MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 4577–4590 doi: 10.1111/j.1365-294X.2009.04379.x

Genetic epidemiology of the Sudden Oak Death pathogen *Phytophthora ramorum* in California

S. MASCHERETTI, P. J. P. CROUCHER, M. KOZANITAS, L. BAKER and M. GARBELOTTO
Department of Environmental Science, Policy and Management, 137 Mulford Hall, University of California, Berkeley, CA 94720-3114, USA

Abstract

A total of 669 isolates of *Phytophthora ramorum*, the pathogen responsible for Sudden Oak Death, were collected from 34 Californian forests and from the ornamental plant-trade. Seven microsatellite markers revealed 82 multilocus genotypes (MGs) of which only three were abundant (>10%). Iteratively collapsing based upon minimum Φ_{ST}, yielded five meta-samples and five singleton populations. Populations in the same meta-sample were geographically contiguous, with one exception, possibly explained by the trade of infected plants from the same source into different locations. Multidimensional scaling corroborated this clustering and identified nursery populations as genetically most distant from the most recent outbreaks. A minimum-spanning network illustrated the evolutionary relationships among MGs, with common genotypes at the centre and singletons at the extremities; consistent with colonization by a few common genotypes followed by local evolution. Coalescent migration analyses used the original data set and a data set in which local genotypes were collapsed into common ancestral genotypes. Both analyses suggested that meta-samples 1 (Santa Cruz County) and 3 (Sonoma and Marin Counties), act as sources for most of the other forests. The untransformed data set best explains the first phases of the invasion, when the role of novel genotypes may have been minimal, whereas the second analysis best explains migration patterns in later phases of the invasion, when prevalence of novel genotypes was likely to have become more significant. Using this combined approach, we discuss possible migration routes based on our analyses, and compare them to historical and field observations from several case studies.

Keywords: coalescent analysis, emergent forest disease, microsatellite, network analysis

Received 26 June 2009; revision received 30 August 2009; accepted 4 September 2009

Introduction

Phytophthora ramorum (Oomycetes, Straminopiles: Peronosporales) (Werres *et al.* 2001) is the causal agent of two types of disease. In the native coastal forests of California and southern Oregon, *P. ramorum* causes Sudden Oak Death (SOD) (Rizzo & Garbelotto 2003), a disease characterized by lethal trunk lesions that affect tanoak (*Lithocarpus densiflora*), coast live oak (*Quercus agrifolia*), Californian black oak (*Quercus kelloggii*) and possibly Canyon live oak (*Quercus chrysolepis*). Additionally, both in North America and Europe, the pathogen causes ramorum blight, a disease characterized mostly by leaf blight and/or branch dieback in over 100 species of both wild and ornamental plants, including California bay laurel (*Umbellularia cailfornica*), California redwood (*Sequoia sempervirens*), Douglas-fir (*Pseudotsuga menziesii*), *Camellia* and *Rhododrendron* species (Davidson *et al.* 2003; Rizzo *et al.* 2005; Grunwald *et al.* 2008).

Symptoms of *P. ramorum* infection were first noticed in 1994–1995 in Californian forests (Garbelotto *et al.* 2001; Rizzo *et al.* 2002) and in European nurseries (Werres *et al.* 2001). In North America, *P. ramorum* is found in Pacific coastal forests, generally <30 km inland, from Monterey County in central California to Curry County, southern Oregon. In Californian forests, the most commonly infected tree is bay laurel (Garbelotto *et al.* 2003;

Correspondence: M. Garbelotto, Fax: +1 510 643 5436;
E-mail: matteo@nature.berkeley.edu

Davidson *et al.* 2005). Sporangia production on its leaves is prolific and bay laurel is therefore the main source of aerial infectious inoculum. In contrast, sporulation on oak wood is extremely rare and so oak species may be considered as a noncontagious dead-end host in the *P. ramorum* life cycle (Davidson *et al.* 2005).

The available genetic data suggest that *P. ramorum* is an exotic pathogen introduced from unknown origins at least three times independently. Populations of *P. ramorum* in Europe and North America belong to three clonal lineages: EU1, NA1 and NA2. EU1 is the only lineage found in Europe and has now also been found in nurseries in California as the result of the movement of infested ornamental plants. NA1 is the clonal lineage responsible for the natural infestation in the North American forests and most of the nurseries in North America while NA2 has only been isolated from nurseries in North America (Rizzo *et al.* 2002; Hansen *et al.* 2003; Ivors *et al.* 2004; Olarte & Garbelotto 2005).

Phytophthora ramorum is self-sterile and sexual reproduction can only occur when individuals of different mating types (A1 and A2) interact (Brasier & Kirk 2004). Mating type A1 occurs commonly in Europe, the EU1 lineage belongs predominantly to the A1 mating type, while NA1 and NA2 belong only to mating type A2 (Brasier & Kirk 2004). Although isolates belonging to both mating types have been found in the same nurseries in western US states (Hansen *et al.* 2003; Ivors *et al.* 2004; Grunwald *et al.* 2008), there is currently no evidence of sexual reproduction in these nurseries. In a recent study, Goss *et al.* (2009) have shown that the three known lineages of the pathogen have originated from three geographically distinct populations, and that they appear to have been isolated from one another and clonally reproducing for *c.* 165 000 years. Because of this long isolation, it is possible that, even in presence of both mating types, the ability to reproduce sexually may either be lost or seriously reduced.

Both amplified fragment length polymorphisms and microsatellite analyses (Ivors *et al.* 2004, 2006; Prospero *et al.* 2007, 2009) have suggested an exclusively asexual mode of reproduction in the forests in California, and Oregon. Considering these genetic data and the fact that no A2 type as been so far detected in the wild, it is very likely that *P. ramorum* propagates in an exclusively clonal manner in California forests. Epidemiological (Davidson *et al.* 2005) and population genetic studies of *P. ramorum* in California, Oregon and Washington (Prospero *et al.* 2004, 2007, 2009; Mascheretti *et al.* 2008) have indicated that although populations of *P. ramorum* are genetically structured due to the limited ability of the pathogen to disperse naturally, there may be widespread long-distance gene flow due to human-mediated transport.

Several microsatellite markers have been developed to study the genetic structure of *P. ramorum* populations (Ivors *et al.* 2006; Prospero *et al.* 2007, 2009), but only a few of these markers exhibit significant variation within the NA1 lineage present in the North American Pacific Costal forests.

Here, we examine the variation in seven microsatellite markers, previously shown to be variable in the NA1 lineage (Ivors *et al.* 2006; Prospero *et al.* 2007), from 669 *P. ramorum* individuals collected in 35 forest populations ranging from Humboldt to Monterey Counties. The present work builds upon our earlier study (Mascheretti *et al.* 2008) where analysis of molecular variance (AMOVA), minimum-spanning network (MSN), and spatial autocorrelation analyses were utilized to make inferences on the origin, the current genetic structure, and potential spread patterns of the pathogen using a much smaller sample size. The current study is based on a greatly expanded data set, both in terms of number of genotyped isolates and sample locations and with the addition of multidimensional scaling (MDS), coalescent analyses, and of neighbour-joining (NJ) analysis using genotype distances rather than Φ_{ST} values. By using the combined results of all analyses performed, we attempt to elucidate the patterns of spread that have lead to the current epidemic by exploring the genetic landscape of *P. ramorum* in California. We identify statistically and geographically homogeneous genetic populations, visualize their relationships in genetic space, and make inferences about the colonization routes that may have led to the current genetic landscape. The latter makes use of Bayesian coalescent approaches for estimating relative migration rates. These inferences are supported at the molecular evolutionary level by a detailed network analysis of multilocus genotype (MG) evolutionary relationships. Finally, this work provides further convincing evidence regarding (i) nursery populations as the original source of the wild infestation in California, (ii) the local generation of novel genotypes through mutation and/or mitotic recombination, and (iii) the bimodal range of spread of the pathogen as previously suggested (Ivors *et al.* 2006; Mascheretti *et al.* 2008).

Materials and methods

Phytophthora ramorum *samples*

A total of 378 isolates of *Phytophthora ramorum* were collected from 340 bay laurels (*Umbellularia californica*) and from 38 coast live oaks (*Quercus agrifolia*) in 20 forest stands in seven Californian counties (Alameda, San Mateo, Sonoma, Marin, Napa, Santa Cruz and Monterey) in spring 2007 and 2008. These were analysed together

with another 278 isolates collected in 2002 and 2005 from bay laurels in 14 forests stands in Humboldt, Marin, Monterey and Santa Cruz and 14 isolates collected from nurseries in 12 US states in 2004 (Fig. 1 and Table 1). Some forests were sampled more than once. Jack London State Park was sampled in 2002 (SO-1a), 2005 (SO-1b) and 2007 (SO-1c). China Camp State Park was sampled in 2005 (MA-3a) and 2007 (MA-3b), and Bean Creek was sampled in 2005 (SC-1a) and 2007 (SC-1b).

Phytophthora ramorum *isolation and DNA extraction*

Several small pieces (~2 mm × 2 mm) of the symptomatic portion of sampled leaves were plated on selective Primaricin–Ampicillin–Rifampicin–Pentachloronitrobenzene agar (Eriwn & Ribeiro 1996), incubated at room temperature in the dark for 3 weeks, transferred into pea broth liquid medium and grown at room temperature for 7 days. Mycelia were harvested by filtering on Whatman paper and lyophilized for 24 h. DNA was extracted using the protocol described by Ivors *et al.* (2006).

Microsatellite analysis

Seven microsatellites, known to be variable in the NA1 lineage, were genotyped: PrMS39a, PrMS39b, PrMS43a, PrMS43b, PrMS45 (Prospero *et al.* 2007), and locus 18 and locus 64 (Ivors *et al.* 2006). The microsatellite loci

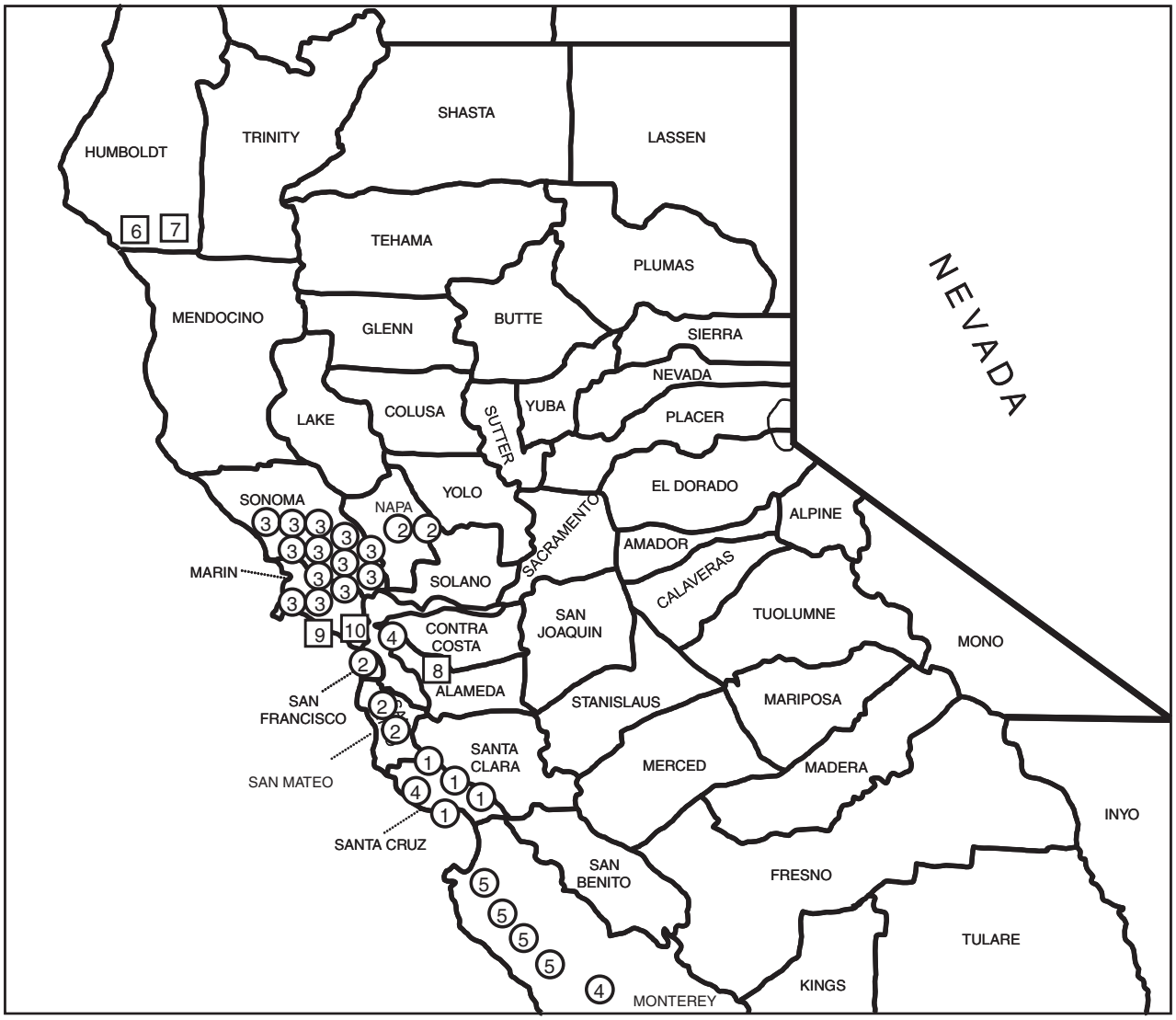

Fig. 1 Map of the *Phytophthora ramorum* localities sampled in Californian forests. The five squares indicate the populations that did not cluster with any other on the basis of minimum Φ_{ST} (6, 7, 8, 9 and 10), the circles indicate the other meta-samples: meta-sample 1 (SC), meta-sample 2 (SM, MA, NA), meta-sample 3 (MA, SO), meta-sample 4 (SC, MA, MO) and meta-sample 5 (MO).

Table 1 Localities where the 669 *Phytophthora ramorum* isolates were sampled, approximate duration of infestation, meta-sample to which each population belongs

	Meta-sample	County	Locality	Longitude	Latitude	*n*	Collection year	Approx. duration of infestation* (years)
HU-1	6	Humboldt	Redwoods	−123.8333	40.1196	19	2005	<5
HU-2	7	Humboldt	Redwoods	−123.8337	40.2583	14	2005	<5
SO-1a	3	Sonoma	Jack London SP	−122.4709	39.2950	14	2002	>10
SO-1b	3	Sonoma	Jack London SP	−122.5521	38.3537	25	2005	>10
SO-1c	3	Sonoma	Jack London SP	−122.5571	38.3486	40	2007	>10
SO-2	3	Sonoma	Austin Creek SP	−123.004	38.5373	8	2007	>10
SO-3	3	Sonoma	Willow Creek SRA	−122.0513	37.0533	14	2007	>5 < 10
SO-4	3	Sonoma	Fairfield Osborn P	122.5948	38.3434	10	2007	>5 < 10
SO-5	3	Sonoma	Freezeout Creek	−123.0505	38.4500	3	2007	>5 < 10
SO-6	3	Sonoma	Sweet Water	−122.9730	38 5537	8	2007	>5 < 10
MA-1	9	Marin	Tomales Bay SP	−122.8723	38.1152	9	2005	<5
MA-2	10	Marin	Angel Island SP	−122.4391	37.8589	19	2005	>5 < 10
MA-3a	3	Marin	China Camp SP	−122.4610	38.0057	23	2005	>10
MA-3b	3	Marin	China Camp SP	−122.4700	38.0012	35	2007	>10
MA-4	2	Marin	Golden Gate NRA	−122.6678	37.9419	24	2005	>5 < 10
MA-5	3	Marin	S.P. Taylor SP	−122.7361	38.0296	24	2005	>10
MA-6	4	Marin	Paradise Drive	−122.4742	37.9002	20	2007	>10
MA-7	3	Marin	Lake Lagunitas	−122.5950	37.9462	10	2007	>10
MA-8	3	Marin	Point Reyes NS	−122.4871	38.0307	3	2007	<5
NA-1	2	Napa	R.L. Stevenson SP	−122.5920	38.7490	8	2007	>5 < 10
NA-2	2	Napa	Napa Skyline Park	−122.2370	38.2663	9	2007	>5 < 10
AL-1	8	Alameda	Castro Valley	−122.0577	37.6922	10	2007	>10
SM-1	2	San Mateo	SF PUC	−122.426	37.7800	50	2008	>10
SM-2	2	San Mateo	Coal Creek P	−122.2033	37.3245	18	2007	>5 < 10
MO-1	5	Monterey	Deetjens	−121.6702	36.1588	22	2005	>5 < 10
MO-2	5	Monterey	Los Padres NF	−121.4399	35.9152	21	2005	<5
MO-3	5	Monterey	Santa Lucia P	−121.8690	36.5279	45	2007	<5
MO-4	5	Monterey	Big Creek R	−121.5975	36.0708	6	2007	<5
MO-5	4	Monterey	Palo Corona RP	−121 8700	36.4508	12	2007	<5
SC-1a	1	Santa Cruz	Bean Creek	−122.0074	37.0831	24	2005	>10
SC-1b	1	Santa Cruz	Bean Creek	−122.0513	37.0533	54	2007	>10
SC-2	4	Santa Cruz	Felton	−122.0262	36.9720	24	2005	>5 < 10
SC-3	1	Santa Cruz	Henry Cowell SP	−122.0727	37.0438	16	2002	>5 < 10
SC-4	1	Santa Cruz	Zyante Road	−122.0438	37.09123	14	2007	>10
Nursery	11					14	2004	

SP, State Park; NF, National Forest; NRA, National Recreational Area; SRA, State Recreation Area; RP, Regional Park; P, Preserve; R, Reserve; SF PUC, San Francisco Public Utility Commission; NS, National Seashore.
*Duration of infestation as estimated at sampling year.

were typed using primers and protocols as described previously (Ivors *et al.* 2006; Prospero *et al.* 2007; Mascheretti *et al.* 2008).

Identification of statistically homogeneous genetic populations

The AMOVA (Excoffier *et al.* 1992) procedure, as implemented by the software Arlequin 3.0 (http://cmpg.unibe.ch/software/arlequin3) (Excoffier *et al.* 2005), was employed to generate pairwise estimates of Φ_{ST} among all 36 *P. ramorum* populations. As previously (Mascheretti *et al.* 2008), all AMOVA analyses were carried out at the 'haplotype level', incorporating the evolutionary distance (the square of the difference in repeat number) between all loci simultaneously into the variance estimates rather than averaging across loci. This maximally efficient approach is possible because of the clonal reproduction of *P. ramorum* and because the alleles at each locus typically either evolve in concert (always homozygous), or with one allele fixed; rendering one allele effectively redundant at each locus. The very few instances of individuals with partial repeats or null alleles (one individual from population MA-3b with 74.5 repeats at MS43b, one individual from population SO-1c with 75.5 repeats at MS43b, and one individual from population SO-4 with a null allele at locus MS18 (-/282) were excluded from these analyses.

As anticipated from previous analyses (Mascheretti *et al.* 2008), many of the populations had very low Φ_{ST} estimates that were not significantly different from zero. Populations were therefore recursively clustered into meta-samples by pooling the pair of populations or clusters that yielded the minimum Φ_{ST} at each round until no further insignificant clustering (i.e. minimum Φ_{ST} $P > 0.05$) was possible (Roewer *et al.* 2005; Mascheretti *et al.* 2008). This greedy algorithm was supervised by applying the clustering in three rounds: first clustering was only applied to populations from the same location, second to populations from the same county, and third to the overall data set. From this point on, we will refer to 'populations' when talking about the original sample populations, and we will refer to 'meta-samples' when talking about the regrouped populations from iterative clustering (including singleton populations that did not collapse with other populations). Finally, a traditional AMOVA was performed to compare the proportion of genetic variation within and among the populations and meta-samples.

The genetic landscape of the *P. ramorum* epidemic in California was graphically visualized by identifying principle coordinates for the pairwise estimates of Φ_{ST} among the 36 populations through a MDS analysis using the ISOMDS function in the R-statistical package (R Development Core Team 2008, www.r-project.org). Individual solutions were iterated until the improvement in stress (Kruskal's *S*) was <0.0001. This process was repeated for one through six dimensions, with the optimum dimensionality being determined by a 'scree' test (Table 2): a clear 'elbow' was detected for the two-dimensional solution with higher-dimensional solutions not providing a substantial reduction in stress.

Spatial autocorrelation analysis was carried out using Moran's *I*-index (Sokal & Oden 1978) to measure the correlation between allelic repeat numbers and given geographical intervals. GPS latitude and longitude coordinates were converted into NAD1986 UTM coordinates using ARCVIEW9 (ESRI) and the autocorrelation was evaluated using SPAGeDI (Hardy & Vekemans 2002). A total of 277 isolates from 15 populations (SO-1b, SO-2, SO-3, SO-5, SO-6, MA-1, MA-2, MA-3a, MA-4, MA-5, MA-7, MO-2, MO-5, SC-1a and SC-1b) were used for this analysis.

Table 2 Multidimensional scaling and 'scree' test of pairwise estimates among the 36 populations of *Phytophthora ramorum*

Dimensionality	Kruskal's *S*
1	36.2321
2	19.0974
3	14.6319
4	11.7149
5	9.9692
6	8.7477

For each of the meta-samples, genotype diversity was calculated in two ways: (i) as the clonal genotype diversity, $R = (1 - G)/(1 - N)$, where G is the number of MGs present in a sample and N is the sample size and (ii) using Stoddart and Taylor's index $G = \Sigma p_i^2$ (Stoddart & Taylor 1988), where p_i is the frequency of the ith MG. For each meta-sample, gene diversity (expected heterozygosity) was also estimated as

$$H_E = 1 - \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{k} p_i^2,$$

where p_i is the frequency of the ith of k alleles, averaged over each l of m loci.

Reconstruction of the evolutionary relationships among the multilocus genotypes

The mutational relationships among the individual diploid MGs were examined by computing matrices of genotype distances, using repeat number differences, according to the method of Bruvo *et al.* (2004). These matrices were converted to MSN using the program MINSPNET (Excoffier & Smouse 1994) and visualized using GRAPHVIZ (http://www.graphviz.org). Because meta-samples comprising multiple undifferentiated populations presumably reflect separate introductions and local spreading of the pathogen, separate networks were calculated for the MGs found in each separate meta-sample and a final network was created by merging the separate sub-networks at the shared MGs (nodes). This approach minimizes spurious edges between MGs that have evolved independently in different meta-samples rather than arriving by dispersal.

In addition, the above genotype distance measure was also employed to calculate a matrix of average pairwise genetic distances among the set of meta-samples. This matrix was used to construct a NJ-tree to display the relationships among the meta-samples in order to corroborate the above Φ_{ST}-based MDS analysis. Programs to calculate the genetic distances were written in Python (PJPC).

Bayesian coalescent inferences of migration and infestation routes

In order to further define the possible routes of spread of *P. ramorum*, the program Migrate-N (Beerli & Felsenstein 1999, 2001; Beerli 2006) was employed to estimate the directional rate of 'migration' events among the meta-samples. Migration events among the genetically

distinct meta-samples that resulted from the iterative collapsing based upon minimum Φ_{ST} were estimated in order to identify inter-regional patterns of spread. This analysis was performed in two ways. First, each analysis was performed using all the genotypes present in each meta-sample. Second, the data were simplified by contracting all singleton MGs and those that occurred more than once but were unique to a single population into their parental (i.e. ancestral) MGs. The contraction was carried out by constructing the MSN for each population to identify the mutational relationships of the MGs. MGs were not collapsed if they were more than two mutational steps away from their parental MGs. In this way, genotypes that were likely to have developed recently and locally, and were therefore unlikely to have been involved in migration events were eliminated.

Microsatellite evolution was modelled under a Brownian motion model, and the program was run assuming a constant mutation rate for all loci, with 200 000 visited parameter values, a burn-in of 10 000 and an adaptive heating scheme with four chains. Migration was estimated as M, the immigration rate m divided by the mutation rate μ (a measure of how much more important immigration is than mutation in bringing new variants into a population).

Results

Variability and genetic diversity

The *Phytopthora ramorum* microsatellites PrMS39a, PrMS39b, PrMS43a, PrMS43b, PrMS45, locus18, and locus64 were analysed for the 669 isolates from 35 populations (Fig. 1; Table 1). PrMS39a was invariant, but the other loci exhibited between 4 (locus64) and 12 (PrMS43b) different alleles, resulting in a total of 82 MGs (see Supporting Information). More than half of the MGs (46; 56%) were singletons, being unique to an individual isolate. The most frequent MGs were MG46 (n = 114; 17.01%), MG42 (n = 96; 14.33%) and MG38 (n = 88; 13.13%), with MG46 and MG38 only differing from MG42 by one mutational step at PrMS43b. The next most frequent MGs were MG24 (n = 41; 6.12%), MG74 (n = 31; 4.63%), MG27 (n = 24; 3.58) and MG16 (n = 21; 3.13%). Among the MGs present more than once, MG1, MG6, MG29, MG55 and MG64 were restricted to single populations. To compare the MGs obtained in this study to the ones obtained in the previous study (which includes 278 of the isolates in the current study), see Supporting Information.

When the forest populations where iteratively collapsed based upon minimum Φ_{ST}, five populations remained as independent singletons: HU-1 (meta-sample 6), HU-2 (meta-sample 7), AL-1 (meta-sample 8), MA-1 (meta-sample 9), MA-2 (meta-sample 10). Five groupings also emerged (Fig. 1 and Table 1). Meta-sample 1 included four populations collected in Santa Cruz County (SC-1a, SC-1b, SC-3, SC-4) together with the nursery population. The latter was however, then excluded and treated as a singleton 'meta-sample 11' for two reasons: (i) the nursery population does not represent a geographically defined population *per se* and (ii) the nursery population is taken to represent the possible source of initial inoculum in California. Meta-sample 2 included the two populations collected in San Mateo, the two collected in Napa and one population collected in Marin County (SM-1, SM-2, NA-1, NA-2, MA-4). Meta-sample 3 amalgamated the eight populations collected in Sonoma County and five of the populations collected in Marin County (SO-1a, So-1b, SO-1c, SO-2, SO-3, SO-4, SO-5, SO-6, MA-3a, MA-3b, MA-5, MA-7, MA-8). Meta-sample 4 collapsed one population collected in Santa Cruz, one in Monterey and one in Marin County (SC-2, MO-5, MA-6). Finally, meta-sample 5 contained all the remaining samples from Monterey County (MO-1, MO-2, MO-3, MO-4). In each case, where the same forest had been sampled more than once in different years, each of the samples from each forest merged into the same meta-sample (SO-1a,b,c and MA-3a,b into meta-sample 3 and SC-1a,b into meta-sample 1). The overall AMOVA based on these groupings, but excluding the nursery population (Table 3), indicated that they are robust with 70.92% of the variation being within samples, 19.83% among the meta-samples, but only 9.25% of the variation being among populations within meta-samples. The overall Φ_{ST} = 0.2908 and the overall Φ_{CT} = 0.1983. In all cases, the permutation P-value was <0.00001.

Table 3 AMOVA results for the 10 meta-samples of *Phytophthora ramorum* in Californian forest (nursery population excluded)

Source of variation	d.f.	Sum of squares	Variance components	% of variation
Among meta-samples	9	106.788	0.1660	19.83
Among populations within meta-samples	25	48.859	0.0748	9.25
Within populations	616	365.712	0.5937	70.92

Table 4 Indices of genetic diversity: R = clonal genotype diversity, G = Stoddart and Taylor's index, H_E = gene diversity (expected heterozygosity) for each meta-sample

Meta-sample	$R = (G - 1)/(N - 1)$	$G = 1/\Sigma(p^2)$	H_E
1	0.1402	4.7223	0.4228
2	0.2963	8.3965	0.5002
3	0.2046	11.2403	0.5242
4	0.2000	5.6403	0.4633
5	0.1828	2.6049	0.3902
6	0.1667	2.2704	0.3700
7	0.3077	2.5789	0.3963
8	0.3333	3.3333	0.4067
9	0.3750	2.0769	0.3200
10	0.2222	3.1391	0.4317
11	0.3077	1.8846	0.3367

Indices of genotype diversity and gene diversity for the meta-samples are given in Table 4. Of particular interest are the estimates of G and H_E, which reveal that the highest diversity was in meta-samples 2, 3 and 4 and the lowest diversity was in meta-samples 5, 6, 7, 9 and 11. With the exception of meta-sample 11 (the nursery population), the latter all represent relatively young infestations, whereas meta-samples in the former group represent the oldest infestations in California. This clearly indicates that genetic diversity within forest stands largely evolves *in situ*, by mutation, over-time, following initial infestation by one or a few genotypes. For a pathogen that relies exclusively on mutations and somatic recombination events to generate new genotypes, lack of high genotypic diversity is not surprising in the controlled and small nursery population, especially when compared to the extremely large wild populations.

The genetic landscape of *P. ramorum* in California was formally examined, independently of geography, in a two-dimensional MDS analysis (Fig. 2). This analysis clearly corroborated the population clustering based upon minimum significant Φ_{ST}, with all populations grouping with their respective meta-sample members. The MDS also revealed subtle patterns that the iterative clustering could not reveal. In particular, the location of meta-sample 2 between two clouds comprising meta-sample 3 is interesting because it suggests some complexity in the relationships among the populations in meta-sample 2 and those in meta-sample 3 (see below). Also, the location of MA-4 (meta-sample 2) closer to meta-sample 1 and the nursery (meta-sample 11), than to the rest of meta-sample 2, is also revealing because MA-4 is one of the populations where SOD was first described and one that has been previously identified as the beginning of the infestation in the wild in California (Mascheretti *et al.* 2008).

The relationships at the meta-sample level were examined by drawing a NJ-tree. Rather than construct this tree on the matrix of pairwise Φ_{ST} values, that themselves defined the meta-samples, the relationships were examined using a matrix of pairwise average repeat number-based genetic distances (Bruvo *et al.* 2004) (Fig. 3). The branches on this tree do not represent splitting events, merely genetic distances. Examination of Figs 2 and 3 shows that the NJ-tree strongly corroborates the relationships among the meta-samples that was indicated by the Φ_{ST}-based MDS plot. The tree

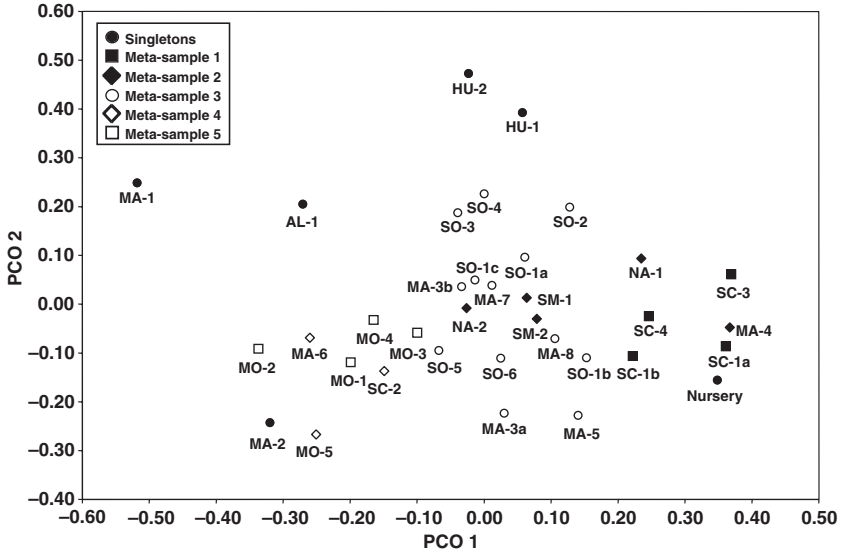

Fig. 2 Multidimensional scaling plot among Californian *Phytophthora ramorum* populations based upon Φ_{ST}. [Correction added after online publication 20 October 2009: the captions for Figs 2 and 3 were transposed.]

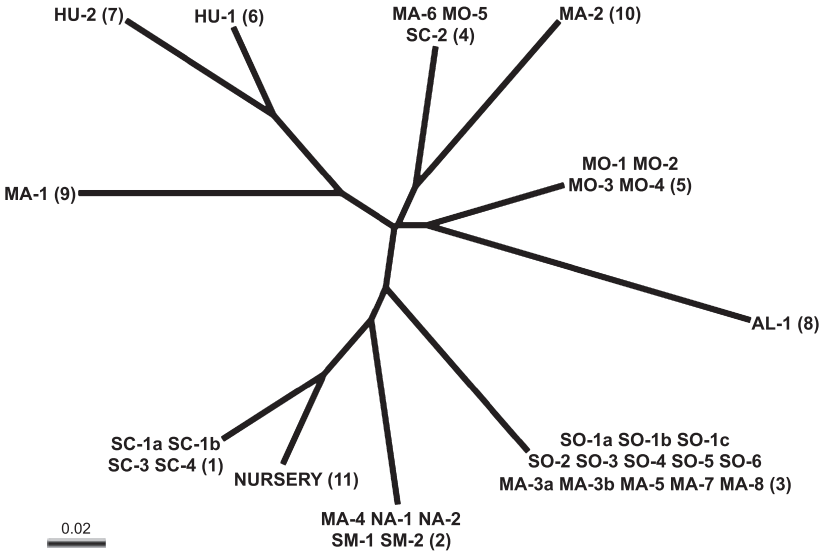

Fig. 3 Neighbour-joining tree showing the relationships among the meta-samples of *Phytophthora ramorum*, defined according to minimum Φ_{ST}. In each case, the name of the meta-sample and its constituent population(s) is indicated. [Correction added after online publication 20 October 2009: the captions for Figs 2 and 3 were transposed.]

also clearly illustrates the affinity between the nursery population (meta-sample 11) and Santa Cruz (meta-sample 1). As in the MDS analysis, meta-sample 2 is the second-most related to the source nursery population: this meta-sample includes MA-4, which as noted above, represents one of the populations where SOD was first described.

Spatial evolution of multilocus genotypes

The mutational relationships among the 82 MGs are illustrated in the MSN in Fig. 4. The common MGs are located within the centre of the network, while the rare or singleton MGs are on the extremities. The three most common MGs, MG46 (17.01%), MG42 (14.33%) and

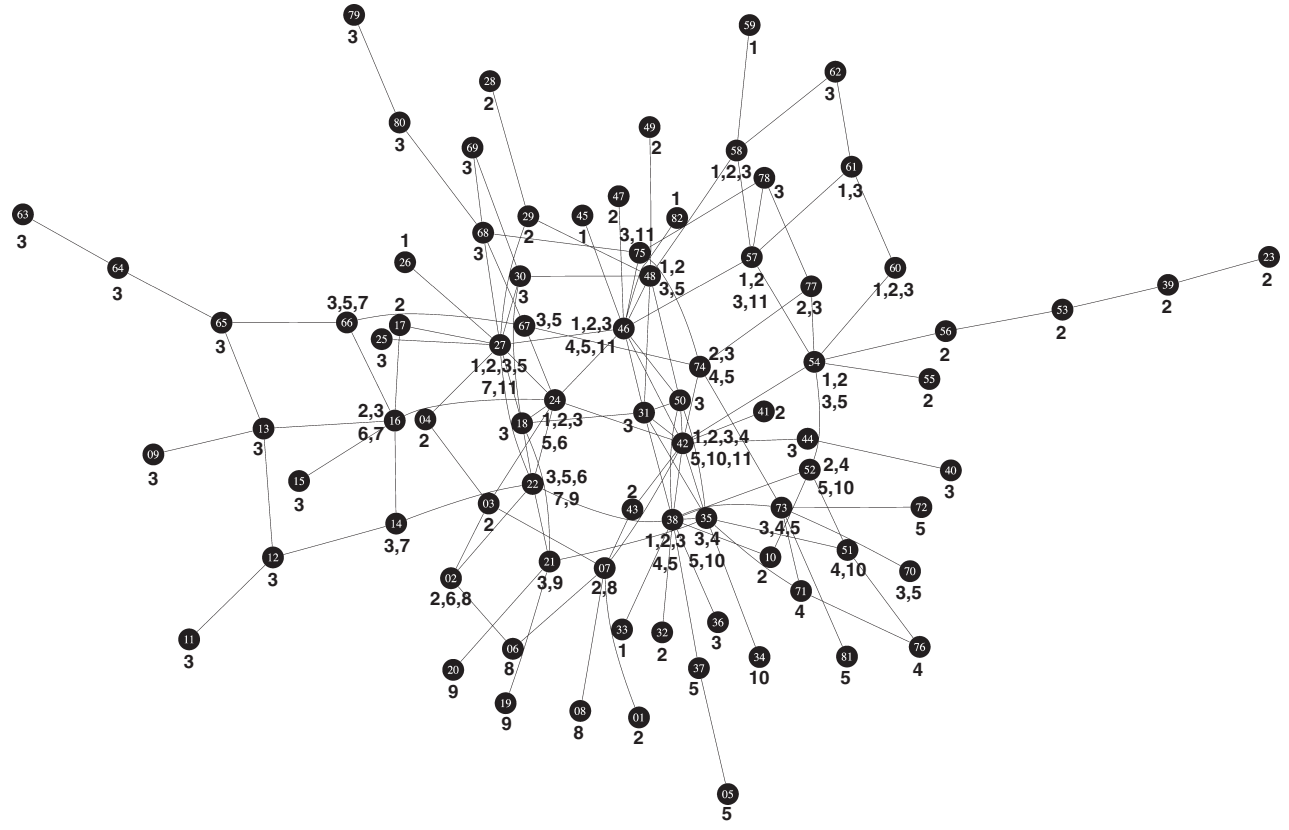

Fig. 4 Minimum-spanning network (MSN) showing the relationships among the individual MGs (the nodes). The meta-samples in which each MG was found are indicated next to each node.

MG38 (13.13%), are in the centre and are present in all five meta-samples comprising multiple populations. The 14 MGs, which are present in more than 10 individuals (MGs 16, 22, 24, 27, 38, 42, 46, 48, 51, 52, 54 57, 73 and 74) (see Supporting Information), are spread among all the meta-samples with the exception of meta-sample 8 (AL-1, Castro Valley RP, $n = 10$) that is located at the edge of the network and includes MGs 2, 6, 7 and the singleton MG8. The rare or singleton MGs towards the perimeter of the network tend to group by meta-sample, supporting (i) the epidemiological reality of the meta-samples, and (ii) that novel MGs evolve by stepwise mutation from common MGs (which likely comprise the majority of inoculum) at local sites and are not themselves being widely dispersed. The local evolution of genotypes in a stepwise fashion is further corroborated by the presence of two tails of several derived novel genotypes, each entirely contained in one of two meta-samples (either meta-sample 2 or meta-sample 3). These two tails represent the first report of novel genotypes accumulating significant mutational differences from the original introduced genotypes, and include a genotype characterized by some alleles that are now identical to those previously reported only from genotypes within the European lineage.

Inferences on possible migration and infestation routes

The inferred migration events (mean estimates) among the genetically distinct meta-samples are shown in Fig. 5A, B. Only 'significant' migration events for which the Bayesian 95% comparison interval did not include zero are presented. Examination of Fig. 5A indicates that meta-sample 1 (Santa Cruz County) and meta-sample 3 (Sonoma and Marin Counties) appear to be ultimately responsible for most of the migrations into other forests and indeed, as noted previously, both of these meta-samples include many of the oldest infestations, including those from Santa Cruz, which have consistently shown to be most genetically similar to the nursery population. The three meta-samples 1, 2 and 3, which appear genetically closest to the nursery source population, both in the Φ_{ST} tree and the MDS analyses, are the only three meta-samples among which migration events occur approximately to the same extent in both directions. The second analysis, conducted using only parental (ancestral) MGs (Fig. 5B), confirmed most of the major routes of migration from meta-samples 1 and 3 with the exception that meta-sample 9 (MA-1, Marin County) appears to have received *P. ramorum* individuals both from meta-sample 3 and 1, instead of only from meta-sample 1. Additionally, the second analysis identifies meta-sample 6 as the major source of migration into meta-sample 7. Some minor but statistically significant migration events highlighted in the first analysis (Fig. 5A) were not confirmed by the second analysis (Fig. 5B): these include migration from meta-sample 5 to 4, 10 to 5, 2 to 4, 2 to 7, 7 to 6. New migration events were identified by the second analysis from meta-sample 5 into 1 and 5 into 6.

The spatial autocorrelation between geographical and genetic distances (with Moran's *I* as the index of genetic similarity) using either the latitude and longitude of the locality or, when available, the coordinates of each tree, yielded a result (data not shown) very similar to the one obtained in the previous study (Mascheretti *et al.* 2008): significant autocorrelation (decreasing positive values of Moran's *I* with increasing geographical distance) until ~630 m, followed by a sharp increase in *I* until ~1200 m, at which point *I* declined, reaching zero at ~10 km.

Discussion

The 35 forest infestations that were investigated revealed 82 MGs of which only three had an overall

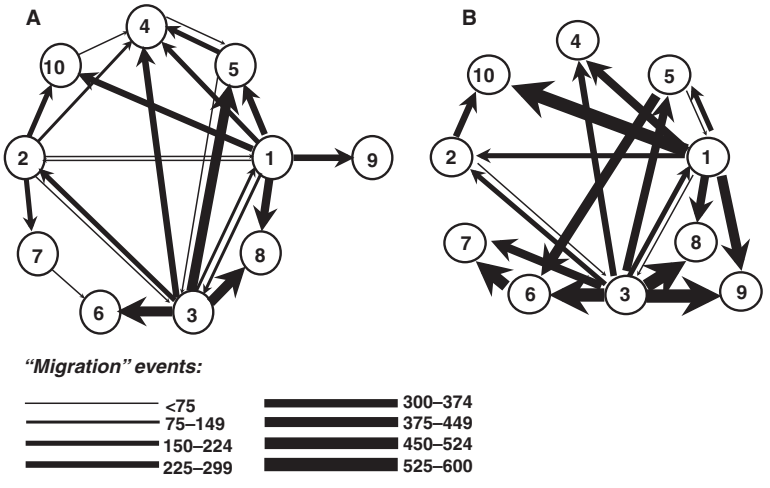

Fig. 5 Visualization of probable migration routes among the meta-samples. (A) Unmodified data. (B) Data after contracting all singleton MGs and those that occurred more than once but were unique to a single population into their parental (ancestral) MGs.

frequency >10% and more than half (46; 56%) were singletons (see Supporting Information). Despite this apparently high-level of genetic diversity, at least as revealed by loci specifically selected to highly variable, a major conclusion of this study is that extensive spatial-genetic structuring is present in Californian *Phytophthora ramorum* populations. Clustering based on minimum Φ_{ST} resulted in five meta-samples comprising more than one population and this clustering was strongly supported both by AMOVA and by Φ_{ST}-based MDS analysis, with the arrangement of the meta-samples in genetic space, in turn, being supported by the non-Φ_{ST}-based NJ-tree. In most cases, the clustered populations are also geographical neighbours, suggesting regional spread from initial foci, either by natural dispersal or local translocation by people. For example, meta-sample 1 includes four populations from Santa Cruz County, meta-sample 3 consists of 13 populations from Sonoma and Marin Counties, meta-sample 5 consists of four populations from Monterey County, and meta-sample 2 includes two disjunct pairs, each comprising two adjacent populations. In contrast though, meta-samples 2 and 4 also contain populations collected from forests that are not geographically adjacent, suggesting human-mediated transport from the same source or, less likely, long-range natural dispersal.

The most likely reason for the appearance of genetically distinguishable meta-samples is their origin from one or a few founding genotypes, followed by the lack of further abundant gene among sites. The lack of effective movement among established infestations is not only in agreement with the presence of genetic structure, but is further corroborated by the observation that three populations re-sampled several years apart showed no change in genetic composition, a fact indicative of the lack of effective immigration into these established infestations. Based on this observation, we conclude that most effective migration may still be occurring only unidirectionally from established infestations into new sites. With time, bidirectional gene flow may be attained among some meta-samples, possibly reducing the overall amount of genetic structure in California. Nonetheless our data are indicative of bidirectional gene flow among the three meta-samples (1, 2, and 3) that are genetically closest to the source nursery population. Human-mediated movement of infected plants occurring during a period in the past when the pathogen causing SOD was still unknown may explain this pattern.

It is difficult to assess with any certainty the origin and pattern of spread of the pathogen within meta-samples for the simple reason that, by definition, the samples within each meta-sample are statistically genetically similar. With few loci, coalescent approaches, such as those implemented by the software Migrate-N, cannot give reliable inferences on migration rates among poorly differentiated populations. However, it is conceivable that the geographically contiguous populations within meta-samples do represent natural spread. Spatial autocorrelation analyses have suggested that the efficiency of spread by *P. ramorum* decreases sharply between 0 and 350 m to then increase between 350 and 1000 m before finally approaching zero at 10 km (Mascheretti *et al.* 2008). This bimodal correlation pattern suggests that although the rather large sporangia of *P. ramorum* (Werres *et al.* 2001) mostly disperse at short-distances, they may also be dispersed over more moderate distances in strong winds and rain. Spatial autocorrelation analyses were repeated with the current more densely sampled data set (data not shown) and were congruent with the previous findings, with significant isolation by distance up to ~630 m—indicating a slightly larger patch size than previously estimated. As previously reported, values of Moran's *I* increased after 630 m, reaching a peak at 1200 m and approaching zero (no correlation between genetic and geographical distance) at around 10 km. This information, now confirmed by the current larger data set, is essential to model the further spread of the disease and to design appropriate containment measures where appropriate.

The MSN (Fig. 4) displays the evolutionary relationships among MGs and shows the more common MGs in the centre of the network while the rare or singleton MGs occupying the extremities. The current network, based upon a greatly expanded data set, strongly corroborates our earlier conclusions (formally supported by an MCMC analysis) regarding the local evolution of novel and rare genotypes from the more common founder MGs (Mascheretti *et al.* 2008). The biggest meta-sample, meta-sample 3 (Sonoma and Marin Counties, $n = 216$), is characterized by 44 MGs including not only most of the common MGs but also some rarer MGs and 24 singletons. Examining the MSN (Fig. 4), it is apparent that all these MGs are immediate neighbours and that meta-sample 3 is responsible for one of the 'tales' of the network including mostly singletons only present in this meta-sample (MGs 9, 11, 12, 13, 15 and 63, 64, 65). The size of meta-sample 3 and the age of its constituent populations may explain the large number of MGs found in this meta-sample. Because no sexual reproduction is known to occur in California forests, all novel genotypes have to be generated by either mutation or mitotic recombination events. Thus, under the assumption that mutation rates are a constant for any given species, a direct relationship would be expected between cumulative number of genotypes and both the size and age of a sample. This relationship was confirmed by the indices of genotype and gene diversity

(Table 4); older (and larger) meta-samples had higher diversity estimates than younger (and smaller) meta-samples. The presence of a tail of derived genotypes in meta-sample 3 could also be partly due to the early establishment of these infestations, and by the large increase in pathogen population size observed during the wet years of 2000 and 2001 in California.

Meta-sample 1 (Santa Cruz, $n = 108$) provides another example of the same microevolutionary pattern leading to the local creation of new genotypes. This meta-sample acted as a major, and probably as the original dispersal source for the California epidemic, given its genetic affinities to the nurseries. The infestation in these four populations in Santa Cruz County is either old or of intermediate duration and this meta-sample includes 16 MGs, all of which are immediate mutational neighbours. Although three MGs are singletons (MGs 33, 45 and 59) all of the remaining 13 MGs are shared with other meta-samples as would be expected for a founding population.

The inferred 'migration' patterns (Fig. 5A, B) indicate that meta-sample 3 and meta-sample 1 may indeed have been a source for dispersal into most of the other forests. Because our analyses indicated bi-directional gene flow between these two meta-samples, and because both populations may be linked to the original introduction of genotypes of this pathogen into the wild from infected ornamental plants, it may be hard to determine which is the ultimate source of inoculum. For example, in several instances both meta-samples 1 and 3 appear to be the source of migration events into other meta-samples, and while this may indeed be the case, it is also likely that the close relatedness of the two sources may confound the analysis. Because these meta-samples appear to be closely related to the nursery samples in both the NJ and the MDS analyses, the dynamics and timing of the introductions related to each of these two sources may be relatively similar. Nonetheless, discriminating between the two would allow us to differentiate between: (i) a likely human-mediated spread of the pathogen if the source of migration events is meta-sample 1 and (ii) either a natural or a human-mediated spread if meta-population 3 was the source. As stated above, meta-samples 3 and 1 appear to exchange *P. ramorum* by multiple dispersal events, and this may be one reason why meta-sample 2, also closely linked to the nursery populations based on the NJ-tree, appears to divide meta-sample 3 in the MDS plot (Fig 3).

Meta-samples that consist of single populations (i.e. that were not merged with other populations) represent interesting case studies to better understand the creation of genetically distinct units and the validity of the coalescent approach in assessing patterns of gene flow in a clonally reproducing organism. These meta-samples are either isolated, as in the case of Castro Valley, Alameda (AL-1, meta-sample 8) and Angel Island, Marin (MA-2, meta-sample 10), or of recent origin, as in the case of Tomales Bay, Marin (MA-1, meta-sample 9) and the two sites (HU1 and HU2, meta-samples 6 and 7, respectively) in Humboldt County. HU2 is a site located a few miles North of the first infestation ever reported for Humboldt County, HU1, and was known to be free of disease when the infestation at HU1 was discovered. Both coalescent analyses using the untransformed data set and the collapsed data set suggest meta-sample 3 to be the source for HU1: this is the first time that the Humboldt county infestation has been linked to a Californian source. However, the coalescent analysis employing the untransformed data set suggests migration events from HU2 into HU1 and a different source for HU2 (Fig. 5A). By eliminating the confounding effect of locally generated genotypes and by performing a second coalescent analysis (Fig. 5B), the major source of migration into HU2 was identified as HU1.

Meta-samples 8 (AL-1) and 9 (MA-1) represent single populations, in Alameda County and in west Marin County, respectively. Although in both cases at least one of the coalescent analyses indicated meta-samples 1 and 3 as equally important sources for these infestations, framing the results in their appropriate historical and geographical contexts, leads to different conclusions. While a nursery (meta-sample 1) origin is most likely for the old and geographically isolated infestation of Alameda County (meta-sample 8), a source from the neighbouring and established infestation represented by meta-sample 3 is most likely for the recent infestation in West Marin County (meta-sample 9).

Finally, meta-sample 10 represents a single population of the pathogen from Angel Island SP, Marin County (MA-2), about a mile off the coast from the Paradise Drive population (MA-6) included in meta-sample 4. Both meta-samples 2 and 3 are also geographically adjacent to the north (meta-sample 3) and south (meta-sample 2). Both coalescent analyses identified meta-samples 1 and 2 as likely sources of migration events, while only the analysis on the untransformed data set identified a connection between Angel Island and meta-sample 4, the one inclusive of Paradise Drive, the site adjacent to the island. However, the direction of migration appears counterintuitive because it is against the predominant direction of the wind: from the island to meta-population 4 and not vice versa as would be expected. Even considering only the results of the second analysis, it is impossible to discern between human-mediated introduction of nursery genotypes, as suggested by meta-sample 1 being the source, and a

probably natural spread of the pathogen from meta-sample 2 in Marin. A third scenario involving human-mediated transport of genotypes from meta-sample 2 would also be plausible, through movement of soil or plant material between the population of meta-sample 2 on the highly visited Mount Tamalpais and the equally popular Angel Island State Park. Although the Migrate-N analysis using the pooled meta-samples is clearly informative (and more tractable than analysing all populations separately), it is likely that the pooling of apparently genetically similar populations but with differing local history, will produce some confounding in the inferred dispersal patterns. This may be the case for meta-samples 2 and 4, both including populations from different areas and both relevant to understand the spread of the disease into Angel Island. Another issue is that some of the meta-samples result in large sample sizes, which may bias estimates of dispersal events from large into small samples. In order to provide an alternative hypothesis for the colonization of Angel Island by the SOD pathogen, we assumed a single introduction event and looked at the network of genotypes to determine whether all genotypes on the island were closely related. MG35 was identified as a possible and as the most likely ancestor for all genotypes present on the island. That same genotype was found only once in meta-sample 3 in Marin, once in Monterey and three times in Santa Cruz Counties, but twice in Paradise Drive just a mile from the island. We suggest that the Angel Island infestation may most likely be explained by the movement of MG35 from Paradise Drive: considering the insular nature of the site, the direction of the prevailing winds towards the island, and the results of the spatial autocorrelation analysis indirectly identifying 1–5 km as a possible range of movement for the pathogen. When the same approach was tried for the other meta-samples that consist of individual populations, the results were not as clear and easy to interpret as for Angel Island. As highlighted by this example, more sampling and more genetic markers will be needed to refine our understanding of the migration patterns of *P. ramorum*.

Many more MGs were isolated in this work than have been previously reported (Ivors *et al.* 2006; Prospero *et al.* 2007; Mascheretti *et al.* 2008), this probably results from the fact that we sampled a larger number of hosts from new localities and that, with time, new, MGs are generated. In particular, it is worth mentioning MG41 (see Supporting Information) from meta-sample 2 (SM-1, San Mateo County), which is characterized by allele sizes 218/264 at locus18. This combination of alleles at locus18 has been previously identified by Ivors *et al.* (2006) only in *P. ramorum* isolates collected in Europe. However, locus 64, as well as additionally typed microsatellites 29, 33 and 82 (Ivors *et al.* 2006), exhibited the presence of typical North American alleles (microsatellite 29: 325/-, microsatellite 33: 315/337 and microsatellite 82: 110/112/114). In addition, a portion of the mitochondrial *CoxI* gene was sequenced, using primers FM35 and FM55, and was typical of the North American strains. Microsatellite MS18, therefore, is not useful for the identification of European strains, since the allele size previously reported to be private to the European lineage is now also present in the North American NA1 lineage.

In conclusion, it appears that *P. ramorum* may have been introduced multiple times in California. Following the NJ and MDS analyses as rough guides, the first introduction appears to have been from the nurseries into Santa Cruz County, as previously reported (Mascheretti *et al.* 2008). This was followed by introductions into Marin, San Mateo, and Napa Counties. Further introductions appear to have occurred in Marin and Sonoma Counties, while all other areas, including the highly impacted Big Sur region, appear to be the result of more recent introduction events. In most cases, natural spread appears to have stemmed from these introductions resulting in undifferentiated meta-populations, as reported here, for large areas of Santa Cruz, San Mateo, Napa, Marin Sonoma, and Monterey Counties. Previously, the presence of a large undifferentiated population was only thought to be present in Monterey County (Mascheretti *et al.* 2008). A combination of human-mediated and natural spread has thus allowed for an effective dispersal of this exotic pathogen over a very large geographical area in California. This study also confirms that all California infestations, including the ones in Humboldt, are linked to one another and probably independent of the Oregon infestation, as also suggested by the recent work of Prospero *et al.* (2009). At the same time, this study provides evidence that localized genetic differentiation of the pathogen is under way in California due to the lack of effective migration among established infestations combined with the local evolution of new genotypes. This micro-evolutionary differentiation process is shown not only by the presence of structure in populations of the pathogen, but by the discovery of new MGs and by the increasing genetic distances detected both among MGs and among populations as shown by the MSN and MDS analyses. The pattern of relatedness among genotypes and populations was the same independent of the analytical approach used and thus, this study provides a convincing depiction of the genetic structure and, indirectly of the history, of the SOD pathogen in California.

Finally, the creation of an MSN network among genotypes, overlaid on the geographical location of these

genotypes, has enabled us to identify clouds of closely related genotypes that are present at a single site and consequently to differentiate between diversity caused by the local rise of novel genotypes versus that caused by the migration of genotypes from a different location. This knowledge, in turn, was instrumental for an improved implementation of the coalescent analysis to infer migration routes of the SOD pathogen among forest sites in California. Results from the coalescent analysis of the untransformed data set are likely to best explain the first phases of the invasion, when the role of novel genotypes may have been minimal, while results of the second analysis in which derived genotypes were collapsed into their ancestral ones are likely to best explain migration patterns during latter phases of the invasion, when prevalence of novel genotypes was likely to have become more significant. This novel approach has yielded inferences on migration patterns at much smaller temporal and spatial scales than those allowed by traditional nested-clade and coalescent approaches used in the study of other plant pathogens (Carbone & Kohn 2001; Brunner *et al*. 2007; Gladieux *et al*. 2008; Zaffarano *et al*. 2009), and may applicable to the study of the genetic epidemiology of other invasive organisms in general.

Acknowledgements

This study was funded by Ecology of Infectious Diseases grant number EF-0622654 provided jointly by NSF and NIH and by the Gordon and Betty Moore Foundation. Collection of samples predating the NSF-NIH grant was possible thanks to funding by USDA-APHIS and by the USFS, State and Private Forestry.

We are grateful to Yana Valacovich, Christopher Lee, Kenneth Peek and David Rizzo for assisting with collections from Humboldt County (YV and CL), Alameda County (KP) and two sites in Monterey County (DR).

References

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.

Brasier C, Kirk S (2004) Production of gametangia by *Phytophthora ramorum* in vitro. *Mycological Research*, **108**, 823–827.

Brunner PC, Schurch S, McDonald BA (2007) The origin and colonization history of the barley scald pathogen *Rhynchosporium secalis*. *Journal of Evolutionary Biology*, **20**, 1311–1321.

Bruvo R, Michiels NK, D' Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distance irrespective of ploidy level. *Molecular Ecology*, **13**, 2101–2106.

Carbone I, Kohn L (2001) A microbial population – species interface: nested cladistic and coalescent inference with multilocus data. *Molecular Ecology*, **10**, 947–964.

Davidson JM, Warres S, Garbelotto M, Hansen EM, Rizzo DM (2003) *Sudden Oak Death and Associated Disaeses Caused by* Phytophthora ramorum. Online Plant Health Progress (PHP-2003-0707-01-DG). Available at, http://www.plantmanagementnetwork.org/php/shared/sod/.

Davidson JM, Wickland aC, Patterson HA, Falk KR, Rizzo DM (2005) Transmission of *Phytophthora ramorum* in mixed-evergreen forest in California. *Phytopathology*, **95**, 587–596.

Eriwn DC, Ribeiro DC (1996) *Phytophthora Diseases Worldwide*. American Phytopathology Society. APS Press, Saint Paul, MN.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software pakage for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics*, **136**, 343–359.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Garbelotto M, Svihra P, Rizzo DM (2001) Sudden Oak Death syndrome fells 3 oak species. *California Agriculture*, **55**, 9–19.

Garbelotto M, Davidson JM, Ivors K *et al.* (2003) Non-oak native plants are main hosts for sudden oak death pathogen in California. *California Agriculture*, **57**, 18–23.

Gladieux P, Zhang X, Afoufa-Bastien D, Valdebenito Sanhueza R, Sbaghi M, Le Cam B (2008) On the origin and spread of the scab disease of apple: out of Central Asia. *PLoS ONE*, **3**, e1455.

Goss EM, Carbone I, Grunwald NJ (2009) Ancient isolation and independent evolution of the three clonal lineages of the exotic sudden oak death pathogen *Phytophthora ramorum*. *Molecular Ecology*, **18**, 1161–1174.

Grunwald NJ, Goss EM, Larsen MM *et al*. (2008) First report of the European lineage of *Phytophthora ramorum* on *Viburnum* and *Osmanthus* spp. in a California Nursery. *Plant Disease*, **92**, 314.

Hansen EM, Reeser PW, Sutton W, Winton L, Osterbauer N (2003) First report of A1 mating type of *Phytophthora ramorum* in North America. *Plant Disease*, **87**, 1267.

Hardy OJ, Vekemans X (2002) SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Ivors KL, Hayden KJ, Bonants PJ, Rizzo DM, Garbelotto M (2004) AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycological Research*, **108**, 378–392.

Ivors K, Garbelotto M, Vries ID *et al.* (2006) Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Molecular Ecology*, **15**, 1493–1505.

Mascheretti S, Croucher PJP, Vettraino A, Prospero S, Garbelotto M (2008) Reconstruction of the Sudden Oak Death epidemic in California through microsatellite analysis of the pathogen *Phytophthora ramorum*. *Molecular Ecology*, **17**, 2755–2768.

Olarte R, Garbelotto M (2005). Genotyping *Phytophthora ramorum* isolates from U.S. nurseries using PC-RFLP and microsatellite analysis. Sudden Oak Death Science Symposium II, 18 January 2005, Monterey, CA.

Prospero S, Black JA, Winton LM (2004) Isolation and characterization of microsatellite markers in *Phytophthora ramorum*, the causal agent of sudden oak death. *Molecular Ecology Notes*, **4**, 672–674.

Prospero S, Hansen EM, Grunwald NJ, Winton LM (2007) Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Molecular Ecology*, **16**, 2958–2973.

Prospero S, Grunwald NJ, Winton LM, Hansen EDM (2009) Migration patterns of the emerging plant pathogen *Phytophthora ramorum* on the west coast of the United States of America. *Phytopathology*, **99**, 739–749.

Rizzo DM, Garbelotto M (2003) Sudden oak death: endangering California and Oregon forest ecosystems. *Frontiers in Ecology and the Environment*, **1**, 197–204.

Rizzo DM, Garbelotto M, Davidson JM, Slaughter GW, Koike ST (2002) Phytophthora ramorum as the cause of extensive mortality of *Quercus* spp. and lithocarpus densiflora in California. *Plant Disease*, **86**, 205–214.

Rizzo DM, Garbelotto M, Hansen EM (2005) Phytophthora ramorum: integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology*, **43**, 309–335.

Roewer L, Croucher PJP, Willuweit S *et al.* (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genetics*, **116**, 279–291.

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society*, **10**, 199–228.

Stoddart JA, Taylor JF (1988) Genotypic diversity: estimation and prediction in samples. *Genetics*, **118**, 705–711.

Werres S, Marwitz R, In' T Veld WA *et al.* (2001) *Phytophthora ramorum sp. nov.*, a new pathogen on Rhododendron and Viburnum. *Mycological Research*, **105**, 1155–1165.

Zaffarano PL, McDonald BA, Linde CC (2009) Phylogeographical analyses reveal global migration patterns of the barley scald pathogen *Rhynchosporium secalis*. *Molecular Ecology*, **18**, 279–293.

SM is a molecular biologist interested in the epidemiology and genetic basis of human and plant diseases, PC is an evolutionary biologist and statistical population geneticist interested in genetic epidemiology and the genomics of selection and adaptive radiation. MK is a doctoral student interested in evolutionary genetic of forest pathogens. LB is currently doctoral student at the University of Hawaii. MG is a mycologist and plant pathologist interested in the ecological and evolutionary consequences of the introduction of exotic diseases.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Table S1 Allele size (indicated in bp of the PCR amplification products) of the 82 multilocus genotypes (MGs) identified in the 669 *Phytophthora ramorum* isolates

Table S2 Distribution of the 82 multilocus genotypes in the 669 isolates grouped by population and meta-sample

Table S3 Correspondence between the 82 multilocus genotypes (MGs) identified in the 669 *Phytophthora ramorum* isolates investigated in the present study and the 35 multilocus genotypes identified in the 292 *P. ramorum* isolates investigated in the previous study (Mascheretti *et al.* 2008)

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.