LETTER

doi:10.1038/nature14372

Phylogenetic structure and host abundance drive disease pressure in communities

Ingrid M. Parker[1,2]*, Megan Saunders[3], Megan Bontrager[1], Andrew P. Weitz[1], Rebecca Hendricks[3], Roger Magarey[4], Karl Suiter[4] & Gregory S. Gilbert[2,3]*

Pathogens play an important part in shaping the structure and dynamics of natural communities, because species are not affected by them equally[1,2]. A shared goal of ecology and epidemiology is to predict when a species is most vulnerable to disease. A leading hypothesis asserts that the impact of disease should increase with host abundance, producing a 'rare-species advantage'[3–5]. However, the impact of a pathogen may be decoupled from host abundance, because most pathogens infect more than one species, leading to pathogen spillover onto closely related species[6,7]. Here we show that the phylogenetic and ecological structure of the surrounding community can be important predictors of disease pressure. We found that the amount of tissue lost to disease increased with the relative abundance of a species across a grassland plant community, and that this rare-species advantage had an additional phylogenetic component: disease pressure was stronger on species with many close relatives. We used a global model of pathogen sharing as a function of relatedness between hosts, which provided a robust predictor of relative disease pressure at the local scale. In our grassland, the total amount of disease was most accurately explained not by the abundance of the focal host alone, but by the abundance of all species in the community weighted by their phylogenetic distance to the host. Furthermore, the model strongly predicted observed disease pressure for 44 novel host species we introduced experimentally to our study site, providing evidence for a mechanism to explain why phylogenetically rare species are more likely to become invasive when introduced[8,9]. Our results demonstrate how the phylogenetic and ecological structure of communities can have a key role in disease dynamics, with implications for the maintenance of biodiversity, biotic resistance against introduced weeds, and the success of managed plants in agriculture and forestry.

Plant pathogens can be important drivers of community diversity, structure and dynamics[1,2,10,11]. A basic premise of epidemiology is that pathogen transmission often increases with host density[12,13]. Density-dependent disease provides a mechanism for the maintenance of plant diversity in natural communities, in which locally uncommon species enjoy a rare-species advantage—based on lower enemy pressure—that mitigates the competitive impacts of dominant species[3–5]. Reports of density-dependent disease dynamics generally infer the potential effects on communities from studies of one or a few species[2], while community-level studies[1] are scarce but essential to evaluate whether such a rare-species advantage predicts patterns of disease across a community.

An ongoing debate concerns how community context influences disease, and particularly whether biodiversity suppresses infection and emerging diseases[14,15]. If increasing the number of species in a community reduces the density of competent hosts or the frequency of infected vectors, then biodiversity shows a suppressive 'dilution effect' on disease[16]. On the other hand, biodiversity may increase disease through a variety of mechanisms[15,17]. Pathogen spillover from one host species to another means that community composition as well as species diversity can influence disease dynamics[18]. An unresolved question is whether these effects of community composition are idiosyncratic, or rather follow general rules that can be used to predict disease pressure.

The particular host species that are susceptible to a pathogen are a non-random subset of the local community, because a pathogen is more likely to be able to infect closely related species than evolutionarily distant ones[6,7,19]. This means that phylogenetic distance among species could be used to predict which local species are likely to be alternative hosts for the same pathogen[6]. By extension, species that are phylogenetically rare—that is, evolutionarily distant from species in their surrounding community—should suffer less disease pressure than other species (Fig. 1), as suggested for herbivory[20].

Working in a coastal grassland dominated by annual plants, we linked abundance to disease across all host species. We quantified the relative abundance of all 43 vascular plant species, which ranged over three orders of magnitude in per cent cover. Foliar pathogens are a ubiquitous aspect of plant life and can have a large impact on plant survival and productivity[21], including in these coastal grassland systems[22]. We measured disease pressure on every species as per cent diseased (necrotic or chlorotic) tissue. All species showed some necrotic or chlorotic disease symptoms; mean diseased leaf tissue varied from 1.4% to 37.5% across species. As predicted, disease pressure was significantly lower on rare species (Fig. 2a).

We then tested for an effect of community context on disease. We created a phylogeny for the community from a new supertree with stable ages for all vascular plants (Supplementary Information and Extended Data Fig. 1). We found that disease pressure on a species was inversely related not only to its numerical rarity but also to its phylogenetic rarity (Fig. 2b). We then constructed a tool to estimate how each member of a community could influence the epidemiology of pathogens on a given host species. We used a global database of associations among 210 host genera and 212 fungal pathogens to model the probability of sharing a pathogen as a function of the phylogenetic distance between two plant species (Extended Data Table 1 and Extended Data Fig. 2). This 'PhyloSusceptibility model' (pS model) was used to predict the probability of sharing a pathogen between all pairs of species in the grassland community. Because pathogen spillover should contribute to disease pressure, we tested whether the overall probability of sharing pathogens with neighbours was a good predictor of disease for each species.

Our accuracy in modelling the observed level of disease in each plant species improved when we included the expected influence of other plant species in the community. The most successful predictive model included the abundance of both the focal host and other species (Fig. 2d). We found that disease increased with the phylogenetically weighted abundance of other plants even in the absence of the effect of the focal host abundance (Fig. 2c). This demonstrates that pathogen

[1]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California 95064, USA. [2]Smithsonian Tropical Research Institute, Apartado 2072, Balboa, República de Panamá. [3]Department of Environmental Studies, University of California Santa Cruz, Santa Cruz, California 95064, USA. [4]Center for Integrated Pest Management, North Carolina State University, Raleigh, North Carolina 27606, USA.
*These authors contributed equally to this work.

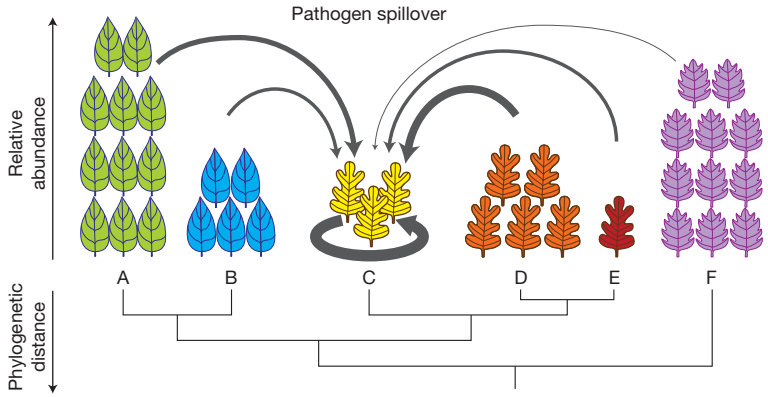


Figure 1 | Pathogen spillover depends on abundance and relatedness. Schematic of the combined influence of phylogenetic distance and relative abundance of community members on pathogen spillover to a focal species. The focal species C is more likely to share pathogens with close relatives (D and E) than phylogenetically distant species (A, B and F), and hosts with greater abundance (indicated by number of leaves) are likely to produce more pathogen inoculum than locally rare species. Arrow thickness reflects the combined effects of phylogenetic distance and relative abundance on spillover from each alternative host in the community.

spillover from other hosts is important, and that phylogenetic structure and ecological structure both play a part in disease pressure across a host community (see Extended Data Fig. 3 for heuristic examples). The PhyloSusceptibility model and the associated stable supertree are tools that can be applied to any plant system to predict pathogen or pest pressure and to generate phylogenetic trees for communities that can be easily compared across studies.

The link between phylogeny and epidemiology implies that pathogens may promote the phylogenetic diversity of plant communities. Our grassland sites comprised assemblages with fewer close relatives than expected by chance from the local species pool (Extended Data Figs 4 and 5). This phylogenetic overdispersion at a local scale has also been observed by others[23], despite the expectation that phylogenetically constrained traits that influence plant distribution will lead to the clustering of related species[24]. The dominant explanation for phylogenetic overdispersion has been competition among similar species[23,24]. However, phylogenetically constrained sharing of pathogens could also provide a local advantage to distantly related species[25].

Furthermore, the role of community structure in disease pressure could help explain why some introduced plants become invasive while others do not. A central hypothesis in invasion ecology is that some species become aggressive weeds and pests because they leave their natural enemies behind in their home range[26,27]. At the same time, the great majority of introduced species do not become invasive, and disease pressure from resident species may contribute 'biotic resistance'[22,27]. Darwin's naturalization hypothesis suggests that those introduced species that are most successful are less closely related to residents[8], a pattern observed in California grasslands[9] although not universally supported[28]. One possible mechanism for this phenomenon is disease pressure originating from closely related resident species. The phylogenetic signal in host range predicts that pathogens that attack introduced species should come primarily from closely related residents, and by extension, introduced plants that are phylogenetically more isolated from the resident community are more likely to escape disease—a phylogenetic rare advantage.

We tested this prediction in our grassland with experimental introductions of novel species. We selected non-horticultural species not present in California from across the angiosperm phylogeny, representing varying degrees of phylogenetic relatedness to resident species (Extended Data Fig. 1). For each of 44 novel hosts, we predicted their susceptibility to local pathogens using the PhyloSusceptibility model and the species lists from the same ten grassland sites. We then grew plants in randomized arrays at the ten sites and quantified per cent symptomatic leaf tissue for each species. For ethical reasons, we removed all novel species before they reproduced.

Disease observed in the novel species was predicted remarkably well by the PhyloSusceptibility model integrated over the resident community (Fig. 3). The introduced species that escaped disease were all species with few close relatives, while those closely related to many residents always showed high levels of disease. In other words, spillover of pathogens from residents onto introduced hosts drove patterns of disease. The nonlinear response in Fig. 3 shows a saturating effect on disease that may reflect a functional limit, such as induced defences that reduce the spread

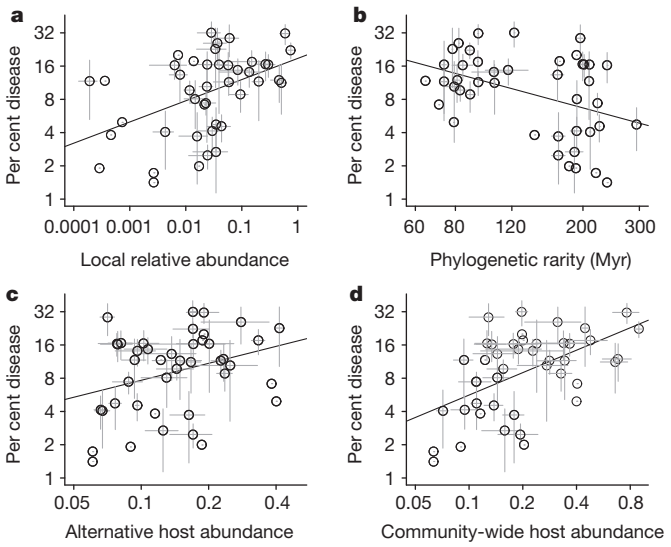


Figure 2 | The joint effects of abundance and phylogenetic signal on disease pressure. **a**, Per cent disease (necrosis and chlorosis) increased with the relative abundance of the focal species ($n = 43$, $r^2 = 0.22$, $P = 0.0017$). **b**, Per cent disease decreased with phylogenetic isolation from the rest of the community (unweighted by abundance), as measured by the 10th quantile of phylogenetic distance from the focal species ($n = 43$, $r^2 = 0.17$, $P = 0.0059$). Myr, million years. **c**, Pathogen spillover influenced disease on focal species; per cent disease was significantly predicted by the sum of the abundance of each other species (i) weighted by its phylogenetically determined probability of sharing a pathogen with the focal species, $p(S)_i$ ($n = 43$, $r^2 = 0.10$, $P = 0.037$). **d**, Disease was most strongly predicted by the focal species abundance combined with the p(S)-weighted abundance of all other species ($n = 43$, $r^2 = 0.28$, $P = 0.00023$). This community-wide, phylogenetically weighted host abundance (**d**; Akaike information criterion (AIC) = 26.6) was a better predictor of disease than the abundance of focal host alone (**a**; AIC = 30.5). Species means (±1 standard error of the mean (s.e.m.)) across ten sites are shown; regressions are based on mean values. Points without error bars represent species measured in only one site.

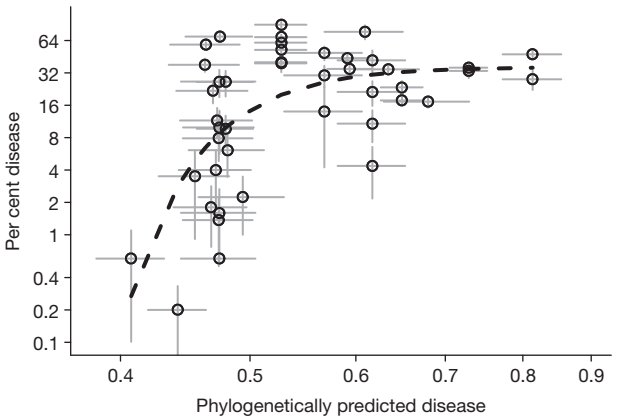


Figure 3 | Prediction of local disease from a global database. A global model of the probability of pathogen spillover as a function of unweighted phylogenetic distance from neighbouring species (PhyloSusceptibility, pS) significantly predicted per cent disease (necrosis and chlorosis) on experimentally introduced plants. The linear model is significant ($\log_{10}(\text{disease}) = 2.27 + 4.16 \times \log_{10}(\text{pS})$, $n = 44$, $r^2 = 0.224$, $P = 0.0011$, residual standard error (RSE) = 0.57, AIC = 79.6), but a three-parameter nonlinear model provided a better fit (shown, $\log_{10}(\text{disease}) = 1.56 - 0.001 \exp(-18.76 \times \log_{10}(\text{pS}))$, RSE = 0.51, AIC = 71.5). Species means (±1 s.e.m.) across ten sites are shown; regressions are based on mean values.

of pathogens into new host tissue, or a constraint on per cent disease imposed by the continuous production of new, uninfected leaves.

In a study at the continental scale, introduced plants that left more pathogens behind and accumulated fewer new pathogens were more likely to be invasive[26]. Our study is the first experimental demonstration, to our knowledge, of a local mechanism for variation in disease that could link Darwin's naturalization hypothesis to this continental pattern.

Crop introductions sometimes fail owing to overwhelming disease pressure from local pathogens[29]. The PhyloSusceptibility model could provide a useful tool for predicting when there is a high probability of elevated disease pressure on proposed horticultural introductions, novel crops or forestry species, biofuels plantations, or new intercropping combinations[6]. Similarly, we can quantify the relative vulnerability of local species of concern to pathogen spillover from these same introduced species. Models to predict the future distribution of plant invasions have been constrained by the difficulty of incorporating biotic drivers, including pathogens and parasites[30]. We show how fairly simple information on the species composition of resident plant communities could help predict both variation in the spread of an invasive species across a landscape and variation in the impact of that species.

Taken together, our results from wild communities and novel host introductions suggest that the structure of communities strongly influences host–pathogen interactions. The role of phylogenetic structure may help explain why loss of species sometimes increases and sometimes depresses disease, shedding new light on an important debate about the value of biodiversity. Finally, pathogens may have a key role in maintaining plant species diversity. Our results indicate that the rare-species advantage should also promote local phylogenetic diversity, leading to communities that capture a broader sampling of evolutionary history.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 September 2014; accepted 5 March 2015.

1. Bagchi, R. *et al.* Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506,** 85–88 (2014).
2. Mordecai, E. A. Pathogen impacts on plant communities: unifying theory, concepts, and empirical work. *Ecol. Monogr.* **81,** 429–441 (2011).
3. Connell, J. in *Dynamics of Numbers in Populations* (eds den Boer, P. J. & Gradwell, G. R.) 298–312 (PUDOC, 1971).
4. Gillett, J. B. Pest pressure, an underestimated factor in evolution. *Syst. Ass. Publication* **4,** 37–46 (1962).
5. Janzen, D. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104,** 501–528 (1970).
6. Gilbert, G. S. & Webb, C. O. Phylogenetic signal in plant pathogen–host range. *Proc. Natl Acad. Sci. USA* **104,** 4979–4983 (2007).
7. Streicker, D. G. *et al.* Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329,** 676–679 (2010).
8. Mack, R. N. in *Proceedings of the IX International Symposium on Biological Control of Weeds* (eds Moran, V. C. & Hoffman, J. H.) 39–46 (Univ. of Cape Town, 1996).
9. Strauss, S. Y., Webb, C. O. & Salamin, N. Exotic taxa less related to native species are more invasive. *Proc. Natl Acad. Sci. USA* **103,** 5841–5845 (2006).
10. Alexander, H. M. Disease in natural plant populations, communities, and ecosystems: insights into ecological and evolutionary processes. *Plant Dis.* **94,** 492–503 (2010).
11. Mangan, S. A. *et al.* Negative plant-soil feedback predicts tree-species relative abundance in a tropical forest. *Nature* **466,** 752–755 (2010).
12. Anderson, R. M. & May, R. M. Population biology of infectious diseases. Part 1. *Nature* **280,** 361–367 (1979).
13. Burdon, J. J. & Chilvers, G. A. Host density as a factor in plant disease ecology. *Annu. Rev. Phytopathol.* **20,** 143–166 (1982).
14. Wood, C. L. *et al.* Does biodiversity protect humans against infectious disease? *Ecology* **95,** 817–832 (2014).
15. Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468,** 647–652 (2010).
16. Johnson, P. T., Preston, D. L., Hoverman, J. T. & Richgels, K. L. Biodiversity decreases disease through predictable changes in host community competence. *Nature* **494,** 230–233 (2013).
17. Randolph, S. E. & Dobson, A. D. Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology* **139,** 847–863 (2012).
18. Power, A. G. & Mitchell, C. E. Pathogen spillover in disease epidemics. *Am. Nat.* **164** (suppl. 5), S79–S89 (2004).
19. Wapshere, A. J. A strategy for evaluating the safety of organisms for biological weed control. *Ann. Appl. Biol.* **77,** 201–211 (1974).
20. Yguel, B. *et al.* Phytophagy on phylogenetically isolated trees: why hosts should escape their relatives. *Ecol. Lett.* **14,** 1117–1124 (2011).
21. Strange, R. N. & Scott, P. R. Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* **43,** 83–116 (2005).
22. Parker, I. M. & Gilbert, G. S. When there is no escape: the effects of natural enemies on native, invasive, and noninvasive plants. *Ecology* **88,** 1210–1224 (2007).
23. Swenson, N. G., Enquist, B. J., Thompson, J. & Zimmerman, J. K. The influence of spatial and size scale on phylogenetic relatedness in tropical forest communities. *Ecology* **88,** 1770–1780 (2007).
24. Webb, C. O., Ackerly, D. D., McPeek, M. A. & Donoghue, M. J. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33,** 475–505 (2002).
25. Liu, X. B. *et al.* Experimental evidence for a phylogenetic Janzen-Connell effect in a subtropical forest. *Ecol. Lett.* **15,** 111–118 (2012).
26. Mitchell, C. E. & Power, A. G. Release of invasive plants from fungal and viral pathogens. *Nature* **421,** 625–627 (2003).
27. Elton, C. S. *The Ecology of Invasions by Animals and Plants* (Univ. of Chicago Press, 1958).
28. Duncan, R. P. & Williams, P. A. Darwin's naturalization hypothesis challenged. *Nature* **417,** 608–609 (2002).
29. Goodell, K., Parker, I. M. & Gilbert, G. S. in *Incorporating Science, Economics, and Sociology in Developing Sanitary and Phytosanitary Standards in International Trade* (ed. Caswell, J.) 87–117 (National Academy Press, 2000).
30. Wisz, M. S. *et al.* The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev. Camb. Philos. Soc.* **88,** 15–30 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Webb for development of the methods of phylogenetic ecology. We thank J. Velzy for greenhouse support and the many undergraduates and volunteers who helped with fieldwork and disease assessments from scanned leaves. Editorial comments were provided by M. Kilpatrick, B. Lyon, C. Fresquez, D. Gordon, S. Grove, S. Gulamhussein, J. Harrower, C. Ray, K. Ross and M. Shu; K. Tanner revised Fig. 1. This research was funded by National Science Foundation grants DEB-0842059 and DEB-1136626, and by a Cooperative Agreement 14-8130-1472-CA between G.S.G. and US Department of Agriculture APHIS-PPQ-CPHST PERAL funded from Section 10201 of the Farm Bill.

Author Contributions I.M.P. and G.S.G. conceived and designed the study, managed the project, performed fieldwork and laboratory work, analysed the data, and wrote the paper. G.S.G. built the supertree R2G2_20140601 and the PhyloSusceptibility model. I.M.P. wrote the initial manuscript. M.S. helped design the study and performed laboratory work. M.B., A.P.W. and R.H. performed fieldwork and laboratory work. R.M. and K.S. contributed the database for the PhyloSusceptibility model.

Author Information Data have been deposited in the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.kf401). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M.P. (imparker@ucsc.edu).

METHODS

Habitat description. Our field sites were located in the extensive grassland associated with the campus of the University of California, Santa Cruz (UCSC) (36° 59′ 18.09′′ N, 122° 3′ 31.29′′ W), on the central coast of California, USA[31]. This area experiences a Mediterranean climate, with a summer dry season and most precipitation coming in the form of winter rains. The area is strongly dominated by herbaceous plants, with only 2% of cover in shrubs. We documented 43 vascular plant species at this location; their abundances showed a characteristic lognormal distribution, with a few very abundant species and 69% of the species with <1% cover. The most common species were introduced Eurasian annual grasses that have dominated California grasslands for over 150 years[32]; *Avena barbata, Bromus diandrus* and *Brachypodium distachyon* together comprised over half of the plant cover.

Plant abundance. We quantified relative abundance of all 43 vascular plant species in the grassland. We located ten sites within the 30-hectare (ha) grassland area, navigating to randomly selected points 44–260 m apart. Each site was sampled in 2011 with a circular plot 20 m in diameter. Within each plot, we quantified per cent cover of all species in eight randomly placed, 50 cm × 100 cm rectangular quadrats. We chose per cent cover as our measure of host abundance rather than number of individual plants, because multiple, independent infections of leaves, combined with the modular growth of plants, point to abundance of plant tissue as the most appropriate measure of host density. From 11 April to 5 May 2011, we used a visual cover estimation method in which each quadrat was first divided into five 20 cm × 50 cm subsections. Two observers independently quantified cover of all species, and combinations of the same three people collected cover data for all ten sites, with high repeatability between observers (mean $r = 0.97$, $n = 81$ quadrats). We calculated post-hoc averages of the two independent cover measures for each site and species. Cover of rare species was estimated at the site scale (314 m^2) by measuring individuals or patches of all species found within the circular plot that were not found in more than two quadrats. Species were identified using the Jepson manual[33].

Disease pressure. Disease pressure was estimated for all species in the ten sites. In each site we laid four 10-m transects in the cardinal directions. At a point every 2 m along each transect, we collected the closest individual plant of all species within a 30-cm radius, thus the maximum sample size per species per site was 20. For any species with fewer than 5 individuals found on the transects, we searched the whole circular plot and collected haphazardly until we reached 5 individuals for that species. To ensure a measure of disease pressure across both young and old leaves, we collected the entire individual, or an entire branch for large perennials such as *Baccharis pilularis*. We then removed every living leaf from the plant, flattened, and glued them onto blue paper. We did not include senescent leaves. The same day, we scanned these pages of leaves at 300 dots per inch (dpi). We analysed scans using Assess: Image Analysis Software for Plant Disease Quantification[34] to estimate the percentage of leaf tissue affected by disease (necrosis and chlorosis).

From our previous extensive work with foliar diseases and herbivory in California grassland plants[22], we know that most necrosis and chlorosis is associated with ascomycete fungal pathogens (for example, *Stemphylium* spp., *Alternaria* spp., *Glomerella* spp., *Phomopsis* spp., *Cladosporium* spp., *Leptosphaerulina* spp.), plus rusts and viruses on the grasses. We carefully examined the scans of leaves at high magnification and excluded damage caused directly by herbivores (for example, necrotic damage from thrips or chewing damage). Although we did not measure the fitness effects of pathogens in this study, our previous work in coastal grasslands documented the effects of these same types of foliar pathogens using statistical modelling, experimental infection under controlled conditions, and fungicide experiments[22]. Because fungicides target different groups of pathogens and have differential effects across plant (and pathogen) species, we could not use a fungicide treatment to uniformly reduce disease pressure across all hosts in the community.

We chose as our measure of foliar disease the proportion of leaf tissue lost to necrosis and chlorosis, because it is directly comparable across host species, and because loss of photosynthetic tissue is a direct index of the impact of disease on host productivity. Our previous work demonstrated a quantitative link between foliar disease and fitness[22]. We captured the most common symptoms caused by above-ground pathogens in the grassland community but not the impacts of vascular wilts, systemic viruses, damping-off, or root rot. These types of diseases could not be incorporated into our assessment of foliar symptoms, but there are no a priori reasons to expect that the pathogens that cause those diseases would have fundamentally different phylogenetic responses than the foliar pathogens[35].

Spatial autocorrelation. We tested for spatial autocorrelation in species composition and overall disease pressure among the ten plots using Mantel tests in R. We calculated dissimilarity matrices among the plots using plant species abundance (quantitative Jaccard), species presence/absence (binary Jaccard), and mean overall disease (quantitative Jaccard). We calculated physical distance between each pair of sites from UTM coordinates (interplot distances ranged from 44 m to 260 m). Mantel tests (based on Pearson's product-moment correlation and 999 permutations) showed no significant spatial autocorrelation for the quantitative Jaccard ($r = 0.140$, $P = 0.226$) or binary Jaccard ($r = 0.146$, $P = 0.17$) for species composition, or for overall disease pressure ($r = -0.067$, $P = 0.63$). In addition, there was no significant relationship between the Jaccard quantitative distances for species similarity and overall disease ($r = -0.054$, $P = 0.59$).

Experimental introductions. We created experimental plant introductions by setting out randomized arrays of novel hosts. We chose plant species that do not grow as native species in California and were not ever, to our knowledge, horticultural introductions in California. We selected herbaceous species from across the angiosperm phylogeny, and for locally important groups (Poaceae, Asteraceae, Fabaceae), we chose species from a number of different tribes. The species were selected to obtain a broad range of phylogenetic rarity in the sample. For each of our candidate experimental species, we used Phylocom (http://phylodiversity.net/phylocom/)[36] to calculate the phylogenetic distances to all resident species using the information on species composition collected in 2011. We then generated histograms to visualize the distributions of those distances. Our final suite of 73 novel species represented a range of phylogenetic similarity to the local community, characterized by the variety of shapes of these histograms. Avoiding horticultural species and plants bred for disease resistance, we ordered seeds from native plant nurseries at a range of locations: Prairie Moon Nursery in Winona County, Minnesota (http://www.prairiemoon.com), Native American Seed in Kimble County, Texas (http://www.seedsource.com), and Ohio Prairie Nursery in Portage County, Ohio (http://www.ohioprairienursery.com). Our final list of 44 species was the subset of these plants that could be germinated successfully and grown within our time frame.

Seeds were sown into flats on 13 January 2012, transplanted after germination into 3.8 cm × 14 cm Ray Leach cone-tainers with potting soil, and grown in the UCSC greenhouse. On 27 March 2012, randomized arrays of 44 species were set out in the field and watered from below using wicks in tubs of water. We used the same ten random locations from which relative abundance was estimated in 2011 and placed one completely randomized array at each site. On 19 April 2012, we quantified disease pressure on each plant, assessing per cent necrotic and chlorotic tissue according to the following categories: 0%, 1%, 5%, 10%, 20%, 50%, 70%, 95% and 100%. No leaves were necrotic at the start of the experiment. Because pathogen infection can increase rates of senescence[37], fully senescent leaves were included as 100% damaged in the results presented here. However, we obtained similar results when we excluded senescent leaves. All arrays were removed from the field before plants could set seed so that no biosafety risk was created by this project.

Estimating phylogenetic distances. We estimated pairwise phylogenetic distances among all plant species in the grassland community and experimentally introduced plants. To make our approach as useful and generalizable as possible, we hand constructed a phylogenetic supertree, R2G2_20140601, with stable ages for all vascular plant families (see Supplementary Information for details of tree construction and sources, as well as a dated Newick file of the full tree). We then used Phylomatic v.4.2 and the Phylocom bladj function (http://phylodiversity.net/phylocom/)[36] to create a dated phylogenetic tree for all taxa in the grassland community and experimentally introduced plants, based on the R2G2_20140601 tree (Extended Data Fig. 1). We calculated pairwise phylogenetic distances among all pairs of species with the cophenetic function in the Picante Package of the R Statistical Framework. Phylogenetic distances are in millions of years (Myr) of independent evolution, which is equivalent to twice the time to most recent common ancestor. We estimated phylogenetic rarity as the 10th quantile distance to other community members in each site, averaged over the ten sites. The 10th quantile was chosen because other measures such as nearest neighbour distance, mean distance, and 25th quantile did not provide as much variability among taxa within which to distinguish patterns. There was no statistical correlation between phylogenetic rarity and local abundance (Spearman $\rho = 0.0014$, $P = 0.93$), which simplifies interpretation of the data.

Evaluating phylogenetic dispersion in plant communities. We compared the phylogenetic dispersion of species within the grassland plots to that expected if species were drawn randomly from a regional metacommunity pool (predicting phylogenetic clustering from habitat filtering) or to the pool of species in the overall grassland community (predicting overdispersion through negative species interactions). For each of the ten sites ($n = 15$–32 species per site), we randomly sampled the same species richness from either the 529 vascular plant species found on the surrounding region of the 810-ha UCSC Campus (including grassland, forest and chaparral habitats[31]) or from the pool of 43 species found across the ten grassland sites. We then calculated the phylogenetic distances (as described earlier) among all pairs of species within a plot (observed) and within each of 9,999 random draws. For each of the 10,000 samples, we calculated a range of quantiles of the pairwise phylogenetic distances from 0% (nearest taxon distance) to 50%

(median distance). The number of species pairs in a plot is $[(n \times n) - n]/2$ where n is the species richness, so a plot with 15 species has 105 phylogenetic distances and a plot with 32 species has 496. We then calculated the difference between each of the phylogenetic distance quantiles from the observed species assemblage to that from each random draw. This difference will be negative if the observed distances are shorter than expected (phylogenetic clustering) and positive if the observed distances are phylogenetically overdispersed (Extended Data Fig. 4). We then recorded the median difference for each of the quantiles for each of the ten sites. We tested whether the phylogenetic distances were greater or less than expected at random using a one-sample t-test (degrees of freedom (df) = 9) against the expected difference of 0 Myr (Extended Data Fig. 5). This approach has an advantage over the more commonly used mean nearest taxon distance or mean pairwise distance metrics[24] because it can uncover shifts in phylogenetic dispersion at many levels.

PhyloSusceptibility estimation of host sharing probabilities. From an independent global data set of associations among 210 plant genera and 212 fungal pathogens, we parameterized a PhyloSusceptibility (pS) model of the probability of sharing a pathogen as a function of phylogenetic distance. Previous work established a broad pattern of phylogenetic signal in host ranges of plant pests and pathogens[6,35]. We re-analysed the same host records for 1,670 pest species (including 212 fungal pathogens) on 210 angiosperm genera from the US Department of Agriculture Global Pest and Disease Database as described previously[35], but using the updated, finer resolution R2G2_20140601 tree with stable dates. We used the same logistic-regression analytical approach as described previously[35], with two exceptions. First, we included all pest species, even those reported from only one host (the 2012 study excluded these specialists, because it focused on polyphagous species). Second, we included the case of zero phylogenetic distance (distance from a known host species to itself). Both of these changes, as well as the finer-scale resolution of the phylogenetic tree, make the estimates of phylogenetic signal more generally applicable and steeper—that is, the probability of a pest being found on two hosts declines more rapidly with phylogenetic distance between the hosts than the estimates from ref. 35. Logistic regressions provide estimates of the probability that any two species of plants would share a pest given the phylogenetic distance between them, following the form $\text{logit}(S) = \beta_0 + \beta_1 \times \log_{10}(\text{PD} + 1)$, where S is whether the target host was susceptible to a pest known from the source host, and PD is the phylogenetic distance (time of independent evolution in Myr) between the source and target host. The probability that the target host is susceptible is then $\text{p(S)} = \exp[\text{logit}(S)]/[1 + \exp(\text{logit}(S))]$. The logistic regression coefficients and confidence intervals for each of nine pest groups is provided in Extended Data Table 1; for fungal pathogens, $\beta_0 = 3.36$, $\beta_1 = -2.86$. The data matrix used for this analysis is available in supplementary data from ref. 35.

For the experimental arrays, we then used this pairwise probability to predict the overall probability that each novel species would share a pathogen with any species in the resident community. A focal species has a probability p(S)_i of sharing a pest with each other species i; the probability that the focal species shares a pest with any resident species can then be calculated as the complement of the product of the probabilities of not sharing a pathogen with each other species $(1 - \text{p(S)}_i)$ using the equation:

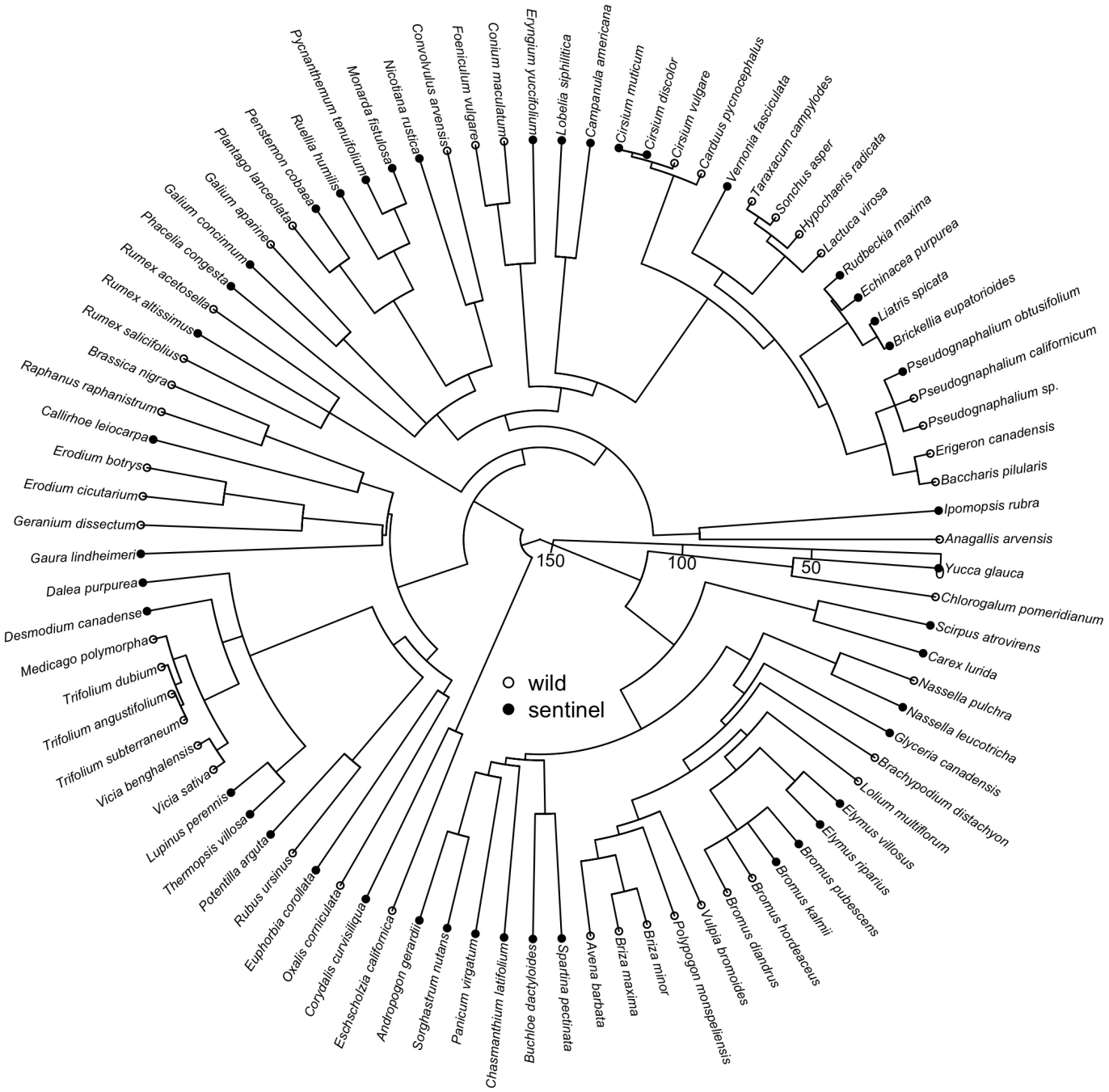
$$1 - \prod_{i=1}^{n-1}(1 - \text{p(S)}_i)$$

for a community with n host species total. We call this the PhyloSusceptibility (pS) model.

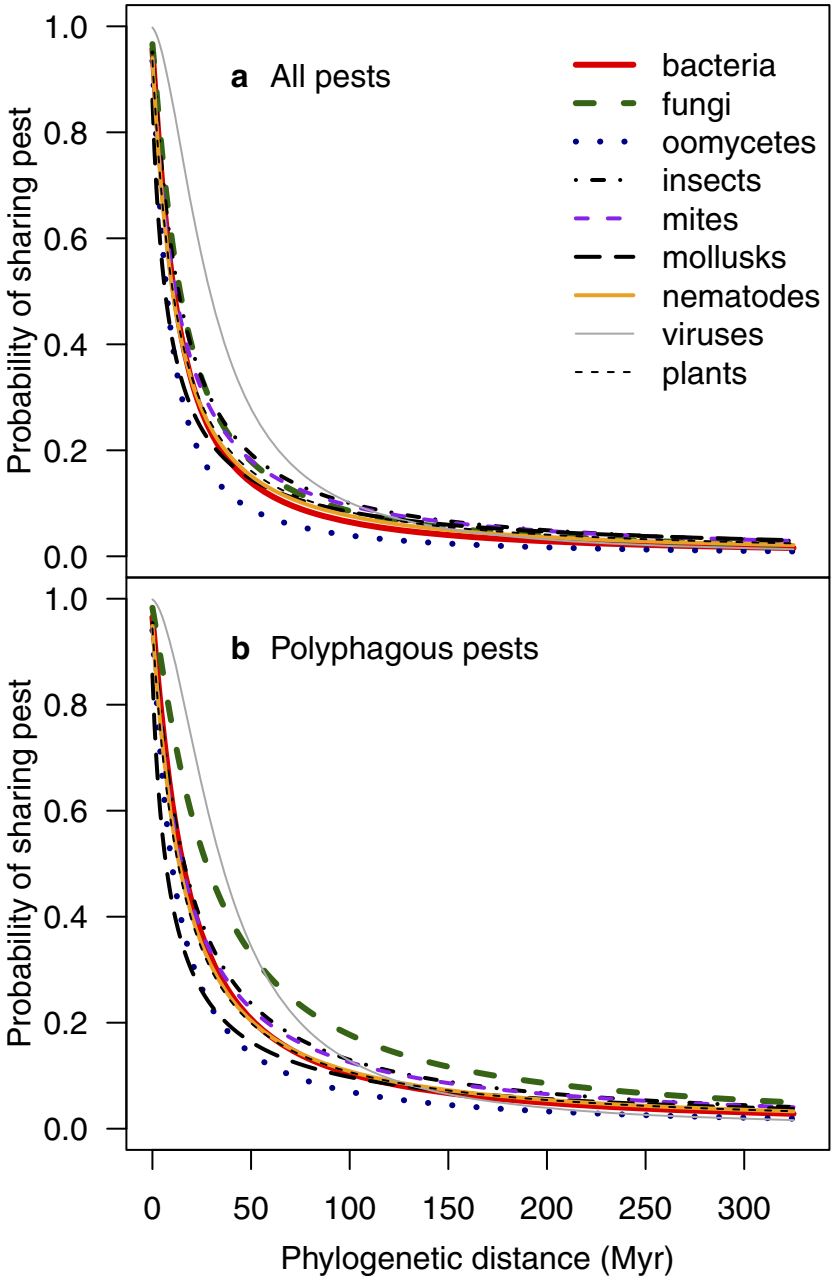
Prediction of disease. For the wild plant community, we tested the predictive power of conspecific abundance, phylogenetic rarity and community structure (the abundance of all other species weighted by their phylogenetic distance) in explaining disease. We used linear regression to test the response of ($\log_{10}$) proportion diseased leaf tissue to ($\log_{10}$) relative abundance of the focal host, and to the 10th quantile phylogenetic distance to the other plant species (without weighting for relative abundance). We also calculated 'alternative host abundance' by multiplying the abundance of each other plant species i by p(S)_i, the probability of sharing pathogens between that species and the focal species (pathogen spillover), then taking the sum over all other species. We added the focal species abundance to alternative host abundance to generate 'community-wide host abundance'. Disease and abundance measures were determined for each of the ten sites and averaged before analysis. We used linear regression to test the ability of ($\log_{10}$) alternative host abundance and ($\log_{10}$) community-wide host abundance to predict ($\log_{10}$) proportion disease.

For the experimental introduction experiment, we fit a nonlinear regression model to quantify the power of the pS model for predicting ($\log_{10}$) proportion diseased tissue on the novel species. Including local abundance of the resident species did not improve the model fit (data not shown). Log transformations were used to improve normality and heteroscedasticity. All analyses were done in R v.3.0.2.

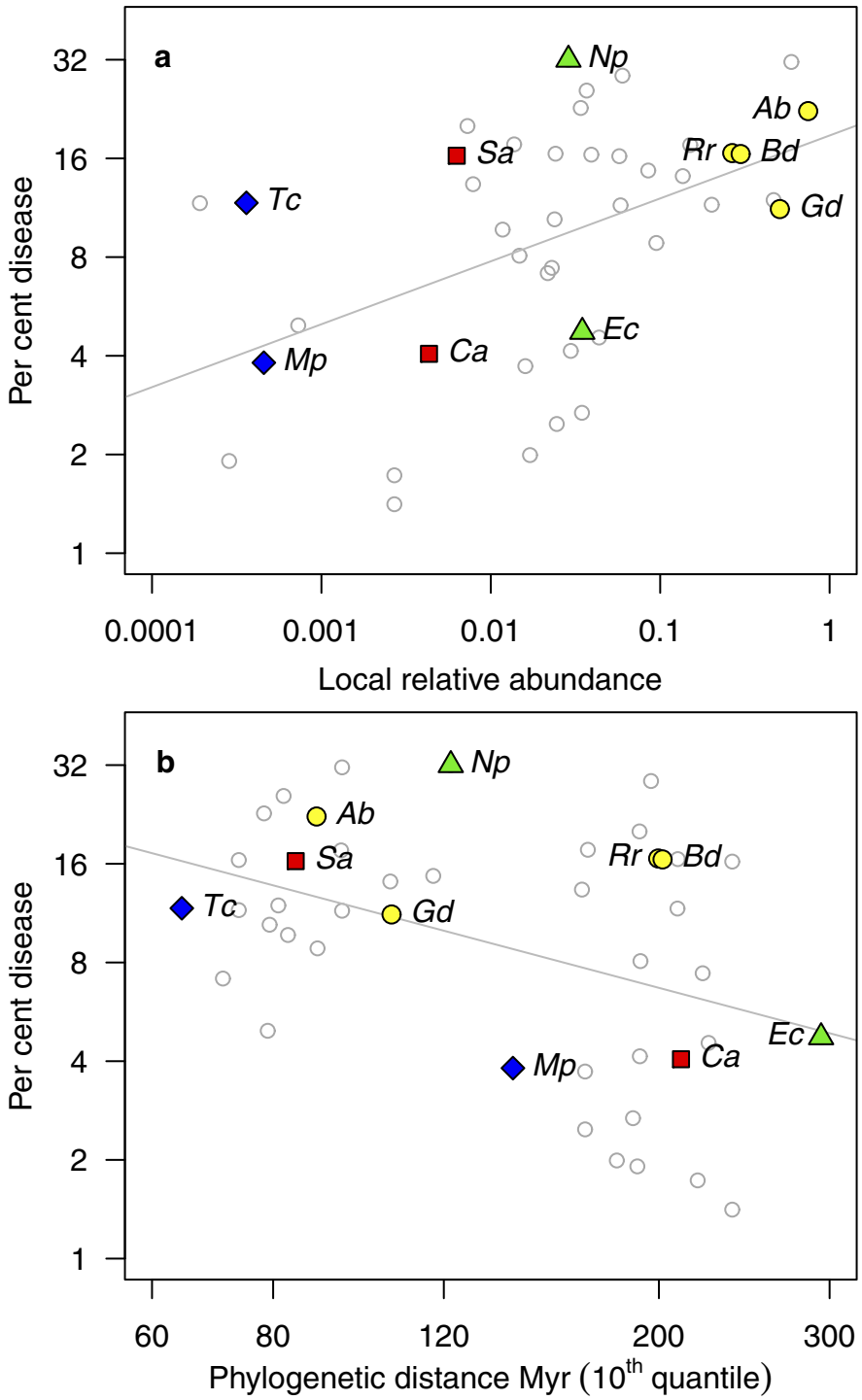
31. Haff, T. M., Brown, M. T. & Tyler, W. B. *The Natural History of the UC Santa Cruz Campus* 2nd edn (Environmental Studies, UC Santa Cruz, 2008).
32. Mack, R. N. in *Biological Invasions: A Global Perspective* (ed. Drake, J. A.) 155–179 (John Wiley & Sons, 1989).
33. Baldwin, B. G. & Goldman, D. H. *The Jepson Manual: Vascular Plants of California* 2nd edn (Univ. of California Press, 2012).
34. Assess 2.0. *Image Analysis Software for Plant Disease Quantification v. 2.0* (APS Press, 2008).
35. Gilbert, G. S., Magarey, R., Suiter, K. & Webb, C. O. Evolutionary tools for phytosanitary risk analysis: phylogenetic signal as a predictor of host range of plant pests and pathogens. *Evol. Appl.* **5,** 869–878 (2012).
36. Webb, C. O. & Donoghue, M. J. Phylomatic: tree assembly for applied phylogenetics. *Mol. Ecol. Notes* **5,** 181–183 (2005).
37. Gilbert, G. S. & Parker, I. M. Rapid evolution in a plant–pathogen interaction and the consequences for introduced host species. *Evol. Appl.* **3,** 144–156 (2010).



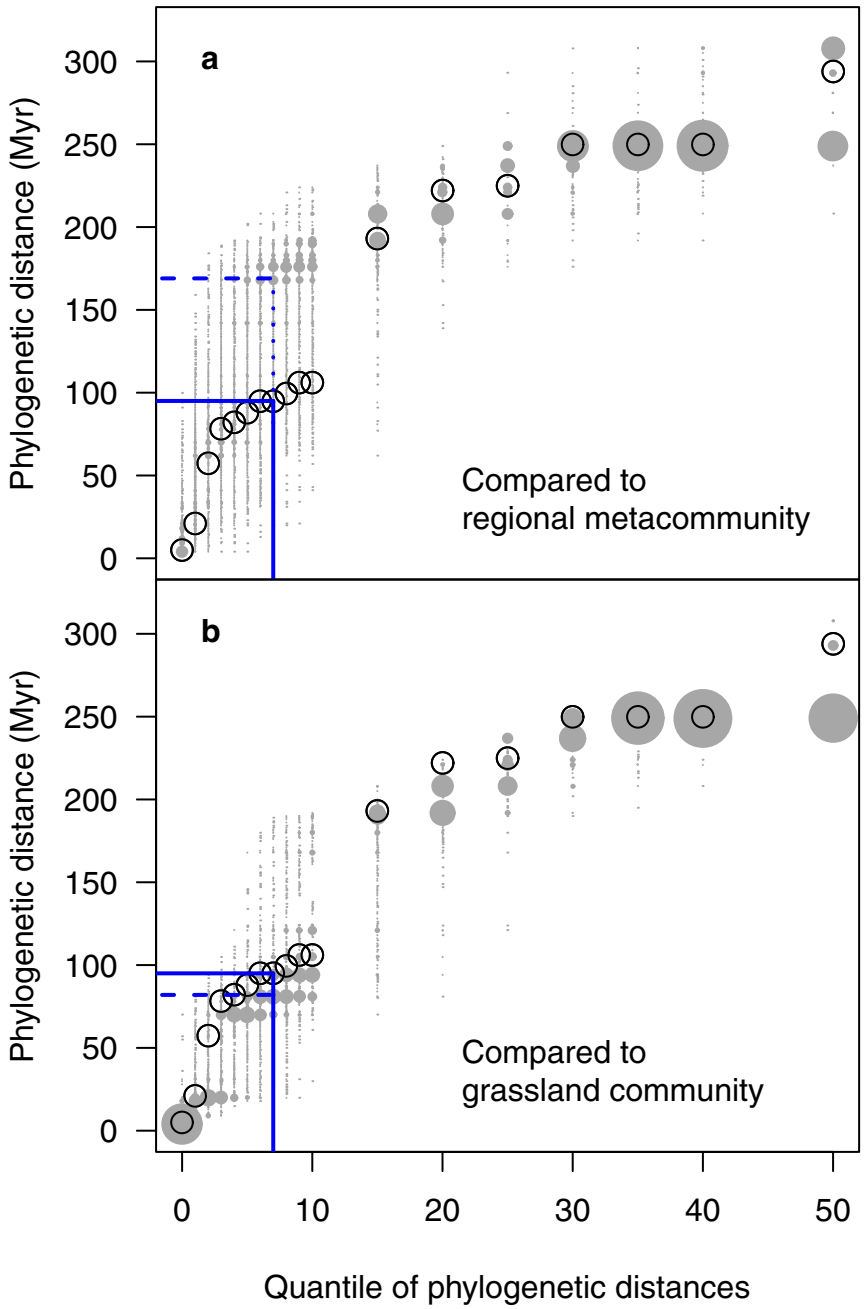
Extended Data Figure 1 | Phylogenetic tree of grassland and experimental species. Phylogenetic tree of 43 wild species (open circles) in the grassland habitat and 44 experimentally introduced species (filled circles) of plants. Phylogenetic topology and distances (in Myr) were derived from supertree R2G2_20140601 (see Supplementary Methods and Supplementary Data).



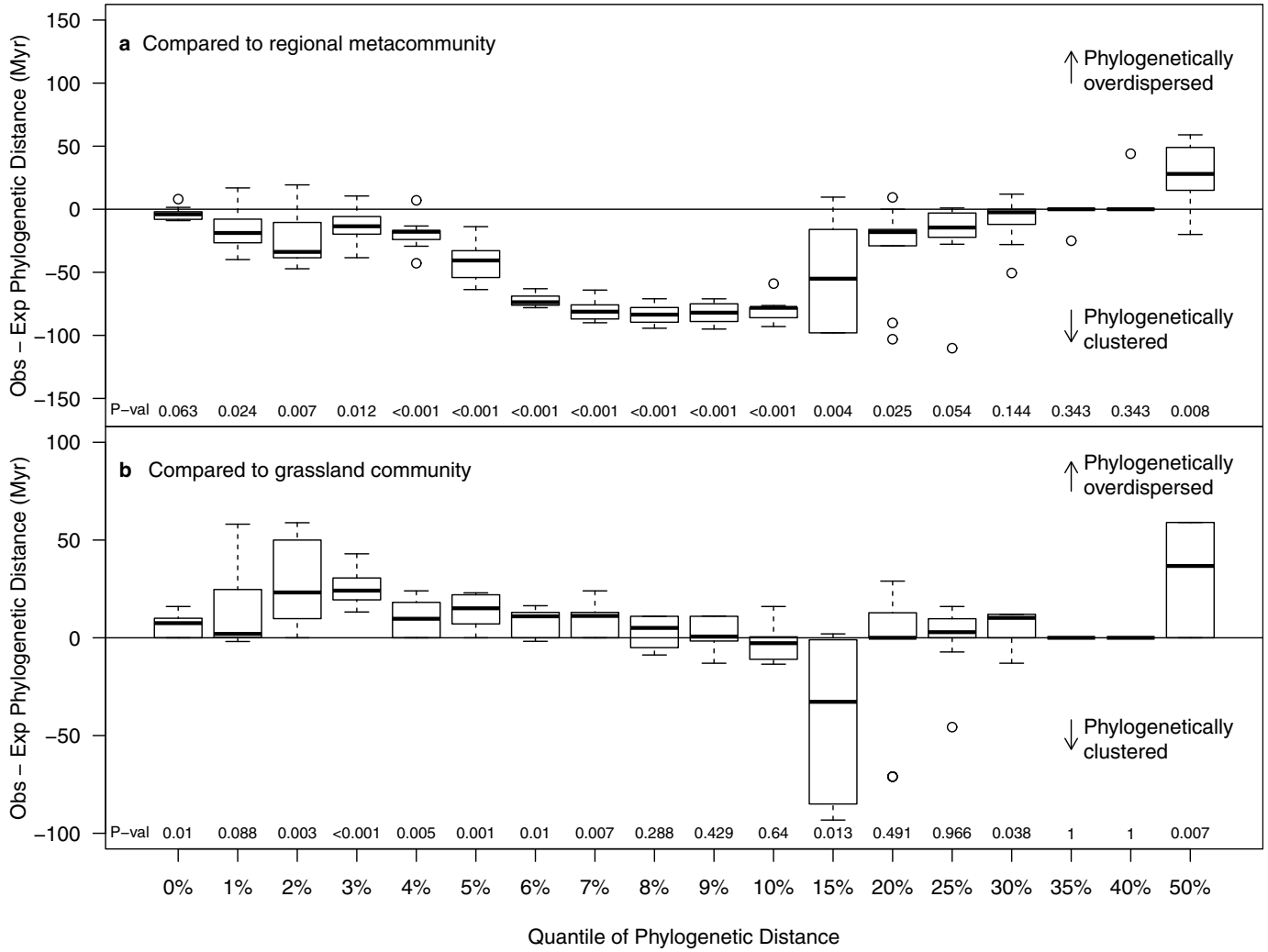
Extended Data Figure 2 | Phylogenetic signal in host sharing. **a**, **b**, The influence of phylogenetic relationship on the probability of sharing a pest between two hosts was modelled from a database of 1,670 pests on 210 angiosperm host genera in the US Department of Agriculture Global Pest and Pathogen Database (described in ref. 35). Phylogenetic distances are calculated from the updated phylogenetic tree R2G2_20140601 (Supplementary Information). The regression takes the form of $\text{logit}(S) = \beta_0 + \beta_1 \times \log_{10}(\text{PD} + 1)$, where S is the probability of susceptibility and PD is the phylogenetic distance (time of independent evolution, in Myr) between the source and target host genera. The probability that a target host is susceptible to a pest from a source host is then $S = \exp(\text{logit}(S))/[1 + \exp(\text{logit}(S))]$. Coefficients were calculated for nine groups of pests/pathogens: bacteria, fungi, oomycetes, insects, mites, molluscs, nematodes, viruses and plants (Extended Data Table 1). **a**, **b**, Coefficients were generated for all pests (**a**), and for the subset of (polyphagous) pests known from more than one host genus (**b**). The latter is how the analysis was originally done in ref. 35.



Extended Data Figure 3 | Heuristic example of the joint influences of focal species abundance and phylogenetic isolation on disease. **a**, **b**, A few examples can be used to illustrate the joint influences of focal host abundance and the presence of close relatives on disease. Most of the most common species, for example, *Avena barbata* (*Ab*), *Bromus diandrus* (*Bd*), *Raphanus raphanistrum* (*Rr*) and *Geranium dissectum* (*Gd*), experienced high levels of diseased tissue, regardless of the identity of their neighbours (**a**, **b**, yellow circles). In contrast, although both *Medicago polymorpha* (*Mp*) and *Taraxacum campylodes* (*Tc*) were very rare (**a**, blue diamonds), *T. campylodes* experienced three times as much disease as *M. polymorpha*, perhaps because it had many more close relatives in the local community (**b**). Similarly, *Convolvulus arvensis* (*Ca*) and *Sonchus asper* (*Sa*) were found at the same abundance (**a**, red squares), but *S. asper* had more close relatives than *C. arvensis* (**b**) and experienced over four times as much disease. At intermediate abundance, *Eschscholzia californica* (*Ec*) and *Nassella pulchra* (*Np*) had very similar cover, but the latter suffered more than six times as much disease as the former (**a**, green triangles). Again, the species with more disease co-occurred with several close relatives while the relatively disease-free species was phylogenetically isolated (**b**). Note that axes are on a log scale.



Extended Data Figure 4 | Measuring phylogenetic dispersion of grassland communities. **a**, **b**, Phylogenetic dispersion analysis from one representative site. Open circles show the observed phylogenetic distance among the 29 species found in grassland site 2, for each of 18 quantiles (from 0 to 50%). Grey circles show phylogenetic distance quantiles for random draws from the 529 species in the regional metacommunity (the UCSC campus) (**a**) or the 43 species found in the surrounding grassland community (**b**). Size of the grey circle is proportional to the number of times that distance was found in 10,000 independent random draws of 29 species. Solid blue lines show the observed phylogenetic distance at the 7% quantile; dotted lines show medians of the random draws. **a**, **b**, The observed distance was 74 Myr less than the median of the random draws from the regional metacommunity at the 7% quantile (that is, phylogenetic clustering, consistent with habitat filtering) (**a**), and the observed distance was 13 Myr greater than expected from random draws of grassland species only (that is, phylogenetic overdispersion, consistent with negative species interactions) (**b**). Statistical tests based on independent estimates of dispersion for each of the ten sites are shown in Extended Data Fig. 5.



Extended Data Figure 5 | Test of phylogenetic dispersion of grassland communities. **a**, Plant species in the grassland were consistently phylogenetically clustered compared to the pattern expected if species were drawn randomly from the regional pool (529 species on the 810-hectare UCSC campus). Values less than zero indicate phylogenetic clustering, while values greater than zero indicate species are more distantly related than random (overdispersed) (Extended Data Fig. 4). One-tailed *t*-tests (df = 9, *P* values given) showed species were more closely related than expected through the first quartile (25%). This is consistent with habitat filtering. **b**, In contrast, species were less closely related than expected if drawn randomly from the 43 species found in the grassland community, especially at shorter phylogenetic distances (through the 7% quantile). Box-and-whisker plots indicate median (dark line), 25th and 75th quartiles (box) and minimum and maximum (whiskers), with the ten sites as replicates.

Extended Data Table 1 | Logistic regression coefficients of phylogenetic signal in pest sharing among plant hosts

		All pests			Only polyphagous		
Group	Statistic	β_0	β_1	# pests	β_0	β_1	# pests
bacteria	median	3.095393	-2.878887	52	3.317268	-2.730410	30
bacteria	L95%CI	4.012584	-2.571435		5.030721	-2.267532	
bacteria	U95%CI	2.394909	-3.276410		2.245152	-3.474872	
fungi	median	3.358610	-2.857656	212	4.053500	-2.788080	95
fungi	L95%CI	4.136312	-2.558613		5.381187	-2.290185	
fungi	U95%CI	2.682188	-3.197864		2.890338	-3.368365	
oomycetes	median	2.660190	-2.921501	61	2.758395	-2.668684	32
oomycetes	L95%CI	3.158198	-2.668083		3.539687	-2.311036	
oomycetes	U95%CI	2.116981	-3.142015		1.953303	-3.016232	
insects	median	2.959036	-2.571142	870	3.032544	-2.461679	637
insects	L95%CI	3.216122	-2.470206		3.346828	-2.337998	
insects	U95%CI	2.729638	-2.682801		2.745062	-2.597728	
mites	median	2.736459	-2.486888	119	2.879907	-2.405970	87
mites	L95%CI	3.435860	-2.191249		3.859141	-2.042179	
mites	U95%CI	2.049466	-2.793569		2.037499	-2.826249	
mollusks	median	1.830065	-2.106471	45	1.789106	-2.003059	37
mollusks	L95%CI	2.443528	-1.863166		2.550251	-1.720545	
mollusks	U95%CI	1.273475	-2.368293		1.129949	-2.329349	
nematodes	median	2.751777	-2.618922	105	2.916024	-2.506838	70
nematodes	L95%CI	3.197427	-2.442838		3.577336	-2.257961	
nematodes	U95%CI	2.349906	-2.815817		2.325130	-2.790456	
viruses	median	6.142249	-4.154148	135	6.691393	-4.295204	108
viruses	L95%CI	6.917711	-3.817072		7.740598	-3.877727	
viruses	U95%CI	5.360523	-4.491828		5.726995	-4.739907	
plants	median	2.668354	-2.528355	71	2.822546	-2.467609	53
plants	L95%CI	3.214360	-2.303952		3.713974	-2.173069	
plants	U95%CI	2.148842	-2.763674		2.120972	-2.845282	

Parameters (median, lower and upper 95% confidence intervals) estimated for a model relating the probability of sharing a pest between two hosts to the phylogenetic distance between the hosts. The model of susceptibility (S, probability the pest will cause disease or damage on the target host) is $\text{logit}(S) = \beta_0 + \beta_1 \times \log_{10}(\text{PD} + 1)$, where PD is the phylogenetic distance (time of independent evolution, in Myr) between the source and target host genera. Phylogenetic distances are calculated from the updated phylogenetic tree R2G2_20140601 (Supplementary Information). Data are from 1,670 pests on 210 angiosperm plant genera in the US Department of Agriculture Global Pest and Pathogen Database. Analyses labelled 'only polyphagous' excluded pests only recorded from one host genus. Confidence intervals are from 1,000 runs of analysis in which a different host was randomly selected as the 'source' host for each pest. Predicted curves are shown in Extended Data Fig. 2.