Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach
Constance Xhaard, Benoît Barres, Axelle Andrieux, Lydia Bousset, Fabien Halkett, Pascal Frey

To cite this version:
Constance Xhaard, Benoît Barres, Axelle Andrieux, Lydia Bousset, Fabien Halkett, et al.. Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. Molecular Ecology, 2012, 21 (10), pp.2383-2398. 10.1111/j.1365-294X.2012.05556.x. hal-01191214

HAL Id: hal-01191214
https://hal.science/hal-01191214
Submitted on 1 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach

CONSTANCE XHAARD,*§ BENOÎT BARRÈS,† AXELLE ANDRIEUX,* LYDIA BOUSSET,‡ FABIEN HALKETT*– and PASCAL FREY*–

*INRA, Université de Lorraine, UMR1136 Interactions Arbres – Microorganismes, F-54280 Champenoux, France, †INRA, Univ. Bordeaux, UMR1202 BIOGECO, F-33610 Cestas France, ‡INRA, Agrocampus Ouest, Université Rennes 1, UMR1349 Institut de Génétique, Environnement et Protection des Plantes, F-35653 Le Rheu, France

Abstract

The advent of molecular epidemiology has greatly improved our ability to identify the population sources and track the pathogen movement. Yet the wide spatial and temporal scales usually considered are useful only to infer historical migration pathways. In this study, Bayesian genetic assignments and a landscape epidemiology approach were combined to unravel genetic origin and annual spread during a single epidemic of a plant pathogen: the poplar rust fungus Melampsora larici-populina. The study focused on a particular area—the Durance River valley—which enabled inoculum sources to be identified and channelled spread of the epidemic along a one-dimensional corridor. Spatio-temporal monitoring of disease showed that the epidemic began in the upstream part of the valley and spread out downstream. Using genetic assignment tests, individuals collected at the end of the epidemic were sorted into two genetic groups; very few hybrids were detected, although individuals from both groups coexisted locally downstream in the valley. The epidemic originated from two genetically distinct inoculum sources. Individuals of each group then dispersed southwards along the Durance River and became mixed in poplar riparian stands. These two genetic groups were found previously at a wider spatial scale and proved to result from distinct evolutionary histories on either wild or cultivated poplars. This study showed that the two groups can mix during an epidemic but do not hybridize because they then reproduce asexually. In general, the methods employed here could be useful for elucidating the genetic origin and retracing the colonization history and migration pathways of recent epidemics.

Keywords: biological invasions, emerging diseases, gene flow, landscape epidemiology, Melampsora larici-populina, poplar rust fungus

Received 28 September 2011; revision received 26 December 2011; accepted 10 January 2012

Introduction

Tracking pathogen movement, identifying source populations and understanding environmental factors (including human activities) that influence pathogen spread are central tasks in disease ecology (Milgroom & Peever 2003; Ostfeld et al. 2005; Archie et al. 2009). The advent of molecular biology has afforded new opportunities in the field of epidemiology (Milgroom 2001). In particular, molecular markers have greatly improved our ability to infer the dispersal routes of pathogens—at the very least corroborating previous classical surveys but most often providing additional clues as to the epidemiological history of disease progression (Milgroom & Peever 2003; de Meeûs et al. 2007; Archie et al. 2009).
Studies on the historical spread of pathogens often rely on phylogeographical analyses (Criscione et al. 2005; Archie et al. 2009; Estoup & Guillemaud 2010), the main aims of which are to (i) highlight the centres of disease origin (e.g. Gómez-Alpizar et al. 2007; Kauserrud et al. 2007; Stukenbrock et al. 2007; Gladieux et al. 2008) and (ii) infer the domestication routes of plant pathogens (Hovmøller et al. 2008; Munkaci et al. 2008; Zaffarano et al. 2008; Gladieux et al. 2010). These classical studies hence provide accurate and valuable information that has helped to unravel the evolutionary history of various diseases. However, they may be less relevant to the design of current management policies. In particular, because they focus on very large temporal and spatial scales, such phylogeographical analyses can decipher only historical migration pathways, and they provide little information on the origin and spread of pathogens during current epidemics (Milgroom & Peever 2003; Archie et al. 2009).

Many plant pathogens experience yearly extensive fluctuations in population size and geographical range (Burdon et al. 1989). This is especially true for crop pathogens in temperate climates. Plant disease epidemics collapse annually because of host senescence (e.g. leaf fall) or crop harvest, and recur the following year when conditions again become favourable (Mundt et al. 2011). These ‘boom and bust’ cycles have important consequences for the population genetics of plant pathogens, in terms of both spatial and temporal genetic differentiation (Bucheli et al. 2001; Zeller et al. 2004). In addition to classical founder effects linked to colonization events (Carlier et al. 1996; Barrès et al. 2008; Halkett et al. 2010), some of this temporal and spatial genetic differentiation can also result from the nonrandom recruitment of colonizing genotypes from different source populations (Burdon et al. 1989; Criscione et al. 2011). In the context of large demographic fluctuations and complex life histories, defining a priori exactly what the parasite population is (i.e. as a homogeneous genetic entity) is a challenging task (Waples & Gaggiotti 2006; Barrett et al. 2008) requiring the disentangling of two processes: pathogen dispersal from inoculum sources, and population genetic mixing during the course of an epidemic. A major impediment to defining pathogen populations is the difficulty of appropriately combining population genetics analyses and epidemiological surveys at a common (hence local) spatial scale. Such molecular epidemiology analysis is required because the timescale considered is too short to accurately decipher dispersal pathways from molecular data only (no accumulation of mutations). As a consequence, the coalescent theory used in phylogeography studies no longer applies at such temporally and spatially restricted scales (Wang 2010). Instead, Bayesian assignment algorithms (Pritchard et al. 2000), which are designed to cluster individuals into homogeneous genetic groups, can be used to delineate pathogen populations, infer the source population of parasites and quantify the gene flow between populations (Waples & Gaggiotti 2006; Archie et al. 2009). Assignment tests became tremendously popular (Kaeuffer et al. 2007), especially with the increasing availability of molecular markers such as microsatellite markers and high-quality genetic data sets, which enable well-resolved multilocus genotypes to be obtained and hence cryptic population structure to be assessed (Criscione et al. 2011; Xhaard et al. 2011). However, whether such assignment tests provide additional value in deciphering pathogen genetic structure resulting from a single epidemic remains to be tested.

A second challenge stems from the fact that following pathogen spread at a landscape scale is a difficult task (Holdenrieder et al. 2004; Plantegenest et al. 2007). In this study, we aimed to benefit from the specificities of a given pathogen life cycle (host alternation) to more easily monitor an epidemic. Host-alternating pathogens switch host species to complete their life cycle, performing their annual sexual reproduction stage on the alternate host and then multiplying asexually on the primary host for several months. It is this asexual multiplication on the primary host that causes the epidemics. Host alternation is common among rust fungi (Pucciniales) (Agrios 2005). In this type of life cycle, the starting point of a recurrent annual epidemic is defined by locations where the two hosts live in sympatry (Bourassa et al. 2007). Provided the geographical distributions of the hosts are known and restricted, it is possible to document the spread of disease on the primary host from starting points located in the sympatry area (Zogg 1949).

The present study focused on the poplar rust fungus *Melampsora larici-populina*, which infects poplar (*Populus* sp.), the primary (or telial) host, and larch (*Larix* sp.), the alternate (or aecial) host. The advantages of focusing on this species are twofold: first, the ecology and distribution areas of the two hosts are well known (Riou-Nivert 2001; Imbert & Lefèvre 2003; Pinon & Frey 2005) and, second, cryptic population genetic structures in this fungal species have been evidenced recently (Xhaard et al. 2011). The Eurasian black poplar, *Populus nigra*, is a pioneer tree species of riparian forests that is distributed widely across Europe and Asia (Imbert & Lefèvre 2003). In Europe, wild poplar populations coexist with cultivated stands (Pinon & Frey 1997). Many poplar cultivars carry qualitative resistance genes selected for their resistance to rust attack. Conversely, qualitative resistance has not been detected so far in *P. nigra* (Legionnet et al. 1999; Pinon & Frey 2005).
Cultivated poplars thus exert a strong and specific selective pressure on *M. larici-populina* individuals, and only those carrying the virulence corresponding to a given resistance gene can infect resistant hosts (Gérard et al. 2006). Evolving according to the gene-for-gene relationship (Flor 1971), *M. larici-populina* populations have overcome all the qualitative resistances deployed to date in poplar cultivation (Pinon & Frey 2005). Consistently, Xhaard et al. (2011) demonstrated recently that poplar rust populations have been greatly impacted by poplar domestication. A cryptic population genetic structure was revealed at a regional scale, with three genetic groups displaying contrasting genetic signatures. The most abundant group (‘cultivated’ group) showed the hallmarks of a history of strong selection by host resistance and subsequent changes in demography. The second group (‘wild’ group) was closer to genetic equilibrium, in accordance with long-standing co-evolution with the wild host. A third group (‘asexual’ group) was shown to result from asexual overwintering of the fungus. Although the ‘cultivated’ and ‘wild’ genetic groups prevail in distinct geographical areas, they were also found mixed in some locations at the end of epidemics. These cases of local coexistence are all the more puzzling because very few hybrids were found. The present study tested the hypothesis that such local mixtures result from the gathering together of poplar rust genotypes during the course of epidemics on the primary host.

A study area—the Durance River valley located in the French Alps (Fig. 1)—was chosen that allows easy completion of the *M. larici-populina* life cycle and accurate monitoring of successive epidemic stages. The Durance River flows north to south and is bordered by a nearly continuous riparian forest of wild poplar (*P. nigra*) of between 0.5 and 2 km wide. The upstream part of the valley encompasses part of the native range of European larch, *Larix decidua* (i.e. the ‘Mélezin’ in the central Alps, Riou-Nivert 2001). The presence of larch is restricted to the northern part of the area. This poplar-larch sympathy area restricts the sexual reproduction of *M. larici-populina* to this zone, which was thus postulated to be the starting point of the annual poplar rust epidemic (Frey et al. 2005). The annual asexual epidemic on poplars is assumed to spread southwards through the riparian stands of wild poplars, channelled along the river by the reliefs bordering the valley.

Hence, the Durance River valley constitutes a one-dimensional ecological corridor in which epidemics of *M. larici-populina* are constrained, thus facilitating monitoring of their course. This simple ecological design is somewhat complicated by the presence of some cultivated poplar stands in both the sympathy and allopatry areas. On wild poplars, ‘wild’ and ‘cultivated’ groups of *M. larici-populina* were found to be segregated spatially in the northern part of the study area (i.e. in the area of poplar-larch sympathy) (Xhaard et al. 2011). Assuming that different foci in the poplar-larch sympathy area exist, these two groups can act as different population sources for the epidemic that spreads each year along the Durance River. In the time-course of an epidemic, it was thus suspected that southern poplar stands could become colonized by individuals with different genetic profiles, leading to the genetic mixture observed (Xhaard et al. 2011). Conversely, if the ‘asexual’ group contributes significantly to the inoculum...
source, then the poplar rust epidemic would develop simultaneously all along the valley, and no pattern of spatio-temporal disease progression would be expected.

In this study, disease progression was monitored and the genetic profiles of *M. larici-populina* individuals collected along the Durance River valley at the end of the annual epidemic were analysed. The aim was to disentangle the genetic origins of this poplar rust epidemic. The hypothesis that two source populations contribute to the epidemic was tested, and the level of genotype mixing along the Durance River was assessed. To achieve this aim, both selected (virulence) and neutral (microsatellites) markers were used, and genetic partition of *M. larici-populina* individuals was tested using several Bayesian assignment tests.

**Materials and methods**

**Site monitoring and sampling details**

The sampling area is a corridor (c. 200 km long) in the Durance River valley in the French Alps (Fig. 1). Fifteen sites along or near the Durance River were sampled: 10 wild stands of *Populus nigra* and five cultivated stands containing the hybrid poplar *P. × interamericana* ‘Beaupré’, from which individuals were sampled (Table 1). Distances between contiguous sampled sites ranged from 5 to 32 km. Two sampled cultivated stands and four wild stands were located within the poplar-larch sympatry area in the upstream part of the river (Fig. 1). Two other cultivated stands were located in the vicinity of sampled wild stands (Valernes and Mirabeau) in the allopatry area (without larch).

The presence of disease symptoms (rust pustules) was monitored for both wild and cultivated sites every 3 weeks between July and October 2004. Hence, the date of first detection of rust symptoms was known for each site. Because the incidence of rust is generally lower in wild than in cultivated stands (Prakash & Thielges 1987; Chen & Harrington 2006), individuals from the wild sites were all collected at the same time in October, at the end of the epidemic. The incidence of rust is much higher in cultivated stands and can lead to early defoliation (Gérard et al. 2006). Therefore, the five cultivated sites were sampled between July and October, depending on local disease incidence. Infected leaves were collected on *P. nigra* in the wild sites and on *P. × interamericana* ‘Beaupré’ in the cultivated stands. To avoid resampling of clonal replicates, each individual was isolated from a distinct leaf collected from different trees and twigs. Whenever possible, 100–110 leaves were collected randomly for each wild site from up to 50 trees. The very focal infection at Valernes allowed the collection of only 30 leaves from five trees. The infection level of *P. nigra* trees at Saint-Clément, Châteauroux, La Brillanne and Mirabeau was also very low, even at the end of the epidemic, and only 10, 18, 22 and 5 infected leaves were collected from each of these sites, respectively. In cultivated stands, about 30 infected leaves were collected from 20 to 30 trees. The

![Table 1. Characteristics of collection sites from wild and cultivated poplar stands and rust disease monitoring (1, rust detected and 0, no rust detected)](image)

<table>
<thead>
<tr>
<th>Collection site</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Host plant</th>
<th>7 July</th>
<th>28 July</th>
<th>18 August</th>
<th>8 September</th>
<th>29 September</th>
<th>13 October</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild stands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prellès</td>
<td>44°51’00” N</td>
<td>6°34’47” E</td>
<td><em>Populus nigra</em></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Saint-Clément</td>
<td>44°39’04” N</td>
<td>6°35’09” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Châteauroux</td>
<td>44°37’18” N</td>
<td>6°32’47” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Embrun</td>
<td>44°33’37” N</td>
<td>6°30’42” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rochebrune</td>
<td>44°27’36” N</td>
<td>6°11’55” E</td>
<td><em>P. nigra</em></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Monêtier</td>
<td>44°22’50” N</td>
<td>5°56’43” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Valernes</td>
<td>44°14’54” N</td>
<td>5°55’18” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Les Mées</td>
<td>44°02’05” N</td>
<td>5°57’52” E</td>
<td><em>P. nigra</em></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>La Brillanne</td>
<td>43°55’21” N</td>
<td>5°53’60” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mirabeau</td>
<td>43°41’26” N</td>
<td>5°40’22” E</td>
<td><em>P. nigra</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cultivated stands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR Pont-du-Fossé</td>
<td>44°40’28” N</td>
<td>6°13’59” E</td>
<td>Beaupré</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BPR Manse</td>
<td>44°37’11” N</td>
<td>6°08’04” E</td>
<td>Beaupré</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BPR Valernes</td>
<td>44°14’58” N</td>
<td>5°55’20” E</td>
<td>Beaupré</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BPR Manosque</td>
<td>43°48’57” N</td>
<td>5°48’59” E</td>
<td>Beaupré</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BPR Mirabeau</td>
<td>43°42’31” N</td>
<td>5°42’56” E</td>
<td>Beaupré</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

© 2012 Blackwell Publishing Ltd
high infection level in cultivated stands (which are never sprayed with fungicide against rust) ensures easy collection of infected leaves, except in Manosque where only eight infected leaves were collected.

**Isolation of individual genotypes**

One single uredinium (sporulating lesion of *Melampsora larici-populina* producing urediniospores) per leaf was selected randomly and grown on fresh leaf discs of *P. × euramericana* ‘Robusta’ as described by Barrès et al. (2008). Mono-uredinal isolation ensures that a single genotype per sampled leaf is isolated and multiplied, so that both virulence phenotype assessment and molecular characterization can be performed on the same individual genotype.

**Assessment of virulence profiles**

All *M. larici-populina* individuals were inoculated on two poplar cultivars: one susceptible cultivar (*P. × euramericana* ‘Robusta’) and the second carrying the *K7* rust resistance (*P. × interamericana* ‘Beaupré’), as described by Barrès et al. (2008). The success of the differential infection revealed the virulence phenotype, denoted Vir7 if both resistant and susceptible leaf discs were infected, and Avr7 if only susceptible (control) leaf discs were infected. After virulence scoring, leaf discs of ‘Robusta’ were stored at −20 °C until DNA extraction.

**Microsatellite analysis**

When more than 50 individuals per site were available, a subset of only 48 individuals was selected randomly for DNA analyses. Otherwise, all individuals were retained (Table 2). DNA was extracted using a DNeasy® 96 Plant kit (Qiagen, Hilden, Germany) as described in the study by Barre’s et al. (2009). The resulting adjusted threshold for the probability values, indicating the multiple copies of the same multilocus genotype that did not occur by chance (true clones). Calculations took into account the population size and allele frequencies.

In order to minimize departure from Hardy–Weinberg (HW) expectations and linkage disequilibrium (LD), on which several analyses are based, only one copy of each multilocus genotype (found to be significantly over-represented) was considered for all subsequent genetic analyses.

**Analysis of genetic data**

**Identification of repeated multilocus genotypes.** Data were checked for misprint and scoring errors using Micro-CHECKER software (Van Oosterhout et al. 2004). Identical genotypes were identified and tested further using MEGA software (Stenberg et al. 2003) to determine whether individuals with identical multilocus genotype were indeed true clones. First, the probability of observing a multilocus genotype *n* times in a population was computed. Second, using a Monte Carlo simulation method, the program determines the significance threshold for the probability values, indicating the multiple copies of the same multilocus genotype that did not occur by chance (true clones). Calculations took into account the population size and allele frequencies.

**Standard population genetics analysis.** Standard population genetics analyses were carried out before Bayesian assignment tests on each sampled location, and after Bayesian assignment tests on each genetic group, to further assess their characteristics. Expected and observed heterozygosities were calculated using GENETIX (4.05) (Belkhir et al. 1996–2004). As the observed number of alleles in a sample is highly dependent on sample size, allelic richness (*A*) was calculated using FSTAT version 2.9.3 (Goudet 1995). *F*-statistics were computed using FSTAT according to Weir & Cockerham (1984). The level of LD between loci was measured by the multilocus estimate disequilibrium (*r*) based on the index of association corrected for the number of loci (Agapow & Burt 2001). Deviations from HW expectations (heterozygote excess and deficit) were calculated using GENEPOP on the web (http://genepop.curtin.edu.au/) (Rousset 2008). In order to adjust the *P*-values for multiple tests for LD, the false discovery rate (FDR) procedure (Benjamini & Yekutieli 2001) was used, which controls for the proportion of significant results (FDR) instead of controlling for the chance of making even a single error. The resulting adjusted *P*-values are called *Q* values. This procedure is implemented in the R package pQVALUE (Storey & Tibshirani 2003). Finally, the correlation between genotypic assignment results and phenotypic
information was tested through a logistic regression between individual memberships and virulence profiles (Boolean data, 1 for Vir7 and 0 for Avr7 individuals).

Population structure and Bayesian assignment analysis. The genetic differentiation between individuals was first visualized with a correspondence analysis performed using GENETIX (version 4.05) (Belkhir et al. 1996–2004). The partition of the data set into different genetic groups was further tested using Bayesian genetic assignment tests. Three different software products were used to check the consistency of the partitioning and the assignment results. First, STRUCTURE version 2.2 (Pritchard et al. 2000; Falush et al. 2003)—the reference software for such analyses (Evanno et al. 2005)—was used with the admixture model with correlated allele frequency and default settings. Iteration parameters were set to a burning-in period of 30 000 iterations followed by 1 200 000 iterations. The likelihood of the assignments [In P(D)] was evaluated for K (the number of genetic groups) varying from 1 to 10. Three independent simulations were performed for each value of K to test for consistency of results. To assess the value of K that best explains the partitioning of the genetic data, the AK estimator was computed as described by Evanno et al. (2005).

Second, GENELOG (version 3.1.4) (Guillot et al. 2005b; Guillot 2008) was used, without including spatial information, to cross-validate the assignment procedure. The strength of this second method is that the best number of groups, K, is inferred during a single run (unlike in STRUCTURE) because of the specific Reverse Jump MCMC (Markov chain Monte Carlo) algorithm implemented, which enables the Markov chain to explore the whole range of putative cluster numbers at each step of the simulation. Hence, the number of groups that best explains the data is given by the dimension value (here the value of K) in which the Markov chain spent most iterations (Guillot et al. 2005a,b). The value of K was allowed to vary from 1 to 10 and initialized at 8. Once the most likely K value was determined, additional runs were performed, with the corresponding value of K fixed, to better infer the membership level of each individual to the K groups, as advised in the study by Guillot et al. (2005b). For runs with varying K, the uncorrelated allele frequency model was used, and for those with fixed K, the correlated allele frequency model (Guillot 2008) with default settings was used, except for the number of nuclei, which, given the number of individuals, was increased to 500 (in the nonspatial mode). Iteration parameters were set to 550 000 iterations with a burning-in period of 50 000 iterations.

Finally, NEWHYBRIDS (version 1.1), which implements a Bayesian genetic assignment approach dedicated to the quantification of hybridization events (Anderson & Thompson 2002), was used. The detection of backcrosses requires many more loci (almost 50) than the 26 used here (Vaha & Primmer 2006). We therefore focused on first-generation hybrids between parental classes. The Bayesian method allows either loading of prior knowledge of the allele frequencies of the parental classes or their estimation from the offspring genetic data set. Here, the program was used without any prior information. The following parameters were used: a short burn-in period of 1000 iterations, followed by 1 000 000 iterations. With each of the three Bayesian clustering methods, three independent runs were performed to check for consistency in the convergence of Markov chains. Whatever the method employed, the same threshold of membership coefficient (q > 0.9) was fixed for assigning individuals to a cluster.

Analysis of isolation by distance. Isolation by distance (IBD) was assessed within each genetic group (considering only locations with at least five assigned individuals) using the continuous method (Rousset 2000) of the ‘Isolde’ procedure implemented in GENEPOP (Rousset 2008; option ‘Analysis of isolation by distance between groups’). The significance of Mantel tests between pairwise genetic distances [calculated as FST/(1 − FST) ratios] and geographical distances was determined using the ISOLDE program (10 000 permutations). For geographical distances, Euclidian distances between pairs of sampling sites were considered. It should be noted that the aim here was not to infer dispersal distance from the slope of the regression line, as this biological system does not meet the requirements of the IBD model (Rousset 1997). Instead, the aim was to decipher whether the observed variation in pairwise genetic distance is gradual (according to geographical distance) or random.

Results

Timing of disease progression along the Durance River valley

Poplar leaf infections were first detected at the beginning of summer in the upstream part of the valley, where larches grow in sympatry with poplar trees (Fig. 1 and Table 1). Other sites then became infected according to their location downstream in the valley, indicating that the epidemic begins in the upstream part of the valley and then spreads out downstream following the river. This spatio-temporal dynamic was particularly obvious with the wild sites. Nonetheless, there were two exceptions to this regular spread. Infection began very early in Les Méès, at the beginning of
the survey period (7 July), although this site is located far from the poplar-larch sympathy area. Conversely, a late infection (8 September) was observed in Saint-Clément, although this site is located within the poplar-larch sympathy area. In the cultivated poplar stands, the epidemic also spread downwards through the valley. Nonetheless, the locations of the cultivated stands—two in the poplar-larch sympathy area in the north and the remaining three in the south part of the valley—did not allow proper quantification of disease progression (compared to the wild sites). Indeed, there was a notable discrepancy in the relative timing of infection between wild and cultivated stands as the latter became infected earlier in the downstream part of the valley (Table 1).

**Site-by-site virulence profiles**

The frequency of individuals carrying Vir7 virulence ranged from 0 to 1 across sampling sites, with a mean proportion of 0.33 (Table 2). All individuals collected on ‘Beaupré’ exhibited a Vir7 phenotype, as expected given the presence of the R7 resistance in this cultivar (Table 2). For individuals sampled on *Populus nigra*, the frequency of Vir7 ranged from 0% (Saint-Clément and Châteauroux) to 44% (Valernes; Table 2). Samples from the wild stands located in the upstream part of the valley (Prelles, Saint-Clément, Châteauroux and Embrun) exhibited a very low percentage (0–3%) of Vir7 individuals. In contrast, the proportion of Vir7 individuals was much higher (17–44%) in all wild sites further downstream in the Durance River valley, except in Les Mées (3%) (Table 2).

**Overall genetic and genotypic diversity**

Of the 478 individuals genotyped, only eight were excluded from further analyses because they failed to amplify. The remaining 470 individuals exhibited only 0.6% missing data on average. All 26 microsatellite markers were highly polymorphic, with the number of alleles ranging from 4 to 19 (mean number across loci: 7.11).

Of the 470 individuals analysed, 331 distinct multilocus genotypes were found. All the repeated multilocus genotypes were found to be significantly over-represented according to the analysis performed using MLGSim. Repetitions of a given multilocus genotype were thus considered to result from clonal reproduction only and were further denoted as clonal copies. Repeated genotypes were found in both wild and cultivated stands (Table 2). Three repeated genotypes were each highly represented, with 34, 30 and 20 clonal copies. The remaining 29 repeated multilocus genotypes occurred in two to ten clonal copies. Almost all repeated genotypes were found within a single site, except for six repeated genotypes that were found at four distinct sites, with a maximal resampling distance of 126 km. Among the wild stands, repeated genotypes predominated in the downstream region of the Durance River valley, especially in Les Mées and La Brillanne (Table 2).

The genetic diversity of *Melampsora larici-populina* samples was significantly higher among wild stands ($H_E = 0.474$) compared to cultivated stands ($H_E = 0.403$, $P < 0.05$). Focusing on the wild stands, the highest genetic diversity was found in the sympathy area at Prelles (Table 2). Significantly more heterozygotes were found in the sympathy area (Prelles to Embrun locations, $H_O = 0.475$) than in the allopatry area (Rochebrune to Mirabeau, $H_O = 0.428$, $P < 0.05$). Conspicuously, the two uppermost sites among the wild and cultivated stands (Prelles and Pont-du-Fosse, respectively) were among the very few sites found to be in genetic equilibrium (both HW equilibrium and random association of alleles across loci, Table 2). Most wild stands displayed significant LD, which, together with high $F_{IS}$ values in the downstream part of the Durance River valley, could denote some genetic mixing (Wahlund effect).

**Genetic structure and assignment results**

Individual genotypes were separated largely according to their Vir7 or Avr7 profile in the correspondence analysis (Fig. 2). Most Vir7 individuals were clustered, regardless of their sampling origin (wild or cultivated stands). While the first axis differentiated individuals according to their Vir7/Avr7 profile, the second differentiated individuals collected mostly from wild sites. Thus, individuals from the cultivated stands were clumped together, in contrast to those from wild stands, which were much more scattered. However, the first two axes explained only 5.8% of the total inertia. The relevance of such a genetic structure among sampled individuals was therefore subsequently explored through Bayesian assignment tests.

All Bayesian assignment tests were conducted to include only one copy of each repeated multilocus genotype in order to minimize departure from the HW expectations on which the assignment procedures were based. Regardless of the method employed, Markov Chains converged rapidly, and the three replicate runs gave consistent results. The STRUCTURE and GENELAND algorithms were in agreement, both gathering individuals into two clusters (Table 3). However, GENELAND seemed to perform better and gave clearer results: (i) Markov Chain for GENELAND runs settled rapidly on
K = 2, whereas likelihood analysis using Structure runs showed bimodal values for the best number of clusters; and (ii) the assignment profile was more clear cut for Geneland (Fig. 3). Applying an assignment coefficient of $q > 0.9$ led to the exclusion of 93 (28%) and 22 (7%) individuals in Structure and Geneland, respectively.

A second step further investigated whether the non-assigned individuals revealed by Structure and Geneland could result from hybridization events between individuals of the two genetic clusters. Interestingly, the assignment profile obtained with NewHybrids was very close to that obtained with Geneland (data not shown). In particular, similar membership values were observed.

### Table 2 Characteristics of Melampsora larici-populina individuals analysed from each location

<table>
<thead>
<tr>
<th>Collection site</th>
<th>$N_P$</th>
<th>Vir7</th>
<th>$N_G$</th>
<th>G</th>
<th>$G/N_G$</th>
<th>$A_r$</th>
<th>$H_e$</th>
<th>$H_o$</th>
<th>$F_{Is}$</th>
<th>$r_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild stands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prelles</td>
<td>103</td>
<td>0.02</td>
<td>48</td>
<td>48</td>
<td>1.00</td>
<td>2.11</td>
<td>0.501</td>
<td>0.486</td>
<td>0.031</td>
<td>0.008</td>
</tr>
<tr>
<td>Saint-Clément</td>
<td>10</td>
<td>0.00</td>
<td>10</td>
<td>10</td>
<td>1.00</td>
<td>2.02</td>
<td>0.467</td>
<td>0.484</td>
<td>−0.040</td>
<td>0.065**</td>
</tr>
<tr>
<td>Châteauroux</td>
<td>16</td>
<td>0.00</td>
<td>16</td>
<td>12</td>
<td>0.75</td>
<td>1.80</td>
<td>0.393</td>
<td>0.467</td>
<td>−0.197**</td>
<td>0.731***</td>
</tr>
<tr>
<td>Embrun</td>
<td>102</td>
<td>0.03</td>
<td>48</td>
<td>44</td>
<td>0.92</td>
<td>2.04</td>
<td>0.482</td>
<td>0.463</td>
<td>0.040</td>
<td>0.019***</td>
</tr>
<tr>
<td>Rochebrune</td>
<td>84</td>
<td>0.42</td>
<td>46</td>
<td>39</td>
<td>0.85</td>
<td>2.03</td>
<td>0.484</td>
<td>0.431</td>
<td>0.110***</td>
<td>0.023***</td>
</tr>
<tr>
<td>Montréal</td>
<td>97</td>
<td>0.35</td>
<td>47</td>
<td>45</td>
<td>0.96</td>
<td>2.06</td>
<td>0.489</td>
<td>0.443</td>
<td>0.094***</td>
<td>0.038***</td>
</tr>
<tr>
<td>Valernes</td>
<td>108</td>
<td>0.44</td>
<td>48</td>
<td>44</td>
<td>0.92</td>
<td>1.96</td>
<td>0.445</td>
<td>0.407</td>
<td>0.088***</td>
<td>0.049***</td>
</tr>
<tr>
<td>Les Mées</td>
<td>108</td>
<td>0.03</td>
<td>48</td>
<td>13</td>
<td>0.27</td>
<td>1.98</td>
<td>0.459</td>
<td>0.439</td>
<td>0.045</td>
<td>0.156***</td>
</tr>
<tr>
<td>La Brillanne</td>
<td>24</td>
<td>0.17</td>
<td>24</td>
<td>5</td>
<td>0.21</td>
<td>1.84</td>
<td>0.413</td>
<td>0.448</td>
<td>−0.098</td>
<td>0.333***</td>
</tr>
<tr>
<td>Mirabeau</td>
<td>5</td>
<td>0.40</td>
<td>5</td>
<td>5</td>
<td>1.00</td>
<td>1.89</td>
<td>0.413</td>
<td>0.437</td>
<td>0.116</td>
<td>0.020</td>
</tr>
<tr>
<td><strong>Cultivated stands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPR pont-du-Fossé</td>
<td>30</td>
<td>1.00</td>
<td>30</td>
<td>27</td>
<td>0.90</td>
<td>1.79</td>
<td>0.387</td>
<td>0.393</td>
<td>−0.015</td>
<td>0.003</td>
</tr>
<tr>
<td>BPR Manse</td>
<td>32</td>
<td>1.00</td>
<td>32</td>
<td>15</td>
<td>0.47</td>
<td>1.92</td>
<td>0.440</td>
<td>0.420</td>
<td>0.047</td>
<td>0.070***</td>
</tr>
<tr>
<td>BPR Valernes‡</td>
<td>29</td>
<td>1.00</td>
<td>29</td>
<td>3</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>—</td>
</tr>
<tr>
<td>BPR Manosque</td>
<td>9</td>
<td>1.00</td>
<td>9</td>
<td>8</td>
<td>0.89</td>
<td>1.83</td>
<td>0.403</td>
<td>0.400</td>
<td>0.008</td>
<td>0.047**</td>
</tr>
<tr>
<td>BPR Mirabeau</td>
<td>30</td>
<td>1.00</td>
<td>30</td>
<td>18</td>
<td>0.60</td>
<td>1.85</td>
<td>0.398</td>
<td>0.387</td>
<td>0.027</td>
<td>0.005</td>
</tr>
</tbody>
</table>

$N_P$, number of individuals phenotyped; Vir7, proportion of Vir7 individuals; $N_G$, number of individuals genotyped; G, number of distinct genotypes; $G/N_G$, genotypic richness; $A_r$, allelic richness, estimated for a sample size of three individuals; $H_e$, unbiased estimate of gene diversity (Nei 1978); $H_o$, observed heterozygosity; $F_{Is}$, departure level and test of Hardy–Weinberg equilibrium; $r_D$, multilocus estimate and test of linkage disequilibrium. Significant levels of $P$-values were corrected for multiple comparisons using false discovery rate (FDR) control (*$Q$ value <0.05; **$Q$ value <0.01; ***$Q$ value <0.001).

1Measures of genetic diversity and genetic equilibrium tests were performed considering only one copy per repeated multilocus genotype.

2Genetic diversity indices and test of genetic equilibrium were not conducted on this site because the number of distinct genotypes was too small.

![Fig. 2](https://example.com/figure2.png) Correspondence analysis performed with Genetix on the 331 multilocus genotypes of Melampsora larici-populina. ▲ cultivated sites (Vir7 individuals), ○ wild sites (Avr7 individuals), ● wild sites (Vir7 individuals).
for individuals previously assigned (membership coefficient \( q > 0.9 \)) to the two clusters. Among the previously nonassigned individuals, a single multilocus genotype (occurring twice in the sample) was undisputedly a first-generation hybrid when considering the same threshold of \( q > 0.9 \). The other nonassigned individuals had varying proportions of their genome attributed to the hybrid and parental clusters (Fig. 4).

**Genetic and phenotypic characteristics of individuals from the two genetic groups**

In the following, we focus on results obtained with **NewHybrids** to further unravel the biological properties of the two genetic entities.
Re-including clonal copies, and applying an assignment coefficient threshold of $q < 0.9$, more than 93% of sampled individuals could be assigned, 251 belonging to the first genetic group and 174 to the second. There was a moderate, albeit significant, genetic differentiation level between the two groups ($F_{ST} = 0.094$, $P < 0.05$).

Interestingly, the distinction made according to virulence phenotype roughly matched a segregation in individual memberships: Vir7 individuals being assigned mostly to genetic group 1 and Avr7 individuals to genetic group 2 (Fig. 4). A highly significant correlation was found between individuals belonging to the first group and their virulence profile (logistic regression, $P < 0.001$). Exceptions to this rule included only three individuals (one from the cultivated stand of BPR Manse and two from the wild stand at Embrun). Reciprocally, five Avr7 individuals were assigned to group 1 (two from Prelles and three from Monetier).

Comparison of genetic characteristics of individuals from the two groups revealed that—in several instances—individuals from group 1 were less genetically diverse than those from group 2. Values of all genetic diversity indices (observed and expected heterozygosity) were significantly lower for individuals from the ‘wild’ group (Table 4). In addition, rare alleles occurred only among group 2 genotypes (data not shown). Groups 1 and 2 were thereafter referred to as ‘cultivated’ and ‘wild’ groups, respectively.

Spatial repartition of individuals belonging to ‘wild’ and ‘cultivated’ groups, and within-group genetic structure

A site-by-site examination of individual membership of the ‘wild’ and ‘cultivated’ genetic groups mirrors the spatial distribution of Vir7/Avr7 phenotypes presented above (Fig. 5). In particular, it is obvious that individuals from the two genetic groups can coexist locally in wild stands without detecting hybrids, whereas cultivated stands consist almost exclusively of individuals from the ‘cultivated’ group. In addition, ‘cultivated’ individuals are not distributed evenly among the wild sites: sites located in the northern part of the valley (Prelles, Saint-Clément, Châteauroux, Embrun) were almost free of ‘cultivated’ individuals (consisting mostly of Avr7 individuals), but the frequency of ‘cultivated’ (and thus Vir7) individuals increased abruptly when considering sites located downstream of the poplar-larch sympatry area (downstream of Rochebrune), reaching up to 44% of sampled individuals in Valernes (Fig. 5). The sole exception to this pattern was the site of Les Mées, where no ‘cultivated’ individuals were found.

Besides, the ‘cultivated’ genetic group is more likely to form a panmictic unit (all sites are in HW equilibrium and only rare pairs of loci in LD; Table 4). Conversely, the ‘wild’ genetic group displayed more evidence of substructuring—which was undetected using Bayesian clustering methods—larger global $F_{ST}$ value, significant HW and LD as illustrated in Table 4. In addition, the variance in genetic and genotypic diversity indices is much greater across locations for the ‘wild’ group (Table 4). Six of the 21 comparisons of pairwise $F_{ST}$ among the ‘cultivated’ group were significant ($P < 0.05$), whereas 21 pairs of sites of 28 comparisons were significantly different among the ‘wild’ group ($P < 0.05$). No IBD was detected among the ‘cultivated’ group ($P = 0.85$). Conversely, a significant positive correlation ($P = 0.02$) was detected between the $F_{ST}/(1 - F_{ST})$ values and geographical distances for individuals from the ‘wild’ group—a result consistent with the substructure evidenced above.

Discussion

Regular spread of a poplar rust epidemic along the Durance River valley

A major result of this study was demonstration of the spread of an annual rust epidemic along the Durance River valley. Monitoring the spatio-temporal occurrence

Table 4 Characteristics of the two genetic groups expressed as the means and ranges (in brackets) across sampling sites of several indices. Only sites with at least nine distinct genotypes were retained

<table>
<thead>
<tr>
<th></th>
<th>$N_{tot}$ (range)</th>
<th>$G/N$ (range)</th>
<th>$A_e$ (range)</th>
<th>$H_e$ (range)</th>
<th>$H_o$ (range)</th>
<th>$r_D$</th>
<th>$F_{IS}$</th>
<th>$F_{ST}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 ('Cultivated') (7 sites)</td>
<td>19.3 (9–31)</td>
<td>0.78 (0.45–1)</td>
<td>2.65* (2.46–2.80)</td>
<td>0.406b (0.388–0.436)</td>
<td>0.400b (0.366–0.421)</td>
<td>0.009**</td>
<td>0.015a</td>
<td>0.035****</td>
</tr>
<tr>
<td>Group 2 ('Wild') (8 sites)</td>
<td>28.2 (9–47)</td>
<td>0.83 (0.23–1)</td>
<td>2.95* (2.17–3.43)</td>
<td>0.473a (0.393–0.507)</td>
<td>0.455a (0.380–0.598)</td>
<td>0.014***</td>
<td>0.039***</td>
<td>0.047***</td>
</tr>
</tbody>
</table>

$N_{tot}$, number of individuals assigned to each genetic groups; $G/N$, genotypic richness; $A_e$, allelic richness (based on sample size of nine individuals); $H_e$, expected heterozygosity; $H_o$, observed heterozygosity; $r_D$, multilocus estimate of linkage disequilibrium; $F$-statistics ($F_{IS}$ and $F_{ST}$) calculated according to Weir & Cockerham (1984). Asterisks indicate the significance level of genetic disequilibrium and genetic structure (**$P$-value <0.01; ***$P$-value <0.001). Letters indicate the results of group comparison tests (different letters indicate significant differences between ‘wild’ and ‘cultivated’ sites, $P < 0.05$).
of *Melampsora larici-populina* on poplar revealed that the epidemic began in the upstream part of the valley, where larches and poplars grow in sympatry, and then spread downstream to the area of allopatry. Due to the reliefs bordering the valley, the spread of *M. larici-populina* is channelled along a one-dimensional corridor. These results again serve to highlight the importance of landscape features in the spread of plant diseases (Mundt et al. 2011).

Epidemiological data allowed the regular spread of the disease on both wild and cultivated poplar stands to be documented, revealing that the timing of first detection was earlier in cultivated stands than in neighbouring wild stands in the allopatry area. Because *M. larici-populina* individuals from cultivated stands can also infect wild poplar stands, the difference in the timing of infection is likely to result from differing ecological conditions in wild vs. cultivated stands. Indeed, cultivated stands are known to be more susceptible to disease (Prakash & Thielges 1987; Chen & Harrington 2006), which could favour detection.

The regular spread of the disease along the Durance River valley revealed by count data is in agreement with results from another host-alternating fungal pathogen: *Puccinia sorghi* (Zogg 1949). This latter study described accurately the spatio-temporal dynamics of a recurrent annual maize rust epidemic at that time in the higher Rhine River valley in Switzerland, over an 80-km-long transect. The alternate host, *Oxalis stricta*, was restricted to a small area of the studied transect, near Lake Constance. In 1945–1947, annual maize rust epidemics were shown to begin in this Maize-Oxalis sympatry area and then to spread regularly on maize along the Rhine River valley. As in the present study, knowledge of the location of the sympatry area between the two host plants enabled the starting point of the annual epidemics to be identified. However, the older study lacked a population genetics component, which in the present study allowed a better understanding of the epidemiological processes involved.

Interestingly, a major discrepancy was found in the regular spread of the disease on wild poplars, which challenges the view that there is a strictly gradual dispersal process for colonization of the Durance River valley. At Les Mées, the first rust symptoms were recorded very early, although this site is located far to the south of the sympatry area. A detailed examination of genotypic profiles revealed that these individuals belonged to the ‘asexual’ genetic group and that the most repeated genotype had also been sampled 5 years later in the southern part of the Durance River valley (Xhaard et al. 2011), suggesting a lack of sexual reproduction at that location. The most likely hypothesis for the early observation of poplar rust at Les Mées in 2004 is therefore that *M. larici-populina* individuals overwintered asexually. Several authors have already suggested asexual overwintering.
Distinct genetic groups persist despite co-occurrence during an asexual epidemic

In this study, the population genetics analysis of individuals provided additional insights into the genetic origin of the epidemic. A detailed Bayesian genetic assignment analysis found that *M. larici-populina* individuals collected after disease spread were clustered into two genetically distinct groups. Hence, this study provides evidence that cryptic population structure can be detected within the course of a single epidemic. It can be concluded that the poplar rust epidemic spreading southwards along the Durance River valley has in fact a dual inoculum source.

The distinction between the two genetic groups was highly robust and supported by the two Bayesian assignment algorithms (*STRUCTURE* and *GENELAND*) used to assess the number of genetic clusters among the samples. The assignment profiles were very similar, although *GENELAND* appeared to be more efficient than the reference software *STRUCTURE* at assigning individuals to genetic groups. Such disparity between software is common (Durand *et al.* 2009) and results mostly from differences in algorithms, as these methods are not based on exact calculus (Rieux *et al.* 2011). The most important result is that this genetic partition accounts for the observed biological differences among the samples; there was a quasi-perfect match between phenotypic profiles and genetic group assignments. Moreover, although the two groups displayed close genetic characteristics, they differed significantly in their level of genetic diversity.

This genetic partition and the specificities of both genetic groups mirror the population structure observed in a previous study at a wider geographical scale (Xhaard *et al.* 2011). Based on their phenotypic and genotypic specificities, groups 1 and 2 were referred to as the ‘cultivated’ and the ‘wild’ group, respectively. Several lines of evidence indicate that individuals from the ‘cultivated’ group derive from a drastic selection event exerted by poplar cultivars, such as ‘Beaupré’, that carry the R7 rust resistance (Gérard *et al.* 2006). In particular, the ‘cultivated’ group consisted almost entirely of *Vir7* individuals; moreover, this group was genetically less diverse than the ‘wild’ group. This latter characteristic is in accordance with the loss of heterozygosity that accompanies pathogen evolution on cultivated crops (Burdon & Thrall 2008; Gladi eux *et al.* 2008), as a result of founder effects due to the limited number of individuals initially able to infect cultivars carrying qualitative resistance genes (Guérin *et al.* 2007). The ‘wild’ group exhibited a higher genetic diversity, consistent with long-standing co-evolution with the wild host.

Comparing the present results with those of Xhaard *et al.* (2011), it is noteworthy that the same partition of genetic variability and level of genetic differentiation was found among individuals sampled after a 5-year interval. The present study thus provides additional evidence that the ‘wild’ and ‘cultivated’ groups of *M. larici-populina* persist for years, despite co-occurrence during asexual epidemics, as observed on wild stands in the southern part of the Durance River valley.

Which processes allow the persistence of genetic differentiation?

In addition to confirming individual assignments, the use of a third assignment software (*NEWHYBRIDS*) permitted the quantification of hybridization events between ‘wild’ and ‘cultivated’ individuals. Overall, very few first-generation hybrids were found, although in many instances individuals from both groups co-occurred at the same location, especially in the southernmost part of the valley. These results stress the absence of hybridization between the two genetic groups, during both asexual epidemics and sexual reproduction in the sympatry area.

The lack of hybrids appearing during asexual epidemics can be interpreted in the light of the fungal life cycle plus information gleaned from disease monitoring. The southernmost sites of the valley are located far from the poplar-larch sympatry area where sexual reproduction takes place. As shown by Bourassa *et al.* (2007), the many asexual cycles that result in disease spread on the primary host are a dead-end process in the absence of an alternate host in the vicinity. In the following spring, the sexual spores resulting from meiosis (basidiospores) produced on dead poplar leaves can infect only larches in the immediate vicinity. Further downstream on the Durance River, individuals of *M. larici-populina* are too far away from the poplar-larch sympatry area and are thus very unlikely to contribute to sexual reproduction in the upstream region of the valley. Indeed, individuals from different genetic origins can co-occur in the same place, but as this co-occurrence takes place during asexual epidemics, they cannot hybridize. This illustrates how epidemiological knowledge informs more precise interpretation of genetic data.

© 2012 Blackwell Publishing Ltd
The lack of hybrids in the poplar-larch sympatry area is more puzzling. The presence of individuals of the ‘cultivated’ group in populations at Pont-du-Fossé and Manse rules out the hypothesis that these individuals cannot survive in the sympatry area. Moreover, the geographical distance between populations within the sympatry area is much shorter than the distance travelled by the epidemic wave in one season. Nonetheless, a clear split in the spatial distribution of the two genetic entities was observed in this area, with no individual from the ‘wild’ group in the western part of the sympatry area, and a quasi-absence of individuals from the ‘cultivated’ group in the eastern part, further upstream on the Durance River. Individuals from these two foci were sampled from different host plants but, as previously discussed, this cannot account for the observed genetic differentiation and the lack of gene flow as individuals from the ‘cultivated’ group can infect wild poplars. The most likely hypothesis to account for this spatial genetic structure is that the sympatry area is a contact zone between the two genetic groups, with ‘cultivated’ genotypes invading the native ‘wild’ distribution area from the northwest (Xhaard et al. 2011). A large mountain range splits this spatial domain, which might shelter ‘wild’ genotypes from hybridization or at least slow down the invasion of the ‘cultivated’ genotypes. Unfortunately, this hypothesis cannot be tested using this data set, and a new dedicated study with extensive sampling in the sympatry area (i.e. a landscape genetic approach) will be required to further decipher local barriers to gene flow.

Taken together, these results confirm that epidemiological studies aiming to infer dispersal capacities during epidemics can benefit from molecular epidemiology data. Unlike demographic quantification only (Zogg 1949), the use of molecular data allows a refined characterization of the inoculum source in the sympatry area. The population genetics analyses performed in this study clearly show that epidemic spread along the Durance River originates from two inoculum sources located in the poplar-larch sympatry area. Individuals of both ‘cultivated’ and ‘wild’ groups then disperse asexually southwards along the Durance River, giving rise to a single but genetically heterogeneous poplar rust epidemic. In the allopatry area, individuals belonging to both genetic groups became mixed on wild Populus nigra stands. The ability to track individuals from the two sources throughout the annual epidemic opens the prospect of refining inference of dispersal by following separately the contributions from the two potential inoculum sources.

Genetic consequences of disease spread

Epidemiology and population genetics approaches also supplement the study of the genetic consequences of demographic fluctuations associated with colonization events during disease spread. Each group detected after Bayesian assignment analysis forms a quasi-panmictic unit, but nonetheless some traces of substructuring were found within both ‘wild’ and ‘cultivated’ groups (positive $F_{IS}$ and significant LD across locations). These traces of genetic disequilibria are very likely to result from demographic changes accompanying disease spread, as found at higher spatial scales in other plant pathogenic fungi (Zeller et al. 2004; Halkett et al. 2010).

In a continental study of the genetic diversity of the North American poplar rust fungus Melampsora medusae f. sp. deltoidae, Bourassa et al. (2007) showed that the genetic diversity was higher in the poplar-larch sympatry area (Great Lakes Region, Ontario and Quebec) than in an area with poplars only (most of the eastern United States). The genetic features of poplar rust individuals collected in the allopatry area suggested that they resulted from annual asexual spread of individuals originating from sexual populations in the poplar-larch sympatry area (Bourassa et al. 2007). The same result, contrasting the genetic signatures of individuals from the sympatry vs. allopatry areas, was found here. Moreover, unlike the case of Bourassa et al. (2007), M. larici-populina individuals were sampled here along the entire epidemiological transect. This study thus provides additional details of the genetic structure resulting from an epidemic spread. A significant correlation between genetic and geographical distances was found among individuals from the ‘wild’ group. This spatial genetic structure is consistent with a gradual dispersal pattern (Slatkin 1985; Leblois et al. 2004; Barrès et al. 2008) resulting from the observed colonization dynamics (Halkett et al. 2010). In addition to deciphering the genetic origins of M. larici-populina individuals collected at the end of a poplar rust epidemic, this population genetics analysis also allowed us to confirm the gradual dispersal process responsible for the spread of disease along the Durance River. Because the source of the populations and the path followed by the epidemic wave can be identified clearly, the Durance River valley constitutes a unique ecological system in which to examine further the genetic consequences of dispersal by combining population genetics and epidemiological data in more detailed spatio-temporal surveys.

Concluding remarks: strengths of combining epidemiology and population genetics

The many population genetics studies on fungal plant pathogens in the literature have tackled a wide array of evolutionary questions. However, gaps between plant pathogen epidemiology and population genetics studies still remain, and there is a paucity of work aiming to
combine these disciplines (Milgroom 2001; Archie et al. 2009). Most of the difficulties stem from differences in the spatial or temporal scales considered. Yet these two disciplines are tightly interconnected because neutral genetic structures derive mostly from the balance between migration and genetic drift, both of which rely on demographic processes. The main originality of this study is that the same ecological system was analysed from both an epidemiological and a population genetics point of view. This molecular epidemiology approach allowed us to append information from both data sets, and hence exceeded the limits of each separate discipline towards achieving a better understanding of this biological system. These data serve to illustrate the strength of combining epidemiology and population genetics data sets. The availability of highly polymorphic and reliable molecular markers, together with the advent of Bayesian clustering methods, allows cryptic population genetic structures to be revealed. As seen here, large genetic units can persist for years in spite of being sometimes locally admixed. It is thus possible to define ‘true’ population genotypes with contrasting evolutionary histories that could serve as management units in crop protection strategies. The methods employed here can define the genetic origin and retrace the colonization history and migration pathways of a recent epidemic. In addition, such detailed analysis of the very fine population genetic structure provides insights into the genetic consequences of contemporaneous demographic events. The Durance River valley therefore represents an ecological system all the more suited to the study of dispersal processes in a plant pathogen. In this respect, this specific ecological scenario will contribute to a better understanding of the evolutionary forces at work during biological invasion or the spread of emerging disease.

Acknowledgements

This research was supported by grants from ECOGER (AIP 00295-298, INTERPOICGER program), the Agence Nationale de la Recherche (ANR 07-BDIV-003, EMERFUNDIS program) and INRA (projet innovant EFPA). C.X. and B.B. were supported by PhD fellowships from the French Ministry of Education and Research (MESR) and the Région Lorraine, respectively. We would like to thank Béranget Bertin and Christine Gehin for their technical help, Cyril Dutech, Jérôme Enjalbert, Frédéric Fabre, Etienne Klein and Jean Pinon for fruitful discussions, two anonymous reviewers for their valuable comments and suggestions, and Helen Rothnie for her careful English language checking.

References


© 2012 Blackwell Publishing Ltd


de Meeüs T, McCoy KD, Prugnolle F et al. (2007) Population genetics and molecular epidemiology or how to “debusquer la bete”. Infection, Genetics and Evolution, 7, 308–332.


This study forms part of the PhD research of C.X., which addressed the gene flow between populations of _Melampsora larici-populina_ and the genetic consequences of dispersal, and of B.B., who studied the inference of dispersal processes in _M. larici-populina_. A.A. is in charge of the molecular techniques used for fungus genotyping. L.B. is an epidemiologist interested in pathogen transmission in relation to adaptation of populations to plant resistances. F.H. is a population biologist interested in the population genetics and evolutionary biology of fungal plant pathogens. P.F. has a long-standing knowledge of the biology and genetics of rust fungi.

**Data accessibility**

Microsatellite data (470 individuals genotyped): data deposited in the Dryad repository: doi: 10.5061/dryad.249kj74s.