Article:
Bieker, Vanessa C., Sánchez Barreiro, Fátima, Rasmussen, Jacob A. et al. (3 more authors) (2020) Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. Molecular ecology resources. ISSN 1755-098X

https://doi.org/10.1111/1755-0998.13174

Received: 21 October 2019 | Revised: 9 February 2020 | Accepted: 14 April 2020

DOI: 10.1111/1755-0998.13174

SPECIAL ISSUE

MOLECULAR ECOLOGY RESOURCES WILEY

Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community

Vanessa C. Bieker[1] | Fátima Sánchez Barreiro[2] | Jacob A. Rasmussen[1,2] | Marie Brunier[1,3] | Nathan Wales[3,4,5] | Michael D. Martin[1]

[1]Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[2]Section for EvoGenomics, GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]School of Industrial Biology (École de Biologie Industrielle - EBI), Cergy, France

[4]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[5]Department of Archaeology, University of York, York, UK

Correspondence
Michael D. Martin, Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
Email: mike.martin@ntnu.no

Funding information
Funding was provided by an NTNU Onsager Fellowship award to M.D.M.

Abstract

Advances in DNA extraction and next-generation sequencing have made a vast number of historical herbarium specimens available for genomic investigation. These specimens contain not only genomic information from the individual plants themselves, but also from associated microorganisms such as bacteria and fungi. These microorganisms may have colonized the living plant (e.g., pathogens or host-associated commensal taxa) or may result from *postmortem* colonization that may include decomposition processes or contamination during sample handling. Here we characterize the metagenomic profile from shotgun sequencing data from herbarium specimens of two widespread plant species (*Ambrosiaartemisiifolia* and *Arabidopsis thaliana*) collected up to 180 years ago. We used BLAST searching in combination with MEGAN and were able to infer the metagenomic community even from the oldest herbarium sample. Through comparison with contemporary plant collections, we identify three microbial species that are nearly exclusive to herbarium specimens, including the fungus *Alternaria alternata*, which can comprise up to 7% of the total sequencing reads. This species probably colonizes the herbarium specimens during preparation for mounting or during storage. By removing the probable contaminating taxa, we observe a temporal shift in the metagenomic composition of the invasive weed *Am. artemisiifolia*. Our findings demonstrate that it is generally possible to use herbarium specimens for metagenomic analyses, but that the results should be treated with caution, as some of the identified species may be herbarium contaminants rather than representing the natural metagenomic community of the host plant.

KEYWORDS

aDNA, *Ambrosia artemisiifolia*, *Arabidopsis thaliana*, genomics, historical herbarium collections, metagenomics

1 | INTRODUCTION

The world's herbaria house a vast number of plant specimens, ~350 million, some of which are up to 400 years old. Due to advances in DNA extraction and sequencing, especially next-generation sequencing (NGS), it is now feasible to include these specimens in genetic studies. This offers the possibility to use them to track changes over time if samples from different time periods are available (e.g., historical herbarium samples and modern populations from the same location; Bieker & Martin, 2018). The incorporation of historical herbarium samples into genetic analyses enables us to better estimate evolutionary metrics, such as the nucleotide substitution rate, than with modern data alone (Gutaker & Burbano, 2017). In combination with the metadata found on herbarium labels (e.g., sampling location, altitude), changes in the distribution, extinctions in specific areas or introductions of new species can be inferred, which can inform conservation and invasive species control efforts (Nualart et al., 2017) in the face of the profound alteration of many of Earth's ecosystems due to anthropogenic activity (Peñuelas et al., 2013).

Herbarium specimens can provide a wealth of genomic information not only about the specimen itself, but also about microorganisms colonizing its tissues. These microbes originate either through colonization of the living plant (*antemortem*), such as host-associated commensal taxa, through colonization around the time of death (*perimortem*), such as pathogens that killed the plant, or through later (*postmortem*) colonization of the herbarium sheets. Thus, the metagenomic community we can observe today may contain DNA from environmental bacteria involved in decomposition or storage conditions (e.g., bacteria and fungi overgrowth), or contamination from sample handling (e.g., skin microbes) and laboratory sources (e.g., reagents contaminated with enzyme expression vectors; Warinner et al., 2017).

Since the 1940s, chemical pesticides such as herbicides and fungicides have become widely used (Hahn, 2014; Vats, 2015). These chemicals can affect not only crops and their associated microbial communities, but also plants that grow close to fields. Herbarium specimens hold potential to shed more light on the impact of these treatments on the metagenomic composition when specimens pre-dating the common use of pesticides are compared to younger samples (Délye, Deulvot, & Chauvel, 2013). When mounted herbarium specimens include the roots, metagenomic analysis of those organs can also provide insights into pollution levels, because for example the overabundance of nitrogen can alter root-associated bacterial composition (Lang, Willems, Scheepens, & Burbano, 2019).

Degradation of DNA is a well-known effect in historical tissue samples such as those from plant and animal natural history collections. Typical ancient DNA (aDNA) damage includes fragmentation to about 40–500bp in length and varied degrees of hydrolytic deamination (Dabney, Meyer, & Pääbo,2013). The deamination of cytosine to uracil plays a major role, leading to apparent C to T transitions in the downstream analysis (Dabney et al.,2013; Staats et al.,2011). Weiß et al.(2016) found that the DNA in preserved plant specimens decays about six times faster than in moa bones (Allentoft et al.,2012), so DNA fragmentation of herbarium samples, despite their relatively young age (up to 400years), can be comparable to those of animal remains several hundreds or thousands of years old. In addition, the endogenous DNA content of herbarium samples can be relatively low. For example, Gutaker et al (2017) found that as little as 16% of Illumina shotgun sequencing reads derived from herbarium specimens collected between 1839 and 1898 mapped to the *Arabidopsis thaliana* reference genome. For these reasons, Shepherd & Perrie (2014) suggest that the same precautions as for truly ancient samples should be used when working with herbarium samples. These include strict separation of pre- and post-PCR laboratories and performing pre-PCR steps in dedicated clean laboratories (Pääbo et al.,2004).

A limited number of studies have successfully characterized particular members of the metagenomic communities contained in historical herbarium specimens. For example, Malmstrom et al. (2007) were able to detect yellow dwarf viruses in up to 100-year-old herbarium specimens from various California grasses (Poaceae). Miller et al (2016) used real-time PCR to detect the fungal pathogen *Discula destructiva* in up to 66-year-old herbarium specimens of Asian dogwood (*Cornus*). In another collection of studies, domestic potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*) herbarium specimens known to be infected by the oomycete plant pathogen *Phytophthora infestans* were used to analyse single nucleotide polymorphisms (SNPs), simple sequence repeats (SSRs) and complete genome sequences of the pathogen (Martin et al., 2013; Saville, Martin, & Ristaino, 2016; Yoshida et al., 2013). In the only study thus far to compare complete metagenomic communities in herbarium specimens, Schubert et al. (2014) showed that the community can differ substantially between herbarium vouchers of the same species.

Following from previous observations that the endogenous DNA in herbarium specimens can be rapidly lost (or contaminated), we sought to better understand this process by characterizing the nature of the metagenomic community of historical herbarium specimens. We also wished to investigate whether particular microbes are responsible for *postmortem* colonization of mounted plants in herbarium collections. For this we used both previously published and novel shotgun sequencing data from modern and historical collections of *Ambrosia artemisiifolia* (Sánchez Barreiro et al., 2017). *Am. artemisiifolia* is an annual weed that is mostly found in disturbed habitats such as grain and cultivated fields and along roadsides (Bassett & Crompton, 1975; Payne, 1970). It is native to North America and became invasive following its introduction to Europe (Chauvel et al 2006). To further elucidate which taxa are specific to herbarium specimens, we also included in the analysis previously published sequencing data from herbarium and modern specimens of *Ar. thaliana*,an annual plant native to Eurasia and Africa (Durvasula et al., 2017; Exposito-Alonso et al., 2018; 1001 Genomes Consortium, 2016) that colonized North America after its probable introduction from Europe in the early 17th century (Exposito-Alonso et al., 2018).

We compare DNA sequenced from herbarium tissues collected between 1835 and 1993 with that from freshly collected, present-day tissues. We show that it is possible to extract metagenomic profiles from shotgun sequencing data even from the oldest samples. We found significant differences between modern and herbarium sample microbial communities and were able to identify three species as possible ubiquitous contamination of plant tissues in herbarium collections.

2 | MATERIAL AND METHODS

2.1 | Previously published data

For this study, we obtained previously published sequence data from leaf tissue destructively sampled from herbarium-mounted (Exposito-Alonso et al., 2018) and modern (1001 Genomes Consortium, 2016) North American samples of *Arabidopsis thaliana*. All the *Ar. thaliana* herbarium samples had been treated with uracil-DNA glycosylase (UDG) to reduce aDNA damage (Exposito-Alonso et al., 2018). For the modern samples, seeds collected from natural populations in North America between 1993 and 2006 were grown in glasshouses, and leaf tissue was collected for DNA extraction as part of the 1001 Genomes project (1001 Genomes Consortium, 2016). We used a subset of these samples to approximately match the sample size of the herbarium *Ar. thaliana* data set. Samples that were collected in close proximity to the locations where the herbarium specimens came from were preferred and we tried to include samples from the whole North American range of *Ar. thaliana*. In addition, we obtained previously published shotgun sequencing data of leaf tissue from North American herbarium specimens of *Ambrosia artemisiifolia* (Sánchez Barreiro et al., 2017).

2.2 | Newly generated data

We generated new shotgun sequence data from wild-collected modern (North American) specimens of *Am. artemisiifolia* as well as from leaf tissue destructively sampled from a selection of historical herbarium specimens (North American and European). Table S1 presents a complete summary of the plant samples and data sources included in this study.

2.3 | DNA extraction, library preparation and sequencing

In this study, we generated additional sequence data for 43 present-day and historical herbarium-derived N. American *Am. artemisiifolia* samples where leaf tissue had been previously collected and DNAextracted according to the methods described in Martin et al. (2014). Twenty-six of these samples had been previously converted into Illumina NGS libraries in Sánchez Barreiro et al. (2017), so for our study these indexed libraries were simply pooled and Illumina-sequenced (see Table S1 for details about the sequencing method). Seventeen of the present-day samples, along with an extraction blank, were converted into NGS libraries and shotgun-sequenced as described below.

For the previously extracted herbarium samples, blunt-end Illumina library preparation was conducted in dedicated, positively pressurized aDNA laboratories at the NTNU University Museum or the University of Copenhagen using either NEBNext library kit E6070L as in Sánchez Barreiro et al. (2017) or the BEST single-tube protocol (Carøe et al., 2017). Both methods involved the ligation of customized blunt-end adapters (Meyer & Kircher, 2010). Sample-specific, single- or dual-indexing barcodes were incorporated into each library using custom, indexed primers during the indexing PCR (Kircher, Sawyer, & Meyer, 2012). Indexing PCR was carried out in 100-µl reactions with 10–20 µl library template, 0.2 mM each dNTP, 0.2 µM forward primer, 0.2 µM reverse primer, 0.05 U/µl AmpliTaq Gold polymerase, 1 × AmpliTaq Gold Buffer, 2.5 mM $MgCl_2$, 0.4 mg/ml bovine serum albumin, with the remaining reaction volume being completed with molecular biology-grade water. For each library, the optimal number of indexing PCR cycles was estimated using qPCR. Indexing PCR was carried out under the following conditions: 95°C for 10 min, 10–15 cycles of 95°C for 30 s, 60°C for 1 min, 72°C for 45 s and a final extension of 72°C for 5 min. The amplified libraries were purified using either Qiagen QIAquick PCR purification columns with a final incubation for 15 min at 37°C prior to eluting in 32 µl Qiagen EB buffer, or SPRI beads prepared as in Rohland and Reich (2012) with a final incubation for 15 min at 37°C prior to eluting into 30 µl EBT buffer (Qiagen EB buffer with 0.05% Tween-20). These libraries were pooled and Illumina-sequenced (see Table S1 for details of the sequencing method).

For the 17 modern wild-collected *Am. artemisiifolia* DNA samples that had not previously been shotgun-sequenced, the DNA was sheared to a mean length of 500 bp using a Covaris LE220, and blunt-end Illumina library preparation was conducted using the BEST single-tube protocol (Carøe et al., 2017) and customized blunt-end adapters. Sample-specific, dual-indexing barcodes were incorporated into each library using custom, indexed primers during the indexing PCR. Indexing PCR was carried out in 100-µl reactions with 7.5 µl library template, 0.25 mM each dNTP, 0.25 µm forward primer, 0.25 µM reverse primer, 1 µl Herculase II Fusion DNA polymerase, 20 µl 5 × Herculase II Reaction Buffer, and 65.3 µl molecular biology-grade water. For each library, the optimal number of indexing PCR cycles was estimated using qPCR. Indexing PCR was carried out under the following conditions: 95°C for 3 min, 10–18 cycles of 95°C for 20 s, 60°C for 20 s, 72°C for 40 s and a final extension of 72°C for 5 min. The amplified libraries were purified using AMPure XP beads (Beckman Coulter) prior to eluting into 27 µl EB buffer (Qiagen). These libraries were pooled equimolarly and Illumina-sequenced (see Table S1 for details of the sequencing method).

In addition, we collected 10 historical herbarium samples of European *Am. artemisiifolia* from the Naturalis herbarium (NHN) collection, converted them into Illumina libraries along with an

extraction blank, and shotgun-sequenced them (see Table S1 for details of the sequencing method). For these samples, leaf tissue was collected from herbarium sheets using gloves and sterile forceps. Tissue homogenization, DNA extraction and library preparation were performed in a dedicated, positively pressurized aDNA laboratory at the NTNU University Museum. Tissue homogenization was achieved using a Qiagen TissueLyser LT and a combination of stainless steel and tungsten carbide beads that had been sterilized in a 10% bleach solution and then rinsed in molecular biology-grade water. The DNA extraction was conducted as in Martin et al. (2014); in brief, the method involved the use of a Qiagen DNeasy Plant Mini Kit, following the manufacturer's protocols except for the addition of proteinase K (2.2 mg/ml final concentration) during the lysis/incubation step, which was conducted overnight for up to 16 hr.

2.4 | Calculating endogenous content

ADAPTERREMOVAL version 2 (Schubert, Lindgreen, & Orlando, 2016) was used to trim/remove residual adapter contamination from the raw sequence data. To estimate the endogenous content, the PALEOMIX version 1.2.13.4 mapping pipeline (Schubert et al., 2014) was used with BWA version 0.7.15 mem (Heng Li, 2013). For *Ar. thaliana* specimens, reads were mapped against the TAIR10 *Arabidopsis thaliana* reference genome assembly (Arabidopsis Genome Initiative, 2000). For *Am. artemisiifolia* specimens, reads were mapped against an unpublished, 1.37-Gbp *Am. artemisiifolia* draft genome assembly. The endogenous DNA content was estimated as the number of raw alignments to the reference genome divided by the total number of raw sequences retained after trimming and quality filtering, and was calculated by PALEOMIX. To test if the endogenous DNA content differs between herbarium and modern specimens of the same species, R version 3.4.2 was used to perform a Welch two-sample *t* test.

2.5 | Metagenomic profiling with BLAST/MEGAN

For removing host sequences prior to the metagenomic profiling, BWA version 0.7.15 aln (Heng Li & Durbin,2009) was used for mapping with a minimum mapping quality (MAPQ) score of zero. Sequences were mapped against their respective reference genome. Unmapped reads were extracted from the resulting BAM files with SAMTOOLS version 0.1.19 (Li,2011; Li et al.,2009) and converted to FASTA files using the PICARDTOOLS version 2.9.1 (http://broadinstitute.github.io/picard) SAMTOFASTQ tool. If the resulting FASTA file of unmapped reads contained more than 1 million reads, 1 million reads were randomly selected using a custom Python script. Otherwise, all unmapped reads were used for the BLAST (Basic Local Alignment Search Tool) search.

The reads were then blasted against the nonredundant nucleotide database from NCBI using BLASTN (Camacho et al., 2009) with maximum e-value = 0.01 and maximum target sequences = 500. For paired-end data, only reads from read 1 were used for the BLAST search. The BLAST results were imported to MEGAN version 6.11.1 (Huson, Auch, Qi, & Schuster, 2007) using the weighted LCA (lowest common ancestor) algorithm with the following parameters: min score = 100, max expected = 0.01, min percent identity = 0, top percent = 10, min support percent = 0.1, min support = 1, percent to cover = 80. In the MEGAN analysis, hits against Viridiplantae, Metazoa and unclassified taxa, such as "environmental samples<Bacteria>" and "unclassified Bacteria," were ignored. For taxonomic-level assignment, MEGAN used the March 2018 version of the US National Center for Biotechnology Information (NCBI) nucleotide to taxonomy mapping database. For the taxonomic binning, MEGAN used the NCBI taxonomy. The weighted LCA algorithm first assigns each reference sequence a weight based on the number of reads that have hits against that sequence. Each read is then placed on the NCBI taxonomy on the node that is above 75% of the total weight of all the references against which the read has significant hits. As reference sequences in the NCBI database can be associated with several species (up to 1,000), a read may be placed on a higher node even if it only has one significant hit. To compare the different samples, MEGAN's compare function was used, ignoring unassigned reads. Counts were normalized to the smallest count. Bray–Curtis distances were used for a principal coordinates analysis (PCoA). This distance method takes the abundance of species into account (Mitra, Gilbert, Field, & Huson, 2010).

To test if the variance in the PCoA plot could be explained by various sample groupings (modern *Ar. thaliana*, herbarium *Ar. thaliana*, modern *Am. artemisiifolia* and herbarium *Am. artemisiifolia*), the distance matrix was exported from MEGAN and a pairwise PERMANOVA test with 9,990 permutations was performed in R version 3.4.2. For the 20 taxa that explain most of the variation in the PCoA, we performed a Welch two-sample *t* test in R version 3.6.2 to test if some taxa are predominantly found in herbarium specimens. Possible herbarium contaminants were identified as those individual microbial taxa for which the presence/absence data indicated a significant difference between both conspecific sample groups (herbarium *Am. artemisiifolia* vs. modern *Am. artemisiifolia* and herbarium *Ar. thaliana* vs. modern *Ar. thaliana*) and a nonsignificant difference between herbarium samples of *Am. artemisiifolia* and *Ar. thaliana*. These taxa (*Alternaria alternata*, *Alternaria solani* and *Eimeria mitis*) were present in more than 30% of herbarium samples from *Ar. thaliana* and *Am. artemisiifolia* specimens and were absent or nearly absent in modern specimens. The MEGAN analysis was repeated with those taxa disabled.

2.6 | Alignments to reference genomes for important microbial taxa

Because the MEGAN analysis identified *Al. alternata*, *Al. solani* and *E. mitis* as a major component of the metagenomic community, especially in herbarium specimens, we sought to better characterize the sequences originating from these genomes. Thus we used PALEOMIX version 1.2.13.4 and ADAPTERREMOVAL version 2 (Schubert

et al., 2016) to remove adapter contamination and then to directly map all raw reads against the 33.5-Mbp *Al. alternata* (Nguyen, Lewis, Lévesque, & Gräfenhan, 2016), the 33.1-Mbp *Al. solani* (Wolters et al., 2018) and the 72.2-Mbp *E. mitis* (GenBank assembly accession: GCA_000499745.2) reference genomes. The mapping was performed with BWA version 0.7.15 aln (Li & Durbin, 2009) with MAPQ 0. Information about the raw genomic sequencing depth was obtained from the PALEOMIX summary files. MAPDAMAGE version 2.0.8 (Jónsson, Ginolhac, Schubert, Johnson, & Orlando, 2013) was used to obtain frequencies of base misincorporation (damage patterns) for mapped reads. MAPDAMAGE results were plotted only for samples with ≥ 0.2 × mean sequencing depth of the *Al. alternata* or *Al. solani* genome and the highest-depth for the *E. mitis* genome since no depth higher than 0.06 × was obtained for this genome.

3 | RESULTS

The samples derived from herbarium collections contained a significantly lower proportion of endogenous DNA content than modern samples (*Arabidopsis thaliana:* mean herbarium samples: 0.77, mean modern samples: 0.95, $p = 5.889 \times 10^{-6}$; *Ambrosia artemisiifolia:* mean herbarium samples: 0.84, mean modern samples: 0.91, $p = 7.999 \times 10^{-6}$, Figure 1). From the metagenomic profiles obtained for all study samples, a total of 205 different species-level taxa could be identified in the entire data set. Only two of the 20 most common species were found in the extraction blanks: *Eimeria mitis* was present in both blanks and *Methylorubrum extorquens* was only present in the extraction blank performed alongside the herbarium *Am. artemisiifolia* specimens. All species identified in the blanks were found in low abundance except *E. mitis*, which was observed in the extraction blank corresponding to modern *Am. artemisiifolia* specimens.

The most abundant metagenomic species observed in the overall data set is *Albugo laibachii*, a known pathogen of *Ar. thaliana* (Kemen et al., 2011). Indeed, we find this species in all *Ar. thaliana* samples, but also in two herbarium and one modern *Am. artemisiifolia* sample (Figure 2). Most of the differentiation between modern and herbarium samples in the PCoA (Figure 3a) is explained by the presence of *Alternaria alternata*, *E. mitis* and *Al. solani* in herbarium specimens. *E. mitis* is present in a majority of *Ar. thaliana* (19 of 34) and *Am. artemisiifolia* (25 of 46) herbarium samples and is almost absent in modern *Ar. thaliana* (2 of 28) and *Am. artemisiifolia* (4 of 17) samples. *E. mitis* is the only taxon among the most common taxa of the entire data set that was found in both the extraction blanks that were performed alongside the *Am. artemisiifolia* samples. Up to 0.8% of raw reads map against the *E. mitis* reference genome. These sequences show no consensus pattern of aDNA damage (Figure S1). For one sample (MB-104), there is no damage observed in *E. mitis*-aligned reads despite there being damage in reads of that sample aligned to *Am. artemisiifolia*, *Al. alternata* and *Al. solani*. For JK401 and BG-102, there seems to be aDNA damage in *E. mitis* reads. For other samples, the *E. mitis* damage patterns show mostly noise, probably because there are too few aligned reads to resolve a pattern (Warinner et al., 2017). *Al. alternata* was detected in a majority of *Ar. thaliana* (24 of 34) and *Am. artemisiifolia* (28 of 46) herbarium samples, was absent in all 28 modern *Ar. thaliana* samples, and was found only in three modern *Am. artemisiifolia* samples (Figure 2). *Al. alternata* was also present in five out of ten European herbarium samples, whereas *E. mitis* was absent in those samples. The three modern *Am. artemisiifolia* samples that contain *Al. alternata* (MOB-4, MOB-7, MOB-8) are all from the same population, and two of them (MOB-7 and MOB-8) also contain *E. mitis*. *Al. alternata* is found in samples from all ten herbaria included in this study. Between 33.3% and 100% of samples from each herbarium are infected. Herbarium CONN has the lowest infection rate (33.3%), and all samples from CFM and UNC are infected. For the NY herbarium, we analysed specimens from both *Am. artemisiifolia* and *Ar. thaliana* and observed the same infection frequency of 71.4% (Table S2). Mapping against the *Al. alternata* reference genome yielded hits for all samples, including those samples for which the species was not identified by MEGAN. Up to 7.4% of the raw reads mapped against the *Al. alternata* reference genome. For three samples (two *Ar. thaliana* herbarium samples and one *Am. artemisiifolia* herbarium sample), the mean sequencing depth of the *Al. alternata* genome was higher than 1× (1.5×, 2.4× and 6.3×). In addition, base

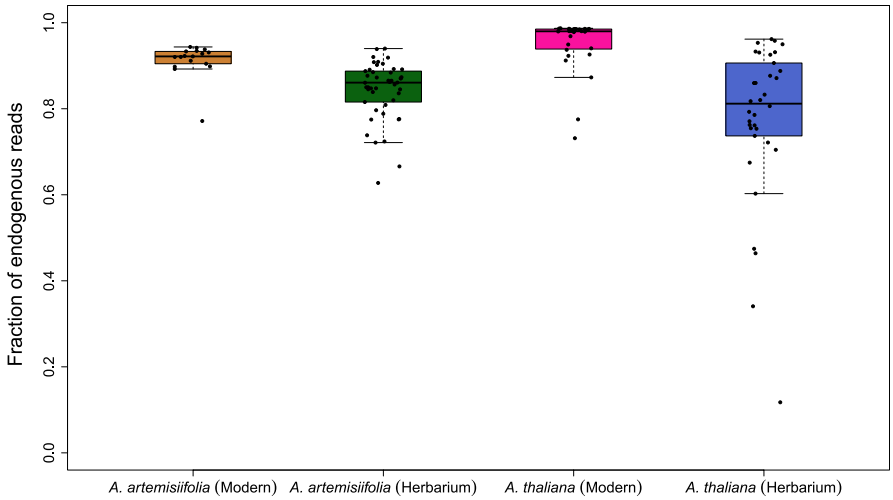


FIGURE 1 Endogenous DNA content (fraction of reads that map against the host reference genome) for herbarium and modern *Ambrosia artemisiifolia* and *Arabidopsis thaliana* samples

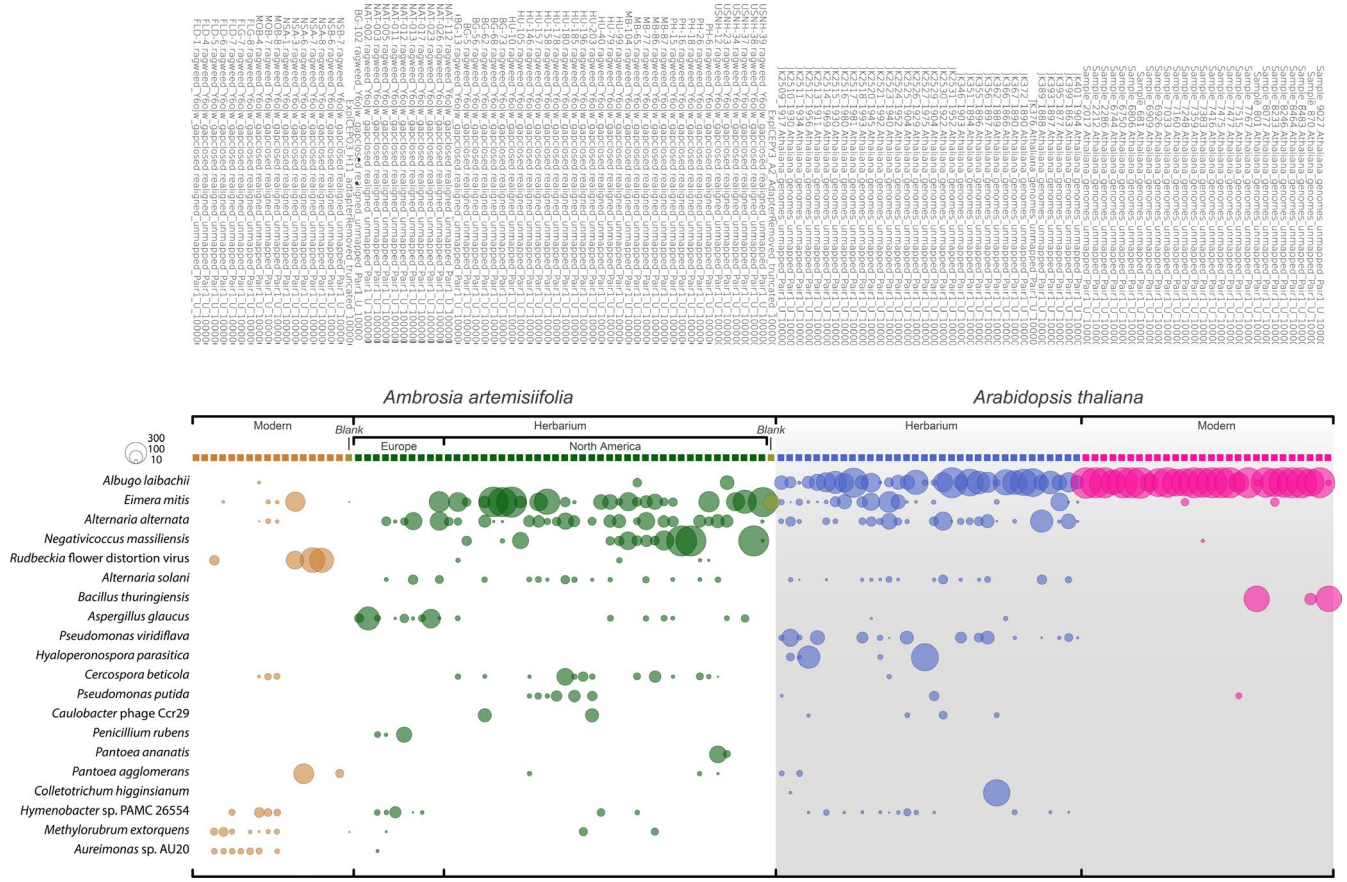


FIGURE 2 Per-sample abundance of most important microbial taxa explaining variation in the PCoA. Circle area scales with the relative abundance of each species in each sample. Species are ordered by their importance for the PCoA, with *Albugo laibachii* accounting for most of the variance. Green: herbarium *Ambrosia artemisiifolia*, orange: modern *Am. artemisiifolia*, blue: herbarium *Arabidopsis thaliana*, pink: modern *Ar. thaliana*, yellow: blank extraction negative controls carried out alongside and for comparison with *Am. artemisiifolia* samples

mismatches in reads mapping to the *Al. alternata* reference genome show a pattern of aDNA damage (Figure 4; Figure S1). *Al. solani* is less abundant than *Al. alternata* and *E. mitis*. It is present in 15 of 46 *Am. artemisiifolia* herbarium samples and 14 of 34 *Ar. thaliana* herbarium samples but is absent from all modern samples. Additionally, it is only present in those samples where *Al. alternata* was identified. Up to 1% of raw reads map against the *Al. solani* reference genome and reads show signs of aDNA damage (Figure S1). Some of the species identified by MEGAN may be host-specific; for example, *Negativicoccus massiliensis*, *Rudbeckia* flower distortion virus (RuFDV) and *Methylorum extorquens* were only identified in *Am. artemisiifolia* samples, whereas *Pseudomonas viridiflava* and *Hyaloperonospora parasitica* were only identified in *Ar. thaliana* samples. Moreover, *N. massiliensis* is only found in herbarium specimens and is less common in specimens from Florida than in specimens collected from other locations in North America and is absent in Europe (Figure 5).

The species-level PCoA of the BLAST/MEGAN analysis shows a clear separation between *Ar. thaliana* and *Am. artemisiifolia* samples, where the modern samples are a subgroup within the herbarium cluster for each sample group (Figure 3a). Pairwise PERMANOVA tests showed significant differences in the metagenomic species composition between all four groups (Figure 6; Table S3). The difference is smaller between herbarium and modern samples from the same species than between species, with herbarium versus modern *Am. artemisiifolia* being more similar than herbarium versus modern *Ar. thaliana*. Herbarium samples of *Am. artemisiifolia* versus *Ar. thaliana* are more similar than any other comparison between species. This difference between the different host species (herbarium/modern *Ar. thaliana* versus herbarium/modern *Am. artemisiifolia*) becomes more pronounced when "herbarium contamination" taxa (*Al. alternata*, *Al. solani*, *E. mitis*) are removed from the analysis (Figure 3b). This also leads to samples from the same species becoming more similar, and in the case of *Ar. thaliana*, the difference between the two groups becomes less significant (Figure 6; Table S3).

4 | DISCUSSION

Principally because it reduces the endogenous DNA content of historical samples, metagenomic "contamination" reduces the achievable sequencing depth-of-coverage of a target host genome. Thus, the metagenomic content of historical samples is a major factor determining which will receive valuable sequencing resources in shotgun sequencing studies of samples from natural history and

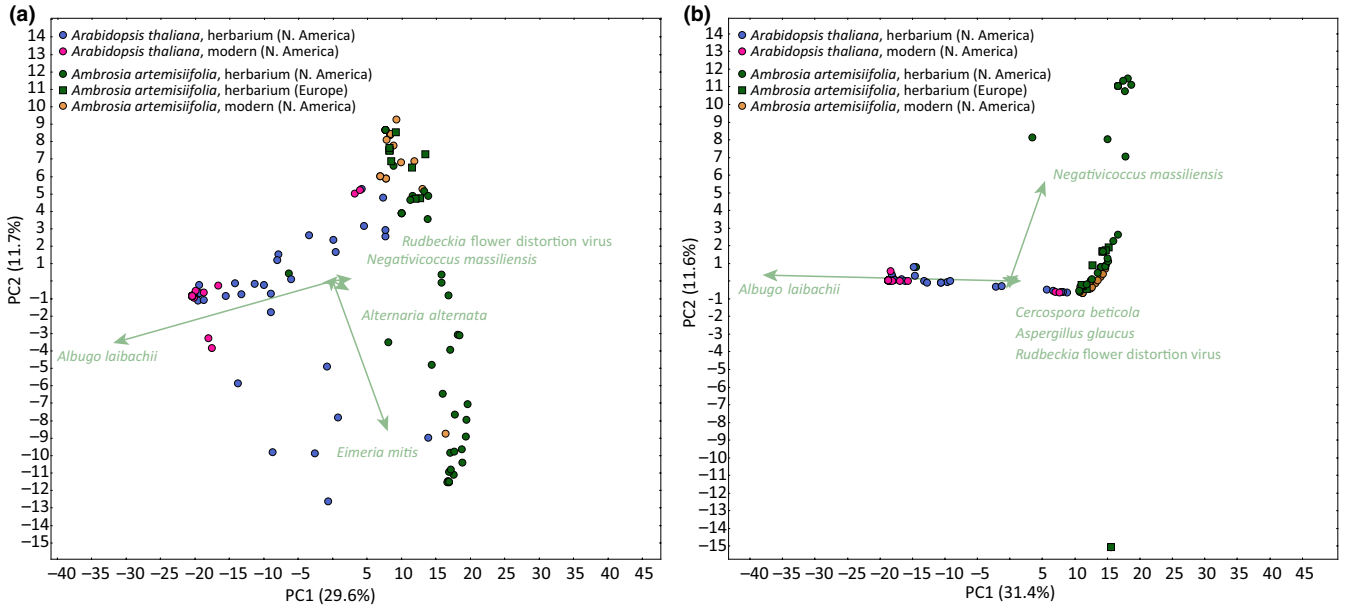
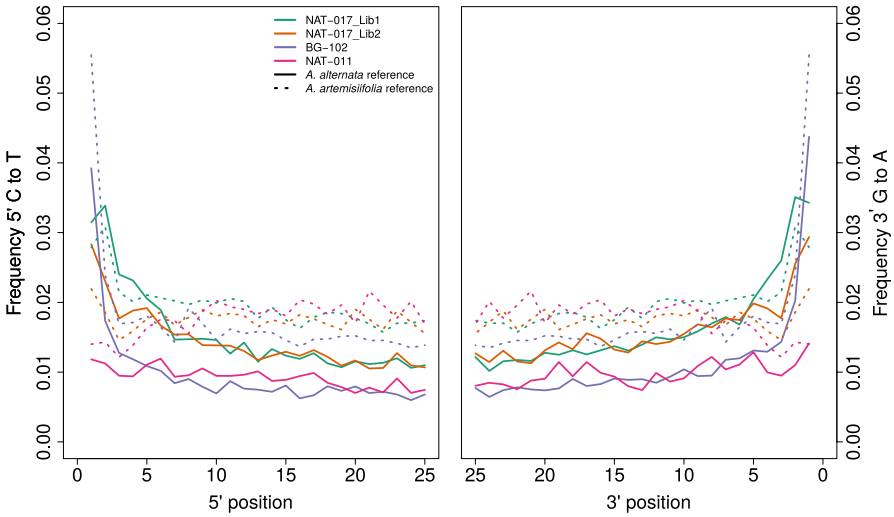


FIGURE 3 Species-level PCoA of the metagenomic composition of modern and historical herbarium plant samples. (a) All identified microbial taxa included. (b) Possible contaminants (*Alternaria alternata*, *Eimera mitis*, *Alternaria solani*) removed from the analysis. The plots are based on Bray–Curtis distances. Green: herbarium *Ambrosia artemisiifolia*, orange: modern *Am. artemisiifolia*, blue: herbarium *Arabidopsis thaliana*, pink: modern *Ar. thaliana*. Spheres indicate samples from North America while points represent samples from Europe. The five microbial species that explain most of the variance are shown as vectors. Vector lengths represent the relative size of the effect

FIGURE 4 Ancient DNA damage (base misincorporation) patterns at fragment ends for *Ambrosia artemisiifolia* herbarium sample metagenomic reads aligned to the *Alternaria alternata* and *Am. artemisiifolia* reference genomes. Left: 5′ C-to-T misincorporations, right: 3′ G-to-A misincorporations. Only libraries with a minimum sequencing depth of 0.2× of the *Al. alternata* genome are shown. Dashed lines show the damage pattern for the *Am. artemisiifolia* reference genome, while solid lines show the damage pattern for the *Al. alternata* reference genome



archaeological collections (Carpenter et al., 2013). In this study, we present a new perspective on the use of herbarium specimens with a comparison of metagenomic composition of shotgun sequencing data from (historical) herbarium and modern plant samples. Our results confirm previous observations that herbarium samples have a lower endogenous DNA content than modern samples. This could indicate that diverse microbes colonize herbarium specimens during preparation or storage, and therefore reduce the proportion of host DNA of the sequenced reads (endogenous content). Indeed, we identify several microbial species that are predominant only in herbarium samples. We find that a particular microbial community develops *postmortem* in plant tissues preserved in herbaria. By removing this "herbarium contamination," the natural metagenomic community of historical specimens can be recovered. Through comparison with contemporary specimens, microbial changes over time can be determined.

The different metagenomic communities we observe between herbarium and modern samples could also arise from differences in sample handling in the laboratory. It is well known that common reagents and extraction kits can be contaminated with microbial DNA (Salter et al., 2014). So far, more than 180 genera have been identified as laboratory contaminants (Glassing, Dowd, Galandiuk, Davis, & Chiodini, 2016). The number and composition of contaminants can differ between extraction kits and even batches of the same

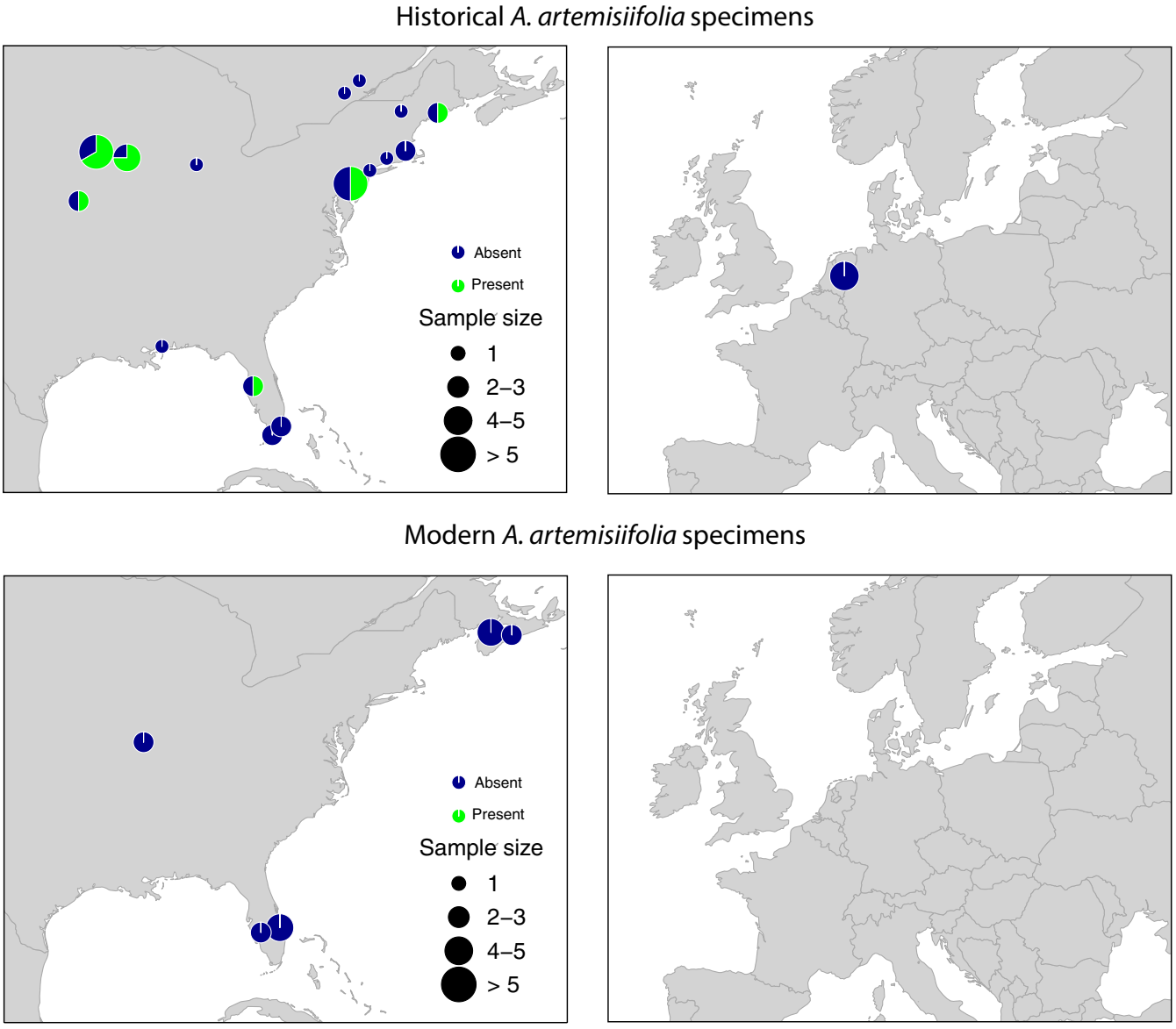


FIGURE 5 Geographical distribution of presence/absence of *Negativicoccus massiliensis* in *Ambrosia artemisiifolia* specimens. Specimens collected within 100 km are grouped together and the centroid of their location is plotted. The frequency of the presence of *N. massiliensis* is shown in green, absent in blue. The size of the circle corresponds to the sample size for each group. Historical herbarium samples are shown in the top panel, while modern samples are shown in the bottom panel

kit. To distinguish the real metagenomic composition from background contamination, it is recommended to sequence an extraction blank (Glassing et al., 2016). From the most abundant species, only *Eimeria mitis* was found in both extraction blanks and *Methylorubrum extorquens* was present in low abundance in one extraction blank. Because only extraction blanks for the *Ambrosia artemisiifolia* specimens were available, we additionally compared the identified species with the list of known laboratory contaminants (Glassing et al., 2016). None of the species that are most abundant in our data set were previously detected as laboratory contamination.

Although the presence of "herbarium contamination" taxa in herbarium but not modern samples might be explained by a change in the genetic diversity of *Am. artemisiifolia* and North American *Arabidopsis thaliana* populations, or by the fact that the modern *Ar. thaliana* samples are not collected from wild populations but grown in a glasshouse, we judge these possibilities unlikely as some of the species that are found in herbarium *Ar. thaliana* samples, but not in modern ones, are also found in herbarium samples of a different species (*Am. artemisiifolia*) and are nearly absent in modern, wild collected *Am. artemisiifolia* samples. Thus, we conclude that most of the abundant species found predominantly in herbarium samples colonized the specimens after collection, during either sample preparation or storage.

Indeed, the third most abundant microbial species in the entire data set, *Alternaria alternata*, is mostly found in herbarium specimens in the MEGAN analysis. *Al. alternata* can live on dead plant material but can also infect living plants and cause several plant diseases (Kimura & Tsuge, 1993). The fungus can also colonize indoor

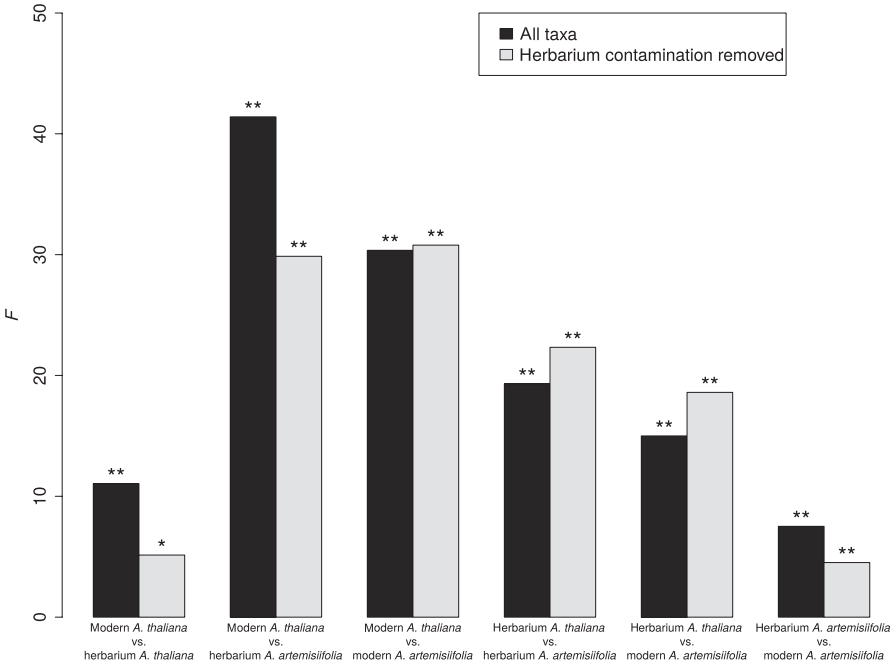


FIGURE 6 *F* value of pairwise PERMANOVA tests with 9,990 permutations between every sample group (modern *Ambrosia artemisiifolia*, herbarium *Am. artemisiifolia*, modern *Arabidopsis thaliana* and herbarium *Ar. thaliana*). Higher values indicate a bigger difference between sample groups. Dark bars show results that include all identified metagenomic taxa, while light bars show results with possible "herbarium contamination" taxa (*Alternaria alternata*, *Alternaria solani*, *Eimeria mitis*) removed from the analysis. Asterisks indicate significance levels: $^{**}p<.01$, $^{*}p<.05$

environments and induce allergic reactions (Gabriel, Postigo, Tomaz, & Martínez, 2016). As the specimens were sampled from ten different herbaria, from both North America and Europe (Table S1), and the absence of *Al. alternata* in herbarium samples is not correlated with source herbarium, the infection of *Ar. thaliana* and *Am. artemisiifolia* herbarium specimens by this fungus seems to be a common occurrence. Mapping against the *Al. alternata* reference genome yielded hits for all samples, including those where the species was not identified by MEGAN. These hits are probably due to nonspecific mapping or mapping of reads to highly conserved regions. We were able to recover three complete genome sequences for *Al. alternata* with up to 6 × coverage that bear the potential for phylogenetic analysis of the fungus. This might inform about when the specimens are infected. If the infection happens in the herbarium, one would expect the same strain to infect all the specimens of one herbarium and differences between herbaria would be expected. If the fungus is already on the plant when they are collected and is able to spread due to the drying process, different strains are expected to be found in the same herbarium and plants collected from the same location are more likely to contain the same strain. In many of the herbarium specimens, the fungus *Al. solani* is found. This pathogenic fungus can cause early blight on potatoes and other solanaceous crops (Chaerani & Voorrips, 2006) and was reported to occur together with *Al. alternata* (Edin, Liljeroth, & Andersson, 2019). Indeed, we find that about half of the specimens infected with *Al. alternata* are also infected with *Al. solani* and *Al. solani* was only found on specimens infected with *Al. alternata*.

Al. alternata is present in specimens from all ten herbaria included in the study and *Al. solani* and *E. mitis* are present in nine out of ten herbaria. Usually, not all specimens of the same herbarium are infected. For the NY herbarium, we have specimens from both *Ar. thaliana* and *Am. artemisiifolia*. The infection rates for *Al. alternata* (71.4%) and *Al. solani* (28.6%) are the same for both host species in this herbarium, indicating that it might be herbarium-specific. For *E. mitis*, the infection rates for *Ar. thaliana* (28.6%) and *Am. artemisiifolia* (85.6%) differ, indicating that the infection rate for this taxon cannot be predicted by the source herbarium. As the samples used in this study were selected for population genetic studies, obviously infected leaves (e.g., those that had dark spots) were avoided and still we find the fungus *Al. alternata* in high abundance in some samples. It is therefore often not possible to know beforehand if a specimen is infected. Knowledge of the infection rate of a given herbarium can therefore help in sample selection for metagenomic studies as heavily infected herbaria can be avoided. To validate if the infection rate is indeed similar across samples from different species from the same herbarium, more samples from different species should be examined.

A possible way to determine if the taxa colonized the herbarium specimens shortly after collection or later during storage is the DNA damage pattern. In older specimens, DNA decay leads to smaller fragment size, and deamination of cytosine to uracil leads to characteristic C to T transitions in the resulting sequences. When reads are mapped to a reference genome, the frequency of these transitions can be measured. If the DNA damage in reads mapping to *Al. alternata* is similar to that of reads mapping to the host specimen (*Ar. thaliana* or *Am. artemisiifolia*), the taxa probably colonized the specimen before or shortly after collection. Significantly lower levels of DNA damage or the absence of DNA damage in microbial reads could indicate that the herbarium specimen was infected more recently (Warinner et al., 2017), during storage in herbaria or handling in the laboratory. It is also possible that due to other factors such as the concentration of nucleases in the cells, the DNA of the fungus and the host specimen age differently. Therefore, lower levels of DNA damage not necessarily result from later infection during storage, but the presence of DNA damage in the host and the absence in the fungus indicates a

more recent infection. This analysis is only possible if there is observable aDNA damage in reads of the host specimen. Due to the age of the herbarium specimen and the conservation and storage methods used, there might be low to no damage observable. The damage pattern for the *Ar. thaliana* samples are generally lower than for *Am. artemisiifolia* specimens because all *Ar. thaliana* herbarium samples were treated with UDG, which reduces the DNA damage prior to sequencing (Exposito-Alonso et al., 2018). In both *Ar. thaliana* and *Am. artemisiifolia* samples, *Al. alternata* and *Al. solani* damage patterns were observed, indicating that the infection with these fungi probably occurred during or shortly after collection. *Al. alternata* is also found in three modern *Am. artemisiifolia* specimens that were collected into silica gel from the same wild population on the same day. It is possible that too little silica gel was used, resulting in slower drying of the plant material. This would indicate that the fungus is already on the plants when they are collected and is able to spread during the relatively slow drying process that is typical in the preparation of herbarium voucher specimens.

Interestingly, the second most abundant species, *E. mitis*, is a pathogen of chickens (Blake, 2015). Species of the genus *Eimeria* can infect diverse animal hosts, including cattle and goats (Chartier & Paraud, 2012). This species is primarily found in herbarium specimens (54% herbarium *Am. artemisiifolia*, 55% herbarium *Ar. thaliana*, 23% modern *Am. artemisiifolia*, and 7% modern *Ar. thaliana*). It is surprising to find an animal pathogen in such high abundance in plant specimens. To our knowledge, neither *Am. artemisiifolia* nor *Ar. thaliana* have been reported to be infected with *Eimeria* species. It is possible that these plants act as a transmitter of the pathogen. However, no clear damage pattern could be observed. For one specimen, there is no damage observed in *E. mitis* reads despite there being damage in reads mapping against the host reference genome. This indicates that *E. mitis* infected this specimen later and was not on the plant by the time of collection. For two other samples, aDNA damage for *E. mitis* reads could be observed. This indicates that *E. mitis* might be present on some specimens by the time of collection or infected them shortly afterwards. In general, due to low numbers of reads mapping against the *E. mitis* read, the aDNA damage could not be accurately estimated as several thousand reads are usually needed (Warinner et al., 2017). However, the fact that this species was also found in the extraction blank indicates it is more likely a laboratory reagent contamination in most of the samples. Herbarium samples are usually more prone to laboratory contamination as they have low DNA concentrations. Therefore, small contaminations are having a greater effect than in modern samples with high DNA concentration. More sequencing effort per sample would be needed to determine if *E. mitis* is rather a laboratory contamination or a herbarum contamination or a combination of both.

Hymenobacter sp. PAMC26554 is common in herbarium samples (in seven *Am. artemisiifolia* and 13 *Ar. thaliana* samples), but is also found in four modern *Am. artemisiifolia* samples. This strain has been isolated from Antarctic lichens and is UV-radiation-resistant and adapted to cold climate (Oh, Han, Ahn, Park, & Kim, 2016). Therefore, this species can probably survive common measures against contamination in herbarium collections, including freezing specimens before archival in the collections. It is also possible that UV-radiation-resistant bacteria survive in dedicated clean laboratory facilities used to handle ancient and historical materials, as these facilities are commonly sterilized using UV light. As the herbarium samples used in this study were processed in three different facilities, and given that these species are not present in all samples and are absent from the extraction blanks we processed, we propose rather that the herbarium specimens themselves are infected.

In addition to *Al. alternata*, other taxa identified in most of the herbarium specimens, while being entirely absent from the modern samples, are *Al. solani* (in 15 *Am. artemisiifolia* and 15 *Ar. thaliana* samples), *Aspergillus glaucus* (in 16 *Am. artemisiifolia* and two *Ar. thaliana* samples), *Negativoccocus massiliensis* (in 14 *Am. artemisiifolia* samples) and *Pseudomonas viridiflava* (in 16 *Ar. thaliana* samples). Because some of these species were identified in herbarium specimens from different host plants (*Ar. thaliana* and *Am. artemisiifolia*), they seem to be specific to herbaria rather than specific to species. Future research comparing the metagenomic communities of herbarium versus modern specimens in a broad panel of plants will be crucial to verifying these findings. Moreover, in future studies the DNA damage pattern of reads mapping to the host plant genome should be compared with those mapping to the herbarium-specific microbial genomes. This may help to determine when the infections developed (e.g., during drying of plant material, during long-term storage or during sample processing in the laboratory). Ultimately, this could inform efforts to prevent these infections or contaminations in the future, which would therefore enable more faithful preservation of these valuable specimens, along with their genomic and metagenomic contents.

The most abundant species in the data set, *Albugo laibachii*, is found in all *Ar. thaliana* herbarium and modern specimens, in two *Am. artemisiifolia* herbarium specimens, and one modern *Am. artemisiifolia* specimen. *Albugo laibachii* is an oomycete and an obligate biotrophic pathogen to *Ar. thaliana* (Thines et al., 2009). *Am. artemisiifolia* is not known to be a host of *Albugo laibachii*, but can be infected by the closely related species *Albugo tragopogonis* (Gerber et al., 2011), which is not represented in the database used for the BLAST search. By removing the originating species from the database in their BLAST search, Warinner et al. (2017) found that reads are then assigned to the closest relative of that species. It is thus possible that reads originated from *Albugo tragopogonis* were falsely assigned to *Albugo laibachii*, the closest related species in the database.

Pseudomonas viridiflava is the most common pathogen of *Ar. thaliana* populations in North America (Jakob et al., 2002) and was indeed found in 47% of the *Ar. thaliana* herbarium specimens in this study, but not in the modern samples. As the pathogen was previously found in modern *Ar. thaliana* specimens, and because the severity of *P. viridiflava* infection is highly related to environmental factors (Everett & Henshall, 1994), the absence of this species in modern *Ar. thaliana* samples in this study probably results from

these specimens being cultivated in glasshouses rather than growing in the wild.

Removal of possible "herbarium contamination" taxa (*Al. alternata*, *Al. solani* and *E. mitis*) from the analysis brought the metagenomic communities of conspecific herbarium and modern samples closer to each other (Figure 3) and increased the differences between species, except for modern *Ar. thaliana* and herbarium *Am. artemisiifolia*, where the differences decrease (Figure 6). This indicates that the original host microbiome can be revealed from herbarium specimens when contaminating taxa are known, and therefore presenting the possibility to study microbial changes over time. Indeed, we find some species that are more common in either modern or herbarium specimens. For *Ar. thaliana*, the bacterium *P. viridiflava* is only found in herbarium specimens (16/34) and absent in all modern specimens, whereas *Bacillus thuringiensis* is only found in modern specimens (3/28). As the modern *Ar. thaliana* specimens were actually grown in glasshouses, these differences probably do not indicate real shifts of the metagenomic composition in wild populations. Moreover, the presence of *B. thuringiensis* only in modern specimens could be a glasshouse contamination and might not be present on wild specimens. To further investigate this, future studies should compare the metagenomic composition of herbarium specimens with wild collected specimens.

For *Am. artemisiifolia*, the bacterium *N. massiliensis* is only found in North American herbarium specimens (14/46) and is more common in specimens from the east coast of the USA (Figure 5). It is only found in four individuals collected from the Midwest, and in one individual from the Southeast. Notably, it was not detected in any of the European herbarium specimens. The bacterium was identified in specimens from five different herbaria and except for the MO herbarium, not all samples from each herbarium are infected. We therefore believe that *N. massiliensis* is not a herbarium contamination and the geographical pattern we observe is due to a real difference in the metagenomic composition. The fungal plant pathogen *Cercospora beticola* is more common in herbarium specimens (13/46) than in modern specimens (3/17) and is also absent in European specimens. Interestingly, the three modern samples containing this pathogen are the same as where *Al. alternata* was found. The geographical distribution of *N. massiliensis* presence/absence in North America, as well as its complete absence in introduced historical populations in Europe, could indicate that the source population is more likely the western or southeastern population than the eastern population. It is also possible that the complete metagenomic community was not transmitted during *Am. artemisiifolia's* introduction to Europe, as the primary introduction vector is seeds (e.g., contaminated birdseed) and not whole plants (Chauvel et al., 2006).

The bacteria *Aureimonas* sp. AU20 and *M.extorquens* are more common in modern specimens (7/17 for both taxa) than in herbarium specimens (1/46 for *Aureimonas* sp. AU20 and 3/46 for *M. extorquens*). Both are found in only one herbarium specimen from Europe. The shift in the metagenomic composition of *Am. artemisiifolia* specimens through time could indicate genetic changes that increase the resistance to species such as *N. massiliensis* and *C. beticola*. Another factor could be the near-ubiquitous use of pesticides and fungicides in modern times. *Am. artemisiifolia* is usually found in disturbed habitats such as next to roads and abandoned and actively used agricultural fields (Bassett & Crompton, 1975; Payne, 1970) and could therefore be exposed to pesticides used in crop production. All *Am. artemisiifolia* herbarium samples were collected before 1940 and therefore pre-date the introduction of chemical herbicides and fungicides. These findings highlight the possibility of extracting at least part of the natural microbiome from shotgun sequences of herbarium specimens. By including younger herbarium samples and linking the presence of certain microbial taxa to host genetic variation, further studies could elucidate what causes temporal changes in the microbial communities associated with wild plant populations.

To reduce contamination of herbarium samples and laboratory reagents with modern DNA, they were extracted in a clean room facility that is subject to extra precautions (Exposito-Alonso et al., 2018). As no specific precautions prior to DNA extraction are taken in herbaria to prevent contamination with human skin and skin microbes (e.g., the herbarium sheets are usually touched without gloves), it is surprising that no human skin microbes such as *Staphylococcus epidermidis* were detected in herbarium specimens. It is possible that these microbes cannot survive on dead plant material, especially when the herbarium sheets were chemically treated for preservation, and are therefore at such low abundances that they could not be detected using our methods.

To ensure that herbarium specimens can be used for metagenomic studies, we suggest the creation of a database of common herbarium contaminants that are found across species and herbaria. This would make it possible to exclude these contaminants from the analysis and observe real shifts in the metagenomic makeup. To get a better understanding of the infection of *Al. alternata* and *Al. solani*, we suggest performing experiments with freshly collected specimens from several species where part of the plant is dried on silica gel and another part is preserved with traditional herbarium methods. It would thus be possible to test if the presence of these metagenomic taxa is due to the preservation method.

Our study is based on leaf material, but could be expanded to other plant parts (e.g., stem or roots). In that case, a similar approach should be used to determine possible "herbarium contamination," as these contaminants could vary between different parts of the plant. We also expect to find different metagenomic communities on other plant parts (see Busby et al., 2017). For example, root samples contain rhizome soil bacteria that are not found on leaves. Herbarium specimens often also contain roots, enabling the comparison of metagenomic root communities over time. This could give insights into, for example, the effects of historical changes in pollution levels (Lang et al., 2019), pesticide use and fertilization practices.

We have shown that it is possible to obtain the metagenomic composition from genome-skimming data sets based on herbarium specimens that were more than 180 years old. As some of the identified species were exclusive to herbarium specimens, found in both tested species and in herbaria from two different continents,

we conclude that they have colonized the specimens during sample preservation or storage. Future studies should include herbarium specimens from additional plant species and compare them to modern wild-collected specimens to investigate the frequency of these species as herbarium specimen contaminants. Investigations of the DNA damage pattern may elucidate when the specimens were colonized, which could inform efforts to preserve specimens in modern herbaria. We suggest that future metagenomic studies on herbarium plant material should consider that some of the identified species may not belong to the natural metagenomic community of the host plant, and thus should be excluded from further analysis and biological interpretation.

ACKNOWLEDGMENTS

We thank our colleagues A. Frisch, M. Bendiksby and M. Nygård for their helpful comments as the study progressed. We thank F. Vieira for his technical help in the bioinformatic analysis. We thank A. L. Kolstad for his help with the statistical analysis. For their generous contributions of tissues destructively sampled for this work, we also thank the herbarium staff and curators at the National Herbarium of the Netherlands (NHN), the Herbarium of the National Museum of Natural History (US), the Herbarium of the Academy of Natural Sciences (PH), the New York Botanical Garden Herbarium (NY), the Harvard University Herbaria (A) and the Herbarium of the Missouri Botanical Garden (MO). Illumina HiSeq4000 sequencing was performed by the NTNU Genomics Core Facility (GCF), which is funded by the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), and the Central Norway Regional Health Authority. We thank three anonymous reviewers for their useful comments which greatly improved the manuscript.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

M.D.M. conceived of the study. F.S.B., N.W., V.C.B., J.A.R., M.D.M. and M.B. performed the sampling and laboratory work. V.C.B. performed the computational analysis. V.C..B wrote the manuscript with input from all authors.

DATA AVAILABILITY STATEMENT

DNA sequences generated for this study can be found under ENA study PRJEB34825. Previously published data for *Am. artemisiifolia* specimens can be found under the ENA study PRJNA339123. Previously published data for modern *Ar. thaliana* specimens can be found under ENA study PRJNA273563. Previously published data for herbarium *Ar. thaliana* specimens can be found under ENA study PRJEB24619. A complete list of accession codes for each sample used in this study can be found in Table S1.

ORCID

Vanessa C. Bieker https://orcid.org/0000-0002-2061-9041
Nathan Wales https://orcid.org/0000-0003-0359-8450
Michael D. Martin https://orcid.org/0000-0002-2010-5139

REFERENCES

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., … Bunce, M. (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proceedings. Biological Sciences / the Royal Society*, *279*(1748), 4724–4733.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, *408*(6814), 796–815.

Bassett, I. J., & Crompton, C. W. (1975). THE BIOLOGY OF CANADIAN WEEDS.: 11. *Ambrosia artemisiifolia* L. and *A. psilostachya* DC. *Canadian Journal of Plant Science. Revue Canadienne De Phytotechnie*, *55*(2), 463–476.

Bieker, V. C., & Martin, M. D. (2018). Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters*, https://doi.org/10.1080/23818107.2018.1458651

Blake, D. P. (2015). *Eimeria* genomics: Where are we now and where are we going?*Veterinary Parasitology*, *212*(1–2), 68–74. https://doi.org/10.1016/j.vetpar.2015.05.007

Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., … Dangl, J. L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biology*, *15*(3), e2001793. https://doi.org/10.1371/journal.pbio.2001793

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M.-H.-H.-S., Samaniego, J. A., … Gilbert, M. T. P. (2017). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution / British Ecological Society*, https://doi.org/10.1111/2041-210X.12871

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American Journal of Human Genetics*, *93*(5), 852–864. https://doi.org/10.1016/j.ajhg.2013.10.002

Chaerani, R., & Voorrips, R. E. (2006). Tomato early blight (*Alternaria solani*): The pathogen, genetics, and breeding for resistance. *Journal of General Plant Pathology: JGPP*, *72*(6), 335–347. https://doi.org/10.1007/s10327-006-0299-3

Chartier, C., & Paraud, C. (2012). Coccidiosis due to *Eimeria* in sheep and goats, a review. *Small Ruminant Research: The Journal of the International Goat Association*, *103*(1), 84–92. https://doi.org/10.1016/j.smallrumres.2011.10.022

Chauvel, B., Dessaint, F., Cardinal-Legrand, C., & Bretagnolle, F. (2006). The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. *Journal of Biogeography*, *33*(4), 665–673.

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7), https://doi.org/10.1101/cshperspect.a012567

Délye, C., Deulvot, C., & Chauvel, B. (2013). DNA analysis of herbarium specimens of the grass weed *Alopecurus myosuroides* reveals herbicide resistance pre-dated herbicides. *PLoS ONE*, *8*(10), e75117. https://doi.org/10.1371/journal.pone.0075117

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., … Hancock, A. M. (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, *114*, 5213–5218. https://doi.org/10.1073/pnas.1616736114

Edin, E., Liljeroth, E., & Andersson, B. (2019). Long term field sampling in Sweden reveals a shift in occurrence of cytochrome b genotype and amino acid substitution F129L in *Alternaria solani*, together with a high incidence of the G143A substitution in *Alternaria alternata*. *European Journal of Plant Pathology / European Foundation for Plant Pathology*, *155*(2), 627–641. https://doi.org/10.1007/s10658-019-01798-9

Everett, K. R., & Henshall, W. R. (1994). Epidemiology and population ecology of kiwifruit blossom blight. *Plant Pathology*, *43*(5), 824–830. https://doi.org/10.1111/j.1365-3059.1994.tb01627.x

Exposito-Alonso, M., Becker, C., Schuenemann, V. J., Reiter, E., Setzer, C., Slovak, R., … Weigel, D. (2018). The rate and potential relevance of new mutations in a colonizing plant lineage. *PLoS Genetics*, *14*(2), e1007155. https://doi.org/10.1371/journal.pgen.1007155

Gabriel, M. F., Postigo, I., Tomaz, C. T., & Martínez, J. (2016). *Alternaria alternata* allergens: Markers of exposure, phylogeny and risk of fungi-induced respiratory allergy. *Environment International*, *89–90*, 71–80. https://doi.org/10.1016/j.envint.2016.01.003

Gerber, E., Schaffner, U., Gassmann, A., Hinz, H. L., Seier, M., & Müller-Schärer, H. (2011). Prospects for biological control of *Ambrosia artemisiifolia* in Europe: Learning from the past. *Weed Research*, *51*(6), 559–573. https://doi.org/10.1111/j.1365-3180.2011.00879.x

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathogens*, *8*(1), 24. https://doi.org/10.1186/s13099-016-0103-7

Gutaker, R. M., & Burbano, H. A. (2017). Reinforcing plant evolutionary genomics using ancient DNA. *Current Opinion in Plant Biology*, *36*, 38–45. https://doi.org/10.1016/j.pbi.2017.01.002

Gutaker, R. M., Reiter, E., Furtwängler, A., Schuenemann, V. J., & Burbano, H. A. (2017). Extraction of ultrashort DNA molecules from herbarium specimens. *BioTechniques*, *62*(2), 76–79. https://doi.org/10.2144/000114517

Hahn, M. (2014). The rising threat of fungicide resistance in plant pathogenic fungi: *Botrytis* as a case study. *Journal of Chemical Biology*, *7*(4), 133–141. https://doi.org/10.1007/s12154-014-0113-1

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, *17*(3), 377–386. https://doi.org/10.1101/gr.5969107

Jakob, K., Goss, E. M., Araki, H., Van, T., Kreitman, M., & Bergelson, J. (2002). *Pseudomonas viridiflava* and *P. syringae* —Natural Pathogens of *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions: MPMI*, *15*(12), 1195–1203.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MAPDAMAGE2. 0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684.

Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A. C., Balmuth, A. L., Robert-Seilaniantz, A., … Jones, J. D. G. (2011). Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biology*, *9*(7), e1001094. https://doi.org/10.1371/journal.pbio.1001094

Kimura, N., & Tsuge, T. (1993). Gene cluster involved in melanin biosynthesis of the filamentous fungus *Alternaria alternata*. *Journal of Bacteriology*, *175*(14), 4427–4435. https://doi.org/10.1128/JB.175.14.4427-4435.1993

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3. https://doi.org/10.1093/nar/gkr771

Lang, P. L. M., Willems, F. M., Scheepens, J. F., Burbano, H. A., & Bossdorf, O. (2019). Using herbaria to study global environmental change. *New Phytologist*, *221*(1), 110–122. https://doi.org/10.1111/nph.15401

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Preprint arXiv:1303.3997.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N. … 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Malmstrom, C. M., Shu, R., Linton, E. W., Newton, L. A., & Cook, M. A. (2007). Barley yellow dwarf viruses (BYDVs) preserved in herbarium specimens illuminate historical disease ecology of invasive and native grasses. *Journal of Ecology*, *95*, 1153–1166. https://doi.org/10.1111/j.1365-2745.2007.01307.x

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., … Gilbert, M. T. P. (2013). Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature Communications*, *4*, 1–7. https://doi.org/10.1038/ncomms3172

Martin, M. D., Zimmer, E. A., Olsen, M. T., Foote, A. D., Gilbert, M. T. P., & Brush, G. S. (2014). Herbarium specimens reveal a historical shift in phylogeographic structure of common ragweed during native range disturbance. *Molecular Ecology*, *23*(7), 1701–1716. https://doi.org/10.1111/mec.12675

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6). https://doi.org/10.1101/pdb.prot5448

Miller, S., Masuya, H., Zhang, J., Walsh, E., & Zhang, N. (2016). Real-time PCR detection of dogwood anthracnose fungus in historical herbarium specimens from Asia. *PLoS ONE*, *11*(4), e0154030. https://doi.org/10.1371/journal.pone.0154030

Mitra, S., Gilbert, J. A., Field, D., & Huson, D. H. (2010). Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. *The ISME Journal*, *4*(10), 1236–1242. https://doi.org/10.1038/ismej.2010.51

Nguyen, H. D. T., Lewis, C. T., Lévesque, C. A., & Gräfenhan, T. (2016). Draft genome sequence of *Alternaria alternata* ATCC 34957. *Genome Announcements*, *4*(1), e01554-15. https://doi.org/10.1128/genomeA.01554-15

Nualart, N., Ibáñez, N., Luque, P., Pedrol, J., Vilar, L., & Guàrdia, R. (2017). Dataset of herbarium specimens of threatened vascular plants in Catalonia. *PhytoKeys*, *77*, 41–62. https://doi.org/10.3897/phytokeys.77.11542

Oh, T. J., Han, S. R., Ahn, D. H., Park, H., & Kim, A. Y. (2016). Complete genome sequence of *Hymenobacter* sp. strain PAMC26554, an ionizing radiation-resistant bacterium isolated from an Antarctic lichen. *Journal of Biotechnology*, *227*, 19–20. https://doi.org/10.1016/j.jbiotec.2016.04.011

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., … Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Payne, W. (1970). Preliminary reports on the floral of Wisconsin. No. 62. Compositae VI. Compositae family VI. The genus *Ambrosia* - the ragweeds. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters*, 353–371.

Peñuelas, J., Poulter, B., Sardans, J., Ciais, P., van der Velde, M., Bopp, L., … Janssens, I. A. (2013). Human-induced nitrogen–phosphorus imbalances alter natural and managed ecosystems across the globe. *Nature Communications*, *4*, 2934. https://doi.org/10.1038/ncomms3934

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. https://doi.org/10.1186/s12915-014-0087-z

Sánchez Barreiro, F., Vieira, F. G., Martin, M. D., Haile, J., Gilbert, M. T. P., & Wales, N. (2017). Characterizing restriction enzyme-associated loci in historic ragweed (*Ambrosia artemisiifolia*) voucher specimens using custom-designed RNA probes. *Molecular Ecology Resources*, *17*(2), 209–220.

Saville, A. C., Martin, M. D., & Ristaino, J. B. (2016). Historic late blight outbreaks caused by a widespread dominant lineage of *Phytophthora infestans* (Mont.) de Bary. *PLoS ONE*, *11*(12), 1–22. https://doi.org/10.1371/journal.pone.0168381

Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., … Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, *9*, 1056–1082. https://doi.org/10.1038/nprot.2014.063

Schubert, M., Lindgreen, S., & Orlando, L. (2016). ADAPTERREMOVAL v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*(1), 88. https://doi.org/10.1186/s13104-016-1900-2

Shepherd, L., & Perrie, L. (2014). Genetic analyses of herbarium material: Is more care required?*Taxon*, *63*(5), 972–973. https://doi.org/10.12705/635.2

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-van Ginkel, R., Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA damage in plant herbarium tissue. *PLoS ONE*, *6*(12), e28448. https://doi.org/10.1371/journal.pone.0028448

The 1001 Genomes Consortium (2016). 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*, *166*(2), 481–491. https://doi.org/10.1016/j.cell.2016.05.063

Thines, M., Choi, Y.-J., Kemen, E., Ploch, S., Holub, E. B., Shin, H.-D., & Jones, J. D. G. (2009). A new species of *Albugo* parasitic to *Arabidopsis thaliana* reveals new evolutionary patterns in white blister rusts (Albuginaceae). *Persoonia*, *22*, 123–128.

Vats, S. (2015). Herbicides: History, Classification and Genetic Manipulation of Plants for Herbicide Resistance. In E. Lichtfouse (Ed.), *Sustainable agriculture reviews*, Vol. *15* (pp. 153–192). Cham: Springer International Publishing.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., … Krause, J. (2017). A Robust framework for microbial archaeology. *Annual Review of Genomics and Human Genetics*, *18*, 321–356. https://doi.org/10.1146/annurev-genom-091416-035526

Weiß, C. L., Schuenemann, V. J., Devos, J., Shirsekar, G., Reiter, E., Gould, B. A., … Burbano, H. A. (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *Royal Society Open Science*, *3*(6), 160239. https://doi.org/10.1098/rsos.160239

Wolters, P. J., Faino, L., van denBosch, T. B. M., Evenhuis, B., Visser, R. G. F., Seidl, M. F., & Vleeshouwers, V. G. A. A. (2018). Gapless Genome assembly of the potato and tomato early blight pathogen *Alternaria solani*. *Molecular Plant-Microbe Interactions: MPMI*, *31*(7), 692–694.

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., … Burbano, H. A. (2013). The rise and fall of the Phytophthora Infestans lineage that triggered the Irish potato famine. *Elife*, *2*, e00731.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bieker VC, Sánchez Barreiro F, Rasmussen JA, Brunier M, Wales N, Martin MD. Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Mol Ecol Resour*. 2020;00:1–14. https://doi.org/10.1111/1755-0998.13174