Populations of *Phytophthora rubi* Show Little Differentiation and High Rates of Migration Among States in the Western United States

Javier F. Tabima,¹ Michael D. Coffey,² Inga A. Zazada,³ and Niklaus J. Grünwald^{3,†}

¹Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, U.S.A.; ²Department of Plant Pathology and Microbiology, UC Riverside, Riverside, CA 92521, U.S.A.; and ³Horticultural Crops Research Laboratory, USDA-ARS, Corvallis, OR 97330, U.S.A.

Accepted 22 January 2018.

Population genetics is a powerful tool to understand patterns and evolutionary processes that are involved in plant-pathogen emergence and adaptation to agricultural ecosystems. We are interested in studying the population dynamics of *Phytophthora rubi*, the causal agent of *Phytophthora* root rot in raspberry. *P. rubi* is found in the western United States, where most of the fresh and processed raspberries are produced. We used genotyping-by-sequencing to characterize genetic diversity in populations of *P. rubi* sampled in the United States and other countries. Our results confirm that *P. rubi* is a monophyletic species with complete lineage sorting from its sister taxon *P. fragariae*. Overall, populations of *P. rubi* show low genetic diversity across the western United States. Demographic analyses suggest that populations of *P. rubi* from the western United States are the source of pathogen migration to Europe. We found no evidence for population differentiation at a global or regional (western United States) level. Finally, our results provide evidence of migration from California and Oregon into Washington. This report provides new insights into the evolution and structure of global and western United States populations of the raspberry pathogen *P. rubi*, indicating that human activity might be involved in moving the pathogen among regions and fields.

Phytophthora rubi and its sister species *P. fragariae* are soil-borne pathogens affecting economically important Rosaceae species across the world. *P. rubi* is the most common soil-borne plant pathogen causing root rot disease in raspberry fields in the western United States (Gigot et al. 2013; Stewart et al. 2014; Wilcox et al. 1993), while *P. fragariae* is an important soil-borne pathogen to the strawberry industry as causal agent of red

stele disease of strawberry (Hickman 1940). *P. rubi* was found in 90% of the surveyed red raspberry fields in Washington state in 2013, with the potential of causing millions of dollars in losses to the industry annually (Gigot et al. 2013; Stewart et al. 2014). This pathogen is difficult to control, as it overwinters in the soil as oospores that can initiate epidemics in subsequent years. Not much is known about the reproductive strategy, genetic diversity, and population dynamics of this pathogen. It is currently unknown if *P. rubi*, a homothallic organism, is predominantly selfing or outcrossing. A recent study showed that *P. rubi* forms a single population across the states of California, Washington, and Oregon (Stewart et al. 2014), but there was no assessment of the degree of selfing or migration across populations at either the local or global scale.

Until recently, *P. rubi* was considered a variety of *P. fragariae* (*P. fragariae* var. *rubi*), as these two species are very similar morphologically and physiologically and are differentiated only by host preference (Man in't Veld 2007). Man in't Veld (2007) showed that both species have a different allozyme profile and are clustered into two distinct phylogenetic groups based on the mitochondrial cytochrome oxidase I (*COI*) gene of four strains. These results were used to promote *P. fragariae* var. *rubi* to the new taxonomic species *P. rubi* (Man in't Veld 2007). Further genetic analyses are needed to determine the degree of lineage sorting between these two species and to confirm the two-species hypothesis. Since a population level study has not been conducted, it is currently not clear whether these two species are reciprocally monophyletic (i.e., all isolates of *P. rubi* are more closely related to *P. rubi* than to *P. fragariae* and vice versa) and whether each of the lineages sort independently from the other.

We studied the population dynamics of the causal agent of *Phytophthora* root rot disease in red raspberries using population genetic approaches. Population genetic approaches have been applied to identifying putative centers of origin for various plant pathogens (Goss et al. 2014; Stukenbrock et al. 2007), to study the role of seed transmission in epidemics (Shah et al. 1995), and to determine the number of introductions or founder events in a forest epidemic (Kamvar et al. 2015a). Population genetics is a fundamental tool for understanding how a pathogen emerges, adapts, and evolves (Grünwald and Goss 2011; Grünwald et al. 2017; Milgroom 2015).

Contemporary genotyping techniques can be used to obtain a high number of single nucleotide polymorphisms (SNPs) at moderate cost (Baird et al. 2008; Boussau and Daubin 2010; Elshire et al. 2011; Grünwald et al. 2017; McCormack et al. 2013). We used genotyping-by-sequencing (GBS), a method that reduces the genome to fragments by using restriction

[†]Corresponding author: Niklaus J. Grünwald;
E-mail: nik.grunwald@ars.usda.gov

Funding: This work was supported by funds from the United States Department of Agriculture (USDA) Agricultural Research Service project 2072-22000-041-00-D and USDA National Institute of Food and Agriculture project 2010-511001-21649.

*The e-Xtra logo stands for “electronic extra” and indicates that six supplementary figures and one supplementary table are published online.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2018.

enzymes, to then massively sequence these fragments, using high throughput sequencing (HTS). The SNPs obtained from GBS can be used to infer population structure, linkage and its associations with reproductive strategies, and migration rates within and across populations (Baird et al. 2008; Davey et al. 2011; Elshire et al. 2011; Goss 2015; Grünwald et al. 2016). GBS or RADseq have recently been adopted to characterize a range of plant pathogens including, for example, *Fusarium graminearum* and *Verticillium dahliae* (Grünwald et al. 2017; Milgroom et al. 2014; Talas et al. 2016).

The goal of this research was to characterize the genetic structure of populations of *P. rubi* using genome-wide SNP data obtained from GBS. First, we tested the hypothesis of complete lineage sorting and reciprocal monophyly between *P. rubi* and its sister taxon *P. fragariae* to support the observation by Man in't Veld (2007) that these two taxa are distinct species. Second, we studied the population structure and migration rates of *P. rubi* at a global and at the western United States scale. Finally, we tested the hypotheses that populations of *P. rubi* are predominantly clonal, are differentiated by geographic region, and that migration occurs among populations. Our work provides novel insights into the evolutionary history and the population dynamics of *P. rubi*.

RESULTS

Number of variants obtained by GBS.

A total of 183 isolates were genotyped using GBS, including 45 *P. fragariae* and 138 *P. rubi* isolates (Table 1; Supplementary Table S1). Of a total of 705,333 raw SNP variants, 5,632 variants were retained after filtering by depth (DP), mapping quality (MQ), missing data, and minimum allele frequency. Six isolates were removed due to poor sequence depth. This filtered set of variants represented isolates from seven countries for *P. fragariae* and from five countries and six states within the United States for *P. rubi* (Table 1). A mean DP of 11× was obtained for all variants, with 54% missing data overall for the filtered dataset. We subsequently limited the data to only *P. rubi* isolates from California, Oregon, and Washington, collected from 10 grower fields: three from California, three from Oregon, and five from Washington. Each of the states was considered a population. Data were further filtered for the population analysis from the western United States isolates to retain high-quality variants by filtering out variants with more than 20% missing data, resulting in a set of 230 SNP loci for 94 isolates.

High genetic differentiation and complete lineage sorting among *P. fragariae* and *P. rubi*.

We tested the hypothesis of reciprocal monophyly between populations of *P. fragariae* and *P. rubi* by constructing a phylogenetic tree and randomly resampling SNPs to evaluate lineage sorting. We first constructed a maximum likelihood (ML) phylogenetic tree using 5,632 variants that supported two monophyletic clades for each taxon. We then reconstructed 1,000 neighbor-joining (NJ) trees using 1,000 subsets of 200 random SNP variants each, to test support for reciprocal monophyly using different genetic variants across these two clades. These 1,000 NJ trees were overlaid on top of the ML tree using DensiTree to detect evidence of incomplete lineage sorting between the *P. rubi* and *P. fragariae* clades (Bouckaert 2010). The NJ trees plotted in DensiTree yielded similar results to the ML tree, separating *P. rubi* and *P. fragariae* isolates into two distinct clusters with 100% support for reciprocal monophyly between *P. rubi* and *P. fragariae* (Fig. 1). Several small subclades were distinguished within *P. rubi* and within *P. fragariae* populations that were a result of a reduced number of variants due to filtering (90 and 100% support). However, the bulk of the isolates within each species formed one clade with little variation and 100% support. Additionally, we calculated population differentiation

using *Gst* (Nei 1972, 1973) and *G'st* (Hedrick 2005) for the filtered dataset. Both indices indicated a very high population differentiation between these two species (*Gst* = 0.42, *G'st* = 0.48), further supporting two genetically isolated species with independent lineage sorting and reciprocal monophyly. Both results support complete lineage sorting between populations of *P. rubi* and *P. fragariae* as observed by Man in't Veld (2007).

Global populations of *P. rubi* show little differentiation and possibly originate from the western United States.

We tested the hypothesis of genetic differentiation across populations of *P. rubi* on a global scale. A population is defined here as the geographic location where the isolates was obtained from (California, Oregon, Washington, New York, Ohio, or Europe). A ML phylogenetic tree showed no clustering by geographic location (Supplementary Fig. S1). A discriminant analysis of principal components (DAPC) further supported this lack of population structure by showing overlapping ellipses representing 95% of the isolates from each of the populations, including the European Union, California, Washington, and Oregon, except for isolates from New York. FastSTRUCTURE returned a range of genetic clusters (K) ranging from K = 1 to K = 4 (Supplementary Fig. S2). The fastSTRUCTURE plots lacked a correlation between cluster assignment and geographic population. The results from the phylogenetic, DAPC, and fastSTRUCTURE analyses grouped the California isolates with some of the European Union and New York isolates, suggesting gene flow between these populations. We tested this hypothesis by evaluating 11 demographic scenarios of migration of *P. rubi* using DIYABC (Supplementary Fig. S3). The demographic analyses were performed with a subset of 12 random isolates from each of the western United States populations to reduce sampling bias. The DIYABC scenario with the highest posterior probabilities supported recent migration from California toward the European Union and showed Washington as an admixed population (posterior probabilities observed for scenario 1: 0.27 ± 0.02, scenario 6: 0.27 ± 0.02, and scenario 7: 0.35 ± 0.2). All scenarios showed recent migration from California to the European Union, admixture in Washington from Oregon and California populations, and differed in support for an ancestral population that was either unsampled, from Oregon, or from California.

Phytophthora rubi populations in the western United States are predominantly clonal.

To test the hypothesis that *P. rubi* populations are predominantly clonal, we calculated the index of association (*I_A*)

Table 1. Origin of isolates obtained for genotyping populations of *Phytophthora fragariae* and *P. rubi*

Species	Country	State	Isolates
<i>P. fragariae</i>	Canada	–	3
	England	–	5
	France	–	3
	Germany	–	1
	Scotland	–	12
	Switzerland	–	1
	United States	–	20
	Total <i>P. fragariae</i>		45
<i>P. rubi</i>	England	–	1
	France	–	4
	Germany	–	3
	Netherlands	–	1
	Scotland	–	3
	United States	California	42
	United States	New York	5
	United States	Ohio	1
	United States	Oregon	34
	United States	Washington	44
Total <i>P. rubi</i>			138
Total of all			183

across all loci and compared the observed value against simulations for populations under strong linkage (100% linked loci), moderate linkage (75 and 50% linked loci), and no linkage (0 loci under linkage) (Fig. 2). The Shapiro-Wilk's normality test showed that all observed and simulated distributions follow normality ($P < 0.1$), with the exception of the simulated 75% linkage dataset ($P = 0.23$, $W = 0.98$). Hence, we performed a non-parametric analysis of variance (ANOVA) analog and a Kruskal-Wallis rank sum test to test for significant differences across treatments. The Kruskal-Wallis test indicated significant differences across I_A distributions ($P < 0.05$, $\chi^2 = 337.17$). A nonparametric multiple comparison of ranks using Kruskal-Wallis was then performed to test for differences across mean ranks. The rank comparison yielded the same results as a Tukey's honest significant difference (HSD) test. The *P. rubi* sample mean was situated between the simulated 75% linkage data and the simulated 50% linkage data. This indicated that populations of *P. rubi* are predominantly clonal and/or selfing.

Population structure of *P. rubi* in the western United States.

We tested the hypothesis that populations of *P. rubi* in the western United States are differentiated. A total of 94 isolates

for which we obtained 230 high quality SNP variants were sampled. The ML tree reconstruction and DAPC clustering suggested low levels of genetic differentiation across populations and fields (Fig. 3; Supplementary Fig. S4). The ML tree showed two clades with isolates from Washington clustering with either isolates from Oregon or California (Fig. 3A). The intermediate clustering of Washington between Oregon and California was also supported in the DAPC analysis (Fig. 3B). Pairwise G_{st} similarly resulted in low genetic differentiation among populations. The G_{st} estimates indicated low differentiation between Oregon and Washington ($G_{st} = 0.024$; $G'_{st} = 0.087$), California and Washington ($G_{st} = 0.023$; $G'_{st} = 0.042$), and California and Oregon populations ($G_{st} = 0.093$; $G'_{st} = 0.143$). The analysis of molecular variance (AMOVA) results indicated little population differentiation across states ($\phi_{st} = 0.098$) and found 69% variation at the local state level ($\phi_{st} = 0.228$), with no significant differences between states ($P = 0.083$) (Table 2).

We conducted a DIYABC analysis to infer the most likely demographic history of the populations collected from California, Oregon, and Washington, testing five scenarios (Fig. 4). The best supported scenarios, scenarios 3 and 5, showed Washington as an admixed population (Fig. 4). Scenarios 3 and 5 resulted in the highest posterior probabilities (0.265 ± 0.02

Fig. 1. Genetic differentiation of *Phytophthora rubi* and *P. fragariae* populations. A maximum likelihood tree was reconstructed using 5,632 single nucleotide polymorphism (SNP) variants (outlined in black). A total of 1,000 neighbor joining tree reconstructions were made with 200 randomly subsampled SNPs and were drawn using DensiTree (shaded). This analysis shows complete lineage sorting between *P. rubi* and *P. fragariae* (100% support).

and 0.264 ± 0.01 , respectively) and were selected as the most plausible scenarios to explain the demographic history of the western United States populations.

A migration analysis was conducted using *migrate-n* to determine the degree of migration among the western United States samples. The results are presented as mutation-scaled migration rates and show a higher migration rate toward Washington from California ($M = 375$ [25 to 75% cylindrical

inclusion: 363.3 to 386.3]) and toward Washington from Oregon ($M = 600$ [594.7 to 642]) than from Washington toward Oregon ($M = 130$ [125.3 to 141.3]), from Washington toward California ($M = 155$ [144.7 to 162.3]), or from Oregon toward California ($M = 210$ [205.3 to 238]), and vice versa ($M = 125$ [109 to 137]).

We also evaluated the degree of linkage among loci and genetic admixture at the field level (or subpopulation), in order to determine if there is evidence of isolates from the same

Fig. 2. Estimation of the degree of linkage disequilibrium in *Phytophthora rubi* populations based on observed and simulated distributions of the index of association (I_A). The boxplot shows the observed distribution of I_A for the population of *P. rubi* compared with the distribution of I_A values for simulated populations with no linkage and 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. rubi* dataset is grouped with the simulated 75% linkage data, supporting a mostly clonal mode of reproduction in United States populations of *P. rubi* in the western United States.

Fig. 3. Population structure of *Phytophthora rubi* in the western United States in the states of California (CA), Oregon (OR), and Washington (WA). Two clustering methods indicate that the populations of *P. rubi* are not clustering by geographic origin. **A**, Maximum likelihood tree with 1,000 bootstrap replicates. There is poor support for clustering of isolates by state of origin. **B**, Scatterplot from a discriminant analysis (DA) of principal components. The ellipses represent the maximum area spanned by 95% of the data in a population by state of origin. There is little differentiation between populations sampled in the three states.

grower field having population assignment to other grower fields as a result of migration among fields. I_A was calculated across fields and resulted in high values of I_A , representing high linkage for most of the fields, with the exception of Washington fields 7 and 8 (Supplementary Fig. S5). A DAPC was performed across all grower fields to detect admixture (Supplementary Fig. S6). The DAPC results showed that the raspberry fields across California include many isolates with subpopulation assignments to fields within California, with the exceptions of isolates 4838 (most plausible assignment to field 8 in Washington) and 4801 (most plausible assignment to field 5 in Oregon). A similar pattern was seen in isolates from fields in Oregon, where most of the isolates were assigned to fields located in Oregon, with the exception of isolates 5086 (most plausible assignment to fields 1 and 2 in California) and 4815 and 4832 (most plausible assignment to field 8 in Washington). We focused on the results from field 7 in Washington. This field indicated high admixture across all isolates. These admixture events have a population assignment from fields in California and Oregon despite being present in a Washington field (Fig. 5). Isolates from field 7 had assignments with high posterior probabilities to fields in California (isolates 5294, 5295, and 5296) and Oregon (isolates 4933, 5094, and 5293), supporting the admixed nature of the isolates. Additionally, fields 7 and 8 were the only fields that resulted in an observed I_A in which the hypothesis of linkage was rejected, but field 7 differed from field 8, as field 7 did not include any isolates that were predominately assigned to a Washington field, while all the isolates from field 8 were predominately assigned to field 8. Field 7 stands out among the subpopulations included in this study as being more diverse, with migration from out of state fields.

DISCUSSION

Little is known about the evolutionary history of the raspberry pathogen *P. rubi*. Our work had two components. First, we wanted to know if *P. rubi* and the strawberry pathogen *P. fragariae* still show evidence of gene flow or if lineage sorting is complete. The prior work by Man in't Veld (2007) was based on a few isolates and one phylogenetic locus, and it was not clear to us if gene flow among these taxa might still occur. Our results support the findings of Man in't Veld (2007)

that *P. rubi* and *P. fragariae* are reproductively isolated. The phylogenetic reconstructions and population differentiation indices revealed that both species are clearly separated from each other. Although these two pathogens coexist in the same agricultural systems and microclimates suitable for small fruit production, development of management strategies and breeding for resistance must be species specific. Second, we were interested in understanding the evolutionary history and population genetic structure of *P. rubi* globally, where we could only source limited isolates, and more intensively in the western United States, where we could sample more deeply. We analyzed the degree of selfing, structure, and migration. Several results stand out.

Our analysis revealed a high degree of linkage among SNP markers, and these results are in line with a pathogen that can reproduce both clonally and sexually. Given that *P. rubi* is homothallic, sexual reproduction likely results in selfing and rarely in outcrossing. Both selfing and clonal reproduction might contribute to the observed linkage disequilibrium. Little is known about the linkage in other homothallic *Phytophthora* species. Recent work based on the index of recombination showed that populations of *P. plurivora* were predominately clonal (Schoebel et al. 2014).

Analysis of global populations, albeit based on a limited number of isolates, provides evidence for migration of *P. rubi* between the United States and Europe. The phylogenetic tree reconstruction of the filtered *P. rubi* variants showed that the populations from California are genetically most similar to the populations from the European Union. The DAPC results corroborate this finding, as do the population assignment plots generated by fastSTRUCTURE, in which the isolates from California and the European Union group with each other. The demographic models compared in DIYABC indicate that a recent migration from California to the European Union is the most plausible scenario explaining the genetic similarities between populations of *P. rubi* from both continents, indicating that California might be a source of the pathogen for Europe via possible transport of contaminated plant material (i.e., bare rooted plants or roots grown in soil with *P. rubi* oospores). California leads the production of red raspberries in the United States, with more than \$460 million in revenue in 2015 compared with Oregon (\$7 million dollars) and Washington (\$89 million dollars) (National Agricultural Statistics Service 2017). Movement of raspberry planting material could be contributing to the movement of the pathogen when infected plants are transported across states and countries, leading to dispersion of the pathogen across the globe. Keep in mind that the number of variants we used for the demographic reconstruction was very small (54 variants across 48 isolates) in order to have balanced sample sizes based on the limited European isolates; thus, we feel that additional sampling in Europe is needed to provide more confidence in the demographic patterns we observed. The DIYABC analysis could not differentiate between two candidate ancestral populations (Oregon versus California). A better sampling worldwide is needed to validate our

Table 2. Analysis of molecular variance (AMOVA) summary of the genetic variation in *Phytophthora rubi*^a

AMOVA	df	SS	MS	CoV (%)	ϕ
Between states	2	297.23	148.61	22	0.304
Between isolates within states	91	776.76	8.53	7	0.098
Within isolates within states	94	658.87	7.009	69	0.228

^a df = degrees of freedom; SS = sum of squares; MS = mean squares; CoV = percentage of component of variance; ϕ = degree of differentiation between population divisions.

Fig. 4. Demographic scenarios tested using DIYABC for western United States populations of *Phytophthora rubi* from California (CA), Oregon (OR), and Washington (WA). Scenarios 1 and 2 reflect the probable hypotheses of population isolation between the states. Scenarios 4 and 5 show Washington as an admixed population with a different state as a root. Scenario 4 assumes that the root is in an unknown location.

results and or provide new insights into the intercontinental migration of *P. rubi* populations.

We observed little structure in the Oregon, California, and Washington populations of *P. rubi*. The phylogenetic reconstruction and the DAPC analysis showed that most isolates from Washington were clustered with either isolates from California or Oregon, indicating lack of population structure. These results are supported by the analyses of population differentiation based on *Gst*. The AMOVA results revealed a similar pattern, in which no significant genetic differentiation was observed among populations. These results provide evidence to suggest that *P. rubi* forms one large, mostly clonal population across the western United States.

Our migration analysis also suggests that Washington might be a sink of the raspberry pathogen *P. rubi*, with migration from Oregon and California. We present multiple sources of evidence for an admixed Washington population. The phylogenetic reconstruction of the western United States populations showed that the Washington isolates cluster within both Oregon and California clades (Fig. 3A). The DAPC results showed a similar pattern of no exclusive clustering among Washington

isolates, where the ellipses that represent 95% of the isolates of each population overlap with both the California and Oregon ellipses (Fig. 3B). In addition, the demographic analysis obtained with DIYABC also supports Washington as an admixed population. The most supported DIYABC scenarios consider Washington an admixed population from isolates from California and Oregon. Based on a migration analysis, Washington received up to four times the influx of migrants than any of the other two states. These patterns are also observed at the field level.

The analysis of isolates across fields in the western United States showed evidence of low population structure and migration across fields. The ML phylogenetic reconstruction did not result in any monophyletic clade comprised of all isolates from the same field. Furthermore, the DAPC showed admixed isolates in every field, suggesting inoculum movement between fields within and across states. These patterns were most notable in the Washington fields, where isolates from Washington had a higher probability of assignment to populations from fields in California or Oregon. In addition, the low I_A values within Washington fields provides more evidence for the hypothesis of high migration toward Washington fields. The

Fig. 5. Example of admixture of *Phytophthora rubi* in a Washington (WA) raspberry field based on the posterior probability of population assignment for field 7 in Washington. Bar plots represent the probability of assignment of each sample to another field as indicated in the legend. Field 7 contains most isolates assigned to out-of-state locations. Sample 4993 has a higher posterior probability assignment to Oregon (OR) field 5, while the other isolates seem to be mostly assigned to California (CA) fields, except for sample 5094, which has equivalent assignments to fields across the western United States.

nonsignificant I_A values within Washington fields can only be explained by migration from other sources, given that this pathogen is predominantly clonal or selfing in the majority of fields sampled (Fig. 2). Our interpretation is that admixture can make populations appear sexual. Field 7 in Washington is a good example, with evidence of migration and admixture with other states; this field does not form an exclusive cluster in the phylogenetic reconstruction and has a low, nonsignificant I_A value. The assignment using DAPC of field 7 isolates showed higher probabilities of assignment to Oregon or California fields; sample 4993 had over 60% assignment to field 5 in Oregon, whereas any of the other isolates had a similar probability to be assigned to fields in either California or Oregon. These results indicate that field 7 is a sink field, potentially, through the introduction of *P. rubi*-infected plant material from locations across the western United States.

We used next-generation sequencing techniques with a GBS approach to characterize the population dynamics of the *P. rubi* plant pathogen. The results show that populations of *P. rubi* are genetically similar globally. We also found evidence of migration of *P. rubi* between the United States and Europe, but our results do not allow for the determination of which population is the source and which population is the sink for the pathogen, and more sampling is required. Our results also indicate a low level of population structure among *P. rubi* populations from the western United States and a high number of migrants from the source populations of California and Oregon toward Washington. Movement via contaminated planting material from in ground nurseries is the most plausible hypothesis that can explain the migration of *P. rubi*. There are several raspberry nurseries located in California and Washington that ship certified planting material (rooted canes and root cuttings) to commercial fields throughout the western United States and world (Zasada et al. 2010), and the reality is that it is difficult to completely prevent disease spread via nursery plants because they can harbor latent or asymptomatic infections. Continued diligence by the raspberry nursery industry to produce planting material free of *P. rubi* will likely prevent the spread of the pathogen to new sites. The transition of some nurseries to the production of tissue-cultured plant material rather than in-ground nursery production is another step in reducing the spread of the economically important raspberry pathogen *P. rubi*.

MATERIALS AND METHODS

Population sampling, DNA extraction, and GBS.

Our sampling strategy consisted of two approaches: i) obtaining a global population sample of *P. rubi* and *P. fragariae* from culture collections to test for gene flow among these two taxa, and ii) intensive hierarchical sampling of populations of *P. rubi* in California, Oregon, and Washington to determine population structure (Table 1). A total of 138 previously isolated isolates of *P. rubi* from six countries and from 10 fields from five states in the western United States (Supplementary Table 1) were used in this study (Stewart et al. 2014) (Table 1). A total of 45 isolates of *P. fragariae* from North America and Europe were also included. For all isolates, mycelium was grown on cellophane filters over a plate of V8-200 media for 7 days and was then harvested and freeze-dried. DNA was extracted using the Qiagen DNEasy plant minikit following the manufacturer's protocol (Germantown, MD, U.S.A.). The restriction endonuclease ApeKI (New England Biolabs, Ipswich, MA, U.S.A.) was used to reduce the complexity of the genome. Each of the isolates was barcoded and the pooled library was subjected to high-throughput sequencing. The first lane contained 19 isolates sequenced on the Illumina HiSeq 2000 platform. These isolates were sequenced in a full lane containing 96 isolates, using 100-bp single end reads. All other isolates were

sequenced on two lanes of the Illumina HiSeq 3000 platform, containing 96 isolates each, using 150-bp paired end reads and a median insert size of 500 bp, following the Illumina protocol (San Diego, CA, U.S.A.).

Read mapping and quality filtering.

Libraries for each lane were demultiplexed by sample, using the Illumina barcode to obtain FASTQ files containing raw reads per sample. Demultiplexing was performed using *sabre* (Byrne et al. 2013). Each FASTQ file was aligned against the reference genome of *P. rubi* (Tabima et al. 2017) using *bowtie 2* (—very careful option) (Langmead and Salzberg 2012). Variants were called using the GATK Haplotype Caller (McKenna et al. 2010) for all isolates, and the resulting variant caller format (VCF) file was quality filtered, using the *vcfR* package (Knaus and Grünwald 2017) in the R statistical framework (R Core Team 2016). Variants were filtered based on a minimum read DP of 4 and a maximum of the 95% percentile of the DP distribution to remove variants with high read depths. Reads were further filtered by mapping quality, where all variants with values lower than the maximum mapping quality (MQ = 44) were removed. Finally, sites with more than 50% missing data were removed from the dataset.

Genetic differentiation between *P. rubi* and *P. fragariae*.

To determine the degree of genetic differentiation between *P. rubi* and *P. fragariae*, we reconstructed a phylogenetic tree using ML in RAxML (Stamatakis et al. 2014) on the filtered dataset with 5,632 SNP variants. The ML tree was calculated using the MULTIGAMMAI model of substitution. To obtain branch support, we performed 1,000 bootstrap samples. Subsequently, to further illustrate the genetic differences between both species, we reconstructed 1,000 trees using random subsets of 200 variants from the filtered dataset. The subsets were sampled using *vcfR* (Knaus and Grünwald 2017). Each tree was reconstructed in *ape* (Paradis et al. 2004), using the NJ algorithm and pairwise distances calculated in *poppr* (Kamvar et al. 2015b). The 1,000 NJ trees were overlaid on the ML tree using *DensiTree* (Bouckaert 2010), to estimate support for reciprocal monophyly between *P. rubi* and *P. fragariae*. Genetic differentiation between both species was estimated by calculating *Gst* (Nei 1972, 1973) and *G'st* (Hedrick 2005) using *vcfR* (Knaus and Grünwald 2017).

Clonality and linkage disequilibrium in *P. rubi*.

To infer the predominant mode of reproduction (e.g., sexual, clonal, or mixed), we used the I_A (Brown et al. 1980; Milgroom 1996). We used the function *sample.ia* from *poppr* (Kamvar et al. 2014) to calculate the I_A for 1,000 random variants within the filtered dataset for all isolates of *P. rubi* and built a distribution of I_A values. To determine the degree of linkage of the *P. rubi* isolates in our dataset, we compared the distribution of the I_A values observed for *P. rubi* against the distribution of 1,000 I_A expected values reconstructed from simulations with 0, 50, 75, and 100% linkage. Simulations were conducted in *adeget* (Jombart 2008), with a dataset consisting of 2,574 loci and 133 samples (analogous to the observed *P. rubi* data). We used the Shapiro-Wilk's normality method to test for normality in each of the distributions of I_A values for the observed and simulated data. An ANOVA test in R was used to test for significant differences across I_A distributions. To determine significant differences among means in pairwise comparisons of the I_A distributions for observed and the three simulated data sets, Tukey's HSD posthoc test was calculated in R. To correct for the absence of normality, we performed a nonparametric Kruskal-Wallis rank sum test to compare across the I_A distributions. A nonparametric, posthoc Kruskal-Wallis rank comparison

test was performed to compare between mean ranks. These nonparametric tests were performed in R.

Analyses of global populations.

The R packages *poppr* (Kamvar et al. 2014; 2015b) and *adeigenet* (Jombart 2008) were used to reconstruct a ML tree, using RAXML (1,000 bootstrap replicates, MULTIGAMMAI model of substitution) and to perform a DAPC (retained 24 principal components and three discriminant components) (Jombart et al. 2010) between the isolates obtained from each of the populations in the study (Oregon, California, Washington, New York, Ohio, the European Union). fastSTRUCTURE (Raj et al. 2014) was used to estimate the population assignments for each sample and to infer the potential number of genetic population clusters. fastSTRUCTURE was run for 1 million iterations, with a range of K clusters from 1 to 10. To determine if isolates from California are the source of the European *P. rubi* population or vice versa, we compared 11 demographic models reconstructed in DIYABC (Cornuet et al. 2008). We removed the isolates from the Ohio and New York populations for all subsequent analysis, due to the low number of isolates for each of these populations. Five scenarios (1, 2, 3, 6, and 7) assumed that the Washington population is a product of admixture between Oregon and California and there has been a recent migration toward Europe from either of the states. Two scenarios (4 and 5) reflect an ancestral migration between California and Europe and a subsequent western United States population establishment from these migrants and posterior admixture of Oregon and California toward Washington. Scenario 5 assumes Europe as the ancestral population and California as the ancestral population. The remaining five scenarios do not assume admixture in Washington. Scenario 8 assumes an ancestral, unknown population that splits early into the European Union population and into the common ancestor of a polytomy between the western United States. Scenarios 9 to 11 assume a recent migration from each of the western United States toward Europe. All populations in the western United States arise from a polytomic, unsampled root population. Each of these scenarios were reconstructed in a subset of 12 isolates from each of the western United States populations (California, Oregon, and Washington) in order to have an identical sample size to the European population. We selected the 12 isolates from each of the United States populations randomly. After subsetting, we filtered the data to retain variants with less than 20% missing data. A total of 54 polymorphic variants were retained. A total of 11 million DIYABC simulations were performed. A direct linear approach using the closest 200 simulations was performed to obtain the scenarios with the highest posterior probability. An analysis of confidence in scenario choice was performed by evaluating the confidence in scenario choice by simulating test datasets, calculating the posterior probabilities of each scenario per simulation, and measuring the proportion of time in which a scenario had the highest posterior probability.

Population genetic analyses across the western United States.

The western United States populations (California, Oregon, and Washington) were sampled in a hierarchical manner, with several grower fields per state (Stewart et al. 2014). The SNP data for the global *P. rubi* and *P. fragariae* populations was filtered to remove *P. fragariae* and all variants with more than 20% missing data. Thus, the data analyzed here is a subset of the data described for the global populations.

To determine the degree of genetic similarity among populations in the western United States, we performed an AMOVA and calculated the *Fst* analogs *Gst* (Nei 1972, 1973) and *G'st*

(Hedrick 2005), using *vcfR* (Knaus and Grünwald 2017). A total of 1,000 permutations were performed to establish significant differences in the AMOVA.

We evaluated five different demographic scenarios using DIYABC to infer demographic history of the California, Oregon, and Washington populations (Cornuet et al. 2008) (Fig. 4): a Washington origin with subsequent migration toward Oregon and California, a California origin with subsequent migration toward Oregon and Washington, a California origin with subsequent migration toward Oregon and admixture of California and Oregon in Washington, an unknown origin with subsequent migration toward Oregon and California and admixture of California and Oregon in Washington, and an Oregon origin with subsequent migration toward California and admixture of California and Oregon in Washington. A total of 11 million DIYABC simulations were calculated. A direct linear approach using the closest 200 simulations was performed to obtain the scenarios with the highest posterior probability. Migrate-N was used to estimate the relative migration rates among populations (Beerli and Felsenstein 2001). Migrate-N was run using 1,000,000 steps with 10 replicates, and one hot chain and three cold Markov chain Monte Carlo chains, using a fixed θ to calculate the mutation scaled migration rate between states.

To determine subpopulation structure among grower fields, we plotted the western United States population ML tree by coloring the tips by grower field, using the *ggtree* R package (Yu et al. 2017). Each of the fields was considered a subpopulation. To determine genetic linkage across fields, we calculated linkage disequilibrium using I_A for each grower field using *poppr*. Finally, to determine admixture across fields, we calculated the posterior probability assignments per grower field for each sample, using DAPC in *adeigenet* (Jombart et al. 2010).

Data deposition and additional materials.

The whole genome shotgun data were previously deposited in GenBank under the accessions MWJK000000000 (*P. fragariae*) and MWJL000000000 (*P. rubi*) and NCBI BioProject accession PRJNA375089 (Tabima et al. 2017). The GBS FASTQ files are deposited in NCBI BioProject under accession PRJNA413437. Various computer scripts, VCF data, supplementary figures, and a supplementary table are archived at one or both the Open Science Framework database and github.

ACKNOWLEDGMENTS

We thank B. Tyler, A. Trippe, M. Peterson, M. Dasenko, and S. O'Neil from the Center for Genome Research and Biocomputing (CGRB) for outstanding advice, technical support, and GBS sequencing. We thank J. Stewart, K. Fairchild, K. Bellingham-Johnston, C. Gray, V. Fieland, and C. Press for their excellent technical support. We also thank B. Knaus for his technical advice and vast expertise. Mention of trade names or commercial products in this manuscript are solely for the purpose of providing specific information and do not imply recommendation or endorsement.

LITERATURE CITED

- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.
- Beerli, P., and Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. U.S.A.* 98: 4563-4568.
- Bouckaert, R. R. 2010. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics* 26:1372-1373.
- Boussau, B., and Daubin, V. 2010. Genomes as documents of evolutionary history. *Trends Ecol. Evol. Amst.* 25:224-232.
- Brown, A. H. D., Feldman, M. W., and Nevo, E. 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96:523-536.

- Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., and Asp, T. 2013. Genome wide allele frequency fingerprints (GWAFs) of populations via genotyping by sequencing. *PLoS One* 8:e57438.
- Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., and Estoup, A. 2008. Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24:2713-2719.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12:499-510.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379.
- Gigot, J., Walters, T. W., and Zasada, I. A. 2013. Impact and occurrence of *Phytophthora rubi* and *Pratylenchus penetrans* in commercial red raspberry (*Rubus idaeus*) fields in northwestern Washington. *Int. J. Fruit Sci.* 13:357-372.
- Goss, E. M. 2015. Genome-enabled analysis of plant-pathogen migration. *Annu. Rev. Phytopathol.* 53:121-135.
- Goss, E. M., Tabima, J. F., Cooke, D. E., Restrepo, S., Fry, W. E., Forbes, G. A., Fieland, V. J., Cardenas, M., and Grünwald, N. J. 2014. The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proc. Natl. Acad. Sci. U.S.A.* 111: 8791-8796.
- 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., and Goss, E. M. 2011. Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. *Annu. Rev. Phytopathol.* 49:249-267.
- Grünwald, N. J., McDonald, B. A., and Milgroom, M. G. 2016. Population genomics of fungal and oomycete pathogens. *Annu. Rev. Phytopathol.* 54:323-346.
- Hedrick, P. W. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633-1638.
- Hickman, C. J. 1940. The red core root disease of the strawberry caused by *Phytophthora fragariae* n. sp. *J. Pomol. Hortic. Sci.* 18:89-118.
- Jombart, T. 2008. *ade4*: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403-1405.
- Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.
- Kamvar, Z. N., Brooks, J. C., and Grünwald, N. J. 2015b. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* 6:208.
- Kamvar, Z. N., Larsen, M. M., Kanaskie, A. M., Hansen, E. M., and Grünwald, N. J. 2015a. Spatial and temporal analysis of populations of the sudden oak death pathogen in Oregon forests. *Phytopathology* 105:982-989.
- 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.
- 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.
- Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Man in't Veld, W. A. 2007. Gene flow analysis demonstrates that *Phytophthora fragariae* var. *rubi* constitutes a distinct species *Phytophthora rubi* comb. nov. *Mycologia* 99:222-226.
- McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., and Brumfield, R. T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* 66:526-538.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297-1303.
- 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*. APS Press, St. Paul, MN.
- Milgroom, M. G., Jiménez-Gasco, M. M., Olivares García, C., Drott, M. T., and Jiménez-Díaz, R. M. 2014. Recombination between clonal lineages of the asexual fungus *Verticillium dahliae* detected by genotyping by sequencing. *PLoS One* 9:e106740.
- National Agricultural Statistics Service. 2017. 2016 Census of Agriculture. United States Department of Agriculture, Washington, D.C.
- Nei, M. 1972. Genetic distance between populations. *Am. Nat.* 106: 283-292.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U.S.A.* 70:3321-3323.
- Paradis, E., Claude, J., and Strimmer, K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289-290.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Raj, A., Stephens, M., and Pritchard, J. K. 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573-589.
- Schoebel, C. N., Stewart, J., Grünwald, N. J., Rigling, D., and Prospero, S. 2014. Population history and pathways of spread of the plant pathogen *Phytophthora plurivora*. *PLoS One* 9:e85368.
- Shah, D., Bergstrom, G., Ueng, P., and others. 1995. Initiation of *Septoria nodorum* blotch epidemics in winter wheat by seedborne *Stagonospora nodorum*. *Phytopathology* 85:452-457.
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Stewart, J. E., Kroese, D., Tabima, J. F., Larsen, M. M., Fieland, V. J., Press, C. M., Zasada, I. A., and Grünwald, N. J. 2014. Pathogenicity fungicide resistance, and genetic variability of *Phytophthora rubi* isolates from raspberry (*Rubus idaeus*) in the western United States. *Plant Dis.* 98: 1702-1708.
- Stukenbrock, E. H., Banke, S., Javan-Nikkhah, M., and McDonald, B. A. 2007. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol. Biol. Evol.* 24:398-411.
- Tabima, J. F., Kronmiller, B., Press, C. M., Tyler, B. M., Zasada, I. A., and Grünwald, N. J. 2017. Whole genome sequences of the raspberry and strawberry pathogens *Phytophthora rubi* and *P. fragariae*. *Mol. Plant-Microbe Interact.* 30:767-769.
- Talas, F., Kalih, R., Miedaner, T., and McDonald, B. A. 2016. Genome-wide association study identifies novel candidate genes for aggressiveness, deoxynivalenol production, and azole sensitivity in natural field populations of *Fusarium graminearum*. *Mol. Plant-Microbe Interact.* 29:417-430.
- Wilcox, W. F., Scott, P. H., Hamm, P. B., Kennedy, D. M., Duncan, J. M., Brasier, C. M., and Hansen, E. M. 1993. Identity of a *Phytophthora* species attacking raspberry in Europe and North America. *Mycol. Res.* 97:817-831.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. 2017. *ggtree*: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8:28-36.
- Zasada, I. A., Walters, T. W., and Hanson, B. D. 2010. Challenges in producing nematode- and pathogen-free fruit and nut nursery crops in the United States. *Outlooks Pest Manage.* 21:246-250.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Open Science Framework database: <https://osf.io/mnc2q>
 Github repository: <https://github.com/grunwaldlab/rubi-gbs>