scientific data

Article in Press

Check for updates

OPEN

DATA DESCRIPTOR

A decade of disease survey data in a progeny-provenance trial: Dothistroma needle blight in Scots pine

Annika Perry[1]✉, Beth Moore[2], Susan Jones[2], Sundeep Kaur[3], Bridget Crampton[3], Amir Gurung[3], Krisztian Nemeth[1], Hattie Barber[1], Jenni Stockan[4], Joan Cottrell[3], Joan Beaton[4] & Stephen Cavers[1]

Long-term data on disease susceptibility in forest trees are rare but essential for understanding host–pathogen dynamics and genetic variation in susceptibility traits. We present a multi-year, multisite common garden dataset quantifying susceptibility of Scots pine (*Pinus sylvestris*) to Dothistroma needle blight. The dataset comprises annual disease assessments collected from the same trees across 11 years, spanning 168 families and 21 Scottish provenances. This design enables partitioning of genetic and environmental sources of variation, evaluation of temporal stability in host response, and estimation of variance components and narrow-sense heritability of susceptibility. The data support analyses of phenotypic plasticity, provenance-level responses, and interactions between disease susceptibility and other adaptive traits. This data resource will facilitate predictive modelling of host susceptibility under current and future environmental conditions.

Background & Summary

Understanding and predicting host susceptibility to disease is critically important for the effective management of existing forests and for informing future forest planning under changing environmental conditions. Disease outbreaks can have profound ecological and economic impacts, and the ability to anticipate variation in host response is therefore increasingly central to both conservation and applied forestry. Despite this, progress in understanding tree disease resistance is often constrained by the limited availability and duration of robust, repeated susceptibility measurements. Visual assessments of the severity of disease symptoms, while widely used, can be subjective and are rarely validated against independent molecular estimates of infection, potentially obscuring true patterns of host response.

Long-term studies involving regular assessment provide a powerful framework for investigating disease dynamics, as they allow temporal variation in host–pathogen interactions to be quantified and enable trait stability to be evaluated across years. Such data are particularly valuable for distinguishing transient responses from consistent phenotypic differences among individuals or genotypes. However, collecting longitudinal data is inherently challenging for large, long-lived organisms such as trees, where disease development, host response, and environmental exposure unfold over extended temporal and spatial scales. Common garden experimental trials are valuable resources for longitudinal studies, enabling repeated measures performed on the same trees. Their design allows genetic and environmental contributions to trait variation to be disentangled, and trials which span multiple sites, families and provenances enable the quantification of phenotypic plasticity alongside genetic components of susceptibility.

Dothistroma needle blight (DNB) is one of the most important diseases of pines[1] caused by the ascomycete fungal pathogen *Dothistroma septosporum* (and occasionally *Dothistroma pini*) and affecting more than 100 species[2] in every continent except Antarctica. Symptoms, including red-brown lesions on needles, can lead to premature needle loss[3,4], growth retardation and, in severe cases, tree death[5]. As pines are major commercial species

[1]UK Centre for Ecology & Hydrology, Penicuik, UK. [2]The James Hutton Institute, Dundee, UK. [3]Forest Research, Alice Holt Lodge, Surrey, UK. [4]The James Hutton Institute, Aberdeen, UK. ✉e-mail: annt@ceh.ac.uk

worldwide, *D. septosporum* is of significant economic concern and extensive efforts to characterise the pathogen and associated host response have been undertaken in North America[4,6], Europe[7,8], and Australasia[9–12].

Scots pine (*Pinus sylvestris*) is the most widely distributed pine species and is of global ecological and economic importance. There is extensive genetic and phenotypic diversity in the species[7,13–15], which is unsurprising given the range of environments it occupies. This is coupled with very low population genetic structure and high gene flow across its range, reflecting its high dispersal capability as a wind-pollinated species[16,17]. In Scotland, the remnant native populations are typically small, highly fragmented and distributed across an environmentally highly heterogeneous landscape. It has been demonstrated that there is heritable genetic variation in susceptibility to infection by *D. septosporum* in Scots pine[7] and there is evidence of coevolution of the pathogen and host[8]. However, despite their critical importance for evaluating the vulnerability of native and planted forests, the long-term dynamics and mechanisms underlying the pathosystem are poorly understood.

The dataset presented here is a valuable resource with which to address that knowledge gap. Its longitudinal structure and experimental design allow: 1) evaluation of the drivers of disease severity; 2) identification of genetic variation in responses among provenances; 3) assessment of the temporal stability of disease resistance; 4) estimation of genetic variance components and narrow-sense heritability associated with susceptibility; 5) exploration of interactions with other key adaptive traits such as growth and reproduction; 6) development of predictive models for host susceptibility; and 7) investigation of the mechanisms underlying host responses to disease.

Methods

Experimental field trials. Seedling source, nursery environments and field sites are detailed by[18].

The field site in the south of Scotland (FS, latitude 55.603625°, longitude −2.893025°) was visually surveyed annually from 2015 to 2025 (inclusive) except for 2019 and 2020. Replicate field sites in the east (FE, latitude 56.893567°, longitude −2.535736°) and west of Scotland (FW, latitude 57.775714°, longitude −5.597181°) were surveyed in 2016 and 2022.

Annual disease surveys. Surveys to record evidence of susceptibility to DNB were performed annually in late summer (between 22nd August and 22nd September: exact dates are recorded in the dataset). Surveyors were experienced in identifying symptoms of DNB. Individual trees were visually assessed for the percentage of needles with symptoms of DNB ('DNB', symptomatic needles) by applying a scale based on 5% increments (where 1% is equivalent to negligible symptoms). It has been observed that trees with high levels of infection lose their needles sooner than those with low levels of infection[3,4], potentially confounding estimates of the proportion of symptomatic needles (very low levels of infection may be recorded where a very susceptible tree has shed all infected needles). For this reason, needle retention ('NR') was also recorded at each survey. As annual growth stages are visible in pines, NR could be estimated in half-year increments (in 2016 it was recorded in quarter-year increments). From 2015 to 2017 a single estimate for each of NR and DNB was recorded for each tree annually. As trees grew larger with time, from 2018 onwards mean values for DNB and NR were obtained from separate estimates for the top, middle and bottom thirds of the canopy of each tree.

Data on symptomatic needles (trait 'DNB') were highly left-skewed, so were log(DNB + 1) transformed. Where multiple values were obtained per tree (from different crown positions), these were averaged prior to log-transformation. We have termed this trait '*direct susceptibility*' as it directly captures disease symptoms on the host tree.

Needle retention values (trait 'NR') were inverted (1-NR) to provide a measure of needle loss: this also ensured that both susceptibility traits were easily and intuitively interpretable (i.e. that low values denote low susceptibility and high values denote high susceptibility). Needle retention, which we assume arises as a cumulative response to infection and may also reflect host response to other needle diseases or a lack of needle drop as a response, we have termed '*indirect susceptibility*'.

Long-term quantification of susceptibility. Direct and indirect long-term disease susceptibility was quantified for each tree at FS using the Area Under the Disease Progress Stairs (AUDPS[19]), which summarises repeated measurements over time into a single value ('*direct-LT susceptibility*', '*indirect-LT susceptibility*'). Trees that died over the period 2015 to 2025 were removed from the analysis. For each trait, AUDPS was calculated using the agricolae package[20] in R[21].

For the field sites in which survey data were only collected in 2016 and 2022 (FE and FW), '*direct-LT susceptibility*' and '*indirect-LT susceptibility*' were represented by the mean value for each trait across the two surveyed years. Where data were missing for one of the years, long-term susceptibility was not estimated.

Finally, direct and indirect susceptibility traits were normalised, to a minimum value of 0 and a maximum value of 1, within each site.

Grouping long-term susceptibility traits. To ensure that these data are also compatible with analyses that require categorical rather than quantitative trait data (for example, when targeting trees to subsample from the extreme ends of trait distribution), we also outline an approach to categorise the long-term susceptibility data into three distinct groups: '*low susceptibility*'; '*moderate susceptibility*' and '*high susceptibility*' for each trait based on their rank.

The approach assumes that the tails of the distributions reflect the extremes of host response. Susceptibility traits were discretised using percentile ranks calculated across the full distribution of observations at each site. For each tree, a percentile rank was computed based on its position within the ordered trait distribution. Individuals falling within the lower and upper 20% of the distribution were classified as low susceptibility and high susceptibility, respectively, while all remaining individuals were assigned to the moderate susceptibility category.

Trait	*Description*	*Unit*
DNBYY_pos	Estimated % needles with symptoms (1% is equivalent to negligible symptoms) subsequent categories are in 5% increments. YY = year surveyed; pos = crown position surveyed (B = bottom third; M = middle third; T = top third; all = whole tree)	%
NRYY_pos	Estimated years needle retention in quarter-year increments (for 2016) or half-year increments. YY = year surveyed; pos = crown position surveyed (B = bottom third; M = middle third; T = top third; all = whole tree)	Years
DirectYY	Annual direct susceptibility. YY = year surveyed (mean value where multiple crown positions surveyed)	log(DNB + 1)
IndirectYY	Annual indirect susceptibility, inverted (1-NR) to represent needle loss. YY = year surveyed (mean value where multiple crown positions surveyed)	1-NR
Direct-LT	AUDPS (site: FS) or mean (sites: FE and FW) of DirectYY across all available surveys, normalised within each site	AUDPS or mean, normalised value(0-1)
Indirect-LT	AUDPS (site: FS) or mean (sites: FE and FW) of IndirectYY across all available surveys, normalised within each site	AUDPS or mean, normalised value(0-1)
Direct-LT_group	Direct long-term susceptibility group (low; moderate; or high susceptibility) based on percentile rank of Direct-LT	Low; Moderate; High
Indirect-LT_group	Indirect long-term susceptibility group (low; moderate; or high susceptibility) based on percentile rank of Indirect-LT	Low; Moderate; High

Table 1. Susceptibility to DNB traits assessed in Scots pine trees at three common garden field trial sites. Where an element of the trait names is in italics, it indicates multiple available options. 'YY' = Year, where e.g. 15 is 2015. 'pos' = which third of the tree was surveyed: T, top; M, middle; B, bottom; all, whole tree surveyed. Long-term susceptibility traits have the suffix '-LT'.

Validation of susceptibility traits using molecular data. During our visual surveys, surveyors made the assumption that the symptoms observed were predominantly due to the pathogen *D. septosporum* (symptoms attributed to other pathogens were not recorded). To validate this molecularly, we used an existing metabarcoding dataset that was generated at the FS site[22]. The raw sequence data can be obtained from the European Nucleotide Archive under PRJEB88228[23]. We describe briefly the approach used but for a full description of molecular methods see 19. A subset of trees (N = 200) was selected to represent the full suite of DNB susceptibility responses while retaining the experimental design of the full trial (described in[18]). Uninfected needles (only those that were visually inspected as green and healthy with no apparent lesions or discolouration) were collected from four separate branches around each tree's crown and sealed within a labelled polythene bag. Needles were lyophilised, homogenised and total cellular DNA was extracted using a modified Qiagen DNeasy 96 kit (10 µl proteinase K was added during the incubation stage). An optimised Illumina metagenomic library preparation protocol was used to amplify the fungal ITS2 region which was sequenced on a NextSeq. 2000 using a 2 × 300 bp paired end strategy. Raw reads were processed to OTU tables using a qiime2 pipeline[24]. Briefly, reads were filtered with fastp[25] to remove short reads (<90 bp), polyG tails, flanking regions surrounding the target sequence, adapters and reads with high error rates. Remaining reads were denoised with dada2[26], and classified with classify-sklearn against the UNITE v9.0 database[27] with a 97.5% Species Hypothesis (SH) threshold to produce OTU tables, which are available from[28].

To align with the disease survey data (conducted on 21-22 September 2023) we use only the timepoint 3 data which was generated from needle samples collected for amplicon sequencing on 27-28 September 2023 and perform the following additional data curation. The assignment of taxa to the OTUs was verified using massBLASTer against the UNITE and INSD databases. Relative abundances of OTUs per tree, where the number of reads per OTU was presented as a proportion of the total number of reads for each tree, were successfully obtained for all but three of the 200 trees (missing tree ids due to poor quality amplification and sequence data: 7291, 7338, 7474). Where multiple OTUs were assigned to a single taxon, total relative abundance was calculated by summing the relative abundance of all assigned OTUs. Additionally, all *Dothistroma* OTUs were binned as *D. septosporum*, as the closely related pathogen, *D. pini*, has not been recorded in the UK. This exercise was completed for all needle pathogens that are known to cause needle loss (*Coleosporium tussilaginis; Cyclaneusma minus; Gremmenia infestans; Gremmeniella abietina; Lophodermella conjuncta; Lophodermium conigenum; Lophodermium seditiosum; Pachyramichloridium pini; Rhizosphaera kalkhoffii; Sirococcus piceicola)* to validate that symptoms in the visual survey dataset were due to infection by *D. septosporum* and not to other needle pathogens. The OTU tables were then subset to only include the relative abundance of these validated needle pathogens providing a data resource suitable for use in molecular validation.

Data Records

Disease survey data are deposited with the Environmental Information Data Centre (https://eidc.ac.uk[29]). Traits measured are detailed in Table 1.

For the dataset, the columns are:

1. PopulationCodeOriginal: Population code originally assigned to the tree
2. FamilyCodeOriginal: Family code (individuals with a shared family code are from the same mother tree) originally assigned to the tree
3. PopulationCodeNew: Population code assigned following genotype checks
4. FamilyCodeNew: Family code (individuals with a shared family code are from the same mother tree) assigned following genotype checks

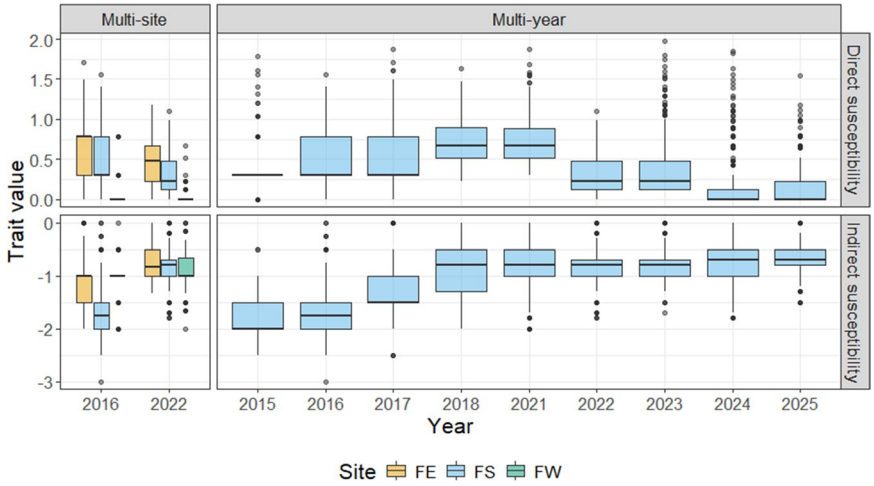


Fig. 1 Boxplots showing variation recorded during annual DNB surveys for direct (top panels) and indirect (bottom panels) susceptibility traits. Left: variation among 3 sites in multi-site surveys conducted in 2016 and 2022. Right: variation among years in surveys conducted over the period 2015 to 2025 (excluding 2019 and 2020) at FS. Solid black lines indicate the median trait value. The bottom and top of boxes indicate the first and third quartile. The upper and lower whiskers extend to the highest and lowest values within 1.5 times the interquartile range. Individual points indicate outliers.

5. FieldSite: Field trial site where the tree is planted (FE: field site in the east of Scotland, Glensaugh; FS: field site in the south of Scotland, Yair; FW: field site in the west of Scotland, Inverewe)
6. FieldCode: the unique four-digit code identifying each tree links this DNB survey dataset directly with other published datasets[30,31]; FE: 5001 to 5672; FW: 6001 to 6004; FS: 7001 to 7672)
7. Block: Randomised block code (A-D)
8. Row: Row that the tree was planted in (numerical)
9. Column: Column that the tree was planted in (numerical)
10. Susceptibility traits: see Table 1 for list of traits and format of column header. Missing data = NA. Data not collected = NC

Relative abundances of needle pathogens for trees within the trial (samples collected 27-28 September 2023) to support molecular validation are deposited in a Zenodo repository[32] along with the intermediate data files needed to generate this table. In this table, 'id' refers to the FieldCode listed above.

Technical Validation

Surveys were performed annually at the same time of year to ensure consistency of the method. Data were checked after each survey and obvious errors were immediately corrected by revisiting the tree and re-surveying the trait in question.

We used boxplots to visualise data range and data distribution for each annual susceptibility trait (direct susceptibility and indirect susceptibility) at each field site (Fig. 1).

We used a pairwise correlation plot to visualise the relationship among normalised long-term susceptibility traits (direct-LT susceptibility and indirect-LT susceptibility) at each field site (Fig. 2). The two traits are positively correlated within one another, but the strength of the relationship differs among the sites.

Discretisation of the long-term susceptibility trait data was applied to allow future analyses that require categorical variables (Fig. 3).

This approach is not recommended for direct-LT and indirect-LT susceptibility traits for trees growing at FW and FE given the skewed and/or patchy distribution of the traits at these sites (Figs. 2,3).

Molecular validation was performed by regressing annual direct and indirect susceptibility traits against the proportion of reads that were identified as *D. septosporum* (Fig. 4). This demonstrated that direct susceptibility was highly significantly positively correlated with the relative abundance of in planta *D. septosporum* when the survey is performed in the same year (2023, Fig. 3). For surveys performed the year before (2022) or the year after (2024) there were no significant associations between trees' relative abundance of *D. septosporum* and their direct susceptibility. In contrast, relative abundance of *D. septosporum* was significantly positively associated with indirect susceptibility in all years compared. We therefore consider the direct susceptibility trait to be a robust evaluation of DNB infection in a given year, but not suitable to infer susceptibility in preceding or subsequent years. In these cases, indirect susceptibility and/or long-term susceptibility traits may be preferable.

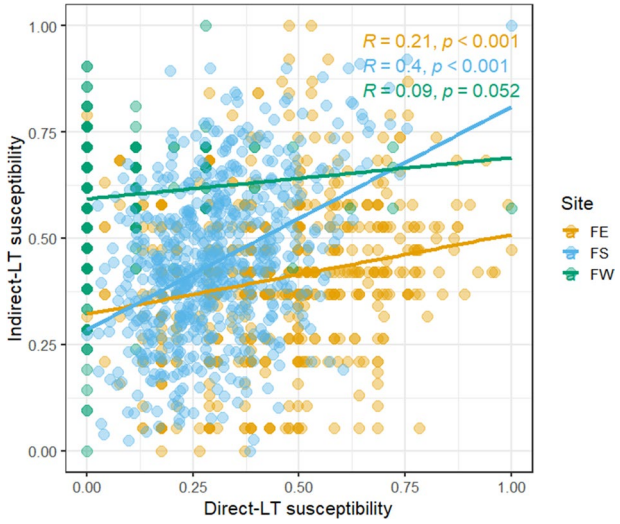


Fig. 2 Pearson's correlation for long-term susceptibility traits (direct-LT susceptibility; indirect-LT susceptibility) quantified for the period 2015–2025 at the FS field site and for 2016 and 2022 at the FE and FW field sites. The correlation coefficient and associated p value are shown for each site.

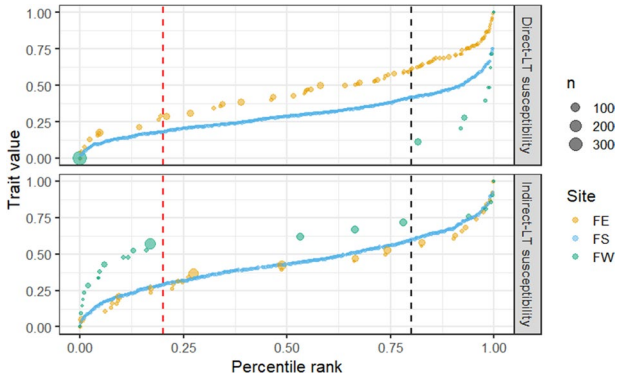


Fig. 3 Discretisation of long-term susceptibility data for each site using percentile ranks. The size of the points is relative to the number of trees (n) represented by each trait value. Vertical dashed lines indicate proposed thresholds to group trees into low susceptibility (trees to the left of the red vertical line) and high susceptibility (trees to the right of the black vertical line) with remaining trees in the centre of the distribution assigned to the moderate susceptibility group.

Finally, we considered the relative abundance of other needle pathogens relative to those assigned to *D. septosporum* to account for the possibility that observed reported symptoms resulted from host susceptibility to other pathogens (Fig. 5). The low relative abundance of other needle pathogens supports the attribution of visual survey of host response to *D. septosporum*.

Data availability

DNB survey data are deposited with the Environmental Information Data Centre (http://eidc.ac.uk): https://doi.org/10.5285/8a3f764c-276e-49e9-9464-36573f48fa62[24]. Modified OTU tables which have been validated for needle pathogens, the fasta sequences for OTUs assigned to needle pathogen genera, and relative abundance of needle pathogens per tree are available on Zenodo: https://doi.org/10.5281/zenodo.21983336[32].

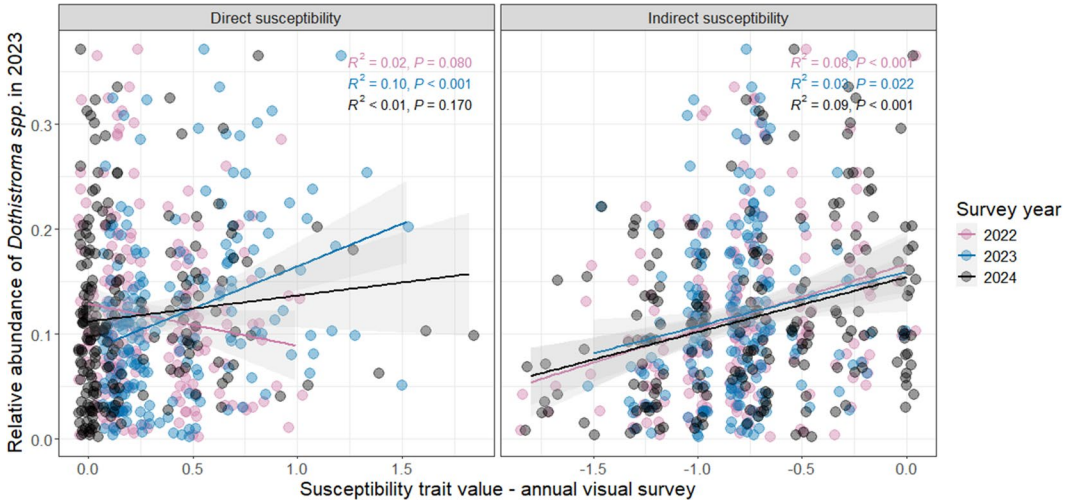


Fig. 4 Linear regressions for in planta relative abundance of D. septosporum against visual survey susceptibility traits (direct and indirect susceptibility). Relative abundance of D. septosporum characterised in 2023. Molecular validation was performed using visual survey susceptibility trait data from the preceding year (2022), the same year (2023) and the following year (2024).

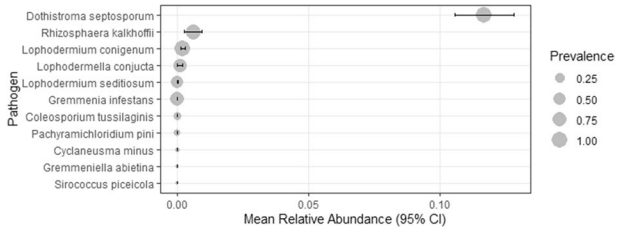


Fig. 5 Comparison of relative abundances of needle pathogens in a subset of trees that underwent metabarcoding from FS trial. Bubble position indicates the mean relative abundance, bubble size represents the prevalence of each needle pathogen taxon among all tested trees (N = 197) and the error bars represent the 95% confidence interval.

Code availability

Code used to verify needle pathogen and for calculating the proportion of reads assigned to needle pathogens from the OTU tables is available on Zenodo: https://doi.org/10.5281/zenodo.21983336[32].

Received: 20 May 2026; Accepted: 28 September 2026;
Published: xx xx xxxx

References

1. Barnes, I., Crous, P. W., Wingfield, B. W. & Wingfield, M. J. Multigene phylogenies reveal that red band needle blight of Pinus is caused by two distinct species of Dothistroma. *D. septosporum and D. pini. Studies in Mycology* **50**, 551–65 (2004).
2. Drenkhan, R. *et al*. Global geographic distribution and host range of Dothistroma species: a comprehensive review. *Forest Pathol* **46**(5), 408–42 (2016).
3. Millberg, H., Hopkins, A. J. M., Boberg, J., Davydenko, K. & Stenlid, J. Disease development of Dothistroma needle blight in seedlings of *Pinus sylvestris* and *Pinus contorta* under Nordic conditions. *Forest Pathol* **46**(5), 515–21 (2016).
4. Lu, M. *et al*. Comparative gene expression analysis reveals mechanism of *Pinus contorta* response to the fungal pathogen *Dothistroma septosporum*. *Mol Plant Microbe Interact* **34**(4), 397–409 (2021).
5. Hocking, D. & Etheridge, D. E. *Dothistroma* needle blight of pines: Effect and etiology. *Annals of Applied Biology* **59**, 133–141 (1967).
6. Capron, A. *et al*. Signatures of Post-Glacial Genetic Isolation and Human-Driven Migration in the Dothistroma Needle Blight Pathogen in Western Canada. *Phytopathology*. **111**(1), 116–27 (2021).
7. Perry, A. *et al*. Substantial heritable variation for susceptibility to *Dothistroma septosporum* within populations of native British Scots pine (*Pinus sylvestris*). *Plant Pathol* **65**(6), 987–96 (2016).
8. Perry, A., Brown, A. V., Cavers, S., Cottrell, J. E. & Ennos, R. A. Has Scots pine (*Pinus sylvestris*) in Scotland? Evidence for spatial heterogeneity in the susceptibility of native provenances. *Evol Appl* **9**(8), 982–93 (2016).
9. Devey, M. E. *et al*. Detection and verification of quantitative trait loci for resistance to Dothistroma needle blight in *Pinus radiata*. *Theor Appl Genet* **108**(6), 1056–63 (2004).
10. Kabir, M. S., Ganley, R. J. & Bradshaw, R. E. Dothistromin toxin is a virulence factor in Dothistroma needle blight of pines. *Plant Pathol* **64**(1), 225–34 (2015).
11. Li, Y., Dungey, H. S., Carson, M. & Carson, S. Genotype by environment interaction for growth and Dothistroma resistance and clonal connectivity between environments in radiata pine in New Zealand and Australia. *PLoS ONE* **13**(10), e0205402 (2018).

12. Ismael, A. *et al*. Indication of quantitative multiple disease resistance to foliar pathogens in *Pinus radiata* D.Don in New Zealand. *Front. Plant Sci.* **11**, 1044 (2020).
13. Wachowiak, W., Perry, A., Donnelly, K. & Cavers, S. Early phenology and growth trait variation in closely related European pine species. *Ecol Evol* **8**(1), 655–66 (2018).
14. Savolainen, O., Bokma, F., Garcia-Gil, R., Komulainen, P. & Repo, T. Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. *Forest Ecology and Management* **197**(1–3), 79–89 (2004).
15. Donnelly, K., Cavers, S., Cottrell, J. E. & Ennos, R. A. Cryptic genetic variation and adaptation to waterlogging in Caledonian Scots pine. *Pinus sylvestris L. Ecol Evol* **8**(17), 8665–75 (2018).
16. Kastally, C. *et al*. Tracing Scots pine expansion in Europe using patterns of rare alleles. *Ecol Evol* **15**(11), e72433 (2025).
17. Bruxaux, J. *et al*. Scots pine - panmixia and the elusive signal of genetic adaptation. *New Phytol* **243**(3), 1231–46 (2024).
18. Beaton, J. *et al*. Phenotypic trait variation in a long-term multisite common garden experiment of Scots pine in Scotland. *Sci Data* **9**(1), 671 (2022).
19. Simko, I. & Piepho, H.-P. The area under the disease progress stairs: calculation, advantage, and application. *Phytopathology.* **102**(4), 381–9 (2012).
20. de Mendiburu, F. agricolae: Statistical Procedures for Agricultural Research. *R package version* **1**, 3–7 (2023).
21. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024 [cited 2024 Aug 20]. Available from: https://www.R-project.org/.
22. Moore B *et al* Metabarcode and transcriptome datasets of *Pinus sylvestris* to assess fungal phyllosphere and disease dynamics. *Sci Data*, https://doi.org/10.1038/s41597-026-08035-z (2026).
23. ENA European Nucleotide Archive https://identifiers.org/ena.embl:PRJEB88228 (2026).
24. Estaki, M. *et al*. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr Protoc Bioinformatics* **70**(1), e100 (2020).
25. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**(2), e107 (2023).
26. Callahan, B. J. *et al*. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**(7), 581–3 (2016).
27. Abarenkov, K. *et al*. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res* **52**(D1), D791–7 (2024).
28. Moore, B. *et al*. PineBiomeDataPaper. *Zenodo.* https://doi.org/10.5281/zenodo.21476744 (2026).
29. Perry A *et al* Long-term multisite Scots pine trial, Scotland: annual Dothistroma needle blight surveys, 2015-2025 v2. NERC EDS Environmental Information Data Centre, https://doi.org/10.5285/8a3f764c-276e-49e9-9464-36573f48fa62 (2026).
30. Perry A *et al* Long-term multisite Scots pine trial, Scotland: field phenotypes, 2013-2020, version 2. *NERC EDS Environmental Information Data Centre*, https://doi.org/10.5285/1c9367fb-ea87-47a1-8257-d9fed54215e7 (2022).
31. Perry A *et al* Long-term multisite Scots pine trial, Scotland: nursery phenotypes, 2007–2011. *NERC EDS Environmental Information Data Centre*, https://doi.org/10.5285/29ced467-8e03-4132-83b9-dc2aa50537cd (2022).
32. Moore, B. *et al*. DNB_Survey_Paper_OTU_processing. *Zenodo.* https://doi.org/10.5281/zenodo.21983336 (2026).

Acknowledgements

We thank Glenn Iason for the original pine trial establishment concept, planning and establishment; Carolyn Riddell, Pete Hedley and Jenny Morris for metabarcoding-related lab work and discussions; Kevin Donnelly, Juan Pablo Lobo-Guerrero Villegas, Luisa Dickenmann, Martin Mullett and Richard Baden for fieldwork.

Author contributions

For the DNB surveys: AP designed, conducted and processed survey results. For the multisite trial: AP, SC, JB, JS and JC contributed to the multi-site common garden design and management. For the molecular validation: AP, BM, KN and HB conducted sample collection and preparation; SK, BC and AG created metabarcoding libraries; BM and SJ processed metabarcoding data. AP, SC and BM wrote the manuscript and all co-authors reviewed and contributed to the final version.

Funding

This work was supported by: the newLEAF project (grant number NE/V019813/1) under the 'Future of UK Treescapes' programme, which was led by UKRI's Natural Environment Research Council (NERC), with joint funding from the Arts and Humanities Research Council (AHRC) and the Economic and Social Research Council (ESRC), and contributions from the UK Government's Department for Environment, Food & Rural Affairs (DEFRA) and the Welsh and Scottish Governments; by the PineBiome project (grant number BB/W020378/1) funded by a grant from UKRI's Biotechnology and Biological Sciences Research Council (BBSRC); project CFP2206 funded by the Department for Environment, Food & Rural Affairs (DEFRA) through the Centre for Forest Protection (CFP); JB and JS's contributions were funded by Scottish Government's Rural & Environment Science & Analytical Services Division, through their Strategic Research Programmes (2006–2011, 2011 2016 and 2016–2022).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.