scientific reports

OPEN

Population genomic analysis of an emerging pathogen *Lonsdalea quercina* affecting various species of oaks in western North America

Olga Kozhar[1], Rachael A. Sitz[2], Reed Woyda[3,4], Lillian Legg[1], Jorge R. Ibarra Caballero[1], Ian S. Pearse[1], Zaid Abdo[3,4] & Jane E. Stewart[1✉]

Understanding processes leading to disease emergence is important for effective disease management and prevention of future epidemics. Utilizing whole genome sequencing, we studied the phylogenetic relationship and diversity of two populations of the bacterial oak pathogen *Lonsdalea quercina* from western North America (Colorado and California) and compared these populations to other *Lonsdalea* species found worldwide. Phylogenetic analysis separated Colorado and California populations into two *Lonsdalea* clades, with genetic divergence near species boundaries, suggesting long isolation and populations that differ in genetic structure and distribution and possibly their polyphyletic origin. Genotypes collected from different host species and habitats were randomly distributed within the California cluster. Most Colorado isolates from introduced planted trees, however, were distinct from three isolates collected from a natural stand of Colorado native *Quercus gambelii*, indicating cryptic population structure. The California identical core genotypes distribution varied, while Colorado identical core genotypes were always collected from neighboring trees. Despite its recent emergence, the Colorado population had higher nucleotide diversity, possibly due to its long presence in Colorado or due to migrants moving with nursery stock. Overall, results suggest independent pathogen emergence in two states likely driven by changes in host-microbe interactions due to ecosystems changes. Further studies are warranted to understand evolutionary relationships among *L. quercina* from different areas, including the red oak native habitat in northeastern USA.

Global forest ecosystems have been increasingly affected by a variety of disturbances, including altered climatic conditions and increased attack of key tree species by pests and pathogens. While the introduction of alien pathogen species to new geographic areas has been considered one of the main drivers of emerging forest diseases during the last century, other factors such as changes in environmental conditions, new host-vector associations (between introduced insect vectors and native tree pathogens or vice versa), cryptic disease agents (e.g., pathogens with a very long latency period or endophytes changing their behavior to pathogenic), hypervirulent strains of known pathogen species, and/or newly emerging species of unknown origin are also recognized as key factors leading to disease emergence in forest ecosystems around the globe[1–4]. The complex nature of many forest diseases, often integrating interactions of several organisms in a single environment, makes them difficult to manage and study. However, knowing underlying causes of disease emergence is crucial to facilitate effective disease management and preventing epidemic outbreaks that can have catastrophic consequences for ecosystems health.

The genus *Quercus* is one of the most important groups of trees in many regions of the Northern Hemisphere[5], with nearly 500 species that have been characterized globally. In addition, oak forests are characterized by high species diversity and great soil fertility[6]. In North America oak trees compose a significant part of many forest

[1]Department of Agricultural Biology, Colorado State University, Fort Collins, CO, USA. [2]Davey Resource Group, Inc., Urban & Community Forestry Services, Atascadero, CA, USA. [3]Program of Cell and Molecular Biology, Colorado State University, Fort Collins, CO, USA. [4]Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO, USA. ✉email: jane.stewart@colostate.edu

ecosystems. For example, in the eastern USA, oak forest types represent roughly half of all forest land[7]. However, the biodiversity of oak species varies within the continent, with roughly 90 species described in eastern North America, 20 species in California[5], and only one native species in Colorado. In North America, oak species are also often planted as shade trees in urban environments because of the ability to withstand harsh urban environments[8]. This has led to the movement of oak species to new geographic areas. For example, northern red oak (*Quercus rubra* L.) that is native to eastern North America, has become a popular shade tree in Colorado, USA. These planted trees may experience new pathogens and/or pathogens may also act as bellwethers of the effects of climate change because they experience environmental conditions outside of their native range.

Oak decline diseases and complex syndromes, caused by combined effects of biotic and abiotic factors, have been reshaping landscapes of deciduous forests in temperate zones around the world[9]. In the United Kingdom, acute oak decline has been associated with galleries of the two spotted oak borer, *Agrilus biguttatus* Fabricius 1777, and several bacterial species (most commonly with *Brenneria goodwinii* Denman et al. 2012, *Gibsiella quercinecans* Brady et al. 2010, and *Rahnella victoriana* Brady et al. 2017, as well as *Pseudomonas daroniae* Bueno-Gonzalez et al. 2019 and *P. dryadis* Bueno-Gonzalez et al. 2019, and *Lonsdalea britannica* Brady et al. 2012)[4,10,11]. *Brenneria goodwinii* and *G. quercinecans*, and *Brenneria* spp. and *R. victoriana* are also associated with oak decline in Spain and Iran, respectively[10,12].

In the early 2000s a significant dieback of northern red oak, pin oak, and Shumard oak of unknown origin appeared in Colorado and was described as drippy blight disease[13]. Recently emerged drippy blight on planted oaks causes abundant ooze on symptomatic tissue and has led to significant increases in tree mortality and removals in public properties in some areas of Colorado[13]. The main symptoms on infected oaks are leaf scorch and premature leaf drop that is followed by dieback of small twigs and the development of branch cankers, from which copious gummosis/bacterial ooze is produced. Based on phenotypic and genetic analyses coupled with pathogenicity tests it was confirmed that the disease was caused by the bacterial pathogen *Lonsdalea quercina* Hildebrand and Schroth 1967 in association with scale kermes insect, *Allokermes galliformis* Riley 1881. In drippy blight studies, symptoms were observed at scale feeding sites[13] and investigated insects were contaminated with the bacteria[14]. Due to the development of abundant ooze on symptomatic tissue, it was called drippy blight disease. The first record of *L. quercina* was published in 1967[15] reporting drippy nut disease on native oak trees in California, USA, that infects acorns and acorn caps of native oaks (*Q. agrifolia* Nee and *Q. wislizeni* A.DC.) probably using wounds made by wasps, weevils, and other common oak pests to enter the host tissue.

Besides *L. quercina*, three other species of *Lonsdalea*, all tree pathogens, have been described around the world: *L. iberica* Brady et al. 2012 causing bark canker and drippy nut disease on native oaks in Spain[16,17], *L. britannica* Brady et al. 2012 associated with oak decline in Britain[16,18], and *L. populi* Toth et al. 2013 causing canker disease of *Populus* in Spain and China[19,20].

While the causal agents of drippy blight have been identified, the causes leading to its (and other *Lonsdalea* spp. oak pathogens) recent emergence remain unknown. To our knowledge, little is known about *L. quercina* population biology and relationship between the bacterial populations from different areas. Since the only two places in North America with published records of *L. quercina* causing disease on oak trees are California and Colorado, we studied the relationship between populations of *L. quercina* from symptomatic oaks in California and Colorado. The objectives of the study were to determine the phylogenetic relationship between *L. quercina* populations from California and Colorado and other *Lonsdalea* species found worldwide and investigate population structure of *L. quercina* in these two locations. Since *L. quercina* was first documented in California, we also assessed whether the Colorado population originated from California, as well as discuss potential processes leading to the disease emergence in Colorado.

Results

Pangenome and clonality. The mean sequence coverage of *Lonsdalea quercina* samples was 82x (range: 50–112×). Genome assemblies consisted of on average of 103 contigs (range: 47–252). Total average length of assembled genomes was 3,827,997 bp (range: 3,700,835–3,999,440 bp), with average GC content 55.21% (range: 54.7–55.6%) and average N50 being 345,548 (range: 209,168–871,573). Our genome assembly and annotation utilized the Prokka and Roary analyses. The pangenome of combined California (CA; n = 52) and Colorado (CO; n = 31) populations consisted of 8624 genes, 2370 of which comprised the core genome (total of 2,460,085 bp) (Fig. 1). The core genomes of each population separately also contained a population specific set of genes that included 245 core genes for CA isolates and 372 core genes in CO (Fig. 1). Among the total 83 isolates sequenced there were 65 unique core genotypes (46 in CA and 19 in CO; genotypic diversity–88.46% in CA and 61.29% in CO) detected (Table 1, Table S1).

Phylogenetic analyses and species boundaries. Phylogenetic analyses, with members from *Enterobacteriacea* Rahn 1937, *Erwiniaceae* Adeolu et al. 2016, *Pectobacteriaceae* Adeolu et al. 2016, and *Yersiniaceae* Adeolu et al. 2016 families of the order *Enterobacterales* ord. nov.[21], separated the CA and CO populations into two distinct clades grouping them with two other previously published genomes of *L. quercina*: members of the CA population were grouped with type strain ATTC 29281 previously collected in CA, and CO population genomes grouped with NCCB 100490 previously collected in CO (Fig. 2, Fig. S1). The phylogenetic analysis on the core gene presence-absence matrix indicated a clear separation of *L. quercina* populations from other closely related *Lonsdalea* species, and a clear distinction between CA and CO populations of *L. quercina* supporting these groupings (Fig. 1). The phylogeny also indicated that the most closely related species to our *L. quercina* populations was *L. iberica*. Interestingly, according to the phylogenetic analyses the *L. quercina* CO population was more closely related to *L. iberica* than the CA population, suggesting possible polyphyletic origin of *L. quercina* CA and CO populations.
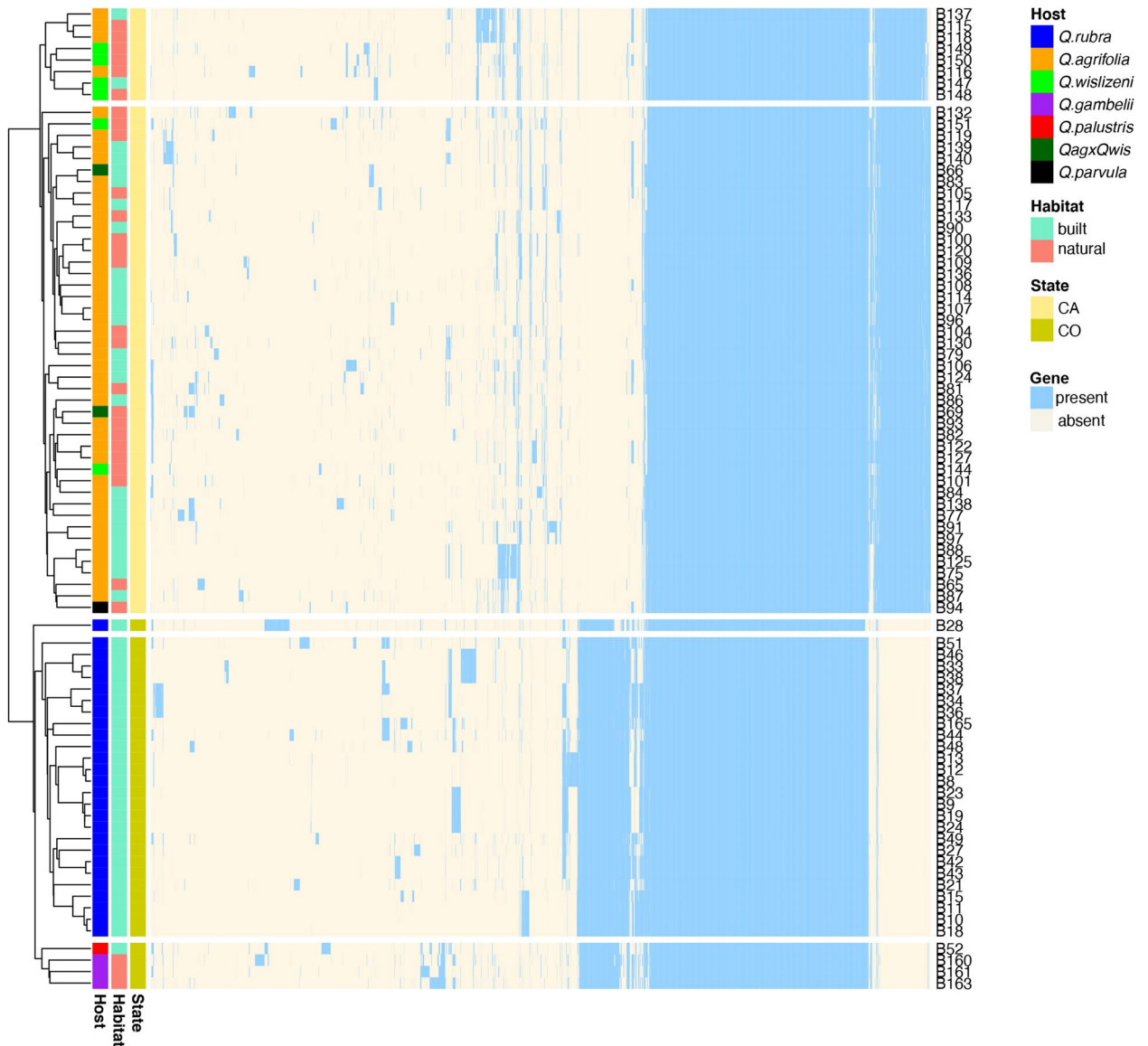
Figure 1. Gene presence-absence matrix from pangenome analysis of 83 isolates of *Lonsdalea quercina* populations collected in California (CA) and Colorado (CO), USA. Each column indicates a gene (blue—present, yellow—absent), each row represents a gene profile of each isolate. Host species, habitat, and state from which each isolate was collected are color coded according to the legend. Pangenome analysis was conducted with Roary v.3.13.0.

Pop	Habitat	Sample #	Genotype #	Genotypic diversity, %[a]	pi[b]
CA	Built	26	23	88.46	0.0048
	Natural	26	23	88.46	0.005
CA total		52	46	88.46	0.0049
CO	Built	28	16	57.14	0.0103
	Natural	3	3	100.00	–
CO total		31	19	61.29	0.0109
CA + CO total		83	65	78.31	

Table 1. Diversity of *Lonsdalea quercina* populations in California (CA) and Colorado (CO), USA. [a]Genotypic diversity was calculated as the percentage of samples carrying unique core genotypes. [b]pi—unbiased nucleotide diversity calculated with Pixy v.0.95.02.
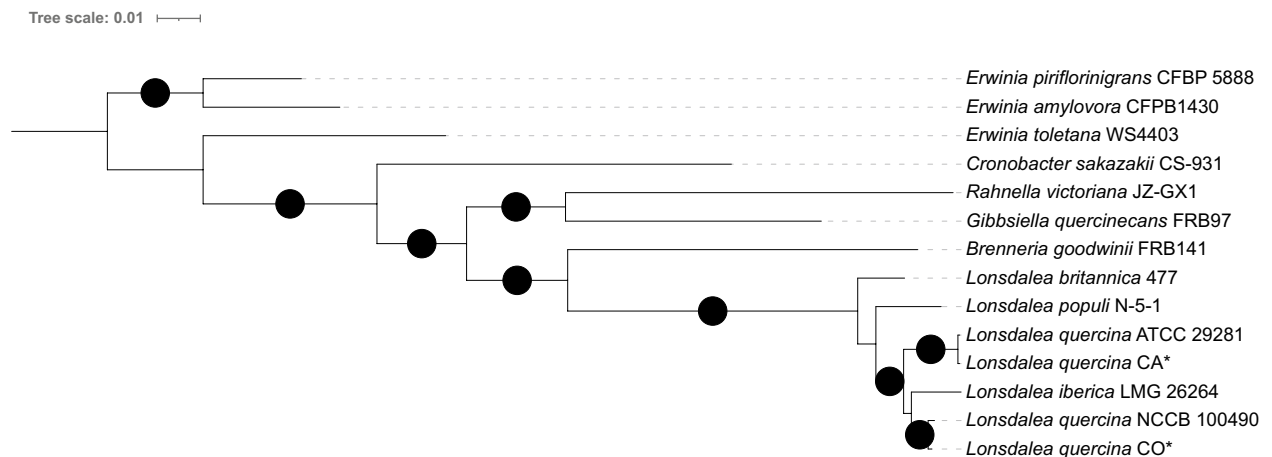
Figure 2. Maximum likelihood phylogenetic tree, based on 20 core genes shared among all individuals in the dataset, showing position of *Lonsdalea quercina* populations sampled in California (CA) and Colorado (CO) within the genus *Lonsdalea* and the order *Enterobacterales*. Branches with > = 95% likelihood support are indicated with black circles. Branch support was calculated with 1000 ultrafast bootstrap replicates. *Erwinia piriflorinigrans* and *Erwinia amylovora* were used as root. *Lonsdalea quercina* CA and *Lonsdalea quercina* CO are consensuses of multi-FASTA alignments of core genes of CA and CO populations, respectively, produced by Roary v.3.13.0.

Average nucleotide identity (ANI) above 95–96% is used to define bacterial species boundaries[22]. Pairwise comparisons of ANI between consensus alignments of the core genomes of either CA or CO populations and one representative each of other *Lonsdalea* spp. were below the species boundaries threshold (between 89.08–93.75%), indicating a clear species boundary between *L. quercina* from other *Lonsdalea* species (Table 2, Table S2). However, the average ANI between CA and CO was 95.18% (range 94.83–95.36) which is right on the threshold of ANI bacterial species boundaries. The CO population had a slightly lower within population average nucleotide identity (ANI) compared to CA population (98.98% for CO and 99.51% for CA, respectively). The lowest ANI between CA and CO populations was observed when a CO genome B28 was compared to any CA genome (average ANI between B28 genome and all CA genomes = 94.87%, range 94.83–95.01). ANI between B28 and CO population was lower than average within CO population ANI but was still within species boundaries (average ANI between B28 and the rest of CO genomes = 97.97%). When B28 genome was excluded from the analysis, the minimum ANI between any CA and CO genomes was 95.06%. These results, together with phylogenetic analyses, suggest that CA and CO populations have been isolated from each other for a long time and may be of mixed evolutionary origin.

Functional annotation and orthology analyses of population specific core genes in California and Colorado populations. Comparative analysis of the predicted protein sequences of core genes unique to each of the populations identified 189 orthologous clusters, 186 of which were shared between CA and CO populations (Fig. S2a), suggesting that most genes, even though differed in sequences identity, produced same metabolic products. Tests for GO enrichment of identified clusters revealed one significantly enriched term in each of the populations (GO:0046690 in CA and GO:0009086 in CO, P < 5e−4, as reported by Orthovenn3). The CA cluster contained three genes. Ten GO terms were associated with these genes, albeit most of them related to uncharacterized or poorly characterized biological or metabolic processes (Fig. S2b). The best Swiss-Prot hits for this cluster were predicted proteins involved in the response to tellurium ion. One significantly enriched cluster containing two genes was identified in CO population with six associated GO terms, also all related to

ID	CA	CO	*L. iberica*	*L. britannica*	*L. populi*
CA*	99.51	95.17	92.18	89.32	90.1
CO*		98.98	93.75	89.72	90.77
L. iberica				89.37	89.08
L. britannica					90.34

Table 2. Estimates of Average Nucleotide Identity (ANI) within *Lonsdalea* genus. CA—*Lonsdalea quercina* population collected in California. CO—*Lonsdalea quercina* population collected in Colorado. ANI estimates between CA and CO are averages all vs all individuals in each population (total of 83 samples), while estimates between CA, CO, and other species are based on CA and CO population core genomes alignments and 1 representative of *L. iberica*, *L. britannica*, and *L. populi* from Table 3.

uncharacterized or poorly characterized biological or metabolic processes (Fig. S2b). The best Swiss-Prot hit characterized the CO population predicted proteins to be involved in methionine biosynthetic process.

The 245 and 372 core genes unique to CA and CO populations, respectively, were also compared in functional annotation databases dbCAN and PHI-base. Of 245 and 372 core genes unique to CA and CO populations, only four (CA) and three (CO) genes and 26 (CA) and 24 (CO) genes were functionally annotated in dbCAN and PHI-base databases, respectively (Fig. S2d, e, Tables S3–S4). Enzymes belonging to families of glycoside hydrolases 23 and 37 and glycosyl transferase family 51 involved in decomposition of polysaccharides were identified in both populations, and glycoside hydrolase family 12 involved in decomposition of xyloglucan was identified in CA population only (Fig. S2d, Table S3). Genes related to pathogenicity were also identified in both populations (Fig. S2e, Table S4), with more virulence factors being identified in CA population.

Population structure and distribution. Based on the gene presence-absence matrix of core and accessory genes CA and CO populations formed two separate clusters (Fig. 1). Both populations contained isolates from trees grown in natural (undeveloped landscapes such as forests, parks, nature preserves, etc.) or built (developed landscapes such as residential neighborhoods of cities, parking lots, etc.) areas (see materials and methods). CA population isolates from either natural or built areas were collected from different host species and these were randomly distributed within the CA cluster (Fig. 1, Table S1). In the CO population all samples, except 1, were collected from planted trees grown in built areas formed a separate group, while three samples were collected from native *Q. gambelii* Nutt in a natural area; all grouped together with 1 sample from a planted tree of a different species (*Q. palustris* Munchh.). Due to the small sample size from the natural area, it remains unclear whether bacterial populations are structured by habitat or host type in CO.

To further study population structure of CA and CO populations, we used an unrooted maximum likelihood phylogenetic tree (Fig. 3a). Consistent with the groupings in the gene presence-absence matrix (Fig. 1), CA and CO populations were separated into two well-supported clades, with isolates from different habitats and host species being randomly distributed within the CA clade. Within CO clade, in turn, three isolates from native *Q. gambelii* (B160, B161, and B163) and 1 isolate from a planted *Q. palustris* (B52) formed a well-supported cluster, distinct from planted *Q. rubra* isolates. Phylogenetic analyses also supported the gene presence-absence matrix grouping results for CO isolate B28 collected from a planted *Q. rubra* in a built area as genetically distinct from the rest of CO population.

Principal component analysis (PCA) based on 2355 thinned core SNPs supported the results of phylogenetic analysis placing CA and CO populations into two distinct groups. CA isolates collected from various species of native oak and different habitats grouped closely together, while CO samples were more spread on the PCA plot (Fig. 3b). In agreement with the groupings in the gene presence-absence matrix and maximum likelihood phylogenetic tree (Figs. 1 and 3a), three CO isolates collected from three plants of a native shrub *Q. gambelii* and 1 isolate from planted *Q. palustris* clustered together on the PCA plot, while the rest of the CO isolates (except B28) collected from planted *Q. rubra* grouped on the opposite side, further suggesting the presence of population structure within CO. Within CA population, no population structure was observed based on host species or habitat, but some strains clustered separately from the rest also indicating presence of some population structure in that region (Fig. 3c, d).

In CA, isolates with identical core genotypes (no SNP variation across the core genomes) varied in their geographical distribution. Some representatives of the same genotype were collected from neighboring trees (Fig. 4a). For example, isolates of the B122 genotype were collected from trees located 4 km from each other (both in natural areas—one in Mare Island Shoreline Heritage Preserve and another in Crockett Hills Regional Park) in San Pablo Bay area. Similarly, two isolates of the B139 genotype were collected from trees 80 m apart in a built area on UC Davis campus. In contrast, representatives of another 4 genotypes were collected from geographically distant areas. Of these, two isolates each representing genotypes B125 and B107 were collected from trees in built areas located 35 and 83 km apart, respectively (trees grown in parking lots or in highly developed city neighborhoods). Two isolates of genotype B100 were collected from trees in natural areas located 197 km from each other (one in Mare Island Shoreline Heritage Preserve in San Pablo Bay area and another in Hastings Natural History Reservation in Carmel Valley). And finally, two isolates of genotype B147 were collected in oak savannah foothills on Rough and Ready highway outside of Yuba City—one alongside the road free of any development (categorized as natural area), and another 13 km down the road next to a private house (categorized as built area) (Fig. 4a). The CO population had higher clonality, but all representatives of the same genotype were always collected from trees in a close proximity to each other (e.g., isolates of either B8, B9, or B10 genotypes were collected around campus of CU Boulder) (Fig. 4b). Per site nucleotide diversity (pi) of CO was twice that of the CA population (0.0109 in CO and 0.0049 in CA, Table 1). Per site nucleotide diversity between isolates from CA built and natural areas did not differ (0.0048 and 0.0050, respectively).

Recombination. The phylogenetic clustering of CA and CO populations was well supported with high bootstrap support, but due to unexpectedly higher nucleotide diversity in the recently emerged CO population compared to the CA population, we explored the possibility of recombination. The Phi test, based on 172,269 informative sites, revealed statistically significant evidence of recombination (P < 0.0001). Split decomposition analysis[23] suggested this was due to recombination among 3 isolates from native *Q. gambelii*, 1 isolate from *Q. palustris*, and 1 isolate (B28) from planted *Q. rubra* in the CO population, and recombination between isolates B48 and B165, B44 and B51, as well recombination among several individuals within CA populations (Fig. S3a–c). However, according to this analysis, no recombination was detected between CA and CO populations. Relative rates of recombination to mutation were the same in CA and CO (R/θ = 0.149 for each), suggesting that genetic changes due to mutation are 7 times more likely than due to recombination in these populations. A

Figure 3. Population structure of *Lonsdalea quercina* sampled in California (CA) and Colorado (CO). (**a**) Maximum likelihood phylogenetic tree showing distinct clades of CA and CO genotypes. Branches with > = 95% likelihood support are indicated with black circles. Size of circle reflects percentage of likelihood support within 95–100% range (smallest circle being 95% and the largest circle being 100%). Branch support was calculated with 1000 ultrafast bootstrap replicates. Circles and triangles next to phylogenetic tree indicate samples from natural and build habitats, respectively. *Q. agrifolia, Q. parvula, Q. wislizeni,* and *Q.agrifolia × Q. wislizeni* (QagxQwis) were sampled in CA, while *Q. rubra, Q. gambelii,* and *Q. palustris*—in CO. (**b**) Principal component analysis (PCA) representing distinct genetic groups of CA and CO populations. (**c, d**) PCA of CA population only representing lack of population structure either by host species (**c**) or habitat (**d**) within the state. Each dot indicates one individual. Percentages in parenthesis on PCA plots indicate the variance explained by each principal component. Different colors in (**a, b**), and (**c**) represent isolates from different host species indicated in "host" legend. Different colors in (**d**) represent isolates from different habitats ("built", "natural").

ClonalFrameML phylogeny with branch length corrected to account for recombination showed longer branches in CO population compared to CA population, supporting results of higher nucleotide diversity in CO population (Fig. S4).

Discussion

An increasing number of tree diseases cause significant, long term, threats to wild and human mediated ecosystems. Clarifying processes leading to the emergence of new diseases is a critical step towards their effective management. In this paper, we aim towards characterizing possible causes of emergence of drippy blight disease caused by *Lonsdalea quercina*, a recently emerged pathogen of oak species in association with kermes scale *Allokermis galliformis*, in Colorado, USA. Using population genomic data, we have revealed that this recently emerged *L. quercina* population in Colorado is genetically distinct from an established *L. quercina* population in California, suggesting that an accidental introduction of the bacteria from CA to CO is unlikely. Despite clear population differentiation, the results of our analyses were inconclusive whether the CO *L. quercina* population comprises a new taxonomic species. Surprisingly, the CO population had a nucleotide diversity that was twice as high compared to the CA population despite a smaller sample size, suggesting bacterial strains isolated from red oak may be moving to CO from elsewhere or that this population has been present in the area for a long time and moved to introduced and planted *Quercus rubra* (red oak) from native species *Q. gambelii* (Shumard oak).

Despite increased access to molecular and genomic data on various microorganisms, including bacteria, taxonomic placement of species in many cases remains a challenging task[24]. Based on previous neighbor-joining phylogenetic analysis of *Lonsdalea* spp. and estimates of average nucleotide identity (ANI) between two previously published genomes of *L. quercina* (*L. quercina* ATCC 29281 sampled in CA and *L. quercina* NCCB 100490 sampled in CO[25], Li et al.[26] hypothesized that these two strains from North America may comprise different species (ANI = 94.04%). In this study using population data from both locations, maximum likelihood genealogies based on either 20 or 51 core genes placed both populations within the genus *Lonsdalea*. In the phylogeny, the CA and CO populations formed two distinct and well supported clades that are closely related to *L. iberica*. Surprisingly, phylogenetic analysis indicated that the *L. quercina* CO population is closer to *L. iberica* than the CA population, suggesting that *L. quercina* may be a polyphyletic group. This pattern was present in both phylogenetic trees, built using either 20 or 51 core genes, however, only the tree based on 51 core genes had a well supported bootstrap highlighting the relationship between *L. iberica* and *L. quercina* CO population. ANI, calculated at the population level (all vs all), indicated higher ANI between CA and CO than the one reported by Li et al. (24). Our analyses demonstrated that these populations show genetic separation just above the proposed ANI species boundaries (average ANI = 95.17%, range 94.83–95.36%) and cannot be separated into two distinct species based on this measurement. At the same time ANI between *L. iberica* and either of studied *L. quercina* populations is below proposed species boundaries. Overall, the results of phylogenetic analyses and ANI estimates suggest that *L. quercina* from CA and CO show clear signs of differentiation, likely due to a long period of isolation from one another, and that *L. quercina* may have a complex evolutionary history. The lack of recombination between the CA and CO populations, based on the results of split decomposition analysis with SplitsTree, is consistent with the hypothesis of a possible polyphyletic origin of *L. quercina* populations. Our pangenome analysis indicated that while CA and CO share most of the core genome (2370 genes), each population has a unique set of core genes (245 core genes in CA and 372 core genes in CO), yet most of the gene clusters identified in the dataset are orthologous, sharing functionality, between the populations according to the orthology analyses (by Orthovenn3). It is worth mentioning that this result could also be affected by the threshold similarity in Roary analysis that is arbitrary set to 95%. Hence, it is possible that these gene clusters are not unique to each population, but rather are common genes diverged beyond 95% threshold. Some epidemiological differences also exist between the CA and CO populations—*L. quercina* in CA infects acorns of native oaks (*Quercus agrifolia, Q. parvula,* and *Q. wislizeni*), while in CO *L. quercina* infects branches in introduced and planted *Q. rubra* and *Q. palustris*, but acorns in natural areas of native *Q. gambelii*. Overall, these results remain inconclusive regarding the taxonomy of CA and CO populations, and further phenotypic and biochemical assessments are required to clarify their taxonomic placement, therefore in this study we refer to them as populations.

Many plant pathogens spread via clonal reproduction[27,28], including important bacterial crop pathogens (for example[29,30]). Phylogenetic and recombination analyses indicated that both populations of *L. quercina* in this study are comprised of mostly clonal modes of reproduction. For bacteria that rely on external factors for dissemination, the distribution of identical core genome genotypes may provide insights into organismal population dynamics. In CA where drippy nut disease caused by *L. quercina* has been established for several decades (first reported in 1967[15], some identical core genome genotypes were randomly distributed across a large sampled
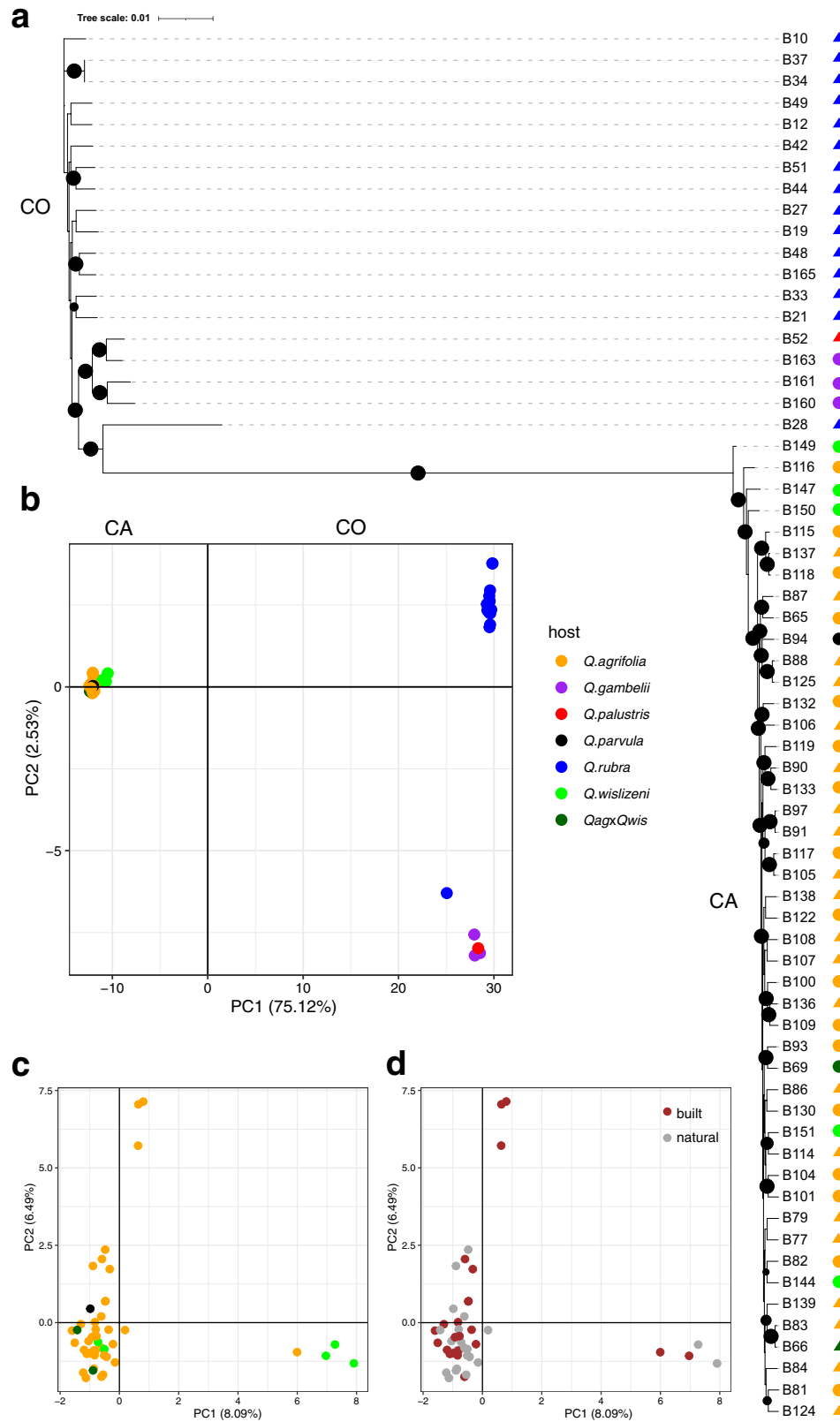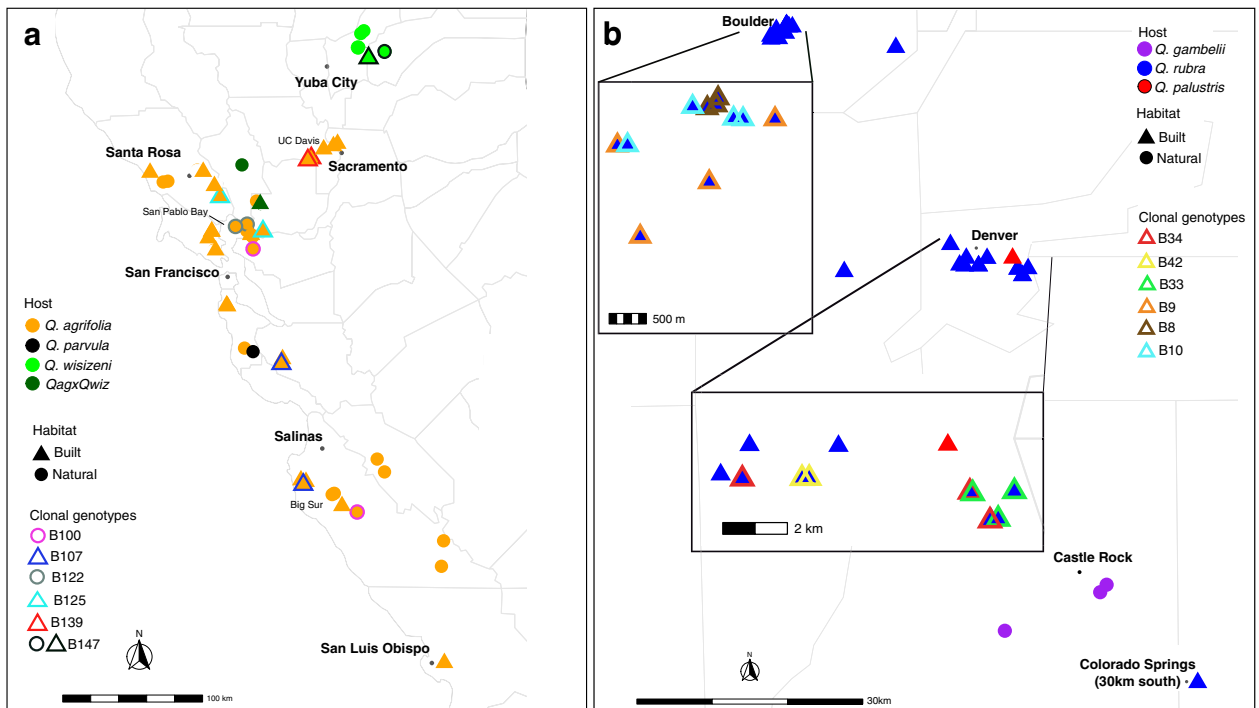
Figure 4. Distribution of *Lonsdalea quercina* populations in California (**a**) and Colorado (**b**), USA. Every symbol on the map represents one isolate. Different colors represent different host species, and figures—habitats (see legends). Clonal genotypes are highlighted with different colors according to the legend "Clonal genotypes". Unhighlighted samples are unique genotypes. Note different scale of each map. On Colorado map (**b**) Boulder and Denver areas are zoomed in to show clonal distribution within cities areas. One isolate from Colorado Springs (B165) is indicated in the right bottom corner (its actual location is 30 km south). Maps were generated in R v4.3.2 in RStudio v1.2.1106 using the following packages: ggplot2 v3.4.2, googleway v2.7.7, ggrepel v0.9.3, ggspatial v1.1.8, libwgeom v0.1-6 and sf v1.0-14.

region in CA, including isolates of a genotype B107 that were sampled from trees of *Q. agrifolia* grown in built areas 85 km from each other. Genotypic diversity did not differ between isolates from natural and built areas in CA (~ 88% in each). There was no association among genotypes with either habitat or host species in CA, suggesting movement of the pathogen between different habitats. These results are not surprising since all oak species in CA are native to the region, and *Q. agrifolia* is a major component of native forests and is also used as a shade tree in residential areas. To date there are no phytosanitary measures established for *L. quercina* in CA, making it possible for the inoculum being moved across large distances by insects, birds, small mammals, and human-mediated activities.

In CO, *L. quercina* of the same core genotype were always sampled from planted trees growing in a close proximity to each other (often neighboring trees). Albeit a small population of three, *L. quercina* collected from native *Q. gambelii* were genetically different from isolates collected from planted *Q. rubra*. Whether these differences are due to geographic distance or host species remains unknown. It is known that *L. quercina* from *Q. rubra* is pathogenic on two other introduced to CO species of oaks, *Q. shumardii* Buckley and *Q. palustris*[13]. In this study, one isolate from *Q. palustris* clustered together with isolates from *Q. gambelii*, suggesting possible movement of bacteria among different host species. Interestingly, the location of sampled *Q. palustris* tree was 45 km away from sampled plants of native *Q. gambelii*. Based on population genetic clustering, none of the sampled isolates from *Q. rubra* were genetically similar to the isolates from *Q. gambelii*, and it remains to be determined whether *L. quercina* from native *Q. gambelii* serves as an inoculum reservoir for planted *Q. rubra* hosts or vice versa. Further studies including isolates of *L. quercina* from different host species and geographical areas need to be done to understand the pathogen movement clearer.

Bacterial pathogens can be associated with insects, whereby entering host tissues via feeding sites, wounds, and/or for dissemination among hosts[19,31–33]. It was suggested that in CA, *L. quercina* may enter host tissue via wounds made by acorn weevils, filbertworms, and some cynipid wasps[34]. Large numbers of free-living nematodes *Panagrellus redivivus* were often found feeding on ooze produced by *L. populi* on poplar trees in Hungary[19]. *Lonsdalea quercina* is also disseminated by several insect species in CO that may be helping the bacteria spread among host trees[14]. Among the associated insects in CO, the kermes scale has a well described close relationship with the pathogen to cause drippy blight disease[13]. It has been suggested that the scale insect may form wounds that allow *L. quercina* to enter the host plant more readily or that there is a more indirect interaction whereby the scale induces host stress, facilitating pathogen growth and disease development. However, this insect does not travel long distances, and this may be one of the explanations why the distribution pattern of recently emerged *L. quercina* genotypes is limited to nearby trees.

Despite more recent emergence, the CO population had ~ 2 times higher nucleotide diversity compared to CA population. This is further unexpected because the sampling area in CA was much larger compared to CO (280 km and 82 km long in CA and CO, respectively and with the majority of CO samples being collected within ~ 40 km radius). At the same time, the relative rate of recombination to mutation was the same in both populations. Higher nucleotide diversity observed in CO may suggest the presence of migrants from different source populations or that the CO population is native and is older than the CA population. As mentioned above, analyses including additional samples from *Q. gambelii* are needed to determine the relationship between *L. quercina* populations infecting native *Q. gambelii* and introduced *Q. rubra* hosts in Colorado.

Most isolates in CO were collected from introduced red oak (*Q. rubra*) from Boulder and Denver, CO which is widely planted as a shade tree in the state. The disease was first documented in Boulder in 2012 when the number of dead oaks doubled and city foresters assessed tree health[13]. The disease has since spread to Denver, CO, but is not present on red oak in other cities across the state. In fact, the disease has not been recognized anywhere else on red oak. Further, red oak seedling are sold in local CO nurseries, but often these seedlings are re-sold from suppliers located in northeastern states of the country where this species is native rather than grown in CO. Several bacterial tree pathogens are known to exist in host tissue without causing symptoms and are therefore spread through grafting or nursery stock [e.g., a bacterial pathogen causing watermark disease in willow trees *B. salicis*[32]]. Other members within the order *Enterobacterales* ord. nov. (formerly members of the *Enterobacteriacea* family[21]), *B. goodwini* and *G. quercinecans* associated with acute oak decline, were found in low abundance in healthy trees[35]. These two species as well as *L. britannica* and *R. victoriana* were detected in healthy tree tissues using metagenomics[36], suggesting their possible endophytic behavior. It could be that *L. quercina* spread via nursery stock or moved from CO native *Q. gambelii* to introduced *Q. rubra* and then in older trees, 15–20 years later, began showing symptoms. This hypothesis needs to be tested in future studies, that will include identifying *L. quercina* populations from northeastern USA, a native habitat of *Q. rubra*, and populations from CO native *Q. gambelii*.

In CO, bacterial isolates were collected from mature oak trees that were planted at least 50 years ago, while first drippy blight diseased trees were noticed in early 2000s. Hence, even if *L. quercina* strains were accidentally introduced with trees to CO, it is unlikely these introductions were the cause of the current epidemic outbreak. In 1990s–2000s three new species of *Lonsdalea* were characterized in different parts of the world. *Lonsdalea britannica* was isolated from symptomatic oak tissues associated with acute oak decline in Great Britain[18], *L. iberica* associated with bark canker of *Quercus* species in Spain[18], and *L. populi*, a causal agent of bark canker of poplar species was reported in Hungary, China, and Spain[19,20,37]. During the same time symptoms of drippy blight disease were first noticed in Colorado, USA[13]. Therefore, it is likely that other environmental factors might have led to the emergence of novel bacterial tree pathogens from the genus *Lonsdalea*, regardless of whether the bacterial strains were introduced to CO from elsewhere or moved from native host *Q. gambelii*.

Bacterial-plant relationships can be altered directly, e.g., via introductions or by changes in environment, or indirectly, e.g., by changes in population dynamics of other organisms with which bacteria interact. According to Filippo et al.[38], in Italy Turkey oak (*Quercus cerris* L.) growth decline is associated with climate change. Sun et al.[39] reported that warming temperatures and increasing drought significantly affect *Quercus* species richness and distribution in China. Using bioclimatic modeling, Mclaughlin & Zavaleta[40] found that saplings of California valley oak (*Q. lobata* Nee) are especially susceptible to warming and drying conditions. Trees stressed by environment conditions can become more susceptible to other stressors, including pathogens[41]. It is possible that rapid changes in climate might have altered relationship between *Lonsdalea* spp. and their hosts. This is particularly likely when considering the drippy blight outbreak in Colorado, as these trees are planted outside of the climatic range where they are native.

The kermes scale (*A. galliformis*), now associated with drippy blight in CO, has been present in the state for a long time but was not recognized as a pest because of a lack of significant damage to oak trees. A spike in insect population size was observed with the first drippy blight outbreak[13]. Sitz et al.[13] hypothesized that interactions between *L. quercina* and *A. galliformis* may be direct or indirect, whereby the scale facilitates pathogen movement into the trees via wounds or that the scale is possibly acting as a stressor of the host and facilitating growth and spread of bacteria within the host tissue. A similar observation was reported about the wood-boring beetle *Agrilus biguttatus* Fabricius, 1777 and *B. goodwini* relationship, in which presence of the beetle larvae triggers the upregulation of tree damaging genes in the bacterium leading to acute oak decline symptoms in Britain[42]. Factors leading to the increase of abundance of *A. galliformis* in CO are unknown. Because of the sessile nature of hard-shelled scale, it seems unlikely that populations are attracted to *L. quercina* infected trees to create new infestations, but it is possible. Trees infected with *L. quercina* may be beneficial to the scale, either through reduced levels of host response or changes in phloem sap composition caused by the bacterial infection. It remains to be determined whether the increased scale population and associated increase in the number of entry points through feeding sites has led to the increase of bacterial populations, or vice versa. But it is possible that increase in population size of one organism (for example, caused by climate change) could have caused increase in population size of another organism.

Conclusions

This study revealed that populations of *L. quercina* in California and Colorado are different, and it is unlikely that the drippy blight epidemic outbreak in CO was the result of accidental introduction of the pathogen from CA. Bacteria were isolated from planted stands of introduced *Q. rubra*, as well as three trees of native *Q. gambelii*, suggesting that the pathogen is also present in natural stands of native CO oak species. However, three isolates from *Q. gambelii* were genetically different from the isolates collected from *Q. rubra*. Due to a small sample size on *Q. gambelii* it remains unclear whether the pathogen moved from native *Q. gambelii* to introduced *Q.*

rubra in CO. While the presence of localized clonality in the CO population is consistent with recent pathogen emergence, higher nucleotide diversity suggests that *L. quercina* isolates from *Q. rubra* might be migrants from elsewhere, possibly from northeastern (NE) states of the USA where red oak seedlings are being produced. Further population genetic analyses, including more populations from *Q. gambelii* in CO and *Q. rubra* in NE USA, are needed to test these two hypotheses of the pathogen origin and clarify possible causes and sources of the recent outbreak of drippy blight in CO, USA.

Methods

Bacterial strains and DNA extraction.
In 2018 symptomatic samples with oozing acorns and/or acorn caps (California, CA) and oozing branches (Colorado, CO; except for *Quercus gambelii* oozing acorns and acorn caps) were collected from total of 83 (52 CA and 31 CO) trees of various oak species native to North America from trees in undeveloped habitats such as forests, parks, nature preserves, etc. (referred here as "natural areas") or developed habitats such as residential neighborhoods of cities, parking lots, yards of private houses in rural areas, etc. (referred here as "built areas"). The tree species were identified by co-author Dr. Ian Pearse (Table S1). Three symptomatic samples, for both branch and acorn/acorn caps were surface disinfested by dipping in 10% bleach for 1 min followed by 1 rinse in distilled H$_2$O for 60 s and blotted dried. Branch bark and cambium were removed at canker margin and canker tissue was streaked onto nutrient agar (Sigma-Aldrich, St. Louis, MO, USA) plate. Plates were incubated at 25C for 3–5 days. All cream-colored bacteria were isolated into pure culture on new plates with nutrient agar.

Lonsdalea bacterial colonies were verified using colony PCR: bacterial colonies were lysed in 20 µl sterile water at 95 °C for 5 min and then used in PCR. Successful amplification of a gryB gene region, which encodes the subunit B protein of DNA gyrase, using species-specific primers (gryB forward: 5′-CTGTACAAGGTG AAGAAAGG-3′, gryB reverse: 5′-CGTCACCAG CATCTCCATCC-3′,[43]) was used in PCR species verification. PCR conditions followed Sitz et al.[14] and were: 94C for 3 min, 30 cycles at 94C for 30 s, 60C for 30 s, and 72C for 1 min, then finally 72C for 7 min using a MJ PTC-200 thermocycler (Bio-Rad Laboratories, Waltham, MA). PCR products were visualized via Sub-Cell GT Wide Mini electrophoresis system (BioRad, Hercules, CA, USA) that yielded a single band (286 kb in length) on 1.5% agarose gel.

Plates with pure cultures of verified *Lonsdalea quercina* isolates were incubated at 25 °C for 3–5 days for DNA extraction. DNA was extracted from pure bacterial colonies using Quick-DNA™ Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, CA, USA) following manufacturer protocol. DNA quality and concentration were estimated with NanoDrop One C spectrophotometer and Qubit 4^{MT} with a dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). Total of 83 *Lonsdalea quercina* strains (one strain per tree) were subjected to whole genome sequencing. Bacterial cultures were put in 15% glycerol for long term storage at −80 °C.

Genome sequencing, assembly, and annotation.
Illumina short read genome sequencing of 83 isolates of *Lonsdalea quercina* was performed at the Novogene Bioinformatics Institute (Beijing, China). DNA libraries were constructed using NEBNext Ultra DNA Library Prep Kit according to the manufacturer's protocol. Libraries were sequenced on Illumina NovaSeq 6000 platform with 150 bp paired end reads. Raw data cleaning (adapter removal) and quality control were performed at Novogene.

Raw reads were concatenated, duplicates were removed with BBtools function "dedupe" followed with "reformat" (BBMap—Bushnell B.—sourceforge.net/projects/bbmap/). Reads quality was assessed with FASTQC v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Genome assembly, assembly quality control, and genome annotation were done using Reads2Resistome pipeline v0.0.2[44]. Within the pipeline, genome assembly was performed using Unicycler v0.4.9[45], genome quality control was assessed using QUAST V5.2.0[46] and genome annotation was done using Prokka v1.14.5[47]. The pangenome analyses were carried out with Roary pipeline v.3.13.0[48] with gff3 input files from Prokka annotation. Three datasets of core genes were used for further analyses that are described in detail below (Fig. 5a and b). All core gene datasets were obtained with individual runs of Roary.

Functional annotation of population specific core genes.
Population specific core genes (245 in CA and 372 in CO population) were compared in functional annotation databases Swiss-Prot and GO through Orthovenn3[49], dbCAN2[50] and PHI-base[51]. The GO term enrichment analysis was done with Orthevenn3. For orthologous analyses in Orthovenn3 the OrthoFinder algorithm with e-value 1e−2 and inflation value 1.50 were used. For annotation of proteins involved in substrate degradation with dbCAN2 protein sequences were annotated using three tools: HMMER: dbCAN (e-value < 1e−15, coverage > 0.35), DIAMOND: CAZy (e-value < 1e−102), and HMMER:dbCAN-sub (e-value < 1e−15, coverage > 0.35). Only sequences annotated with all three tools were recorded. The annotation of genes involved in host–pathogen interactions within PHI-base was conducted with default parameters except e-value was set to 1e−15 and the percent identity threshold was set to ≥ 50%.

Species boundaries.
The maximum likelihood phylogenetic trees based on core genes were constructed with IQTREE v.2.0.3[52]. First, we determined the phylogenetic position of CA and CO populations within the genus *Lonsdalea* and the order *Enterobacterales* ord. nov.[21]. The genomes of members of four families from *Enterobacterales* ord. nov. (*Erwinia piriflorinigrans*, *E. amylovora*, *E. toletana* from the family *Erwiniaceae*; *Cronabacter sakazakii* from the family *Enterobacteriaceae*; *Rahnella victoriana* and *Gibbsiella quercinecans* from the family *Yersiniaceae*; and *Brenneria goodwinii*, *Lonsdalea populi*, *L. quercina*, and *L. iberica* from the family *Pectobacteriaceae*[21]) were retrieved from the National Center of Biotechnology Information (Table 3). A total of 51 and 20 core genes present in 100% of samples were used to build phylogenetic trees with all the California (CA),
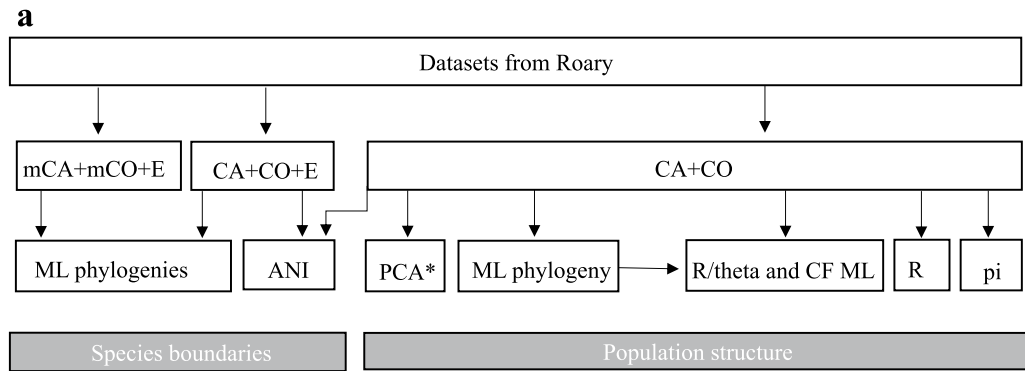
a

b

Dataset name	Content	# core genes	Best-fit model
CA + CO	Genes present in all 83 samples from Table S1	2370	GTR+F+R3
CA+CO+E	Genes present in all 83 samples from Table S1 and genomes from Table 3	51	TIM3+F+R4
mCA+mCO+E	Genes present in multi-fasta alignments of each CA and CO populations and genomes from Table 3	20	TIM+F+R10

Figure 5. Analyses pipeline using three core genome datasets (**a**) and the datasets description (**b**). (**a**) Arrows indicate what analyses were performed with corresponding datasets. ML phylogeny stands for maximum likelihood phylogenetic analysis with IQTREE v.2.0.3; ANI—average nucleotide identity estimated with JSpeciesWS; PCA—principal component analysis with Adegenet v. 2.1.3; Recombination—Split Decomposition Analysis for evidence of recombination with SplitTree; R/Ө and phylogenetic tree with corrected branch length (CF ML) with ClonalFrameML; pi—unbiased nucleotide diversity estimated with Pixy v.0.95.2. (**b**) Datasets were obtained with Roary v.3.13.0. Best-fit model for phylogenetic analyses was selected with ModelFinder. *PCA was conducted on thinned CA + CO dataset (see Materials and methods for details).

Family	Species	Strain	Accession No.	GenBank No.
Enterobacteriaceae	*Cronobacter sakazakii*	CS-931	ASM351612v3	GCA_003516125
Erwiniaceae	*Erwinia amylovora*	CFPB1430	ASM9156v1	GCA_000091565
	Erwinia piriflorinigrans	CFBP 5888	ASM105051v1	GCA_001050515
	Erwinia toletana	WS4403	ASM1787546v1	GCA_017875465
Pectobacteriaceae	*Brenneria goodwinii*	FRB141	ASM229144v1	GCA_002291445
	Lonsdalea britannica	477	ASM351598v1	GCA_003515985
	Lonsdalea iberica	LMG 26264	ASM211158v1	GCA_002111585
	Lonsdalea populi	N-5–1	ASM1599946v1	GCA_015999465
	Lonsdalea quercina	ATCC 29281	JIBO01000001	JIBO01000001
	Lonsdalea quercina	NCCB 100490	JIBQ00000000	GCA_000688695
Yersiniaceae	*Gibbsiella quercinecans*	FRB97	ASM229142v1	GCA_002291425
	Rahnella victoriana	JZ-GX1	ASM2127628v1	GCA_021276285

Table 3. List of genomes of members of four families from order *Enterobacterales* ord. nov. used in this study.

Colorado (CO), and *Enterobacteriales* (E) samples (CA + CO + E dataset) and with multi-fasta alignments of core genes of each population (mCA + mCO + E dataset), respectively (Fig. 5b). The best-fit model (TIM3 + F + R4—for CA + CO + E dataset, TIM + F + R10—for mCA + mCO + E dataset, Fig. 5b) selected by ModelFinder[53], was used for the ascertainment bias correction (+ ASC) model[54]. Tree branch support was assessed via 1000 ultrafast bootstrap replicates[55] with hill-climbing nearest neighbor interchange search bootstrap optimization to reduce the risk of overestimating branch supports (-bnni) due to severe violations of the model. The concatenated trees were visualized with iTOL online tool v.5.7 (itol.embl.de[56]). Average nucleotide identity (ANI) among all strains was estimated with JSpeciesWS using a BLAST-based approach (ANIb)[57], using datasets CA + CO + E (51 core genes total) for comparisons among *L. quercina* populations from this study and other species of genus *Lonsdalea*, and CA + CO (2370 core genes total), and to estimate ANI among *L. quercina* strains analyzed in this study.

Population structure.   Since some of the strains were clonal (see results), further analyses were performed on the dataset containing 1 representative of each core genome genotype. Population structure was assessed with principal component analysis (PCA) and phylogenetic analysis. PCA was conducted using 2355 thinned SNPs retrieved from core genes of CA + CO dataset. SNPs were extracted with SNP-sites software[58] followed by filtering with VCFtools v.0.1.16[59]: 1 SNP in every 1000 SNPs (–thin 1000) with minimal frequency 0.05 (–maf 0.05). Analysis was run within R package Adegenet v.2.1.3[60]. An unrooted phylogenetic tree of CA + CO dataset was built to assess population structure within samples obtained in this study. The phylogenetic analysis was conducted as described above, with the best-fit model GTR + F + R3 (Fig. 5b). Unbiased nucleotide diversity (pi) was estimated with Pixy v.0.95.02[61].

Recombination.   Evidence of recombination was obtained with Split Decomposition analysis and the Phi test implemented in SplitsTree[23]. ClonalFrameML[62] was used to estimate relative rates of recombination to mutation (R/θ) in each CA and CO populations and to construct phylogeny with branch lengths corrected to account for recombination. Maximum likelihood phylogenetic tree with CA + CO dataset was used as input for ClonalFrameML.

Ethics statement.   We declare that collection of plant material for this study complies with the relevant institutional, national, and international guidelines and legislations. There were no rare plants used in this study. All symptomatic plant material in California was collected from native to the region, commonly growing oak species *Quercus agrifolia*, *Q. parvula*, *Q. wislizeni*. In Colorado, symptomatic plant material was collected from commonly planted as shade trees *Quercus rubra* and three plants of wild *Q. gambelii*, an oak shrub commonly growing on the side of roads in the area. Moreover, all methods listed in this study were carried out in accordance with relevant guidelines and regulations.

Data availability
Genome sequence data used in this study have been deposited to the NCBI Sequence Read Archive under project number PRJNA924752 (SAMN32769485 to SAMN32769567). Three datasets used in this study are available on Dryad https://doi.org/10.5061/dryad.m63xsj46h.

References
1. Ghelardini, L., Pepori, A. L., Luchi, N., Capretti, P. & Santini, A. Drivers of emerging fungal diseases of forest trees. *For. Ecol. Manag.* **381**, 235–246 (2016).
2. Koskella, B., Meaden, S., Crowther, W. J., Leimu, R. & Metcalf, C. J. E. A signature of tree health? Shifts in the microbiome and the ecological drivers of horse chestnut bleeding canker disease. *New Phytol.* **215**, 737–746 (2017).
3. Adams, A. S. *et al.* Mountain pine beetles colonizing historical and naïve host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. *Appl. Environ. Microbiol.* **79**, 3468–3475 (2013).
4. Denman, S. *et al.* Microbiome and infectivity studies reveal complex polyspecies tree disease in Acute Oak Decline. *ISME J* **12**, 386–399 (2018).
5. Nixon, K. C. The oak (Quercus) biodiversity of California and adjacent regions. In: Standiford, Richard B., et al, tech. editor. *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Challenging Landscape*. Gen. Tech. Rep. PSW-GTR-184, Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture: 3–20 184 (2002).
6. Shrestha, B. B. *Quercus semecarpifolia* Sm. in the Himalayan region: Ecology, exploitation and threats. *Himal. J. Sci.* **1**, 126–128 (2003).
7. Dey, D. C. Sustaining oak forests in Eastern North America: Regeneration and recruitment, the pillars of sustainability. *For. Sci.* **60**, 926–942 (2014).
8. Thomsen, S., Reisdorff, C., Gröngröft, A., Jensen, K. & Eschenbach, A. Responsiveness of mature oak trees (*Quercus robur* L.) to soil water dynamics and meteorological constraints in urban environments. *Urban Ecosyst* **23**, 173–186 (2020).
9. Pautasso, M., Schlegel, M. & Holdenrieder, O. Forest health in a changing world. *Microb. Ecol.* **69**, 826–842 (2015).
10. Bueno-Gonzalez, V. *et al. Pseudomonas daroniae* sp. nov. and *Pseudomonas dryadis* sp. nov., isolated from pedunculate oak affected by acute oak decline in the UK. *Int. J. Syst. Evol. Microbiol.* **69**, 3368–3376 (2019).
11. Brady, C., Hunter, G., Kirk, S., Arnold, D. & Denman, S. Description of *Brenneria roseae* sp. nov. and two subspecies, *Brenneria roseae* subspecies *roseae* ssp. nov. and *Brenneria roseae* subspecies *americana* ssp. nov. isolated from symptomatic oak. *Syst. Appl. Microbiol.* **37**, 396–401 (2014).
12. Moradi-Amirabad, Y., Rahimian, H., Babaeizad, V. & Denman, S. *Brenneria* spp. and *Rahnella victoriana* associated with acute oak decline symptoms on oak and hornbeam in Iran. *For. Pathol.* **49**, 12535 (2019).
13. Sitz, R. A. *et al.* Drippy blight, a disease of red oaks in Colorado, U.S., produced from the combined effect of the scale insect *Allokermes galliformis* and the bacterium *Lonsdalea quercina* subsp. *quercina. Arboricul. Urban For.* **43**, 146–153 (2018).
14. Sitz, R. A., Aquino Vincent, M., Tisserat Ned, A., Cranshaw Whitney, S. & Stewart, J. E. Insects visiting drippy blight diseased red oak trees are contaminated with the pathogenic bacterium *Lonsdalea quercina. Plant Dis.* **103**, 1809–2147 (2019).
15. Hilderbrand, D. C. & Schroth, M. N. A new species of Erwinia causing the drippy nut disease of live oaks. *Phytopathology* **57**, 250–253 (1967).
16. Biosca, E. G. *et al.* Isolation and characterization of *Brenneria quercina*, causal agent for bark canker and drippy nut of *Quercus* spp. in Spain. *Phytopathology* **93**, 485–492 (2003).
17. Poza-Carrión, C. *et al. Brenneria quercina* and *Serratia* spp. isolated from Spanish oak trees: Molecular characterization and development of PCR primers. *Plant Pathol.* **57**, 308–319 (2008).
18. Brady, C. L. *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., emendation of the description of the genus *Brenneria*, reclassification of Dickeya dieffenbachiae as *Dickeya dadantii* subsp. dieffenbachiae comb. nov., and emendation of the description of *Dickeya dadantii. Int. J. Syst. Evol. Microbiol.* **62**, 1592–1602 (2012).

19. Tóth, T., Lakatos, T. & Koltay, A. *Lonsdalea quercina* subsp. *populi* subsp. nov., isolated from bark canker of poplar trees. *Int. J. Syst. Evol. Microbiol.* **63**, 2309–2313 (2013).

20. Li, Y. *et al.* A canker disease of populus × euramericana in China caused by *Lonsdalea quercina* subsp. *populi*. *Plant Dis.* **98**, 368–378 (2014).

21. Adeolu, M., Alnajar, S., Naushad, S. & S. Gupta, R. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for *Enterobacterales* ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **66**, 5575–5599 (2016).

22. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).

23. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

24. Margos, G. *et al.* Controversies in bacterial taxonomy: The example of the genus Borrelia. *Ticks Tick-borne Dis.* **11**, 101335 (2020).

25. Caballero, J. I. *et al.* Genome sequences of strain ATCC 29281 and pin and northern red oak isolates of *Lonsdalea quercina* subsp. *quercina*. *Genome Announc.* **2**, 10–1128 (2014).

26. 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.* **67**, 4680–4684 (2017).

27. Montes, M. S. *et al.* Population genetics of Phytophthora infestans in Denmark reveals dominantly clonal populations and specific alleles linked to metalaxyl-M resistance. *Plant. Pathol.* **65**, 744–753 (2016).

28. Shakya, S. K. *et al.* Phylogeography of the wide-host range panglobal plant pathogen Phytophthora cinnamomi. *Mol. Ecol.* **30**, 5164–5178 (2021).

29. Monteil, C. L. *et al.* Population-genomic insights into emergence, crop adaptation and dissemination of Pseudomonas syringae pathogens. *Microb. Genom.* **2**, e000089 (2016).

30. McCann, H. C. *et al.* Genomic analysis of the kiwifruit pathogen *Pseudomonas syringae* pv. actinidiae provides insight into the origins of an emergent plant disease. *PLoS Pathog.* **9**, e1003503 (2013).

31. Ghosh, D. K., Motghare, M. & Gowda, S. Citrus greening: Overview of the most severe disease of citrus (2018).

32. Almeida, R. P. P. & Nunney, L. How do plant diseases caused by *Xylella fastidiosa* emerge?. *Plant Dis.* **99**, 1457–1467 (2015).

33. Johnson, K. B. & Stockwell, V. O. Management of fire blight: A case study in microbial ecology. *Annu. Rev. Phytopathol.* **36**, 227–248 (1998).

34. Swiecki, T. J. & Bernhardt, E. A. *A field guide to insects and diseases of California oaks*. Gen. Tech Rep. PSW-GTR-197. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture. 151 p. 197, (2006).

35. Denman, S., Plummer, S., Kirk, S., Peace, A. & McDonald, J. E. Isolation studies reveal a shift in the cultivable microbiome of oak affected with Acute Oak Decline. *Syst. Appl. Microbiol.* **39**, 484–490 (2016).

36. Gathercole, L. A. P. *et al.* Evidence for the widespread occurrence of bacteria implicated in acute oak decline from incidental genetic sampling. *Forests* **12**, 1683 (2021).

37. Berruete, I. M. *et al.* First report of bark canker disease of poplar caused by *Lonsdalea quercina* subp. *populi* in Spain. *Plant Dis.* **100**, 2159–2159 (2016).

38. Di Filippo, A. *et al.* Climate change and oak growth decline: Dendroecology and stand productivity of a Turkey oak (*Quercus cerris* L.) old stored coppice in Central Italy. *Ann. For. Sci.* **67**, 706–706 (2010).

39. Sun, S. *et al.* The effect of climate change on the richness distribution pattern of oaks (*Quercus* L.) in China. *Sci. Total Environ.* **744**, 140786 (2020).

40. Mclaughlin, B. C. & Zavaleta, E. S. Predicting species responses to climate change: demography and climate microrefugia in California valley oak (*Quercus lobata*). *Glob. Change Biol.* **18**, 2301–2312 (2012).

41. Kolb, T. E. *et al.* Observed and anticipated impacts of drought on forest insects and diseases in the United States. *For. Ecol. Manag.* **380**, 321–334 (2016).

42. Doonan, J. M., Broberg, M., Denman, S. & McDonald, J. E. Host-microbiota-insect interactions drive emergent virulence in a complex tree disease. *Proc. Biol. Sci.* **287**, 20200956 (2020).

43. Shang, J., Liu, B. L. & He, W. A new method to detect Lonsdalea quercina in infected plant tissues by real-time PCR. *For. Pathol.* **45**, 28–35 (2015).

44. Woyda, R., Oladeinde, A. & Abdo, Z. Reads2Resistome: An adaptable and high-throughput whole-genome sequencing pipeline for bacterial resistome characterization. 2020.05.18.102715 Preprint at https://doi.org/10.1101/2020.05.18.102715 (2020).

45. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).

46. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

47. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

48. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

49. Sun, J. *et al.* OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* **51**, W397–W403 (2023).

50. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).

51. Urban, M. *et al.* PHI-base: the pathogen–host interactions database. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkz904 (2019).

52. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

53. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

54. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).

55. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

56. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

57. Richter, M., Rosselló-Móra, R., Oliver Glöckner, F. & Peplies, J. JSpeciesWS: A web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**, 929–931 (2016).

58. Page, A. J. *et al.* SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* **2**, e000056 (2016).

59. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

60. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

61. Korunes, K. L. & Samuk, K. PIXY: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).

62. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).

Acknowledgements

Author contributions

J.E.S., R.A.S., and I.S.P. designed the study, R.A.S. and I.S.P. collected the samples, R.A.S. and L.L. processed samples, OK analyzed data and wrote the manuscript, R.W. and J.I.C. contributed to data analyses and data visualization, J.E.S., R.A.S., I.S.P., O.K., R.W., J.I.C., and Z.A. contributed to editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-41976-8.

Correspondence and requests for materials should be addressed to J.E.S.

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.