Insect Systematics and Diversity, (2018) 2(5): 4; 1–14
doi: 10.1093/isd/ixy014
Research

OXFORD

Molecular Phylogenetics, Phylogenomics, and Phylogeography

Opening the Door to the Past: Accessing Phylogenetic, Pathogen, and Population Data From Museum Curated Bees

Anthony D. Vaudo,[1,3] Megan L. Fritz,[2] and Margarita M. López-Uribe[1]

[1]Department of Entomology, The Pennsylvania State University, 501 ASI Building University Park, PA 16802, [2]Department of Entomology, University of Maryland, 4112 Plant Sciences Building, College Park, MD 20742, and [3]Corresponding author, e-mail: adv124@psu.edu

Subject Editor: Jeffrey Lozier

Received 9 May 2018; Editorial decision 12 August 2018

Abstract

Tens of thousands of insects are deposited in collections every year as a result of survey-based studies that aim to investigate ecological questions. DNA-based techniques can expand the utility of these collections to explore their demographic and evolutionary history, temporal changes in their abundance, and pathogen dynamics. Using museum collections of the non-model bee species *Eucera* (*Peponapis*) *pruinosa* Say 1837 (Hymenoptera: Apidae: Eucerini), we developed a standard minimally-destructive and budget-friendly protocol to extract DNA and amplify common gene-fragments for barcoding, phylogenetic analysis, and pathogens. We also generated genome-wide single nucleotide polymorphism (SNP) data from DNA sequencing (ddRADseq) libraries for population structure analyses. We systematically studied the effect of specimen age (≤10 years ago) and tissue type (whole bees vs. abdomen) on DNA quality, single gene-fragment amplification, and SNP calling. We found that all analyses were achievable with both tissue types, yet with variable levels of efficiency because of general DNA degradation. Specifically, we found that not all samples yielded satisfactory results for molecular studies; however, we did not find a systematic effect of specimen age on DNA quality which is encouraging for future studies involving historical specimens. We report the first evidence for the presence of the microsporidian pathogen *Nosema* spp. in squash bees, opening a window for the study of historical changes in disease pressure in this important agricultural pollinator. Our protocols can be used as a template for the design of future experiments that extract multiple pieces of information using DNA-based methods from insect museum stored specimens.

Key words: pollinator, ddRAD, museum specimen, population genetics, *Nosema* spp.

Biological museum collections are indispensable resources for recording and analyzing patterns of biodiversity and the recent evolutionary history of species in an explicit spatio-temporal context (see Holmes et al. 2016 for an eloquent review). Physical specimens are essential for species morphological identification, assessments of intra and interspecific phenotypic variation, species phenology, geographic distribution, and biotic interactions (e.g., host–plants, prey–predator) among others. Additionally, information from molecular markers can add important biological and ecological data to these collection specimens. For example, modern genetic analyses of collection specimens could expand our knowledge into population demography and evolutionary changes for both model and non-model species (e.g., Bi et al. 2013, Lozier et al. 2016). However, the applicability of molecular tools to museum specimens can be limited by the degradation of DNA in museum specimens and the risk of destructive sampling for DNA extractions.

Molecular data from museum specimens could be particularly useful for studying 'at-risk', declining, and endangered taxa. The current concern for the status and health of bee populations over the globe has created a burgeoning field of research dedicated to identifying causes and possible solutions to their decline (Goulson et al. 2015). Many of these studies aim to characterize bee communities in different types of landscapes to identify the drivers explaining differences and changes in bee abundance, richness and composition across different communities (e.g., Brosi et al. 2007, Cameron et al. 2011, Fortel et al. 2014, Winfree et al. 2014, Gezon et al. 2015, Rollin et al. 2015, Russo et al. 2015, Kammerer et al. 2016, Hamblin et al. 2018). These research efforts have resulted in thousands of bee specimens deposited in collections every year with analyses limited to those describing community diversity and composition (but see Lozier and Cameron 2009, Cameron et al 2016). One of the main factors hindering the detection of bee declines over time is the lack

© The Author(s) 2018. Published by Oxford University Press on behalf of Entomological Society of America.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

of long-term population data evaluating demographic declines in bee species (Bartomeus et al. 2013, Goulson et al. 2015). The potential application of molecular tools on the millions of bee specimens available in stored insect collections opens a number of possibilities to gather data about current and past bee population-level demographic processes that can be linked to environmental changes.

Among the questions that can be addressed using molecular techniques from bee specimens stored in collections include: species identification through DNA barcoding (Sheffield et al. 2017), resolving phylogenetic relationships of groups that comprise rare species (Danforth et al. 2013, Kahnt et al. 2017), reconstructing the demographic history of species populations by identifying population bottlenecks (Lozier and Cameron 2009), determining effective population sizes (López-Uribe et al. 2014) or detection of geographic range expansions (Dellicour et al. 2013, López-Uribe et al. 2016). Insights gained from the application of molecular techniques to bees stored in insect collections can be diverse and far-reaching. At the intersection of between and within species level, studies of bee genetic diversity that include specimens collected from wide geographic areas have often resulted in the identification of cryptic species (i.e., determining *Bombus spp*. with overlapping mimicry complexes etc.; Murray et al. 2007, Carolan et al. 2012, Williams et al. 2012), or provide additional support to describe new species with monotonous morphologies (e.g., *Lasioglossum* (Halictidae) (Gibbs 2018); the subfamily Euryglossinae and genera *Liphanthus* and *Lopnchopria* (Colletidae) (Hogendoorn et al. 2015, Packer and Ruz 2017). Alternatively, molecular markers may identify the presence of physical polymorphisms within a species (Granto et al. in press). At the species level, population genetic analyses can identify the presence of population structure facilitating the discovery of independent evolutionary units that deserve different conservation efforts (Lecocq et al. 2013). Additionally, museum specimens can also be used to test hypotheses about emerging diseases and the historical role of pathogen dynamics in changes of bee populations over time (e.g., Cameron et al. 2016). Therefore, bee collections provide a promising but relatively untapped source for phylogenetics, phylogeography, population genetics, and pathogen dynamics data to examine individuals over time in natural and anthropogenic changing landscapes.

One of the main challenges of applying molecular techniques in museum specimens is that DNA degrades after specimen death (Frampton et al. 2008, Zimmermann et al. 2008), thus genetic information may be limited and difficult to obtain. In addition, most DNA extraction methods require the partial or total destruction of the specimen, which conflicts with the purpose of long-term storage of specimens in insect collections (Mitchell 2015, Sproul and Maddison 2017), though successful nondestructive techniques have been used (Gilbert et al. 2007, Tin et al. 2014, Cameron et al. 2016). Currently, hybridization-based methods to recover target conserved gene sequences of fragmented DNA (e.g., exon capture) work well with degraded DNA samples from collection but are costly to develop and limited in their use to species level phylogenetic studies given the focus on conserved protein coding sequences (Lemmon et al. 2012, Kanda et al. 2015, Mitchell 2015, Hamilton et al. 2016, Sproul and Maddison 2017). Studies of population genetics that aim to reconstruct recent demographic events rely on calling a large quantity of neutral single nucleotide polymorphisms (SNPs) across the genome for optimal statistical power (Helyar et al. 2011). Restriction-based protocols (also known as RAD-seq approaches; reviewed in Andrews et al. 2016) are widely used for the simultaneous discovery and genotyping SNPs in non-model species that lack reference genomes (Helyar et al. 2011, Peterson et al. 2012, Fritz et al. 2016, Alter et al. 2017, Nunziata et al. 2017, Raffini et al. 2017). However, these methods generally require high molecular weight DNA to decrease the proportion of missing data across samples (Graham et al. 2015, Andrews et al. 2016). Therefore their use for specimens stored in insect collections has been questioned.

In this study, we modified previously published DNA extraction methods (Gilbert et al. 2007) to create a minimally destructive (no tissue homogenization), and budget-friendly protocol for molecular studies of museum stored bee specimens. We used the squash bee, *Eucera (Peponapis) pruinosa* Say 1837 (Hymenoptera: Apidae: Eucerini; Dorchin et al. 2018) as our model system. *Eucera pruinosa* is an obligate pollen specialist and important pollinator of plants in the genus *Cucurbita* (family Cucurbitacae), which includes crops such as pumpkin, squash, and zucchini (Hurd et al. 1974). This solitary bee is native to Central America, has expanded its geographic range to North America following the widespread cultivation of *Cucurbita* crops, and their populations show signatures of recent demographic bottlenecks (López-Uribe et al. 2016). Because of its association with agricultural landscapes, we leverage the availability of large numbers of *E. pruinosa* in collections to conduct an analysis of population structure from the regions where bees were collected. We also used the extracted DNA to detect common pathogens of managed pollinators that share the same floral resources with *E. pruinosa* in agroecosystems.

Specifically, we describe a protocol to extract genomic DNA, amplify bee and pathogen DNA fragments and build ddRAD libraries while returning all physical identifying insect material to the collection. We systematically studied the effect of specimen age and bee tissue type used for extraction on DNA quality, yield, single polymerase chain reaction (PCR) fragment amplification and number of SNPs recovered from ddRAD libraries. We tested whether bees collected 1, 3, 5 or 10 yr ago and DNA extracted from the whole bee or just the abdomen affected the PCR amplification of nuclear and mitochondrial genes, and the number of genome-wide SNPs for estimation of population differentiation through *F*-statistics. We found that all single gene amplification and ddRAD-based analyses were achievable from specimens of all ages with both types of tissue (whole bees and abdomens), yet with variable levels of efficiency. We did not find a general effect of age on our ability to amplify and analyze gene segments, and we demonstrated the potential to build and assemble ddRAD libraries de novo (i.e., without a reference genome) for a non-model bee species, which is encouraging for future studies involving historical specimens.

Methods

Bee Specimen Collection and Storage

We selected specimens of male *E. pruinosa* individuals that were collected 1, 3, 5, and 10 yr prior to the study (years 2016, 2014, 2012, and 2007, respectively). The bees were collected from Adams (2012, 2014, 2016), Centre (2007), and Lancaster (2012) counties in Pennsylvania, United States (Fig. 1, Supp. Table S1). Specimens from 2007 were collected in blue pan traps while the remaining bees were collected in blue vane traps (Joshi et al. 2015). Briefly, the pan traps were filled with soapy water, left outside for 1 d, and bees were collected and stored in 70% ethanol. The blue vane traps were filled with 60:40 ethylene-glycol:water and left in the field for 1 wk to collect bees, after which the bees were stored in 70% ethanol. Within 4 wk of storage in alcohol, the bees were pinned and allowed to air dry, then placed in entomological drawers for permanent storage. These specimens were collected for various projects by the Penn State's Fruit Research and Extension Center (FREC)

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

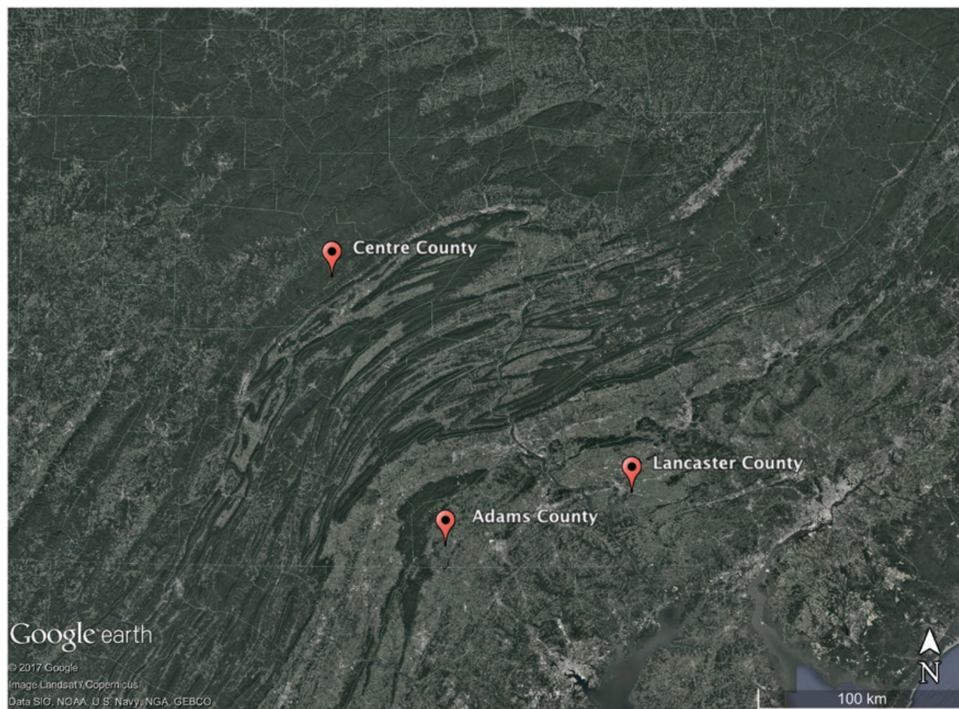

Fig. 1. Map of sites sampled in the state of Pennsylvania (United States) with labels for the counties in which bee specimens were collected.

and stored in the FREC Entomological Collection and Penn State's Frost Entomological Museum (see Supp. Table S1 for specimen collection data and identification numbers). Following DNA extraction (see below), we washed the whole bees or bee abdomens in soapy water, rinsed them with 100% ethanol, rinsed them again with 70% ethanol, then allowed the specimens to air dry overnight (following Droege 2015). This protocol allowed the specimens to recover their appearance after extraction. All specimens and abdomen (which were stored in clear pill capsules) were stored with their respective labels in the López-Uribe lab collection at the Pennsylvania State University (Fig. 2).

DNA Extraction and Quality Verification

We used two tissue types for DNA extractions: 'whole bee' or 'abdomen', but neither protocol required tissue homogenization. The minimally destructive technique for whole bees involved soaking the entire specimen in extraction buffer. For the whole bees, we removed the labels but left the bees pinned to prevent breaking the thorax when removing pins from the specimen. For the 'abdomen' method we aimed to reduce any alteration that might be caused by the extraction buffer and preserve head and thorax material for future genetic use. For this set of bees, we carefully wiggled the abdomen to separate the bees' metasoma from the thorax at the petiole and soaked only the abdomen in extraction buffer. After soaking the different tissues in proteinase K (250 μg/ml) extraction buffer for ~20 h, we followed a phenol:chloroform DNA extraction method for all samples (Gilbert et al. 2007; detailed protocol in Supp. Material SI1; modifications included total volume of extraction buffer for samples). DNA was extracted from 120 bee specimens: 24 bees from each year (2007, 2012, 2014, 2016), and 12 bees within each year assigned to whole bee or abdominal extractions. We extracted DNA from abdomen of an additional 12 specimens from both 2012 and 2014 to corroborate results from our initial extractions.

To determine the molecular weight of the extracted DNA, we used electrophoresis (1.2% agarose gel, 5 μl DNA extract mixed with 1 μl SYBR-green solution [1 μl in 250 μl 10× loading dye], 100V for 60 min) alongside Bio-Rad EZ Load 500 bp Molecular Ruler (Bio-Rad Laboratories, Inc., Hercules, CA). We categorized each sample as a binary variable for visual presence or absence of high molecular weight DNA (i.e., clear bands of DNA >500 bp; see Fig. 3). DNA purity was determined by 260/280 and 260/230 ratios on a SpectraMax Microplate Reader (Molecular Devices, LLC, San Jose, CA) with TE buffer as a blank. We quantified DNA concentration (ng/μl) through a Qubit Fluorometer using a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fischer Scientific, Inc., Waltham, MA). To determine if the age of bee specimens ('year') and tissue type ('whole body' or 'abdomen') affected DNA quality and efficiency of extraction, we ran independent analysis of variances (ANOVAs) for DNA 260/280 and 260/230 ratios and DNA yield, and contingency analysis for DNA quality. Because the potential confounding effect of collection type (blue vane vs. pan trap), we also determined the potential effect of collection on DNA quality as above, and for downstream analyses.

PCR Screening for Gene Fragments

To verify our ability to amplify single gene fragments from extracted DNA from museum stored bees, we amplified the nuclear ribosomal gene 28s (1000 bp), nuclear protein coding gene CAD (Carbamoyl-Phosphate Synthetase 2, Aspartate Transcarbamylase, and Dihydroorotase; 800 bp), and mitochondrial protein COI (cytochrome oxidase I; 850 bp). Because our initial screen showed no amplification for the 800 bp fragment of CAD, we designed primers for a shorter fragment size (130 bp) using Geneious v10.1.3 (Biomatters LTD, Aukland, New Zealand). We also tested the squash bees for evidence of two well-characterized bee pathogens: *Ascosphaera apis* (Maasen ex Claussen) L. S. Olive & Spiltoir (Onygenales: Ascosphaeraceae)—a fungal disease that attacks larvae

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

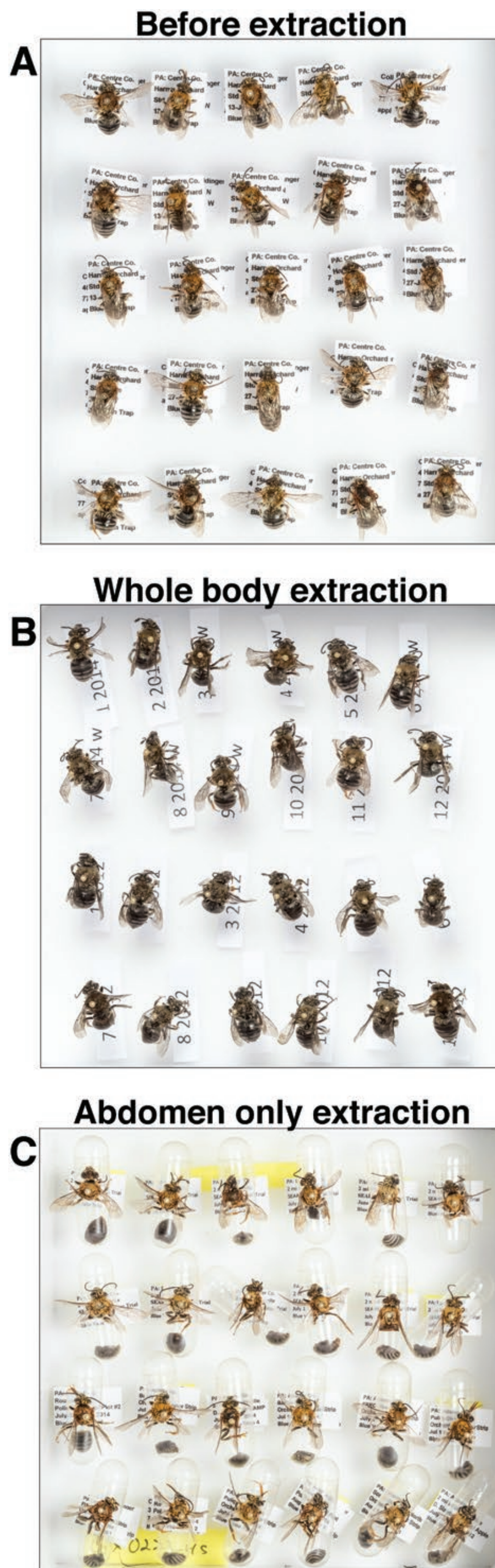

Fig. 2. Photographs of museum curated *Eucera pruinosa* males before (A) and after whole body (B) and abdomen only (C) DNA extraction. After extractions, all identifying morphological features of the bees remained intact. The main damage caused by whole body extractions was the darkening of orange/yellow thoracic hairs (B). The only damage from abdomen only extractions was the removal of the abdomen which were stored and pinned with specimens in pill capsules after extraction. Photographs by Nick Sloff.

(Gilliam et al. 1988, Jensen et al. 2013) and *Nosema* spp. Nägeli (Microsporida: Nosematidae)—a microsporidian gut disease common in bee adults (McIvor and Malone 1995, Paxton et al. 2007, Forsgren and Fries 2010, Zheng et al. 2015). We used species-specific primers for *A. apis* and general primers for *Nosema* spp.

We amplified all gene fragments using PCR reactions in a 96-well plate format and all PCR reactions were conducted with positive and negative controls (see Supp. Table S2 for list of primers and PCR conditions). PCR amplification was verified by gel electrophoresis (1.2% agarose gel, 2 µl PCR product mixed with 1 µl SYBR-green solution [1 µl in 250 µl 10× loading dye], 100V for 60 min) and categorized as 'positive' if a band was visible at the expected fragment size and 'negative' if there was no visible band. We conducted two-way contingency analysis on the number of positive and negative results from each PCR with year and tissue as independent variables. For the detected pathogens, we sequenced the DNA fragments using Sanger sequencing technology. Sequences were quality trimmed, aligned, and BLAST searched from the NCBI database for species identity using Geneious v10.1.3.

ddRAD Library Preparation

We developed ddRAD libraries from 95 samples (plus one negative control with no genomic DNA) and assessed the number of reads and SNPs that could be obtained from bees of different ages and between whole bodies or abdomen. We used all samples from 2007 and 2016, all whole body samples from 2012 and 2014, and abdomen samples from 2012 and 2014 in which DNA extraction was successful and had high molecular weight (>500 bp) when possible (see Supp. Table S1 for list of specimens used). We modified previous protocols to prepare the ddRAD libraries (Peterson et al. 2012, Fritz et al. 2016), using a 20 µl aliquot of our DNA extractions. The detailed ddRAD library preparation protocol is provided in Supp. Material SI2.

Because we were working with generally degraded DNA, we developed a specific protocol for these samples to equalize ddRAD libraries downstream starting with 20 µl of DNA extract. To preserve DNA extracts for long term storage, we limited our use to only 20 µl; using higher volumes (>20 µl) for ddRAD preparation may reduce opportunities for future use. Using the DNA concentration data obtained through the Qubit assay above, we first normalized all samples with total DNA >200ng to 200ng in 10 µl (20 ng/µl). However, not all samples had >200ng total DNA. Therefore, samples with total DNA amounts between 100 ng and 200 ng were concentrated to 100 ng in 10 µl, and samples with total DNA amounts between 50 ng and 100 ng were concentrated to 50 ng in 10 µl.

We randomized our 96 samples into six labeled groups of 16 samples and digested all samples (25 µl reaction) with *Eco*RI and *Msp*I (New England Biolabs, Inc., Ipswich, MA). We then ligated a unique barcoded adapter to the *Eco*RI overhang region for each of the 16 samples in each group, while all samples had the same adapter annealed to the *Msp*I end (25 µl reaction). 10 µl of digested DNA was used for single adapter ligation reactions for samples normalized to 200 ng/µl. 10 µl of digested DNA was used for single adapter ligation reactions for samples normalized to 100 ng/µl. Finally, two separate adapter ligations using 10 µl each of digested DNA was used for samples normalized to 50 ng/µl. Then, from the adapter ligation reactions, we pooled 10 µl for samples normalized to 200 ng/µl, 20 µl for samples normalized to 100 ng/µl, and 20 µl from both separate ligation reactions for samples normalized to 50 ng/µl. These steps were taken to create equal representation of all samples' digested and ligated DNA. There were six total pools with 16 samples from each group.

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

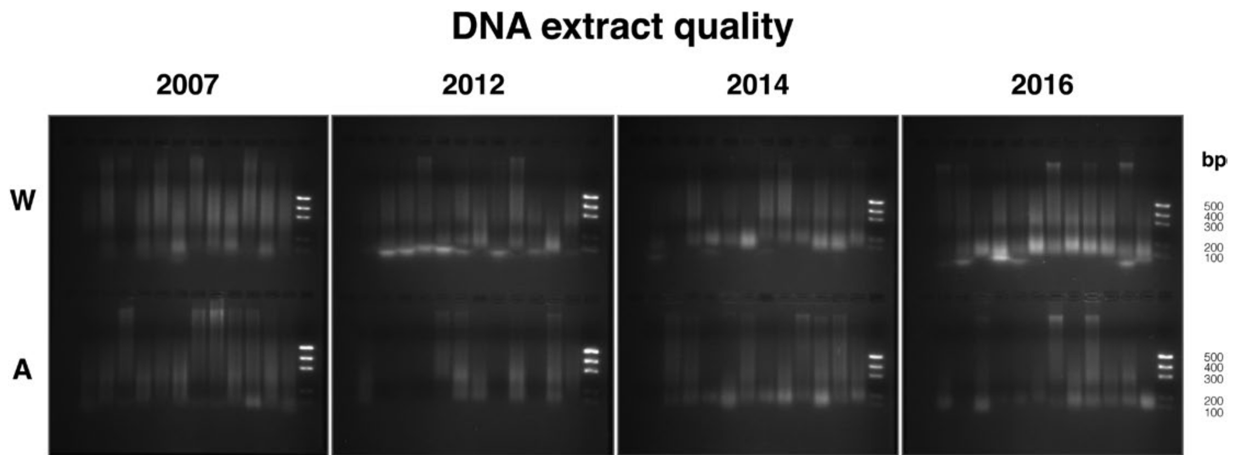

Fig. 3. Gel electrophoresis of DNA extracts of *Eucera pruinosa* museum curated bees from whole bodies (W) and abdomen only (A). Note the large amount of variation from intact high molecular weight to highly degraded DNA that was not specific to any year or tissue type (electrophoresis conditions: 1.2% agarose gel, 5 µl DNA extract mixed with 1 µl SYBR-green solution [1 µl in 250 µl 10× loading dye], 100V for 60 min, 500 bp ladder).

Pooled DNA fragments were size selected with BluePippin (Sage Science, Inc., Beverly, MA) and quality control conducted with High Sensitivity D1000 Screentape on an Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA). Library fragment size was optimized to 250–350 bp to maximize the number of small fragments recovered given the DNA degradation of some of our samples. We amplified the DNA fragments in each pool while attaching Illumina indices (Illumina, Inc., San Diego, CA) in four separate high-fidelity PCR reactions (18 cycles, Phusion DNA polymerase) per pool (to avoid PCR bias of single reactions). Each reaction contained a common forward primer, and six different reverse primers (associated with each pool) containing different Illumina indices. We verified PCR success and checked for contamination with gel electrophoresis. We purified the PCR products with AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN). We verified the range of fragment size, purity, and concentrations of our libraries with Agilent Bioanalyzer. Following the results of the Bionalyzer, we diluted each of the six pools to 20 ul at 10 nM DNA then combined 5 ul of each of our six pools of barcoded and indexed DNA, followed by one final QIAquick PCR Purification (Qiagen, Germantown, MD) eluted in 30 µl H_2O. The final pool was sequenced on Illumina HiSeq 2500 using 100 bp single-end reads.

SNP Calling

Demultiplexing and Filtering

We demultiplexed sequencing reads, aligned contigs, and called SNPs in the program ipyrad v.0.7.11 (Eaton 2014). In steps one and two of ipyrad, we demultiplexed, trimmed, and filtered our sequencing reads with default parameters except for the 'filter_adapters' parameter which we set to '2'; this setting ensures that Illumina adapters, reverse complement of the second cut site, and barcodes are removed. We used 6 separate 'params' files, using a separate barcode file for each.fastq file associated Illumina index. Sequence quality was verified using fastqc and multiqc on all samples. We then merged all data using the merge function, and branched our parameters into two separate workflows. We deemed these parameters 'strict' or 'relaxed' as we used more or less stringent filtering parameters respectively (see Supp. Material SI3A for ipyrad parameter settings). We set the minimum read depth to be called in a consensus sequence (Parameter #12, 'min_depth_maj_rule') to 3 or 6 reads and the clustering threshold for de novo assembly (parameter #14, 'clust_threshold') to 0.85 or 0.95 for the relaxed and strict datasets respectively. For parameter #18, 'max_alleles_consensus', we set 1 for both datasets because we were analyzing haploid males. We used the default settings for the remaining ipyrad parameters. We first conducted clustering analysis (ipyrad step 6) separately by age of specimen, tissue, and collection technique. We compared the effect of these factors on number of quality reads, consensus reads, and clustered loci per sample using ANOVA. Because there were no significant differences in any factor in the library assembly (see Results and Table 2), we followed by clustering all samples together for SNP calling and population analysis.

We verified the presence of endogenous bee DNA for SNP filtering. We converted the .loci output file from ipyrad to fasta format, keeping one representative sequence per locus, using the pyrad2fasta function in the radseq Python script package. We used the Megablast function (which allows for fast searches of large datasets at 95% similarity threshold) in the Geneious v11.1.4 software to search for our >120,000 loci from the strict dataset against the complete NCBI nucleotide database. From the list of matching sequences, we determine the percentage of sequences that were of non-Hymentopteran origin in our dataset.

SNP Filtering

We sorted our ipryrad stats summary output file for both relaxed and strict datasets by the number of 'loci_in_assembly' for each sample. Deemed as 'failed' samples, we created a list of samples that had less than 1,000 loci, in which there was an obvious set of samples that had <1,000 loci (relaxed: 5–869 loci, strict: 2–669 loci) and the rest (relaxed: ≥1,624 loci, strict: ≥1,945 loci). This totaled 13 samples for the relaxed and 12 for the strict datasets. We then used VCFtools (Danecek et al. 2011) for filtering SNP data using the same protocol for both relaxed and strict datasets and removed the failed samples (see Supp. Material SI3B for the annotated script). First, to remove loci and samples with high levels of missing data, we applied two filtering criteria: 1) removing SNPs missing from over 50% of individuals, and 2) removing SNPs that had a minor allele count of less than 3 (which requires alleles to be present in at least three individuals per locus). We set the minimum read depth of genotypes to three reads per genotype. We then printed out the frequency of missing data for each sample using 'missing-indv', and imported the dataset into JMP Pro 13 (SAS Institute, Inc., Cary, NC).

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

We determined the distribution of proportion of missing data per sample. There an obvious increase in missing data per sample (i.e., proportion of total SNP loci that were *not* called in an individual) at the 75th percentile of samples which was 64% missing data for relaxed and 63% for strict data sets. Therefore, all samples that had higher missing data than these values were removed from the datasets. This resulted in 62 samples remaining from our relaxed and 63 samples remaining from our strict datasets. Then, to prepare the SNP dataset for analysis, we then set the minimum minor allele frequency for each locus to 0.05 and only included loci with a mean depth value (over all included individuals) to 20. We removed indels and thinned the dataset to only include loci with one SNP per sequence. Finally, we determined which SNP loci were possibly from contaminant origin by using megablast of the retained sequences and subsequently removed the specific loci from the final .vcf file (provided as Supp. Material SI4).

Structural Equation Model

We used piecewise structural equation modeling (SEM) using the piecewiseSEM package in R (Lefcheck 2015) to analyze the effect of specimen age (1, 3, 5, 10 yr old) and tissue type (whole body or abdomen) on our ability to call quality SNP data from museum specimens. Because of the potential effect of the different collection methods (pan or vane trap) on DNA quality, we also included collection method as a predictor variable in our models (see Results below). We chose SEM because of the stepwise nature of the wetlab and bioinformatic protocol on SNP calling. We grouped our variables into general categories to analyze how methods (age, tissue type, and collection method) influence DNA quality (high molecular weight DNA, 260/280 ratio as a proxy for purity, and concentration). The second set of variables that we tested influenced loci assembly (number of quality reads per sample and number of assembled loci), which would finally affect SNP calling (total SNPs and average depth of reads per sample; Fig. 6). We included our strict versus relaxed parameters for assembly in the model to determine if it affected total SNPs and read depth per sample. Because SEM uses linear regression analyses, we coded binary data for the following variables: tissue type (0-whole body, 1-abdomen), collection (0-vane trap, 1-pan trap), high molecular weight DNA >500 bp (0-no, 1-yes), and parameter strictness (0-relaxed, 1-strict). We report standardized coefficients to compare relative effects of each input variable and R^2 values for each regression.

Estimating Population Structure

To determine if there was population structure among individuals collected from different sites and years, we estimated *F*-statistics using the SNP markers that we generated from *E. pruinosa* ddRAD libraries. We imported our output .vcf file from VCFtools and table of population data into R with the adegenet v2.1.0 package (Jombart 2008). To assess possible genetic differentiation between bees collected in different years from the same locality (Adams county) and between samples collected from counties (Supp. Table S1), we estimated Nei's Gst and Jost's D and generated confidence intervals (CIs) using 100 bootstrap replications in the R packages hierfstat (functions: pairwise.nfst and boot.ppfst; Goudet 2005) and mmod (functions: D_Jost, chao_bootstrap, update_summarise_bootstrap, and pairwise_D; Winter 2012). We also used a multivariate distance approach to estimate population differentiation through a discriminant analysis of principal components (DAPC) using the R package adegenet. For DAPC analysis, we retained 20 principal components and 2 discriminant functions which explained 50% of variance in the data. We used Bayesian Information Criterion analysis (BIC) to determine how many genetic clusters (*k*) were represented among our samples. We used DAPC to visualize probability of group/county membership for each individual bee.

Results

We successfully extracted DNA from all specimens using both whole body and abdomen extraction protocols, however with varying levels of DNA quantity and quality due to degradation. We found minimal physical damage to the bee specimens. Whole bees lost some hair color pigmentation but no other identifiable morphological features were altered (Fig. 2). The only physical damage to bees in the abdomen extraction method was the separation of the metasoma from the rest of the body, yet all identifying morphological features remained intact; and all the tissue in the mesosoma and heads are available for further extractions.

DNA Quality, Quantity and PCR Success

DNA purity, measured by 260/280 and 260/230 ratios, did not significantly differ with year. However, DNA concentration (ng/μl) was significantly different between bees with different ages; surprisingly, the oldest bees, collected in 2007 (10 years ago [ya]), exhibited the highest DNA concentrations (Table 1). We found no significant difference in the percentage of high molecular weight extractions (determined by gel electrophoresis) between samples collected in different years, with and overall 39% of all samples resulting in high molecular weight DNA (Table 1; Fig. 3). The presence of high molecular weight DNA significantly predicted PCR amplification success for the longer fragment 28S (1000 bp; $\chi^2 = 22.7$, $P < 0.001$) while the long CAD fragment (800 bp) did not amplify in any samples. The proportion of extractions leading to positive PCR products did differ by year for amplification of the nuclear genes 28s and short fragment CAD (130 bp). Bees collected in 2007 (10 ya) exhibited the highest percentage of positive amplification results (Table 1; Fig. 4). Bees collected in 2016 (1 ya) had the second highest positive PCR amplifications of 28s, while 2016, 2014, and 2012 bees all had similar results when amplifying the short fragment (130 bp) CAD. For the mitochondrial gene COI, we did not find significant difference between years in the proportion of successful amplification (Fig. 4). The chalkbrood pathogen, *A. apis* was not detected in any samples. We detected *Nosema* spp. pathogen in samples collected across all years. Although there was no significant difference in proportions between years; 5- and 10-yr-old bees showed the lowest number of samples amplifying *Nosema* DNA (Table 1; Fig. 5). We corroborated the identify of *Nosema* spp. among our samples through Sanger sequencing, but we were not able to identify these pathogens at the species level because we only amplified a ~140 bp segment from a conserved SSU rRNA region.

Tissue type had a significant effect on the quantity and quality of DNA extracted. Abdomens resulted in purer DNA (higher 260/280 ratios) but lower DNA quantity than whole body extractions (Table 1). However, there was no difference in percentage of extractions with high molecular weight DNA (Table 1; Fig. 3). Although there were differences in quantity and quality of DNA extracted between abdomen and whole bodies, we did not detect significant differences in the percent of successful PCR amplifications for 28s and CAD (Fig. 4). COI, a longer gene fragment, amplified twice as frequently in whole body extractions than in abdomen extractions (Table 1; Fig. 4).

We detected a difference in the quality and quantity between bees collected in pan traps and blue vane traps. Even though the samples from the pan traps had been collected 10 yr ago, DNA extractions from these bees showed the highest overall purity (260/280 ratios) and quantity of DNA. However, the only difference in PCR amplification success was found in the 2007 pan trapped bees, which amplified for the CAD gene at a higher proportion (100%) than vane trapped bees (83%; Table 1).

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

Table 1. Results of DNA extraction quality, quantity, and PCR amplification for success of *Eucera pruinosa* by the year bee was collected, tissue extracted and collection method

		DNA quality							Gene PCR screening			
		260/280		260/230		ng/µl		High Mol. Wt.	28s	CAD	COI	Nosema
		Mean	Std Err	Mean	Std Err	Mean	Std Err	#Success/*N*	#Amplified/*N*	#Amplified/*N*	#Amplified/*N*	#Amplified/*N*
Year	2007	1.79	0.05	4.41	0.33	32.86	2.75	10/24	16/24	24/24	11/24	7/24
	2012	1.68	0.05	2.4	0.52	15.9	1.67	13/36	11/24	19/24	15/24	8/24
	2014	1.69	0.04	4.1	0.31	12.26	1.67	13/36	8/24	21/24	13/24	14/24
	2016	1.69	0.06	3.85	0.34	23.09	3.03	10/24	15/24	20/24	16/24	11/24
		$F_{3,116} = 0.96$ $P = 0.42$		$F_{3,109} = 1.16$ $P = 0.33$		$^*F_{3,116} = 16.5$ $P < 0.001$		$\chi^2 = 0.38$ $P = 0.95$	$\chi^2 = 6.95$ $P = 0.07$	$^*\chi^2 = 8.06$ $P = 0.05$	$\chi^2 = 2.52$ $P = 0.47$	$\chi^2 = 5.17$ $P = 0.16$
Tissue	Abdomen	1.87	0.02	4.08	0.34	17.31	1.6	28/72	28/48	41/48	37/48	18/48
	Whole	1.46	0.02	2.82	0.15	23.13	2.02	18/48	22/48	43/48	18/48	22/48
		$^*F_{1,118} = 143.8$ $P < 0.001$		$^*F_{1,108} = 53.9$ $P < 0.001$		$^*F_{1,118} = 5.18$ $P = 0.02$		$\chi^2 = 0.02$ $P = 0.88$	$\chi^2 = 1.51$ $P = 0.22$	$\chi^2 = 0.38$ $P = 0.54$	$^*\chi^2 = 15.9$ $P < 0.001$	$\chi^2 = 0.69$ $P = 0.41$
Collection	Pan	1.79	0.05	4.41	0.33	32.86	2.75	10/24	16/24	24/24	11/24	7/24
	Vane	1.69	0.03	3.39	0.26	16.33	1.23	36/96	34/72	60/72	44/72	33/72
		$F_{1,118} = 2.87$ $P = 0.09$		$F_{1,108} = 1.94$ $P = 0.17$		$^*F_{1,118} = 34.4$ $P < 0.001$		$\chi^2 = 0.14$ $P = 0.71$	$\chi^2 = 2.78$ $P = 0.10$	$^*\chi^2 = 7.46$ $P = 0.006$	$\chi^2 = 1.7$ $P = 0.19$	$\chi^2 = 2.12$ $P = 0.15$

Note that Pan collected bees were only from 2007 while Vane trapped bees were from all other years. *F* values are given for ANOVA analyses and χ^2 for contingency analyses. Significant differences are indicated by asterisks (*) at $P < 0.05$.

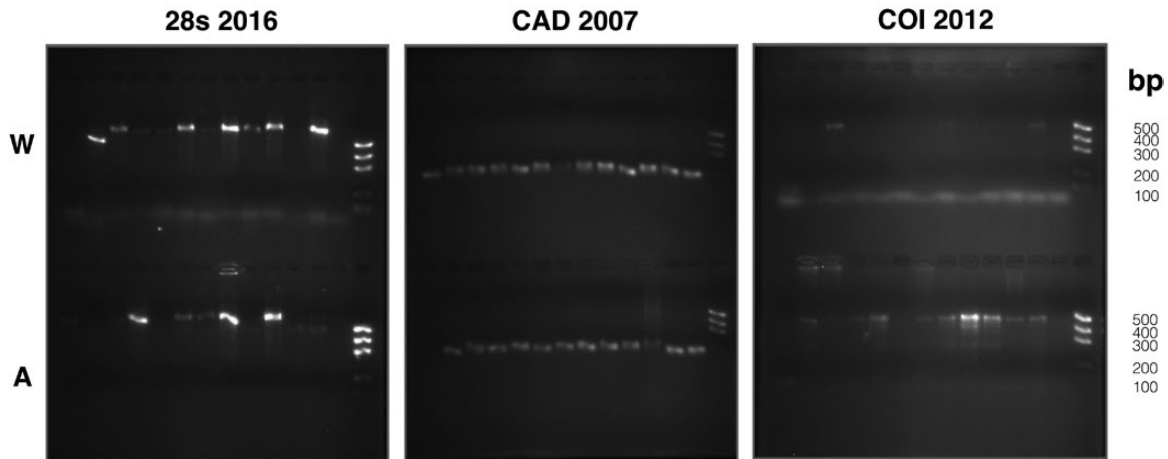

Fig. 4. Examples of results of PCR amplification of 28s, CAD, and COI genes from whole bodies (W) or abdomen only (A) extractions. Note the variability in 28s amplification, consistency in CAD amplification, and difference in efficiency between whole body versus abdomen extractions in COI amplification (electrophoresis conditions: 1.2% agarose gel, 2 µl PCR product mixed with 1 µl SYBR-green solution [1 µl in 250 µl 10× loading dye], 100V for 60 min, 500 bp ladder).

SNP Calling

After initial filtering of failed samples from our libraries, we retained 82 out of 95 samples and recovered 156,429 total SNP sites from the relaxed assembly protocol. We retrieved 83 out of 95 samples and 73,438 total SNP sites from the strict assembly protocol—defined by higher minimum depth and clustering thresholds (Table 2). We determined that our loci were only 1% non-Hymenopteran, indicating our data were mostly free from contaminants prior to SNP filtering. Specimen age, tissue type, nor collection technique had significant effects on loci assembly of our ddRAD libraries (Table 2) therefore we clustered all samples for final SNP calling and analysis.

After SNP filtering, we retained 62 individuals and 849 total SNPs from our relaxed library and 63 individuals and 912 SNPs from our strict library (using a missing data threshold of 64 and 63% for relaxed and strict datasets, respectively). We subsequently removed the loci associated with potential contaminants (bacteria, fungi, mollusk, or plant) which was one locus for the relaxed dataset and three loci for the strict dataset resulting in a final 848 loci and 909 loci for the relaxed and strict datasets, respectively. For our final relaxed dataset, the samples had an average of 87 ± 3% of all SNPs (734 ± 25 SNPs) and our strict dataset had an average of 77 ± 3% of all SNPs (695 ± 26 SNPs), but no samples had all SNPs of the data

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

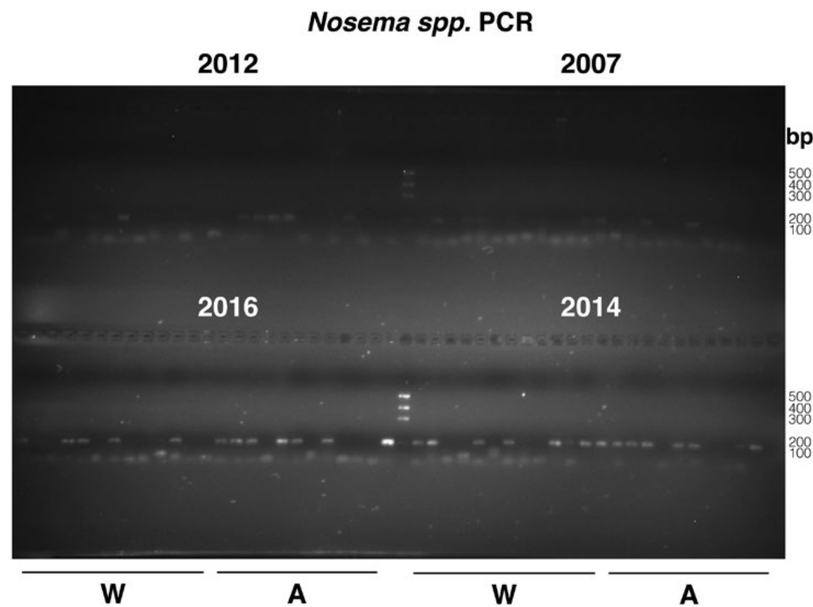

Fig. 5. The results of PCR of *Nosema spp.* from all specimens from each year collected and from whole body (W) versus abdomen only (A) extractions (electrophoresis conditions: 1.2% agarose gel, 2 µl PCR product mixed with 1 µl SYBR-green solution [1 µl in 250 µl 10× loading dye], 100V for 60 min, 500 bp ladder).

matrix. Discarded samples had lower raw reads, quality reads, and total loci in assembly than those that passed through all SNP filters (raw reads: 261,204 ± 303,960 vs. 2,234,554 ± 216,632, $F_{1,94}$ = 28, P < 0.01; quality reads: 254,793 ± 301,461 vs. 2,196,896 ± 214,850, $F_{1,94}$ = 28, P < 0.01, 6,288 ± 2,325 vs. 39,421 ± 1,657, $F_{1,94}$ = 135, P < 0.01), yet had higher DNA concentration than samples that passed SNP filters (29 ± 2 vs. 20 ± 1.7 ng/µl, $F_{1,94}$ = 7.5, P = 0.007).

We found neither specimen age, collection technique, nor collection method resulted in differences in quantity or quality of sequences of ipyrad loci assembly (Table 2). We, therefore, used structural equation modeling to test for relationships (correlation coefficients and R^2) between each step of library assembly and SNP calling. SEM revealed that specimen age, tissue type, and collection technique had small downstream effects on SNP calling (Fig. 6). Age did not influence concentration, purity, or frequency of detecting high molecular weight DNA. Also, tissue type (abdomen vs. whole body) did not affect concentration (r = −0.5, P = 0.63) or high molecular weight DNA (r = 0.15, P = 0.18; Fig. 6). Extractions from abdomen resulted in purer DNA, which was quantified as higher 260/280 ratios (r = 0.78, P < 0.01; Fig. 6). Collection technique had a significant effect on the ratio (r = 0.39, P < 0.01) and concentration (r = 0.75, P < 0.01) of DNA extractions, where pan trapping yielded greater amounts of DNA than vane trapping (Fig. 6).

We were able to call SNP data from museum specimens, independent of age and tissue type. DNA quality (concentration, purity, or high molecular weight) did not positively affect our ability to detect high-quality reads per sample prior to assembly or number of SNPs per sample (loci per sample) post-assembly. The larger number of initial loci called per sample via ipyrad increased the number of retained SNPs after filtering (r = 0.37, P < 0.03) and the number of quality reads per sample positively affected the average depth of each called SNP per sample (r = 0.67, P < 0.01; Fig. 6). When calling SNPs, stricter levels of ipyrad parameters initially increased number of loci retained per individual during ipyrad filtering (r = 0.43, P < 0.01) yet ultimately resulted in fewer SNPs per individual after filtering (r = −0.41, P < 0.01).

Although there were no SNP loci found in all individuals, the average level of missing data per individual was low (relaxed: 13%, strict: 23% missing). Levels of missingness did not differ by sites (relaxed: $F_{2,59}$ = 0.7, P = 0.5; strict: $F_{2,60}$ = 0.5, P = 0.6) suggesting that missingness was distributed randomly throughout the dataset. We did not detect significant genetic differentiation among groups of individuals collected from Adams County in 2012, 2014, and 2016 with Nei's G_{ST} (mean = −0.008; 95% CI = −0.01 to 0.001), yet Jost D was significant (mean = 0.1; 95% CI = 0.095–0.106; Table 3A). We did not detect significant genetic differentiation between different counties with Nei's G_{ST} (mean = −0.001; 95% CI = −0.004 to 0.006); but again small differences were detected with Jost D (mean = 0.061; 95% CI = 0.059–0.063; Table 3B), and highest differentiation between Lancaster and Centre Counties. The significant population differentiation found with Jost D may be the result of this statistic being biased upwards in situations of populations with high migration rates (Whitlock 2011). For the k-means clustering analysis, there was a small numerical change in BIC score at increasing numbers of clusters, though the lowest was at k = 1 (BIC = 431) versus k = 3 (for three counties; BIC = 436). The DAPC scatterplot (Fig. 7A) and group membership probabilities (Fig. 7B) revealed that the population was largely admixed, yet the majority of bees showed highest probability of membership to the county in which they were collected (proportion of successful assignment: Adams = 0.86, Centre = 0.54, Lancaster = 0.62).

Discussion

In this study, we successfully used a minimally destructive technique to access molecular information from non-model museum curated bees. The DNA quality and quantity recovered from these specimens were suitable for PCR amplification of a number of commonly used genes for DNA barcoding, phylogenetic reconstruction and pathogen identification. Most notably, we successfully built a ddRAD library that generated a number of SNPs suitable for estimation of population structure analyses. Our results show consistently high

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

Table 2. ddRAD library summary statistics (mean, SE and sample size [*N*]) arranged by the age of specimen, extraction technique, and collection technique

		Quality reads			# Consensus reads per sample (ipyrad)						# Loci (ipyrad)						# SNPs (post filtering)						Mean depth					
					Relaxed			Strict			Relaxed			Strict			Relaxed			Strict			Relaxed			Strict		
		Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*	Mean	SE	*N*
Age	1	2,050,048	431,228	23	60,135	7,567	23	36,848	5,177	23	22,529	3,084	23	28,318	2,142	23	650	73	16	630	57	16	34	11	16	37	10	16
	3	1,002,484	229,544	24	46,740	7,204	24	27,421	4,519	24	21,916	3,213	24	23,157	3,597	24	755	30	16	693	50	16	30	9	16	30	8	16
	5	1,578,741	292,138	24	67,901	8,204	24	40,939	5,043	24	27,371	2,954	24	33,898	3,842	24	798	20	17	760	41	18	40	9	17	38	8	18
	10	1,560,721	550,431	24	51,604	10,196	24	28,981	5,552	24	17,616	2,936	24	23,197	4,037	24	730	57	13	688	64	13	54	27	13	50	20	13
		$F_{3,94} = 1.17$ $P = 0.33$			$F_{3,94} = 1.25$ $P = 0.3$			$F_{3,94} = 1.61$ $P = 0.19$			$F_{3,94} = 1.73$ $P = 0.17$			$F_{3,94} = 1.74$ $P = 0.17$														
Tissue	Whole body	1,744,535	312,652	48	62,237	6,093	48	37,454	3,700	48	17,302	1,704	48	30,947	3,022	48	762	26	32	696	37	34	44	12	32	42	9	34
	Abdomen	1,336,598	241,426	47	50,758	5,753	47	29,487	3,499	47	16,930	1,938	47	25,105	2,812	47	704	43	30	695	38	29	33	7	30	34	6	29
		$F_{1,94} = 1.06$ $P = 0.30$			$F_{1,94} = 1.87$ $P = 0.17$			$F_{1,94} = 2.44$ $P = 0.12$			$F_{1,94} = 0.02$ $P = 0.89$			$F_{1,94} = 2.0$ $P = 0.16$														
Collection	Vane	1,536,627	192,015	71	58,232	4,492	71	35,044	2,878	71	15,763	1,333	71	29,464	2,376	71	735	28	49	697	29	50	35	6	49	35	5	50
	Pan	1,560,721	550,431	24	51,604	10,196	24	28,981	5,552	24	17,616	2,936	24	23,197	4,037	24	730	57	13	688	64	13	54	27	13	50	20	13
		$F_{1,94} = 0.003$ $P = 0.96$			$F_{1,94} = 0.46$ $P = 0.5$			$F_{1,94} = 1.05$ $P = 0.31$			$F_{1,94} = 0.42$ $P = 0.52$			$F_{1,94} = 1.77$ $P = 0.19$														

Differences in values are presented for comparison between results obtained from relaxed versus strict parameters used in ipyrad. Strict parameters included higher minimum depth and clustering thresholds in library loci assembly. Note that there were no significant differences ($P < 0.05$) between specimen age, extraction technique, or collection technique in any of the ipyrad library assembly output variables. SEM was therefore used to detect relationships between variables for downstream SNP calling (see Results and Fig. 6).

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

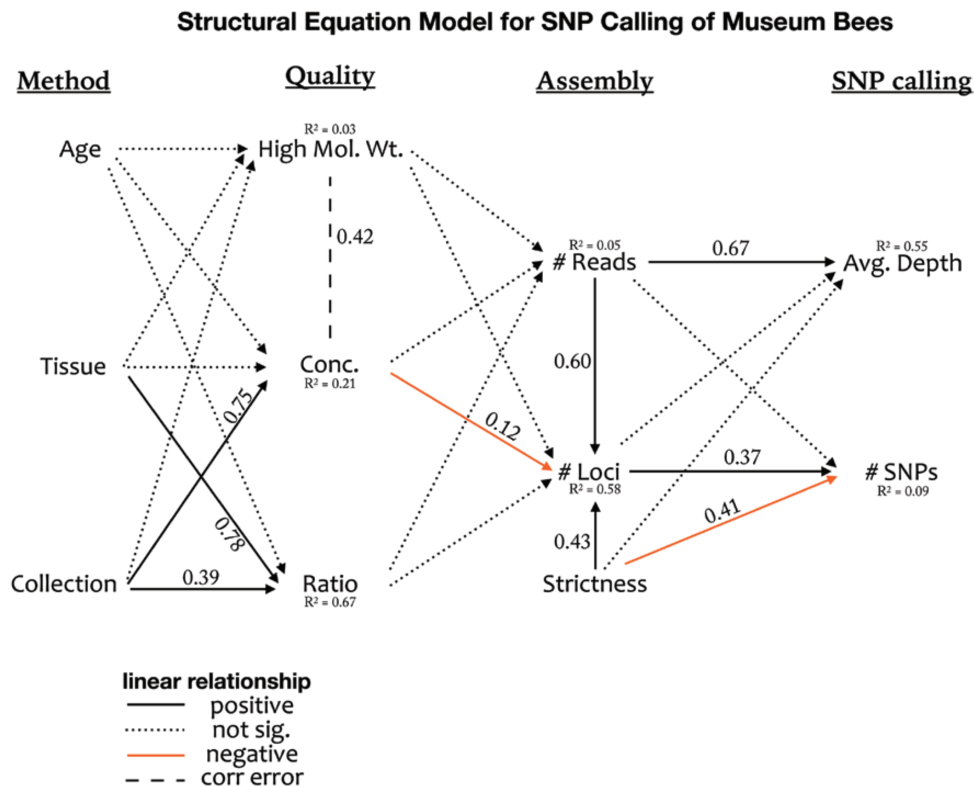

Fig. 6. SEM evaluating the effects of methods on DNA extraction quality and downstream SNP calling. Direction of individual paths are indicated by arrows. Each group of variables ('Method', 'Quality', and 'Assembly') were used in multiple regression equations for the output variables in the following group. Solid black lines indicate significant positive correlations between variables ($P < 0.05$), while solid red lines indicate significant negative correlations ($P < 0.05$). Dotted lines indicate nonsignificant relationships. Standardized coefficients of correlation to indicate strength of effect are provided for significant correlations and R^2 values provided of linear regressions for each output variable. Correlated error indicating bidirectionality between high molecular weight and concentration were included to improve the model.

Table 3. Population structure statistics for museum *Eucera pruinosa* ddRAD SNP data

A	Adams ('12 v '14 v '16)		Global Nei's G_{ST} (CI)			Global Jost's D (CI)	
			−0.008 (−0.01 to 0.001)			0.1 (0.095 to 0.106)	
			Pairwise Nei's G_{ST}			Pairwise Jost's D	
			Adams 2012	Adams 2014		Adams 2012	Adams 2014
		Adams 2014	−0.012		Adams 2014	0.117	
		Adams 2016	−0.016	−0.0037	Adams 2016	0.118	0.067
B	All Counties		Global Nei's G_{ST} (CI)			Global Jost's D (CI)	
			0.001 (−0.004 to 0.006)			0.061 (0.059–0.063)	
			Pairwise Nei's G_{ST}			Pairwise Jost's D	
			Adams	Centre		Adams	Centre
		Centre	−0.0008		Centre	0.054	
		Lancaster	0.004	−0.003	Lancaster	0.053	0.075

Global and pairwise Nei's G_{ST} and Jost's D are reported with 95% CIs after 100 bootstrap replications for global statistics and pairwise statistics between between tested populations. (A) Population structure between bees collected in 2012, 2014, and 2016 in Adams County, Pennsylvania. (B) Population structure between three sampled counties: Adams, Centre, and Lancaster County, Pennsylvania, United States. See Figs. 1 and 7.

DNA molecular weight from approximately 40% of the museum specimens without differences between age of the bees or tissue type used. In addition, we report for the first time the presence of *Nosema* spp. in squash bees (*E. pruinosa*).

Although we expected lower quality DNA and PCR efficiency from older samples, the oldest specimens from 2007 showed the highest yields and success of PCR amplification of 28s and short CAD amplicon. This suggest that one of the strongest predictors of DNA quality from museum bee specimens is collection technique. Specifically, we found that bees collected from pan traps and placed in alcohol within 24 h show better results for DNA-based studies than bees collected in propylene glycol in vane traps, which are usually left outside for up to 7 d. DNA degrades quickly and can become unsuitable for DNA studies soon after the death of invertebrate and vertebrate specimens (Frampton et al. 2008, Graham et al. 2015). However, this degradation can be slow for some insect specimens depending on how samples are collected (Zimmermann et al. 2008). For example, in our study we found that the same proportion

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

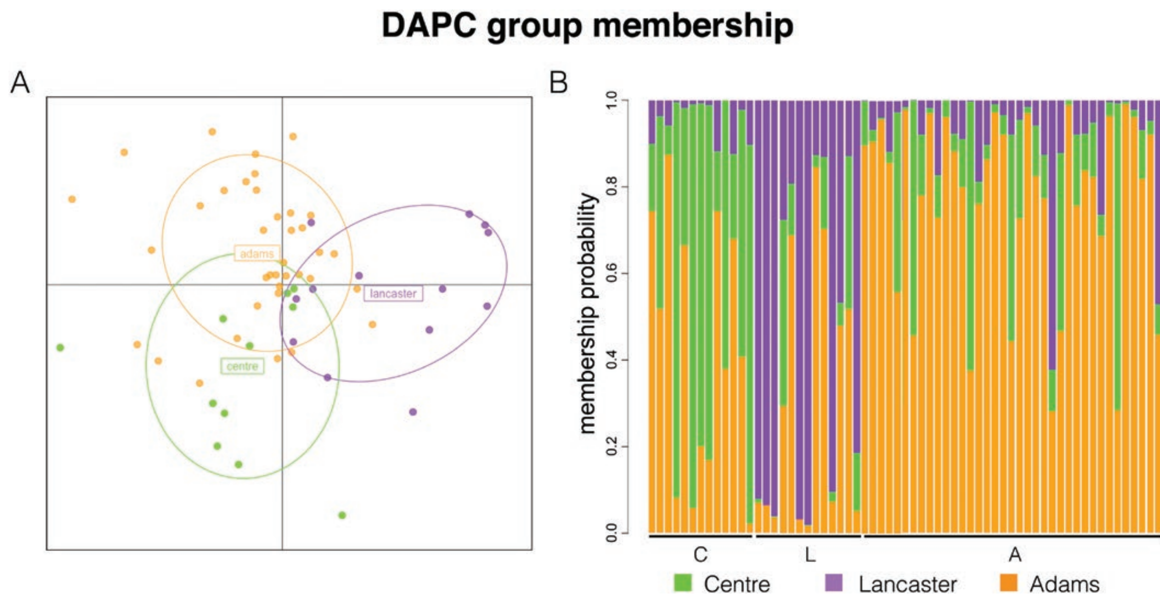

Fig. 7. DAPC visualization of *Eucera pruinosa* museum specimens by county. (A) Scatterplot of group membership. (B) Group membership probability by sample. Colors represent membership probability by County in legend. Bars under samples represent actual location of collection for each specimen (A = Adams, C = Centre, and L = Lancaster Counties).

of individuals of all years exhibited high molecular weight DNA. Even though the specific drivers of this variation in DNA quality among individuals that were collected at the same time and with similar techniques remains to be determined, our results corroborate with recent research investigating the effect of collection method on DNA quality of bees (Ballare et al. unpublished, personal communication). Previous studies have been able to accurately sequence DNA extracted from museum pinned beetles that were more than 100 yr old supporting our findings (Mitchell 2015, Sproul and Maddison 2017). However, the innovation of our study is that, unlike previous studies using insect museum samples, our protocol does not include grinding the whole or parts of the specimen (e.g., legs) for the discovery and genotyping of genomic markers (but see Gilbert et al. 2007 and Cameron et al. 2016 for examples of minimally destructive techniques). This is particularly challenging for most insect species that provide more limited amounts of starting material per specimen (this study: ~10 mg).

We found that presence of high molecular weight DNA quality significantly predicted PCR amplification success for the longer fragment 28S. And, indeed our longer CAD fragment (~800 bp) did not amplify in any samples where our short CAD fragment (~130) amplified in nearly all samples. Because of this effect, Mitchell (2015) used a series of primers for COI at decreasing amplicon lengths to coincide with decreasing fragment sizes associated with specimen age; then used PCR reamplification to recover barcode length sequences. Our evaluation, however, produced a similar proportion of specimens successfully amplifying COI. Other studies have used DNA repairing and enrichment for museum stored specimens before Illumina sequencing followed by bioinformatic recovery of reference genes for phylogenetic reconstructions of beetles (Kanda et al. 2015, Sproul and Maddison 2017). These techniques can be costly but effective for sequencing, however with similar gene fragments (COI, CAD) we were able to extract and amplify our target genes at similar efficiency. The choice of technique for research may depend on costs, type of equipment available and the research goals.

We tested the effectiveness of two minimally destructive and budget-friendly DNA extraction techniques modified from (Gilbert et al. 2007), one where the bee's whole body was not damaged and soaked in extraction buffer and one where only the abdomen was used for extraction (Fig. 2). Neither technique damaged identifying morphological features of the bees. Abdominal extractions produced higher purity samples than whole body extractions, possibly because the enamel of the pins disintegrated into the extraction buffer during incubation. To minimize the impact of contaminants from the pin, we recommend using stainless steel pins for preserving specimens in museums.

Our general recommendation is to use the abdomen extraction protocol for work with museum bee specimens for the following reasons. The quality and purity of the DNA is high, with minimal sacrifice of DNA yield. The slightly lower yield did not negatively affect the ability to amplify DNA fragments, while amplification was higher for COI abdomen extractions. Additionally, the head and thorax of specimens used for abdominal extractions are preserved intact for future genetic studies, while the abdomens are still returned to the collections for additional morphological characterization (Fig. 2). From a laboratory perspective, significantly lower volumes of reagents were used for abdomen only extractions than whole body extractions (Supp. Material SI1). Finally, abdomen extractions avoid the risk of losing pollen associated with the scopae of specimens except for Megachilidae (for which care should be taken to remove and/or identify pollen prior to extractions). Keeping in mind the preservation of DNA in museum specimens, we suggest collections are made with care toward the preservation of genetic material. This includes immediate freezing or dehydration of samples after killing and storage in dark, cool, and dry facilities (Ballare et al. unpublished, personal communication).

One important and recently identified cause of bee decline is the spillover of various pathogens, including *Nosema* spp., from managed to wild bee populations especially from managed honey bees (*Apis mellifera* L. (Hymenoptera: Apidae)) and bumble bees (*Bombus* spp.) (Meeus et al. 2011; Graystock et al. 2013a,b; Cameron et al. 2016). Cameron et al. (2011) analyzed *Nosema* spp. prevalence in different bumble bee species over time using museum stored samples, which highlights the usefulness of DNA markers to

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

identify pathogens in bee specimens from collections. Here, for the first time, we detected *Nosema* in *E. pruinosa* (42% of our samples). The presence of *Nosema* may be a result of pathogen pressure that *E. pruinosa* populations experience in the presence of honey bees and bumble bees used for pollination services. Further research on this topic is necessary to determine the potential role of pathogen spillover from managed bees into these wild bee populations (e.g., Colla et al. 2006; Graystock et al. 2013a, 2016). However, we were not able to accurately identify the specific *Nosema* species because we amplified a small fragment in a conserved region of the SSU rRNA gene. Because detection of pathogens from museum specimens is likely due to an interaction of both DNA quality from extractions and pathogen presence, inferences of actual percentage prevalence in populations must be made with discretion.

DNA quality was not predictive of downstream analyses of our ddRAD libraries regarding ability to obtain quality reads or call SNPs. Even though it seems counterintuitive that the strict assembly protocol resulted in more loci called in ipyrad and total SNPs in our finally libraries, stricter assembly protocols resulted in more SNPs filtered out per sample used for population analyses (Table 2; Fig. 6). Our strict assembly protocol resulted in higher number of total SNPs called in our library possibly because this protocol may have split similar sequences into separate loci because of the higher clustering thresholds. We show that our SNP loci libraries were largely free from contaminants (only ~1% non-hymenopteran DNA in the ipyrad ouput, and only 0.1–0.3% after SNP filtering) which is promising for building de novo ddRAD libraries from museum specimens. Of the potential contaminants in the initial ipyrad loci assembly, the majority were bacteria, fungi, and a small number of loci from a variety of different aquatic organisms. These contaminants can be a concern when genotyping non-targeted genomic markers from museum specimens (such as fungi growing on stored specimens), but can be removed from SNP data sets initially by filtering demultiplexed sequences or consensus loci prior to SNP filtering.

SNP calling from ddRAD data is prone to genotyping errors as a result of PCR artifacts such as allele dropout, or over-representation of homozygotes (Arnold et al. 2013, Schweyen et al. 2014, but also see Ebbert et al. 2016 for assessment of PCR duplicate removal). In our dataset, the presence of genotyping errors from PCR duplicates is low because we sequenced haploid males allowing us to remove heterozygous sites that could be the result of PCR artifacts. We used a published standard lab protocol using 18× PCR cycles for amplifying our double digested gene fragments from fresh samples (Fritz et al. 2016, 2017), which resulted in more than sufficient DNA concentrations for sequencing (70.4 ± 7.2 nM per sample pool). In future studies using diploid females, researchers could lower the number of PCR cycles and use degenerate base regions in sequencing adapters to help detect and remove PCR duplicates that may affect allele frequencies estimation (Schweyen et al. 2014).

An important aspect to consider when building ddRAD library from museum specimens is that the number of individuals that fail may be higher than with freshly collected samples. We discarded the data of one-third of the bee specimens that were originally included in the SNP library because of high levels of missing data. These samples had fewer raw and quality reads than samples that passed all SNP calling filters. This could be the result of high quantities of degraded DNA in these samples, though we could not predict this result from our initial DNA quality screening. Our final SNP data matrix did not contain any SNPs that were shared by all individuals, but the data displayed low mean levels of missingness on a per sample basis (relaxed: 14%, strict: 23% missing). Population structure estimates via *F*-statistics are robust to moderate levels of missing data (Fu 2014, Chattopadhyay et al. 2016, Fritz et al. 2016), which facilitates the use of these techniques with museum samples. Recently, a new technique, hyRAD has been developed to obtain SNP data from museum stored insects and birds (Suchan et al. 2016, Linck et al. 2017). hyRAD uses a double digested DNA library to hybridize and capture orthologous DNA fragments from degraded samples, and uses the original library as a reference sequences for calling SNPs. Although an effective and promising strategy, at the time hyRAD requires the extra cost of laboratory resources and use of fresh collected model organisms to build probes. An advantage of ddRAD analysis is the high-throughput use of de novo sequence assembly that can be used for population-level studies of different species (Peterson et al. 2012, Eaton 2014). Here, we successfully discover and genotyped genome-wide markers (~900 SNPs) for a non-model bee species.

The large number of recent studies involving the collection of numerous bee specimens for pollinator community studies have been stored in insect collections and critically underused given the technological advancements for molecular analyses. In this study, we present a standard protocol to extract DNA from museum curated bees and obtain material to make phylogenetic, disease, and population-based inference on non-model bee species from the past. We reveal that minimally destructive techniques are effective at obtaining important genetic information while preserving specimens for future use. This protocol works for specimens of different ages and methods of collection while preserving the actual specimen in great condition. We hope to encourage new and innovative studies to make deeper use of the wealth of information housed within historical museum specimens.

Supplementary Data

Supplementary data are available at *Insect Systematics and Diversity* online.

Acknowledgments

We would like to thank David Biddinger Penn State's Fruit Research and Extension Center (FREC) and Penn State's Frost Entomological Museum for allowing us to use their specimens for analyses. We thank Nolan Amon for lab assistance throughout the experiment. We thank Alexandra DeYonke for assistance and advice for ddRAD library preparations. We thank Nathaniel Pope for assistance and advice in bioinformatics preparation of SNP data. Finally, we appreciate the members of the López-Uribe lab, Shalene Jha and Kimberly Ballare for critical review of our manuscript.

Data Availability Statement

Data from this study are available from the Dryad Digital Repository: doi:10.5061/dryad.3528sj6 (Vaudo et al. 2018).

References Cited

Alter, S. E., J. Munshi-South, and M. L. J. Stiassny. 2017. Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. Mol. Ecol. 26: 1401–1419.

Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17: 81–92.

Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies. 2013. RADseq underestimates diversity and introduced genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22: 3179–3190.

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

Bartomeus, I., J. S. Ascher, J. Gibbs, B. N. Danforth, D. L. Wagner, S. M. Hedtke, and R. Winfree. 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. Proc. Natl. Acad. Sci. USA 110: 4656–4660.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next generation museum population genomics. Mol. Ecol. 22: 6018–6032.

Brosi, B. J., G. C. Daily, and P. R. Ehrlich. 2007. Bee community shifts with landscape context in a tropical countryside. Ecol. Appl. 17: 418–430.

Cameron, S. A., J. D. Lozier, J. P. Strange, J. B. Koch, N. Cordes, L. F. Solter, and T. L. Griswold. 2011. Patterns of widespread decline in North American bumble bees. Proc. Natl. Acad. Sci. USA 108: 662–667.

Cameron, S. A., H. C. Lim, J. D. Lozier, M. A. Duennes, and R. Thorp. 2016. Test of the invasive pathogen hypothesis of bumble bee decline in North America. Proc. Natl. Acad. Sci. USA 113: 4386–4391.

Carolan, J. C., T. E. Murray, Ú. Fitzpatrick, J. Crossley, H. Schmidt, B. Cederberg, L. McNally, R. J. Paxton, P. H. Williams, and M. J. F. Brown. 2012. Colour patterns do not diagnose species: quantitative evaluation of a DNA barcoded cryptic bumblebee complex. PLoS One 7: e29251.

Chattopadhyay, B., K. M. Garg, A. K. V. Kumar, D. P. S. Doss, F. E. Rheindt, S. Kandula, and U. Ramakrishnan. 2016. Genome-wide data reveal cryptic diversity and genetic introgression in an oriental cynopterine fruit bat radiation. BMC Evol. Biol. 16: 41.

Colla, S. R., M. C. Otterstatter, R. J. Gegear, and J. D. Thomson. 2006. Plight of the bumble bee: pathogen spillover from commercial to wild populations. Biol. Conserv. 129: 461–467.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27: 2156–2158.

Danforth, B. N., S. Cardinal, C. Praz, E. A. B. Almeida, and D. Michez. 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. Annu. Rev. Entomol. 58: 57–78.

Dellicour, S., P. Mardulyn, O. J. Hardy, C. Hardy, S. P. M. Roberts, and N. J. Vereecken. 2013. Inferring the mode of colonization of the rapid range expansion of a solitary bee from multilocus DNA sequence variation. J. Evol. Biol. 27: 116–132.

Dorchin, A., M. M. López-Uribe, C. J. Praz, T. Griswold, and B. N. Danforth. 2018. Phylogeny, new generic-level classification, and historical biogeography of the *Eucera* complex (Hymenoptera: Apidae). Mol. Pylogenet. Evol. 119: 81–92.

Droege, S. 2015. The very handy manual: how to catch and identify bees and manage a collection. ftp://ftpext.usgs.gov/pub/er/md/laurel/Droege/Handy%20Bee%20Manual.pdf (accessed 4 October 2018).

Eaton, D. A. R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 30: 1844–1849.

Ebbert, M. T. W., M. E. Wadsworth, L. A. Staley, K. L. Hoyt, B. Pickett, J. Miller, J. Duce, J. S. K. Kauwe, and P. G. Ridge. 2016. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinf. 17: 491–500.

Forsgren, E., and I. Fries. 2010. Comparative virulence of *Nosema ceranae* and *Nosema apis* in individual European honey bees. Vet. Parasitol. 170: 212–217.

Fortel, L., M. Henry, L. Guilbaud, A. L. Guirao, M. Kuhlmann, H. Mouret, O. Rollin, and B. E. Vaissière. 2014. Decreasing abundance, increasing diversity and changing structure of the wild bee community (Hymenoptera: Anthophila) along an urbanization gradient. PLoS One 9: e104679.

Frampton, M., S. Droege, T. Conrad, S. Prager, and M. H. Richards. 2008. Evaluation of specimen preservatives for DNA analyses of bees. J. Hymenopt. Res. 17: 196–200.

Fritz, M. L., S. Paa, J. Baltzegar, and F. Gould. 2016. Application of a dense genetic map for assessment of genomic responses to selection and inbreeding in *Heliothis virescens*. Insect Mol. Biol. 25: 385–400.

Fritz, M. L., A. M. DeYonke, A. Papanicolaou, S. Micinski, J. Westbrook, F. Gould. 2017. Contemporary evolution of a Lepidopteran species, *Heliothis virescens*, in response to modern agricultural practices. Mol. Ecol. 27: 167–181.

Fu, Y.-B. 2014. Genetic diversity analysis of highly incomplete SNP genotype data with imputations: an empirical assessment. G3 - Genes Genom. Genet. 4: 891–900.

Gezon, Z. J., E. S. Wyman, J. S. Ascher, D. W. Inouye, and R. E. Irwin. 2015. The effect of repeated, lethal sampling on wild bee abundance and diversity. Methods Ecol. Evol. 6: 1044–1054.

Gibbs, J. 2018. DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome. 61: 21–31.

Gilbert, M. T. P., W. Moore, L. Melchior, and M. Worobey. 2007. DNA extraction from dry museum beetles without conferring external morphological damage. PLoS One 2: e272.

Gilliam, M., S. Taber III, B. J. Lorenz, and D. B. Prest. 1988. Factors affecting development of chalkbrood disease in colonies of honey bees, *Apis mellifera*, fed pollen contaminated with *Ascosphaera apis*. J. Invertebr. Pathol. 52: 314–325.

Goudet, J. 2005. Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol. Ecol. Resour. 5: 184–186.

Goulson, D., E. Nicholls, C. Botías, and E. L. Rotheray. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science. 347: 1255957.

Graham, C. F., T. C. Glenn, A. G. McArthur, D. R. Boreham, T. Kieran, S. Lance, R. G. Manzon, J. A. Martino, T. Pierson, S. M. Rogers, et al. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol. Ecol. Resour. 15: 1304–1315.

Grando, C., N. D. Amon, S. Clough, N. Guo, W. Wei, P. Azevedo, M. M. López-Uribe, and M. I. Zucchi. In press. Two colors, one species: the case of *Melissodes nigroaenea* (Apidae: Eucerini), an important pollinator of cotton fields in Brazil. Sociobiology doi:10.13102/sociobiology.v65i4.3464

Graystock, P., K. Yates, B. Darvill, D. Goulson, and W. O. H. Hughes. 2013a. Emerging dangers: deadly effects of an emergent parasite in a new pollinator host. J. Invertebr. Pathol. 114: 114–119.

Graystock, P., K. Yates, S. E. F. Evison, B. Darvill, D. Goulson, and W. O. H. Hughes. 2013b. The Trojan hives: pollinator pathogens, imported and distributed in bumblebee colonies. J. Appl. Ecol. 50: 1207–1215.

Graystock, P., E. J. Blane, Q. S. McFrederick, D. Goulson, and W. O. H. Hughes. 2016. Do managed bees drive parasite spread and emergence in wild bees? Int. J. Parasitol. 5: 64–75.

Hamblin, A. L., E. Youngsteadt, and S. D. Frank. 2018. Wild bee abundance declines with urban warming, regardless of floral density. Urban Ecosystems 21: 419–428.

Hamilton, C. A., A. R. Lemmon, E. M. Lemmon, and J. E. Bond. 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evol. Biol. 16: 212.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, G. E. Maes, E. Diopere, G. R. Carvalho, et al. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol. Ecol. Resour. 11 (Suppl 1): 123–136.

Hogendoorn, K., M. Stevens, and R. Leijs. 2015. DNA barcoding of euryglossine bees and the description of new species of *Euhesma Michener* (Hymenoptera, Colletidae, Euryglossinae). Zookeys. 520: 41–59.

Holmes, M. W., T. T. Hammond, G. O. U. Wogan, R. E. Walsh, K. LaBarbera, E. A. Wommack, F. M. Martins, J. C. Crawford, K. L. Mack, L. M. Bloch, et al. 2016. Natural history collections as windows on evolutionary processes. Mol. Ecol. 25: 864–881.

Hurd, P. D. J., E. G. Linsley, and A. D. Michelbacher. 1974. Ecology of the squash and gourd bee, *Peponapis pruinosa*, on cultivated cucurbits in California (Hymenoptera: Apoidea). Sm. C. Zool. 68: 1–17.

Jensen, A. B., K. Aronstein, J. M. Flores, S. Vojvodic, M. A. Palacio, and M. Spivak. 2013. Standard methods for fungal brood disease research. J. Api. Res. 52: 1–20.

Jombart, T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24: 1403–1405.

Joshi, N. K., T. Leslie, E. G. Rajotte, M. A. Kammerer, M. Otieno, and D. J. Biddinger. 2015. Comparative trapping efficiency to characterize bee abundance, diversity, and community composition in apple orchards. Ann. Entomol. Soc. Am. 108: 785–799.

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018

Kahnt, B., G. A. Montgomery, E. Murray, M. Kuhlmann, A. Pauw, D. Michez, R. J. Paxton, and B. N. Danforth. 2017. Playing with extremes: origins and evolution of exaggerated female forelegs in South African *Rediviva* bees. Mol. Phylogenet. Evol. 115: 95–105.

Kammerer, M. A., D. J. Biddinger, E. G. Rajotte, and D. A. Mortensen. 2016. Local plant diversity across multiple habitats supports a diverse wild bee community in Pennsylvania apple orchards. Environ. Entomol. 45: 32–38.

Kanda, K., J. M. Pflug, J. S. Sproul, M. A. Dasenko, and D. R. Maddison. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. PLoS One 10: e0143929.

Lecocq, T., N. J. Vereecken, D. Michez, S. Dellicour, P. Lhomme, I. Valterová, J.-Y. Rasplus, and P. Rasmont. 2013. Patterns of genetic and reproductive traits differentiation in Mainland vs. Corsican populations of bumblebees. PLoS One 8: e65642.

Lefcheck, J. S. 2015. PiecewiseSEM: piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol. Evol. 7: 573–579.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biol. 61: 727–744.

Linck, E. B., Z. R. Hanna, A. Sellas, and J. P. Dumbacher. 2017. Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. Ecol. Evol. 7: 4755–4767.

López-Uribe, M. M., K. R. Zamudio, C. F. Cardoso, and B. N. Danforth. 2014. Climate, physiological tolerance and sex-biased dispersal shape genetic structure of neotropical orchid bees. Mol. Ecol. 23: 1874–1890.

López-Uribe, M. M., J. H. Cane, R. L. Minckley, and B. N. Danforth. 2016. Crop domestication facilitated rapid geographical expansion of a specialist pollinator, the squash bee *Peponapis pruinosa*. Proc. Biol. Sci. 283: 20160443.

Lozier, J. D., and S. A. Cameron. 2009. Comparative genetic analyses of historical and contemporary collections highlight contrasting demographic histories for the bumble bees *Bombus pensylvanicus* and *B. impatiens* in Illinois. Mol. Ecol. 18: 1875–1886.

Lozier, J. D., J. M. Jackson, M. E. Dillon, and J. P. Strange. 2016. Population genomics of divergence among extreme and intermediate color forms in a polymorphic insect. Ecol. Evol. 6: 1075–1091.

McIvor, C. A., and L. A. Malone. 1995. *Nosema bombi*, a microsporidian pathogen of the bumble bee *Bombus terrestris* (L.). New. Zeal. J. Zool. 22: 25–31.

Meeus, I., M. J. F. Brown, D. C. De Graaf, and G. Smagghe. 2011. Effects of invasive parasites on bumble bee declines. Conserv. Biol. 25: 662–671.

Mitchell, A. 2015. Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. Mol. Ecol. Resour. 15: 1102–1111.

Murray, T. E., Ú. Fitzpatrick, M. J. F. Brown, and R. J. Paxton. 2007. Cryptic species diversity in a widespread bumble bee complex revealed using mitochondrial DNA RFLPs. Conserv. Genet. 9: 653–666.

Nunziata, S. O., S. L. Lance, D. E. Scott, E. M. Lemmon, and D. W. Weisrock. 2017. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. Mol. Ecol. 26: 1060–1074.

Packer, L., and L. Ruz. 2017. DNA barcoding the bees (Hymenoptera: Apoidea) of Chile: species discovery in a reasonably well known bee fauna with the description of a new species of *Lonchopria* (Colletidae). Genome. 60: 414–430.

Paxton, R. J., J. Klee, S. Korpela, and I. Fries. 2007. *Nosema ceranae* has infected *Apis mellifera* in Europe since at least 1998 and may be more virulent than *Nosema apis*. Apidologie. 38: 558–565.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.

Raffini, F., C. Fruciano, P. Franchini, and A. Meyer. 2017. Towards understanding the genetic basis of mouth asymmetry in the scale-eating cichlid *Perissodus microlepis*. Mol. Ecol. 26: 77–91.

Rollin, O., V. Bretagnolle, L. Fortel, L. Guilbaud, and M. Henry. 2015. Habitat, spatial and temporal drivers of diversity patterns in a wild bee assemblage. Biodivers. Conserv. 24: 1195–1214.

Russo, L., M. Park, J. Gibbs, and B. Danforth. 2015. The challenge of accurately documenting bee species richness in agroecosystems: bee diversity in eastern apple orchards. Ecol. Evol. 5: 3531–3540.

Schweyen, H., A. Rozenberg, and F. Leese. 2014. Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. Biol. Bull. 227: 146–160.

Sheffield, C. S., J. Heron, J. Gibbs, T. M. Onuferko, R. Oram, L. Best, N. deSilva, S. Dumesh, A. Pindar, and G. Rowe. 2017. Contribution of DNA barcoding to the study of the bees (Hymenoptera: Apoidea) of Canada: progress to date. The Canadian Entomologist. 10: 1–19.

Sproul, J. S., and D. R. Maddison. 2017. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. Mol. Ecol. Resour. 17: 1183–1201.

Suchan, T., C. Pitteloud, N. S. Gerasimova, A. Kostikova, S. Schmid, N. Arrigo, M. Pajkovic, M. Ronikier, and N. Alvarez. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One 11: e0151651.

Tin, M. M.-Y., E. P. Economo, and A. S. Mikheyev. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. PLoS One 9: e96793.

Vaudo, A. D., M. L. Fritz, and M. M. López-Uribe. 2018. Data from: opening the door to the past: accessing phylogenetic, pathogen, and population data from museum curated bees. Dryad Digital Repository. doi:10.5061/dryad.3528sj6

Whitlock, M. C. 2011. G'_{ST} and D do not replace F_{ST}. Mol. Ecol. 20: 1083–1091.

Williams, P. H., M. J. F. Brown, J. C. Carolan, J. An, D. Goulson, A. M. Aytekin, L. R. Best, A. M. Byvaltsev, B. Cederberg, R. Dawson, et al. 2012. Unveiling cryptic species of the bumblebee subgenus *Bombus s. str.* worldwide with COI barcodes (Hymenoptera: Apidae). Syst. Biodivers. 10: 21–56.

Winfree, R., N. M. Williams, J. Dushoff, and C. Kremen. 2014. Species abundance, not diet breadth, drives the persistence of the most linked pollinators as plant-pollinator networks disassemble. Am. Nat. 183: 600–611.

Winter, D. J. 2012. MMOD: an R library for the calculation of population differentiation statistics. Mol. Ecol. Resour. 12: 1158–1160.

Zheng, H.-Q., H.-R. Gong, S.-K. Huang, A. Sohr, F.-L. Hu, and Y. P. Chen. 2015. Evidence of the synergistic interaction of honey bee pathogens *Nosema ceranae* and deformed wing virus. Vet. Microbiol. 177: 1–6.

Zimmermann, J., M. Hajibabaei, D. C. Blackburn, J. Hanken, E. Cantin, J. Posfai, and T. C. Evans. 2008. DNA damage in preserved specimens and tissue samples: a molecular assessment. Front. Zool. 5: 18.

Downloaded from https://academic.oup.com/isd/article-abstract/2/5/4/5133258 by ESA Member Access user on 17 October 2018