TECHNICAL ADVANCE

Application of Target Enrichment Sequencing for Population Genetic Analyses of the Obligate Plant Pathogens *Pseudoperonospora cubensis* and *P. humuli* in Michigan

Julian C. Bello,¹ Mary K. Hausbeck,¹ and Monique L. Sakalidis^{1,2,†}

¹ Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, U.S.A.

² Department of Forestry, Michigan State University, East Lansing, MI 48824, U.S.A.

Accepted 30 June 2021.

Technological advances in genome sequencing have improved our ability to catalog genomic variation and have led to an expansion of the scope and scale of genetic studies over the past decade. Yet, for agronomically important plant pathogens such as the downy mildews (*Peronosporaceae*), the scale of genetic studies remains limited. This is, in part, due to the difficulties associated with maintaining obligate pathogens and the logistical constraints involved in the genotyping of these species (e.g., obtaining DNA of sufficient quantity and quality). To gain an evolutionary and ecological perspective of downy mildews, adaptable methods for the genotyping of their populations are required. Here, we describe a targeted enrichment (TE) protocol to genotype isolates from two *Pseudoperonospora* species (*P. cubensis* and *P. humuli*), using less than 50 ng of mixed pathogen and plant DNA for library preparation. We were able to enrich 830 target genes across 128 samples and identified 2,514 high-quality single nucleotide polymorphism (SNP) variants. Using these SNPs, we detected significant genetic differentiation (analysis of molecular variance [AMOVA], $P = 0.01$) between *P. cubensis* subpopulations from *Cucurbita moschata* (clade I) and *Cucumis sativus* (clade II) in the state of Michigan. No evidence of location-based differentiation was detected within the *P. cubensis* (clade II) subpopulation in Michigan. However, a significant effect of location on the genetic variation of the *P. humuli* subpopulation was detected in the state (AMOVA, $P = 0.01$). Mantel tests found evidence that the genetic distance among *P. humuli* samples was associated with the physical distance of the hop yards from which the samples were collected ($P = 0.005$). The differences in the distribution of

genetic variation of the Michigan *P. humuli* and *P. cubensis* subpopulations suggest differences in the dispersal of these two species. The TE protocol described here provides an additional tool for genotyping obligate biotrophic plant pathogens and the execution of new genetic studies.

Keywords: cucurbit downy mildew, cucurbits, environmental sampling, genotyping, hop downy mildew, hops, population genetics, *Pseudoperonospora cubensis*, *Pseudoperonospora humuli*, sequence capture

Downy mildew (DM) pathogens (family *Peronosporaceae*) cause foliar disease in several agronomically important plant species (Gent et al. 2009; Kanetis et al. 2014; Kunjeti et al. 2016; Lane et al. 2005; Rivera et al. 2016; Wallace and Quesada-Ocampo, 2017; Wong and Wilcox 2001). The group is comprised of at least 20 different genera representing a large portion of the described pathogenic species of oomycetes (Choi and Thines 2015; Thines 2014). All DM species are considered obligate biotrophs, growing only in association with living host tissue (Bourret et al. 2018). Over the past decade, several studies have provided key insights into DM biology, including virulence mechanisms (Baxter et al. 2010; Burkhardt and Day 2016; Purayannur et al. 2020; Savory et al. 2012a), host specificity (Choi and Thines 2015; Rivera et al. 2016; Summers et al. 2015; Wallace et al. 2020), and fungicide resistance (Blum et al. 2011; Gisi and Sierotzki 2008). Despite this progress, several unresolved research questions in ecology and evolution remain, many of which could be addressed with emerging genomic and genetic approaches.

Recent technological advances in high-throughput sequencing have facilitated the genotyping of dozens of samples and the analysis of whole plant-pathogen genomes (Cui et al. 2019; Fletcher et al. 2019; Rahman et al. 2019; Rivera et al. 2016; Savory et al. 2012b; Sharma et al. 2015; Withers et al. 2016). The recent availability of genome-wide sequence information from multiple individuals of the same species has facilitated population genomic studies of several plant pathogens (Carleson et al. 2019; Gent et al. 2019; Grünwald et al. 2017; Tabima et al. 2018). Whole or reduced representation genome sequencing technologies have become increasingly popular to monitor genetic changes in plant-pathogen populations. However, these technologies have not been broadly used to study DM populations due to difficulties in obtaining sufficient amounts of DNA from single obligate biotrophic individuals (Milgroom 2015).

[†]Corresponding author: M. L. Sakalidis; sakalidi@msu.edu

Funding: This project was supported, in part, by funds from the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2015-51181-24285, the Agricultural Research Fund of Pickle Packers International, Inc., and the Michigan State University Project GREEN GR15-020.

*The e-Xtra logo stands for “electronic extra” and indicates there is supplementary material published online.

The author(s) declare no conflict of interest.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Traditionally, other genotyping technologies that require significantly lower inputs of DNA, such as Sanger sequencing and microsatellites (simple sequence repeats [SSRs]), are used with DNA extracted from fresh and stored symptomatic tissue to study DM populations (Kitner et al. 2015; Naegele et al. 2016; Quesada-Ocampo et al. 2012; Rivera et al. 2016; Wallace and Quesada-Ocampo 2017; Wallace et al. 2020). However, these approaches are limited by the relatively low number of variants that can be obtained when compared with genome-wide sequencing options. DNA from symptomatic tissue can be used for sequencing with high-throughput sequencing technologies, but the low concentration of pathogen DNA and large amounts of exogenous material (e.g., plant and bacterial DNA) can add further expense and complexity to the sequencing and bioinformatic analysis (Stassen et al. 2012).

The quality of high-throughput sequencing data depends on the quality and purity of the DNA sample, which means that the target pathogen is ideally present in higher amounts when compared with the host plant before DNA extraction (Jouet et al. 2019). A spore propagation approach using detached susceptible leaves can be used to propagate bulk amounts of DM sporangia from which additional pathogen DNA can be recovered (Gent et al. 2019; Summers et al. 2015; Thomas et al. 2017); however, multiple growing cycles (seven to 10 days per cycle) are needed, making this approach time-consuming and labor-intensive (Ali et al. 2011; Gent et al. 2019). Spore propagation is most successful when using freshly collected symptomatic leaves, but propagating sporangia from samples stored over long periods of time or under poor conditions (e.g., samples stored above -80°C) can be complicated. Additionally, sporangial propagation requires continuous maintenance of fresh sporangia on a highly susceptible host, which may bias the genetic composition of the populations under study due to the selection of genotypes by the host and the propagation conditions (Jones et al. 2014; Thomas et al. 2017). While sporangial propagation can provide large amounts of high-quality (HQ) DNA needed for sequencing, this approach is impractical for population studies requiring a large number of samples collected over time (i.e., years).

Sequence capture methods may provide a solution for genotyping DM pathogens that excludes nontarget DNA while also

facilitating high coverage sequencing of several target loci (Kozarewa et al. 2015; Lim and Braun 2016). These techniques use affinity probes (RNA/DNA) to isolate particular sequences of interest (called target regions) out of a larger pool of DNA fragments (DNA library). Sequence capture methods such as target enrichment (TE) have been used across a variety of genomic studies with model (Clark et al. 2011; Gnirke et al. 2009) and nonmodel organisms (Faircloth et al. 2015; McCormack et al. 2016; Starrett et al. 2017) and have also facilitated the study of museum specimens with low amounts of poor-quality DNA (Cruz-Dávalos et al. 2017). Additionally, sequence capture methods have been used to study the genetic variation of other plant pathogens, such as *Phytophthora infestans* (Lin et al. 2020), *Phytophthora capsici* (Thilliez et al. 2019), and the obligate biotroph *Albugo candida* (Jouet et al. 2019). These techniques could facilitate the study of DM pathogen populations using symptomatic tissue samples, alleviating the need for sporangial propagation and expanding the type and condition of the samples used.

In this study, we evaluated the genotypic variation of two DM species of genus *Pseudoperonospora* using TE sequencing. *P. cubensis* and *P. humuli* infect cucurbits and hops, respectively, worldwide and are considered the most economically important species of the genus *Pseudoperonospora* (Choi et al. 2005; Mitchell et al. 2011). *P. cubensis* (clades I and II) causes foliar blight (DM) of cucumber (*Cucumis sativus*), melon (*Cucumis melo*), pumpkin (*Cucurbita maxima*), watermelon (*Citrullus lanatus*), and squash (*Cucurbita moschata*) (Summers et al. 2015; Thomas et al. 2017; Wallace et al. 2020). *P. humuli* negatively impacts hop (*Humulus lupulus*) cone yield (Gent et al. 2010). We describe an optimized TE procedure for sequencing DM and a bioinformatic pipeline for population genetic analyses using TE data. Evaluation of the structure, diversity, and reproduction of *P. cubensis* and *P. humuli* populations in Michigan was of particular interest (Fig. 1).

RESULTS

Library sequencing and read mapping.

The sequencing and alignment results of all libraries used in this study are summarized in Table 1. A total of 270 samples

Fig. 1. Number of acres of *Cucumis sativus* and *Humulus lupulus* planted in Michigan (adapted from Neufeld 2017). **A**, *C. sativus* planted acreage, location, and number of samples collected in Michigan ($n = 40$). A total of 26 samples were collected outside of Michigan in Ontario Canada ($n = 6$) and the states of Florida ($n = 2$), Alabama ($n = 1$), North Carolina ($n = 2$), South Carolina ($n = 2$), California ($n = 1$), New York ($n = 1$), Wisconsin ($n = 1$), Ohio ($n = 3$), Iowa ($n = 1$), Indiana ($n = 5$), and Georgia ($n = 1$). **B**, *H. lupulus* planted, location, and number of samples collected from hop yards in Michigan ($n = 33$) and Oregon ($n = 1$). Hop yards with more than 800 plants (*H. lupulus*) are represented by stars, those with fewer are represented by circles. Hop yards sampled (A to G) are colored in red. **C**, Population assignment by geographic location of *Pseudoperonospora* spp. samples in the Midwest (Wisconsin, Indiana, Ohio, Michigan [Mi]), and Ontario, Canada).

containing from 2 to 500 ng of DNA were used for library preparation and enrichment. However, we could not generate enough enrichment products from 110 libraries, including 51 libraries prepared from samples with less than 50 ng of DNA. The remaining libraries ($n = 160$) were enriched successfully and were prepared from samples containing from 2 to 500 ng of DNA; these libraries included 118 libraries of *P. cubensis* and 42 libraries of *P. humuli* that were sequenced using three lanes of HiSeq4000.

TE libraries. Approximately 800 million paired end reads (2×150 bp) were obtained from the sequencing of these libraries (162) generating approximately 10 million reads ± 9.2 million (standard deviation) per library. A range of 13 to 95% of the reads per sample aligned to the target reference genome of *P. cubensis*. The total number of reads obtained per library was influenced by the hybridization time of the libraries with the capture probes for enrichment and the number of samples sequenced per lane. The enrichment sets 1 and 2 (TES1 and TES2) contained approximately the same number of libraries (62 and 72); however, the libraries of TES2 that hybridized for 48 h generated almost two times more aligned reads (2.64 ± 2.6 million) than the libraries that hybridized for 24 h (1.34 ± 1.0 million) (Table 1). Five percent of the reads per library that did not align to the reference genome of *P. cubensis* aligned to the *C. sativus* reference genome.

Low-coverage whole-genome sequencing (Lc-WGS) libraries. Seven libraries of *P. cubensis* and two libraries of *P. humuli* were sequenced without enrichment (Lc-WGS), using MiSeq, generating a total of 41 million sequencing reads (2×150 bp). Approximately 0.55 ± 0.37 million reads were obtained per library, but only 25% of them aligned to the reference genome of *P. cubensis*. Twenty-five percent of the remaining reads aligned to the *C. sativus* reference genome.

WGS libraries. A total of 350 million paired-end reads (2×150 bp) were obtained from Bioproject PRJNA360426, which included WGS data of nine libraries of *P. cubensis* and one library of *P. humuli*. An average of 15.6 ± 6 million reads per sample were obtained after sequencing. Eighty-nine percent of the total reads aligned to the *P. cubensis* reference genome and approximately 2% of the reads per sample aligned to the reference genome of *C. sativus*.

Genotyping and variant calling.

Single nucleotide polymorphisms (SNPs) were retrieved from the sequencing reads generated from 160 enriched DNA libraries

and the reads obtained from the 10 libraries sequenced using WGS (Bioproject PRJNA360426). Ninety percent of the enriched libraries were prepared from input DNA amounts between 2 to 100 ng (Fig. 2B). After mapping the reads to the target reference genome of *P. cubensis*, 5,214 biallelic SNPs were identified across 70% of the 170 samples. These SNPs were located within 170 of the regions targeted for enrichment. To retain the maximum number of SNPs, we excluded 42 of the samples prepared using TE, as they contained more than 40% missing data (24 of these were hybridized for 24 h, with the remaining libraries hybridized for 48 h) (Table 1). SNPs with more than 75% of the missing data across the entire data set were also excluded. This resulted in a total of 2,514 HQ SNPs with <7% missing data for downstream analysis. All HQ SNPs were located within 107 of the 830 genes targeted, 66% of the SNPs were found within secreted and effectors proteins. The remaining 34% of the SNPs were contained within esterases (16.3%), glucanases (3%), lyases (3%), SSRs (4%), and other (7%) genes.

The coverage depth (CD) across the 2,514 HQ SNPs was variable among samples and significant differences in the CD were observed between the libraries genotyped using TE and the libraries genotyped using WGS (Kruskal-Wallis $P = 2.2e-16$) (Fig. 2A). The enriched libraries with a hybridization time of 48 h (TES3) exhibited a significantly higher CD than the enriched libraries generated with 24 h of hybridization time (TES2) or the libraries genotyped using WGS (Kruskal-Wallis $P = 2.2e-16$) (Fig. 2A). The CD across all HQ SNPs averaged 262 ± 546 , and the CD was 358 ± 546 for the libraries with hybridization times of 24 and 48 h, respectively (Table 1; Fig. 2A). Among the samples sequenced using WGS, the CD across HQ SNPs averaged 180 ± 116 (Table 1). No correlation was observed between the amount of input DNA used for library preparation and the CD across HQ SNPs ($P > 0.05$) (Fig. 2B).

Characterization of technical error.

Technical replicates were used to characterize the number of errors introduced into a sample during TE. Of the five samples submitted for enrichment, only one sample and its technical replicate were excluded after the quality filtering of the SNPs. Differences in the average coverage between technical replicates of the same sample were observed (Fig. 3A); however, only a small proportion of loci were different between replicates (Fig. 3B). On average, 18 SNPs (out of 2,514 total SNPs) differed between technical replicates of *P. cubensis*, suggesting an error rate of roughly 0.7% within the HQ variants. A total of

Table 1. Sequencing and alignment results from libraries sequenced using target enrichment, low-coverage whole-genome sequencing (Lc-WGS), and WGS

Sequencing characteristics	TES1 ^a	TES2	TES3	Lc-WGS	WGS ^b
Sequencing platform	HiSeq4000	HiSeq4000	HiSeq4000	MiSeqV2.300	HiSeq
Sequencing format	2×150 bp	2×150 bp	2×150 bp	2×150 bp	2×100 bp
Sequencing lanes	1	1	1	2	1
Expected sequencing per lane (Gbp)	105	105	105	3.6 – 4.5	105
Input amount of DNA per library	5 to 400 ng	5 to 400 ng	5 to 400 ng	<50 ng	>1 μ g
Hybridization time with probes (baits)	24 h	48 h	24 h	–	–
Total number of libraries sequenced	62	72	26	9	10
Total number of clean reads (millions)	211	290	300	41	350
Reads aligned to the <i>P. cubensis</i> genome (%)	35	54	72	25	89
Reads aligned to the <i>Cucumis sativus</i> genome (%)	<5	<5	<5	>25	2
Aligned reads per samples $\pm$ SD (millions) ^c	1.34 ± 1.0	2.64 ± 2.6	16.6 ± 12	0.55 ± 0.37	16.6 ± 18
Libraries with high quality SNPs (%) ^d	61	84	92	–	100
Average coverage among HQ-SNPs ^e	262 ± 446	358 ± 546	542 ± 346	–	180 ± 116

^a Target enrichment sets 1 and 2 (TES1 and TES2) were hybridized for 24 and 48 h, respectively.

^b All the samples genotyped using WGS were retrieved from Bioproject PRJNA360426 published by Thomas et al. (2017).

^c The value given in millions of reads aligned to the reference genome of *Pseudoperonospora cubensis* per sample $\pm$ standard deviation (SD).

^d Percent of libraries retained containing 2,514 high-quality single nucleotide polymorphisms (SNPs).

^e Average coverage among high-quality SNPs within the libraries retained $\pm$ SD.

30 SNPs differed between the two replicates of *P. humuli* (one sample), suggesting a slightly higher error rate of 1.2% within the HQ variants.

Genetic differentiation and population structure.

In the genetic distance tree of HQ SNPs (2,514) all 128 samples genotyped fell into one of two clades (100% bootstrap support); either *P. cubensis* (94 samples) or *P. humuli* (34 samples) (Fig. 4). Within the *P. cubensis* clade, samples were either clade I or II of *P. cubensis* (Thomas et al. 2017). The majority of samples from *Cucumis* spp. aligned with clade II of *P. cubensis*, while the remaining samples collected from *Cucurbita moschata* and *Citrullus lanatus* aligned with clade I (Fig. 4).

Samples in the ordination plot formed three clusters corresponding to *P. humuli*, *P. cubensis* clade I, and *P. cubensis* clade II; these clusters were consistent with the stratification by species and within *P. cubensis* by host in the genetic distance tree (Fig. 5A). F-statistics ($F_{ST} > 0.25$) and analysis of

molecular variance (AMOVA) detected significant genetic differentiation ($P = 0.01$) between pathogen species, supporting the clustering of the ordination plot (Table 3). Stratification by host was supported by a high level of genetic differentiation among samples collected from *Cucumis* spp. and samples collected from *Cucurbita moschata* ($F_{ST} = 0.08$) (Table 3). AMOVA revealed that 78% of the total genetic variance was significantly associated with differences between host within *P. cubensis* ($P = 0.01$) (Table 4).

All *P. cubensis* samples collected from *C. sativus* in the Midwest, including Michigan, and the reference sample from Florida clustered loosely together across all four quadrants of the ordination plot (Fig. 5B). However, the reference samples from North Carolina and California clustered separately from each other and were not contained within the 90% confidence ellipses of the other subpopulations (Fig. 5B). More than 70% of the samples from the subpopulations of central and east Michigan were contained within the 90% confidence ellipse of the Midwest-w/o-

Fig. 2. Depth coverage of high-quality single nucleotide polymorphisms (SNPs) across libraries. **A**, Depth coverage of high-quality SNPs ($\log_{10}$) across libraries of *Pseudoperonospora cubensis* and *P. humuli* sequenced using target enrichment with 24 (TES2) and 48 h (TES3) of hybridization and whole-genome sequencing (WGS). **B**, Depth coverage of high-quality SNPs ($\log_{10}$) across libraries sequenced using target enrichment with different amounts of input DNA for library preparation. Only samples that were retained after quality filtering were included in the analysis. All the sample genotyped using WGS were retrieved from Bioproject PRJNA360426 published by Thomas et al. (2017).

Fig. 3. Characterization of sequencing error among technical replicates. **A**, Coverage distribution of the technical replicates sequenced. Box plots are filled with 2,514 high-quality single nucleotide polymorphisms (SNPs). Each sample is colored according to the enrichment set (TES). TES1 and TES2 were hybridized for 24 and 48 h, respectively. **B**, Genetic differentiation among technical replicates. An unweighted pair group method with arithmetic means tree was reconstructed using 2,514 SNPs. The genetic distance represents the number of SNPs that are different among samples. Pairwise distances among technical replicates are shown in the top left corner.

MI subpopulation (Fig. 5B); however, only 40% of the samples from west Michigan were contained within the ellipse of the of the Midwest-w/o-MI subpopulation. AMOVA did not support significant genetic differentiation among the subpopulations of *P. cubensis* from *C. sativus* in Michigan and the Midwest-w/o-MI subpopulation ($P = 0.39$). Less than 1% of the total genetic variance was associated with difference among subpopulations (or regions) (Tables 6 and 7).

The clone correction of all *P. cubensis* samples collected from *C. sativus* and *C. melo* ($n = 78$) resulted in 34 multilocus

genotypes (MLGs), 35% of which were detected in multiple subpopulations (Fig. 6A and B). MLGs 103 and 114 were detected in all the subpopulations of Michigan (east, west, and central) and the Midwest-w/o-MI subpopulation. MLGs 44, 55, 68, 82, 115, and 125 were detected in more than two subpopulations within Michigan. The Midwest-w/o-MI subpopulation also shared MLG 51 and MLG 54 with the subpopulation of east Michigan (Fig. 6A). Additionally, MLG 33 from Florida was also detected in the east Michigan subpopulation (Fig. 6A). Supporting these findings, the Mantel test did not detect a significant

Fig. 4. Genetic differentiation of *Pseudoperonospora cubensis* and *P. humuli* samples. An unweighted pair group method with arithmetic means tree was reconstructed using 2,514 single nucleotide polymorphism (SNP) variants ($n = 128$). Bootstrap support values are indicated above the branches. The genetic distance represents the proportion of loci that are different between samples. One asterisk (*) indicates samples previously classified as clade I members and two (**) indicate samples previously classified as clade II members. Technical replicates are enclosed in red squares.

Fig. 5. Ordination plots of *Pseudoperonospora* spp. based on 2,514 single nucleotide polymorphisms. **A**, Ordination plot of *P. cubensis* and *P. humuli* samples according to host species. All points represent samples collected in Michigan unless indicated otherwise. **B**, Ordination plot of *P. cubensis* samples from *Cucumis sativus* from 2007 to 2017 in the United States ($n = 57$). **C**, Ordination plot of *P. humuli* samples from hop yards of Michigan in 2017 and 2018 ($n = 34$).

relationship between the genetic and physical distances of the *P. cubensis* samples collected from *C. sativus* in the Midwest ($P = 0.454$) (Fig. 7A).

The level of genetic differentiation among *P. humuli* samples from Michigan and Oregon ($F_{ST} = 0.014$) was higher when compared with the genetic differentiation detected among *P.*

cubensis samples from *C. sativus* originating in the Midwest, California, Florida, and North Carolina ($F_{ST} = 0.0018$). In the ordination plot of all the *P. humuli* samples, most samples from the subpopulation of west Michigan were loosely dispersed in the left quadrants, while most samples from the subpopulation of north Michigan (75%) were clustered tightly in the right

Table 2. Plant hosts and the location of the 128 samples of *Pseudoperonospora* spp. samples used for population analyses

Pathogen and host species	Michigan				Other U.S. states	Total
	West	East	North	Central		
<i>P. cubensis</i> clade I						
<i>Cucurbita</i> spp. and <i>Citrullus lanatus</i>	3	1	0	0	7	11
<i>P. cubensis</i> clade II						
<i>Cucumis</i> spp.	27	19	0	11	26	83
<i>P. humuli</i>						
<i>Humulus lupulus</i>	8	0	20	5	1	34
Total ^a	39	20	20	15	34	101

^a Total reflects the sum of the samples retained after quality filtering.

Table 3. Pairwise F-statistics (F_{ST}) comparisons among *Pseudoperonospora cubensis* (clade I and II) and *Pseudoperonospora humuli*^a

Pairwise comparison	F_{ST}	P_{value} ^b
<i>P. cubensis</i> clade II (<i>Cucumis</i> spp.) vs. <i>P. cubensis</i> clade I (<i>Cucurbita</i> spp.)	0.08	0.01
<i>P. cubensis</i> clade II (<i>Cucumis</i> spp.) vs. <i>P. humuli</i> (<i>Humulus lupulus</i>)	0.27	0.01
<i>P. cubensis</i> clade I (<i>Cucurbita</i> spp.) vs. <i>P. humuli</i> (<i>H. lupulus</i>)	0.25	0.01

^a Sample sizes used in this analysis included 83, 34 and 11 samples collected from *Cucumis* spp., *Humulus lupulus* and *Cucurbita* spp., respectively.

^b P_{values} were calculated using an analysis of molecular variance.

Table 4. Analysis of molecular variance for *Pseudoperonospora* spp. grouped by host^a

Source	Df	SSD	MSD	Sigma	Variance (%)	P_{value}
Between host	5	3,649.1124	729.82249	51.209771	78.720821	0.01
Between samples within host	16	342.3743	21.39839	1.463212	2.239468	0.01
Within samples	106	1,373.5466	12.95799	12.950063	19.749711	0.01
Total	127	5,365.0333	42.24436	65.631446	100	

^a The significance of variance was tested from 999 permutations of the data. Sample sizes used in this analysis included 83, 34, and 11 samples collected from *Cucumis* spp., *Humulus lupulus*, and *Cucurbita* spp., respectively. Df = degrees of freedom; SSD = sum of squared differences; and MSD = mean of squared differences. Variance (%) was adjusted to zero for negative sigma values.

Table 5. Genotypic diversity estimates for *Pseudoperonospora* spp. samples grouped by host (clade)^a

Clade/host species	N	MLG	eMLG	SE	H	G	λ	E_5	Hexp
<i>P. cubensis</i> clade I									
<i>Cucurbita</i> spp. and <i>Citrullus lanatus</i>	11	9	9	0.00	2.10	7.12	0.860	0.855	0.215
<i>P. cubensis</i> clade II									
<i>Cucumis</i> spp.	83	37	9.25	1.11	3.30	20.08	0.950	0.728	0.192
<i>P. humuli</i>									
<i>Humulus lupulus</i>	34	25	9.19	1.14	2.96	12.30	0.919	0.620	0.217
Total	128	70	10.01	0.923	3.94	35.93	0.972	0.692	0.206

^a N = number of individuals sampled of each region; MLG = multilocus genotypes observed; eMLG = expected MLGs at a sample size based on rarefaction; SE = standard error; H = Shannon-Wiener Index; G = Stoddart and Taylor's index; λ = Simpson's index; E_5 = evenness; Hexp = within population gene diversity.

Table 6. Analysis of molecular variance of Midwest subpopulations of *Pseudoperonospora cubensis* collected from *Cucumis sativus*^a

Source	Df	SSD	MSD	Sigma	Variance (%)	P_{value}
Between subpopulations	3	19.94	6.646	0.000	-0.127	0.39
Between samples within subpopulations	20	126.08	6.304	0.1495	2.475	0.32
Within samples	54	318.56	5.899	5.899	97.65	0.42
Total	77	464.58	6.033	6.0411	100	

^a The significance of variance was tested from 999 permutations of the data. Sample sizes used in this analysis included 20, 27, and 11 samples collected from *Cucumis* spp. in the east, west, and central Michigan, respectively. The Midwest-w/o-MI subpopulation (20 samples) was formed by all the samples collected in midwest states of the United States as well as Ontario, Canada, not including Michigan. Df = degrees of freedom; SSD = sum of squared differences; and MSD = mean of squared differences. Variance (%) was adjusted to zero for negative sigma values.

quadrants (Fig. 5C). Only two samples from the north subpopulation were contained within the 90% confidence ellipse of the west subpopulation (Fig. 5C). The samples from central and east Michigan were either contained in the ellipses from the north or west subpopulations. The reference sample from Oregon was not contained within the 90% confidence ellipse of Michigan subpopulations (Fig. 5C). These patterns were supported by AMOVA, in which significant genetic differentiation between the subpopulations of north and west Michigan was detected ($P = 0.01$) (Table 9). A total of 7.1% of the genetic variance was significantly associated with differences among subpopulations and 9.3% of the variance was associated with differences among hop yards within a subpopulation (Table 9). This suggests that geographic region may have a significant effect on the structure of the *P. humuli* population in Michigan.

Clone correction of the 34 *P. humuli* samples resulted in 23 MLGs, none of which were detected in two different regions (Fig. 6C and D). However, some MLGs were detected in at least two hop yards within the same region. MLGs 21 and 16 were detected in two hop yards sampled in north Michigan (Fig. 6C). The Mantel test supported the hypothesis that the genetic distance among *P. humuli* samples collected in Michigan was significantly associated ($P = 0.0042$) with the geographic distance among hop yards (Fig. 7B).

Genotypic diversity.

The highest level of genotypic diversity was observed among the samples of *P. humuli* followed by *P. cubensis* clade I and *P. cubensis* clade II (Table 5). Each species or clade consisted of different MLGs spread relatively evenly. This was reflected in the estimates of genotypic richness (eMLG = 9 to 9.25) and evenness ($E_5 > 0.8$) (Table 5). The diversity indexes of Shannon-

Wiener, Stoddart and Taylor, and Simpson were highest among *P. cubensis* clade II samples, followed by *P. humuli* samples, and then *P. cubensis* clade I samples (Table 5). The greatest expected heterozygosity or proportion of heterozygous genotypes expected was observed among *P. humuli* samples, followed by *P. cubensis* clade I samples and, then, *P. cubensis* clade II samples (Table 5).

The Michigan subpopulations and the Midwest-w/o-MI subpopulation of *P. cubensis* collected from *C. sativus* exhibited similarly low levels of expected heterozygosity and genotypic diversity (Table 8). Expected heterozygosity ranged between 0.178 and 0.197, while genotypic richness (eMLG) ranged between 8.91 to 9.7. The east Michigan subpopulation had the greatest expected heterozygosity (Hexp = 0.197), followed by Midwest-w/o-MI (Hexp = 0.190), west Michigan (Hexp = 0.189), and central Michigan subpopulations (Hexp = 0.178) (Table 8). The east Michigan subpopulation had the highest genotypic richness (eMLG), while west Michigan and the Midwest-w/o-MI subpopulations had larger sample sizes and lower numbers of MLGs (Table 8). The diversity indexes of Shannon-Wiener, Stoddart and Taylor, and Simpson were all highest for the east Michigan subpopulation, followed by the subpopulations of west Michigan, Midwest-w/o-MI, and then, central Michigan (lowest indices) (Table 8). The highest E_5 was detected for the central Michigan subpopulation followed by the subpopulations of east Michigan, west Michigan, and Midwest-w/o-MI (lowest E_5) (Table 8).

The genotypic diversity varied more widely across the *P. humuli* subpopulations of Michigan (eMLG = 5 to 12) compared with the variation detected among *P. cubensis* subpopulations (eMLG = 8.91 to 9.7) (Table 10). The expected heterozygosity among the subpopulations of *P. humuli* was higher and ranged between 0.206 to 0.220. The subpopulations of north (Hexp = 0.224) and east-central Michigan (Hexp = 0.220) had the greatest expected heterozygosity, followed by the subpopulation of west Michigan (Hexp = 0.206) (Table 10). The subpopulation of west Michigan had the highest genotypic richness, while the north Michigan subpopulation had the largest sample size and the most MLGs. The diversity indexes of Shannon-Wiener, Stoddart and Taylor, and Simpson were all highest for the north Michigan subpopulation, followed by the subpopulations of west and east-central Michigan (lowest indices) (Table 10). The highest E_5 was detected in the subpopulations of west and east-central Michigan, followed by the north Michigan subpopulation (lowest E_5) (Table 10).

Reproductive mode.

To determine whether allelic variants were randomly associated as expected in populations with a sexual mode of

Table 7. Pairwise F-statistics comparisons among subpopulations of *Pseudoperonospora cubensis* collected from *Cucumis sativus* in the Midwest^a

Subpopulation	Michigan			Midwest-w/o-Michigan
	East	Central	West	
East Michigan	—			
Central Michigan	0.0005	—		
West Michigan	0.0003	0.00142	—	
Midwest-w/o-Michigan	0.0019	-0.00001	0.0047	—

^a Sample sizes used in this analysis included 20, 27, and 11 samples collected from *Cucumis* spp. in the east, west, and central Michigan, respectively. The Midwest-w/o-MI subpopulation (20 samples) was formed by all the samples collected in Midwest states well as Ontario, Canada, not including Michigan.

Fig. 6. Frequency and geographic distribution of *Pseudoperonospora cubensis* and *P. humuli* multilocus genotypes (MLGs). **A**, Frequency of *P. cubensis* genotypes collected from *Cucumis sativus* in the Midwest. Unique genotypes are colored in gray. **B**, Frequency of *P. humuli* genotypes collected from *Humulus lupulus* in Michigan. Unique genotypes are colored in gray.

reproduction (low linkage), we calculated the mean index of association (I_A) across loci for each species in Michigan and compared with the estimated I_A of simulated populations under strong linkage (100% linked loci), moderate linkage (75 and 50% linked loci), and low linkage (25% and 0 loci under linkage) (Fig. 8). Upon comparison, significant differences were observed between the mean I_A of each species and the mean values estimated for the simulated populations (*P. cubensis*, $P < 2.2 \times 10^{-16}$ and *P. humuli*, $P < 2.2 \times 10^{-16}$). The I_A mean value of the *P. cubensis* samples was situated between the I_A mean values of the simulated data with 50 and 75% linkage (Fig. 8A). This indicated that populations of *P. cubensis* have a mixed mode of reproduction with a predominantly clonal phase. The I_A mean value of the *P. humuli* samples was not significantly different

from the simulated data with 25% linkage (Fig. 8B), suggesting that the populations of *P. humuli* may have a mixed mode of reproduction that could be predominantly sexual.

DISCUSSION

Technological advances in genome sequencing have accelerated the cataloging of genomic variation of plant pathogens; however, extracting large amounts of HQ DNA from obligate pathogens responsible for DM is challenging and hampers their genotyping using next-generation sequencing. We adapted a TE method that facilitated the sequencing of 712 genes annotated as virulence factors in the genome of *P. cubensis*. We used this method to sequence *P. cubensis* and *P. humuli* DNA extracted

Fig. 7. Relationship between genetic differences and geographic distances among samples. **A**, *Pseudoperonospora cubensis* samples from *Cucumis sativus* and **B**, *P. humuli* samples from *Humulus lupulus* in Michigan.

Table 8. Genotypic diversity estimates and index of association of *Pseudoperonospora cubensis* subpopulations collected from *Cucumis sativus* in the Midwest^a

Subpopulation	N	MLG	eMLG	SE	H	G	λ	E_s	Hexp
East Michigan	20	16	9.7	0.854	2.69	13.3	0.925	0.896	0.197
Central Michigan	11	9	9	0	2.15	8.07	0.876	0.936	0.178
West Michigan	27	14	7.75	1.149	2.38	8.85	0.883	0.769	0.189
Midwest-w/o-Michigan ^b	20	14	8.91	0.973	2.51	10.5	0.905	0.842	0.19
Total	78	34	9.07	1.14	3.21	18.2	0.945	0.727	0.192

^a N = number of individuals sampled of each region; MLG = multilocus genotypes observed; eMLG = expected MLGs at a sample size based on rarefaction; SE = standard error; H = Shannon-Wiener Index; G = Stoddart and Taylor's index; λ = Simpson's index; E_s = evenness; Hexp = within population gene diversity.

^b The subpopulation formed by all the samples collected in the midwest states of the United States as well as Ontario, Canada, not including Michigan.

Table 9. Analysis of molecular variance of Michigan subpopulations of *Pseudoperonospora humuli*^a

Source	Df ^b	SSD ^c	MSD ^d	Sigma	Variance (%)	P _{value}
Between subpopulations	1	36.187	36.18681	1.136518	7.249556	0.01
Between samples within subpopulations	2	45.042	22.68075	1.489694	9.346595	0.03
Within samples	24	319.12	13.22243	13.32243	83.503849	0.28
Total	27	401.34	14.42018	15.95864	100	

^a The significance of variance was tested from 999 permutations of the data. Sample sizes used in this analysis included 8 and 20 samples from two and three hop yards in west and north Michigan, respectively. Df = degrees of freedom; SSD = sum of squared differences; and MSD = mean of squared differences. Variance (%) was adjusted to zero for negative sigma values.

from sporangia collected from plant tissue with signs of the pathogen. This approach facilitated the genotyping of samples that contained very low amounts of pathogen DNA mixed with other environmental contaminants (i.e., plant DNA, bacteria DNA). After aligning the sequenced DNA, we identified 2,514 SNPs and resolved the population structure of *P. cubensis* and *P. humuli* in Michigan. An effect of location on the genetic variation of *P. humuli* was detected and the genetic distance among samples was associated with the physical distance of the hop yards. Evidence of location-based differentiation within Michigan was not detected for the *P. cubensis* population.

By using affinity probes and several amplification cycles, our TE protocol was designed to provide high coverage sequencing of specific loci in the *P. cubensis* and *P. humuli* genome. This facilitated the sequencing of a low amount of target (pathogen) DNA from environmental samples containing a significant amount of contaminant DNA from plant tissues and other microorganisms. Sequencing DNA directly from environmental samples reduces time, labor, and DNA input that are required for other sequencing approaches, such as genotyping by sequencing (GBS) or WGS (Gent et al. 2019; Summers et al.

2015; Thomas et al. 2017). This is particularly advantageous for DM, as other sequencing approaches (GBS, WGS, Rad-Seq) rely on large amounts of HQ DNA that can only be obtained using a laborious propagation procedure (Gent et al. 2019; Thomas et al. 2017). Using TE, several samples with less than 50 ng of DNA were successfully genotyped. Most of the samples that failed were single leaf lesions containing few or no sporangia (*P. cubensis* had been previously confirmed via quantitative PCR [qPCR]). The TE protocol also enabled the genotyping of samples that were no longer viable for propagation, expanding the number of samples that could be genotyped, including those collected more than 10 years ago.

Using a reduced amount of DNA (2 to 100× less), the samples sequenced after TE reached similar sequencing coverage when compared with the samples sequenced using a WGS approach, at the region's target (250×). However, the enriched samples required six to seven times less space for sequencing (60 to 70 samples per lane of HiSeq) compared with the space used for the sequencing of libraries using WGS (10 samples per lane of HiSeq). This was possible because TE reduced the complexity of the *P. cubensis* genome from 88.22 MB to <1 MB

Table 10. Genotypic diversity estimates and index of association of Michigan subpopulations of *Pseudoperonospora humuli*^a

Subpopulation	N	MLG	eMLG	SE	H	G	λ	E _s	Hexp
West	8	8	8	0	2.08	8	0.875	1	0.206
North	20	12	7.02	1.05	2.11	5.88	0.83	0.676	0.224
East-central	5	5	5	0	1.61	5	0.8	1	0.220
Total	33	25	8.33	1.08	2.86	12.52	0.92	0.701	0.217

^a N = number of individuals sampled of each region; MLG = multilocus genotypes observed; eMLG = expected MLGs at a sample size based on rarefaction; SE = standard error; H = Shannon-Wiener Index; G =Stoddart and Taylor's index; λ = Simpson's index; E_s = evenness; Hexp = within population gene diversity.

Fig. 8. Estimation of the degree of linkage disequilibrium within *Pseudoperonospora cubensis* and *P. humuli* of Michigan. **A**, Observed index of association (I_A) distribution of *P. cubensis* compared with the I_A distributions of 0, 25, 50, 75, and 100% linkage. **B**, Observed I_A distribution of *P. humuli* compared with the I_A distributions of 0, 25, 50, 75, and 100% linkage. Groupings based on the Kruskal-Wallis rank-sum test are noted by the letters over the boxplots, in which the *P. cubensis* and *P. humuli* population datasets are grouped with the simulated 50 and 25% linkage data, respectively.

(i.e., 1.13% of the genome). This reduction in the genome complexity and the enrichment of up to four samples in a single enrichment experiment (MyBaits) make TE a cost-effective alternative to WGS.

However, we found that our TE protocol can result in a higher number of sequencing errors per sample, as compared with WGS. The high number of amplification cycles used for the enrichment of samples with low amounts of DNA may have led to the introduction of PCR errors and a subsequent reduction in genotyping accuracy. Typically, high-throughput sequencing (e.g., WGS and GBS) without enrichment of samples with low amounts of DNA (<250 ng) requires an average of 14 PCR cycles, however, six to 14 extra cycles were used after library construction in our TE protocol (for a total of 20 to 28 cycles). Thus, possibly due to the high number of extra amplification cycles, the samples genotyped using TE were subject to an error rate between 0.7 and 1.2%, an error rate 10 times higher than the estimated error rate of >0.1% for high-throughput sequencing data (Grünwald et al. 2017; Ma et al. 2019). We calculated this error rate using only three and one technical replicate from two different species; this is far from ideal as sequencing error rates can show sample-specific effects and sequence context (genome) dependency (Ma et al. 2019). Therefore, a higher number of technical replicates (approximately three technical replicates of 10% of the total number of samples) is recommended for future studies to more precisely assess the genotyping reproducibility of TE. In this study, despite an error rate of about 1%, the genetic distance between samples in 95% of cases was greater than the distance generated due to PCR errors.

Using TE, we detected genetic patterns linked to pathogen biology. For example, a significant effect of host in the population structure of *P. cubensis* was detected, and the samples collected from *C. sativus* were genetically different from the samples collected from *Cucurbita moschata*. These results support previous studies that have shown that the structure of the *P. cubensis* population is driven by host preference, with samples from *Cucumis* spp. and *Cucurbita moschata* belonging to two distinct evolutionary clades (Thomas et al. 2017; Wallace et al. 2020). Thomas et al. (2017) identified two *P. cubensis* clades in the United States; clade I occurs on *Cucurbita pepo*, *Cucurbita moschata*, *Cucurbita maxima*, and *Citrullus lanatus*, and clade II occurs on *Cucumis* spp. In our analyses, all *P. cubensis* samples clustered by clades according to that distribution.

Annual cucurbit downy mildew (CDM) infections in the northern United States are driven by an influx of airborne *P. cubensis* sporangia from overwintering sites (Bello et al. 2020; Goldenhar and Hausbeck 2019; Naegele et al. 2016). CDM limits cucumber yield in Michigan, where more than 37,000 acres of cucumbers are planted every year (United States Department of Agriculture 2020). Small, yet significant genetic differences among cucumber production regions in Michigan were previously reported (Naegele et al. 2016). However, a significant effect of location in the subpopulation structure of *P. cubensis* was not detected in the current study, suggesting no differentiation within and among regions in Michigan. An exchange of migrants has a homogenizing effect on subpopulations (Milgroom 2015) and may occur in Michigan due to the availability of susceptible crops across the state. The detection of the same MLG in multiple locations supports this hypothesis and also suggests that *P. cubensis* sporangia may be disseminated unrestricted in the state, as there is no geographical barrier (e.g., mountain range or water body) that may limit the homogenizing of geographically distant populations. The wide host availability in Michigan may also facilitate the establishment of incoming MLG from other subpopulations in the Midwest. The exchange of migrants may occur among Midwestern states outside of Michigan, but the absence of susceptible crops between geographically distant

populations may result in lower rates of exchange and, subsequently, more genetically differentiated populations. Additional sampling is needed to test this hypothesis.

Significant differences were detected between the *P. humuli* populations from the north and west regions of Michigan, and we did not find any evidence of genetic differentiation among hop yards within the same region. In Michigan in 2018, approximately 990 acres of hops were planted across more than 50 commercial hop yards (Michigan Department of Agriculture & Rural Development, Lansing, MI, U.S.A.) and, despite the potential for airborne dispersal of hop DM (Bello et al. 2020; Gent et al. 2009), none of the 12 MLGs detected in the north region ($n = 20$) were detected in west ($n = 8$) or central ($n = 5$) Michigan. This is consistent with the restricted pattern of dispersal suggested for this pathogen; new infections of hop plants by *P. humuli* are less likely to occur far from their inoculum source (Johnson et al. 1991). However, given the limited sample size per subpopulation, these results may be an artifact, and additional samples are needed to corroborate the patterns of genetic distribution observed in this study. Similarly, additional samples are also needed to corroborate the significant correlation detected between the genetic and geographic distance of *P. humuli* samples that suggest a higher probability for new *P. humuli* infections to occur close to their inoculum source.

In Oregon, Gent et al. (2019) found significant genetic differences between hop yards planted within 10 km of each other and a low amount of population differentiation between hop yards established from the same planting material. This suggests that infected hop plant material may be a more important source of primary inoculum than airborne migration for DM. Most of the hop yards sampled for this study were among the first yards establish from rhizomes in the mid-2000s, when commercial hop production returned to Michigan, and very few sources of propagation material were available (Sirrime et al. 2014). As the hop industry matured in the state, new sources of propagation material became available, but still in limited numbers. The lack of genetic differentiation among hop yards within regions in Michigan may be a result of the introduction of a reduced number of genotypes via planting material from the same origin source into multiple hop yards. The detection of the same MLGs in multiple hop yards and relatively low genetic diversity of north Michigan partially support this hypothesis, but further sampling and more information on sources of plant material are needed to verify this hypothesis in Michigan.

The restricted airborne dispersal of *P. humuli* sporangia compared with *P. cubensis* could be attributed to the cultivation practices and geographic distribution of hops. Generally, diseased basal shoots are close to the ground and during most of the season are sheltered within a canopy of healthy basal shoots that hamper dispersal (Johnson et al. 1991). Additionally, the area planted with hops in the state is equivalent to only 2% of the area planted with cucumbers. This lower host availability may result in reduced *P. humuli* sporangia production compared with *P. cubensis* and lower aerial exchange of MLG among subpopulations. The restricted exchange of MLG can result in genetically differentiated subpopulations (Milgroom 2015).

The relatively low level of genetic differentiation and genetic diversity detected within the *Pseudoperonospora* spp. populations of Michigan is consistent with the clonal reproductive mode of *P. cubensis* and inbreeding reproductive mode of *P. humuli* (Gent et al. 2019; Naegele et al. 2016). Although we estimated relatively low I_{AS} (65%) for both species, suggesting they experience a sexual phase, especially *P. humuli*, we believe these values are an artifact of our sampling strategy. Our samples were collected mainly from individual leaf lesions (cucurbits) or bulked inoculum of diseased shoots (hops), so it is likely that most of our *P. humuli* sample units may contain

multiple genotypes creating an effect of random mating (no linkage disequilibrium). An earlier study of *P. humuli* populations in the Pacific Northwest that used a sampling strategy similar to that in our study suggested that the *P. humuli* population of Oregon was reproducing sexually (Chee et al. 2006). However, in a genetic study of *P. humuli* using more precise sample units in the same region, GBS revealed strong evidence of linkage disequilibrium (Gent et al. 2019). This is consistent with the non-random mating expected from species with an asexually reproducing or highly inbred mode (Gent et al. 2019). Similarly, Wallace et al. (2020) provided evidence of nonrandom mating or recombination consistent with selfing or asexual reproduction for *P. cubensis* clade II. A strong signal of nonrandom mating for *P. cubensis* clade II was also detected in this study.

The use of more precise sample units (e.g., individual leaf lesions or single cells) and the sequencing of selectively neutral regions should be considered for studies focusing on capturing fine population structure (Grünwald et al. 2016). We included 94 neutral regions (genes) containing small SSR markers among the regions targeted for sequencing. However, the sequencing of these regions was of poor quality and we could not detect SSRs within them. Instead, we found 115 SNPs within these regions that were not significantly differentiated among or within subpopulations (data not shown). Most of the signal of population differentiation was found within 66% of the HQ SNPs used; these SNPs were contained within secreted proteins or effector genes that corresponded to 10% of the genes targeted. The SNPs contained within these regions are likely to be under strong selective pressure and may be acceptable to describe genetic differentiation among populations; however, these markers may limit the detection of ancient divergence among populations. We encourage the use of neutral polymorphic regions more evenly distributed across the genome for future genetic studies using TE, as they may provide more accurate estimates of population differentiation at multiple temporal and spatial scales (Milgroom 2015).

We also recommend the use of a higher number of baits for enrichment when working with low concentrations of target DNA and the quantification of plant and pathogen DNA, using qPCR before library preparation. More baits should increase the sequencing coverage across samples, facilitating the calling of variants. Similarly, the utilization of samples with a higher amount of pathogen DNA relative to the amount of contaminating DNA could increase the chances for the successful preparation of libraries using TE. Future studies should also include more technical replicates to assess error due to batch effects, as the high number of amplification cycles resulted in lower genotyping accuracy. The introduction of technical replicates from multiple generations of clones is also advised. This is critical when working with high-throughput sequencing data, due to the introduction of false mutations in the data that could create additional MLGs (Gent et al. 2019; Potapov and Ong 2017).

In summary, our results reveal the key strengths of TE for the genotyping of DM. This approach provides a solution for the genotyping of obligate biotrophic pathogens, for which high-throughput sequencing is typically constrained by the low amounts of target DNA and high amounts of nontarget (contaminating) DNA. TE provides a cost-effective approach for the genotyping of unpurified field samples and the assessment of sequence polymorphisms across a large number of individuals. This method could be adapted to a diverse group of pathogens, even without a reference genome. Future population studies using TE should carefully consider the sampling strategy and the size and complexity of the genomic regions targeted. Including technical replicates will also be important to ensure the accurate genotyping of the samples after enrichment and the reproducibility of the experiments.

MATERIALS AND METHODS

Sample collection and DNA extraction.

Over a total of 12 years (2007 to 2009, 2012 to 2013, 2015, and 2018 to 2019), *P. cubensis* sporangia were obtained from symptomatic cucurbit tissue collected in Michigan, other states (Indiana, Iowa, Ohio, Florida, and Wisconsin), and one Canadian province, Ontario (Fig. 1A; Supplementary Table S1). Approximately 85% of *P. cubensis* samples were processed by excising single lesions using a sterile scalpel. Sporangia from each lesion were harvested by vigorously shaking a 2-ml microcentrifuge tube containing the tissue and 1 ml of distilled H₂O for 30 s. Dislodged sporangia were pelleted (14,000 rpm for 5 min; 5424 Centrifuge, Eppendorf) and were subjected to DNA extraction. The remaining samples were processed by gently rinsing sporangia from infected leaves with multiple lesions, using a Preval spray power unit (Preval, Chicago) filled with distilled water as described by Mitchell et al. (2011). Dislodged sporangia were transferred into a 2-ml centrifuge tube, were pelleted (14,000 rpm for 5 min; 5424 Centrifuge, Eppendorf), and were subjected to DNA extraction.

P. humuli samples were processed using methods similar to those described by Chee et al. (2006). Sporangial suspensions were obtained from diseased basal shoots collected from six commercial hop yards and a hop research plot located in Michigan. The three hop yards in the north region were located within a radius of approximately 25 km (hop yards A, B, and C) (Fig. 1B). The two hop yards in the west region of the state were separated by approximately 100 km and were operated by the same producer (hop yards D and E) (Fig. 1B). The sixth commercial hop yard was located in the east region (hop yard F) and was located approximately 80 km from the research plot (hop yard G) located in the central region (Fig. 1B). Diseased basal shoots were brought to the laboratory, stems were placed into beakers of water, and shoots were individually covered with a plastic bag overnight, to induce sporulation. The following day, sporangia were washed from the abaxial leaf surface using a Preval spray power unit filled with distilled water. Sporangial suspensions from single shoots were transferred into 2-ml centrifuge tubes and were pelleted by centrifugation (14,000 rpm for 5 min). All DNA extractions were performed on the pelleted sporangial suspensions using the NucleoSpin Plant II isolation kit (Macherey-Nagel), following manufacturer instructions.

DNA library construction.

TE libraries. Whole DNA was extracted from 305 samples as described above and was quantified using a Qubit 2.0 fluorimeter (Thermo Fisher Scientific). Before fragmentation of DNA, using a M220 Focused-ultrasonicator (Covaris), the presence of *P. cubensis* and *P. humuli* in cucurbit or hop samples was confirmed by Sanger sequencing of the intergenic transcribed spacer ribosomal DNA region, as described by Quesada-Ocampo et al. (2012). A subset of five samples was divided into two individual tubes and was submitted independently for library preparation and enrichment. These samples were used to estimate the number of errors introduced into a sample during the TE workflow. Sequencing libraries were prepared using the KAPA HyperPrep library sequencing preparation kit (Kapa/Roche), following manufacturer recommendations, and an additional KAPA Pure Bead (Kapa/Roche) cleanup to ensure the libraries were free from any adapters. Insert size and library quality was verified using the Agilent 2100 Bioanalyzer (Agilent Technologies). To optimize the use of the MYbaits custom kit reactions (80-bp baits; Arbor Biosciences, formerly MYcroarray Inc.), good quality libraries were pooled into groups of four libraries for enrichment. Probes (baits) were designed according to manufacturer protocols (probe compatibility, repeat masking,

and melting temperature filters) to cover 118 population-informative DNA regions and 712 DNA regions annotated as protein genes at a 2× coverage (GitHub database). Protein gene sequences were obtained using bedtools 2.26.0 (Quinlan and Hall 2010), with coordinate information obtained from the genome annotation of *P. cubensis* available online from Oregon State University Libraries (Burkhardt et al. 2015). All genes annotated as extracellular toxins, hydrolytic enzymes, enzyme inhibitors, cell-entering RxLR effectors, cell-entering crinkler effectors, secreted proteins, and transcription factors were targeted for sequencing. These regions were selected to identify annotated genes containing SNPs that may play a role in controlling host-specificity and selectively neutral regions for population level analyses (Kitner et al. 2015; Quesada-Ocampo et al. 2012; Wallace and Quesada-Ocampo 2017).

Hybridization of DNA libraries and probes were performed at 65°C for 24 or 48 h (Supplementary Table S1). After hybridization, library pools were bound to Dynabeads MyOne Streptavidin C1 magnetic beads (Life Technologies) for enrichment. Amplification of the enriched libraries was performed using the 2× KAPA HiFi HotStart ReadyMix (Kapa/Roche) and 5 µM of each TruSeq forward and reverse primer. Amplification conditions were 98°C for 2 min, followed by six to 14 cycles of 98°C for 20 s, 60°C for 30 s, and 72°C for 60 s, and, then, a final extension of 72°C for 5 min. To retain libraries with at least 0.5 ng of DNA with fragments narrowly distributed around 300 bp, the libraries were quantified using a Qubit 2.0 fluorimeter (Invitrogen) and were analyzed using a Caliper LabChipGX HS DNA assay (PerkinElmer) after PCR. Based on these quantifications, the libraries were combined in equimolar amounts into three groups for sequencing. The first set of samples (TES1) included 62 enriched libraries with a hybridization time of 24 h. The second set (TES2) of samples included 72 enriched libraries with a hybridization time of 48 h (Supplementary Table S1). The third set (TES3) of samples included 24 enriched libraries with a hybridization time of 48 h (Supplementary Table S1). All sets were quantified using the Qubit 2.0 and the Kapa Biosystems Illumina library quantification qPCR kit. Each set was loaded onto one lane of an Illumina HiSeq 4000 flow cell and was sequenced in a 2 × 150 bp paired-end format.

Lc-WGS libraries. In order to assess the performance of TE in comparison with Lc-WGS and to estimate the amount of exogenous material in the DNA samples, 10 samples of genomic DNA directly extracted from sporangial suspensions derived from individual lesions were submitted to the Michigan State University Research Technology Support Facility for DNA library preparation and sequencing. Due to the low DNA concentration of some of the samples, the Rubicon ThruPLEX DNA-Seq library preparation kit (Takara Bio) was used to prepare the sequencing libraries. Library preparation was performed following manufacturer recommendations with an additional AMPureXP bead cleanup (Beckman Coulter). To retain libraries with at least 0.5 ng of DNA with fragments narrowly distributed around 300 bp, the libraries were quantified using a Qubit 2.0 fluorimeter (Invitrogen) and were analyzed using a Caliper LabChipGX HS DNA assay (PerkinElmer) after PCR. Based on these quantifications, the libraries were grouped in equimolar amounts into two pools that were quantified using the Kapa Biosystems Illumina library quantification qPCR kit (Kapa/Roche). Each pool was loaded into one lane of an Illumina MiSeq standard flow cell (v2) (Illumina) and sequencing was performed in a 2 × 150 bp paired-end format, using a v2 500 cycle MiSeq reagent cartridge (Illumina).

Genotyping and variant calling.

In order to assess the performance of TE for the genotyping of *P. cubensis* and *P. humuli*, we analyzed the data in

combination with previously published WGS data generated for a comparative genomic analyses of *P. cubensis*. This data set included one *P. humuli* sample collected from hop in Oregon and nine *P. cubensis* samples collected from different cucurbits in Alabama, South Carolina ($n = 2$), Florida, North Carolina ($n = 2$), Georgia, California, and New York. The datasets can be found in Bioprojects PRJNA360426 (Thomas et al. 2017) and PRJNA675756 (this project), available online through the National Center for Biotechnology Information. Sequencing raw reads of all libraries were trimmed and quality was filtered, using Trimmomatic 0.32 (Bolger et al. 2014). The publicly available *P. cubensis* (Savory et al. 2012b) and *C. sativus* genome sequences were used as references for alignment of the reads. To estimate plant contamination, HQ reads were initially aligned to the *C. sativus* genome, using the Burrows-Wheeler aligner (BWA-0.7.17) (Li and Durbin 2009). All reads including those that aligned to the reference genome of *C. sativus* were then aligned to a reference genome of *P. cubensis* containing only the contigs in which the target genes are located, hereafter referred to as the target reference genome of *P. cubensis*. Clean reads of Bioproject PRJNA360426 were also aligned to this subset of the genome. Samtools-0.1.19 (Li et al. 2009) was used to sort the alignments and PCR duplicates were removed with Picard. Coverage depth was calculated using BEDTools-2.27.1.

The next generation sequencing plugin (NGSEP) for analysis of high-throughput sequencing data was used for calling SNPs and producing the variant call format file (VCF) (Perea et al. 2016). The VCF file generated was filtered, using the NGSEP plugin, in order to retain the most informative SNPs of sufficient quality. Only biallelic SNPs, present in 70% of samples, with a minimum quality of 40 were retained. Indels and copy number variants were excluded to facilitate the analysis. The resulting VCF file was imported into R using the package vcfr (version 1.5.0) (Knaus and Grünwald 2017). Missing genotype data were deleted before further analysis (Tabima et al. 2018). Samples with more than 40% of missing variants were removed from the analysis to retain the maximum number of SNPs. Variants with more than 75% of missing information across samples were also removed. To increase the stringency for the genotyping, we excluded all the samples with an average coverage below 50×. A nonparametric Kruskal-Wallis rank-sum test was used to assess the differences among the mean coverage depth across libraries.

Population assignment.

P. cubensis samples were assigned to several different subpopulations depending on the analyses. To study the distribution of the genetic variation by host, *P. cubensis* samples were assigned to subpopulations according to host species. To study the distribution of the genetic variation by geographic location, only the *P. cubensis* samples collected from *Cucumis* spp. were used. The samples collected in the states of Michigan, Wisconsin, Indiana, and Ohio and the Canadian province of Ontario were assigned to the Midwest subpopulation (Fig. 1C). Midwest samples excluding those from Michigan were assigned to the Midwest-w/o-MI subpopulation (Fig. 1C). Samples from Michigan were further divided into three subpopulations according to the region of collection (east, west, and central) (Fig. 1C). Samples collected from Iowa, Alabama, Oregon, California, South Carolina, Florida, New York, North Carolina, and Georgia were assigned to their respective state of origin.

P. humuli samples were assigned to either the north or west subpopulation based on their location of collection in Michigan (Fig. 1B and C; Table 2; Supplementary Table S1). Three hop yards (A, B, and C) were located within a radius of approximately 25 km and comprised the north subpopulation (Fig. 1C). Hop yards D and E in southwest Michigan are operated by one producer and make up the west subpopulation (Fig. 1C).

Representative samples from central and east Michigan and Oregon were also included in our analysis.

Genetic differentiation and population structure.

The filtered SNPs were used to construct a genetic tree based on Nei's genetic distance (Nei 1972) and, to visualize the genetic relationship among all samples; 1,000 bootstrap replicates were performed to obtain branch support (Kamvar et al. 2015). Clone-correction was performed by collapsing samples to the average genetic distance detected between technical replicates of sequencing. For this, the "mlg.filter" function of the R package Poppr (version 2.4.1) was used with a bitwise distance equal to 0.02 and the farthest neighbor algorithm. Ordination plots based on principal component analysis (PCA) were also constructed to visualize the differentiation of samples within and among subpopulations from different host or regions across the Midwest (Fig. 1C). Three distinct PCAs were constructed to visualize the relationship among i) all *P. cubensis* and *P. humuli* samples by host, ii) all *P. cubensis* samples collected from *C. sativus*, and iii) all *P. humuli* samples collected in Michigan and Oregon. The PCAs were performed using the "glPca" function of R package ADEGENET (Jombart and Ahmed 2011), and the ordination plots were constructed using the R package Ggplot2. Samples were colored according to host or subpopulations, and ellipses of 90% confidence intervals were drawn on the ordination plots.

Genetic differentiation between and within subpopulations was quantified using hierarchical F_{ST} (hierfstat version 0.04.22) (Goudet 2005) and AMOVA (poppr version 2.4.1) (Kamvar et al. 2014) in R version 3.6.1. To test the hypothesis that the *Pseudoperonospora* spp. populations are differentiated by host, we performed an AMOVA and calculated F_{ST} statistics using all the samples collected by *P. cubensis* and *P. humuli* grouped by host. Similarly, to test the hypothesis that the *P. cubensis* subpopulation from *C. sativus* in Michigan is differentiated by geographic location, we performed an AMOVA and calculated F_{ST} statistics using the samples from the Michigan subpopulations (east, west, and central) and the Midwest-w/o-Mi subpopulation. Population differentiation among the *P. humuli* subpopulations of Michigan (north and west) was analyzed in the same way. A Mantel test (Mantel 1967) was performed to evaluate the relationship between the genetic distance of the samples and the physical distance of the location where the samples originated. This analysis was performed for each species independently (*P. cubensis* and *P. humuli*), using only the samples collected in Michigan. The Mantel test was performed as described by Gent et al. (2019), using the "mantel.test" function of the ade4 package (Dray and Dufour 2007). The genetic and Euclidean-geographic distance matrices were created with the function "bitwise.dist" (Kamvar et al. 2014) and "dist", respectively. Physical distances were calculated using the coordinates of each sample location. All scripts and functions for SNP filtering and population analyses are deposited in the GitHub database. These scripts are annotated to provide more in-depth information for how quality control was performed.

Genotypic diversity.

Genotypic diversity was determined using the filtered SNP data (not clone-corrected) using the R package Poppr (Kamvar et al. 2014). The diversity estimates calculated included genotypic richness (MLG), eMLG based on rarefaction, E_s , and the Shannon-Wiener (Shannon 1948), Stoddart and Taylor (Stoddart and Taylor 1988), and Simpson (Simpson 1949) indices. The within-population gene diversity (H_{exp}) (Nei 1972) that describes the proportion of heterozygous genotypes was calculated using the function "basic.stats" of the R package hierfstat. The diversity estimates were calculated using i) all *P. cubensis*

and *P. humuli* samples grouped by host, ii) all *P. cubensis* samples collected from *C. sativus* in each Michigan subpopulation and the Midwest-w/o-Michigan subpopulation, and iii) the *P. humuli* samples from the north and west subpopulations of Michigan. Only categories or subpopulations with at least five individuals were included for this analysis.

Reproductive mode.

To infer the predominant reproductive mode of *P. cubensis* and *P. humuli* in Michigan (e.g., sexual, clonal, or mixed), we calculated the I_A (Brown et al. 1980; Milgroom 1996) according to Tabima et al. (2018), using only the *P. cubensis* and *P. humuli* samples collected in the state from *C. sativus* and *H. lupulus*, respectively. For this, we calculated the I_A value, using a subset of 1,000 random HQ SNPs and compared it with the observed values of simulated populations with 0, 25, 50, 75, and 100% linkage. The mean I_A values of each species were compared independently against the mean I_A values of the simulated populations. Simulations were conducted with a dataset consisting of 2,514 loci analogous to the observed *P. cubensis* and *P. humuli* data. A nonparametric Kruskal-Wallis rank-sum test was used to assess the differences among means for all I_A distributions. Multiple comparisons between I_A distribution were performed using the nonparametric Kruskal-Wallis test and Tukey's honest significant difference test.

ACKNOWLEDGMENTS

We thank the following undergraduate research assistants: J. Chan and A. Job. We also thank D. Higgins for helpful discussions.

AUTHOR-RECOMMENDED INTERNET RESOURCES

GitHub database: https://github.com/jcbello131/TE_Pseudoperonospora
Oregon State University Libraries dataset:
<https://dx.doi.org/10.7267/N9TD9V7M>
Picard: <https://broadinstitute.github.io/picard>

LITERATURE CITED

- Ali, S., Gautier, A., Leconte, M., Enjalbert, J., and de Vallavieille-Pope, C. 2011. A rapid genotyping method for an obligate fungal pathogen, *Puccinia striiformis* f.sp. *tritici*, based on DNA extraction from infected leaf and multiplex PCR genotyping. BMC Res. Notes 4:240.
- Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R. H. Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den Ackerveken, G., Jones, J. D. G., McDowell, J. M., Beynon, J., and Tyler, B. M. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330:1549-1551.
- Bello, J. C., Sakalidis, M. L., Perla, D. E., and Hausbeck, M. K. 2020. Detection of airborne sporangia of *Pseudoperonospora cubensis* and *P. humuli* in Michigan using Burkard spore traps coupled to quantitative PCR. Plant Dis. 105:1373-1381.
- Blum, M., Waldner, M., Olaya, G., Cohen, Y., Gisi, U., and Sierotzki, H. 2011. Resistance mechanism to carboxylic acid amide fungicides in the cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Pest Manag. Sci. 67:1211-1214.
- Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
- Bourret, T. B., Choudhury, R. A., Mehl, H. K., Blomquist, C. L., McRoberts, N., and Rizzo, D. M. 2018. Multiple origins of downy mildews and mito-nuclear discordance within the paraphyletic genus *Phytophthora*. PLoS One 13:e0192502.

- Brown, A. H. D., Feldman, M. W., and Nevo, E. 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96:523-536.
- Burkhardt, A., Buchanan, A., Cumbie, J. S., Savory, E. A., Chang, J. H., and Day, B. 2015. Alternative splicing in the obligate biotrophic oomycete pathogen *Pseudoperonospora cubensis*. *Mol. Plant-Microbe Interact.* 28:298-309.
- Burkhardt, A., and Day, B. 2016. Transcriptome and small RNAome dynamics during a resistant and susceptible interaction between cucumber and downy mildew. *Plant Genome* 9.
- Carleson, N. C., Fieland, V. J., Scagel, C. F., Weiland, J. E., and Grünwald, N. J. 2019. Population structure of *Phytophthora plurivora* on rhododendron in Oregon nurseries. *Plant Dis.* 103:1923-1930.
- Chee, H. Y., Nelson, M. E., Grove, G. G., Eastwell, K. C., Kenny, S. T., and Klein, R. E. 2006. Population biology of *Pseudoperonospora humuli* in Oregon and Washington. *Plant Dis.* 90:1283-1286.
- Choi, Y.-J., Hong, S.-B., and Shin, H.-D. 2005. A re-consideration of *Pseudoperonospora cubensis* and *P. humuli* based on molecular and morphological data. *Mycol. Res.* 109:841-848.
- Choi, Y.-J., and Thines, M. 2015. Host jumps and radiation, not co-divergence drives diversification of obligate pathogens. A case study in downy mildews and Asteraceae. *PLoS One* 10:e0133655.
- Clark, M. J., Chen, R., Lam, H. Y. K., Karczewski, K. J., Chen, R., Euskirchen, G., Butte, A. J., and Snyder, M. 2011. Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* 29:908-914.
- Cruz-Dávalos, D. I., Llamas, B., Gaunitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Scheu, A., Beneke, N., Ludwig, A., Cooper, A., Willerslev, E., and Orlando, L. 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* 17:508-522.
- Cui, C., Herlihy, J. H., Bombarely, A., McDowell, J. M., and Haak, D. C. 2019. Draft assembly of *Phytophthora capsici* from long-read sequencing uncovers complexity. *Mol. Plant-Microbe Interact.* 32:1559-1563.
- Dray, S., and Dufour, A. 2007. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* 22:1-20.
- Faircloth, B. C., Branstetter, M. G., White, N. D., and Brady, S. G. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol. Ecol. Resour.* 15:489-501.
- Fletcher, K., Gil, J., Bertier, L. D., Kenefick, A., Wood, K. J., Zhang, L., Reyes-Chin-Wo, S., Cavanaugh, K., Tsuchida, C., Wong, J., and Michelmore, R. 2019. Genomic signatures of heterokaryosis in the oomycete pathogen *Bremia lactucae*. *Nat. Commun.* 10:2645.
- Gent, D. H., Adair, N., Knaus, B. J., and Grünwald, N. J. 2019. Genotyping-by-sequencing reveals fine-scale population differentiation in populations of *Pseudoperonospora humuli*. *Phytopathology* 109:1801-1810.
- Gent, D. H., Nelson, M. E., Farnsworth, J. L., and Grove, G. G. 2009. PCR detection of *Pseudoperonospora humuli* in air samples from hop yards. *Plant Pathol.* 58:1081-1091.
- Gent, D. H., O'camb, C. M., and Farnsworth, J. L. 2010. Forecasting and management of hop downy mildew. *Plant Dis.* 94:425-431.
- Gisi, U., and Sierotzki, H. 2008. Fungicide modes of action and resistance in downy mildews. *Eur. J. Plant Pathol.* 122:157-167.
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27:182-189.
- Goldenhar, K. E., and Hausbeck, M. K. 2019. Fungicides for control of downy mildew on pickling cucumber in Michigan. *Plant Health Prog.* 20:165-169.
- Goudet, J. 2005. Hierfstat, a package for *r* to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5:184-186.
- Grünwald, N. J., Everhart, S. E., Knaus, B. J., and Kamvar, Z. N. 2017. Best practices for population genetic analyses. *Phytopathology* 107:1000-1010.
- Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* 54:323-346.
- Johnson, D. A., Alldredge, J. R., Allen, J. R., and Allwine, R. 1991. Spatial pattern of downy mildew in hop yards during severe and mild disease epidemics. *Phytopathology* 81:1369-1374.
- Jombart, T., and Ahmed, I. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070-3071.
- Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C. H., McGuire, B., Gubler, W. D., Walker, M. A., and Cantu, D. 2014. Adaptive genomic structural variation in the grape powdery mildew pathogen, *Erysiphe necator*. *BMC Genomics* 15:1081.
- Jouet, A., Saunders, D. G. O., McMullan, M., Ward, B., Furzer, O., Jupe, F., Cevik, V., Hein, I., Thilliez, G. J. A., Holub, E., van Oosterhout, C., and Jones, J. D. G. 2019. *Albugo candida* race diversity, ploidy and host-associated microbes revealed using DNA sequence capture on diseased plants in the field. *New Phytol.* 221:1529-1543.
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208.
- Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281.
- Kanetis, L., Vasilou, A., Neophytou, G., Samouel, S., and Tsaltas, D. 2014. First report of downy mildew caused by *Peronospora belbahrii* on sweet basil (*Ocimum basilicum*) in Cyprus. *Plant Dis.* 98:283.
- Kitner, M., Lebeda, A., Sharma, R., Runge, F., Dvořák, P., Tahir, A., Choi, Y.-J., Sedláková, B., and Thines, M. 2015. Coincidence of virulence shifts and population genetic changes of *Pseudoperonospora cubensis* in the Czech Republic. *Plant Pathol.* 64:1461-1470.
- Knaus, B. J., and Grünwald, N. J. 2017. vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17:44-53.
- Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., and Hendrickson, C. L. 2015. Overview of target enrichment strategies. *Curr. Protoc. Mol. Biol.* 112:7.21.1-7.21.23.
- Kunjeti, S. G., Anchieta, A., Martin, F. N., Choi, Y.-J., Thines, M., Michelmore, R. W., Koike, S. T., Tsuchida, C., Mahaffee, W., Subbarao, K. V., and Klosterman, S. J. 2016. Detection and quantification of *Bremia lactucae* by spore trapping and quantitative PCR. *Phytopathology* 106:1426-1437.
- Lane, C. R., Beales, P. A., O'Neill, T. M., McPherson, G. M., Finlay, A. R., David, J., Constantinescu, O., and Henricot, B. 2005. First report of Impatiens downy mildew (*Plasmopara obducens*) in the UK. *Plant Pathol.* 54:243.
- Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Lim, H. C., and Braun, M. J. 2016. High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Mol. Ecol. Resour.* 16:1204-1223.
- Lin, X., Song, T., Fairhead, S., Witek, K., Jouet, A., Jupe, F., Witek, A. I., Karki, H. S., Vleeshouwers, V. G. A. A., Hein, I., and Jones, J. D. G. 2020. Identification of AvrM1 from *Phytophthora infestans* using long read and cDNA pathogen-enrichment sequencing (PenSeq). *Mol. Plant Pathol.* 21:1502-1512.
- Ma, X., Shao, Y., Tian, L., Flasch, D. A., Mulder, H. L., Edmonson, M. N., Liu, Y., Chen, X., Newman, S., Nakitandwe, J., Li, Y., Li, B., Shen, S., Wang, Z., Shurtleff, S., Robison, L. L., Levy, S., Easton, J., and Zhang, J. 2019. Analysis of error profiles in deep next-generation sequencing data. *Genome Biol.* 20:50.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209-220.
- McCormack, J. E., Tsai, W. L. E., and Faircloth, B. C. 2016. Sequence capture of ultraconserved elements from bird museum specimens. *Mol. Ecol. Resour.* 16:1189-1203.
- Milgroom, M. G. 1996. Recombination and the multilocus structure of fungal populations. *Annu. Rev. Phytopathol.* 34:457-477.
- Milgroom, M. G. 2015. Population biology of plant pathogens: genetics, ecology, and evolution. The American Phytopathological Society, St. Paul, MN, U.S.A.
- Mitchell, M. N., O'camb, C. M., Grünwald, N. J., Mancino, L. E., and Gent, D. H. 2011. Genetic and pathogenic relatedness of *Pseudoperonospora cubensis* and *P. humuli*. *Phytopathology* 101:805-818.
- Naegle, R. P., Quesada-Ocampo, L. M., Kurjan, J. D., Saude, C., and Hausbeck, M. K. 2016. Regional and temporal population structure of *Pseudoperonospora cubensis* in Michigan and Ontario. *Phytopathology* 106:372-379.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106:283-292.
- Neufeld, K. N. 2017. Epidemiology of cucurbit downy mildew: predicting risk of disease infection and validation of the CDM ipmPIPE forecasting system. Ph.D. thesis. North Carolina State University, Raleigh, NC, U.S.A.

- Perea, C., De La Hoz, J. F., Cruz, D. F., Lobaton, J. D., Izquierdo, P., Quintero, J. C., Raatz, B., and Duitama, J. 2016. Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics* 17:498.
- Potapov, V., and Ong, J. L. 2017. Examining sources of error in PCR by single-molecule sequencing. *PLoS One* 12:e0169774.
- Purayannur, S., Cano, L. M., Bowman, M. J., Childs, K. L., Gent, D. H., and Quesada-Ocampo, L. M. 2020. The effector repertoire of the hop downy mildew pathogen *Pseudoperonospora humuli*. *Front. Genet.* 11:910.
- Quesada-Ocampo, L. M., Granke, L. L., Olsen, J., Gutting, H. C., Runge, F., Thines, M., Lebeda, A., and Hausbeck, M. K. 2012. The genetic structure of *Pseudoperonospora cubensis* populations. *Plant Dis.* 96:1459-1470.
- Quinlan, A. R., and Hall, I. M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.
- Rahman, A., Góngora-Castillo, E., Bowman, M. J., Childs, K. L., Gent, D. H., Martin, F. N., and Quesada-Ocampo, L. M. 2019. Genome sequencing and transcriptome analysis of the hop downy mildew pathogen *Pseudoperonospora humuli* reveal species-specific genes for molecular detection. *Phytopathology* 109:1354-1366.
- Rivera, Y., Salgado-Salazar, C., Gulya, T. J., and Crouch, J. A. 2016. Newly emerged populations of *Plasmopara halstedii* infecting rudbeckia exhibit unique genotypic profiles and are distinct from sunflower-infecting strains. *Phytopathology* 106:752-761.
- Savory, E. A., Adhikari, B. N., Hamilton, J. P., Vaillancourt, B., Buell, C. R., and Day, B. 2012a. mRNA-Seq analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection. *PLoS One* 7:e35796.
- Savory, E. A., Zou, C., Adhikari, B. N., Hamilton, J. P., Buell, C. R., Shiu, S.-H., and Day, B. 2012b. Alternative splicing of a multi-drug transporter from *Pseudoperonospora cubensis* generates an RXLR effector protein that elicits a rapid cell death. *PLoS One* 7:e34701.
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379-423.
- Sharma, R., Xia, X., Cano, L. M., Evangelisti, E., Kemen, E., Judelson, H., Oome, S., Sambles, C., van den Hoogen, D. J., Kitner, M., Klein, J., Meijer, H. J., Spring, O., Win, J., Zipper, R., Bode, H. B., Govers, F., Kamoun, S., Schornack, S., Studholme, D. J., Van den Ackerveken, G., and Thines, M. 2015. Genome analyses of the sunflower pathogen *Plasmopara halstedii* provide insights into effector evolution in downy mildews and *Phytophthora*. *BMC Genomics* 16:741.
- Simpson, E. H. 1949. Measurement of diversity. *Nature* 163:688.
- Sirrione, R., Lizotte, E., Brown, D., O'Brien, T., and Leach, A. 2014. Estimated costs of producing hops in Michigan. Extension report no. E3236. Michigan State University, Lansing, MI, U.S.A. Published online.
- Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W., Jr., McCormack, J. E., and Faircloth, B. C. 2017. High phylogenetic utility of an ultra-conserved element probe set designed for Arachnida. *Mol. Ecol. Resour.* 17:812-823.
- Stassen, J. H., Seidl, M. F., Vergeer, P. W., Nijman, I. J., Snel, B., Cuppen, E., and Van den Ackerveken, G. 2012. Effector identification in the lettuce downy mildew *Bremia lactucae* by massively parallel transcriptome sequencing. *Mol. Plant Pathol.* 13:719-731.
- Stoddart, J. A., and Taylor, J. F. 1988. Genotypic diversity: Estimation and prediction in samples. *Genetics* 118:705-711.
- Summers, C. F., Gulliford, C. M., Carlson, C. H., Lillis, J. A., Carlson, M. O., Cadle-Davidson, L., Gent, D. H., and Smart, C. D. 2015. Identification of genetic variation between obligate plant pathogens *Pseudoperonospora cubensis* and *P. humuli* using RNA sequencing and genotyping-by-sequencing. *PLoS One* 10:e0143665.
- Tabima, J. F., Coffey, M. D., Zazada, I. A., and Grünwald, N. J. 2018. Populations of *Phytophthora rubi* show little differentiation and high rates of migration among states in the western United States. *Mol. Plant-Microbe Interact.* 31:614-622.
- Thilliez, G. J. A., Armstrong, M. R., Lim, T.-Y., Baker, K., Jouet, A., Ward, B., van Oosterhout, C., Jones, J. D. G., Huitema, E., Birch, P. R. J. and Hein, I. 2019. Pathogen enrichment sequencing (PenSeq) enables population genomic studies in oomycetes. *New Phytol.* 221:1634-1648.
- Thines, M. 2014. Phylogeny and evolution of plant pathogenic oomycetes—A global overview. *Eur. J. Plant Pathol.* 138:431-447.
- Thomas, A., Carbone, I., Choe, K., Quesada-Ocampo, L. M., and Ojiambo, P. S. 2017. Resurgence of cucurbit downy mildew in the United States: Insights from comparative genomic analysis of *Pseudoperonospora cubensis*. *Ecol. Evol.* 7:6231-6246.
- United States Department of Agriculture. 2020. Vegetables Annual Summary 2019. United States Department of Agriculture, Washington, DC.
- Wallace, E. C., D'Arcangelo, K. N., and Quesada-Ocampo, L. M. 2020. Population analyses reveal two host-adapted clades of *Pseudoperonospora cubensis*, the causal agent of cucurbit downy mildew, on commercial and wild cucurbits. *Phytopathology* 110:1578-1587.
- Wallace, E. C., and Quesada-Ocampo, L. M. 2017. Analysis of microsatellites from the transcriptome of downy mildew pathogens and their application for characterization of *Pseudoperonospora* populations. *PeerJ* 5:e3266.
- Withers, S., Góngora-Castillo, E., Gent, D., Thomas, A., Ojiambo, P. S., and Quesada-Ocampo, L. M. 2016. Using next-generation sequencing to develop molecular diagnostics for *Pseudoperonospora cubensis*, the cucurbit downy mildew pathogen. *Phytopathology* 106:1105-1116.
- Wong, F. P., and Wilcox, W. F. 2001. Comparative physical modes of action of azoxystrobin, mancozeb, and metalaxyl against *Plasmopara viticola* (grapevine downy mildew). *Plant Dis.* 85:649-656.