Host ecotype generates evolutionary and epidemiological divergence across a pathogen metapopulation

Anna-Liisa Laine1, Jeremy J. Burdon2, Adnane Nemri2 and Peter H. Thrall2

1Metapopulation Research Group, Department of Biosciences, University of Helsinki, PO Box 65, Helsinki 00014, Finland
2CSIRO Plant Industry, GPO Box 1600, Canberra, Australian Capital Territory 2601, Australia

The extent and speed at which pathogens adapt to host resistance varies considerably. This presents a challenge for predicting when—and where—pathogen evolution may occur. While gene flow and spatially heterogeneous environments are recognized to be critical for the evolutionary potential of pathogen populations, we lack an understanding of how the two jointly shape coevolutionary trajectories between hosts and pathogens. The rust pathogen Melampsora lini infects two ecotypes of its host plant Linum marginale that occur in close proximity yet in distinct populations and habitats. In this study, we found that within-population epidemics were different between the two habitats. We then tested for pathogen local adaptation at host population and ecotype level in a reciprocal inoculation study. Even after controlling for the effect of spatial structure on infection outcome, we found strong evidence of pathogen adaptation at the host ecotype level. Moreover, sequence analysis of two pathogen infectivity loci revealed strong genetic differentiation by host ecotype but not by distance. Hence, environmental variation can be a key determinant of pathogen population genetic structure and coevolutionary dynamics and can generate strong asymmetry in infection risks through space.

1. Introduction

Pathogens pose threats to global food security and on the health of human, plant and animal populations, and can exert strong selection on their hosts [1–3]. The evolutionary potential of pathogen populations underlies risks of disease, and hence understanding evolutionary trajectories of pathogens is critical for any disease prevention and control strategy [4–6]. However, the strength and outcome of coevolutionary interactions is highly variable across space and time, ranging from hotspots with rapid reciprocal coevolution to coldspots where the two species do not coevolve [7–9]. This spatial and temporal heterogeneity has presented a challenge for generating predictions about when—and under what circumstances—we are most likely to observe evolutionary responses in pathogens.

There is substantial theory demonstrating that the spatial scale of host–pathogen interactions and resulting rates of gene flow are critical for how coevolutionary dynamics are played out between hosts and their pathogens [7,10–12]. A meta-analysis found that the relative rate of gene flow in hosts versus parasites was the strongest predictor of local adaptation, with significant parasite local adaptation detected only when parasites had greater gene flow rates than their hosts [13]. In the interaction between the fungal pathogen Podosphaera plantaginis and its host plant, Plantago lanceolata, Laine [14] showed that the pathogen adapts to local clusters of host populations rather than individual host demes. As the pathogen frequently disperses up to a kilometre, it is likely that patterns of adaptation would be swamped at smaller spatial scales. Further support for the importance of spatial scale of pathogen dispersal on local adaptation comes from a recent cross-species comparison that showed hosts to be least resistant to their local parasites, and most resistant to parasites collected several tens to hundreds kilometres away [15]. However, the lack of gene flow among widely separated populations may reduce their evolutionary potential and hence,
prevent local adaptation at larger spatial scales [16,17]. Together, these studies suggest that local adaptation is most likely to occur at ‘intermediate’ spatial scales—where the definition of ‘intermediate’ will depend on the balance between gene flow, relative dispersal ability of host and parasite and the strength of natural selection [18].

In addition to heterogeneity in the spatial scales separating local host and pathogen demes, spatially structured systems are likely to experience a range of different environmental conditions. The importance of environmental heterogeneity for host–parasite interactions was recognized over 50 years ago by the ‘disease triangle’, which identifies host genotype, pathogen genotype and the environment as the primary determinants of infection outcome [19]. In support of this, empirical studies across a range of different biological systems have demonstrated that many key components of host–parasite interactions are environmentally mediated [20], including host resistance, costs of resistance, parasite infectivity, parasite latency, transmission and virulence [21–26].

Despite recent advances in our understanding of how gene flow and spatially variable selection gradients drive patterns of local adaptation, theory has largely outpaced empirical studies in this field of research [7,27–30]. Hence, we lack data on local adaptation across multiple populations and environments at the metapopulation scale, as pointed out in a recent review by Blanquart et al. [31]. Here, we use the well-characterized interaction between the native Australian flax, Linum marginale, and its fungal rust pathogen Melampsora lini to examine the potential for environmental factors to mediate coevolutionary dynamics between these species. The system is ideal for this purpose as the interaction takes place on two distinct host ecotypes that occur in close proximity in subalpine landscapes in Kosciuszko National Park, Australia (figure 1). The host, L. marginale, grows on both dry, well-drained hillsides (hill ecotype) and flatter bog areas found near watercourses (bog ecotype). Previous studies have shown that host plants in these two habitats have diverged into two ecotypes that differ morphologically and genetically, and that pathotypes occurring on the two ecotypes also vary in their infection profiles, despite being in many cases separated only by few hundreds of metres [32–35].

Of particular interest is that despite the clear potential for among-population gene flow (the rust fungus is aerially dispersed), both average host resistance and pathogen infectivity are consistently higher in hill ecotype populations than in bog ecotype populations [35]. Previous studies have not examined whether this differentiation is strong enough to generate patterns of local adaptation at the ecotype level in the M. lini pathogen. To tease apart the relative importance of gene flow and environmental habitat quality on coevolutionary dynamics, we carried out a large cross-inoculation study across four bog and four hill populations to determine whether resistance and infectivity vary according to host ecotype, and whether the pathogen is adapted at the host ecotype level. We find that even after controlling for the distances separating individual host and pathogen populations,
there is strong evidence of pathogen adaptation at the host ecotype level rather than at the individual host population level. The differences may be strongly driven by host resistance, which has evolved to be higher in the hill host ecotype. Moreover, sequence analysis of two avirulence loci in the pathogen revealed strong genetic differentiation by host ecotype but not by distance.

2. Material and methods
(a) The Linum–Melampsora pathosystem

*Linum marginale* is a perennial herb endemic to southern, temperate areas in Australia. In the mountainous Kosciuszko National Park where the study populations are situated, *L. marginale* has an herbaceous perennial life form dying back each winter to a perennial rootstock, with a few surviving short shoots. Populations in the Kosciuszko region are selving [36,37]. The coming of spring induces new growth, with plants flowering in mid-to-late summer. 

*Melampsora lini* is a rust fungal species that infects *L. marginale* in the Kosciuszko National Park (New South Wales, Australia). Unlike many rust fungi, *M. lini* completes its complex life cycle on a single host species. Field surveys of infection are highly feasible as the rust fungus can attack all aerial parts of the plant, and infection is visible as bright orange lesions called uredia. These lesions release the clonal rust urediospores that can be aerially dispersed over long distances during epidemics. At the end of the growing season, the pathogen population goes through a major decline. While survival in the arid inland areas of Australia is ensured by the initiation of the sexual cycle and the formation of dormant teleiospores, there is no evidence of the pathogen undergoing the sexual phase in the Kosciuszko region [38]. There reproduction appears to be achieved solely through the production of asexual urediospores.

The interaction between *L. marginale* and *M. lini* is governed by gene-for-gene specificity [39] in which plant resistance (R) gene products (immune receptors) recognize a subset of pathogen effector proteins, known as avirulence (Avr) gene products, and trigger defence [40]. Long-term coevolutionary studies across multiple populations of this wild host–pathogen interaction have documented rapid genetic change at corresponding plant resistance and pathogen avirulence loci consistent with expectations of negative frequency-dependent selection [41].

(b) Selection of study populations, field surveys of infection and sampling

For this study, we chose eight *L. marginale* populations in the Kosciuszko region. Four of these populations (CBL, CEM, G3 and SH2) occurred on slopes and are hereafter referred to as hill populations. Four populations (CBL bog, G2, P1 and PS) were on flatter bog habitats and are called bog populations. As the plants from the two habitat types can be distinguished by their morphology [32–35], we refer to them as plant ecotypes. While the pathogens from the two habitat types cannot be distinguished morphologically (but are genetically differentiated: see Results), we henceforth follow the same terminology and refer to pathogens from the hill and bog habitats as hill and bog ecotypes, respectively. The study populations are situated in three areas of the Kosciuszko region that are separated by some tens of kilometres (see map in figure 1; coordinates of the populations are given in the electronic supplementary material, table S1). Each population was visited a total of six times, first on 14 December 2009, then at approximately two week intervals starting on 11 January (subsequent survey dates were 21 January, 3 February and 18 February) until 3 March 2010 when epidemic development had ceased and plants were dying back. The progress of rust epidemics in these populations was monitored at each visit by checking 100 *L. marginale* plants haphazardly chosen from across the population and scoring their infection status (presence/absence). Seeds were haphazardly collected during the last two of these visits from both healthy and infected individuals from at least 20 plants in each population. At the peak of the epidemics during the last two surveys, altogether 10 infected plants were sampled for rust by lightly rubbing cotton buds across the surface of sporulating uredia in each population. In the laboratory, samples were separately inoculated onto 10–15 cm tall seedlings of the universally susceptible *L. usitatissimum* ‘Hoshangabad’ by gently wiping the cotton buds with urediospores over the surface of a number of leaves. Inoculated plants were left in a saturated atmosphere overnight before being transferred to a naturally lit glasshouse. One week later, single infections were isolated and put through repeated cycles of increase to produce sufficient inoculum for the local adaptation study [35]. Seeds were planted and grown in the glasshouse under natural light conditions.

(c) Characterizing bog and hill plant community composition

To determine whether there were systematic differences in vegetation among bog and hill habitats, the plant community composition was surveyed on 3 February 2010 (mid-summer) at each study site. At each site, vegetation was quantified within seven 1 m radius quadrats centred on a haphazardly chosen *L. marginale* plant. All plants (including *L. marginale*) within the quadrat boundary were identified (to species level if possible) and their per cent vegetative cover estimated.

(d) Quantifying pathogen local adaptation

To assess the level of local adaptation of pathogen populations to their local host populations and ecotypes, we performed a fully reciprocal inoculation study. Inoculations yield a direct measure of pathogen infectivity (ability to infect a given host genotype or not) and host resistance (ability of a host line to resist a given pathogen isolate). For the study, we chose 10 host lines and 10 pathogen isolates from each of the eight populations. Out of the 640 possible host genotype × pathogen genotype combinations, the experiment consisted of 6262 inoculations owing to low germination of seed in seven host lines (lines 8 and 9 from CBL, line 6 from P1, lines 5 and 8 from PS, line 8 from SH2, and line 3 from G3). For the inoculations, shoots cut from young, vigorously growing plants of randomly selected host lines plus the universally susceptible *L. usitatissimum* ‘Hoshangabad’ as a control were exposed to a single pathogen isolate. Stems of each plant were placed upright in holes pierced through the lid of a 12 cm disposable plastic tub filled with tap water. The tubs were sealed at the bottom of 12 cm diameter, 35 cm tall plastic towers into which 10 mg of urediospores of *M. lini* dispersed in 100 mg of t alc was injected with compressed air. The spore–talc mix was allowed to settle for 2–5 min before each set (14 host lines plus the control) was lightly sprayed with water. The following day, tubs were transferred to a naturally lit glasshouse, where infections were scored 12–14 days later. Any tests giving ambiguous results were repeated [32,42]. Five infection categories could be distinguished, ranging from those showing many large, freely sporulating uredia (+) through various forms of restricted growth (±, −, −−) to those in which there were no macroscopic signs of infection (=) (for details, see [43,44]).

(e) Genetic characterization of pathogen populations

For each of the eight pathogen populations, the 10 isolates used in the local adaptation study were genotyped at the AvrP4 and
AvrP123 loci as per Barrett et al. [45]. Genomic DNA was extracted from 100 mg of spores per isolate using a DNeasy 96 plant kit (QIAGEN). For AvrP4, a 568 bp PCR product was amplified using 5'-CATCATAATCTACCCCGTAC (forward) and 5'-GCAC TATCTAACTGACCA (reverse) primer pair. For AvrP123, a 598 bp PCR product was amplified using 5'-ATGGTAACCT TTTGAAGGC (forward) and 5'-GCCATGTATGTTGACAC (reverse). Methods for PCR and sequencing were as previously described in Barrett et al. [45]. Sequence alignments were performed using predicted protein sequences.

(f) Statistical analyses

To estimate the botanical diversity of the bogs and hills, we calculated Shannon’s diversity index [46] for each study site, and the Bray–Curtis similarity index [47] was calculated to compare vegetation similarity in the bog and hill sites.

We analysed the development of epidemics as a repeated measures generalized linear model (GLM) in JMP (v. 8.0.2) with the proportion of infected host plants as the response variable with habitat type (bog versus hill) and population (nested under ‘habitat type’) as fixed explanatory variables, and observation date as a covariate. This model allowed us to account for the non-independence between our temporal observations. We included two-way interactions to investigate changes in the epidemiological curve among ecotypes and populations. We used GENALEX v. 6.5 to perform analyses of molecular variance (AMOVA) on sequence variation of AvrP4 and AvrP123 loci in the pathogen and to estimate Nei’s unbiased genetic distance between pathogen populations.

To study patterns of local adaptation between M. lini and its L. marginale host populations, the infection data were analysed as the ability of a given pathogen isolate to infect (1 = infection category +) or not (0 = infection categories ±, −) a given host line. These data were fitted using a hierarchical GLM model (GLMM) assuming a binomial error distribution and a logit link function as implemented in SAS v. 9.1 [48]. Specifically, we were interested in the contributions of population of plant and pathogen origin, the pairwise distances between host and pathogen populations, and host and pathogen habitat type of origin (bog versus hill) to infection outcomes. Hence, as fixed explanatory variables, we had host population and population and pathogen population (nested under their respective habitat of origin), and host and pathogen habitats types of origin. Host line and pathogen isolate (hierarchically nested under population and habitat type) were defined as random effects in the model. To account for the effect of spatial structure on the strength of local adaptation, we fitted the model with pairwise distances between host and pathogen populations (m; log-transformed) as a covariate. We started out with a full model, and in a stepwise manner checked for non-significant interactions which would be excluded from the final model.

3. Results

(a) Vegetation survey

A total of 37 plant species were found across the bog and hill sites. The bogs and hills differed in terms of their dominant plant species. Although grasses dominated both habitats, Poa costiniana and Poa hiemata (two relatively long tussock grasses) tended to dominate bog sites while the hills were predominantly covered by Poa fawcettiae and Rytidosperma nivicola, the latter being a relatively short grass. Using species diversity derived from Shannon’s index as a measure of biodiversity, on average hill sites supported nearly double the botanical diversity of bog sites (electronic supplementary material, figure S1a). The Bray–Curtis index revealed that, on average, bogs tended to be about 45% similar to one another, hills approximately 37% similar to one another, while a comparison of bogs with hills gave rise to only 5% similarity (electronic supplementary material, figure S1b). It is therefore readily apparent from our data that the hills and bogs differed significantly in terms of plant community structure.

(b) Epidemiology and pathogen population structure

Although there was significant variation among populations within the bog and hill habitats in the proportion of infected plants (figure 2; $F_{7,47} = 12.17$, $p < 0.0001$), disease prevalence was significantly higher in the bog populations than in the hill populations (figure 2; $F_{1,47} = 5.28$, $p = 0.0001$). There were also significant differences in the rate at which epidemics developed among populations within the two habitat categories (figure 2; time $\times$ population (habitat type) $F_{6,47} = 4.58$, $p = 0.0017$). As expected, the proportion of infected plants increased through time (figure 2; $F_{1,47} = 269.12$, $p < 0.0001$).

Genetic analysis of the two avirulence loci AvrP4 and AvrP123 in the eight pathogen populations revealed strong differentiation by host ecotype but not by distance. The hill populations were primarily composed of two alternative haplotypes AvrP123-Lm1/AvrP4-Lm1 and AvrP123-Lm2/AvrP4-Lm8 (figure 3), consistent with observations from other hill populations in the same region of the Kosciuszko National Park.
Table 1. Results of a hierarchic GLMM analysing infectivity of M. lini on L. marginale host plants according to host and pathogen origin.

<table>
<thead>
<tr>
<th>source</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance</td>
<td>26.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>host population (habitat type)</td>
<td>9.69</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pathogen population (habitat type)</td>
<td>5.35</td>
<td>0.0001</td>
</tr>
<tr>
<td>host habitat type</td>
<td>20.62</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>pathogen habitat type</td>
<td>3</td>
<td>0.0873</td>
</tr>
<tr>
<td>host habitat × pathogen habitat</td>
<td>251.61</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

[41]. Interestingly, out of 40 hill isolates tested, seven out of nine isolates that do not belong to the two haplotypes described above come from the CBL hill population. By contrast, bog pathogen populations were almost exclusively composed of the AvrP123-Lm1/AvrP4-Lm7 haplotype (figure 3).

In the AMOVA analyses of pathogen populations, the total differentiation among populations was 0.448 (PhiPT; p = 0.001) of which 0.346 was structured by host ecotype (PhiRT; p = 0.001). Significant pathogen differentiation was observed at the habitat level among boggs and hills (PhiPR = 0.157; p = 0.002). Conversely, grouping bog and hill population pairs to account for geography did not significantly contribute to explaining molecular variance (0%; p > 0.05), indicating that the physical proximity of a hill and bog ecotype had no significant effect on the structure of the pathogen populations they carried. Overall, based on the limited number of genetic markers assayed here, the bog pathogen populations appeared more closely related than their hill counterparts (electronic supplementary material, figure S2). Together, these results demonstrate that the pathogen populations from the bog and hill habitats are genetically differentiated.

(c) Local adaptation

Within habitats, there were significant differences in infection outcomes among both host and pathogen populations as well as their interaction (table 1). Notably, even after controlling for the significant effect of distance separating host and pathogen populations (table 1), habitat was a significant determinant of infection outcome. Host habitat type (i.e. bog or hill) had a significant direct effect on infection outcome (p < 0.0001; table 1) while the effect of pathogen habitat was marginally significant (p = 0.0873; table 1). The interaction between host and pathogen habitats was statistically significant indicating that bog pathogen populations interact differently with bog hosts than with hill hosts, and vice versa (figure 4 and table 1; electronic supplementary material, figures S3–S4). This interaction is consistent with the expectation of local adaptation at the habitat level as is evident in figure 4. Both bog and hill pathogens infected a higher proportion of their sympatric host ecotypes (i.e. bog pathogens on bog hosts and hill pathogen on hill hosts) than allopatric host ecotypes (table 1 and figure 4; electronic supplementary material, figures S3–S4).

4. Discussion

While the impact of the environment on disease dynamics is widely recognized [20], our study is, to our knowledge, the first demonstration of how the effects of habitat differentiation can be so pronounced that they drive coevolutionary dynamics and genetic structure of pathogen populations with direct implications for epidemiological dynamics. This finding is in line with the predictions of the theory of the geographical mosaic of coevolution [9].

Theory predicts spatial structure to be a key force driving pathogen local adaptation via effects on realized rates of gene flow [10,12,28,49]. We did not find strong evidence of pathogen local adaptation at the host population level but the pairwise distances separating the host and pathogen populations explained some of the variation we found in patterns of pathogen population infectivity. However, infectivity on sympatric versus allopatric host ecotype was the most important factor explaining patterns of local adaptation. There was a tendency for pathogens originating from the bog populations to infect a consistently lower proportion of hosts from the hill populations than from bog populations, although there was some variation in this trend. On the other hand, pathogens from the hill populations were in some cases equally or even better at infecting hosts from bog populations than from hill populations. These results suggest that patterns of adaptation are, at least to some extent, driven by higher resistance in the hill populations than in the bog populations. Higher resistance in the hill populations have also been documented by earlier studies suggesting that the differences in resistance and infectivity have been maintained over an extended period of time [33,35]. While such coevolutionary hotspots have been identified for several other types of interspecific interactions, this is among the first examples of the environment generating coevolutionary disease hotspots across the landscape [50,51].

Despite the fact that there are no obvious barriers to dispersal or gene flow among bog and hill populations, our genetic analysis of pathogen populations from the bogs and hills revealed strong differentiation by host ecotype but not by distance. The bog pathogen populations were almost exclusively composed of the AvrP123-Lm1/AvrP4-Lm7 haplotype, while two other haplotypes dominated the hill pathogen populations. The observed differences in host resistance may be enough to maintain differentiation among the bog and hill pathogen populations. In M. lini, the ability to infect a large number of host lines comes at a cost of lower spore production, and hence, pathotypes with unnecessary virulence genes may
fail to establish in susceptible host populations where they are inferior competitors against less infective isolates [52]. Our finding of more infectious isolates in the hill populations with higher resistance is also consistent with this hypothesis. Importantly, the differences we measure between the bog and hill habitats in pathogen population genetic structure and patterns of adaptation are also reflected in epidemiological dynamics. While there was considerable variation among the eight M. lini populations in their disease progression trajectories, on average infection prevalence was higher in the bog habitats than in the hill habitats. While the two habitat types did not differ in their ambient temperature during the growing season (A.-L. Laine 2010, unpublished data), it is possible that some unmeasured abiotic variation has an impact on disease dynamics in these habitats. Alternatively, the trade-off between infectivity and spore production in M. lini [52] may explain why populations dominated by particularly infectious isolates had lower disease prevalence than populations dominated by less infective isolates [41].

A future challenge will be to identify the key variables that distinguish the bog and hill habitats with regard to generating the patterns of ecotype local adaptation we report here. The clear differences we observed in the plant communities between the bog and hill environments suggest that these habitats may differ in both above- and below-ground conditions, and in both biotic and abiotic factors. Given that water drains from the hill sites, soil moisture and its direct effects on plant water availability and potential impacts on soil microbiota [53] associated with the host may be important for generating variation between these habitats. Below-ground communities may have powerful effects on how species interactions are played out above-ground [54–56]. A current ongoing study sequencing the soil microbiomes of the hill and bog habitats will help understand the role of a below-ground species community in generating spatial variation in the L. marginale—M. lini interaction.

Understanding when and where evolution takes place underlies our ability to manage risks imposed by diseases on the health of humans and food security. Our study provides direct evidence of environmental variation having such a profound effect on host–pathogen coevolution—and disease dynamics—that patterns of local adaptation are seen at the host ecotype level even after accounting for spatial structure. Spatial heterogeneity in the environment of the L. marginale—M. lini interaction has resulted in spatially divergent coevolutionary selection, driving higher resistance and infectivity in the hill sites than in the bog sites. Establishing direct links between key components of environmental variation, its impact on coevolutionary and realized epidemiological dynamics offers an exciting avenue for future research and is needed to truly understand how risks of disease evolve.

Acknowledgements. We thank Michael Goddard for his assistance in the field and Caritta Eliasson for technical assistance with the glasshouse inoculation studies.

Funding statement. This research was supported by the National Institutes of Health (NIH grant no. 5RO1 GM074265-01A2).

References

22. Laine A-L. 2007 Pathogen fitness components and genotypes differ in their sensitivity to nutrient and


