rhodes, M.K.; Wiens, J.J. Inordinate Fondness Multiplied and Redistributed: The Number of Species on Earth and the New Pie of Life. *Q. Rev. Biol.* **2017**, *92*, 229–265. [[CrossRef](#)]
2. Flint, H.J.; Scott, K.P.; Louis, P.; Duncan, S.H. The Role of the Gut Microbiota in Nutrition and Health. *Nat. Rev. Gastroenterol. Hepatol.* **2012**, *9*, 577–589. [[CrossRef](#)] [[PubMed](#)]
3. Hopkins, S.R.; Wojdak, J.M.; Belden, L.K. Defensive Symbionts Mediate Host–Parasite Interactions at Multiple Scales. *Trends Parasitol.* **2017**, *33*, 53–64. [[CrossRef](#)] [[PubMed](#)]
4. Li, H.; Young, S.E.; Poulsen, M.; Currie, C.R. Symbiont-Mediated Digestion of Plant Biomass in Fungus-Farming Insects. *Annu. Rev. Entomol.* **2020**, *66*, 16–17. [[CrossRef](#)] [[PubMed](#)]
5. Leung, T.; Poulin, R. Parasitism, Commensalism, and Mutualism: Exploring the Many Shades of Symbioses. *Vie Milieu* **2008**, *58*, 107–115.
6. Hatcher, M.J.; Dick, J.T.A.; Dunn, A.M. Diverse Effects of Parasites in Ecosystems: Linking Interdependent Processes. *Front. Ecol. Environ.* **2012**, *10*, 186–194. [[CrossRef](#)]
7. Okamura, B.; Hartigan, A.; Naldoni, J. Extensive Uncharted Biodiversity: The Parasite Dimension. *Integr. Comp. Biol.* **2018**, *58*, 1132–1145. [[CrossRef](#)]
8. Engel, M.S.; Ceriaco, L.M.P.; Daniel, G.M.; Dellapé, P.M.; Löbl, I.; Marinov, M.; Reis, R.E.; Young, M.T.; Dubois, A.; Agarwal, I.; et al. The Taxonomic Impediment: A Shortage of Taxonomists, Not the Lack of Technical Approaches. *Zool. J. Linn. Soc.* **2021**, *193*, 381–387. [[CrossRef](#)]

9. Löbl, I.; Klausnitzer, B.; Hartmann, M.; Krell, F.-T. The Silent Extinction of Species and Taxonomists—An Appeal to Science Policymakers and Legislators. *Diversity* **2023**, *15*, 1053. [\[CrossRef\]](#)
10. McManus, D.P.; Bowles, J. Molecular Genetic Approaches to Parasite Identification: Their Value in Diagnostic Parasitology and Systematics. *Int. J. Parasitol.* **1996**, *26*, 687–704. [\[CrossRef\]](#)
11. Poulin, R. The Rise of Ecological Parasitology: Twelve Landmark Advances That Changed Its History. *Int. J. Parasitol.* **2021**, *51*, 1073–1084. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Johnson, K.P.; Weckstein, J.D.; Virrueta Herrera, S.; Doña, J. The Interplay between Host Biogeography and Phylogeny in Structuring Diversification of the Feather Louse Genus *Penenirmus*. *Mol. Phylogenet. Evol.* **2021**, *165*, 107297. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Johnson, K.P.; Matthee, C.; Doña, J. Phylogenomics Reveals the Origin of Mammal Lice out of Afrotheria. *Nat. Ecol. Evol.* **2022**, *6*, 1205–1210. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kwong, W.K.; Engel, P.; Koch, H.; Moran, N.A. Genomics and Host Specialization of Honey Bee and Bumble Bee Gut Symbionts. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11509–11514. [\[CrossRef\]](#)
15. Richardson, E.J.; Bacigalupe, R.; Harrison, E.M.; Weinert, L.A.; Lycett, S.; Vrieling, M.; Robb, K.; Hoskisson, P.A.; Holden, M.T.G.; Feil, E.J.; et al. Gene Exchange Drives the Ecological Success of a Multi-Host Bacterial Pathogen. *Nat. Ecol. Evol.* **2018**, *2*, 1468–1478. [\[CrossRef\]](#)
16. Whibley, A.; Kelley, J.L.; Narum, S.R. The Changing Face of Genome Assemblies: Guidance on Achieving High-Quality Reference Genomes. *Mol. Ecol. Resour.* **2021**, *21*, 3. [\[CrossRef\]](#)
17. Ekblom, R.; Galindo, J. Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms. *Heredity* **2011**, *107*, 1–15. [\[CrossRef\]](#)
18. Schlötterer, C.; Tobler, R.; Kofler, R.; Nolte, V. Sequencing Pools of Individuals—Mining Genome-Wide Polymorphism Data without Big Funding. *Nat. Rev. Genet.* **2014**, *15*, 749–763. [\[CrossRef\]](#)
19. Huyse, T.; Poulin, R.; Théron, A. Speciation in Parasites: A Population Genetics Approach. *Trends Parasitol.* **2005**, *21*, 469–475. [\[CrossRef\]](#)
20. Fuentes-Pardo, A.P.; Ruzzante, D.E. Whole-Genome Sequencing Approaches for Conservation Biology: Advantages, Limitations and Practical Recommendations. *Mol. Ecol.* **2017**, *26*, 5369–5406. [\[CrossRef\]](#)
21. Zhu, Y.; Bergland, A.O.; González, J.; Petrov, D.A. Empirical Validation of Pooled Whole Genome Population Re-Sequencing in *Drosophila melanogaster*. *PLoS ONE* **2012**, *7*, e41901. [\[CrossRef\]](#)
22. Chen, C.; Parejo, M.; Momeni, J.; Langa, J.; Nielsen, R.O.; Shi, W.; Vingborg, R.; Kryger, P.; Bouga, M.; Estonba, A.; et al. Population Structure and Diversity in European Honey Bees (*Apis mellifera* L.)—An Empirical Comparison of Pool and Individual Whole-Genome Sequencing. *Genes* **2022**, *13*, 182. [\[CrossRef\]](#)
23. Gautier, M.; Foucaud, J.; Gharbi, K.; Cézard, T.; Galan, M.; Loiseau, A.; Thomson, M.; Pudlo, P.; Kerdelhué, C.; Estoup, A. Estimation of Population Allele Frequencies from Next-Generation Sequencing Data: Pool-versus Individual-Based Genotyping. *Mol. Ecol.* **2013**, *22*, 3766–3779. [\[CrossRef\]](#)
24. Proctor, H.C. Feather Mites (Acari: Astigmata): Ecology, Behavior, and Evolution. *Annu. Rev. Entomol.* **2003**, *48*, 185–209. [\[CrossRef\]](#)
25. Doña, J.; Proctor, H.; Mironov, S.; Serrano, D.; Jovani, R. Global Associations between Birds and Vane-Dwelling Feather Mites. *Ecology* **2016**, *97*, 3242. [\[CrossRef\]](#)
26. Diaz-Real, J.; Serrano, D.; Pérez-Tris, J.; Fernández-González, S.; Bermejo, A.; Calleja, J.A.; De la Puente, J.; De Palacio, D.; Martínez, J.L.; Moreno-Opo, R.; et al. Repeatability of Feather Mite Prevalence and Intensity in Passerine Birds. *PLoS ONE* **2014**, *9*, e107341. [\[CrossRef\]](#)
27. Doña, J.; Moreno-García, M.; Criscione, C.D.; Serrano, D.; Jovani, R. Species MtDNA Genetic Diversity Explained by Intrapopulation Size in a Host-Symbiont System. *Ecol. Evol.* **2015**, *5*, 5801–5809. [\[CrossRef\]](#)
28. Matthews, A.E.; Larkin, J.L.; Raybuck, D.W.; Slevin, M.C.; Stoleson, S.H.; Boves, T.J. Feather Mite Abundance Varies but Symbiotic Nature of Mite-Host Relationship Does Not Differ between Two Ecologically Dissimilar Warblers. *Ecol. Evol.* **2018**, *8*, 1227–1238. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Doña, J.; Potti, J.; De La Hera, I.; Blanco, G.; Frías, O.; Jovani, R. Vertical Transmission in Feather Mites: Insights into Its Adaptive Value. *Ecol. Entomol.* **2017**, *42*, 492–499. [\[CrossRef\]](#)
30. Peet, R.; Kirk, A.; Behnke, J.M. A Long-term Study of Temporal Variation in Wing Feather Mite (Acari: Astigmata) Infestations on Robins, *Erithacus rubecula*, in Nottinghamshire, UK. *J. Zool.* **2022**, *316*, 296–306. [\[CrossRef\]](#)
31. Matthews, A.E.; Barnett, C.J.; Boves, T.J. Differential Survival and Dispersal of Avian Feather Mites with Contrasting Host Specificities. *Ecol. Entomol.* **2022**, *47*, 864–871. [\[CrossRef\]](#)
32. Doña, J.; Sweet, A.D.; Johnson, K.P.; Serrano, D.; Mironov, S.; Jovani, R. Cophylogenetic Analyses Reveal Extensive Host-Shift Speciation in a Highly Specialized and Host-Specific Symbiont System. *Mol. Phylogenet. Evol.* **2017**, *115*, 190–196. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Matthews, A.E.; Wijeratne, A.J.; Sweet, A.D.; Hernandez, F.A.; Toews, D.P.L.; Boves, T.J. Dispersal-Limited Symbionts Exhibit Unexpectedly Wide Variation in Host Specificity. *Syst. Biol.* **2023**, *72*, 802–819. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Dabert, J.; Mironov, S.V.; Dabert, M. The Explosive Radiation, Intense Host-Shifts and Long-Term Failure to Speciate in the Evolutionary History of the Feather Mite Genus *Analges* (Acariformes: Analgidae) from European Passerines. *Zool. J. Linn. Soc.* **2022**, *195*, 673–694. [\[CrossRef\]](#)

35. Fernández-González, S.; Pérez-Rodríguez, A.; Proctor, H.C.; De La Hera, I.; Pérez-Tris, J. High Diversity and Low Genetic Structure of Feather Mites Associated with a Phenotypically Variable Bird Host. *Parasitology* **2018**, *145*, 1243–1250. [CrossRef]
36. Doña, J.; Osuna-Mascaró, C.; Johnson, K.P.; Serrano, D.; Aymí, R.; Jovani, R. Persistence of Single Species of Symbionts across Multiple Closely-Related Host Species. *Sci. Rep.* **2019**, *9*, 17442. [CrossRef]
37. Gaud, J.; Atyeo, W.T. *Feather Mites of the World (Acarina, Astigmata): The Supraspecific Taxa. Part 1 Text*; Musee Royal de L'Afrique Central: Tervuren, Belgium, 1996; Volume 277, pp. 1–193.
38. Matthews, A.E.; Klimov, P.B.; Proctor, H.C.; Dowling, A.P.G.; Diener, L.; Hager, S.B.; Larkin, J.L.; Raybuck, D.W.; Fiss, C.J.; McNeil, D.J.; et al. Cophylogenetic Assessment of New World Warblers (Parulidae) and Their Symbiotic Feather Mites (Proctophyllodidae). *J. Avian Biol.* **2018**, *49*, e01580. [CrossRef]
39. Hernandez, F.A.; Matthews, A.E.; Boves, T.J. Four New Feather Mite Species of the Genus *Amerodectes* Valim & Hernandez (Acariformes: Proctophyllodidae) from New World Warblers (Passeriformes: Parulidae) in the USA. *Syst. Appl. Acarol.* **2018**, *23*, 946–968. [CrossRef]
40. Matthews, A.E.; Boves, T.J.; Percy, K.L.; Wijeratne, A.J. Draft Genome Sequencing Data of a Feather Mite, *Amerodectes protonotaria* Hernandez 2018 (Acariformes: Proctophyllodidae). *Data Br.* **2023**, *46*, 108835. [CrossRef]
41. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.J.* **2011**, *17*, 10–12. [CrossRef]
42. Allen, J.M.; LaFrance, R.; Folk, R.A.; Johnson, K.P.; Guralnick, R.P. ATRAM 2.0: An Improved, Flexible Locus Assembler for NGS Data. *Evol. Bioinforma* **2018**, *14*, 1–4. [CrossRef] [PubMed]
43. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]
44. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [CrossRef] [PubMed]
45. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
46. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Von Haeseler, A.; Jermini, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
47. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [CrossRef]
48. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
49. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156. [CrossRef]
50. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* **2015**, *4*, 7. [CrossRef]
51. Koboldt, D.C.; Chen, K.; Wylie, T.; Larson, D.E.; McLellan, M.D.; Mardis, E.R.; Weinstock, G.M.; Wilson, R.K.; Ding, L. VarScan: Variant Detection in Massively Parallel Sequencing of Individual and Pooled Samples. *Bioinformatics* **2009**, *25*, 2283–2285. [CrossRef]
52. Kofler, R.; Pandey, R.V.; Schlötterer, C. PoPoolation2: Identifying Differentiation between Populations Using Sequencing of Pooled DNA Samples (Pool-Seq). *Bioinformatics* **2011**, *27*, 3435–3436. [CrossRef]
53. Raineri, E.; Ferretti, L.; Esteve-Codina, A.; Nevado, B.; Heath, S.; Pérez-Enciso, M. SNP Calling by Sequencing Pooled Samples. *BMC Bioinform.* **2012**, *13*, 239. [CrossRef]
54. Hivert, V.; Leblois, R.; Petit, E.J.; Gautier, M.; Vitalis, R. Measuring Genetic Differentiation from Pool-Seq Data. *Genetics* **2018**, *210*, 315–330. [CrossRef]
55. Guirao-Rico, S.; González, J. Benchmarking the Performance of Pool-Seq SNP Callers Using Simulated and Real Sequencing Data. *Mol. Ecol. Resour.* **2021**, *21*, 1216–1229. [CrossRef] [PubMed]
56. Gautier, M.; Vitalis, R.; Flori, L.; Estoup, A. F-Statistics Estimation and Admixture Graph Construction with Pool-Seq or Allele Count Data Using the R Package Poolstat. *Mol. Ecol. Resour.* **2022**, *22*, 1394–1416. [CrossRef]
57. Matthews, A.E.; Boves, T.J.; Sweet, A.D.; Ames, E.M.; Bulluck, L.P.; Johnson, E.I.; Johnson, M.; Lipshutz, S.E.; Percy, K.L.; Raybuck, D.W.; et al. Novel Insights into Symbiont Population Structure: Globe-trotting Avian Feather Mites Contradict the Specialist-Generalist Variation Hypothesis. *Mol. Ecol.* **2023**, *32*, 5260–5275. [CrossRef] [PubMed]
58. Team, R.C. R: A Language and Environment for Statistical Computing. Available online: <http://www.r-project.org> (accessed on 13 October 2022).
59. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
60. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 978-3-319-24277-4.
61. Kassambara, A. *Ggpubr: “ggplot2” Based Publication Ready Plots* 2020. Available online: <https://CRAN.R-project.org/package=gpubr> (accessed on 13 October 2022).

62. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: Fast Processing of NGS Alignment Formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef]
63. Lex, A.; Gehlenborg, N.; Strobel, H.; Vuilleumot, R.; Pfister, H. UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1983–1992. [CrossRef] [PubMed]
64. Ahlmann-Eltze, C. Ggplot2: Combination Matrix Axis for ‘Ggplot2’ to Create ‘UpSet’ Plots 2020. Available online: <https://CRAN.R-project.org/package=ggupset> (accessed on 13 October 2022).
65. Gramates, L.S.; Agapite, J.; Attrill, H.; Calvi, B.R.; Crosby, M.A.; dos Santos, G.; Goodman, J.L.; Goutte-Gattat, D.; Jenkins, V.K.; Kaufman, T.; et al. FlyBase: A Guided Tour of Highlighted Features. *Genetics* **2022**, *220*, iyac035. [CrossRef]
66. Andrews, E.S. Analyzing Arthropods for the Presence of Bacteria. *Curr. Protoc. Microbiol.* **2013**, *28*, 1E.6.1–1E.6.14. [CrossRef] [PubMed]
67. Zélé, F.; Weill, M.; Magalhães, S. Identification of Spider-Mite Species and Their Endosymbionts Using Multiplex PCR. *Exp. Appl. Acarol.* **2018**, *74*, 123–138. [CrossRef] [PubMed]
68. Arakawa, K.; Yoshida, Y.; Tomita, M. Genome Sequencing of a Single Tardigrade *Hypsibius dujardini* Individual. *Sci. Data* **2016**, *3*, 160063. [CrossRef] [PubMed]
69. Boothby, T.C.; Tenlen, J.R.; Smith, F.W.; Wang, J.R.; Patanella, K.A.; Nishimura, E.O.; Tintori, S.C.; Li, Q.; Jones, C.D.; Yandell, M.; et al. Evidence for Extensive Horizontal Gene Transfer from the Draft Genome of a Tardigrade. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15976–15981. [CrossRef] [PubMed]
70. Koutsovoulos, G.; Kumar, S.; Laetsch, D.R.; Stevens, L.; Dau, J.; Conlon, C.; Maroon, H.; Thomas, F.; Aboobaker, A.A.; Blaxter, M. No Evidence for Extensive Horizontal Gene Transfer in the Genome of the Tardigrade *Hypsibius dujardini*. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5053–5058. [CrossRef]
71. Davis, N.M.; Di Proctor, M.; Holmes, S.P.; Relman, D.A.; Callahan, B.J. Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data. *Microbiome* **2018**, *6*, 226. [CrossRef]
72. Schmieder, R.; Edwards, R. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* **2011**, *6*, 0017288. [CrossRef]
73. Challis, R.; Richards, E.; Rajan, J.; Cochrane, G.; Blaxter, M. BlobToolKit—Interactive Quality Assessment of Genome Assemblies. *G3 Genes Genomes Genet.* **2020**, *10*, 1361–1374. [CrossRef]
74. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments. *Genome Biol.* **2014**, *15*, R46. [CrossRef]
75. Bergeron, L.A.; Besenbacher, S.; Turner, T.; Versoza, C.J.; Wang, R.J.; Price, A.L.; Armstrong, E.; Riera, M.; Carlson, J.; Chen, H.; et al. The Mutationathon Highlights the Importance of Reaching Standardization in Estimates of Pedigree-Based Germline Mutation Rates. *eLife* **2022**, *11*, e73577. [CrossRef]
76. Bailly, X.; Migeon, A.; Navajas, M. Analysis of Microsatellite Variation in the Spider Mite Pest *Tetranychus turkestani* (Acari: Tetranychidae) Reveals Population Genetic Structure and Raises Questions about Related Ecological Factors. *Biol. J. Linn. Soc.* **2004**, *82*, 69–78. [CrossRef]
77. Criscione, C.D.; Poulin, R.; Blouin, M.S. Molecular Ecology of Parasites: Elucidating Ecological and Microevolutionary Processes. *Mol. Ecol.* **2005**, *14*, 2247–2257. [CrossRef]
78. Fain, A. A Review of the Family Epidermoptidae Trouessart Parasitic on the Skin of Birds (Acarina: Sarcoptiformes). *Konink Acad. Wetensch. Let. Schone Kunst.* **1965**, *27*, 1–144.
79. Hernandez, F.A.; O'Connor, B.M. *Cystoidosoma hermaphroditus* Sp. n., the First Representative of the Quill Mite Family Ascouracaridae (Acari: Astigmata: Pterolichoidea) from an Owl (Aves: Strigiformes). *Folia Parasitol.* **2015**, *62*, 37. [CrossRef] [PubMed]
80. Gregory, T.R.; Young, M.R. Small Genomes in Most Mites (but Not Ticks). *Int. J. Acarol.* **2020**, *46*, 1–8. [CrossRef]
81. Demastes, J.W.; Hafner, D.J.; Hafner, M.S.; Light, J.E.; Spradling, T.A. Loss of Genetic Diversity, Recovery and Allele Surfing in a Colonizing Parasite, *Geomydoecus aurei*. *Mol. Ecol.* **2019**, *28*, 703–720. [CrossRef] [PubMed]
82. Virrueta Herrera, S.; Johnson, K.P.; Sweet, A.D.; Ylinen, E.; Kunasranta, M.; Nyman, T. High Levels of Inbreeding with Spatial and Host-associated Structure in Lice of an Endangered Freshwater Seal. *Mol. Ecol.* **2022**, *31*, 4593–4606. [CrossRef]
83. Mimee, B.; Duceppe, M.-O.; Véronneau, P.-Y.; Lafond-Lapalme, J.; Jean, M.; Belzile, F.; Bélair, G. A New Method for Studying Population Genetics of Cyst Nematodes Based on Pool-Seq and Genomewide Allele Frequency Analysis. *Mol. Ecol. Resour.* **2015**, *15*, 1356–1365. [CrossRef]
84. del Labrador, M.M.; Doña, J.; Serrano, D.; Jovani, R. Quantitative Interspecific Approach to the Stylosphere: Patterns of Bacteria and Fungi Abundance on Passerine Bird Feathers. *Microb. Ecol.* **2021**, *81*, 1088–1097. [CrossRef]
85. Doña, J.; Proctor, H.; Serrano, D.; Johnson, K.P.; van Oploo, A.O.; Hugueta-Tapia, J.C.; Ascunce, M.S.; Jovani, R. Feather Mites Play a Role in Cleaning Host Feathers: New Insights from DNA Metabarcoding and Microscopy. *Mol. Ecol.* **2019**, *28*, 203–218. [CrossRef]
86. Papkou, A.; Guzella, T.; Yang, W.; Koepfer, S.; Pees, B.; Schalkowski, R.; Barg, M.C.; Rosenstiel, P.C.; Teotónio, H.; Schulenburg, H. The Genomic Basis of Red Queen Dynamics during Rapid Reciprocal Host–Pathogen Coevolution. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 923–928. [CrossRef] [PubMed]
87. Lee, C.-C.; Lin, C.-Y.; Tseng, S.-P.; Matsuura, K.; Yang, C.-C.S. Ongoing Coevolution of *Wolbachia* and a Widespread Invasive Ant, *Anoplolepis gracilipes*. *Microorganisms* **2020**, *8*, 1569. [CrossRef] [PubMed]

88. Sugihara, Y.; Abe, Y.; Takagi, H.; Abe, A.; Shimizu, M.; Ito, K.; Kanzaki, E.; Oikawa, K.; Kourelis, J.; Langner, T.; et al. Disentangling the Complex Gene Interaction Networks between Rice and the Blast Fungus Identifies a New Pathogen Effector. *PLoS Biol.* **2023**, *21*, e3001945. [[CrossRef](#)] [[PubMed](#)]
89. Fair, J.M.; Paul, E.; Jones, J.; Clark, A.B.; Davie, C.; Kaiser, G. *Guidelines to the Use of Wild Birds in Research*; The Ornithological Council: Washington, DC, USA, 2010.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.