HHS Public Access

Author manuscript

Mol Ecol. Author manuscript; available in PMC 2020 October 09.

Published in final edited form as:

Mol Ecol. 2019 October ; 28(19): 4513–4528. doi:10.1111/mec.15231.

Host And Geography Together Drive Early Adaptive Radiation Of Hawaiian Planthoppers

Kari Roesch Goodman[1], **Stefan Prost**[2,3], **Ke Bi**[4,5,6], **Michael S. Brewer**[7], **Rosemary G. Gillespie**[1]

[1]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, 94720, USA

[2]Department of Integrative Biology, University of California, Berkeley, CA, 94720, USA

[3]Senckenberg Research Institute · LOEWE-Centre for Translational Biodiversity Genomics, Senckenberganlage 25 D-60325 Frankfurt/Main, Germany

[4]Computational Genomics Resource Laboratory (CGRL), California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA

[5]Ancestry, 153 Townsend St., Ste. 800 San Francisco, CA 94107

[6]Museum of Vertebrate Zoology, University of California, Berkeley, 94707

[7]Department of Biology, East Carolina University, Greenville, NC, USA

Abstract

The interactions between insects and their plant host have been implicated in driving diversification of both players. Early arguments highlighted the role of ecological opportunity, with the idea that insects “escape and radiate” on new hosts, with subsequent hypotheses focusing on the interplay between host shifting and host tracking, coupled with isolation and fusion, in generating diversity. Because it is rarely possible to capture the initial stages of diversification, it is particularly difficult to ascertain the relative roles of geographic isolation versus host shifts in initiating the process. The current study examines genetic diversity between populations and hosts within a single species of endemic Hawaiian planthopper, *Nesosydne umbratica* (Hemiptera, Delphacidae). Given that the species was known as a host generalist occupying unrelated hosts, *Clermontia* (Campanulaceae) and *Pipturus* (Urticaceae), we set out to determine the relative importance of geography and host in structuring populations in the early stages of differentiation on the youngest islands of the Hawaiian chain. Results from extensive exon capture data showed that *N. umbratica* is highly structured, both by geography, with discrete populations on each

Name, address, fax number and email of corresponding author: Rosemary G. Gillespie, Dept. Environmental Science, Policy, & Management, University of California, Berkeley, CA, 94720; gillespie@berkeley.edu.

Author Contributions

KRG conceived and designed the experiments, performed the experiments, gathered and analyzed the data, wrote the paper, prepared figures and/or tables. **SP** analyzed the data, reviewed drafts of the paper, and helped prepare figures. **KB** analyzed the data, reviewed drafts of the paper. **MSB** analyzed the data, reviewed drafts of the paper. **RGG** helped conceive the work, write the paper and prepare figures and tables.

Data Accessibility

The raw Illumina reads from the exon capture experiment, raw Illumina reads from the transcriptome and the assembled transcriptome are archived at NCBI under BioProject number PRJNA341388.

volcano, and by host plant, with parallel radiations on *Clermontia* and *Pipturus* leading to extensive co-occurrence. The marked genetic structure suggests that populations can readily become established on novel hosts provided opportunity; subsequent adaptation allows monopolization of the new host. The results support the role of geographic isolation in structuring populations and with host shifts occurring as discrete events that facilitate subsequent parallel geographic range expansion.

Keywords

Exon capture; population genomics; island biogeography; planthopper; *Nesosydne*

Introduction

Associations between insects and their host plants have long been considered to play a key role in the tremendous diversity of phytophagous insects (Nakadai & Kawakita 2016; Wiens *et al.* 2015), but the actual causes remain elusive. Key hypotheses as to how diversity might arise within phytophagous insects have focused on the role of cycles of diet breadth expansion and contraction (Janz & Nylin 2008), with higher rates of diversification in taxa that show greater lability in host use (Hardy & Otto 2014). In particular, although the prevailing hypothesis is that there is broad-scale conservatism in insect host use (Winkler & Mitter 2008), there is an ever increasing number of examples of relatively rapid host shifts (Janz 2011), in particular given the availability of novel hosts. However, perhaps the greatest unknown is the relative importance of geographic isolation versus shifts or expansion of host use for achieving initial divergence (Doellman *et al.* 2018). In studies that have examined the interplay between these factors (Forbes *et al.* 2017), considerable work has highlighted the importance of ecological shifts prior to isolation (e.g., Matsubayashi *et al.* 2010); yet other studies show the reverse (e.g., Goodman *et al.* 2012). The current study uses a system of diversifying phytophagous insects in Hawaii to examine the relative importance of geography versus host association in isolating populations.

The relative importance of geographic isolation and host shifts in driving diversification can perhaps most readily be assessed by examining the early stages of population differentiation and speciation. It has already been shown that colonization of new habitats is clearly tied to host shifting followed by specialization in butterflies (Janz 2011), and most speciation events associated with shifts in host use appear to follow periods of expanded host ranges (Janz *et al.* 2001; Vamosi *et al.* 2014). So, periods of multiple host use may precede specialization on a new host (Janz & Nylin 2008; Weingartner *et al.* 2006). Although novel mutations may provide the needed fitness advantage in such events, another explanation is that when an insect colonizes a novel environment, provided the colonist can survive through ecological fitting (Janzen 1985), establishment will be facilitated by existing plasticity (West-Eberhard 2003) and similarity of a new host to the old one or because “sloppy fitness space” may allow exploitation of a more novel host (Agosta & Klemens 2008). Whether that host acquisition results in wider diet breadth or a host shift, potentially with speciation, may depend on population structure and gene flow among ancestral and novel associated populations, and the selective pressure associated with the novel host.

Given limited gene flow, alleles dictating performance on the ancestral host could be lost through drift in small, isolated populations (Gompert *et al.* 2015). Because of the association here between geographic range and host evolution, it is important to understand how the two may operate together. However, both geographic range and ancestral host shifts can be difficult to reconstruct phylogenetically when host associations are labile (Losos & Glor 2003; Stireman 2005). As a result, both pattern and processes of geographic-and host-range evolution, and associated factors involved in driving diversification, remain obscure (Nosil & Mooers 2005; Nyman 2010; Stireman 2005).

Hotspot islands have provided ideal systems for looking at associations between insects and their hosts (Jordal *et al.* 2006), providing insights into early stages of diversification (Goodman 2010) and how potential co-occurrence of close relatives might be achieved through parallel episodes of speciation (Hernández-Teixidor et al. 2016). Here, we make use of the Hawaiian archipelago as a system to identify how small-scale eco-evolutionary patterns may play out into large-scale evolutionary processes. The Hawaiian archipelago is particularly suitable for such studies because it provides a precisely dated time progression from older substrates (islands and volcanoes) in the northwest to younger substrates in the southeast (Lim & Marshall 2017). On a geologic timescale, the current high islands are young, with the oldest being Kauai at ca. 5 million years, and with areas of the youngest, Hawaii Island, still growing today. This, together with the fact that, after initial colonization of a lineage, by far the majority of colonization events occur within the archipelago (rather than from mainland sources), allows us to observe evolutionary processes at different stages and ages, and in particular can provide insights into the early stages of differentiation (Shaw & Gillespie 2016).

The genus *Nesosydne* (Delphacidae) is known only from islands in the eastern Pacific Ocean, with an adaptive radiation of more than 80 species endemic to the Hawaiian Islands (Asche 1997; Fennah 1958; Zimmerman 1948). The origin of the lineage outside Hawaii remains obscure (Asche 1997). Whereas delphacid planthoppers in the rest of the world are primarily specialized on monocots, the arrival of the *Nesosydne* lineage in the Hawaiian Islands coincides with an expansion of host range onto a wide variety of dicotyledonous host plants in 20 different families (Asche 1997; Drew & Roderick 2005; Fennah 1958; Hasty 2005; Roderick 1997; Roderick & Metz 1997; Wilson *et al.* 1994; Zimmerman 1948). Within the Hawaiian Islands, most species of *Nesosydne* are monophagous with very limited population sizes (Asche 1997). However, one species, *Nesosydne umbratica,* appears to be polyphagous as it has been documented on several very different host plants. Moreover, there is a spectrum of wing forms from brachypterous to fully macropterous forms (Zimmerman 1948), with some species dimorphic, having both brachypterous and macropterous individuals (Denno & Perfect 1994); such dimorphsm, which is found *N. umbratica,* has been shown to lead to increased reproductive success in brachypterous forms relative to their winged counterparts (Langellotto *et al.* 2000) and might be advantageous in the dynamic Hawaiian Island landscape.

In order to study the early stages of diversification and the interplay between geographic isolation and host shifting, we focus on *Nesosydne umbratica*. This species is limited to the wet, mid-elevation forests of the two youngest islands of the Hawaiian chain, Maui and

Hawaii. The apparent polyphagy of the species coupled with the very young geological substrates on which it is found (Figure 1), provide an extraordinary opportunity to study the interplay between geographic isolation and host transitions in the origins of specialization. We found the taxon in abundance primarily on *Clermontia spp*. (Campanulaceae) and *Pipturus spp*. (Urticaceae) on both islands. To understand the demographic history within the taxon, we couple population-level sampling across both islands and host plants with transcriptome-based exon capture to generate a SNP data set. This approach allows us to investigate SNPs within exonic and neutral flanking regions. The specific questions we ask relate to the relative roles of geographic isolation versus host shifts and conservatism in driving diversification. Given that populations of this species are known to be broadly distributed across the islands, with several known hosts, we first ask whether the species uses both hosts interchangeably within populations or if we can detect the beginnings of specialization among the hosts. We then ask whether island (or volcano) colonization is associated with (i) expanded host range; or (ii) host shifting.

Materials & Methods

Ecology of *Nesosydne umbratica*:

While little scientific work has been performed on *Nesosydne umbratica,* collection records indicate that it occurs on seven plant genera in four families *Charpentiera obovata* (Campanualaceae), *Clermontia clermontioides* (Campanualaceae), *Cyanea bamatiflora* (Campanualaceae), *Cyrtandra* (Gesneriaceae), *Pipturus* (Urticaceae), *Stenogyne* (Lamiaceae) and *Urera glabra* (Urticaceae) (Denno & Perfect 1994; Giffard 1922; Kirkaldy 1902, 1910; Zimmerman 1948). In our fieldwork, we have found the taxon in abundance primarily on *Clermontia spp*. (Campanulaceae) and *Pipturus spp*. (Urticaceae) on both Hawaii and Maui, and focus our collecting efforts on these relatively common plants. However, when we encountered the other, less common, plants from which *N. umbratica* has been recorded in the past (*Charpentiera, Cyrtandra, Stenogyne,* and *Urera*), we searched them as well, but never found planthoppers.

Specimen Collection:

Specimens of *N. umbratica* were collected unsystematically and placed directly into 95% ethanol from as many plants as possible within stands of *Clermontia spp.* (Campanulaceae), *Cyanea spp.* (Campanulaceae), *Dubautia sp.* (Asteraceae) and *Pipturus spp.* (Urticaceae) from wet forests throughout East Maui and Hawaii Island between 2013 and 2015 (Figure 1). The planthoppers collected on *Cyanea* were from two stands that were outplanted in 2009, one each on Maui and Hawaii (PEP 2009). Where it was possible to do so given plant presence and access at each site, at least 15 plants were sampled at each location (Table 1); collections of *N. umbratica* on West Maui were made only from *Clermontia* populations. Insect specimens for analysis were subsampled from within these collections, with an attempt made to provide multiple individuals per host plant per site. Clearly associated groups of adults and nymphs were collected from individual host plants and adult males were identified using the key in Zimmerman (1948). 184 individuals were used in the final analysis, and specimens were deposited in the Essig Museum of Entomology at U.C. Berkeley.

Transcriptome Sequencing and Probe Design:

To obtain reference transcriptomes for exon capture probe design, we first extracted and then pooled RNA from 5 frozen specimens of *N. umbratica*. The RNA was isolated using a trizol extraction protocol. Library preparation and sequencing were performed by HudsonAlpha (Huntsville, AL. USA). A single RNAseq library was created using the Illumina Truseq RNA v2 kit and sequenced on a HiSeq 2500 using a fraction of single lane designed to yield approximately 50 million paired-end reads (25 million forward and 25 million reverse) at 50 bases in length. Downstream transcriptome quality control and assembly followed a pipeline provided on the Computational Genomics Resource Laboratory (CGRL)-QB3-UCBerkeley Github site (https://github.com/CGRL-QB3-UCBerkeley/DenovoTranscriptome). In short, raw reads were quality trimmed and adapter contamination was removed using TRIMMOMATIC (Bolger *et al.* 2014) and CUTADAPT (Martin 2011). Pre-and post-trimming read quality was assessed using FASTQC (Andrews 2010). The reads that mapped (BOWTIE2; default parameters; (Langmead *et al.* 2009)) to a custom bacterial database were removed as potential contamination. Overlapping forward and reverse paired-end reads were merged into a single longer read using FLASH (Magoč & Salzberg 2011). The resulting cleaned forward, reverse, and merged reads were assembled into a reference transcriptome using TRINITY (Haas *et al.* 2013) with minimum contig length of 100 bases and minimum kmer coverage of two. Transcriptome sequencing and assembly resulted in 73,510 transcripts comprising 36,337,547 bp (median contig length 354 bp; mean contig length 494.32 bp).

The resulting files were searched (BLASTX; e-value 1e-03) against the NCBI non-redundant (nr) protein database for additional, non-read-based contaminant removal (MEGAN4; (Huson *et al.* 2007)) and functional annotation (BLAST2GO; (Conesa *et al.* 2005)). Due to the lack of planthopper genomic resources at the time of the work, all sequences assigned by MEGAN4 to the Neoptera and daughter taxa were retained as putative *Nesosydne* sequences; all sequences assigned to other taxa, including bacteria, fungi, viruses, and non-neopteran animals, were of potentially exogenous origins and were removed. In an attempt to reduce including multiple paralogs in probe design, TRINITY components with multiple members were removed. Open reading frames (ORFs) were predicted using the TRANSDECODER script bundled with TRINITY. A custom python script was used to remove sequences with aberrant GC percentages (i.e., >70 % or <30%). These values were used to remove contigs likely resulting from contamination. Analyses of the published brown planthopper (*Nilaparvata lugens*) transcripts (Xue *et al.* 2014; GCF_000757685.1) show an average GC content of 42.69% (1% quantile: 31% and 99% quantile: 62.6%; Figure S1). Only Trinity assembly “components” with a single transcript were included in probe design as a means of removing multi-paralog gene families.

The final assembly after the filtering steps was used to design probes for exon capture (SeqCap EZ, Roche Nimblegen, which included a total of 5 reactions). Mitochondrial genes were identified by BLASTing known mitochondrial gene regions (COI, COII, COIII, AFP6, ND1, ND4, ND5, ND6, CYTB) against the filtered transcriptomic data and were also included as probes. The target contained 3,982 transcripts and 2,166,784 bp of sequence

(median length 462 bp; mean contig length 543.40 bp). This approach allows us to not just capture SNPs in exonic but also in neutral regions, flanking the exons.

DNA extraction and library preparation:

DNA was extracted from individual full bodies using the Qiagen DNeasy® Blood and Tissue kit. As a first step during DNA extraction, adults were incubated overnight at 56°C in proteinase K. Their exoskeletons were then removed from the solution and preserved for vouchering. Nymphs are tiny so were destructively sampled in order to yield enough DNA. They were frozen at −80°C for 5 minutes, then pulverized using a sterile pipette tip prior to extraction. DNA quality was assessed using agarose gel electrophoresis in TBE buffer, and quantity was measured using a Qubit® 2.0 Fluorometer (Life Technologies).

184 genomic libraries were prepared following the protocol in Meyer and Kircher (2010). Starting material ranged from 300–500ng per individual, which was fragmented on a Bioruptor® UCD-200 (Diagenode) until it yielded fragments distributed between 100bp and 500bp. The samples were dual-indexed with 8 PCR cycles, using indexing oligos P5 #1–5 (Kircher *et al.* 2011) and P7 #1–50 (Meyer & Kircher 2010) so that each individual had a unique combination of two indexes. Completed libraries were then quantified using a NanoDrop® (Thermo Scientific). Individuals indexed with the same P5 oligo were pooled together in equal amounts (by mass) into a single multiplexed group of 4μg (measured on a NanoDrop®), yielding 5 multiplexed samples.

Exon capture, hybridization & sequencing:

Five exon capture hybridization experiments were performed corresponding to the 5 multiplexed DNA pools. These were accomplished using a Nimblegen SeqCap EZ Library® custom design (probe selection described above) following the manufacturer's protocol, with two exceptions. First, in the step hybridizing the multiplexed samples and the custom SeqCap EZ Library®, custom xGEN P5 and P7 7-bp indexes (Integrated DNA Technologies) were used instead of the SeqCap oligos and, as suggested by the Nimblegen protocol, SeqCap EZ Developer Reagent was used in place of COT Human DNA. Finally, in the step purifying the amplified captured multiplex DNA sample, lab-prepared SeraMag beads (Rohland & Reich 2012) were used in place of Agencourt AMPure XP Beads.

The success of the capture experiment was verified by measuring shifts in the enrichment curve for two positive and one negative assay using qPCR (Table S1). Here we report the crossing threshold (Ct) value for the enrichment curve, which indicates the number of PCR cycles at which the amount of sample becomes detectable above the baseline. Positive assays amplify genomic regions that are targeted by the capture and are expected to shift toward a lower Ct value after capture due to having more copies of the targeted region per ng of DNA, while negative assays attempt to amplify regions excluded by the capture probes and are expected to shift towards a higher Ct value. Primers for the negative assay were designed from a sequence from the transcriptome-designed random probes that was not included in the probe set. Thus, it is known to be present in the organism, but was not targeted for enrichment. Primers for positive assays were designed from transcriptome-designed random probes that were included and so were targeted for enrichment. The post-capture library

pools were sequenced on an Illumina HighSeq 4000 platform on one lane with 150 paired-end reads at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Bioinformatics:

Raw sequencing data processing.—Raw sequence data were cleaned by following the general protocol outlined by Singhal (2013) and Bi et al. (2012) with some modifications. Raw FASTQ reads were filtered using SKEWER (Jiang *et al.* 2014) and TRIMMOMATIC (Bolger *et al.* 2014) to trim adapter contaminations and low quality reads. Exact PCR and optical duplicates were removed using SUPER-DEDUPER (https://github.com/dstreett/Super-Deduper). We used BOWTIE2 (Langmead & Salzberg 2012) to align the resulting reads against the *Escherichia coli* genome to remove potential bacteria contamination that was present in the data. Overlapping paired reads were merged using FLASH (Magoč & Salzberg 2011). For the resulting cleaned reads (paired, merged and orphan reads), we selected 12 representative libraries that had the most amount of data for assembly. We assembled each of the 12 individuals using ABySS (Simpson *et al.* 2009). We first assembled the data using various k-mer sizes (31, 41, 51, 61 and 71) and then used blat (ref. https://www.ncbi.nlm.nih.gov/pubmed/11932250), CD-HIT-EST (Li & Godzik 2006) and CAP3 (Huang & Madan 1999) to cluster and merge all raw assemblies into reduced, less-redundant assemblies. We used BLASTN (e-value cutoff = 1e-20, similarity cutoff = 85%) to compare these assembles against the 3,982 target loci to extract the set of contigs that were associated with targets. For contigs that were derived from the same targeted locus, we joined them together with Ns based on their relative BLAST hit positions to that locus. Most of the final in-target assemblies contain multiple contigs, and each may include complete or partial coding and flanking sequences. We used these in-target assemblies as a pseudo-reference genome. Cleaned sequence data from each individual library were then aligned to this reference using NOVOALIGN (http://www.novocraft.com/products/novoalign/) and we only kept reads that mapped uniquely to the reference. We used PICARD (http://broadinstitute.github.io/picard/) to add read groups and GATK (McKenna *et al.* 2010) to perform re-alignment around indels. We then used SAMTOOLS/BCFTOOLS(Li *et al.* 2009) to generate a raw VCF (variant call format) that contains all potential variable and invariable sites. The sites in the VCF file were then filtered using a custom filtering program, SNPcleaner (https://github.com/fgvieira/ngsClean) by following the protocol specified in Bi et al. (2013) to filter sites based on missing data (at least 70% of the samples having at least 3× coverage), strand bias, map quality bias, base quality bias, end distance bias and HWE. Individuals with an average coverage less than 5× were dropped from the analysis. After these filters, 1,646,142 sites from 1,872 markers were used in downstream analyses.

Data analysis:

Low coverage data introduces uncertainty into analyses and as such, sequencing and mapping errors can have a proportionally larger effect if not accounted for carefully in the analysis (Crawford & Lazzaro 2012). Because our data set contains primarily low to medium coverage data (between 5–15×), we used ANGSD (Korneliussen *et al.* 2013) to compute genotype likelihoods upon which most of the downstream population genetic analyses were based. We first investigated the shape of the site frequency spectrum (SFS), calculating the folded SFS using realSFS (part of the ANGSD package). In the absence of

strong demographic changes or selection, we would expect the SFS to follow an exponential decrease (see e.g. Li *et al.* 2012; Pavlidis & Alachiotis 2017) and indeed that was what we found. Furthermore, the shape of the SFS can indicate whether the amount of SNP data will be sufficient and whether the data shows any signatures of selection or demographic effects (e.g. based on the amount of SNPs with medium allele frequencies).

In some populations it was difficult to obtain large numbers of specimens; however, we chose to include all populations and analyze them using a variety of approaches in order to study the range of genetic variation across the sampled landscape. Phylogenetic methods are not sensitive to sample size issues, and comparison among results from our phylogenetic and population-based analyses indicated similar patterns.

Population Variation.—To estimate genetic variation within and among populations, we calculated Fst, nucleotide diversity (Pi), Tajima's D and Watterson Theta. To calculate Fst we first estimated allele frequency likelihoods using ANGSD (-doSaf 1). We then calculated the pairwise 2dSFS between all population pairs, separated by host plant, using realSFS (part of the ANGSD package). Pairwise Fst values were then calculated based on the indexed 2dSFS as a prior. We performed multidimensional scaling (to 2 dimensions) with R (using the cmdscale function) to visualize Fst distances. We used the empirical Bayes (EB) method implemented in ANGSD to calculate nucleotide diversity (pi), population mutation rate (Watterson's Theta), and test for neutrality (Tajima's D). We first estimated site allele frequency likelihood using ANGSD and then computed maximum likelihood estimate of the SFS (folded) using realSFS for each population. The folded SFS for each population was then used a prior to obtain log-scaled per-site estimates of the thetas. We computed per population statistics (i.e. Tajima's D) using thetaStat. We used R to plot the genome-wide distributions of Tajima's D and Watterson Theta, and extracted average per-site estimate of Watterson Theta and Pi values using an in-house perl script.

Population Structure.—We investigated our first approximation of population structure using PCA. We calculated genotype likelihoods using the samtools model implemented in ANGSD, using the function-doGeno 32 to output genotype posterior probabilities assuming a uniform prior (-doPost 2). The output was then used to compute a genotypic covariance matrix using ngsCovar (part of the ngsTools package, Fumagalli *et al.* 2013). To minimize the effect of very rare alleles that may be due to sequencing error, we set-minmaf to 0.002. We plotted comparisons between the first three PCAs using R.

Admixture Analysis.—We used NGSADMIX (Skotte *et al.* 2013) to investigate signatures of admixture in our data. To avoid problems caused by having distances among populations that are variable, where very large distances among populations might overshadow more shallow ones, we split the data into closely related groupings based on the results of the phylogenetic analyses below. We analyzed *Pipturus*-associated individuals together, and all remaining specimens (mostly from *Clermontia*) except for the West Maui individuals together. We output the genotype likelihoods using the beagle format using-doGlf 2 and only used SNPs with a p-value lower than 1e-6. We then fit our data to different numbers of clusters (K = 2–7 for the *Pipturus*-associated group and K = 2–10 for the *Clermontia*-associated group) using NGSADMIX (using-minmaf 0.002), running ten

repetitions at each value of K and calculating log likelihoods for each run. Admixture proportions and log likelihoods were visualized using R.

Phylogenetic Analyses Using the Exome-capture Data.—To investigate phylogenetic relationships in our sampling, we calculated an unrooted network and also constructed a tree based on genetic distances. We used SPLITSTREE (Huson & Bryant 2006) to calculate the unrooted network. Genotype calls were performed using ANGSD using SNP pval of 1e-06 and a minimum depth of 3. A genetic distance matrix was then calculated based on the genotype calls using R package ADEGENET (Jombart 2008). We then used a custom perl script to convert the file to the SPLITSTREE input format, and SPLITSTREE to calculate and visualize the phylogenetic network. We then used NGSDIST (Vieira *et al.* 2015) to calculate a distance matrix based on genotype posteriors and FASTME (Lefort *et al.* 2015) to convert the matrix into a tree structure. NGSDIST estimates pairwise distances based on two-dimensional site-frequency spectra (described in Korneliussen *et al.* 2013; Vieira *et al.* 2015). In order to calculate genetic distances within our low-coverage data and to subsequently compute a tree, we first calculated posterior probabilities for all possible genotypes using ANGSD (with the –doGeno 8 function). We then ran NGSDIST with 100 bootstrap replicates and a block size of 5. The phylogenetic tree was then calculated using FASTME and visualized with the plotTree.R script, which is provided as part of the ngsTools package.

Phylogenetic Analyses using Mitochondrial DNA Data.—We further investigated phylogenetic relationships and divergence times between populations using mitochondrial DNA data. We inferred the best fitting substitution model using jModeltest2 (Darriba *et al.* 2012). The Akaike Information Criterion (AIC) showed the highest support for a GTR+I+G model. We subsequently used this model in the Bayesian phylogenetic reconstruction and divergence dating. We used the mutation rate previously estimated for plant hoppers of 2.7% sequence divergence per million years (Goodman *et al.* 2012) for the analysis. We reconstructed the phylogenetic relationships using a Yule prior-based strict molecular clock model as implemented in BEAST2 (Bouckaert *et al.* 2014) The analysis was carried out running a MCMC chain length of 100M iterations, sampling every 50,000 iteration. We then checked the convergence of the parameters and effective sample sizes using Tracer 1.7.1 (Rambaut *et al.* 2018). We did not perform divergence time estimates for the nuclear data due to uncertainty related to genotype calls from low coverage data. Furthermore, there is no genome-wide mutation estimate available for our species or planthoppers in general.

Ancestral State Reconstruction for Host-use and Island Origins, and Relative Timing of Divergence.—To quantify the numbers of host and island shifts, we performed ancestral state reconstructions, using mitochondrial DNA and nuclear DNA data separately. To do so, we loaded the two respective phylogenetic trees (see section above) into Mesquite 1.0 (Maddison 2008). We then imported the host-use and the island origins as character states and carried out ancestral state reconstruction using ‘Parsimony Ancestral States’. Parsimony reconstruction infers ancestral states that minimize the number of required character changes based on the phylogenetic tree and a character states matrix.

Results

Bioinformatics:

The capture experiment resulted in 77,143Mb of raw data (188 – 1075 Mb per sample), 37,395Mb (101 – 470 Mb per sample) after sequences were trimmed, low quality sequences were dropped, and adapters and duplicates were removed. Average coverage in the dataset was 8×, with any individual with an average coverage of less than 5× removed, with the exception of 5 individuals with 4× coverage. There were 184 individuals for analysis (Table 1). The specificity (average percentage of reads mapped) was 24.25%, while sensitivity (average percentage of the target captured) was 50%. After filtering by SNPCLEANER, 1,646,142 sites were passed into ANGSD. We obtained an average coverage of 273× for the mitochondrial loci.

Data analysis:

Population Variation and Structure.—The SFS had a shape that is typical for a neutral scenario (exponential decrease: Figure S2) (Marth *et al.* 2004). Across population and host plant comparisons, average per-site *pi* ranged from 0.0028 (E. Maui, *Clermontia*) – 0.0054 (Kohalas, *Pipturus*), Watterson's *theta* ranged from 0.0028 (E. Maui, *Clermontia*) – 0.0068 (Kohalas, *Clermontia*) and Tajima's D values ranged from −1.0034 (Kohalas, *Clermontia*) – 0.8211 (Hualalai, *Clermontia*) (Table S2). Usually, Tajima's D values >2 and <−2 are considered likely significant. Fst values among population pairs, separated by geographic location and by host plant, ranged from 0.07 (Kohalas, *Pipturus* vs. Mauna Kea, *Pipturus*) – 0.53 (E. Maui, *Clermontia* vs. Kohalas, *Dubautia*) (Table 2, Figure S3), and there were clear clusters evident in the PCA plots comparing the first three principal components (Figure S4a–c in Supporting Information).

Admixture Analysis.—The admixture analysis showed the highest convergence of the parallel runs for a grouping of 6 populations for the *Pipturus*-associated individuals (Figure S5a, Figure 2 top) and 8 for the *Clermontia*-associated individuals (Figure S5b, Figure 2 bottom). We selected these values because they show the least amount of variation among runs while explaining much of the structure in the data. Simultaneously, the slope of the log likelihood curve begins to level off at this K value in both groups. The West Maui population was not included in this analysis because exploratory analyses indicated that it was very different from the others. Including highly distinct populations together with populations that are less distinct can overwhelm the signal in the data and make it difficult for the algorithm to converge. This makes it difficult to interpret what is happening among the less distinct populations.

Within the Hawaii Island *Pipturus*, on Mauna Kea there are two distinct genetic lineages (yellow and green, Figure 2). The Kohala group was the most distinct from the rest of the populations, and there was evidence of genetic mixing within this site, with individuals that appear backcrossed between Kohala (purple), both Mauna Kea groups (yellow and green) and Kau (blue) individuals. East Maui was a very distinct group from the Hawaiian specimens.

Within *Clermontia*-associated individuals, East Maui again made up a very distinct group, and again there is evidence of genetic mixing in the Kohalas in that backcrossed individuals between the Mauna Kea and Kohala sites are present (Figure 2). Furthermore, there were three distinct genetic groups within this site: one only found in the Kohalas (green), one found both in the Kohalas and Mauna Kea (purple), and one found both in the Kohalas and Puu Makaala (light blue). In addition, there was genetic mixing in Kau and South Kona. In Kau, there is a distinct genetic group and also backcrossed individuals between Puu Makaala, the Kohalas, Mauna Kea and South Kona. In South Kona there is a distinct genetic group and there also were mixed individuals with genes representing Puu Makaala, Kau, the Kohalas, and Mauna Kea. Both in Puu Makaala and in Hualalai, there were individuals present from two very distinct genetic populations, suggesting early colonization by an older lineage followed by recolonization by younger lineages.

Phylogenetic Analyses.—The approach using genotype calls and the approach using genotype posteriors yielded similar topologies (Figure 3, Figure S6 in Supporting Information). There were three main lineages present within our data: (1) a West Maui (*Clermontia*-associated) lineage that was most closely related to individuals collected from *Pipturus* from both islands; (2) a set of taxa collected from *Pipturus* in East Maui and Hawaii; and (3) a Hawaii (*Clermontia*-+ *Dubautia*-+ *Cyanea*-associated*)* clade that included dispersal back to Maui (Figure 3).

In the first clade, individuals collected from *Clermontia* on West Maui formed a very distinct genetic lineage that was most closely related to a genetic lineage that was collected from *Pipturus*. This *Pipturus* lineage had an East Maui population at its base, largely followed by Kohala, after which it spread across Hawaii Island. Within the second clade, individuals from Kohala were at the base of the clade. The *Dubautia* group was genetically distinct and sister to the remaining groups, including a group collected from *Clermontia*. This led to a group on Hawaii Island that was primarily on *Clermontia* with shifts to other hosts and substantial geographic movement. Within this Hawaii Island *Clermontia*-associated group, there were two genetic lineages: One genetic lineage was composed of sister populations from Hualalai and Puu Makaala. The second contained populations that appeared to have diversified in the following sequence: from Kohala, dispersal to South Kona with one population dispersing from Hawaii Island back to East Maui, while another spread across Hawaii to Mauna Kea, Saddle Road and Kau, with colonization back to Hualalai and Puu Makaala. However, not all nodes supporting this sequence had strong bootstrap support. There were four examples of adventitious capture apparent in the data (Kau.Cl.37, Koh.Pi. 77, Koh.Pi.76 and SKO.Cl.41), suggesting a high degree of mobility within the group, despite its tendency to produce pronounced geographic structure.

Ancestral state reconstruction for host-use and island origins, and relative timing of divergence.—Ancestral state reconstruction of island origins and host-use reconstructed markedly different histories for both nuclear and mitochondrial DNA regions (Figure S7). Most notably, the mitochondrial DNA tree showed two *Clermontia*-associated lineages on West Maui, while there was only one in the nuclear tree. Moreover, the mtDNA showed two main *Pipturus*-associated lineages on Hawaii Island that were not closely related

to the Maui *Pipturus*-associated lineage. In contrast, the nuclear DNA showed a single *Pipturus*-associated lineage that appeared on Maui and then went on to colonize Hawaii Island. Importantly, the shift to *Pipturus* itself is not associated with colonization of Hawaii Island. Rather, the shift to *Pipturus* occurred on East Maui, geographically separated from its closest *Clermontia*-associated sister population on West Maui. This shift may have preceded the secondary colonization of Hawaii Island. The BEAST analysis of the mitochondrial data indicated that *N. umbratica* began to diverge 1.77 million years ago, which is consistent with the early formation of Maui (Figure 1, Figure S8).

Discussion

The most striking result from the current study is the highly structured nature of populations within *Nesosydne umbratica* across the young islands of Maui and Hawaii (inferred using nuclear DNA data). The structure is shaped by both host plant and geographic locality, with host specialization playing the strongest role in shaping structure. This pattern is consistent with other studies that highlight the role of host associations in shaping species diversification (e.g., Bennett & O'Grady 2012; Condamine *et al.* 2012; Futuyma & Agrawal 2009; Jermy 1984).

Geographic Structure

N. umbratica shows marked geographic structure across the very young Hawaii Island. There are several mechanisms that could account for this. First, a population might have established on the growing young island, going on to colonize younger volcanoes as they formed. Once a population became established at any one site, further immigration may have been prevented as a result of priority or monopolization effects, in which an early colonizing species in an area has an advantage over subsequent colonizers (Fukami 2015). Indeed, the first colonists are likely to be able to monopolize new habitat, which can lead to long-term establishment and associated adaptive evolution (De Meester *et al.* 2016). Despite the highly structured nature of the populations, it is very clear that they have moved between the islands multiple times, although the ancestral state reconstruction is ambiguous about the island of origin (Figures 3, S5). However, the dating analysis suggests that diversification must have begun on Maui given the 1.77 my age of the oldest divergence in the lineage (Figure S8).

Host shifting and host-associated structure

In continental regions, the majority of species in the Delphacidae are restricted to monocots, especially grasses and sedges. The Hawaiian *Nesosydne* are similarly restricted by host, with 88% of the 82 species feeding on species within a single plant family and 77% of species feeding on a single plant species. However, the Hawaiian radiation, taken together, has been recorded on a total of 28 plant families, mostly dicots. The plant family with the most recorded *Nesosydne* species is the Asteraceae with 25 species, while 12 plant families support a single species each. For the current study, the primary host plants of *N. umbratica* were *Clermontia* (Campanulaceae) and *Pipturus* (Urticaceae); *N. umbratica* is one of 9 species found on Campanulaceae, and one of 8 species found on Urticaceae. The *N. umbratica* from *Clermontia* and *Pipturus* on East Maui and Hawaii Island fall in to clearly

divergent lineages. *Clermontia* is shown to be the ancestral host, *Pipturus*, the derived host (Figure S7). Based on the data there are two possibilities as to where and how the shift to *Pipturus* occurred. (1) Colonization of a new geographic location (East Maui from West Maui) was associated with host expansion, resulting in a generalist host association, "*Clermontia* + *Pipturus*". This generalist population subsequently gave rise to a specialist *Pipturus*-associated species which then colonized Hawaii Island. (2) Colonization of the new host (*Pipturus*) occurred on West Maui (not sampled for *Pipturus*) with subsequent colonization of East Maui.

Both scenarios emphasize the importance of geographic isolation, a result previously reported for the related *Nesosydne chambersi* from Hawaii Island (Goodman *et al.* 2012). The first scenario would support arguments in which episodes of increasing host generalization alternate with specialization on particular hosts (Janz & Nylin 2008); such events, when coupled with taxon pulses (cyclical episodes of expansion and isolation in geographical range) are known drivers of diversification (Hoberg & Brooks 2008). The second scenario would support hypotheses in which diversification is driven simply by shifts in host-plant use (Hardy & Otto 2014). In the case of *N. umbratica*, the evidence is stronger for (2), because populations tend to be highly structured by both host and geographic locality, even though individuals are occasionally found on alternate hosts (Figure 3). The shift from *Clermontia* to the new host (*Pipturus*) on Maui may have been facilitated by some measure of 'ecological fitting' with subsequent selection on that new host that allowed adaptation. Gompert et al. (2015) have argued that the tendency to shift host is dictated by low levels of gene flow between the ancestral and derived populations, in this case facilitated by geographic isolation at a different site (East Maui), allowing alleles affecting performance on the ancestral host to be lost through drift; this effect might be expected to be enhanced by small population sizes on novel hosts.

Parallel biogeographic patterns subsequent to host shifts

The host shift to *Pipturus* on Maui preceded the establishment of this lineage on Hawaii Island, potentially facilitating establishment and allowing parallel radiations throughout the wet forests on both *Pipturus* and *Clermontia*. It is clear that populations within *N. umbratica* are (1) geographically discrete; and (2) largely host specialists within a given geographic area. Specialization on different hosts is evidently associated with the parallel geographic expansions of the different host lineages (Figure 4) such that the different host lineages co-occur at many sites throughout Hawaii Island as well as on East Maui (Figure 1). Tantalizing evidence suggests there might even be a third parallel radiation within this group: Collections were also taken from *Dubautia* (a host plant in the Asteraceae, a different family entirely) in the Kohalas on Hawaii Island. The phylogenetic analysis indicated that this population's position is sister to the Hawaii Island *Clermontia*-associated group, and that it is deeply divergent from them. This suggests that future collections made from other parts of the island from *Dubautia* would fall within this group.

Discrepancies between mitochondrial and nuclear data and divergence times

Comparison of the phylogeny generated from mitochondrial DNA show some marked differences compared to nuclear DNA. Overall, the mtDNA suggests much more movement

than the nuclear tree, in relation to both the number of host shifts (16 from the mtDNA tree, 7 from the nuclear tree) and number of geographic shifts between volcanoes (25 inferred from the mtDNA tree, 17 from the nuclear tree). The finding of higher inferred movement from mtDNA markers as compared to nuclear makers has been found in other Hawaiian insects (Shaw 2002) and the general phenomenon has been discussed extensively (Toews & Brelsford 2012). One possible explanation is that shared hosts facilitate hybridization between species and that mitochondrial introgression occurs more frequently due to differences in abundance between the two species (Linnen & Farrell 2007). Such biased mitochondrial introgression is expected to be more pronounced when hybridizing species differ in abundance (Chan & Levin 2005). Thus, in the sawfly genus *Neodiprion*, this effect of a predisposition of mtDNA to introgression was used to highlight the importance of interspecific hybridization events in the evolutionary history of the lineage.

Our data for *N. umbratica* shows two *Clermontia*-associated lineages on West Maui for the mt DNA tree, while there is only one in the nuclear tree, possibly indicative of hybridization. The potential role of admixture is supported by the NGSadmix results, at least in Kohala and Kau *Clermontia*-associated populations, though more needs to be done to unravel the role of admixture. Moreover, the nuclear DNA shows that the shift to *Pipturus* occurred on East Maui, and this lineage was the progenitor for the entire *Pipturus*-associated lineage on Hawaii Islands. In contrast, the mtDNA shows that the East Maui *Pipturus*-associated lineage is embedded within *Clermontia*-and *Cyanea*-associated populations from the same location. This raises the intriguing possibility that the shift to *Pipturus* on East Maui was associated with introgressive hybridization between populations from different hosts. In terms of the divergence dating, the mitochondrial tree we find that some of the early divergences between lineages are old relative to the island ages. However, calibrations based on recent and population-level splits are known to yield very high mutation rates, which has been attributed to the persistence of mutations as transient polymorphisms early while populations are diverging (Ho & Lo 2013; Papadopoulou *et al.* 2010). As a result, divergence dates may be pushed back.

Conclusions

The study highlights some fundamental attributes of the ongoing and recent divergence among populations of *N. umbratica*. First, they are highly structured by host and second they are highly structured by geography, having radiated in parallel throughout the island of Hawaii. Based on these results we suggest that host shifting can occur quite readily given the availability of hosts to which they may show some preadaptation, colonization of new geographic areas then being associated with adaptation to the new host. Once *N. umbratica* populations have become established on a given host, they tend to be resistant to invasion by additional individuals, consistent with the concept of monopolization (De Meester *et al.* 2016). As a result, the pattern of colonization and adaptation in *N. umbratica* appears to be one of geographic isolation coupled with host shifting (Figure 4). It may be that over time after an initial shift, adaptation to a specific host may increase, making host shifts less likely as host associations are structuring the populations just as much as geographic separation. The study highlights the tight interplay between geographic isolation and host shifting in fostering divergence in these insects. Overall, it is clear that geographic isolation plays a

major role in isolating populations and potentially initiating speciation, as might be expected when the association between insects and their host plants is largely species-specific (Althoff *et al.* 2012). However, there is no evidence of expansion of host range within populations (Janz & Nylin 2008). Rather, host-shifting occurs as a discrete event given the availability of host plants that are sufficiently similar and are not already occupied (a phenomenon well documented in *Rhagoletis* flies, Powell *et al.* 2014). Such shifting events can allow secondary waves of colonization, with populations becoming highly structured by geography on the secondary host.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Some of the laboratory work was performed in the Evolutionary Genomics Laboratory in the Museum of Vertebrate Zoology at U.C. Berkeley. This study would only have been possible with support from Nature Conservancy of Hawaii (Pat Bily), the State Department of Land and Natural Resources and the Hawaii Natural Areas Reserve System (Cynthia King) who provided permits and logistic support for KRG to collect on their lands. We are very grateful to Lydia Smith for her advice and support throughout the project. We also thank Cerise Chen for her careful help with sample preparation. This work used the Vincent J. Coates Genomics Sequencing Laboratory at U.C. Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. This project was supported by the NSF's Dimensions of Biodiversity program grant DEB1241253.

References

Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. Ecology Letters 11, 1123–1134. [PubMed: 18778274]

Althoff DM, Segraves KA, Smith CI, Leebens-Mack J, Pellmyr O (2012) Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. Molecular Phylogenetics and Evolution 62, 898–906. [PubMed: 22178365]

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Asche M (1997) A review of the systematics of Hawaiian planthoppers (Hemiptera: Fulgoroidea). Pacific Science 51, 366–376.

Bennett GM, O'Grady PM (2012) Host–plants shape insect diversity: Phylogeny, origin, and species diversity of native Hawaiian leafhoppers (Cicadellidae: Nesophrosyne). Molecular phylogenetics and evolution 65, 705–717. [PubMed: 22884527]

Bi K, Linderoth T, Vanderpool D, et al. (2013) Unlocking the vault: next-generation museum population genomics. Molecular ecology 22, 6018–6032. [PubMed: 24118668]

Bi K, Vanderpool D, Singhal S, et al. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC genomics 13, 403. [PubMed: 22900609]

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. [PubMed: 24695404]

Bouckaert R, Heled J, Kühnert D, et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. . PLoS computational biology 10, e1003537.

Chan KM, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59, 720–729. [PubMed: 15926684]

Condamine FL, Sperling FA, Wahlberg N, Rasplus JY, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. Ecology Letters 15, 267–277. [PubMed: 22251895]

Conesa A, Götz S, García-Gómez JM, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. [PubMed: 16081474]

Crawford JE, Lazzaro BP (2012) Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. Frontiers in genetics 3.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature methods 9.

De Meester L, Vanoverbeke J, Kilsdonk LJ, Urban MC (2016) Evolving perspectives on monopolization and priority effects. Trends in Ecology & Evolution 31, 136–146. [PubMed: 26778169]

Denno RF, Perfect TJ (1994) Planthoppers as models for ecological study and effective pest management In: Planthoppers, pp. 1–4. Springer.

Doellman M, Ragland G, Hood G, et al. (2018) Genomic differentiation during speciation-with-gene-flow: comparing geographic and host-related variation in divergent life history adaptation in Rhagoletis pomonella. Genes 9, 262.

Drew AE, Roderick GK (2005) Insect biodiversity on plant hybrids within the Hawaiian silversword alliance (Asteraceae: Heliantheae-Madiinae). Environmental Entomology 34, 1095–1108.

Fennah RG (1958) Fulgoroidea of south-eastern Polynesia. Transactions of the Royal Entomological Society of London 110, 117–220.

Forbes AA, Devine SN, Hippee AC, et al. (2017) Revisiting the particular role of host shifts in initiating insect speciation. Evolution 71, 1126–1137. [PubMed: 28052326]

Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. Annual Review of Ecology, Evolution, and Systematics 46, 1–23.

Fumagalli M, Vieira FG, Korneliussen TS, et al. (2013) Quantifying population genetic differentiation from next-generation sequencing data. Genetics 195, 979–992. [PubMed: 23979584]

Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences 106, 18054–18061.

Giffard WM (1922) The distribution and island endemism of Hawaiian Delphacidae (Homoptera) with additional lists of their food plants. Proc. Hawaii. Entomol. Soc. 5, 103–118.

Gompert Z, Jahner JP, Scholl CF, et al. (2015) The evolution of novel host use is unlikely to be constrained by trade-offs or a lack of genetic variation. Molecular ecology 24, 2777–2793. [PubMed: 25877787]

Goodman KR (2010) Emerging biodiversity: diversification of the Hawaiian Nesosydne planthoppers, UC Berkeley.

Goodman KR, Welter SC, Roderick GK (2012) Genetic divergence is decoupled from ecological diversification in the Hawaiian Nesosydne planthoppers. Evolution 9, 2798–2814.

Haas BJ, Papanicolaou A, Yassour M, et al. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature protocols 8.

Hardy NB, Otto SP (2014) Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. Proceedings of the Royal Society of London B: Biological Sciences 281, 20132960.

Hasty GL (2005) Host-plant specialization affects species diversity and vulnerability to endangerment in a group of island-endemic phytophagous insects (Delphacidae, Hawaiian Islands), University of California at Berkeley.

Hernández-Teixidor D, López H, Pons J, Juan C, Oromí P (2016) Host plant associations and geographical factors in the diversification of the Macaronesian Rhopalomesites beetles (Coleoptera: Curculionidae). Journal of Biogeography.

Ho SY, Lo N (2013) The insect molecular clock. Australian Journal of Entomology 52, 101–105.

Hoberg EP, Brooks DR (2008) A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. Journal of Biogeography 35, 1533–1550.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome research 9, 868–877. [PubMed: 10508846]

Huson D, Bryant D (2006) Application of phylogenetic networks in evolutionary studies, Mol Boil Evol, 23: 254–267. Software available from www.splitstree.org.

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. Genome research 17, 377–386. [PubMed: 17255551]

Janz N (2011) Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. Annual Review of Ecology, Evolution, and Systematics 42, 71.

Janz N, Nyblom K, Nylin S (2001) Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. Evolution 55, 783–796. [PubMed: 11392396]

Janz N, Nylin S (2008) The oscillation hypothesis of host-plant range and speciation In: Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects (ed. Tilmon KJ), pp. 203–215. University of California Press, Berkeley, CA.

Janzen DH (1985) Dan Janzen's thoughts from the tropics 1: On ecological fitting. Oikos 45, 308–310.

Jermy T (1984) Evolution of insect/host plant relationships. The American Naturalist 124, 609–630.

Jiang H, Lei R, Ding S-W, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC bioinformatics 15, 182. [PubMed: 24925680]

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. [PubMed: 18397895]

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric'speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. Molecular ecology 15, 2935–2947. [PubMed: 16911212]

Kircher M, Sawyer S, Meyer M (2011) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic acids research 40, e3–e3.

Kirkaldy GW (1902) Hemiptera In: Fauna Hawaiiensis (ed. Sharp D), pp. 93–174. Cambridge University Press, Cambridge.

Kirkaldy GW (1910) Supplement to Hemiptera In: Fauna Hawaiiensis (ed. Sharp D), pp. 531–599. Cambridge University Press, Cambridge.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC bioinformatics 14, 289. [PubMed: 24088262]

Langellotto GA, Denno RF, Ott JR (2000) A trade-off between flight capability and reproduction in males of a wing-dimorphic insect. Ecology 81, 865–875.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357–359. [PubMed: 22388286]

Lefort V, Desper R, Gascuel O (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Molecular biology and evolution 32, 2798–2800. [PubMed: 26130081]

Li H, Handsaker B, Wysoker A, et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Li J, Li H, Jakobsson M, et al. (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology 21, 28–44. [PubMed: 21999307]

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. [PubMed: 16731699]

Lim JY, Marshall CR (2017) The true tempo of evolutionary radiation and decline revealed on the Hawaiian archipelago. Nature 543, 710–713. [PubMed: 28297717]

Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. Evolution: International Journal of Organic Evolution 61, 1417–1438. [PubMed: 17542850]

Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. Trends in Ecology & Evolution 18, 220–227.

Maddison WP (2008) Mesquite: a modular system for evolutionary analysis. Evolution 62, 1103–1118. [PubMed: 18298648]

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. [PubMed: 21903629]

Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166, 351–372. [PubMed: 15020430]

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17, pp. 10–12.

Matsubayashi KW, Ohshima I, Nosil P (2010) Ecological speciation in phytophagous insects. Entomologia experimentalis et applicata 134, 1–27.

McKenna A, Hanna M, Banks E, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20, 1297–1303. [PubMed: 20644199]

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010, pdb. prot5448.

Nakadai R, Kawakita A (2016) Phylogenetic test of speciation by host shift in leaf cone moths (Caloptilia) feeding on maples (Acer). Ecology and Evolution 6, 4958–4970. [PubMed: 27547326]

Nosil P, Mooers A (2005) Testing hypotheses about ecological specialization using phylogenetic trees. Evolution 59, 2256–2263. [PubMed: 16405168]

Nyman T (2010) To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. Biol. Rev. 85, 393–411. [PubMed: 20002390]

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Molecular Biology and Evolution 27, 1659–1672. [PubMed: 20167609]

Pavlidis P, Alachiotis N (2017) A survey of methods and tools to detect recent and strong positive selection. Journal of Biological Research-Thessaloniki 24, 7.

PEP PEPP (2009) Plant Extinction Prevention (PEP) Program-Annual Report. State of Hawaii, Division of Forestry and Wildlife, Native Ecosystems Protection & Management, Honolulu, HI.

Powell TH, Forbes AA, Hood GR, Feder JL (2014) Ecological adaptation and reproductive isolation in sympatry: genetic and phenotypic evidence for native host races of R hagoletis pomonella. Molecular Ecology 23, 688–704. [PubMed: 24351094]

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology 67, 901–904. [PubMed: 29718447]

Roderick GK (1997) Herbivorous insects and the Hawaiian silversword alliance: coevolution or cospeciation? Pacific Science 51, 440–449.

Roderick GK, Metz EC (1997) Biodiversity of planthoppers (Hemiptera: Delphacidae) on the Hawaiian silversword alliance: effects of host plant phylogeny and hybridisation. Memoirs of the Museum of Victoria 56, 393–399.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome research 22, 939–946. [PubMed: 22267522]

Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences 99, 16122–16127.

Shaw KL, Gillespie RG (2016) Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. Proceedings of the National Academy of Sciences 113, 7986–7993.

Simpson JT, Wong K, Jackman SD, et al. (2009) ABySS: a parallel assembler for short read sequence data. Genome research 19, 1117–1123. [PubMed: 19251739]

Singhal S (2013) De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set. Molecular Ecology Resources 13, 403–416. [PubMed: 23414390]

Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. Genetics 195, 693–702. [PubMed: 24026093]

Stireman J (2005) The evolution of generalization? Parasitoid flies and the perils of inferring host range evolution from phylogenies. Journal of evolutionary biology 18, 325–336. [PubMed: 15715839]

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21, 3907–3930. [PubMed: 22738314]

Vamosi JC, Armbruster WS, Renner SS (2014) Evolutionary ecology of specialization: insights from phylogenetic analysis. Proc. R. Soc. B

Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M (2015) Improving the estimation of genetic distances from Next-Generation Sequencing data. Biological journal of the Linnean Society 117, 139–149.

Weingartner E, Wahlberg N, Nylin S (2006) Dynamics of host plant use and species diversity in Polygonia butterflies (Nymphalidae). Journal of evolutionary biology 19, 483–491. [PubMed: 16599924]

West-Eberhard MJ (2003) Developmental plasticity and evolution Oxford University Press.

Wiens JJ, Lapoint RT, Whiteman NK (2015) Herbivory increases diversification across insect clades. Nature communications 6.

Wilson WW, Mitter C, Denno RF, Wilson MR (1994) Evolutionary patterns of host plant use by delphacid planthoppers and their relatives In: Planthoppers: Their Ecology and Management (eds. Denno RF, Perfect TJ), pp. 7–113. Chapman & Hall, New York.

Winkler IS, Mitter C (2008) The phylogenetic dimension of insect-plant interactions: a review of recent evidence. Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects, 240–263.

Xue J, Zhou X, Zhang C-X, et al. (2014) Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome Biology 15, 521. [PubMed: 25609551]

Zimmerman EC (1948) Homoptera: Auchenorhyncha University of Hawaii Press, Honolulu, HI.

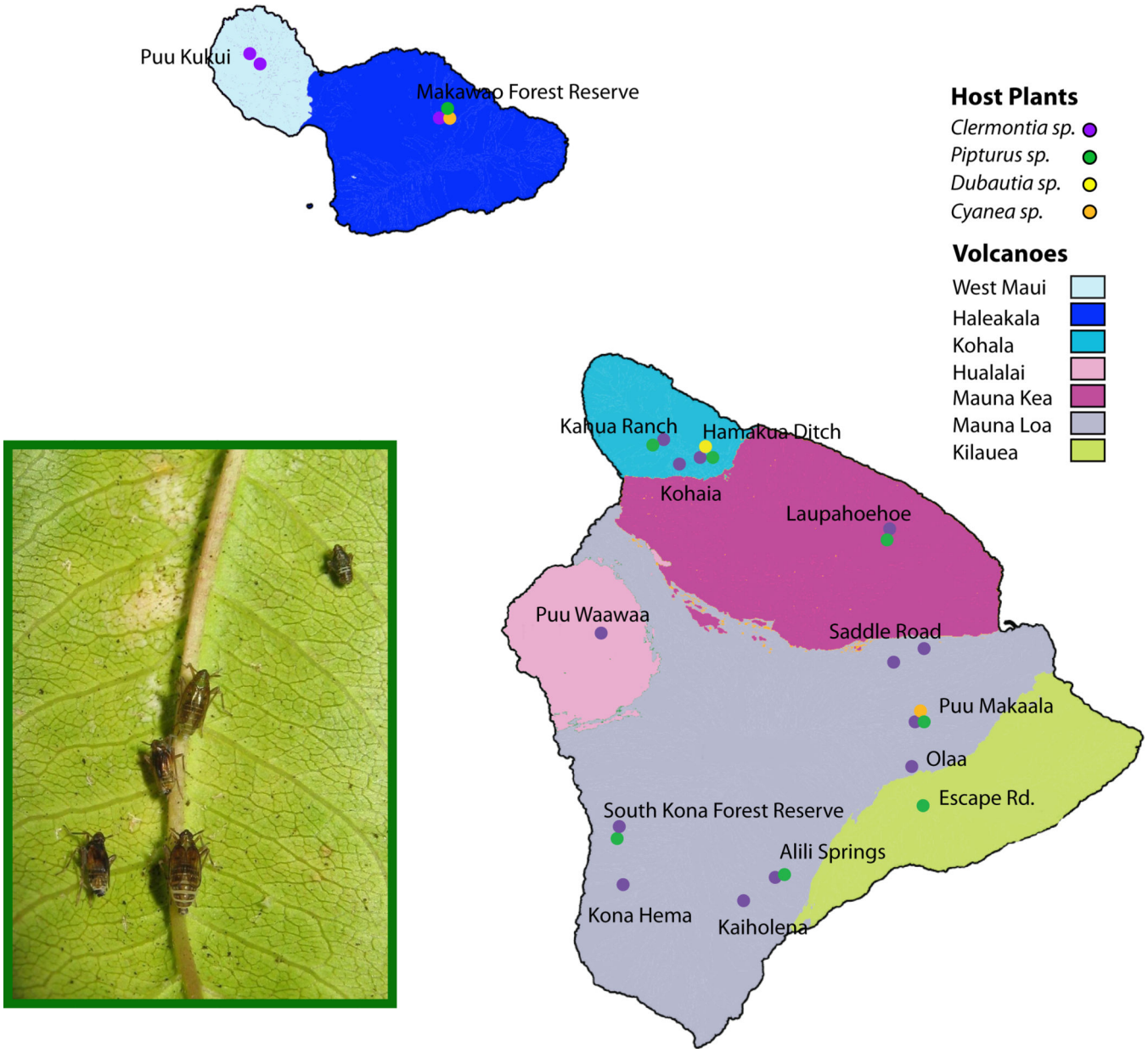

Figure 1.

Collecting localities and specimens. Collecting sites for *Nesosydne umbratica*, and the hosts on which they were collected, are shown as circles. Background colors indicate the different volcanoes, ranging from Kilauea (approximate upper limit of time available for colonization, 0.10 Ma), Mauna Loa (0.20 Ma), Mauna Kea (0.38 Ma), Hualalai (0.40 Ma), Kohala (0.43 Ma), Haleakala (0.8 Ma), and West Maui (1.3 Ma). Inset: Specimens of *N. umbratica* in their natural setting under the leaf, from Hawaii Volcanoes National Park. Photo credit: Karl Magnacca.

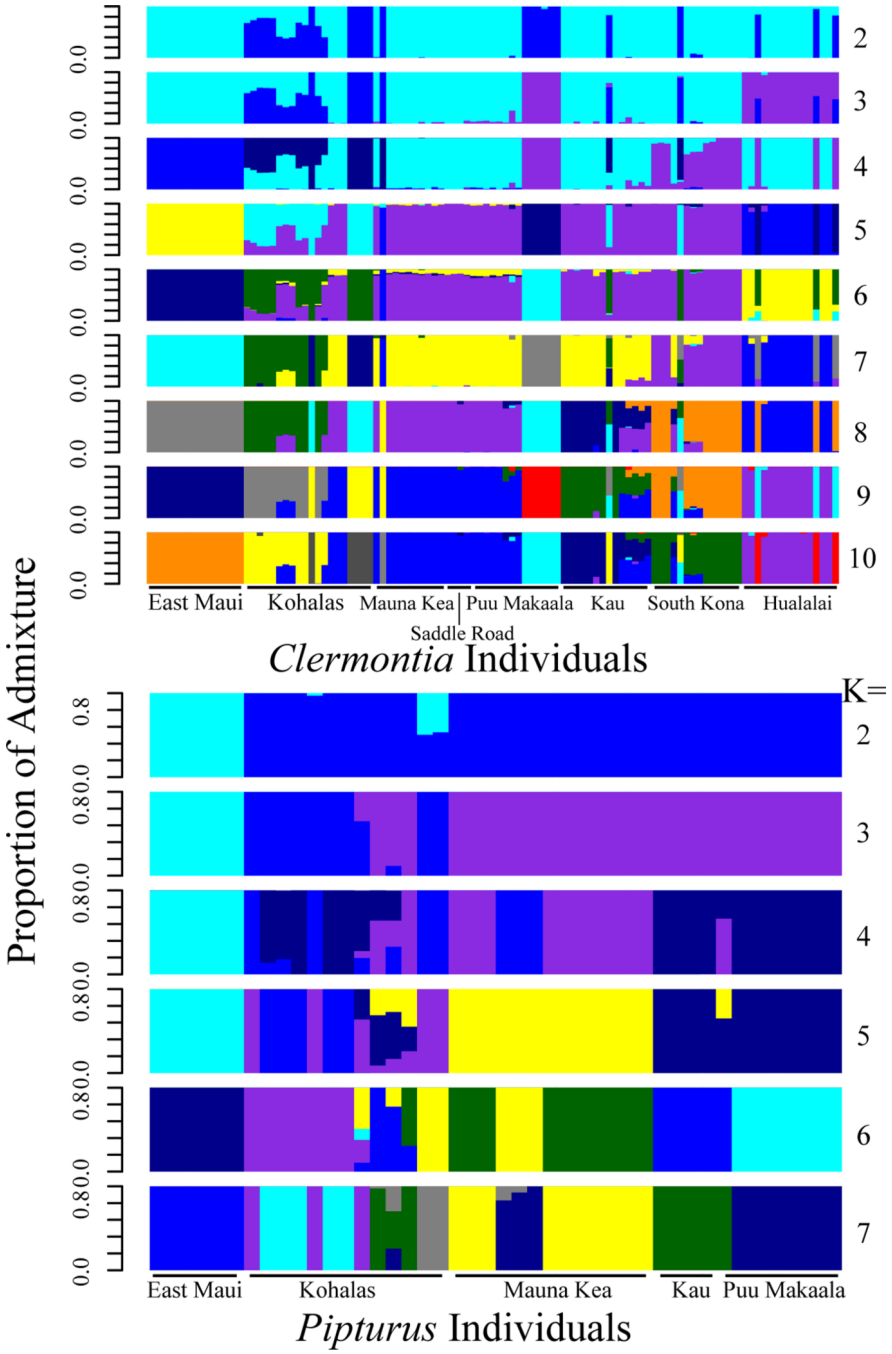

Figure 2.

Results of Bayesian clustering admixture analysis for different values of K (putative ancestral populations: 2–7 for insects from *Pipturus*, 2–10 for insects from *Clermontia*) using genotype likelihood data from the populations shown in Figure 1. Each bar corresponds to an individual; colors within bars represent admixture proportions. Thothe analysis identified 6 populations as the optimum number for the *Pipturus*-and 8 for the *Clermontia*-associated clades (see also Figure S5).

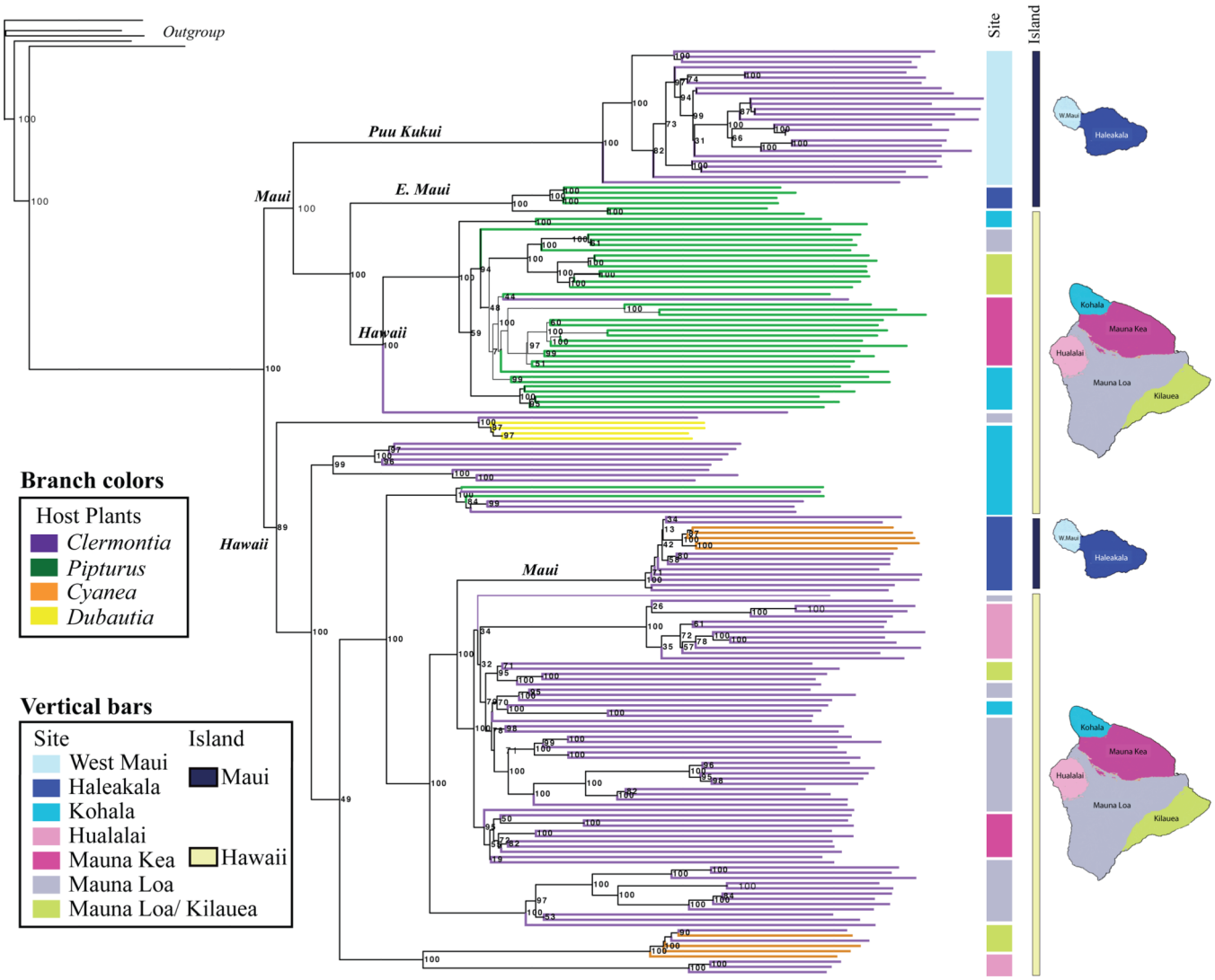

Figure 3.

Population tree of *Nesosydne umbratica*, based on ngsDist analysis. Analysis run with 100 bootstrap replicates and a block size of 5. The phylogenetic tree shows three distinct lineages: (1) a West Maui (*Clermontia*-associated) lineage that is most closely related to individuals collected from *Pipturus* from both islands; (2) a set of taxa collected from *Pipturus* in East Maui and Hawaii; and (3) a Hawaii (*Clermontia*-+ *Dubautia*-+ *Cyanea*-associated) clade that includes dispersal back to Maui. Circles at nodes indicate support. Colors of vertical bars indicated on the right reflect locations of samples (sites) as shown in Figure 1. Samples shown with the same color were collected from broadly overlapping sites.

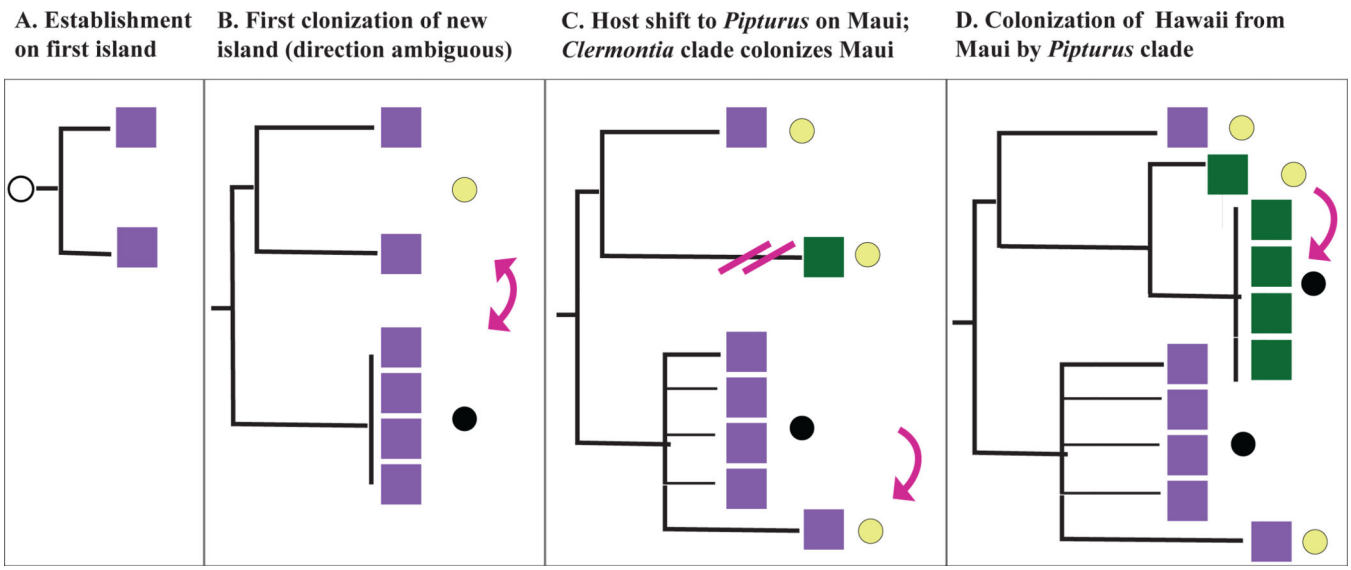

Figure 4.

Implied Sequence of Events Relating to Host Shifting and Geographic Isolation. This shows the scenario implied by the relative sequences of island colonization and host shifting events. Colors as in Figure 3: Purple squares indicate *Clermontia*-associated, green squares indicate *Pipturus*-associated. Pink arrows indicate island shifts; pink slashes indicate host shifts. Black circles, Maui Island; yellow circles, Hawaii Island.

Table 1:

Sampling table. *Nesosydne umbratica* specimen collection locations, with numbers of individuals by host plant. **Host plants** indicates the number of *Nesosydne* individuals used in the analysis from each genus of host plant at each locality. Collections were made across multiple plants at each site. **Total** is the total # of specimens used in the analysis.

Locality name	Host plants	Total
West Maui: Puu Kukui	*Clermontia arboresens waihiae* (26)	26
East Maui: Haleakala, Makawao Forest Reserve	*Clermontia arborescens waihiae* (10) *Cyanea duvalliorum* (5) *Pipturus forbesii* (6)	21
Hawaii: Kohalas: Koaia, Hamakua Ditch & Kahua Ranch	*Clermontia sp.* (16) *Dubautia laxa* (4) *Pipturus albidus* (13)	33
Hawaii: Mauna Kea, Laupahoehoe Forest Reserve	*Clermontia parviflora* (13) *Pipturus albidus* (13)	26
Hawaii: Mauna Loa, Saddle Road	*Clermontia parviflora* (4)	4
Hawaii: Mauna Loa/ Kilauea, Puu Makaala: Puu Makaala, Olaa & Escape Road	*Clermontia sp.* (10) *Cyanea sp.* (4) *Pipturus albidus.* (8)	22
Hawaii: Mauna Loa, Kau: Alili Springs & Kaiholena	*Clermontia sp.* (14) *Pipturus albidus* (4)	18
Hawaii: Mauna Loa, South Kona: Kona Hema & South Kona Forest Reserve	*Clermontia sp.* (14)	14
Hawaii: Hualalai, Puu Waawaa	*Clermontia clermontiodes* (15)	15
Outgroup: *Nesosydne raillardiae* Hawaii: Mauna Kea Mauna Loa: Hawaii Volcanoes National Park Mauna Loa: Ocean View Estates	*Dubautia ciliolata x D. arborea* (3) *Dubautia ciliolata x D. scabra* (1) *Dubautia scabra* (1)	5
Total		184

Table 2.

Pairwise Fsts among sampling localities. Host plants the insect populations were collected from are indicated by initials: C=*Clermontia,* D=*Dubautia,* P=*Pipturus*. Locations on Mauna Loa – PM, Puu Makaala; SK, South Kona.

	E. Maui (C)	**E. Maui (P)**	**W. Maui (C)**	**Kohala (C)**	**Kohala (D)**	**Kohala (P)**	**Mauna Kea (C)**	**Mauna Kea (P)**	**Saddle (C)**	**Mauna Loa PM (C)**	**Mauna Loa PM (P)**	**Mauna Loa Kau (C)**	**Mauna Loa Kau (P)**	**Mauna Loa SK (C)**
E. Maui, *Clermontia*	0													
E. Maui, *Pipturus*	0.507	0												
W. Maui, *Clermontia*	0.477	0.432	0											
Kohala, *Clermontia*	0.219	0.304	0.324	0										
Kohala, *Dubautia*	0.527	0.455	0.463	0.264	0									
Kohala, *Pipturus*	0.373	0.264	0.357	0.175	0.328	0								
Mauna Kea, *Clermontia*	0.212	0.407	0.412	0.089	0.411	0.263	0							
Mauna Kea, *Pipturus*	0.38	0.286	0.368	0.192	0.349	0.072	0.275	0						
Saddle, *Clermontia*	0.258	0.433	0.437	0.109	0.461	0.277	0.068	0.29	0					
Mauna Loa PM, *Clermontia*	0.24	0.366	0.381	0.104	0.368	0.239	0.107	0.253	0.125	0				
Mauna Loa PM, *Pipturus*	0.459	0.325	0.416	0.251	0.407	0.084	0.349	0.091	0.37	0.314	0			
Mauna Loa Kau, *Clermontia*	0.283	0.479	0.462	0.159	0.491	0.335	0.121	0.343	0.139	0.163	0.422	0		
Mauna Loa Kau, *Pipturus*	0.501	0.351	0.437	0.265	0.445	0.114	0.379	0.122	0.412	0.336	0.116	0.461	0	
Mauna Loa SK, *Clermontia*	0.311	0.456	0.447	0.174	0.459	0.316	0.181	0.318	0.218	0.192	0.398	0.228	0.433	0
Hualalai, *Clermontia*	0.305	0.453	0.443	0.167	0.457	0.316	0.169	0.326	0.197	0.227	0.398	0.227	0.431	0.25