INFERRING EVOLUTIONARY SIGNALS FROM ECOLOGICAL DATA IN A PLANT–PATHOGEN METAPOPULATION

OTSO OVASKAINEN1 AND ANNA-LIISA LAINE

Metapopulation Research Group, Department of Biological and Environmental Sciences, P.O. Box 65, Viikinkaari 1, University of Helsinki, FI-00014 Finland

Abstract. We followed the dynamics of local epidemics in three populations of a natural plant–pathogen system for four sequential years. We characterize the overwintering process with spatial statistics and use a stochastic, spatially explicit, modeling approach with Bayesian parameter estimation to study the spread of the infection during the growing season. Our modeling approach allows us to infer coevolutionary signals from spatiotemporal data on pathogen prevalence. Most importantly, we are able to assess the distribution of resistant hosts within the distribution of all host plants. We show that resistant hosts occur in areas with high pathogen encounter rates, and that the occurrence of resistance correlates with overwintering probability of the pathogen. The estimates for essentially all model parameters are characterized by a large amount of variation over the years and the populations. While the variation in the fraction of resistant hosts and in the force of infection is to a large extent explained by the population, the other model parameters (two parameters describing the shape of the dispersal kernel) vary essentially in an unpredictable manner, suggesting that much of the variation may occur at very fine spatial and temporal scales.

Key words: Bayesian data analysis; epidemiology; host–pathogen metapopulation; plant–pathogen interaction; stochastic modeling.

INTRODUCTION

Development of epidemiological models has been driven by the need to understand and predict the dynamics, invasion, and persistence of plant and animal disease (Anderson and May 1991, Gilligan 2002, Shaw 2002). The inherent variable nature of epidemics has presented a challenge for this work and has brought forward the need to use stochastic models (Renshaw 1991, Andersson and Britton 2000), which may be used to predict the severity of disease and/or to clarify the types of interventions that are likely to be useful in disease control (Shaw 2002, Gibson et al. 2004). However, although parameter estimation has been recognized as the critical link between theory and application, it has tended to lag behind model development (Gibson et al. 2004), which imposes a pressing need for further research on the methodology of parameter estimation.

One of the principal challenges in epidemiological analysis is to understand the causes of the variability that occurs between different epidemics of the same disease (Gibson et al. 1999). Most analytical work on stochasticity in epidemics has focused on demographic variability (Renshaw 1991, Andersson and Britton 2000). However, in many instances, environmental variation can be a critical influence on the development of an epidemic. A wide range of economically and environmentally important fungi and invertebrates are known to have a strong sensitivity to environmental factors. For example, many species of fungi have threshold temperatures and humidity levels for germination to occur (Truscoot and Gilligan 2003). Variability of physical conditions, such as temperature and rainfall, may induce pronounced fluctuations in population dynamics of pathogenic fungi (Burdon 1993). Another factor that has been largely ignored in models of epidemic development in plant–pathogen systems is the impact of genetic variability of host and pathogen, although there is empirical and experimental evidence showing that it may, in many cases, have a substantial effect (Schmid 1994, Thrall and Jarosz 1994, Alexander et al. 1996, Thrall et al. 2001).

While the majority of data on plant–pathogen interactions come from agricultural systems, much may be gained from studying natural pathosystems. Those characteristics that particularly set these two systems apart embrace the concepts of host population size, density and spatial distribution, genetic variability in the host population, and population continuity or predictability through time. All these factors are likely to have a major impact on the occurrence and transmission of pathogens. Remarkably little effort has been directed at investigating causal links between disease dynamics and host population genetic structure in plant–pathogen systems. This is particularly surprising given the potential for such variation to affect pathogen evolution and the emergence of new diseases (Thrall and Burdon 2003). To be able to establish such links, we must define

Manuscript received 21 March 2005; revised 14 September 2005; accepted 23 September 2005. Corresponding Editor: M. Lavine.

1E-mail: otso.ovaskainen@helsinki.fi
the spatial scale at which hosts and pathogens interact in order to assess variation in contact rates and the resulting selection pressure. This calls for data on disease in natural pathosystem where both hosts and pathogens are free to evolve, in contrast to agricultural systems where changes in host resistance result from human imposed selection. Such data is needed to identify and estimate critical parameters associated to epidemic development, to test the fit of stochastic epidemiological models, and to understand the evolution of both virulence and resistance.

We describe a case study of a natural plant–pathogen metapopulation where the dynamics of local epidemics are followed in three different populations for four subsequent years. In addition to performing spatial statistical analyses, we model the spread of the infection as an individual-based stochastic process. This allows us to assess the distribution of resistant plants within the host populations, to test the fit of the model against the data, and to assess the variability in model parameters among the populations and the years.

THE FIELD STUDY

Our study system is located in the Åland Islands in southwest Finland where populations of the host plant, Plantago lanceolata, are abundant, but fragmented in their distribution. Only a fraction (<5%) of these populations are infected by the pathogen, Podosphaera plantaginis, at a given point in time. In the Åland islands P. plantaginis persists as a metapopulation with frequent local extinction and colonization events (Laine 2004, Laine and Hanski 2006). Podosphaera plantaginis belongs to the powdery mildews (Erysiphales) which are obligate fungal pathogens that require living host tissue throughout their life cycle. Powdery mildew fungi propagate vegetatively with short-lived abundantly produced spores that are aerially dispersed (Bushnell 2002). During the growing season, generations of the pathogen follow one another in quick succession of approximately 10 days (Laine 2004), leading to local epidemics. Powdery mildews are not considered lethal pathogens but they exploit the host’s nutrient supplies for their own growth and spore production. This causes a reduction in the host’s photosynthesis, impairs growth, and reduces yields to the extent that when infection coincides with other stressful environmental conditions, it may lead to mortality of the infected host individual (Agrios 1997, Bushnell 2002, Laine 2004).

The host, P. lanceolata, is capable of clonal reproduction via the production of side rosettes as well as sexual reproduction. In contrast to the pathogen, the seeds of P. lanceolata have no specialized dispersal mechanism, but are simply dropped to the ground, close to the mother plant (Bos 1992, van Damme 1992). The pollen flow appears to be restricted mainly to short distances (Tonsor 1985, Bos et al. 1986) and proximate Plantago populations have been shown to be highly differentiated and variable in their resistance to the pathogen (Laine 2004). In the Åland Islands, the phenology of the host results in distinct crashes in pathogen numbers as plants die back to underground rootstocks at the end of the summer. The primary infection focus/foci may be initiated from overwintering mycelial growth or resting spores on the host or from a spore colonizing the population from outside.

The data were collected during the years 2000–2003 from three host populations (1, 2, 3). The host populations are located relatively close to each other; populations 1 and 2 are separated by 198 m, populations 2 and 3 by 303 m, and populations 1 and 3 by 500 m. The locations of the host populations have been mapped with a GPS in the context of metapopulation studies of the Glanville fritillary butterfly where all known populations of P. lanceolata are surveyed annually in the Åland Islands (Hanski 1999). In all three populations, the entire coverage of P. lanceolata was surveyed in a grid of 1-m² quadrats. The size of the grid for populations 1, 2, and 3 was 600, 793, and 420 m², respectively. The number of quadrats with P. lanceolata varied among the populations in the four successive years (population 1, 320, 293, 341, 287; population 2, 235, 237, 242, 190; population 3, 133, 125, 126, 108). The three populations were surveyed twice each year, first in early July when the epidemic starts to pick up and again in early September when the spread of the disease has ceased. In the surveys, the number of hosts and the number of infected individuals were recorded for each quadrant. Fig. 1 gives a representative example of the data (year 2001, population 1).

We analyze the data in two parts, first going through the infection process that takes place during the summer period, and then proceeding to the overwintering process.

THE INFECTION PROCESS

We start with a statistical characterization of the data (Table 1). The prevalence of the mildew, i.e., the fraction of infected hosts, depends on Plantago density and the year both at the beginning of the summer and at the end of the summer. In addition, mildew prevalence at the beginning of the summer depends on the population. The probability of gaining the infection during the summer depends on Plantago density, the year, the population, and the prevalence of the mildew in the focal quadrat at the beginning of the summer. As expected, Plantago density had a positive effect in all cases, and mildew prevalence at the beginning of the summer had a positive effect for the probability of gaining the infection.

Fig. 2 quantifies some spatial statistics of the data for the year 2001, population 1 (other years and populations are comparable, data not shown). As is evident based on visual inspection of the data (Fig. 1), the distributions of infected hosts are highly aggregated in space. The upper panels (A and B) describe the pattern at the beginning of the summer, when the average prevalence was 0.033. Panel A shows that prevalence was highest around
quadrats where \textit{Plantago} density was high, as the prevalence next to a randomly selected plant was 0.04, the difference being statistically significant up to the distance of 12 m. The infected plants occurred in a highly aggregated manner, the probability of finding an infected host next to an infected host being 0.12, the high correlation however extending only to the distance of 1–2 m. The middle panels (C and D) show the corresponding analysis for the pattern at the end of the summer. The results are analogous to the ones for the beginning of the summer, so that prevalence is somewhat higher around randomly selected plants (i.e., around quadrats where \textit{Plantago} density is high) and much higher around randomly selected infected plants (i.e., the pattern is highly aggregated). The panels E and F show that the probability of gaining the infection is somewhat higher (from 0.24 to 0.28) around quadrats where \textit{Plantago} density is high and still higher (from 0.24 to 0.32) around quadrats where the mildew was present at the beginning of the season.

\textit{The individual-based model}

While the statistics given above are useful in characterizing the observed patterns, they fail to give

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Estimate</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>log(density of plants)</td>
<td>0.77</td>
<td>1, 1236</td>
<td>190.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>B</td>
<td>population</td>
<td>(0, 0.13, −0.49)</td>
<td>2, 1236</td>
<td>8.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>B</td>
<td>year</td>
<td>(0, −0.61, −0.58, −3.1)</td>
<td>3, 1236</td>
<td>74.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>E</td>
<td>log(density of plants)</td>
<td>0.46</td>
<td>1, 1021</td>
<td>4.9</td>
<td>0.027</td>
</tr>
<tr>
<td>E</td>
<td>population</td>
<td>(0, 0.30, −0.14)</td>
<td>2, 1021</td>
<td>1.2</td>
<td>0.31</td>
</tr>
<tr>
<td>E</td>
<td>year</td>
<td>(0, −0.97, −1.70, −5.8)</td>
<td>3, 1021</td>
<td>47.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>P</td>
<td>log(density of plants)</td>
<td>0.18</td>
<td>1, 1756</td>
<td>30.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>P</td>
<td>logit mildew prevalence</td>
<td>0.66</td>
<td>1, 1756</td>
<td>871.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>P</td>
<td>population</td>
<td>(0, −0.49, −0.59)</td>
<td>2, 1756</td>
<td>17.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>P</td>
<td>year</td>
<td>(0, −0.61, −0.50, −3.25)</td>
<td>3, 1756</td>
<td>49.0</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Notes: Year and population were set as fixed factors, density of plants and mildew prevalence at the beginning of summer as covariates, and the quadrat as a random factor. The estimates for population and year refer to the coefficients for populations (1, 2, 3) and for the summers (2000, 2001, 2002, 2003), respectively. The error distributions were set to negative binomial (model B with a high frequency of zeros in the data), Poisson (model E), and binomial (model P). We assumed a log link function in models B and E and a logit link function in model P. A within-population spatial covariance structure assuming a Gaussian distribution was included in all models. The covariance term was significant ($P < 0.0001$) for all three models, the estimated range $\rho$ (in m) and sill $\sigma^2$ being $\rho = 0.91$, $\sigma^2 = 2.1$ (model B), $\rho = 0.70$, $\sigma^2 = 0.42$ (model E), and $\rho = 0.90$, $\sigma^2 = 0.16$ (model P).
much insight on the processes that are responsible in creating the patterns. In order to gain such insight, we model the infection process with a stochastic individual-based discrete-time model as follows. We assume either that all host plants are susceptible to the infection (model A), or that some of the host plants are resistant to the pathogen (model B). Our notion of resistance includes not just genetic resistance, but also induced resistance and other possible reasons that may prohibit the infection to take place.

The statistical analysis of the data presented above (Table 1) suggested that the infection process varies considerably among the years and the populations. In order to quantify more precisely the aspects that vary in space and time, we first model the infection process separately for each year and population, and then compare the estimates among the years and populations.

We let $P_i$ and $H_i$ denote the number of host plants and nonresistant host plants in quadrat $i$, so that in model A, $H_i = P_i$; and, in model B, $H_i \leq P_i$. We let $I_{i,t}$ denote the number infected individuals in quadrat $i$ at time $t$, where $t = 0, 1, \ldots, t_{\text{max}}$ represents discrete time steps, the length of which we set to 10 d, corresponding to the germination period of the mildew. The discrete time model should be considered as an approximation of a more realistic continuous-time model, in which the

Fig. 2. Spatial statistics (K-functions; Bailey and Gatrell 1995) for population 1, year 2001. Panels (A) and (B), which refer to the pattern at the beginning of the summer (Fig. 1B), show (A) the probability that a plant at distance $d$ from a randomly chosen plant or (B) a randomly chosen infected plant is infected. Panels (C) and (D) show the corresponding results for the pattern at the end of the summer (Fig. 1C). Panels (E) and (F) show the probability that an initially uninfected plant got the infection during the summer, given that it was at distance $d$ from (E) a randomly chosen plant or (F) a randomly chosen plant that was infected at the beginning of the summer. In all panels, the dashed lines represent the mean and one-sided 95% confidence intervals based on a random distribution.
germination period would be fixed to 10 d but in which the mildew generations could otherwise overlap.

We assume that each plant individual is nonresistant with probability $h$ (fixed to $h = 1$ in model A). Each infected host is assumed to produce spores at a constant rate $N$ per time unit, and the spores are assumed to be dispersed according to a dispersal kernel $D(g, d)$, the shape of which is described by the parameters $g$ and $d$. We approximate a diffusion-like infection process by assuming that a fraction $1 - g$ of the spores land close to their mother, the shape of the local kernel being given by the two-dimensional Gaussian distribution $(\propto \exp(-(x^2 + y^2)/d^2))$. In addition, we allow for the possibility of long-ranged jumps (induced, for example, by wind or insects) by letting a fraction $g$ of the spores to disperse “globally,” which we approximate by assuming that they land uniformly in an area of 1000 m$^2$ consisting of the host population and its surroundings. Based on our data, we assume that each infected host remains infected until the end of the growing season.

We assume that the density of host plants is low enough to avoid any crowding effects, so that the amount of spore deposition experienced by a focal plant is not reduced by shading due to neighboring plants. In other words, a spore that lands in a quadrat is assumed to infect a nonresistant host in the quadrat with a probability $\epsilon$ that is independent of the quadrat. Letting $X_{ijt}$ denote the number of spores that land in quadrat $i$ between times $t$ and $t + 1$, the probability that a given nonresistant host in quadrat $i$ will become infected is $p_{ijt} = 1 - (1 - \epsilon)^{X_{ijt}}$. Assuming that the number of spores is large and that the probability $\epsilon$ is small, $p_{ijt}$ can be approximated by $p_{ijt} = 1 - \exp(-\beta s_{ijt})$, where $s_{ijt} = \sum_j I_{ij}, D(g, d)(x_i - x_j)$ is proportional to the number of spores that are expected to land in quadrat $i$ and $\beta = N\epsilon$ is the force of infection. By the above assumptions, the number of new cases in quadrat $i$ that appear between times $t$ and $t + 1$ is distributed as $\text{bin}(H_{ij,t} - I_{ij,t}, p_{ijt})$.

### Bayesian parameter estimation

To summarize the above, the data for a given year and population consists of the vector $y = (P_0, I_{ij})$, where the number of infected plants $I_{ij}$ is included only for those time steps $t \in T^d \subset \{0, \ldots, t_{\text{max}}\}$ for which there is data. The parameters of the model are given by the vector $\theta = (h, g, d, \beta, H_0, I_{ij})$, where $I_{ij}$ represents missing data, i.e., the number of infected plants for those time steps $t \in T^d = \{0, \ldots, t_{\text{max}}\} \setminus T^m$ for which there is no data. According to the Bayesian inference (Gelman et al. 2004), the posterior distribution of the parameter vector is given by $p(\theta|y) \sim p(\theta)p(y|\theta)$, where $p(\theta)$ represents the prior distribution and $p(y|\theta)$ is the sampling distribution. In the present case, the sampling distribution is given by

$$p(y|\theta) \sim \prod_{t \in T^d} \prod_{i} P(H_{ij,t} - I_{ij,t-1}, I_{ij,t} - I_{ij,t-1}, p_{ij,t-1})$$

(1)

where $P(n, k, p)$ represents the binomial probability of having $k$ successful trials among a total of $n$ trials, given that the success rate is $p$.

The parameters of the model were estimated using a Metropolis-Hastings-Gibbs algorithm, in which each parameter was updated in turn, conditional on the values of the other parameters. In addition to the actual model parameters $g$, $d$, $\beta$, and $h$, we estimated full posterior distributions also for missing data (time steps for which the number of infected individuals were not counted) and for the number of nonresistant hosts within each quadrat (which were not known as only the total number of host plants was known). For the sake of illustration, we extracted the mean and the 95% highest posterior density interval from the full posterior distributions. In addition, we calculated the mode of the posterior distribution by conditional maximization. The prior distributions (which were chosen to be essentially noninformative for $\beta$ and $d$ and completely noninformative for the other parameters) and the method by which the samples from the conditional posterior distributions were drawn are described in more detail in the Appendix. Fig. 3 gives an example of posterior mode estimates for missing data and the density of nonresistant hosts.

### Model fit

In order to evaluate the fit of the model, we examine in Fig. 4 how the model corresponds to data in predicting the probability of gaining the infection during the summer for population 1, year 2001 (other populations and years are comparable, not shown). We evaluated the model performance with respect to two biologically relevant characteristics, which are the density of *Plantago* (left-hand panels) and connectivity to infected hosts at the beginning of the summer (right-hand panels). The upper panels correspond to model A, where all host plants are assumed to be nonresistant ($h = 1$). There is no systematic bias with respect to *Plantago* density (the lines in panel A), implicating that the assumption of no crowding effects at the range of natural host densities of this system is consistent with the data. The model however slightly overestimates the infection probability in quadrats with high connectivity (the lines in panel B). We first expected that the bias would result from the shape of the dispersal kernel (global + Gaussian), which may underestimate the infection pressure at intermediate distances. We tested a number of alternative dispersal kernels (power law, global + power law, student $t$), but they did not improve the model fit. Based on these negative results and previous laboratory inoculation trials, which have shown host populations to be variable in disease resistance (Laine 2004), we extended the model to account for the possibility that some of the host plants may be resistant to the pathogen (model B). This did not change the already good fit as a function of *Plantago* density (the lines in panel C), but somewhat reduced the bias with respect to connectivity (the lines in panel D). An unbiased fit is obtained by using in the simulation
the model estimated distribution of nonresistant host plants (the lines in panels E and F) instead of randomizing the locations of nonresistant host plants (as was done in panels C and D). The dots in panel F illustrate the way in which the estimation scheme takes the advantage of the additional degrees of freedom provided by the estimated distribution of nonresistant hosts. In quadrats with high connectivity, the model predicted infection probability is almost 1, which forces the estimated number of nonresistant hosts to be close to the number of plants that actually gained the infection. This is seen in panel F as an almost complete match between the data and simulations for quadrats with high connectivity. In quadrats with low connectivity, the model predicted infection probability is low, which leaves more uncertainty in the estimate for the number of nonresistant hosts in these quadrats. The difference between panels D and F suggests that the distribution of

Fig. 3. The fraction of plants that are (A) nonresistant and (B–F) the fraction of infected plants for the time steps $t = 1, \ldots, 5$ according to the mode of the posterior distribution: population 1, year 2001.
resistance might be nonrandom, which feature we analyze in more detail in the following section.

**Inferences from the model**

Fig. 5 illustrates the posterior distributions for all feasible combinations of populations and years. While part of the data is too sparse to provide highly concentrated posterior distributions, it is clear that all of the parameters show both statistically significant and biologically relevant variation over the years and the populations. Most strikingly, the posterior mean for the fraction of nonresistant hosts varies from 0.3 to 0.9, the differences being statistically significant both between the years within a population and between the populations within a year. The force of infection and the fraction of spores that is dispersed globally are also highly variable. The length scale parameter $d$ was estimated most consistently, most posterior means
attaining a value around $d \approx 0.3$, which would suggest that some 75% of the spores that are dispersed locally fall within the 1-m$^2$ quadrat from which they originated.

In order to assess possible differences between the three populations, we examined the posterior distributions for the years 2000 and 2002, which provided the highest quality data. For the year 2002 we obtain $P(h_3 > h_1 > h_2) \geq 0.99$ and $P(\beta_3 < \beta_1 < \beta_2) \geq 0.99$, where the subscripts refer to the populations. The year 2000 gives exactly the same ordering, though now $P(h_3 > h_1) = 0.92$. As these two factors (fraction of nonresistant hosts and force of infection) can counteract each other to produce any given average prevalence level, these results might seem as an artefact. In order to validate the result against independent data, Laine (2006) performed inoculation trials using both sympatric and allopatric mildew strains. The results are well in line with the results presented above: while the average susceptibility to sympatric strains was 53% in population 1, it was only 37% in population 2.

The difference between Fig. 4D and Fig. 4F suggested that the spatial pattern of resistance within the plant population might be nonrandom. We examine this in Fig. 6, which assesses the distribution of nonresistant hosts based on the mode of the posterior distribution. While panel A shows that the fraction of nonresistant hosts does not depend on Plantago density, panel B indicates that the fraction of nonresistant hosts decreases with connectivity to infected plants at the beginning of the summer, the result being statistically significant ($P = 0.005$; regression weighted by the number of plants), but the effect size being small. Interestingly, independent inoculation trials (Laine 2006) repeated the same pattern, so that plants from sites with high infection pressure turned out to be least susceptible to sympatric mildew strains. Panel C shows that the nonresistant hosts appear to be aggregated, but only at a very small spatial scale (<1 m), and with a very small effect size (from 0.44 to 0.455). While we otherwise postpone the study of the overwintering process to the next section, we show in panel D that the distribution of nonresistant hosts can correlate with overwintering probability in an interesting way: in areas with high overwintering probability the fraction of nonresistant hosts was smaller than on average in population 1 and year 2001, the result being both statistically and biologically significant (from 0.44 to 0.39). This result does however not generalize to other populations and years, but the pattern varies. For example, in the next year (2002), overwintering probability was positively correlated with the fraction of nonresistant hosts.

**Winter Survival**

Table 2 presents the results of a generalized mixed model which examines how winter survival of the mildew depends on the population and the year, on the density of Plantago, and on the prevalence of the mildew. The overwintering probabilities varied significantly between the years and the populations. Both the density of plants and the fraction of infected plants had
a negative effect on overwintering probability, which could be explained by the fact that in sites where the density of infected plants is high, the plants that are infected are likely to be especially severely infected, measured, e.g., by the fraction of infected leaves. In such a case, Plantago mortality during the winter can be especially high, which reduces also the overwintering probability of the mildew.

Fig. 7 examines the spatiotemporal aspects of winter survival for the year 2001–2002 in population 1 (other years and populations give comparable results, not shown here). While panel A shows that winter survival did not depend much on the density of mildew around the quadrat, panel B shows that it was spatially correlated: survival was five times more likely if it is known that a randomly selected mildew within the same quadrat survived. Thus, in simple terms, there are areas in which surviving over the winter succeeds and areas where it fails. Furthermore, while panel C shows that winter survival was not much higher close to quadrats where there was mildew at the end of season a year ago, panel D indicates that it was two times higher close to quadrats where winter survival succeeded a year ago, suggesting that the favorable and unfavorable overwintering sites remain close to their locations from the previous winter. Note that the spatial scale is small, extending just up to some 2 m.

**DISCUSSION**

We assessed the spread of the infection within the growing season with an individual-based model supplemented with Bayesian parameter estimation. This allowed us to account not only for the uncertainty in the parameter estimates, but also for some complica-
tions present in the data, including missing data and an unknown distribution of resistant hosts. Despite the simplicity of the model and variation in parameter estimates among the years and sites, the model was able to capture the main characteristics of the spread of infection in the field. Here, we identify some causes of the observed variability and discuss how local disease epidemics translate into selection driving host resistance in natural plant populations.

The parameter $\beta$, the force of infection, varied over the years and especially over the populations. Although there may be “good” and “bad” years and sites for the pathogen, there is likely to be so much variation in local soil and microclimatic conditions, to which the pathogen is sensitive to, that the overall weather conditions may affect the pathogen differently in different sites even within the same population or year. For example, low rainfall and high temperatures should on principle favor both germination and spore production of powdery mildews (Bushnell 2002), but on the most shallow soils, the hosts may suffer from drought, in which case the pathogen does also poorly. The observed variation may occur at such a fine scale that attempts to quantify it would be impossible. Despite this unpredictable variation, the overall spread of the infection followed a similar pattern over the different years and host populations. Even following the exceptionally low prevalence of the mildew at the beginning of 2003, the parameter values that were estimated for the slow and only partial recovery that happened during the summer 2003 are not strikingly different from the other years.

The pathogen population crashes every winter as most hosts die back to rootstock. In contrast to the diffusion-like summer infection process, winter survival of the pathogen was much more unpredictable. The crash of the pathogen population was especially severe in the winter 2002–2003, which is linked to an overall decline in the density of Plantago caused by the lowest rainfall recorded for the past 150 years. As the entire life cycle of the mildew is obligate, it is dependent on living host tissue. In those areas of the host population where some host plants remain above ground, the mildew is more likely to survive as mycelia in dormant buds. The areas where the pathogen was most likely to succeed in surviving the winter were aggregated over space and time. The occurrence of such sites is most likely due to surrounding environmental conditions, such as higher vegetation that provided more shelter for the plants, favoring winter survival of the pathogen in these areas.

The infection process was well explained by dividing the dispersal of spores into local and global components. As pointed out by Shaw (1995), wind transport can occur over all scales, and thus for organisms moved by wind no single dispersal scale that declines exponentially.

Fig. 7. Spatial analysis of overwintering success from year 2001 to year 2002 in population 1. All panels show the probability that the mildew succeeded in overwintering, i.e., that a plant that was infected at the end of 2001 was found to be infected at the beginning of 2002. The x-axis measures the distance from a randomly chosen infected mildew that was present (A) at the end of 2001, (B) at the beginning of 2002, (C) at the end of 2000, and (D) at the beginning of 2001. In all panels, the dashed lines represent the mean and one-sided 95% confidence intervals based on a random distribution.
with distance can be characterized. While the high degree of aggregation of the mildew may seem somewhat surprising for a wind-dispersed pathogen, it is explained by a combination of several factors. Most importantly, the majority of the spores were allocated into local spore dispersal, around one half of the spores being estimated to land within the same 1-m² quadrat where the infected individual was situated.

As a result of the aggregation of the favorable overwintering sites, the initial foci from which the epidemic begins to establish do not change much over time. As pathogen transmission occurs mainly over short distances, it follows that infection prevalence within the host populations is aggregated over space and time across the entire growing season. This pattern of aggregation has evolutionary implications, as it results in different encounter rates between the host and the pathogen in different areas of the host population. Selection intensity for resistance will be much higher in areas where encounter rates are high. It has long been recognized that selection intensity and even direction may vary considerably among populations (Wright 1943, Thompson 1999). The theory of the geographical mosaic of coevolution identifies coevolutionary hotspots, where reciprocal coevolution actually takes place, intermixed with sites with less activity, the so-called coldspots (Thompson 1994, 1999).

Our first model that considered all hosts to be nonresistant, overestimated the infection probability in areas of the host population where the force of infection was high. We then included the parameter \( h \) in our model to account for genetic variability in host resistance, which improved the fit of the model. The model estimated distribution of nonresistant hosts correlated with two ecological characteristics of the epidemic, as the model predicted that resistant hosts occurred especially in areas with high pathogen encounter rates, and that the occurrence of resistance correlated with overwintering probability of the pathogen. These results trace back to the idea of coevolutionary hotspots and coldspots, so that in areas where encounter rates are high between host and pathogen there is more selection for hosts to evolve resistance. A laboratory inoculation experiment (Laine 2006) with the material originating from populations 1 and 2 confirmed that plants from areas where encounter rates with the pathogen had been high in the past were more resistant than plants from areas where the pathogen had rarely occurred during the survey period. This pattern of higher resistance was consistent in both populations: hosts from hotspot areas were more resistant to the tested pathogen strains than hosts from the coldspot areas. This result strongly suggests that high encounter rates with the pathogen have produced asymmetric selection for resistance in the host within these populations (Laine 2006).

The fraction of hosts differed strikingly among the populations and to some extent among the years. Variation among the years is to be expected, as the composition of both the host and the pathogen population is likely to change from one year to the next. Because host resistance is race specific (Laine 2004) the same perennial host may express resistance against strains present in a particular year that are absent or less common the next. Particular host genotypes may become locally abundant following clonal proliferation, yet the mortality of young ramets may be high (Mook et al. 1992) so that the relative composition of host genotypes may vary even on relatively short time scales of a few years. Furthermore, while a host genotype may be resistant or susceptible to a particular strain of the pathogen, it is important to note that in reality also a myriad of other forces will affect its infection probability. Resistance may be induced if the spore load exceeds a certain threshold resulting in leaf necrosis that prevents germination of spores (Nicot et al. 2002). Infection probability may be further complicated by the age, size, nutritional status, and morphology of the host plant. While crowding effects were not visible in the data, this does not rule out the occasional possibility that a nonresistant individual was shaded, thus reducing the likelihood of infection.

A novel feature in our modeling approach was that it allowed us to infer coevolutionary signals from purely ecological data. Most importantly, we were able to assess the distribution of susceptible hosts within the distribution of plants, the results being supported by independent direct experiments. This is nontrivial, as data about the overall prevalence level does not allow for the separation of the parameters \( \beta \) (force of infection) and \( h \) (fraction of hosts), and thus the signal about the distribution of hosts is inferred solely from the spatiotemporal pattern of disease spread. Furthermore, the signal is expected to be partly filtered out by other factors such as variation in dispersal parameters. While direct measurements remain the most reliable method when assessing the patterns of host susceptibility, our study makes the point that small scale variation in resistance may be so pronounced that its effects are visible even in such integrated measures as the spread pattern of a disease.

**Acknowledgments**

We thank Ilkka Hanski, Atte Moilanen, Saskya van Nouhuys, Tord Snäll, and two anonymous reviewers for valuable comments. This study was supported by the Academy of Finland (grant number 205376 and the Finnish Centre of Excellence Programme 2000–2005, grant number 44887) and LUOVA-graduate school (Ministry of Education).

**Literature Cited**


APPENDIX

Technical details on parameter estimation (Ecological Archives E087-051-A1).