

## Article

# Evidence for the widespread occurrence of bacteria implicated in acute oak decline from incidental genetic sampling

Louise A. P. Gathercole <sup>1,2</sup>, Gabriele Nocchi <sup>1,2</sup>, Nathan Brown <sup>3</sup>, Timothy Coker <sup>2</sup>, William J. Plumb <sup>1,2</sup>, Jonathan J. Stocks <sup>2</sup>, Richard A. Nichols <sup>1</sup>, Sandra Denman <sup>3</sup>, and Richard J. A. Buggs <sup>1,2</sup>

<sup>1</sup> Queen Mary University of London, Mile End Road, London E1 4NS, UK

<sup>2</sup> Royal Botanic Gardens, Kew, Richmond TW9 3AE, UK

<sup>3</sup> Forest Research, Centre for Ecology, biosecurity and Society, Farnham, Surrey GU10 4LH, UK

\* Correspondence: r.buggs@kew.org; Tel. +44(0)2083325755

Acute Oak Decline (AOD) is complex syndrome affecting Britain's keystone native oak species, (*Quercus robur* L. and *Q. petraea* L. (Matt.) Liebl.), in some cases causing mortality within five years of symptom development. The most distinguishable symptom is weeping stem lesions, from which four species of bacteria have been isolated: *Brenneria goodwinii*, *Gibbsiella quercinecans*, *Lonsdalea britannica* and *Rahnella victoriana*. We do not yet know where else these bacteria exist, and little is known about the relationship of the wider oak leaf microbiome (phyllosphere) to acute oak decline. Here we investigate whether incidental evidence from a large oak genome re-sequencing dataset could be used to detect these bacteria in oak foliage, and whether bacterial incidence co-varied with AOD status or location. Oak leaves and buds were sampled from 421 trees at five sites in England. Whole genomic DNA from these samples was shot-gun sequenced with short reads. Non-oak reads were extracted from these data and queried to microbial databases. Reads uniquely matching AOD-associated bacterial genomes were found to be present on trees from all five sites and included trees with active lesions, trees with historic lesions and trees without AOD symptoms. The abundance of the AOD-associated bacteria did not differ between tree health categories but did differ among sites. We conclude that the AOD-associated bacteria may be members of the normal oak microbiome, whose presence on a tree is not sufficient to cause AOD symptoms.

**Keywords:** *Quercus*, acute oak decline, phyllosphere, *Brenneria goodwinii*, *Gibbsiella quercinecans*, *Lonsdalea britannica*, *Rahnella victoriana*

## 1. Introduction

The native oak trees, *Quercus robur* and *Q. petraea* are the most common native broad-leaf species in Britain [1]. They have historical, cultural, and economic importance and are ecological keystone species for other organisms [1-3]. A recent study found 2300 macro-species of arthropods, fungi and lichens, mosses and liverworts, birds and mammals associated with British native oaks, of which 326 were obligate (found only on *Q. robur/petraea*) [4]. Oak in the UK is currently experiencing damage or elevated mortality from oak processionary moth (*Thaumetopoea processionea*), powdery mildew (*Erysiphe alphitoides*) [5, 6], and complex multi-factorial decline disorders [7]. Decline disorders are complex interactions of multiple factors such as long and short term adverse environmental conditions and the actions of secondary pests and diseases [8]. With no clear universal key factors, Manion defined tree decline disorders as "an interaction of interchangeable, specifically ordered abiotic and biotic factors to produce a gradual general deterioration, often ending in death of trees" [9]. In the UK, two separate decline disorders have been described: Chronic Oak Decline (COD) and Acute Oak Decline (AOD) [10]. Chronic Oak Decline is a

slow decline syndrome thought to be related to root health with root rot fungi as a key biotic factor [10].

Acute Oak Decline has been described as a distinctive decline condition that can rapidly kill trees [7, 11]. The key observable symptoms of AOD are cracks in the bark plate and stem bleeds, together with signs of the presence of the beetle *Agrilus biguttatus* (larval galleries, and adult exit holes in the bark). It is a complex episodic decline syndrome associated with several predisposing environmental factors. Affected sites are more likely to be found in warmer areas of Britain and areas of low rainfall [11, 12]. Soil pH, and the rhizosphere microbiome have been found to differ between symptomatic and asymptomatic trees at some sites [13, 14]. A study that included Hatchlands Park and Sheen Wood (both of which are studied here, see below) found that the bacterial composition of the rhizosphere differed between sites and tree health conditions [13]. Observations at AOD affected sites show a fluctuating pattern of visible symptoms, which may relate to weather patterns. In many affected trees, callus tissue overgrows lesions and the trees enter remission, where external stem bleeds are no longer visible. However, some trees die within 3 to 5 years of the first infection [11, 12].

The biotic cause of the lesions characteristic of Acute Oak Decline is a subject of ongoing investigation. A complex polyspecies microbiome has been found in necrotic lesion tissue. Three bacterial species, *G. quercinecans*, *B. goodwinii* and *R. victoriana*, have been consistently found, with *L. britannica* sometimes present. [15-23]. Comparisons of the genomes of these species with orthologous genes in canonical phytopathogens and non-pathogenic symbionts suggest that *B. goodwinii* and *L. britannica* have the genomic potential to cause tissue necrosis, with *B. goodwinii* the most likely to be the key causal necrogenic phytopathogen in AOD [16]. The role of the apparently less pathogenic *G. quercinecans* may be to release necrotizing enzymes that contribute to tissue necrosis or to help the more pathogenic *B. goodwinii* to activate [16].

Below the bark, lesions are associated with the galleries of larvae of a wood-boring beetle, the two-spotted buprestid *Agrilus biguttatus*, which may play a role in tree death through girdling and disrupting vascular flow [24-26]. Co-inoculations of logs with *A. biguttatus* eggs and Acute Oak Decline associated bacteria found that the necrogenic bacterium *B. goodwinii* appears to be triggered by the presence of the larvae, with upregulation of damaging genes increased 10-fold [27]. Whilst colonisation by *Agrilus* species, can occur without AOD symptoms, the lesion symptoms of AOD are not found outside of the range of *A. biguttatus*, of which Britain is the northernmost limit [25].

A significant difference in microbial composition has been found in the cultivable bacteria isolated from diseased stem tissue compared with healthy stem tissue. Diseased tissue yielded a greater abundance of bacteria in general, and higher *G. quercinecans* and *B. goodwinii* abundance in particular [28]. Microbial isolation and cultivation from stem tissue from both healthy and symptomatic trees at Acute Oak Decline affected and unaffected sites, with identification of bacteria using DNA amplicon sequencing, showed that *G. quercinecans* and *B. goodwinii* were significantly associated with each other, but were not found in AOD free sites and were low in abundance in the stems of healthy trees at AOD sites. *Rahnella* species were identified at AOD sites, but *R. victoriana* was found only in diseased trees and not at all sites [15], although it has previously been found in both healthy and diseased trees [28]. *Lonsdalea britannica* was not widespread [15]. *Gibbsiella quercinecans* and *B. goodwinii* have recently been isolated from stem bleeds in *Q. robur* in Spain [29], whilst *Brenneria* and *Rahnella* species have been isolated from AOD-like stem bleeds in *Q. castaneifolia*, *Q. brantii* and *Carpinus betulus* in Iran [30] suggesting that these genera may be involved in similar decline syndromes in other tree species in different parts of the world.

At present little is known about the wider occurrence of the four Acute Oak Decline-associated bacteria outside of AOD lesions. Studies to date have focused on cultures from oak stems [15, 28, 31]. However, little data is available about other parts of the oak phyllosphere: the above-ground parts of the plant that can be colonised by micro-organisms. Leaf surfaces form the largest component of the phyllosphere, providing a highly variable

environment where bacteria are estimated to be the most abundant colonisers [32-34]. In deciduous plants like *Q. robur* and *Q. petraea*, leaves are a temporary habitat, available for only part of the year. However, when leaves fall and decompose there is an opportunity for microbes to enter the soil and possibly be taken up via the roots and deposited in the stem or on new flushing leaves. Laboratory experiments assessing the ability of *B. goodwinii* and *G. quercinecans* to survive in water and soil showed that *G. quercinecans*, although not culturable from soil, was detectable using qPCR up to 84 days after inoculation in both rainwater and soil. *Brenneria goodwinii* was not able to survive in either environment suggesting that it is an endophytic or epiphytic plant pathogen, whereas *G. quercinecans* is more generalist and able to survive in environmental reservoirs [35]. Recent unpublished research into potential transmission mechanisms of AOD found *B. goodwinii* and *G. quercinecans* on samples of leaves of both healthy and symptomatic trees and at AOD affected and unaffected sites (Emma Bonham, personal communication 20<sup>th</sup> June 2019). It is therefore possible that these species are usually present on the leaves of the oak and other forest trees. In contrast, Meaden et al [31] found no evidence of *B. goodwinii* in healthy trees in a woodland with no AOD symptoms.

Metagenomics can be applied to analyse microbiomes from high-throughput sequencing data. Originally proposed as a way of capturing the unculturable genomes and chemistry of microbes in soil samples [36], metagenomics has increasingly been used as a culture-independent method to characterise the microbial communities of a range of environments [37-39]. A marker gene approach, such as that used in the analysis of the oak rhizosphere at sites with Acute Oak Decline, is one method [13]. An alternative approach analyses shotgun sequence data from whole sample DNA, either assembling the raw reads into contigs for functional gene annotation and taxonomic profiling or, by taking an assembly free approach, using raw reads for reference-based classification. Metagenomic classification software exist to rapidly provide taxonomic identification of raw reads using a database of microbial reference genomes. The recent publication of whole genome assemblies of the four bacteria implicated in AOD [16] means that they can be detected using Kraken2, an ultrafast classification tool designed to assign taxonomic labels to sequences of DNA using exact alignments [40].

Here, we analyse a large short-read dataset based on high-throughput sequencing of DNA extracted from oak leaves and buds from five sites, including Acute Oak Decline symptomatic, asymptomatic and 'in remission' native oaks. This dataset was generated for the study of oak genomic variation [41], but here we examine the incidental evidence it may provide about the phyllosphere of the sequenced trees. We aim to discover whether the four AOD-associated bacteria, previously identified in stem bleeds, could be identified in these foliar data using metagenomic classification of reads that do not map to the oak genome. We wish to understand whether read numbers suggest that these bacteria differ in presence or abundance between affected and unaffected trees, or between geographical locations. We also aim to identify the composition of the wider microbiome of the oak phyllosphere and to ascertain whether there are differences between geographical locations and between health status groups.

## 2. Materials and Methods

This study makes use of data initially collected for an investigation of oak tree genomes in relation to Acute Oak Decline and a wider population study of oak trees at four sites where research and monitoring of AOD was already taking place [41]. These sites are Attingham Park in Shropshire, Langdale Wood in Herefordshire, Hatchlands Park in Surrey and Sheen Wood in Richmond Park, Greater London. In addition, samples and limited data were available from a set of trees at Chestnuts Wood in the Forest of Dean, Gloucestershire. Samples from Chestnuts Wood were either healthy or showing signs of Chronic Oak Decline. Although AOD is present in parts of the Forest of Dean, none of the sampled trees from Chestnuts Wood had AOD symptoms and these symptoms are not present within 50m of the sampled area. Table 1 summarises the features of each site, and the

sample sizes of trees in each health category, based on annual Forest Research and Woodland Heritage monitoring data. Some of the trees at Attingham Park were planted in the 1980s and therefore are less likely to be affected by AOD which mainly affects mature trees [10].

*Table 1: Summary of sites and numbers of trees in each health category at each site. n=421. In addition, 15 technical replicates were sequenced. These were used in the PCA analysis but not in any subsequent analysis*

			Health groups sampled at each site (Asymptomatic <sup>1</sup> , Remission <sup>2</sup> , Acute Oak Decline <sup>3</sup> or Chronic Oak Decline <sup>4</sup> Symptoms) - number of trees in each health category				
	Features	AOD first recorded [42]	A <sup>1</sup>	R <sup>2</sup>	AOD <sup>3</sup>	COD <sup>4</sup>	Total
<b>Attingham Park</b>	Small area of closely planted oak dominated amenity woodland within parkland. Area of new planting with trees approximately 30 years old	2007	58	12	15	0	85
<b>Hatchlands Park</b>	Landscaped parkland with shelterbelts	After 2001	37	26	19	0	82
<b>Langdale Wood</b>	Managed open forest, not closely planted, mowed grass below	Before 2006	77	30	20	0	127
<b>Sheen Wood</b>	Oak dominated woodland in a Royal Park, trees not closely planted. Location as a boundary belt on the edge of the park in an urban area	1991	17	60	25	0	102
<b>Chestnuts Wood</b>	Area of plantation 50-60 years old in the Forest of Dean. Close to natural woodland. COD present. No AOD in the woodland sampled, but AOD is present in the wider area	N/A	14	0	0	11	25
<b>Total</b>			203	128	79	11	421

Samples of small branches containing leaves and buds were collected by the Forest Research Technical Services Unit and placed into plastic bags, which were delivered the same day, or the day after, to RBG Kew. Sampling took place from 1/11/2017 to 14/11/2017. At RBG Kew, samples were placed in a cold room upon receipt. A small amount of leaf and/or bud material from each sample was placed into a small zip-lock bag with silica gel. Another portion of each sample was pressed in a plant press and dried using blotting paper. Another portion was placed in a -80 °C freezer.

Health status, categorised as Acute Oak Decline or Chronic Oak Decline symptomatic, in remission or asymptomatic was recorded at the time of sampling. Trees were also allocated an age category of semi-mature, mature or over-mature based on a combination of a guide diameter, physiological stages based on a model by Raimbault for early tree development [43] and whether the tree is perceived to be in the first, second or third stage of life.

The sampling for this study was not originally designed for a metagenomic study and therefore did not follow all the best practice protocols for microbiome and metagenomic studies as outlined by Knight, Vrbanac et al. [44]. Specifically, the samples were taken from two tissue types and were dried using two different methods. Samples were not collected and processed in the sterile conditions normally associated with studies of microbiomes, but there was little opportunity for cross contamination from other samples as each sample was cut at the stem and bagged immediately. Detailed information was not collected about the exact position in the canopy or the part of the leaf from which each sample was taken, both of which are associated with variability in microbial community abundance and composition [34, 45]. However, since samples were collected by hand or long-arm pruner the maximum height is likely to be 3m from the ground.

Whole genomic DNA was extracted in most cases from leaf or bud tissues that had been dried using silica gel, and in a few cases from bud material dried in a plant press. Extractions were done using the Qiagen DNeasy protocol and whole genomic DNA was sent to Novogene (Hong Kong) for library preparation and whole-genome shotgun sequencing. A total of 436 DNA samples with full phenotypic information were successfully sequenced, including 15 technical replicates from separate DNA extractions. Of these 330 were from leaf tissue dried on silica gel, 21 from bud tissue dried on silica gel, and 78 from bud tissues dried in plant presses. Additionally, 4 samples were extracted from both bud and leaf tissues dried in silica gel and 3 samples were a combination of press dried bud and silica gel dried leaf. Of the successfully sequenced individual trees, 351 were *Q. robur*, 10 were *Q. petraea* and 16 were hybrids, according to genome-based analyses. A further 45 trees not included in Nocchi et al 2021 [41] were identified by morphology, 43 as *Q. robur* and one as *Q. petraea*, making a total of 394 *Q. robur*, 11 *Q. petraea* and 16 hybrids.

Sequencing libraries were prepared with NEBNext DNA Library Prep Kit. Whole genome short read sequencing data was generated with the Illumina Novaseq 6000 platform with read lengths of 2x150bp to produce 16.5 Gbp of data per sample. All subsequent analysis utilised Queen Mary University of London's Apocrita HPC facility, supported by QMUL Research-IT [46]. The WGS data was cleaned and aligned to the reference *Q. robur* genome V2.3 [47] using the bwa mem algorithm from bwa version 0.7.15 with the -M option [48]. Reads that did not align to the reference genome were extracted from the bwa output with samtools v1.9 using the fastq function with flag (-f) 12 [49] and written to separate files. To remove potential human tissue contamination, these non-oak reads were aligned to the human genome build37 using bowtie2 v2.3.4 with the default parameters [50]. The non-human reads were separated and returned as paired read files using samtools v1.9 and bedtools v2.28 with default parameters. Raw reads for each sample are available in the European Nucleotide Archive, accession PRJEB30573.

The non-oak, non-human reads from each sample were taxonomically classified with the bioinformatics software Kraken2, as follows. A custom library was created from all available fungal (1562 genomes), bacterial (56460 genomes), viral (13466 genomes), protozoal (11151 genomes) and archaeal (614 genomes) genomes along with the UniVec\_Core library of common contaminants downloaded from the NCBI refseq database [51] on



9/5/21. Three of the four Acute Oak Decline-associated genomes were present in the library and a *Rahnella victoriana* genome was downloaded separately and added to the library. The library contained a total of 5 *Brenneria* species (*B. goodwinii*, *B. nigrifluens*, *B. rubrifaciens*, *B. sp. IR50* and *B. sp. EniD312*), 2 *Lonsdalea* species (*L. britannica* and *L. populi*), and 5 *Rahnella* species (*R. britannica*, *R. aquatilis*, *R. aceris*, *R. sp. Y9602*, and *R. sp. ERMR1:05*). The only *Gibbsiella* species in the library was *G. quercinecans*. From this library, kraken2 created substrings (or *k-mers*), each 31 bases in length.

This custom Kraken database was used to identify the biotic sources of the reads that had not aligned to the *Q. robur* or *H. sapiens* genomes. Kraken2 queried all the possible 31-mers in our paired reads against the 31-mers in the custom database and matched them to the lowest common ancestor (i.e. the highest taxonomic level within which all lower levels contained that 31-mer). Only exact matches for full 31-mers were classified. Each matched read was then allocated an appropriate taxonomic classification using the set of taxa matched to the 31-mers, and a report produced, showing the number and percentage of reads classified at each clade and taxon level for each sample. Kraken-biom (<https://pypi.org/project/kraken-biom/>) was used to create BIOM files [52] for downstream analysis. One BIOM file included classification of reads to all possible taxonomic levels for each sample. A second BIOM file classified reads only to phylum level. Statistical analysis was carried out using R version 3.5.3, in RStudio [53, 54]. The full package citation is available in Supplementary Materials.

To compare the microbiomes among sampled trees, read counts for each classified taxon at all taxonomic levels were normalised by dividing by the total number of all classified non-oak reads per sample. We ran a Principal Components Analysis (PCA) on these data to identify any potential patterns in variation across the whole microbiome between tissue type, drying method, tree species, sampling site, or tree health status as well as examining the technical replicates.

The software Bracken (Bayesian Reestimation of Abundance after Classification with Kraken) [55] can be used with Kraken classifications to estimate abundance at species level by re-classifying higher level taxonomic reads at species level. A Bracken analysis was run on the Kraken allocations. No additional reads were allocated to the AOD-associated bacterial classifications, so this approach was not pursued further.

Each DNA sample was scored as having presence or absence of the Acute Oak Decline-associated bacterial according to the Kraken results using two thresholds: (1) presence of one or more reads assigned to an AOD-associated bacterium; (2) presence of ten or more reads assigned to an AOD-associated bacterium. These allocations of the DNA samples are referred to hereafter as the one-read and the ten-read scorings respectively. The value of ten, the default cut-off for Bracken re-estimation, is somewhat arbitrary. It eliminates two thirds of the samples, hence the analysis was carried out on both scorings. The percentage of trees with AOD-associated bacteria scored as present at each site and in each health group (AOD Symptomatic, COD Symptomatic, Remission, Asymptomatic) was calculated for each scoring. Chi-squared tests were carried out to assess whether the differences in the percentage of trees with each AOD bacteria present was significant between sites or health categories.

In addition to these analyses of presence / absence we analysed the abundance of each Acute Oak Decline-associated taxon (where present) as follows. A relative abundance score ( $a_{ij}$ ) was obtained from the counts of each AOD-associated taxon ( $x_{ij}$ ) expressed as a proportion of the total number of bacterial reads in sample ( $t_i$ ). A  $\log(\log)$  transformation was applied after inspection of the residuals in a preliminary analysis:  $a_{ij} = \log(\log(x_{ij} / t_i) + 20)$ . For the one-read data set, 305 of the original 421 trees were included in this analysis once Chestnuts Wood and trees with none of the AOD-associated bacteria were removed. For the ten-read data set, 97 trees were included in the analysis. To investigate whether the relative abundance score varied as a function of multiple factors, we ran mixed effects models using the `lme` function from the `nlme` package in R [56]. In all models, the random effect was the sample identifier allowing for sample specific effects on the counts of up to 4 taxa for each sample. The response variable was the proportioned

and log(log) transformed count of reads classified by Kraken as *G. quercinecans*, *B. goodwinii*, *L. britannica* or *R. victoriana*. Fixed effects were bacterial species, site (population) and health status. Trees were categorised as asymptomatic, in remission, or AOD symptomatic – according to the surveyed health status in 2017, which was the year in which the leaves were collected. Previous work has shown an association between tree size as an approximation of age, being a predictor of bacterial community composition [31]. Diameter at breast height (DBH) could be used as a continuous variable to approximate age, but it was not available for all samples. Instead, age was included in the models as a categorical variable with three levels (semi-mature, mature, or over-mature). An initial additive model was run as follows:

*Random effects: individual*

*Abundance\_score ~ (fixed effects): drying\_method + tissue + bacterial\_species + age + health + site*

The initial model was run three times. First with the health status as three separate categories, secondly with health status as two categories - remission trees were included with symptomatic trees, thirdly with remission trees included with asymptomatic trees. There was little difference in outcome, so health was included as a three-category variable for the final model. (See supplementary material for results).

Health status, having been found to be not significant, was removed for a final model. Tissue type and drying method were significant, but closely correlated, therefore tissue type was included in the final model. The final model for the one-read scoring included tissue, age, bacterial species, site and the interaction between bacterial species and site, as follows:

*Random effects: individual*

*Abundance\_score ~ (fixed effects): tissue + bacterial\_species + age + health \*site*

The interaction between species and site was included to investigate whether the combination of a particular bacterial species and a site had an effect on the proportion of reads classified as Acute Oak Decline-associated bacteria. Both of these independent variables had been found to be significant in earlier models.

The missing values for some of the bacteria in the smaller ten-read scoring meant that it was not possible to include the interaction between bacterial species and site, therefore an additive model was run, as follows:

*Random effects: individual*

*Abundance\_score ~ (fixed effects): tissue + bacterial\_species + age + health +site*

In order to further confirm the classifications made by Kraken of some reads to the four Acute Oak Decline-associated bacteria on the basis of 31-mers, we took the full-length reads and queried them against a custom BLAST database of the four AOD-associated genomes using magic-BLAST [57]. We calculated the percentage of these reads, classified by Kraken as being from AOD-related bacteria, that had a 95% or greater alignment with the genome of the species that Kraken had classified them to.

### 3. Results

Within each of the short-read sequence datasets for each oak tree, a mean of 96.44% (standard deviation = 4.4) of reads mapped to the *Q. robur* genome. A further 0.22% (standard deviation = 1.13) of reads mapped to the human genome. This left a mean of 3.33% (standard deviation = 3.97) of non-oak, non-human reads. Of these reads, the mean percentage of reads classified by Kraken was 6.05% (range = 0.49% to 51.14%, sd = 5.25): thus between 2724 and 6626088 reads were classified per sample. Of these, across all samples, 44.55% were classified as Proteobacteria, 33.9% as Actinobacteria, 12.13% as Bacteroidetes, 6.2% as Ascomycota and 1.28% as Firmicutes. The relative abundance of reads classified as fungi differed among sites (Figure 1). Notably, ascomycetes were unusually common in the samples from Hatchlands Park and amongst the bacteria, proteobacteria was unusually common at Chestnuts Wood. Basidiomycota only accounted for 1% of the classified reads across all reads. These were in greater abundance at Hatchlands Park (2.4%) where the predominant environment is wood pasture. There was a relatively low percentage (0.1%) of the reads classified as Basidiomycota at Chestnuts Wood, which is a more closely planted woodland with COD, often associated with the Basidiomycota genus *Armillaria* causing problems with the tree roots. These seems not to have translated into the presence of basidiomycota on the leaves. Table 2 lists the most commonly identified phyla.

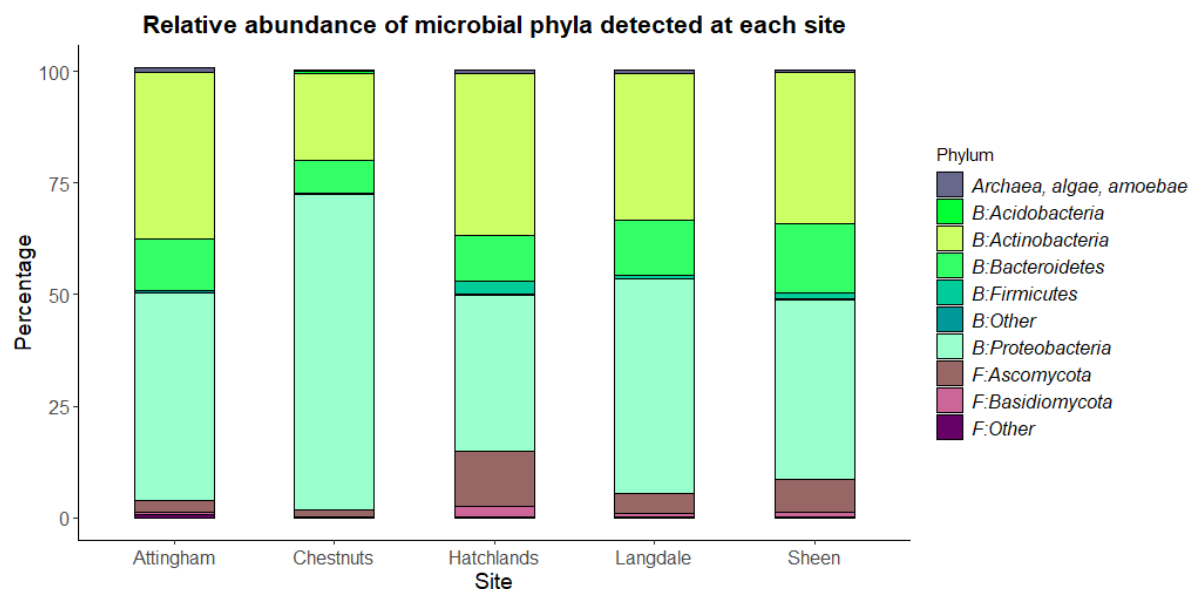


Figure 1: The relative abundance of microbial phyla detected by Kraken classification software at each site. Fungi (F) are at the bottom of the bars and bacterial phyla (B) at the top. Archaea and other eukaryotic microbes comprise less than 0.05% of the total classified. The most abundant phylum is the proteobacteria to which the AOD-associated bacteria belong.



Table 2: Most numerous classifications at phylum level

Phylum	% reads
Proteobacteria	44.55
Actinobacteria	33.9
Bacteroidetes	12.13
Ascomycota	6.2
Firmicutes	1.28

A Principal Components Analysis based on all samples (including technical replicates, but excluding two extreme outliers) and all classified microbiome read counts across all taxonomic levels is shown in Figure 2. Each panel of this figure shows PC1 and PC2 for the same PCA, but with different labels. The first two Principal Components explained just 3.2% of the variability due to the large number of classified taxa. In Figure 2A the technical replicates are labelled. Most technical replicates appeared close to each other with the exception of two of the Sheen Wood trees. In Figure 2B, the different species of oak are labelled; showing little clustering due to oak species, whether identified morphologically or genetically. In Figures 3C and 3D, the sample drying method and tissue source are labelled. Samples dried in the plant press and those dried in silica gel fall separately, the pattern being very similar to that of the tissue source. Most bud samples were dried in the press and most leaf samples in silica gel. In Figure 2E, the health status of the trees is labelled. There is no clustering of trees with different Acute Oak Decline health statuses, although those in remission have a wider scatter. In Figure 2F, the sample sites are labelled. There is no apparent clustering by site, although Sheen Wood has a wider scatter and seems to be confounded with the wider scatter of remission trees. Chronic Oak Decline is confounded with site as it was only found in Chestnuts Wood. Comparison of Figures 3C,3D and 3F shows that the effects of site, tissue source and drying method are largely confounded. Principal Components 3 and 4 showed even less clustering (not shown).

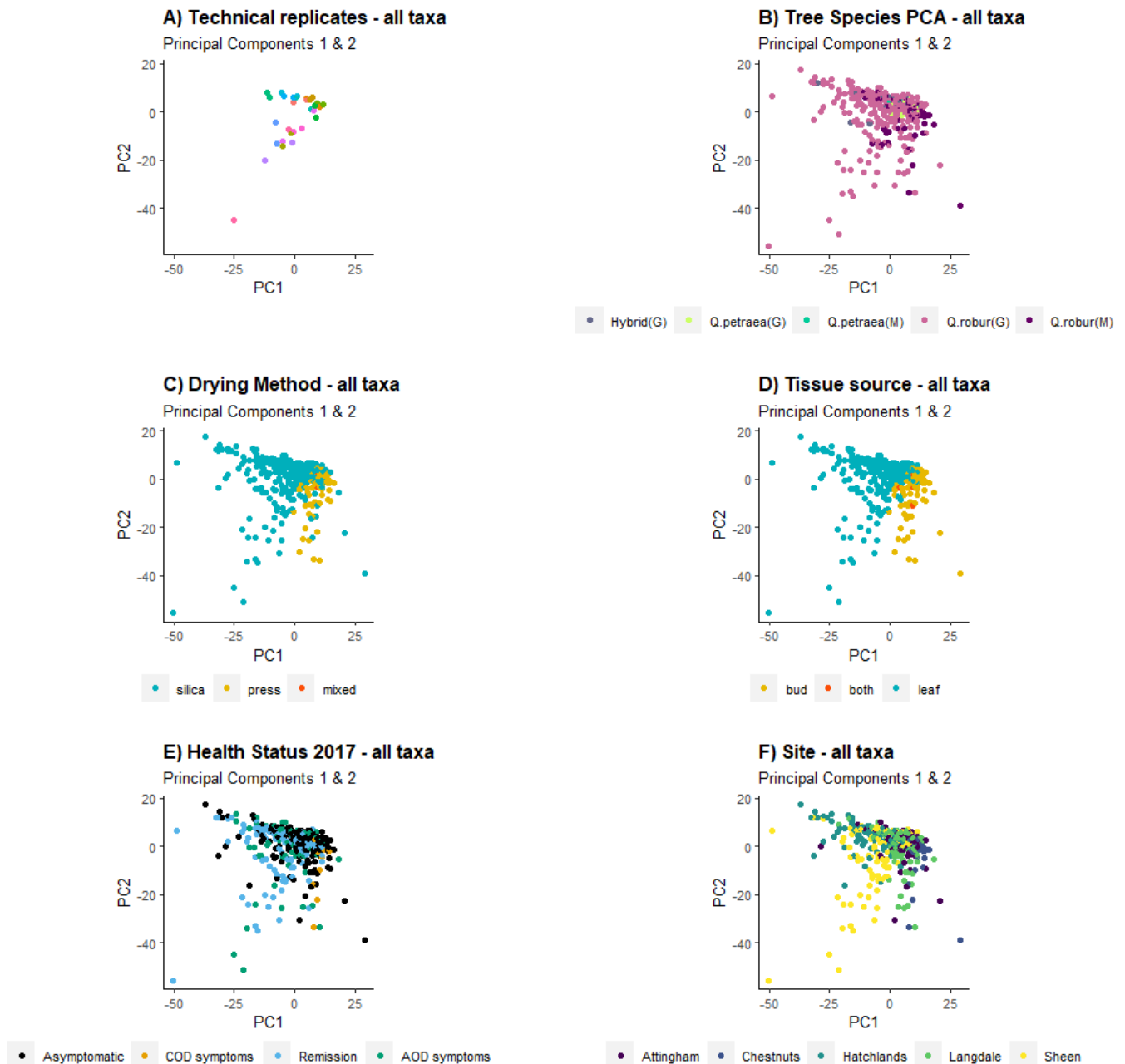


Figure 2 (previous page): A PCA based on all samples (including technical replicates, but excluding two extreme outliers) and all classified microbiome read counts across all taxonomic level. Each panel shows PC1 (explaining % variability) and PC2 (explaining % variability) for the same PCA, but with different labels. A) technical replicates. Most fall close to each other with the exception of two trees from Sheen Wood. B) oak tree species; showing little clustering. (Species are categorised as Hybrid, Q. robur or Q. petraea (according to genome-based analysis, or morphologically categorised as Q. robur or Q. petraea). C) the sample drying method. Samples dried in the plant press and those dried on silica gel fall on different sides of the plot, but are very similar to D) tissue source. Most bud samples were dried in the press and leaf samples in silica gel. The differences in microbial classification may be due to differences between the bud and open leaf, or be from differences in drying method]. E) Health Status: There is no clustering of trees by health status, although those in remission have a wider scatter. F) Sample site: There is no clustering by site, although Sheen Woods has a wider scatter which seems to be confounded with a wider scatter of remission trees. COD health status is confounded with site as COD was only found in Chestnuts Wood. Another PCA where reads were classified at the phylum level showed similar patterns of clustering and PC 3 and 4 showed even less clustering (not shown).

Low numbers of reads classified by Kraken as *Brenneria goodwinii*, *Gibbsiella quercinecans*, *Lonsdalea britannica* and *Rahnella victoriana* were found from both leaf and bud samples at all sites and across all health categories. Total reads for all four taxa per tree (including technical replicates) ranged from 0 to 1,607,824 with a mean of 3955 and a median of 3 classified reads. There were two very high outliers, and apart from these the highest count was 7961. Only 35 samples in total had 100 reads or more.

In the one-read scoring, *B. goodwinii* was detected in 42 trees (10%), *G. quercinecans* in 132 (31%), *L. britannica* in 195 (46%) and *R. victoriana* in 237 (56%). In the more stringent ten-read scoring *B. goodwinii* was detected in just 5 trees (1%), *G. quercinecans* in 12 (3%), *L. britannica* in 48 (11%) and *R. victoriana* in 66 (16%).

Figure 3 shows the proportion of trees at each site where each species of bacteria was detected in each scoring. It is notable that reads from all four Acute Oak Decline-associated bacteria were found at Chestnuts Wood where Acute Oak Decline is not present, but Chronic Oak Decline is. The lowest incidence of the four bacteria was at Hatchlands, where AOD is present. Figure 4 shows the proportion of samples in the four sites (grouped together) where AOD is present. The bars show the proportion of trees in the different health statuses where AOD-bacterial reads were found in their samples. This shows no noticeable difference between symptomatic, asymptomatic and remission trees.

We carried out chi-squared tests on the differences in the proportion of trees in which each bacterial species was detected across the four sites with Acute Oak Decline. These found significant differences across sites for all of the bacteria except *B. goodwinii*, which was detected on the lowest number of trees. Hatchlands Park had few trees with more than 10 reads classified as each AOD bacteria. Removing Hatchlands reduced the differences between sites, but this still remained significant (Table 3).

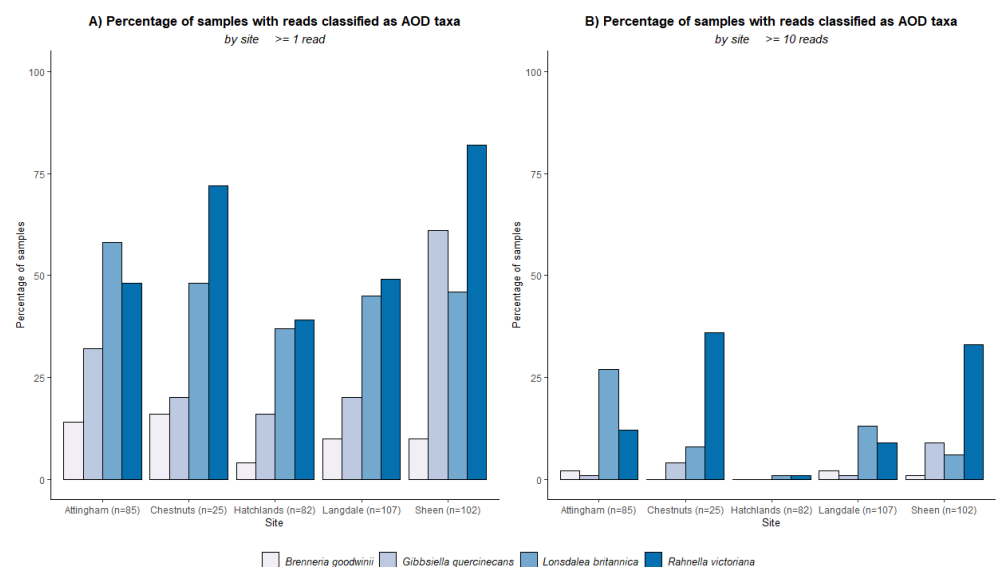


Figure 3: Percentage of samples with AOD-associated bacteria present at each of the sites. 4A shows the percentage of trees where 1 or more reads were classified by Kraken2. 4B shows the percentage where at least 10 reads were classified. Hatchlands Park has very few trees in which more than 10 reads were classified as AOD-associated bacterial genomes. Reads that matched AOD genomes were found at all sites, including Chestnuts Wood which did not have any AOD symptomatic trees

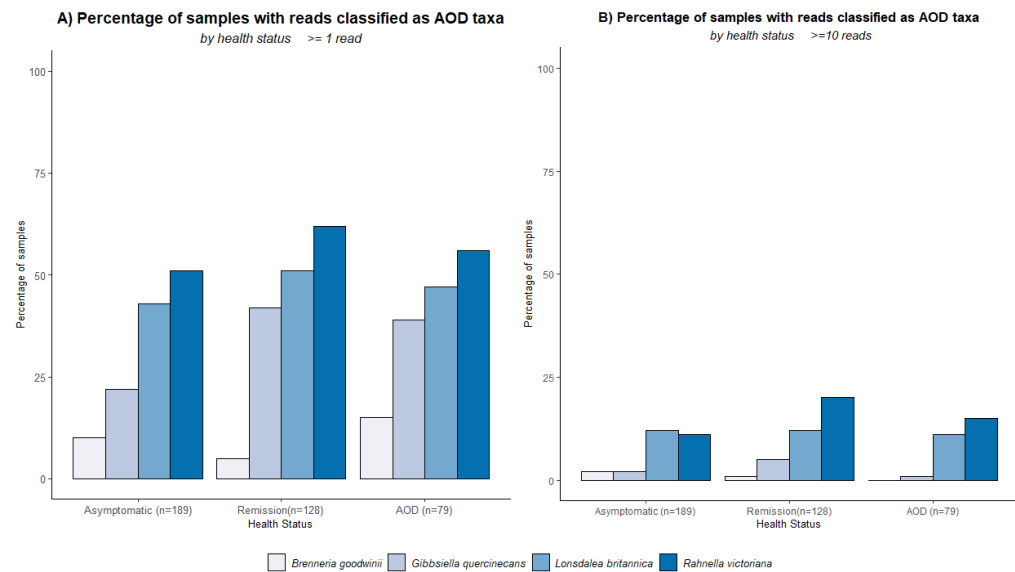


Figure 4: : Percentage of samples with Acute Oak Decline-associated bacteria present in each health group across all sites with AOD-symptomatic trees. 4A shows the percentage of trees where 1 or more reads were classified by Kraken2. 4B shows the percentage where at least 10 reads were classified. Reads that matched AOD genomes were found on Asymptomatic, AOD symptomatic trees and trees in remission. *Brenneria goodwinii*, the putative necrotic agent in AOD stem lesions was detected on the leaves of very few trees in any health group.

Table 3: Pearson's chi-squared test of differences between the proportion of trees showing presence of each bacteria across the four sites with Acute Oak Decline symptomatic trees. Presence is considered to be one or more classified reads or 10 or more classified reads. There are significant differences across all bacterial species except *B. goodwinii*, which was detected on a low number of trees. Hatchlands Park had few trees with more than 10 reads classified as each AOD bacteria. Removing Hatchlands reduced the differences between sites, but results still remained significant. Chestnuts Wood was not included in this analysis. Statistically significant results are highlighted in **bold**.

	Presence = one-read scoring				Presence = 10-read scoring			
	Pearson's chi-squared test with Hatchlands (df=3)		Pearson's chi-squared test with Hatchlands removed (df=2)		Pearson's chi-squared test with Hatchlands (df=3)		Pearson's chi-squared test with Hatchlands removed (df=2)	
	$\chi^2$	p-value	$\chi^2$	p-value	$\chi^2$	p-value	$\chi^2$	p-value
<i>G. quercinecans</i>	57.448	<b>&lt;0.001</b>	42.359	<b>&lt;0.001</b>	18.818	<b>&lt;0.001</b>	12.673	<b>0.002</b>
<i>B. goodwinii</i>	5.400	0.145	1.049	0.591	2.024	0.568	0.558	0.757
<i>L. britannica</i>	7.619	0.054	3.745	0.154	31.762	<b>&lt;0.001</b>	17.456	<b>&lt;0.001</b>
<i>R. victoriana</i>	42.861	<b>&lt;0.001</b>	32.608	<b>&lt;0.001</b>	44.24	<b>&lt;0.001</b>	24.952	<b>&lt;0.001</b>

The abundance of the Acute Oak Decline-associated bacteria, measured by the number of reads assigned to the taxon by Kraken, was analysed for any patterns in abundance related to health status, geographical site, tissue and drying method. In an initial additive

mixed effect models on the samples from the four AOD sites we found no significant differences between symptomatic, remission and healthy trees in the proportion of genomic sequences classified as AOD-associated bacteria. This was the case when samples were categorised in three groups (symptomatic, remission and healthy), or when trees in remission were included with healthy trees, and when trees in remission were included with AOD symptomatic trees (see Supplementary Table 1). However, we did find a significant effect by including site in the models of the one-read scoring ( $P=0.024$ ), but not the lower powered comparison for the ten-read scoring ( $P = 0.967$ ) (see Supplementary Table 1).

In a second mixed-effects model, health status was not included, and we included interactions between bacterial species and site. We found a highly significant interaction between bacterial species abundance and site in the one-read scoring (Table 4). Site and age category were both significant even if semi-mature (too young to be affected by Acute Oak Decline) trees were removed from the dataset. Site and age also remained significant if the site Attingham, which has a large number of young trees, was removed from the dataset. In the ten-read scoring, tree age had a significant effect on the read counts but not site. In all analyses tissue type was a significant factor.

*Table 4: A) ANOVA results of mixed effect model of abundance scores for Acute Oak Decline-associated bacteria using one- (A) and ten-read scoring (B). Random effect was the sample number, fixed effects were tissue (leaf or bud), age category (semi-mature, mature, over-mature), bacterial species and site. Model (A) included an interaction between bacterial species and site, which was lacking in (B) due to different numbers of samples in each category. In addition to the significant effect of site on the abundance of reads classified in the (A) (1 or more reads classified), an interaction was found between the abundance of the different bacterial species and the site. Site was not significant in (B) after accounting for the other effects. Statistically significant effects ( $P<0.05$ ) are highlighted in **bold**.*

	Fixed Effect	Numerator DF	Denominator DF	F-value	P-value
A) One-read scoring (n=305)	Tissue	2	297	28.46	<b>&lt;0.001</b>
	Age	2	297	16.95	<b>&lt;0.001</b>
	Bacterial Species	3	250	25.58	<b>&lt;0.001</b>
	Site	3	297	2.98	<b>0.0316</b>
	Bacterial Species : Site	9	250	7.34	<b>&lt;0.001</b>
B) Ten-read scoring (n=97)	Tissue	2	89	10.274	<b>&lt;0.001</b>
	Age	2	89	0.107	0.899
	Bacterial Species	3	19	3.870	<b>0.0257</b>
	Site	3	89	0.835	0.478

Figure 5 shows the abundance score for each bacterial species by health status and by population in both the one- and ten-read scorings. Figures 5A and 5C show the abundance scores for each species in each health category. There are scores for all of the bacteria in all health categories in the one-read scoring, and also in the ten-read scoring, with the exception of *B. goodwinii* which had no samples with 10 or more reads in the Acute Oak Decline symptomatic category. Figures 5B and 5D show the scorings by site, where differences in abundance can be seen. The ten-read scoring has no scores for *B. goodwinii* or *G. quercinecans* at Hatchlands Park. *L. britannica* and *R. victoriana* have greater abundance



scores in the more stringent ten-read scoring than *B. goodwinii* or *G. quercinecans*. In the BLAST searches of the reads identified by Kraken2 as being from AOD-related bacteria, 57% of *B. goodwinii* classified reads, 60% of *G. quercinecans*, 57% of *L. britannica* and 61% of *R. victoriana* matched their respective genomes with 95% or more of the full read aligning to the genome.

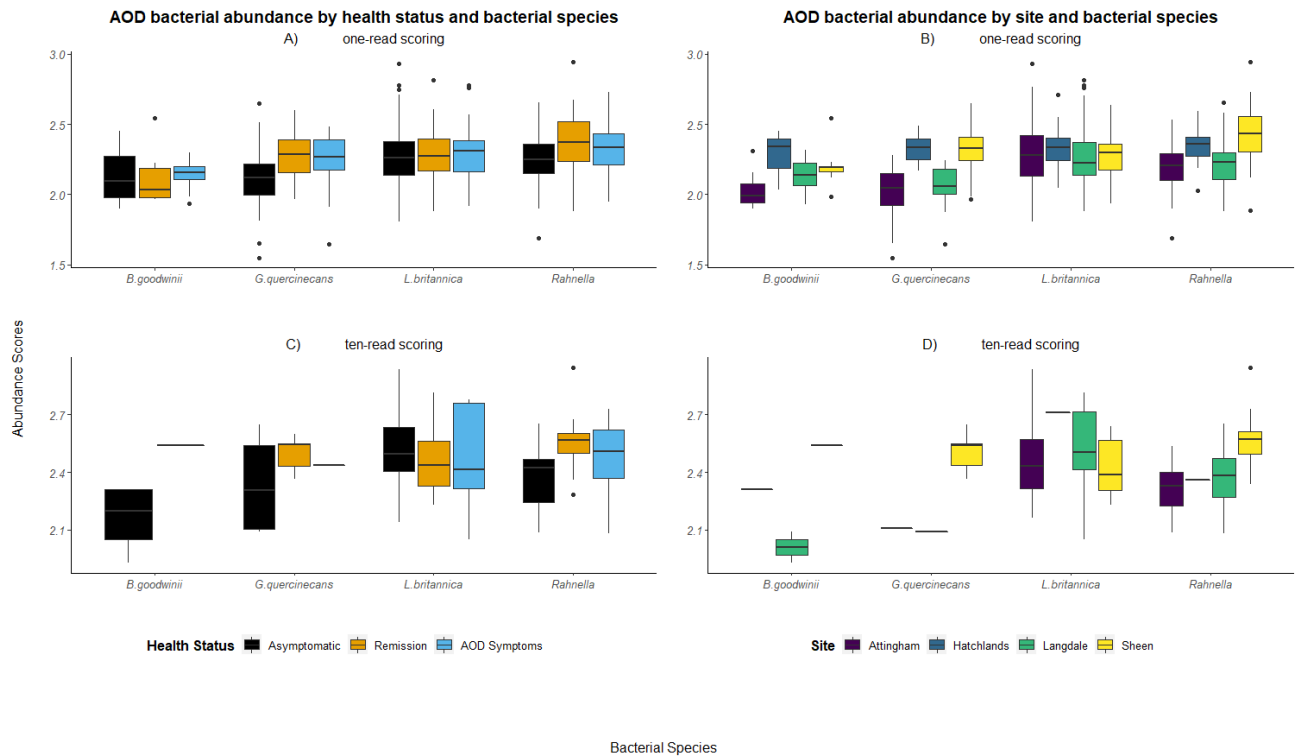


Figure 5: The abundance score of each of the Acute Oak Decline-associated bacteria detected on the samples shown by health status and by population using both one- (figure 5A and 5B) and ten-read (figure 5C and 5D) scorings. The one-read scoring shows all abundance scores where one or more reads were classified as the AOD-associated bacteria by Kraken2. The ten-read scoring required a minimum of 10 reads classified by Kraken2. Figures 5A and 5C show the abundance scores for each species in each health category. There are abundance scores for all bacteria in each health category in the one-read scoring, and also in the ten-read scoring, with the exception of *B. goodwinii* which had no samples with 10 or more reads in the AOD symptomatic category. Figures 5B and 5D show the scorings by site, where differences in abundance can be seen. The ten-read scoring has no scores for *B. goodwinii* or *G. quercinecans* at Hatchlands Park. *L. britannica* and *R. victoriana* have greater abundance scores in the more stringent ten-read scoring than *B. goodwinii* or *G. quercinecans*.

#### 4. Discussion

In this study we were able to use incidental read data and the metagenomic classifier Kraken2, to characterise the foliar microbiome, find evidence for the presence of Acute Oak Decline-associated bacteria across 5 geographical sites and all AOD health categories.

In total, Kraken2 classified reads were assigned to 56 phyla. The bacterial phyla *Proteobacteria* and *Actinobacteria* accounted for over 78% of the read classifications and *Bacteroidetes* a further 12%. Previous 16s RNA taxonomic studies of *Q. robur* metagenomes from the trunk and the rhizosphere (root microbiome) have found that *Acidobacteria* were highly abundant [31, 58]. The present study found that *Acidobacteria* made up only 0.07% of the total reads classified, but this is unsurprising as *Acidobacteria* are abundant in soil and less likely to be dispersed in the phyllosphere. A similar pattern was found in a study of bacterial communities in *Ginkgo biloba* which found that *Acidobacteria* were abundant in trunk tissue, but not in leaf tissue, with *Bacteroidetes* also one of the four most abundant taxon across the different tissues [59]. Therefore, our findings seem to be in line with previous bacterial microbiome studies. Fungi made up only a small percentage of the

classified reads compared with bacteria. The two major fungal phyla, *Ascomycota* and *Basidiomycota* are both represented in our classifications, with *Ascomycota* being the most abundant phylum. This is a similar finding to an ITS1 amplicon sequencing study of several leaf samples from a single *Q. robur*, which found that 50% of the operational taxonomic units (OUT)s were *Ascomycota* and 11% *Basidiomycota* [60] (<https://www.ncbi.nlm.nih.gov/refseq/>). Principle Components Analysis of the full Kraken taxonomic classifications of the phylum level classifications did not identify any patterns in the data according to health status. Samples from different sites do not cluster separately in the PCA plot.

The bacteria associated with Acute Oak Decline had previously been identified from isolations taken from lesions in the stem tissue of affected trees [17, 19-21, 23]. Higher abundances of these bacteria have been found in lesion tissue than unaffected stem tissue from symptomatic trees and than stem tissue from healthy trees [28]. We found evidence of their genome sequences in the leaf and leaf-bud tissues, in low and variable abundances, suggesting that these bacteria are in fact ubiquitous. These results imply that the presence of the bacteria on an individual is not a sufficient cause for Acute Oak Decline symptoms to appear. It is of particular interest that in our more stringent ten-read scoring, there was no evidence for the necrogenic species *B. goodwinii* in symptomatic trees. It may therefore be the case that the proliferation of these bacteria at AOD lesion sites is triggered by abiotic factors and/or boring by the beetle *A. biguttatus*, as has been suggested by Doonan et al [27].

The detection of these bacterial sequence data in DNA extracted from leaves does not distinguish between bacteria living as endophytes, epiphytes or present due to rain or insect dispersal. Pettifor et al [35] found that rainwater, and to a lesser extent, soil, could act as a reservoir for *G. quercinecans* but not for *B. goodwinii*. This evidence indirectly suggests that *B. goodwinii* is an endosymbiont, but *G. quercinecans* may have a different niche [35]. We found *R. victoriana* and *L. britannica* to be more widespread than either *G. quercinecans* or *B. goodwinii*, which could suggest a wider dispersal in the environment for these two species.

Our analysis of the abundance of Acute Oak Decline-associated bacteria showed differences between the geographical sites. This is in line with a study of the oak rhizosphere, which found differences in the bacterial composition of the root microbiome between sites [13]. It is also in line with phyllosphere studies on other tree species, reviewed by Rastogi et al. [61] in which geographical location has been shown to be a driver of bacterial community composition. To our knowledge, this has not previously been shown in the oak phyllosphere.

One site, Hatchlands Park, had a lower proportion of trees where the bacteria were present than the other sites. The ecology of the selected trees at Hatchlands Park is different from the other sites in that they are spaced further apart in wood pasture and hedgerow settings rather than in closer planted woodland. The samples were also taken from a wider geographical area in the site. This may have led to greater differences in soil condition or exposure to rainwater, sunlight, and insect activity. Differences in the microbiome and the abundance of Acute Oak Decline-associated bacteria are likely to be partly due to environmental and climatic factors, but there may also be an effect of genetics if the trees were naturally grown or planted from the same seed source at each plantation site. Genotype has been shown to have an effect on the microbiome in poplar, particularly of fungal foliar communities [62, 63] and may have an impact on the geographical differences in the full microbiome. AOD symptomatic trees at some of these sites have previously been found in physical clusters [42], which could suggest a mechanism passing the bacteria between the trees. However, the presence of AOD-associated bacteria across many trees in the woodland or parkland could suggest that these clusters may be related more to the action of adult *A. biguttatus*, or to soil or other abiotic conditions in specific parts of the site.

The number of reads classified at species level for the Acute Oak Decline-associated bacteria of interest were low. The developers of Kraken tested the original Kraken1

program against a simulated Hi-seq metagenome and found that precision at genus level, defined as the proportion of correctly classified reads from all the classification attempts, was 99.2% [64]. Although this precision is high, and Kraken and Kraken2 have been found by other authors to be precise and accurate and in line with other metagenomic pipelines [65, 66], the low numbers of total AOD-associated bacterial reads classified may put our results within the potential margin of error for false positives. The trends we detected in the species level comparisons should therefore be verified, for example using qPCR studies on the oak phyllosphere. Our classifications at the phylum level analysis are more robust due to the greater likelihood of a correct classification at this level [66] and the higher numbers of reads we classified in each phylum.

What is not clear from this study is whether the AOD-associated bacteria are part of the leaf microbiome only at sites where Acute Oak Decline is present. The bacteria were found in Chestnuts Wood, which is not currently showing any AOD symptoms, although AOD is present in other parts of the Forest of Dean. A microbiome study by Meaden et al. [31] of the *Q. robur* trunk microbiome in 64 trees at Wytham Woods, which has no cases of AOD, found no 16S RNA signals for *Brenneria goodwinii*, but did not look for the other AOD-associated bacteria. It is therefore possible that presence of the AOD-associated bacteria indicates a symptomatic site, rather than symptomatic individuals, and that these bacteria are not a normal part of the oak microbiome outside of these sites. A useful further study would be to collect leaf samples from trees in woodlands with no AOD present and use the recently developed real-time PCR assay to look for AOD-associated bacteria at these sites [67]. In addition, a metagenomic study of oak trees and other species from a variety of sites could shed further light on the range and habitats of these bacteria. The presence of bacteria on the leaves in non-AOD woodlands would confirm that these bacteria are indeed to be part of the normal oak microbiome. An absence could indicate that AOD is a site-wide infection in which trees with predisposing factors become symptomatic. A systematic survey across the *Q. robur* and *Q. petraea* species range in Britain or Europe, including outside of the range of *A. biguttatus*, could shed greater light on whether these species are symbionts of the oak tree.

## 5. Conclusions

In this study, we were able to identify the most abundant fungal and bacterial phyla in the oak phyllosphere, and detect signatures of Acute Oak Decline-associated bacterial genomes over a much larger number of samples than are typically used for metagenomic studies. This was achieved using incidental read data from a whole-genome-sequencing population genomics project on oak trees in England. We found sequence-read evidence for the presence of AOD-associated bacteria on leaf and bud material in four Acute Oak Decline affected sites and additionally at a site where AOD symptoms are not present in the immediate vicinity. We found evidence that the AOD-associated bacteria were present on trees in all health categories. There were no significant differences in either the proportion of trees with AOD-associated bacteria present, or in the abundance of any of the four AOD-associated bacteria between the symptomatic, remission and healthy trees. There were significant differences between geographical sites in both the proportion of trees with the bacteria present and the abundance of bacteria, although the abundance trend could not be demonstrated in the lower powered analysis of the more stringent ten-read scoring. Interestingly, *Brenneria goodwinii*, the bacterial species believed to have a necrogenic affect in AOD lesions was not detected on the leaves of symptomatic trees in the more stringent ten-read scoring, but was found on healthy trees and those in remission suggesting that its presence on the tree is not sufficient for infection to occur.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Table S1: ANOVA results of additive mixed effect models of abundance scores for AOD-associated bacteria with samples in three health categories. Table S2: List of R packages (see also references).

**Author Contributions:**

Conceptualization, LG, RB, RN; methodology, LG, GN, RB, RN.; formal analysis, LG, RN.; investigation, LG NB, SD, TC, JS, WP, Forestry Commission Technical Services Unit; resources, Forest Research, RBG Kew, QMUL ITS Research Team.; data curation, LG, GN, NB; writing—original draft preparation, LG; writing—review and editing, RB, RN, SD, NB, GN.; visualization, LG; supervision, RB,RN; project administration, RB; funding acquisition, Richard Buggs. All authors have read and agreed to the published version of the manuscript.

Louise Gathercole <https://orcid.org/0000-0001-9784-9526>

Gabriele Nocchi <https://orcid.org/0000-0001-7119-5181>

Nathan Brown <https://orcid.org/0000-0003-3590-0538>

Timothy Coker <https://orcid.org/0000-0002-2516-0639>

William Plumb <https://orcid.org/0000-0002-1498-0113>

Jonathan Stocks <https://orcid.org/0000-0003-4958-4580>

Richard Nichols <https://orcid.org/0000-0002-4801-9312>

Sandra Denman <https://orcid.org/0000-0001-6138-3402>

Richard J. A. Buggs <https://orcid.org/0000-0003-4495-3738>

**Funding:** This research was funded by Defra via Action Oak, The Forestry Commission and Woodland Heritage.

**Data Availability Statement:** DNA sequences: trimmed sequencing reads have been deposited in the European Nucleotide Archive under the Project Accession No. PRJEB30573, which can be found at: <https://www.ebi.ac.uk/ena/browser/view/PRJEB30573>.

**Acknowledgments:** Samples were collected by the Forestry Commission Technical Services Unit. Thanks to Raphaella Jackson for helpful advice on microbial analysis and to James Crowe for help with bioinformatics tools. We also acknowledge the assistance of the ITS Research team at Queen Mary University of London.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**References**

1. Quine, C.P., Atkinson, N., Denman, S., Desprez-Loustau, M-L., Jackson, R., Kirby, K. (eds), *Action Oak Knowledge Review: an assessment of the current evidence on oak health in the UK, identification of evidence gaps and prioritisation of research needs*. 2019, Haslemere, UK: Action Oak.
2. Leroy, T., C. Plomion, and A. Kremer, *Oak symbolism in the light of genomics*. New Phytologist, 2019. 0(ja).
3. Rackham, O., *Trees and woodland in the British landscape: the complete history of Britain's trees, woods & hedgerows*. Rev. ed. 2001, London: Phoenix Press.
4. Mitchell, R.J., et al., *Collapsing foundations: The ecology of the British oak, implications of its decline and mitigation options*. Biological Conservation, 2019. **233**: p. 316-327.
5. Tomlinson, I., C. Potter, and H. Bayliss, *Managing tree pests and diseases in urban settings: The case of Oak Processionary Moth in London, 2006–2012*. Urban Forestry & Urban Greening, 2015. **14**(2): p. 286-292.
6. Lonsdale, D., *Review of oak mildew, with particular reference to mature and veteran trees in Britain*. Arboricultural Journal, 2015. **37**(2): p. 61-84.
7. Denman, S., et al., *A description of the symptoms of Acute Oak Decline in Britain and a comparative review on causes of similar disorders on oak in Europe*. Forestry, 2014. **87**(4): p. 535-551.
8. Manion, P.D., Lachance, Denis, *Forest decline concepts*. 1992, St. Paul, Minn.: The American Phytopathological Society.

9. D Manion, P., *Tree disease concepts* / Paul D. Manion. 1981, Englewood Cliffs, NJ: Prentice-Hall.
10. Denman, S. and J. Webber, *Oak declines - new definitions and new episodes in Britain*. Q. J. For., 2009. **103**: p. 285-290.
11. Brown, N., et al., *Integrating regulatory surveys and citizen science to map outbreaks of forest diseases: acute oak decline in England and Wales*. Proc Biol Sci, 2017. **284**(1859).
12. Brown, N., et al., *Predisposition of forests to biotic disturbance: Predicting the distribution of Acute Oak Decline using environmental factors*. Forest Ecology and Management, 2018. **407**: p. 145-154.
13. Pinho, D., et al., *Linking Tree Health, Rhizosphere Physicochemical Properties, and Microbiome in Acute Oak Decline*. Forests, 2020. **11**(11).
14. Scarlett, K., et al., *Relationships between nitrogen cycling microbial community abundance and composition reveal the indirect effect of soil pH on oak decline*. ISME J, 2020.
15. Denman, S., et al., *Microbiome and infectivity studies reveal complex polyspecies tree disease in Acute Oak Decline*. ISME J, 2018. **12**(2): p. 386-399.
16. Doonan, J., et al., *Genomic analysis of bacteria in the Acute Oak Decline pathobiome*. Microb Genom, 2019. **5**(1).
17. Brady, C., et al., *Taxonomy and identification of bacteria associated with acute oak decline*. World J Microbiol Biotechnol, 2017. **33**(7): p. 143.
18. Brady, C., et al., *Proposal to reclassify Brenneria quercina (Hildebrand and Schroth 1967) Hauben et al. 1999 into a new genus, Lonsdalea gen. nov., as Lonsdalea quercina comb. nov., descriptions of Lonsdalea quercina subsp. quercina comb. nov., Lonsdalea quercina subsp. iberica subsp. nov. and Lonsdalea quercina subsp. britannica subsp. nov.* Vol. 62. 2011. 1592-602.
19. Brady, C., et al., *Description of Gibbsiella quercinecans gen. nov., sp. nov., associated with Acute Oak Decline*. Systematic and Applied Microbiology, 2010. **33**(8): p. 444-450.
20. Brady, C., et al., *Rahnella victoriana sp. nov., Rahnella bruchi sp. nov., Rahnella woolbedingensis sp. nov., classification of Rahnella genomospecies 2 and 3 as Rahnella variigena sp. nov. and Rahnella inusitata sp. nov., respectively and emended description of the genus Rahnella*. Vol. 37. 2014.
21. Li, Y., et al., *Elevation of three subspecies of Lonsdalea quercina to species level: Lonsdalea britannica sp. nov., Lonsdalea iberica sp. nov. and Lonsdalea populi sp. nov.* Int J Syst Evol Microbiol, 2017. **67**(11): p. 4680-4684.
22. Soutar, C.D. and J. Stavriniades, *Phylogenetic analysis supporting the taxonomic revision of eight genera within the bacterial order Enterobacterales*. Int J Syst Evol Microbiol, 2020.
23. Denman, S., et al., *Brenneria goodwinii sp. nov., associated with acute oak decline in the UK*. Int J Syst Evol Microbiol, 2012. **62**(Pt 10): p. 2451-6.
24. Vansteenkiste, D., et al., *Predispositions and symptoms of Agrilus borer attack in declining oak trees*. Annals of Forest Science, 2004. **61**(8): p. 815-823.
25. Brown, N., et al., *A review of Agrilus biguttatus in UK forests and its relationship with acute oak decline*. Forestry, 2014. **88**(1): p. 53-63.
26. Reed, K., et al., *The lifecycle of Agrilus biguttatus : the role of temperature in its development and distribution, and implications for Acute Oak Decline*. Agricultural and Forest Entomology, 2018. **20**(3): p. 334-346.
27. Doonan, J.M., et al., *Host-microbiota-insect interactions drive emergent virulence in a complex tree disease*. Proceedings of the Royal Society B: Biological Sciences, 2020. **287**(1933): p. 20200956.
28. Denman, S., et al., *Isolation studies reveal a shift in the cultivable microbiome of oak affected with Acute Oak Decline*. Syst Appl Microbiol, 2016. **39**(7): p. 484-490.
29. González, A.J. and M. Ciordia, *Brenneria goodwinii and Gibbsiella quercinecans isolated from weeping cankers on Quercus robur L. in Spain*. European Journal of Plant Pathology, 2020. **156**(3): p. 965-969.



30. Moradi-Amirabad, Y., et al., *Brenneria spp. and Rahnella victoriana associated with acute oak decline symptoms on oak and hornbeam in Iran*. Forest Pathology, 2019. **49**(4): p. e12535.
31. Meaden, S., C.J.E. Metcalf, and B. Koskella, *The effects of host age and spatial location on bacterial community composition in the English Oak tree (Quercus robur)*. Environmental Microbiology Reports, 2016. **8**(5): p. 649-658.
32. Stone, B.W., Weingarten, E. A. and Jackson, C. R., *The Role of the Phyllosphere Microbiome in Plant Health and Function*, in *Annual Plant Reviews online*, J.A. Roberts, Editor. 2018.
33. Leveau, J., *Microbial communities in the phyllosphere*, in *Biology of the Plant Cuticle*, M.C. Riederer M, Editor. 2006, Blackwell: Oxford. p. 334-367.
34. Vorholt, J.A., *Microbial life in the phyllosphere*. Nature Reviews Microbiology, 2012. **10**: p. 828.
35. Pettifor, B.J., et al., *Survival of Brenneria goodwinii and Gibbsiella quercinecans, associated with acute oak decline, in rainwater and forest soil*. Systematic and Applied Microbiology, 2020. **43**(2): p. 126052.
36. Handelsman, J., et al., *Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products*. Chemistry & Biology, 1998. **5**(10): p. R245-R249.
37. Müller, T. and S. Ruppel, *Progress in cultivation-independent phyllosphere microbiology*. FEMS Microbiology Ecology, 2014. **87**(1): p. 2-17.
38. Zolfo, M., et al., *Profiling microbial strains in urban environments using metagenomic sequencing data*. Biology Direct, 2018. **13**(1): p. 9.
39. Roumpeka, D.D., et al., *A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data*. Frontiers in Genetics, 2017. **8**(23).
40. Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2*. Genome Biology, 2019. **20**(1): p. 257.
41. Nocchi, G., Brown, N., Coker, T., and W. Plumb, Stocks, J., Denman, S., & Buggs, R. J. A. , *Genomic structure and diversity of oak populations in British parklands*. Plants, People, Plan. Plants, People, Planet, 2021.
42. Brown, N., et al., *Spatial and temporal patterns in symptom expression within eight woodlands affected by Acute Oak Decline*. Forest Ecology and Management, 2016. **360**: p. 97-109.
43. Raimbault, P. *Physiological Diagnosis*. 1995. Versailles: Societe Francaise d'Arboriculture.
44. Knight, R., et al., *Best practices for analysing microbiomes*. Nat Rev Microbiol, 2018. **16**(7): p. 410-422.
45. Herrmann, M., et al., *Canopy Position Has a Stronger Effect than Tree Species Identity on Phyllosphere Bacterial Diversity in a Floodplain Hardwood Forest*. Microb Ecol, 2021. **81**(1): p. 157-168.
46. King, T., Butcher, Simon, & Zalewski, Lukasz., *Apocrita - High Performance Computing Cluster for Queen Mary University of London*. 2017, March 24, Zenodo: London.
47. Plomion, C., et al., *Oak genome reveals facets of long lifespan*. Nature Plants, 2018. **4**(7): p. 440-452.
48. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.
49. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
50. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. Nature Methods, 2012. **9**(4): p. 357-359.
51. NCBI, *The NCBI Handbook [Internet]*. 2013-: National Center for Biotechnology Information (US).
52. McDonald, D., et al., *The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome*. GigaScience, 2012. **1**(1).
53. R Core Team, *R: A Language and environment for statistical computing*. 2013, Vienna, Austria: R Foundation for Statistical Computing.
54. RStudio Team, *RStudio: Integrated Development for R*. 2019, RStudio, Inc.: Boston, MA.
55. Lu, J., et al., *Bracken: estimating species abundance in metagenomics data*. PeerJ Computer Science, 2017. **3**: p. e104.
56. Pinheiro, J., et al., *R Core Team (2019). \_nlme: linear and nonlinear mixed effects models. R package version 3.1-148*. 2020.

57. Boratyn, G.M., et al., *Magic-BLAST, an accurate RNA-seq aligner for long and short reads*. BMC Bioinformatics, 2019. **20**(1): p. 405.
58. Uroz, S., et al., *Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil*. Environmental Microbiology Reports, 2010. **2**(2): p. 281-288.
59. Leff, J.W., et al., *Spatial structuring of bacterial communities within individual Ginkgo biloba trees*. Environmental Microbiology, 2015. **17**(7): p. 2352-2361.
60. Jakuschkin, B., *Genetic architecture of the interactions between English oak (Quercus robur L.) and the microbial community of its phyllosphere*. 2015, Université de Bordeaux.
61. Rastogi, G., G.L. Coaker, and J.H. Leveau, *New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches*. FEMS Microbiol Lett, 2013. **348**(1): p. 1-10.
62. Bálint, M., et al., *Host Genotype Shapes the Foliar Fungal Microbiome of Balsam Poplar (Populus balsamifera)*. PLOS ONE, 2013. **8**(1): p. e53987.
63. Cregger, M.A., et al., *The Populus holobiont: dissecting the effects of plant niches and genotype on the microbiome*. Microbiome, 2018. **6**(1): p. 31.
64. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments*. Genome Biology, 2014. **15**(3): p. R46.
65. Kibegwa, F.M., et al., *A Comparison of Two DNA Metagenomic Bioinformatic Pipelines While Evaluating the Microbial Diversity in Feces of Tanzanian Small Holder Dairy Cattle*. BioMed Research International, 2020. **2020**: p. 2348560.
66. Menzel, P., K.L. Ng, and A. Krogh, *Fast and sensitive taxonomic classification for metagenomics with Kaiju*. Nature Communications, 2016. **7**(1): p. 11257.
67. Crampton, B.G., et al., *A multiplex real-time PCR assay enables simultaneous rapid detection and quantification of bacteria associated with acute oak decline*. Plant Pathology, 2020. **69**(7): p. 1301-1310.
68. Vu, V.Q., *ggbiplot: A ggplot2 based biplot*. 2011.
69. Arnold, J.B., *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. 2019.
70. Auguie, B., *gridExtra: Miscellaneous Functions for "Grid" Graphics*. 2017.
71. Wilke, C.O., *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. 2020.
72. Henry, L. and H. Wickham, *purrr: Functional Programming Tools*. 2020.
73. Müller, K. and H. Wickham, *tibble: Simple Data Frames*. 2021.
74. Wickham, H., *The split-apply-combine strategy for data analysis*. Journal of statistical software, 2011. **40**(1): p. 1-29.
75. Wickham, H., *forcats: Tools for Working with Categorical Variables (Factors)*. 2021.
76. Wickham, H., *plyr: Tools for Splitting, Applying and Combining Data*. 2019.
77. Wickham, H., *stringr: Simple, Consistent Wrappers for Common String Operations*. 2019.
78. Wickham, H., *tidyr: Tidy Messy Data*. 2021.
79. Wickham, H., et al., *Welcome to the Tidyverse*. Journal of Open Source Software, 2019. **4**(43).
80. Wickham, H., et al., *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. 2020.
81. Wickham, H., et al., *dplyr: A Grammar of Data Manipulation*. 2021.
82. Wickham, H. and J. Hester, *readr: Read Rectangular Text Data*. 2020.
83. Wickham, H. and D. Seidel, *scales: Scale Functions for Visualization*. 2019.
84. Hope, R.M., *Rmisc: Rmisc: Ryan Miscellaneous*. 2013.
85. Sarkar, D., *lattice: Trellis Graphics for R*. 2018.