High-resolution Profiling of Bacterial and Fungal Communities using Pangenome-Informed Taxon-Specific Amplicons and Long-Read Sequencing

Luzia Stalder[1], Monika Maurhofer[2,*], Daniel Croll[1,*]

[1] Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland.
[2] Plant Pathology, Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland.

[*] Correspondence: monika.maurhofer@usys.ethz.ch, daniel.croll@unine.ch

Author contributions: LS, MM and DC conceived the study, LS performed the research and analyzed the data, MM and DC supervised the work and acquired funding. LS and DC wrote the manuscript with input from MM.

Competing interests: the authors declare that no competing interests exist.

Keywords: Amplicon sequencing, long read sequencing, PacBio, Sequel II, multiplexing, pangenome, microbiome, wheat, phyllosphere, *Pseudomonas*, *Zymoseptoria tritici*

ABSTRACT

High-throughput sequencing technologies have greatly advanced our understanding of microbiomes, but resolving microbial communities at species and strain levels remains challenging. In this study, we present a pipeline for designing, multiplexing, and sequencing highly polymorphic taxon-specific amplicons using PacBio circular consensus sequencing. We focus on the wheat microbiome as a proof-of-principle and demonstrate unprecedented resolution for the wheat-associated *Pseudomonas* microbiome and the ubiquitous fungal pathogen *Zymoseptoria tritici*. Our approach achieves an order of magnitude higher phylogenetic resolution compared to existing ribosomal amplicons. We show that the designed amplicons accurately capture species and strain diversity outperforming full-length 16S and ITS amplicons. We track microbial communities in the wheat phyllosphere across time and space to establish fine-grained species and strain-specific dynamics. The high-resolution profiling provided by taxon-specific amplicons opens new avenues for studying microbial community dynamics in complex environments. Moreover, the cost-effectiveness and scalability of our approach make it applicable to diverse ecological studies. Overall, our work demonstrates the potential of long-read amplicon sequencing to overcome limitations in phylogenetic resolution and enhance our understanding of microbiome diversity and dynamics.

INTRODUCTION

High-throughput sequencing technologies (HTS) have revolutionized our understanding of microbiomes revealing insights into structures, functions and dynamics. Microbial community dynamics are influenced by various factors such as the identity of associated hosts, health status, and environmental variables (Vorholt *et al.*, 2017; Fitzpatrick *et al.*, 2020; Trivedi *et al.*, 2020; Choi, Khan and Lee, 2021; Song *et al.*, 2021; Xu *et al.*, 2021). Targeted amplicon sequencing is the most commonly employed HTS method for microbiome analyses interrogating conserved loci such as 16S ribosomal DNA for prokaryotes and internal transcribed spacers (ITS) for eukaryotes. Amplicon sequencing is widely applicable and overcomes challenges such as low input biomass or contamination by host DNA, making amplicon sequencing particularly suitable for plant microbiome analyses where contaminating DNA is often present in substantial amounts (Liu *et al.*, 2021). In contrast, untargeted metagenomic sequencing and assembly can only provide rich information about microbial communities if the targeted organisms are abundant (Liu *et al.*, 2021). A major limitation of amplicon sequencing has been the lack of resolution below the genus level in most studies based on either 16S or ITS (Earl *et al.*, 2018; Liu *et al.*, 2021).

Long-read sequencing based on PacBio or Oxford Nanopore opened the possibility to sequence long amplicons (Karst *et al.*, 2021). PacBio circular consensus sequencing (CCS) generates highly accurate (>99.9%) reads comparable to Illumina and Sanger sequencing but reaching up to 20 kb in length (Wenger *et al.*, 2019). Studies based on PacBio sequencing of full-length 16S or ITS sequences achieved high taxonomic resolution (Wagner *et al.*, 2016; Earl *et al.*, 2018; Callahan *et al.*, 2019; Escapa *et al.*, 2020). Yet, increased resolution would often be necessary to define functional roles of microbiome components. For example, closely related species of the ubiquitous *Pseudomonas* bacteria can have drastically different effects on plant microbiome functions. Certain *P. syringae* strains are pathogenic, but *P. fluorescens*, *P. protegens*, or *P. putida* species generally promote plant growth (Großkinsky *et al.*, 2016; Xin, Kvitko and He, 2018). Closely related Pseudomonas strains can also differ significantly in volatiles and metabolites production underpinning distinct niche exploitation in the environment (Passera *et al.*, 2019). Despite the progress, achieving species and strain-level resolution remains challenging though. Thus, sequencing approaches capturing long and highly polymorphic regions beyond 16S and ITS loci are needed to reveal microbiome diversity at an ecologically relevant resolution. Furthermore, multiplexing large numbers of samples and multiple amplicons is required to tackle complex experimental designs and deep analyses of microbiomes across environments.

In this study, we present a pipeline for designing, multiplexing, and sequencing highly polymorphic taxon-specific amplicons using PacBio CCS. Our new approach allows to track complex bacterial and fungal communities at the species and strain level from environmental samples. As a proof-of-principle,

we focus on the microbiome of wheat (Laino *et al.*, 2015). The fungal microbiome of wheat is often dominated by the pathogenic fungus *Zymoseptoria tritici* causing septoria tritici blotch (Kerdraon *et al.*, 2019; Barroso-Bergada *et al.*, 2022). The bacterial microbiome is dominated by Gammaproteobacteria including the genus *Pseudomonas* both above and below ground (Chen *et al.*, 2021; Kavamura *et al.*, 2021). The fungal pathogen *Z. tritici* significantly impacts the composition of microbial communities associated with wheat (Kerdraon et al., 2019). For instance, *Z. tritici* suppresses the host immune system to facilitate colonization by strains of the *P. syringae* group (Seybold *et al.*, 2020), while specific *P. fluorescens* strains inhibit the growth of *Z. tritici* (Levy, Eyal and Chet, 1988).

In this study, we introduce a pipeline to establish and validate a novel suite of highly multiplexed amplicons with over an order of magnitude higher phylogenetic resolution compared to existing ribosomal amplicons. As a proof-of-principle, we target the wheat-associated *Pseudomonas* microbiome and the major fungal pathogen *Z. tritici*. We achieve unprecedented species and strain-level resolution for both groups in mixed samples. We highlight the substantial gains in phylogenetic resolution by tracking strains across the wheat canopy and over time.

RESULTS

PANGENOME-INFORMED DESIGN OF TAXON-SPECIFIC AMPLICONS

We developed new 3-kb long amplicons to enhance the resolution of bacterial and fungal community profiling while maintaining universal amplification within the targeted group of organisms (Figure 1A). To identify highly polymorphic *Pseudomonas* amplicons, we constructed a comprehensive pangenome of 18 high-quality genomes representing all subgroups of the genus. From this analysis, we identified 28 core regions conserved in the *Pseudomonas* pangenome. These core regions served as candidate regions for primer development. For each core fragment, we designed all possible amplicons ranging from 2.7-3.2 kb resulting in 224 amplicon candidates. Nucleotide diversity in aligned core regions was used as a metric to prioritize amplicon candidates.

We selected the ten most polymorphic amplicon candidates for PCR evaluation using both *Pseudomonas* reference strains and naturally colonized wheat leaf samples. Candidate primer sequences were adjusted based on alignment against all available sequences of *Pseudomonas* strains to maximize recovery of the *Pseudomonas* diversity. We allowed primer candidates to include up to five degenerate bases. To reduce the number of sequence variants to be considered in primer candidates, we reduced degenerate positions to match known combinations in *Pseudomonas* genomes. Finally, we selected two amplicons that consistently showed the strongest PCR amplification for further analyses. The first amplicon spanned a locus including the *rpoD* gene, previously utilized for taxonomic classification

(Gomila *et al.*, 2015; Peix, Ramírez-Bahena and Velázquez, 2018). The second amplicon encompassed genes encoding an ABC transporter (APE98195.1) and an ABC transporter permease (APE98194.1), respectively. We pursed a parallel approach to identify primers suitable for amplifying intra-specific diversity of the fungal pathogen *Z. tritici*. We based our analysis on a recently established reference-quality global pangenome for the species (Badet *et al.*, 2020). The two best performing amplicons were located on chromosome 9 and 13, respectively.

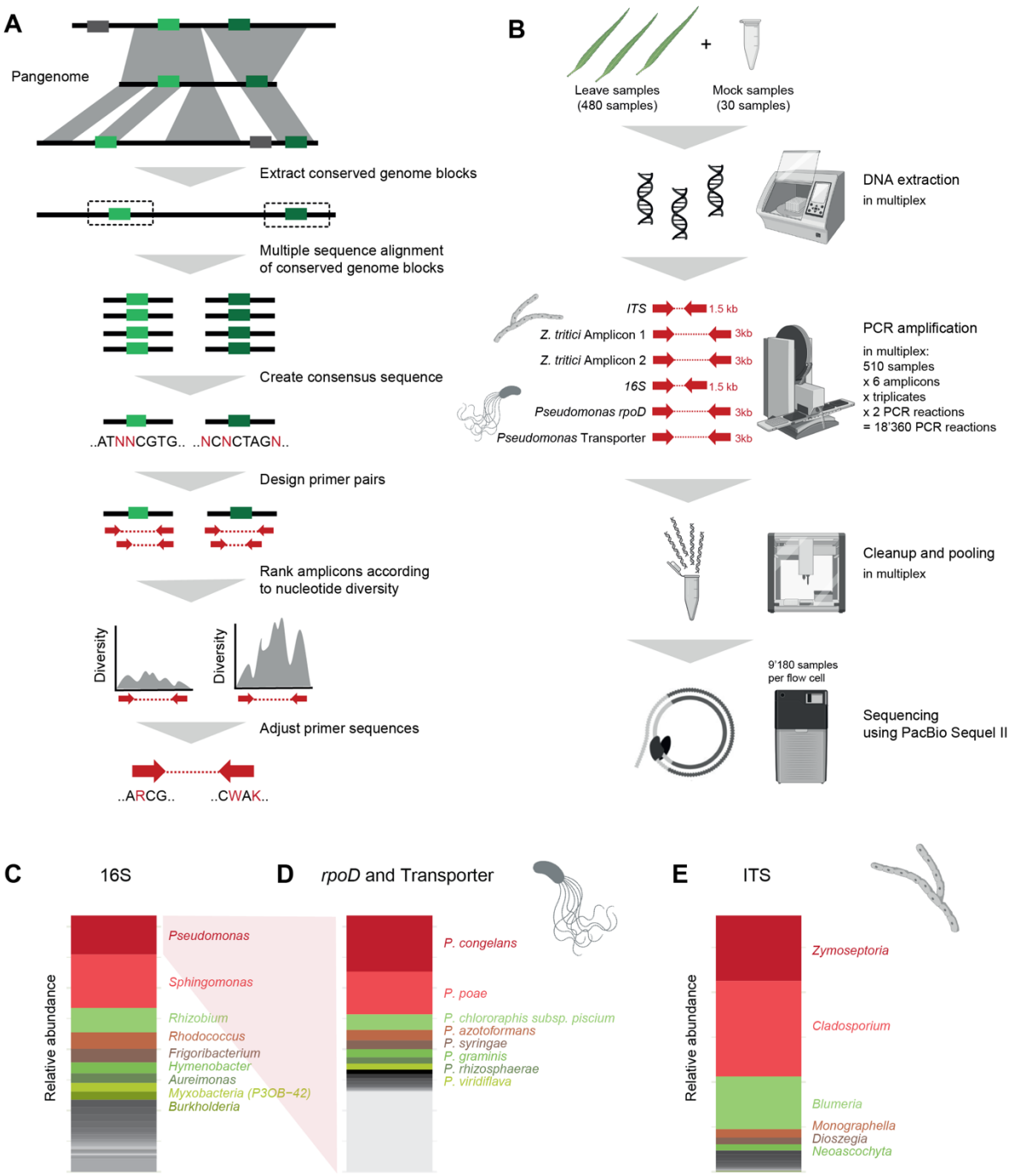


Figure 1: Pan-genome informed design of taxon-specific amplicons. (A) Schematic overview of amplicon design pipeline. Conserved blocks in the *Pseudomonas* and *Z. tritici* pangenomes were extracted. Consensus sequences of the core blocks were used to design amplicons. Amplicon candidates were ranked according to

nucleotide diversity, and primer sequences were adjusted. Polymorphic amplicons were tested on reference strains and environmental samples. (B) Overview of multiplexed amplicon evaluation on environmental samples. DNA from mock samples and wheat leave samples were automatically extracted. Two *Pseudomonas*-specific amplicons, two *Z. tritici*-specific amplicons, as well as the full-length 16S and ITS were amplified in each sample and prepared using robotic liquid handling for accuracy and throughput. Reactions were performed in triplicates and barcoded individually using asymmetric M13 barcodes resulting in a total of 18'360 PCR reactions. All reactions were pooled and sequenced in two PacBio Sequel II 8M cells. (C) Relative abundances of bacterial genera on wheat leaves determined by the full-length 16S amplicon. (D) Relative and mean abundances of *Pseudomonas* species on wheat leaves assessed by the *Pseudomonas*-specific *rpoD* and transporter amplicons. (E) Relative abundances of fungal genera on wheat leaves determined by the full-length ITS amplicon.

To evaluate the performance of the two *Pseudomonas*-specific and two *Z. tritici*-specific amplicons, we conducted tests using an extensive mock community of well-characterized laboratory strains. In addition to the new amplicons, we amplified the full-length 16S and ITS loci to contrast performance on the same sample pools. Additionally, we tested the *Pseudomonas* and *Z. tritici*-specific amplicons on a diverse set of wheat leaf samples collected at five timepoints during the growing season at an experiment field site near Zurich, Switzerland (Figure 1B). We optimized DNA extraction methods to maximize bacterial and fungal yield from wheat leaves. Subsequently, we amplified the two selected *Pseudomonas*-specific amplicons and the two *Z. tritici*-specific amplicons in 30 mock and 480 wheat leaf samples. Each reaction was performed in triplicate, resulting in 9,180 amplifications. We tagged each sample with a unique asymmetric M13 barcode combination. Liquid handling was performed with a robot and samples were processed in a fully randomized layout. We sequenced all 9,180 reactions on two PacBio Sequel II flow cells resulting in a total of 1.5M high-quality reads used in all subsequent analyses (Supplementary Table 1).

The wheat leaf samples revealed a diverse assembly of bacteria using the full-length 16S with *Pseudomonas* and *Sphingomonas* being the dominant genera (Figure 1C and Supplementary Figure S1). The *Pseudomonas*-specific *rpoD* and transporter amplicons revealed a diverse assembly of species, primarily belonging to the *P. fluorescens* and *P. syringae* groups (Figure 1D and Supplementary Figure S1). The mycobiome analyzed using the ITS was dominated by the fungal genera *Zymoseptoria* and *Cladosporium*, broadly consistent with recent findings in wheat fields (Figure 1E and Supplementary Figure S1) (Barroso-Bergada *et al.*, 2022).

DETECTION LIMITS OF TAXON-SPECIFIC AMPLICONS

To assess the performance and detection limits of the new amplicons, we analyzed both defined mock communities and dilution series. For the *Pseudomonas* amplicons, we examined a panel of ten isolates representing the phylogenetic diversity of the genus (Supplementary Table 2). Our results showed that both the full-length 16S and the *Pseudomonas*-specific *rpoD* and transporter amplicons correctly distinguish all isolates (Figure 2A). All isolates consistently amplified for all three amplicons, except for the *P. putida* Leaf59 isolate failing for the transporter amplicon. No genome sequence is available for verification, but amplification failure may be caused by primer mismatches. The *rpoD* and transporter amplicons exhibited clustering of the eight *P. fluorescens* group isolates according to their subclade described by (Flury *et al.*, 2016). While the *rpoD* and transporter amplicons predominantly produced a single amplicon sequence variant (ASV) per culture, the 16S amplicon exhibited multiple ASVs per culture. This can be attributed to the multi-copy nature of the 16S gene (Johnson *et al.*, 2019).

To evaluate amplicon performance in more complex strain mixtures, we created equimolar mixtures containing two to ten *Pseudomonas* isolates. The 16S, the *rpoD* and the transporter amplicons successfully recovered ASVs from individual isolates even in complex mixtures (Figure 2B). However, the *rpoD* amplicon showed reduced amplification efficiency for the *P. thivervalensis* isolate PITR in the mixture. The transporter amplicon exhibited lower amplification efficiency for the *P. fluorescens* isolate SS101. Furthermore, we observed greater variation in abundance between replicates for the 16S amplicon compared to the *rpoD* and transporter amplicons.

We managed unprecedented levels of multiplexing for microbial barcoding and recovered, hence, relatively low read numbers per sample. To assess detection limits in our experimental setup, we analyzed a two-strain dilution series consisting of the *P. syringae* Leaf129 and *P. thivervalensis* PITR isolates (Figure 2C, Supplementary Figure S2, Supplementary Table 3). Both *Pseudomonas*-specific amplicons identified the diluted isolates down to a concentration of 10^{-3}, while the 16S amplicon detected the diluted isolates down to a concentration of 10^{-2} (Figure 2C). Similarly, we performed a comparable analysis for the fungal pathogen *Z. tritici* using a two-strain dilution series comprising *Z. tritici* isolates 1E4 and IR01_48b (Figure 2C, Supplementary Figure S2, Supplementary Table 3). The ITS and *Z. tritici*-specific amplicons successfully detected diluted strains down to a concentration of 10^{-2}. Polyphenols contained in plant DNA extracts can inhibit amplification. To test for this, we replicated the dilution series by diluting reference isolates with leaf samples. We obtained similar amplification yields from culture samples compared to plant samples indicating that the remaining plant extracts did not significantly affect the amplification (Figure 2D, Supplementary Figure S2, Supplementary Table 3).

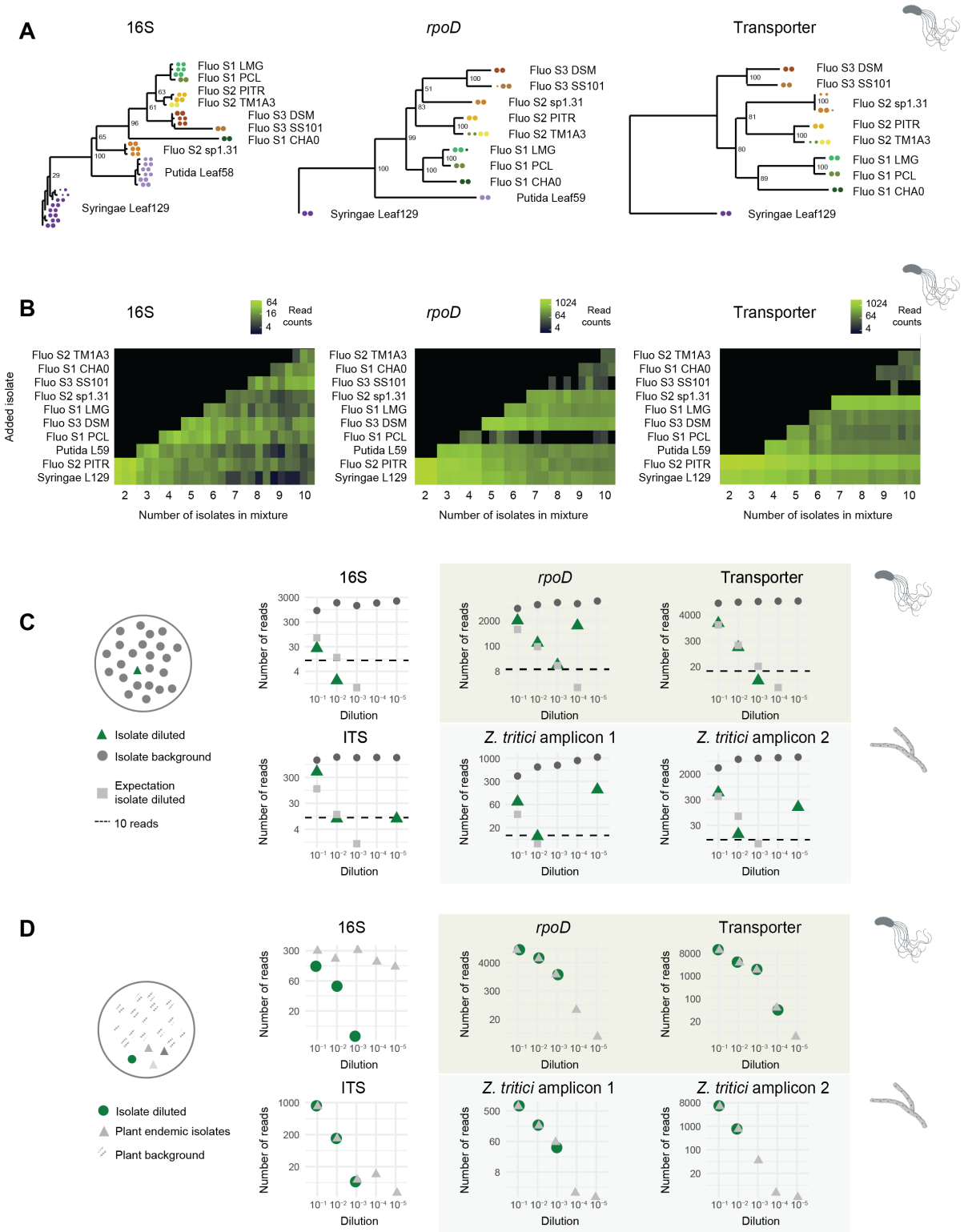


Figure 2: Evaluation of detection limit of taxon-specific amplicons using mock communities. (A) Phylogenetic trees of amplified sequence variants (ASVs) from ten individual *Pseudomonas* cultures for the full-length 16S amplicon, the *Pseudomonas*-specific *rpoD* and transporter amplicon. Colors indicate different cultures, and the dot size represents the abundance of the ASVs individually for each of the two replicates. For the isolates from the *P. fluorescens* group, subclades S1, S2 and S3 are labelled according to Flury et al., 2016. The transporter amplicon failed to amplify the *P. putida* Leaf59 isolate. (B) Mixed equimolar combinations of the ten *Pseudomonas* cultures showing read counts of the main ASV for each mixture. The three replicates are shown individually. (C) Top panel: Serial dilution series of *P. syringae* Leaf129 (green) mixed into

P. thivervalensis S2 PITR (grey). Read numbers are shown for the full-length 16S, and the *Pseudomonas*-specific *rpoD* and transporter amplicons. The expected read numbers of the diluted *P. syringae* Leaf129 isolate are indicated in light grey and calculated from the background read numbers of *P. thivervalensis* S2 PITR. Bottom panel: Serial dilution series of *Z. tritici* 1E4 (green) in *Z. tritici* IR01_48b (grey) isolates. Read numbers are shown for the full-length ITS, and the *Z. tritici*-specific amplicons on chromosomes 9 and 13. The expected read numbers of the diluted *Z. tritici* 1E4 isolate are indicated in light grey and calculated from the background read numbers of the *Z. tritici* IR01_48b isolate. (D) Top panel: Read numbers of a serial dilution series of *P. thivervalensis* S2 PITR (green) mixed into a wheat leaf DNA extract background (grey) for the full-length 16S amplicon, and the *Pseudomonas*-specific *rpoD* and transporter amplicons. Bottom panel: Read numbers of a serial dilution series of *Z. tritici* 1E4 (green) mixed into a wheat leaf DNA extract background (grey) for the full-length ITS and the *Z. tritici*-specific amplicons.

DISCRIMINANT POWER OF AMPLICONS TO RESOLVE STRAIN GENOTYPES

To assess gains in phylogenetic resolution of the wheat microbiome, we compared the *Pseudomonas* and *Z. tritici* amplicons to 16S and ITS amplicons, respectively. *Pseudomonas*-specific amplicons revealed 933 and 538 ASVs at the *rpoD* and transporter locus, respectively. In contrast, the 16S amplicon revealed only 86 ASVs matching the genus *Pseudomonas* (Figure 3A and B, Supplementary Figure S3). This represents a three-fold (2.7X for *rpoD* and 3.3X for the transporter) increase in ASVs for the *Pseudomonas*-specific amplicons compared to the 16S and based on relative number of reads per amplicon (Figure 3B).

We assigned all ASVs to *Pseudomonas* species using 1071 available genomes from nine different groups for BLASTn analyses (Supplementary Table 4). The *Pseudomonas*-specific amplicons produced significantly better matches for species assignment compared to 16S sequences (Figure 3C). To assess the consistency of species assignments, we extracted the smallest subtree containing >90% of the ASVs for each *Pseudomonas* species and examined the proportion of the species within this subtree, as well as the Shannon diversity index of the ASVs matching each species (Figure 3D-F). *Pseudomonas*-specific amplicons showed higher proportions and lower Shannon diversity indices compared to the 16S as expected for more accurate species assignments. For this analysis we considered four species for the 16S, 14 for *rpoD*, and ten for the transporter amplicon, where the low species number for 16S stems from the requirement of at least five detected ASVs per species (Supplementary Figure S4). Among the species resolved by the *rpoD* and the transporter amplicons, but failing with the 16S, were subspecies of *P. chlororaphis* and various species from the *P. syringae* group (Figure 3A and Supplementary Figure S3).

For *Z. tritici*, we found 869 and 757 ASVs for the chromosome 9 and 13 amplicon, respectively, compared to 13 ASVs obtained by ITS (Figure 3G and H). Based on the total number of reads per amplicon, this represents a more than tenfold increase of ASVs for the *Z. tritici*-specific amplicons. These findings indicate a remarkably high strain diversity within a single field in agreement with previous studies (Singh, Karisto and Croll, 2021; McDonald *et al.*, 2022).

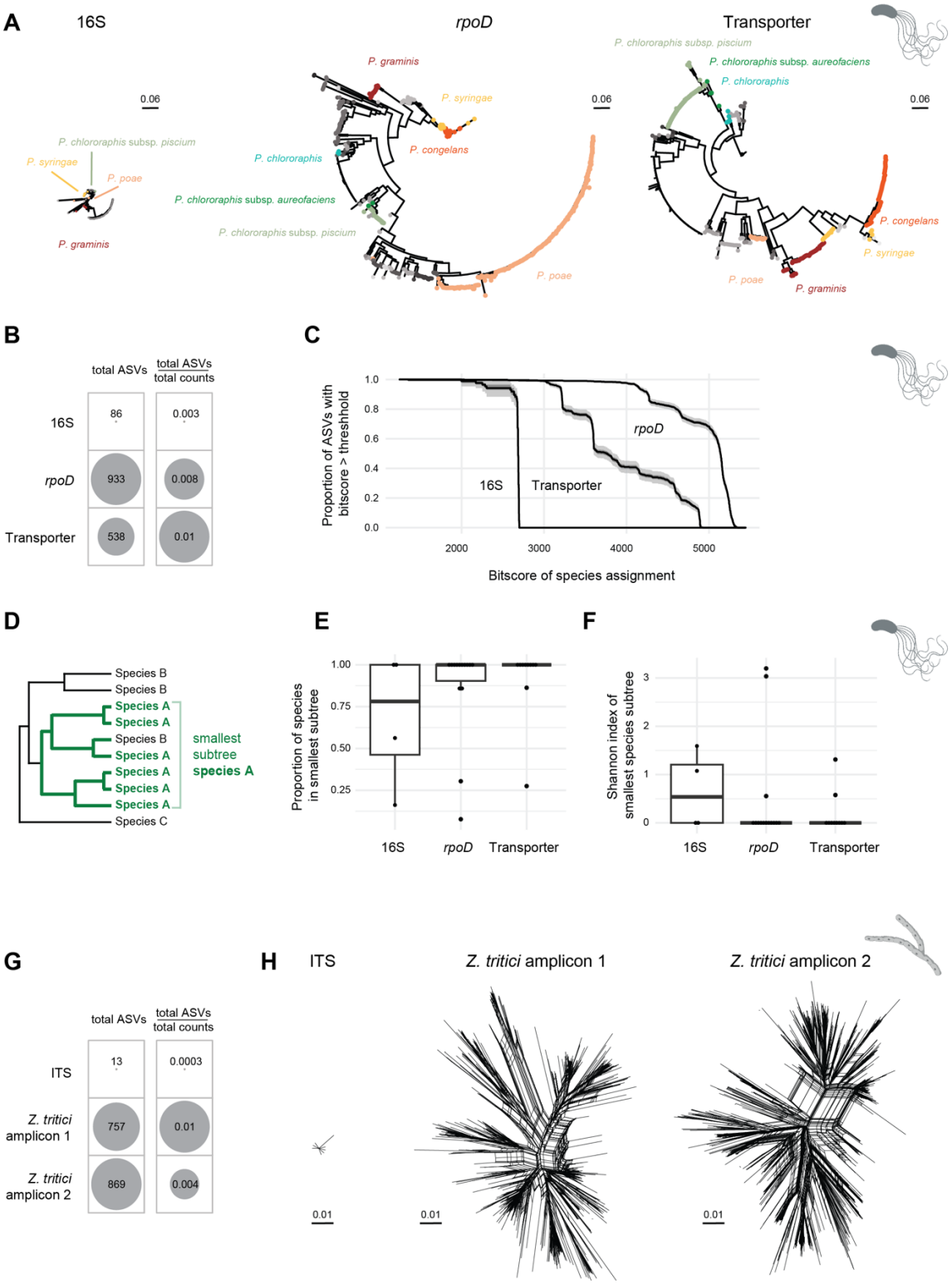


Figure 3: Phylogenetic resolution of taxon-specific and ribosomal DNA amplicons. (A) Phylogenetic trees of *Pseudomonas* amplified sequence variants (ASVs) of the full-length 16S amplicon as well as the *Pseudomonas*-specific *rpoD* and transporter amplicons. ASVs are colored by *Pseudomonas* species. (B) Total number of ASVs detected by the full-length 16S amplicon, and the *Pseudomonas*-specific *rpoD* and transporter amplicons relative to the total read count per amplicon. (C) Proportion of ASVs with a BLASTn bitscore above a threshold for *Pseudomonas* species assignment. The 50% and 95% confidence intervals based on permutations are shown in grey. (D) Identification of the smallest subtree containing >90% of the ASVs for each *Pseudomonas* species shown in (E) and (F). Only species with >5 ASVs were considered. (F) Proportion of species in the smallest

subtree containing >90% of the species ASVs assigned to specific species. (E) Shannon diversity indices of the subtree species composition for the smallest subtree containing >90% of the ASVs assigned to specific species. (G) Total number of ASVs detected by the full-length ITS amplicon, and *Z. tritici*-specific amplicons on chromosome 9 and 13 relative to the total read count per amplicon. (H) Phylogenetic networks (Splitstree) of *Z. tritici* ASVs identified using the full-length ITS amplicon, and the *Z. tritici*-specific amplicons.

HIGH-RESOLUTION TRACKING OF NATURAL PLANT COLONIZATION BY PSEUDOMONADS

A major aim of high-resolution microbiome analyses is to resolve strain-level interactions. To assess the power of the newly developed amplicons, we performed a hierarchical sampling across space and time tracking expected microbiome shifts in a wheat field. Sampling was conducted at five different timepoints during the wheat growing season, from May (first node appearance) to July (prior to harvest). We sampled eight different wheat cultivars replicated each in two plots. Furthermore, leaves were sampled at three different canopy heights to capture developmental patterns of the plant. In total, 480 samples were collected.

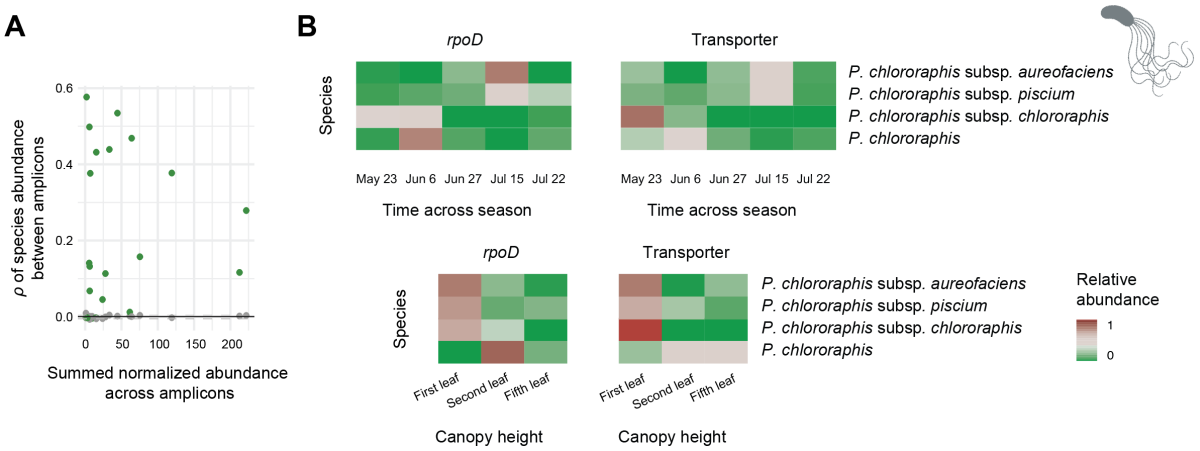


Figure 4: Differential abundance of wheat phyllosphere *Pseudomonas* species across season and canopy height. (A) Spearman correlation of *Pseudomonas* species abundance across samples between the *Pseudomonas*-specific *rpoD* and transporter amplicons. Correlation values were compared to null expectations based on permutations with the grey dotted line indicating the 95% confidence interval of the null expectation. (B) Relative abundances of *P. chlororaphis* subspecies across time and canopy heights revealed by the *Pseudomonas*-specific *rpoD* and transporter amplicons.

We first examined whether *Pseudomonas* species abundance patterns were consistently reproduced by both the *rpoD* and transporter amplicons. For this, we calculated correlation coefficients of *Pseudomonas* species abundance across samples and between the two amplicons (Figure 4A and Supplementary Figure S5). We compared correlation values to null expectations based on permutations. We observed that 68% and 77% using Spearman and Pearson correlations, respectively, of the species correlations were higher than the 95% confidence interval of the null expectation.

286 Next, we investigated species abundance changes across time and space. We identified different
287 subspecies of *P. chlororaphis* exhibited variable abundance patterns throughout the season and across
288 canopy heights in both amplicons (Figure 4B). *P. chlororaphis* subsp. *aurofaciens* and subsp. *piscium*
289 were most abundant in July, whereas *P. chlororaphis* and *P. chlororaphis* subsp. *chlororaphis* were
290 most abundant in May and June. *P. chlororaphis* subsp. *aurofaciens*, subsp. *piscium* and subsp.
291 *chlororaphis* were most prominent at the bottom of the canopy, whereas *P. chlororaphis* was most
292 abundant on upper leaves.

293 WITHIN-SPECIES DIVERSITY OF A PATHOGEN

294 Genetic diversity within pathogen species can underpin rapid breakdowns of fungicide efficacy or host
295 resistance (Lucas, Hawkins and Fraaije, 2015; Ebert and Fields, 2020). How genotypic diversity
296 changes over pathogen life cycles remains largely unknown though (Barrett *et al.*, 2008; Tack *et al.*,
297 2012). Here, we examined changes in genotypic diversity across the epidemic phase for the fungal
298 pathogen *Z. tritici* based on species-specific amplicons. The two *Z. tritici*-specific amplicons showed
299 consistent numbers of ASVs per leaf throughout the season, indicating minimal turnover in pathogen
300 diversity within a single epidemic phase (Figure 5A). Similarly, the number of ASVs did not differ
301 significantly across different canopy heights (Figure 5A). Tracking individual ASVs across the season,
302 we observed two distinct groups based on their abundance patterns (Figure 5B). The first group
303 consisted of a small number of highly abundant strains persisting throughout the entire growing season.
304 In contrast, the second group consisted of numerous strains that are predominantly scarce and only
305 detected at specific timepoints.

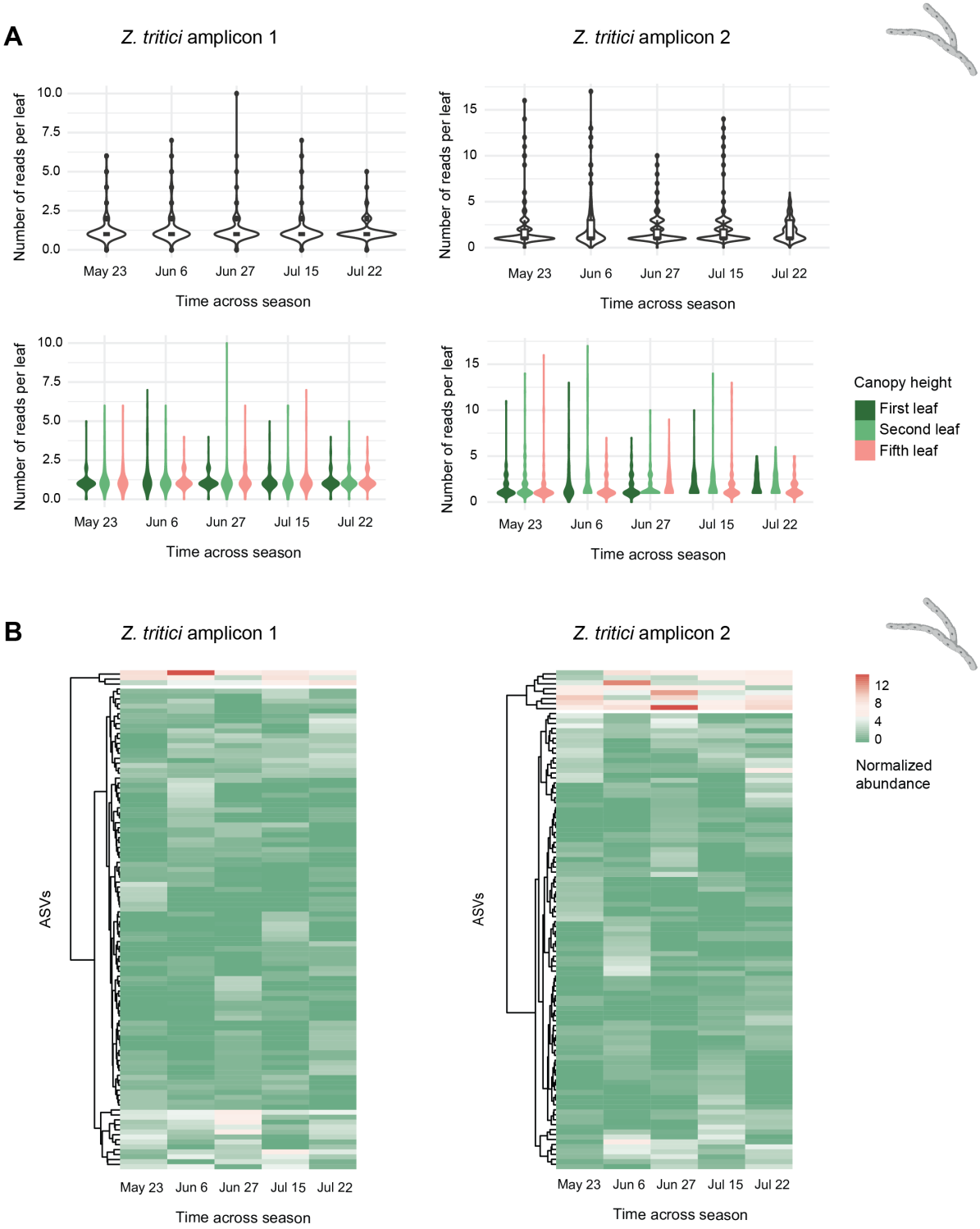


Figure 5: Genotypic diversity of the wheat phyllosphere pathogen *Z. tritici* across season. (A) Number of unique *Z. tritici* amplicon sequence variants (ASVs) per leaf across timepoints based on the *Z. tritici*-specific amplicons on chromosome 9 and 13. The same analysis is shown highlighting canopy differentiation. (B) Heatmap of the 100 most abundant *Z. tritici* ASVs clustered based on their abundance pattern across the season.

311 # DISCUSSION

312 We demonstrate that limitations in phylogenetic resolution of microbial community profiling can be
313 overcome using systematically designed taxon-specific 3-kb amplicons. We find that the new loci
314 provide species and strain-level insights into subsets of bacterial and fungal communities coexisting in
315 the plant microbiome exceeding full-length 16S or ITS amplicons by an order of magnitude.

316 The pangenome-informed design of amplicons for pseudomonads was optimized to capture diversity
317 specifically and exclusively in this ubiquitous group of bacteria. Compared to full-length 16S
318 amplicons, the new amplicons provide consistently higher confidence in species assignments across
319 clades. The confidence in species assignments stems from the high degree of species resolved to
320 monophyletic clades for most *Pseudomonas* species. This enhanced resolution enabled us to
321 differentiate *Pseudomonas* at the subspecies level, which is not feasible using 16S alone. We also
322 focused on the dominant wheat phyllosphere pathogen *Z. tritici* and achieved a strain-level resolution
323 inaccessible with existing single-locus genotyping approaches. The recovered diversity reflects
324 genotyping resolution typically only achievable by whole-genome sequencing (Singh, Karisto and
325 Croll, 2021; McDonald *et al.*, 2022).

326 We evaluated the discriminant power of the new bacterial and fungal amplicons by tracking microbial
327 communities in the wheat phyllosphere. We contrasted the resolution to the universal barcoding loci
328 16S and ITS in samples covering different time points, cultivars, and canopy heights. The increased
329 resolution indeed revealed species and strain-specific changes across time and canopy height
330 inaccessible by ribosomal barcoding datasets. One noteworthy finding was the differential abundance
331 patterns observed for the *P. chlororaphis* subspecies. Strains vary in their expression of phenotypic
332 traits and can act as powerful biocontrol against pathogens (Anderson and Kim, 2020; Raio and
333 Puopolo, 2021). We were also able to resolve genotypic diversity in a highly diverse fungal pathogen
334 and track changes throughout a single epidemic season.

335 The replicated monitoring of *Pseudomonas* and plant pathogen diversity with two independent
336 amplicons each enabled us to assess the reproducibility of species and strain identification as well as
337 abundance estimates. Certain *Pseudomonas* species were over- or underrepresented in a particular
338 amplicon and across samples. Such discrepancies could stem from differences in primer binding
339 efficiency and variations in amplification success due to amplicon sequence variability (Yeh *et al.*,
340 2021). Our findings align with previous studies investigating reproducibility of different 16S amplicons
341 (Wear *et al.*, 2018). This underscores the importance of implementing controls in quantitative
342 microbiome studies. The availability of two independent 3-kb amplicons for *Pseudomonas* in addition
343 to a control by the 16S provides a powerful toolset to track the reproducibility of microbiome
344 community assessments across samples and taxonomic groups.

The high quality of PacBio HiFi reads enables microbial community profiling using long amplicons with a high level of resolution. With a large number of high-quality genomes available, leveraging pangenomes to perform more integrative analyses of genetic diversity within taxonomic groups and identify the most informative amplicons. Although pangenomes have been increasingly used to analyze genetic diversity, to the best of our knowledge, they have not been utilized for amplicon design (Brockhurst *et al.*, 2019; Tettelin and Medini, 2020; Cummins *et al.*, 2022). With the rapidly growing number of genomes, this approach holds broad applicability.

In our proof-of-principle demonstration we multiplexed nearly 10,000 reactions in single sequencing runs. This achievement was made possible by the recent improvements in Sequel II. The sequence quality suffices to recover sequences at a very high rate based on asymmetric barcoding, which allows to multiplex nearly 10,000 samples using the standard set of barcodes. Additionally, we improved sample handling efficiency and reproducibility through robot liquid handling.

Overall, our approach yielded similar costs per high-quality sequence as the commonly used 16S and ITS Illumina-based amplicon sequencing. This gain in throughput opens cost-effective avenues to explore microbial community dynamics with complex experimental designs and in heterogeneous environments. Furthermore, designing, optimizing, and multiplexing long amplicons for other ecologically relevant taxa is clearly in reach with the availability of high-quality genomes. In conclusion, our work shows that long-read amplicon sequencing based on purpose-designed taxon-specific amplicons overcomes limitations in phylogenetic resolution associated with ribosomal amplicon sequencing.

METHODS

WHEAT LEAVES SAMPLING

We analyzed eight elite European winter wheat (*Triticum aestivum*) varieties sampled at five different timepoints over the growing season (i.e. cultivars Aubusson, Arobase, Lorenzo, CH Nara, Zinal, Simano, Forel and Titlis on the 23.05.2019, 06.06.2019, 27.06.2019, 15.07.2019 and 22.07.2019, respectively). Two biological replicates of the wheat panel were grown in two complete block designs separated by approximately 100m at the field phenotyping platform site of the Eschikon Field Station of the ETH Zurich, Switzerland (coordinates 47.449°N, 8.682°E). The cultivars were grown in plots of 1.2-by-1.7-m, with the genotypes arranged randomly within each block. No fungicides were applied. For each cultivar, block and timepoint, two plants were collected. From each plant three leaves were collected: the bottom leaf touching the ground (first leaf), the lowest leaf not touching the ground (second leaf) and then the top flag leaf (fifth leaf). Each leaf was immediately stored in a plastic foil to

avoid contamination, stored at 4°C overnight before processing. In total, 480 leaves were collected (8 cultivars x2 blocks x5 timepoints x2 plants x3 leaves).

MOCK COMMUNITY CREATION

To establish three sets of mock communities, we used ten different *Pseudomonas* strains and two *Z. tritici* isolates (Supplementary Table 2). The first set was composed of a ten-fold serial dilution series of a two-strain mixture up to 10^{-5}. Specifically, we combined two *Pseudomonas* isolates (*P. syringae* Leaf129 and *P. thivervalensis* PITR2) and two *Z. tritici* isolates (1E4 and IR01_48b). The second set consisted of a ten-fold serial dilution series up to 10^{-5} diluting *P. thivervalensis* PITR2 and 1E4 in a leaf DNA extraction sample (Leaf 1: Timepoint 2, Cultivar Zinal, Plot2, Plant 2, Leaf height: top; Leaf 2: Timepoint 3, Cultivar Titlis, Plot 2, Plant 2, Leaf height: top). For the third set, we combined increasing numbers of *Pseudomonas* strains in equimolar concentrations (two to ten isolates).

PANGENOME CREATION AND PRIMER DESIGN

To construct the *Pseudomonas* pangenome, we gathered 18 high-quality genomes representing all subgroups of the genus (Gomila *et al.*, 2015; Peix, Ramírez-Bahena and Velázquez, 2018) (Supplementary Table 5). We used Panseq (Laing *et al.*, 2010) to construct the pangenome with the following settings: minimumNovelRegionSize = 500, novelRegionFinderMode = no_duplicates, percentIdentityCutoff = 60, coreGenomeThreshold = 16, runMode = pan, storeAlleles = 1, allelesToKeep = 2, frameshift = 1, overwrite = 1, maxNumberResultsInMemory = 500, blastWordSize = 11, nucB = 200, nucC = 65, nucD = 0.12, nucG = 90, nucL = 20, cdhit = 0, sha1 = 0. We considered Panseq loci as core if they were shared by >88% of the genomes (i.e. >=16/18). We created a multiple sequence alignment of each core fragment using muscle v.3 with default parameters (Edgar, 2004). For each multiple sequence alignment, we created a consensus sequence using the EMBOSS v.6 cons function, setting the identity to ten and plurality to 0.8 (Rice, Longden and Bleasby, 2000). We then ran Primer3 v. 2.5 on each core consensus sequence (Untergasser *et al.*, 2012). We used the following Primer3 settings: PRIMER_PICK_LEFT_PRIMER=1, PRIMER_PICK_INTERNAL_OLIGO=0, PRIMER_PICK_RIGHT_PRIMER=1, PRIMER_OPT_SIZE=20, PRIMER_MIN_SIZE=18, PRIMER_MAX_SIZE=22, PRIMER_PRODUCT_SIZE_RANGE=1500-3100, PRIMER_NUM_RETURN=8, PRIMER_EXPLAIN_FLAG=1. For each primer pair, we cut out the respective amplicon from the multiple sequence alignment using the EMBOSS v. 6 extractalign function (Rice, Longden and Bleasby, 2000). We calculated the nucleotide diversity from each amplicon multiple sequence alignment. For this, we first converted the multiple sequence alignment to vcf format using snp-sites (Page *et al.*, 2016), and then vcftools v0.1.16 with the options –sites-pi –haploid to calculate the nucleotide diversity (Danecek *et al.*, 2011). We selected the ten amplicons with the greatest summed nucleotide diversity for further evaluation. Each candidate primer pair was blasted against all available *Pseudomonas* genomes from the ncbi nucleotide collection using BLASTn

413 (Altschul *et al.*, 1990). All hits were extracted and aligned using MAFFT v7.427 with the option –auto
414 (Katoh *et al.*, 2002; Katoh and Standley, 2013). Then, we assessed allele frequencies at every position
415 of the multiple sequence alignment using base R functions (R Core Team, 2022). We manually adjusted
416 primer sequences to account for the specific combinations of alleles necessary to amplify all known
417 *Pseudomonas* sequences.

418 For the primer development of *Z. tritici*, we based our analysis on the 19-isolate pangenome from Badet
419 et al., 2020. Specifically, we extracted all core genes present in all genomes and extracted 6-kb windows
420 around each core gene as our input core loci. To extract core regions, we used the bedtools v.2.28.0
421 function intersect and getfasta (Quinlan and Hall, 2010). We used BLASTn to extract core regions for
422 all genomes and created a multiple sequence alignment using MAFFT v7.427 with the option –auto
423 (Katoh *et al.*, 2002; Katoh and Standley, 2013). We then created consensus sequences and developed
424 primer sequences analogously to the *Pseudomonas* amplicon primer procedure described above. To
425 evaluate amplicon diversity, we analyzed variant call files (i.e. vcf) for each amplicon based on a
426 worldwide collection of 1109 *Z. tritici* isolates (Feurtey *et al.*, 2023). We selected the ten amplicons
427 with highest nucleotide diversity and adjusted primer sequences analogously to the *Pseudomonas*
428 amplicon primer procedure.

429 All ten *Pseudomonas* and ten *Z. tritici* primer candidates were tested on reference cultures of
430 *Pseudomonas* and *Z. tritici*, respectively, as well as on naturally infected wheat leaf samples. The two
431 *Pseudomonas* and two *Z. tritici* primer pairs with most consistent amplification success were retained
432 for further steps.

433 SAMPLE HOMOGENIZATION, DNA EXTRACTION, AMPLIFICATION, POOLING AND
434 CLEANUP

435 Leaves were lyophilized for 48h and weighed. Then, the complete leaves were homogenized using
436 0.5mm and 0.2mm zirconium beads in the Bead Ruptor bead mill homogenizer (OMNI) using the
437 following settings: speed 5.00, number of cycles 2, time of cycle 1:00, time distance between cycles
438 1:00. DNA extraction was performed with automated magnetic-particle processing using the
439 KingFisher Flex Purification Systems (Thermo Scientific). To enhance the DNA extraction of fungal
440 and bacterial DNA, lyticase and lysozyme was added to the first lysis step with PVP buffer. Specifically,
441 for 10mg dry leaf mass 3.9µl lyticase (200,000 U/mg, diluted to 6.5mg/ml), 3.9µl lysozyme (22,500
442 U/mg, diluted to 10mg/ml) and 98µl PVP lysis buffer was added, and samples were incubated at 55°C
443 for 30min. Then, 3.9µl proteinase K (30 U/mg, diluted to 10mg/ml) for 10mg dry mass was added and
444 incubated at 55°C for 30min. From each sample, 150µl of clear lysate was transferred to an empty
445 binding plate (KingFisher Flex, Thermo Scientific). For each sample, 360µl PN binding buffer, 30µl
446 well suspended Sbeadex beads were added. Using the KingFisher Flex, the first washing step was

447 performed using 400µl PN1 buffer per sample, then a second wash using 390µl buffer PN1 with 10µl
448 RNase A (diluted to 10mg/µl in water) and a third wash using 400µl PN2 buffer. Each sample was
449 eluted in 100µl nuclease-free water. The DNA concentration was measured using the Spark Microplate
450 reader (Tecan). Then, DNA concentrations were diluted to 5ng/µl using the Liquid Handling Station
451 (BRAND). PCR reactions were pipetted using the Mosquito HV liquid handling robot (SPT Labtech).
452 The first amplicon PCR reaction was performed in a 15µl reaction volume. Specifically, 7.5µl KAPA
453 HiFi HotStart ReadyMix (2x), 1.5µl forward primer (3µM), 1.5ul reverse primer (3µM), 3ul DNA
454 (5ng/µl) and 1.5µl HPLC water were combined. Primer sequences and cycling protocols are
455 documented in Supplementary Table 6 and 7. *Pseudomonas* and *Z. tritici*-specific amplicon PCR
456 products were diluted 1:5, PCR products from 16S and ITS 1:10. The second barcoding PCR reaction
457 was performed in a 25µl reaction volume. Specifically, 12.5µl KAPA HiFi HotStart ReadyMix (2x),
458 2.5µl M13 forward barcode (3µM), 1.5µl M13 reverse barcode (3µM), 2µl diluted PCR product and
459 5.5µl HPLC water were combined. Barcode sequences are available in Supplementary Table 8. Samples
460 were pooled by amplicon taking 1.5µl from each barcoded product. Each amplicon pool was cleaned
461 using AMPure XP beads using a bead ratio of 0.8x.

462 LIBRARY PREPARATION AND SEQUENCING

463 Library preparation and sequencing was carried out at the Functional Genomics Centre Zurich (FGCZ).
464 Two SMRTbell libraries were prepared for each amplicon length using the SMRTbell prep kit 3.0. One
465 for the long 3-kb amplicons, one for the 1.5kb 16S and ITS amplicons. Size selection was performed
466 using BluePippin (Sage Science) with a 0.75% dye-free cassette for each library. PacBio sequencing
467 was performed on a Sequel II machine with the SMRT 8M cell. Two sequencing runs were performed,
468 the first one using SMRT Link version 10.1 and the second one SMRT Link version 11.1. The first run
469 encompassed all leaf samples, and the second one all mock samples.

470 PACBIO RAW READ PROCESSING

471 CCS were extracted from raw reads using the ccs software from the bioconda package pbccs provided
472 by the manufacturer (Pacific Biosciences). For the first run pbccs v. 6.0.0 was used and for the second
473 run v. 6.3.0. We split the CCS reads by barcodes using the software lima 2.0.0 (Pacific Biosciences)
474 with the following parameters lima –log-level INFO –per-read –min-passes 0 –split-bam-named –ccs
475 –different -A 1 -B 3 -D 2 -I 2 -X 0. We assigned the reads to the respective amplicons using BLASTn
476 assignments to reference amplicon from the *P. fluorescens* CHA0 and *Z. tritici* isolate 1E4 (Altschul *et*
477 *al.*, 1990). All 6 reference amplicons were blasted against each read and reads were then assigned to
478 the reference hit with the lowest e-value, the highest length and the highest identity. We removed primer
479 sequences using cutadapt v. 3.4 with the following parameters: cutadapt -a
480 FORWARD_PRIMER_SEQ...REVERSE_PRIMER_SEQ –discard-untrimmed –revcomp (Martin,
481 2011). Primers were treated as linked, i.e. reads without primers at both ends were discarded. The R

482 package dada2 v. 1.28.0 was used to infer ASVs for each amplicon separately (Callahan *et al.*, 2016).
483 In the following, dada2 steps for each amplicon are described. Reads were filtered and trimmed using
484 the function filterAndTrim with the parameters minLen=minLength, maxLen=maxLength,
485 rm.phix=FALSE, maxEE=2, qualityType = "FastqQuality", multithread=TRUE. Length ranges for
486 each amplicon are described in Supplementary Table 9. Reads were dereplicated using the function
487 derepFastq with the parameters verbose=TRUE, qualityType="FastqQuality". Error models were
488 estimated with the function learnErrors with the parameters
489 errorEstimationFunction=dada2:::PacBioErrfun, BAND_SIZE=32, multithread=TRUE. Reads were
490 denoised using the function dada with the parameters BAND_SIZE=32, multithread=TRUE,
491 pool=FALSE. Importantly, we did not used the option pool=TRUE, as this introduced spurious ASVs
492 in the reference cultures of the mock community. ASV sequence tables were generated using the
493 function makeSequenceTable. To remove chimeras from the sequence table, the function
494 removeBimeraDenovo with parameters method="consensus", minFoldParentOverAbundance=3.5,
495 multithread=TRUE, verbose=TRUE was used. Supplementary Figure 6 shows the read tracking
496 through the dada2 pipeline for each amplicon.

497 TAXONOMIC CLASSIFICATION

498 We assigned 16S reads using the dada2-formatted Silva database v. 138 (Yilmaz *et al.*, 2014) and ITS
499 reads to the UNITE database v. 8.3 (Nilsson *et al.*, 2019). To perform taxonomy assignment, we used
500 the function assignTaxonomy from the dada2 package. As the UNITE database comprises mostly ITS1-
501 ITS2 reference sequences and only few full-length ITS-LSU sequences, we truncated all ITS reads to
502 the ITS1-ITS2 fragment for assignment. We blasted the reads against the ITS1-ITS2 fragment of
503 *Z.tritici* strain S-46 (KT336200.1) and used the hit coordinates to truncate reads with the seqkit software
504 function subseq (Shen *et al.*, 2016). We assigned *Pseudomonas* reads of the *rpoD*, transporter and 16S
505 amplicons to *Pseudomonas* species using BLASTn against all 1071 full-length *Pseudomonas* genomes
506 available from the *Pseudomonas* db v. 21.1 (2022-11-20) (Supplementary Table 4) (Altschul *et al.*,
507 1990; Winsor *et al.*, 2016). The best assignment was chosen according to BLASTn bitscores. We
508 assigned *Z. tritici* reads of the *Z. tritici* amplicons on chromosomes 9 and 13, and of the ITS amplicon
509 to a database of previously sequenced *Z. tritici* strains. For this, we used draft assemblies of previously
510 collected 177 genomes from the Eschikon Field Station of the ETH Zurich, Switzerland (Singh, Karisto
511 and Croll, 2021), as well as genomes from the reference pangenome (Badet *et al.*, 2020). The best
512 assignment was chosen according to bitscores.

513 ## ALIGNMENT, PHYLOGENETIC TREE AND NETWORK CONSTRUCTION

514 We performed multiple sequence alignment using PASTA v.1.9.0 with the following MAFFT
515 arguments --leavegappyregion --6merpair --maxiterate 0 --adjustdirection --reorder and FastTree model
516 -gtr -gamma -fastest (Mirarab *et al.*, 2015). We built phylogenetic trees of leaf samples using FastTree
517 v.2.0.0 (Price, Dehal and Arkin, 2009), and of the *Pseudomonas* mock samples using raxml-ng v. 1.2.0
518 with the options --all --model GTR+G --opt-model on --threads 4 --seed 2 --outgroup "Main ASV of *P.*
519 *syringae* Leaf129" (Stamatakis, 2014). To create unrooted phylogenetic networks for the *Z. tritici*
520 amplicon 1 on chromosome 13, amplicon 2 on chromosome 9 and the ITS, the software SplitsTree v.
521 4.19.1 was used using uncorrected p distances (Huson and Bryant, 2006).

522 ## STATISTICS

523 Regression analyses were performed using the R function lm from the package stats v. 4.2.2 (Best and
524 Roberts, 1975). Correlations were calculated using the R function cor.test from the package stats
525 v. 4.2.2 (Best and Roberts, 1975). Shannon diversity was calculated using the function diversity from
526 the R package vegan v. 2.6.4 (Oksanen Jari, 2022). Permutations were calculated for 1000 iterations
527 using the base R function sample with replace=TRUE (R Core Team, 2022). Subtrees were identified
528 using the function subtrees from the R package ape v. 5.7.1 (Paradis and Schliep, 2019).

529 ## VISUALIZATION

530 To visualize ASV abundances, the R package phyloseq v. 1.42.0 was used (McMurdie and Holmes,
531 2013). Counts were normalized by sample using the phyloseq function
532 transform_sample_counts(phyloseq_object, function (x) x / sum(x)). The heatmap was created using
533 the R package pheatmap v. 1.0.12 (Kolde, 2019). Phylogenetic trees were visualized using the R
534 package ggtree v. 3.6.2 (Yu, 2020; Curry and Zhang, 2022). All other plots were created using the R
535 package ggplot2 v. 3.4.2 (Wickham, 2016).

ACKNOWLEDGMENTS

Data produced in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. We thank Silvia Kobel and Aria Minder for the helpful discussions regarding the DNA extraction and PCR automations. We thank all the people who helped with the field sampling and processing, specifically Dominic Stalder, Nikhil Kumar Singh, Leen Abraham, Thomas Badet, Stéphanie Ruaud, Camille Kessler, Simone Fouché, Ursula Oggenfuss, Anna Spescha, Florence Gilliéron, Jana Schneider, Michael Brunner and Hansueli Zellweger. We thank Dominic Stalder and Sabina Tralamazza for all the helpful discussions on the analysis and Hanspeter Stalder for the input on PCR design and troubleshooting.

REFERENCES

Altschul, S.F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderson, A.J. and Kim, Y.C. (2020) 'Insights into plant-beneficial traits of probiotic Pseudomonas chlororaphis isolates', *Journal of Medical Microbiology*, 69(3), pp. 361–371. Available at: https://doi.org/10.1099/JMM.0.001157.

Badet, T. *et al.* (2020) 'A 19-isolate reference-quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici', *BMC Biology*, 18(1), pp. 1–18. Available at: https://doi.org/10.1186/s12915-020-0744-3.

Barrett, L.G. *et al.* (2008) 'Life history determines genetic structure and evolutionary potential of host-parasite interactions', *Trends in ecology & evolution*, 23(12), pp. 678–685. Available at: https://doi.org/10.1016/J.TREE.2008.06.017.

Barroso-Bergada, D. *et al.* (2022) 'NGS data reveals the phyllosphere microbiome of wheat plants infected by the fungal pathogen Zymoseptoria tritici', *bioRxiv* [Preprint]. Available at: https://doi.org/10.1101/2022.02.07.479402.

Best, D.J. and Roberts, D.E. (1975) 'Algorithm AS 89: The Upper Tail Probabilities of Spearman's Rho', *Applied Statistics*, 24(3), p. 377. Available at: https://doi.org/10.2307/2347111.

Brockhurst, M.A. *et al.* (2019) 'The Ecology and Evolution of Pangenomes', *Current biology : CB*, 29(20), pp. R1094–R1103. Available at: https://doi.org/10.1016/J.CUB.2019.08.012.

Callahan, B.J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods 2016 13:7*, 13(7), pp. 581–583. Available at: https://doi.org/10.1038/nmeth.3869.

Callahan, B.J. *et al.* (2019) 'High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution', *Nucleic acids research*, 47(18), p. E103. Available at: https://doi.org/10.1093/NAR/GKZ569.

Chen, J. *et al.* (2021) 'Wheat Microbiome: Structure, Dynamics, and Role in Improving Performance Under Stress Environments', *Frontiers in Microbiology*, 12. Available at: https://doi.org/10.3389/FMICB.2021.821546.

Choi, K., Khan, R. and Lee, S.W. (2021) 'Dissection of plant microbiota and plant-microbiome interactions', *Journal of microbiology (Seoul, Korea)*, 59(3), pp. 281–291. Available at: https://doi.org/10.1007/S12275-021-0619-5.

Cummins, E.A. *et al.* (2022) 'Prokaryote pangenomes are dynamic entities', *Current opinion in microbiology*, 66, pp. 73–78. Available at: https://doi.org/10.1016/J.MIB.2022.01.005.

Curry, E. and Zhang, H. (2022) *Data Integration, Manipulation and Visualization of Phylogenetic Trees*, *Data Integration, Manipulation and Visualization of Phylogenetic Trees*. Chapman and Hall/CRC. Available at: https://doi.org/10.1201/9781003279242.

Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics (Oxford, England)*, 27(15), pp. 2156–2158. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTR330.

Earl, J.P. *et al.* (2018) 'Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes', *Microbiome*, 6(1). Available at: https://doi.org/10.1186/S40168-018-0569-2.

Ebert, D. and Fields, P.D. (2020) 'Host-parasite co-evolution and its genomic signature.', *Nature reviews. Genetics* [Preprint]. Available at: https://doi.org/10.1038/s41576-020-0269-1.

Edgar, R.C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*, 32(5), p. 1792. Available at: https://doi.org/10.1093/NAR/GKH340.

Escapa, I.F. *et al.* (2020) 'Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets', *Microbiome*, 8(1), p. 65. Available at: https://doi.org/10.1186/S40168-020-00841-W.

Feurtey, A. *et al.* (2023) 'A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen', *Nature communications*, 14(1). Available at: https://doi.org/10.1038/S41467-023-36674-Y.

Fitzpatrick, C.R. *et al.* (2020) 'The Plant Microbiome: From Ecology to Reductionism and Beyond', pp. 81–100.

Flury, P. *et al.* (2016) 'Insect pathogenicity in plant-beneficial pseudomonads: Phylogenetic distribution and comparative genomics', *ISME Journal*, 10(10), pp. 2527–2542. Available at: https://doi.org/10.1038/ismej.2016.5.

Gomila, M. *et al.* (2015) 'Phylogenomics and systematics in Pseudomonas', *Frontiers in Microbiology*, 6(MAR), pp. 1–13. Available at: https://doi.org/10.3389/fmicb.2015.00214.

Großkinsky, D.K. *et al.* (2016) 'Cytokinin production by Pseudomonas fluorescens G20-18 determines biocontrol activity against Pseudomonas syringae in Arabidopsis', *Scientific reports*, 6. Available at: https://doi.org/10.1038/SREP23310.

Huson, D.H. and Bryant, D. (2006) 'Application of phylogenetic networks in evolutionary studies', *Molecular biology and evolution*, 23(2), pp. 254–267. Available at: https://doi.org/10.1093/MOLBEV/MSJ030.

Johnson, J.S. *et al.* (2019) 'Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis', *Nature Communications 2019 10:1*, 10(1), pp. 1–11. Available at: https://doi.org/10.1038/s41467-019-13036-1.

Karst, S.M. *et al.* (2021) 'High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing', *Nature Methods*, 18(2), pp. 165–169. Available at: https://doi.org/10.1038/s41592-020-01041-y.

Katoh, K. *et al.* (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Research*, 30(14), pp. 3059–3066. Available at: https://doi.org/10.1093/NAR/GKF436.

Katoh, K. and Standley, D.M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780. Available at: https://doi.org/10.1093/MOLBEV/MST010.

Kavamura, V.N. *et al.* (2021) 'Defining the wheat microbiome: Towards microbiome-facilitated crop production', *Computational and Structural Biotechnology Journal*, 19, p. 1200. Available at: https://doi.org/10.1016/J.CSBJ.2021.01.045.

Kerdraon, L. *et al.* (2019) 'Differential dynamics of microbial community networks help identify microorganisms interacting with residue-borne pathogens: The case of Zymoseptoria tritici in wheat', *Microbiome*, 7(1). Available at: https://doi.org/10.1186/S40168-019-0736-0.

Kolde, R. (2019) 'pheatmap: Pretty Heatmaps'.

Laing, C. *et al.* (2010) 'Pan-genome sequence analysis using Panseq: An online tool for the rapid analysis of core and accessory genomic regions', *BMC Bioinformatics*, 11, pp. 1–14. Available at: https://doi.org/10.1186/1471-2105-11-461.

Laino, P. *et al.* (2015) 'Morpho-physiolological and qualitative traits of a bread wheat collection spanning a century of breeding in Italy', *Biodiversity Data Journal 3: e4760*, 3(1), pp. e4760-. Available at: https://doi.org/10.3897/BDJ.3.E4760.

Levy, E., Eyal, Z. and Chet, I. (1988) 'Suppression of septoria tritici blotch and leaf rust on wheat seedling leaves by pseudomonads', *Plant Pathology*, 37(4), pp. 551–557. Available at: https://doi.org/10.1111/J.1365-3059.1988.TB02114.X.

Liu, Y.X. *et al.* (2021) 'A practical guide to amplicon and metagenomic analysis of microbiome data', *Protein & cell*, 12(5), pp. 315–330. Available at: https://doi.org/10.1007/S13238-020-00724-8.

Lucas, J.A., Hawkins, N.J. and Fraaije, B.A. (2015) *The Evolution of Fungicide Resistance*, *Advances in Applied Microbiology*. Elsevier Ltd. Available at: https://doi.org/10.1016/bs.aambs.2014.09.001.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal*, 17(1), pp. 10–12. Available at: https://doi.org/10.14806/EJ.17.1.200.

McDonald, B.A. *et al.* (2022) 'How large and diverse are field populations of fungal plant pathogens? The case of Zymoseptoria tritici', *Evolutionary Applications*, 15(9), p. 1360. Available at: https://doi.org/10.1111/EVA.13434.

McMurdie, P.J. and Holmes, S. (2013) 'phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data', *PLOS ONE*, 8(4), p. e61217. Available at: https://doi.org/10.1371/JOURNAL.PONE.0061217.

Mirarab, S. *et al.* (2015) 'PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences', *Journal of Computational Biology*, 22(5), p. 377. Available at: https://doi.org/10.1089/CMB.2014.0156.

Nilsson, R.H. *et al.* (2019) 'The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications', *Nucleic Acids Research*, 47(D1), pp. D259–D264. Available at: https://doi.org/10.1093/NAR/GKY1022.

Oksanen Jari (2022) 'vegan: Community Ecology Package'.

Page, A.J. *et al.* (2016) 'SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments', *Microbial genomics*, 2(4), p. e000056. Available at: https://doi.org/10.1099/MGEN.0.000056/CITE/REFWORKS.

Paradis, E. and Schliep, K. (2019) 'ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R', *Bioinformatics*, 35(3), pp. 526–528. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTY633.

Passera, A. *et al.* (2019) 'Not Just a Pathogen? Description of a Plant-Beneficial Pseudomonas syringae Strain', *Frontiers in Microbiology*, 10(June). Available at: https://doi.org/10.3389/fmicb.2019.01409.

Peix, A., Ramírez-Bahena, M.H. and Velázquez, E. (2018) 'The current status on the taxonomy of Pseudomonas revisited: An update', *Infection, Genetics and Evolution*, 57(November 2017), pp. 106–116. Available at: https://doi.org/10.1016/j.meegid.2017.10.026.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) 'FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix', *Molecular Biology and Evolution*, 26(7), p. 1641. Available at: https://doi.org/10.1093/MOLBEV/MSP077.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. Available at: https://doi.org/10.1093/bioinformatics/btq033.

R Core Team (2022) *R: A Language and Environment for Statistical Computing*. Available at: https://www.r-project.org/ (Accessed: 13 July 2023).

Raio, A. and Puopolo, G. (2021) 'Pseudomonas chlororaphis metabolites as biocontrol promoters of plant health and improved crop yield', *World Journal of Microbiology and Biotechnology*, 37(6), pp. 1–8. Available at: https://doi.org/10.1007/S11274-021-03063-W/FIGURES/2.

Rice, P., Longden, L. and Bleasby, A. (2000) 'EMBOSS: the European Molecular Biology Open Software Suite', *Trends in genetics : TIG*, 16(6), pp. 276–277. Available at: https://doi.org/10.1016/S0168-9525(00)02024-2.

Seybold, H. *et al.* (2020) 'A fungal pathogen induces systemic susceptibility and systemic shifts in wheat metabolome and microbiome composition', *Nature Communications 2020 11:1*, 11(1), pp. 1–12. Available at: https://doi.org/10.1038/s41467-020-15633-x.

Shen, W. *et al.* (2016) 'SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation', *PLOS ONE*, 11(10), p. e0163962. Available at: https://doi.org/10.1371/JOURNAL.PONE.0163962.

Singh, N.K., Karisto, P. and Croll, D. (2021) 'Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen Zymoseptoria tritici', *Microbial Genomics*, 7, p. 678. Available at: https://doi.org/10.1099/mgen.0.000678.

Song, S. *et al.* (2021) 'Mechanisms in plant-microbiome interactions: lessons from model systems', *Current opinion in plant biology*, 62. Available at: https://doi.org/10.1016/J.PBI.2021.102003.

Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics*, 30(9), pp. 1312–1313. Available at: https://doi.org/10.1093/BIOINFORMATICS/BTU033.

Tack, A.J.M. *et al.* (2012) 'Variation in infectivity and aggressiveness in space and time in wild host–pathogen systems: causes and consequences', *Journal of Evolutionary Biology*, 25(10), pp. 1918–1936. Available at: https://doi.org/10.1111/J.1420-9101.2012.02588.X.

Tettelin, H. and Medini, D. (2020) 'The Pangenome', *The Pangenome: Diversity, Dynamics and Evolution of Genomes*, pp. 1–307. Available at: https://doi.org/10.1007/978-3-030-38281-0.

Trivedi, P. *et al.* (2020) 'Plant–microbiome interactions: from community assembly to plant health', *Nature Reviews Microbiology*, 18(11), pp. 607–621. Available at: https://doi.org/10.1038/s41579-020-0412-1.

Untergasser, A. *et al.* (2012) 'Primer3—new capabilities and interfaces', *Nucleic Acids Research*, 40(15), p. e115. Available at: https://doi.org/10.1093/NAR/GKS596.

Vorholt, J.A. *et al.* (2017) 'Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research', *Cell Host & Microbe*, 22(2), pp. 142–155. Available at: https://doi.org/10.1016/J.CHOM.2017.07.004.

Wagner, J. *et al.* (2016) 'Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification', *BMC microbiology*, 16(1), pp. 1–17. Available at: https://doi.org/10.1186/S12866-016-0891-4.

Wear, E.K. *et al.* (2018) 'Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton', *Environmental Microbiology*, 20(8), p. 2709. Available at: https://doi.org/10.1111/1462-2920.14091.

Wenger, A.M. *et al.* (2019) 'Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome', *Nature Biotechnology 2019 37:10*, 37(10), pp. 1155–1162. Available at: https://doi.org/10.1038/s41587-019-0217-9.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer International Publishing (Use R!). Available at: https://doi.org/10.1007/978-3-319-24277-4.

Winsor, G.L. *et al.* (2016) 'Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database', *Nucleic acids research*, 44(D1), pp. D646–D653. Available at: https://doi.org/10.1093/NAR/GKV1227.

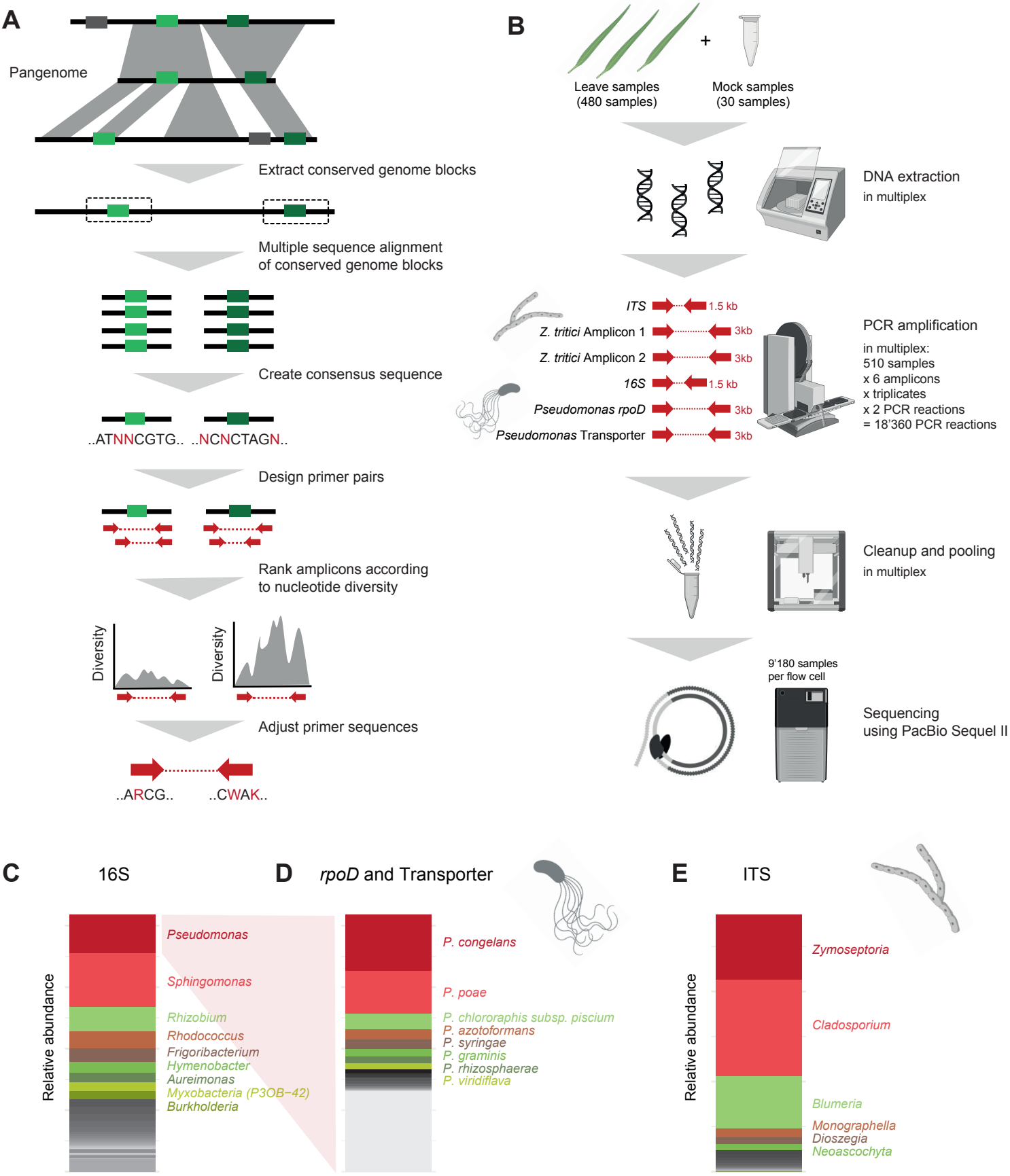
Xin, X.F., Kvitko, B. and He, S.Y. (2018) 'Pseudomonas syringae: What it takes to be a pathogen', *Nature Reviews Microbiology*, 16(5), pp. 316–328. Available at: https://doi.org/10.1038/NRMICRO.2018.17.

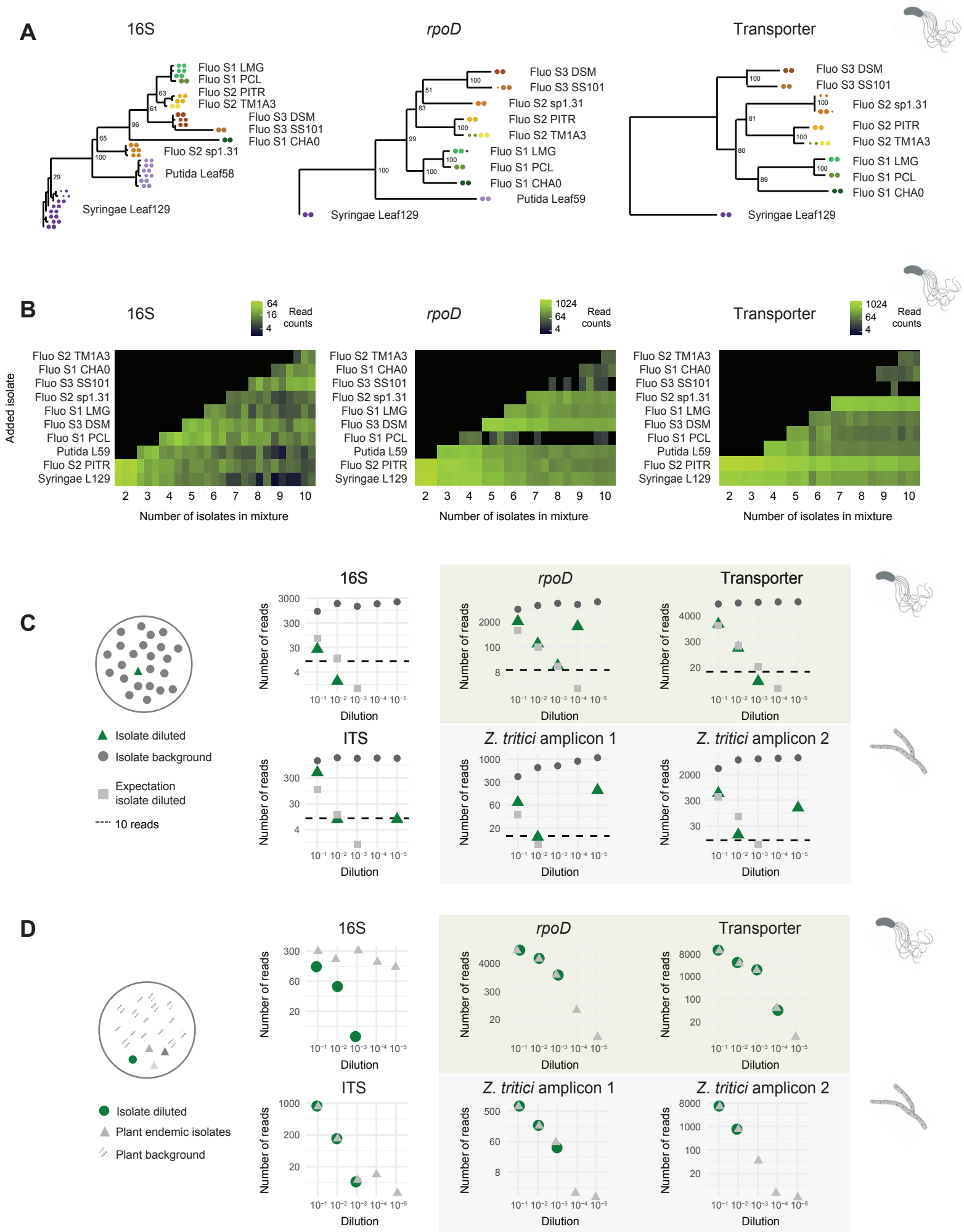
Xu, L. *et al.* (2021) 'Holo-omics for deciphering plant-microbiome interactions', *Microbiome*, 9(1). Available at: https://doi.org/10.1186/S40168-021-01014-Z.

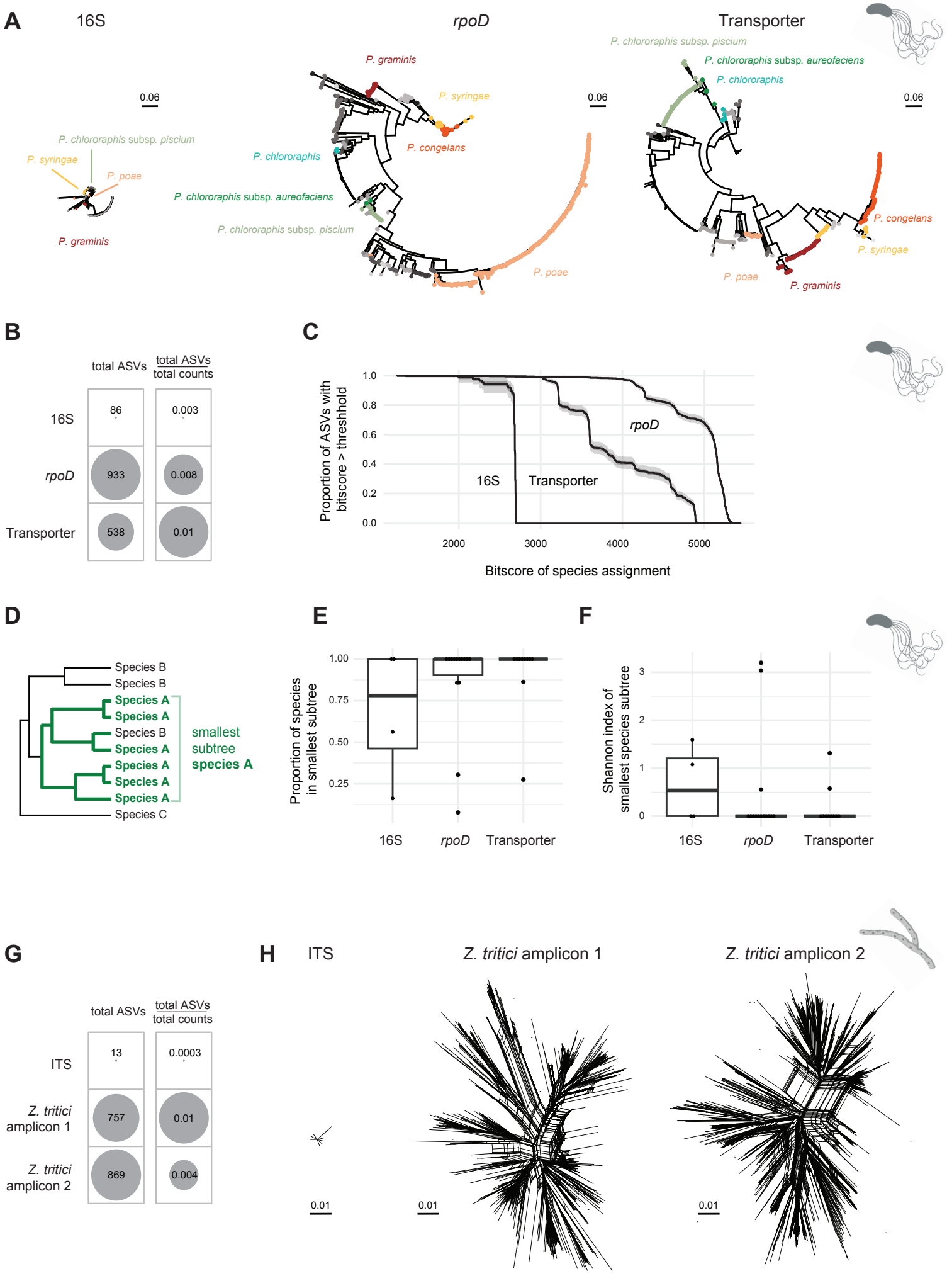
Yeh, Y.C. *et al.* (2021) 'Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S', *Environmental Microbiology*, 23(6), pp. 3240–3250. Available at: https://doi.org/10.1111/1462-2920.15553.

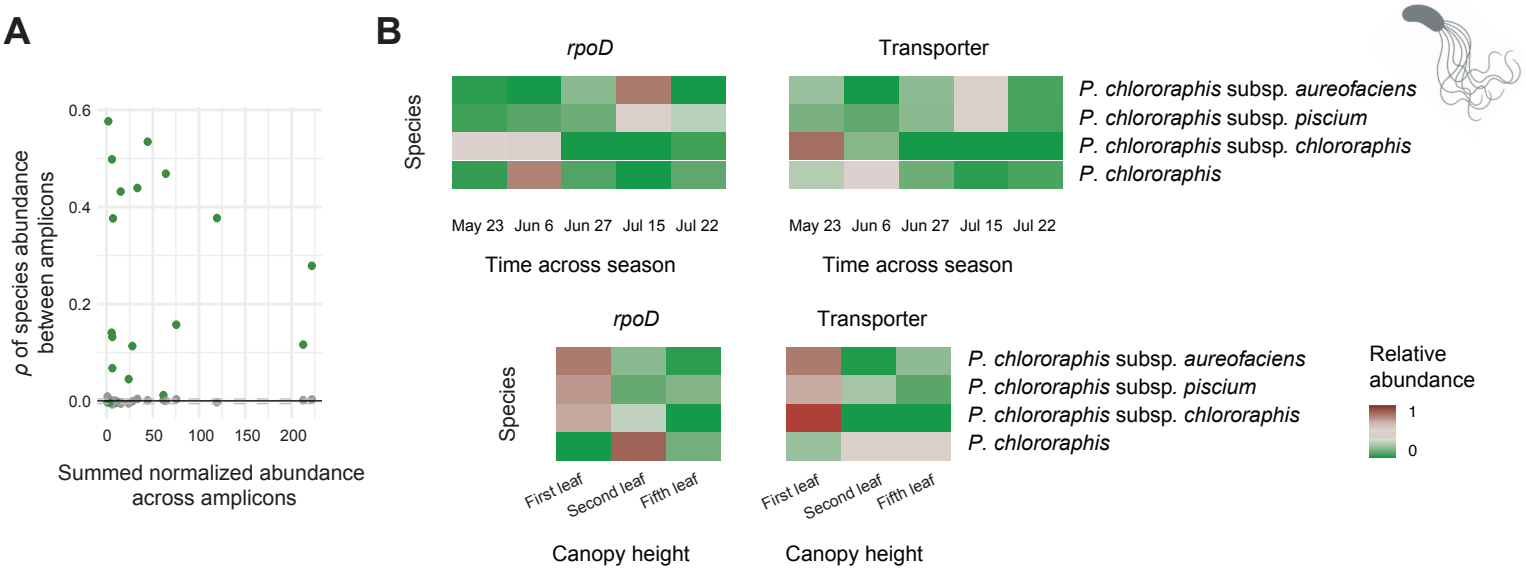
Yilmaz, P. *et al.* (2014) 'The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks', *Nucleic Acids Research*, 42(D1), pp. D643–D648. Available at: https://doi.org/10.1093/NAR/GKT1209.

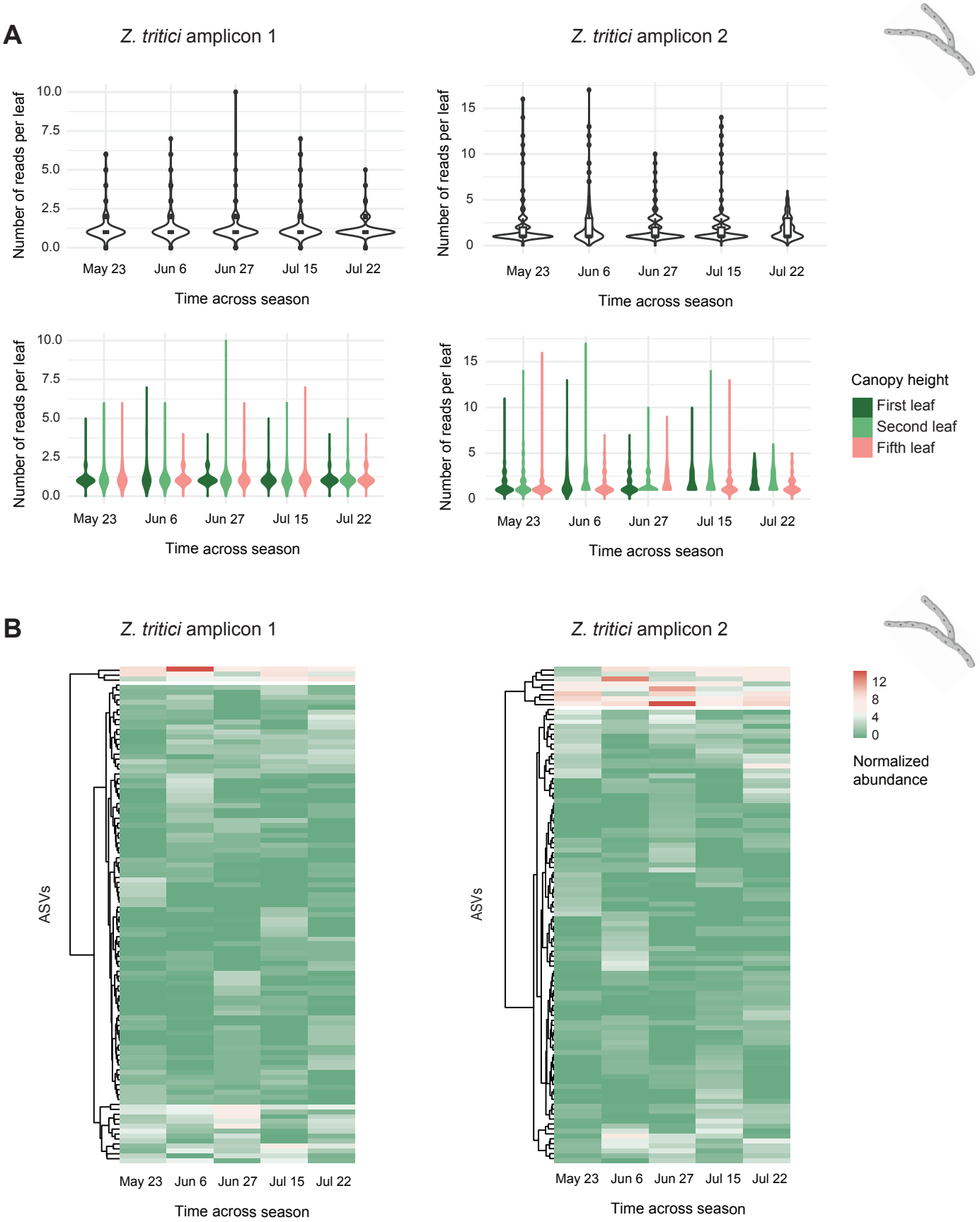
Yu, G. (2020) 'Using ggtree to Visualize Data on Tree-Like Structures', *Current protocols in bioinformatics*, 69(1). Available at: https://doi.org/10.1002/CPBI.96.


A
Pangenome
Extract conserved genome blocks
Multiple sequence alignment
of conserved genome blocks
Create consensus sequence
..ATNNCGTG.. ..NCNCTAGN..
Design primer pairs
Rank amplicons according
to nucleotide diversity
Diversity
Diversity
Adjust primer sequences
..ARCG.. ..CWAK..
B
Leave samples
(480 samples)
+
Mock samples
(30 samples)
DNA extraction
in multiplex
ITS
1.5 kb
Z. tritici Amplicon 1
3kb
Z. tritici Amplicon 2
3kb
16S
1.5 kb
Pseudomonas rpoD
3kb
Pseudomonas Transporter
3kb
PCR amplification
in multiplex:
510 samples
x 6 amplicons
x triplicates
x 2 PCR reactions
= 18'360 PCR reactions
Cleanup and pooling
in multiplex
9'180 samples
per flow cell
Sequencing
using PacBio Sequel II
C
16S
Relative abundance
Pseudomonas
Sphingomonas
Rhizobium
Rhodococcus
Frigoribacterium
Hymenobacter
Aureimonas
Myxobacteria (P3OB−42)
Burkholderia
D
rpoD and Transporter
P. congelans
P. poae
P. chlororaphis subsp. piscium
P. azotoformans
P. syringae
P. graminis
P. rhizosphaerae
P. viridiflava
E
ITS
Relative abundance
Zymoseptoria
Cladosporium
Blumeria
Monographella
Dioszegia
Neoascochyta


A
16S
rpoD
Transporter
Fluo S1 LMG
Fluo S1 PCL
Fluo S2 PITR
Fluo S2 TM1A3
Fluo S3 DSM
Fluo S3 SS101
Fluo S1 CHA0
Fluo S2 sp1.31
Putida Leaf58
Syringae Leaf129
Putida Leaf59
B
Read counts
Added isolate
Fluo S2 TM1A3
Fluo S1 CHA0
Fluo S3 SS101
Fluo S2 sp1.31
Fluo S1 LMG
Fluo S3 DSM
Fluo S1 PCL
Putida L59
Fluo S2 PITR
Syringae L129
Number of isolates in mixture
C
Isolate diluted
Isolate background
Expectation isolate diluted
10 reads
Number of reads
Dilution
ITS
Z. tritici amplicon 1
Z. tritici amplicon 2
D
Isolate diluted
Plant endemic isolates
Plant background


A
16S
rpoD
Transporter
0.06
P. chlororaphis subsp. piscium
P. syringae
P. poae
P. graminis
P. chlororaphis
P. chlororaphis subsp. aureofaciens
P. congelans
B
total ASVs
total ASVs / total counts
16S
rpoD
Transporter
86
0.003
933
0.008
538
0.01
C
Proportion of ASVs with bitscore > threshold
Bitscore of species assignment
16S
Transporter
D
Species B
Species A
Species C
smallest subtree species A
E
Proportion of species in smallest subtree
F
Shannon index of smallest species subtree
G
ITS
Z. tritici amplicon 1
Z. tritici amplicon 2
13
0.0003
757
0.01
869
0.004
H
0.01


A
ρ of species abundance between amplicons
0.6
0.4
0.2
0.0
0
50
100
150
200
Summed normalized abundance across amplicons
B
rpoD
Transporter
Species
May 23
Jun 6
Jun 27
Jul 15
Jul 22
Time across season
P. chlororaphis subsp. aureofaciens
P. chlororaphis subsp. piscium
P. chlororaphis subsp. chlororaphis
P. chlororaphis
First leaf
Second leaf
Fifth leaf
Canopy height
Relative abundance
1
0


A
Z. tritici amplicon 1
Z. tritici amplicon 2
Number of reads per leaf
10.0
7.5
5.0
2.5
0.0
15
10
5
0
May 23
Jun 6
Jun 27
Jul 15
Jul 22
Time across season
Canopy height
First leaf
Second leaf
Fifth leaf
B
Z. tritici amplicon 1
Z. tritici amplicon 2
ASVs
12
8
4
0
Normalized abundance