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 developed and validated a pipeline for designing, multiplexing, and sequencing highly polymorphic taxon-specific amplicons using PacBio circular consensus sequencing. We focused 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 achieved 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. Furthermore, we tracked microbial communities in the wheat phyllosphere across time and space to establish fine-grained species and strain-specific dynamics. To expand the utility of our approach, we generated pangenome-informed amplicon templates for additional key bacterial and fungal genera. The strain-level microbiome profiling enables the tracking of microbial community dynamics in complex environments and is applicable to diverse ecological niches. Overall, our work demonstrates how pangenome-informed amplicons overcome limitations in phylogenetic resolution to unravel microbial strain 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 [1–6]. 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 [7]. In contrast, untargeted metagenomic sequencing and assembly can only provide rich information about microbial communities if the targeted organisms are abundant [7]. 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 [7,8].

Long-read sequencing based on PacBio or Oxford Nanopore opened the possibility to sequence long amplicons [9]. 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 [10]. Studies based on PacBio sequencing of full-length 16S or ITS sequences achieved high taxonomic resolution [8,11–13]. 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*, *P. chlororaphis* or *P. putida* species generally promote plant growth [14–20]. Closely related *Pseudomonas* strains can also differ significantly in volatiles and metabolites production underpinning distinct niche exploitation in the environment [19,21–23]. 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.

Here we present an approach for designing, multiplexing, and sequencing highly polymorphic taxon-specific amplicons using PacBio CCS. Our pipeline 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 [24]. The fungal microbiome of wheat is often dominated by the pathogenic fungus *Zymoseptoria tritici* causing septoria tritici blotch [25–27]. The bacterial microbiome is dominated by Gammaproteobacteria including the genus *Pseudomonas* both above and below ground [28,29]. 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 [30], while specific *P. fluorescens* strains inhibit the growth of *Z. tritici* [31].

In this study, we introduce and validate a pipeline for a novel suite of highly multiplexed amplicons that accomplish 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, and we highlight the substantial gains in phylogenetic resolution by tracking strains across the wheat canopy and over time. Furthermore, we show that this pipeline can be applied to other ecologically relevant genera in need of high-resolution amplicons including *Rhizobia*, *Streptomyces*, and *Aspergillus*.

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 [32,33]. The second amplicon encompassed genes encoding an ABC transporter (APE98195.1) and an ABC transporter permease (APE98194.1), respectively. We pursued a parallel approach to identify primers suitable for amplifying the intra-specific diversity of the fungal pathogen *Z. tritici*. We based our

analysis on a recently established reference-quality global pangenome for the species [34]. The two best-performing amplicons were located on chromosomes 9 and 13, respectively.

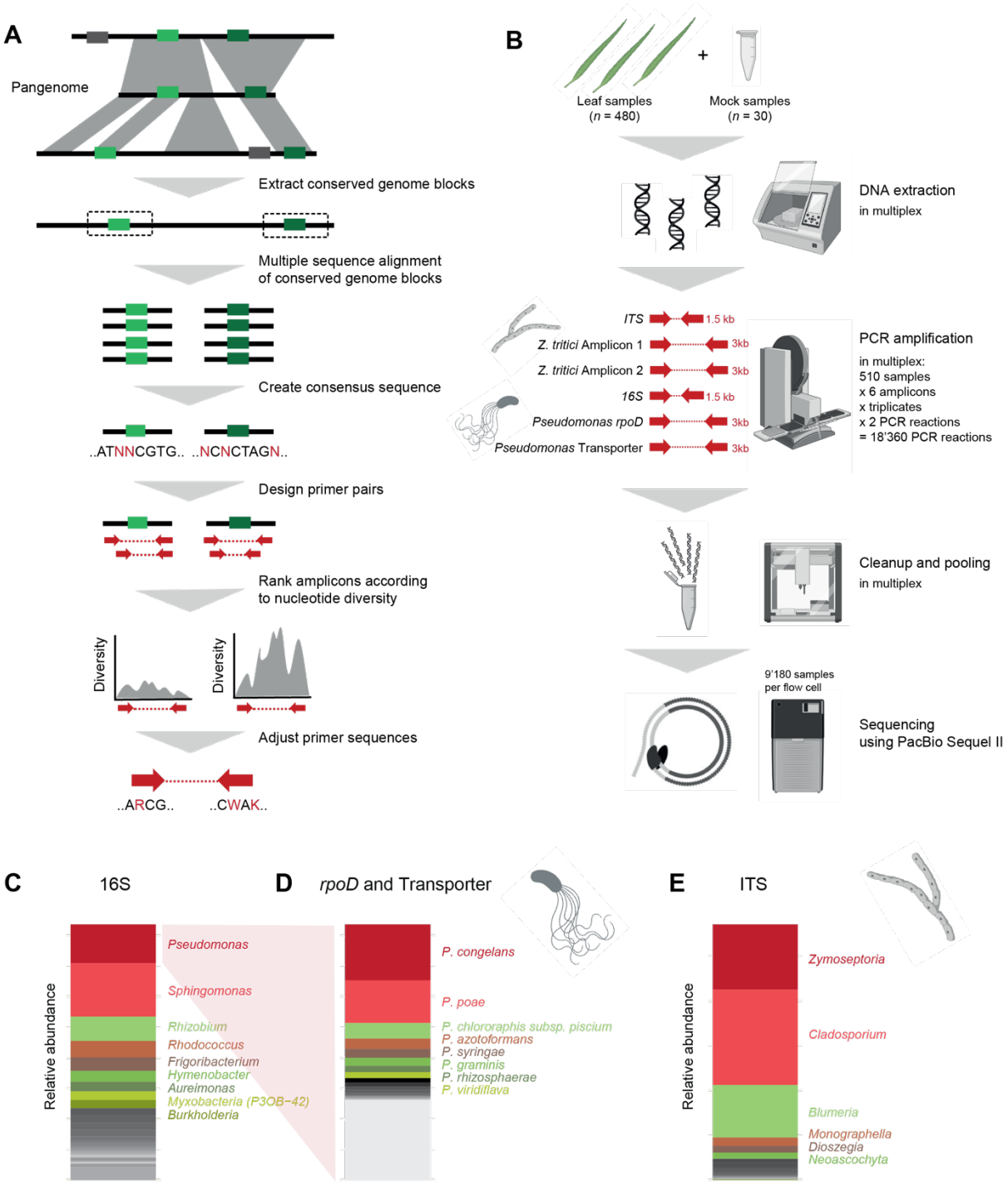


Figure 1: Pangenome-informed design of taxon-specific amplicons. (A) Schematic overview of amplicon design pipeline. Conserved blocks in the *Pseudomonas* and *Z. tritici* pangenomes were extracted. Consensus sequences of the core blocks were used to design amplicons. Amplicon candidates were ranked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

137 and barcoded individually using asymmetric M13 barcodes resulting in a total of 18'360 PCR reactions. All
138 reactions were pooled and sequenced in two PacBio Sequel II 8M cells. (C) Relative abundances of bacterial
139 genera on wheat leaves determined by the full-length 16S amplicon. (D) Relative and mean abundances of
140 *Pseudomonas* species on wheat leaves assessed by the *Pseudomonas*-specific *rpoD* and transporter amplicons.
141 (E) Relative abundances of fungal genera on wheat leaves determined by the full-length ITS amplicon.

142

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

156 Our analysis of 480 wheat leaf samples revealed a diverse assembly of bacteria. Based on the full-length
157 16S amplicon, we found *Pseudomonas* and *Sphingomonas* to be the dominant genera (Figure 1C and
158 Supplementary Figure S1). The *Pseudomonas*-specific *rpoD* and transporter amplicons revealed a
159 diverse assembly of species, primarily belonging to the *P. fluorescens* and *P. syringae* groups (Figure
160 1D and Supplementary Figure S1). The mycobiome, analyzed using the full-length ITS, was
161 predominantly composed of the fungal genera *Zymoseptoria* and *Cladosporium*. This is broadly
162 consistent with recent findings in wheat fields [25,27]. However, our dataset showed a higher abundance
163 of *Cladosporium*. This could be attributed to the humid conditions during our sampling season or the
164 limited resolution of the order *Capnodiales* in other amplicon studies on wheat leaves. These studies all
165 analyzed the short ITS1 amplicon (approximately 350bp), compared to our use of the full-length ITS
166 amplicon (approximately 1500bp) (Figure 1E and Supplementary Figure S1).

167

168 DETECTION LIMITS OF TAXON-SPECIFIC AMPLICONS

169 To assess the performance and detection limits of the new amplicons, we analyzed both defined mock
170 communities and dilution series. For the *Pseudomonas* amplicons, we examined a panel of ten isolates
171 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* Leaf58 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 [35]. 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 [36].

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 isolate *P. chlororaphis ssP. piscium* PCL1391 and *P. fluorescens* SS101 in the mixture. The transporter amplicon exhibited lower amplification efficiency for the isolate *P. fluorescens* SS101. Furthermore, we observed greater variation in abundance between replicates for the 16S amplicon compared to the *rpoD* and transporter amplicons.

We aimed to achieve very high levels of multiplexing and, as a consequence, recovered 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* PITR2 isolates (Figure 2C, Supplementary Figure S2, Supplementary Table 3). Both *Pseudomonas*-specific amplicons were able to identify the diluted isolates at a concentration as low as 10^{-3}, which corresponds to $7.5*10^{-3}$ ng input DNA or approximately 1050 cells. On the other hand, the 16S amplicon could detect the diluted isolates only at a slightly higher concentration of 10^{-2}, representing $7.5*10^{-2}$ ng input DNA or approximately 10´520 cells (Figure 2C). Similarly, we performed a comparable analysis for the fungal pathogen *Z. tritici* using a two-strain dilution series comprising *Z. tritici* isolates ST99CH 1E4 and ST01IR 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 $10^{-2,}$ corresponding to $7.5*10^{-2}$ ng input DNA or approximately 1720 cells. 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).

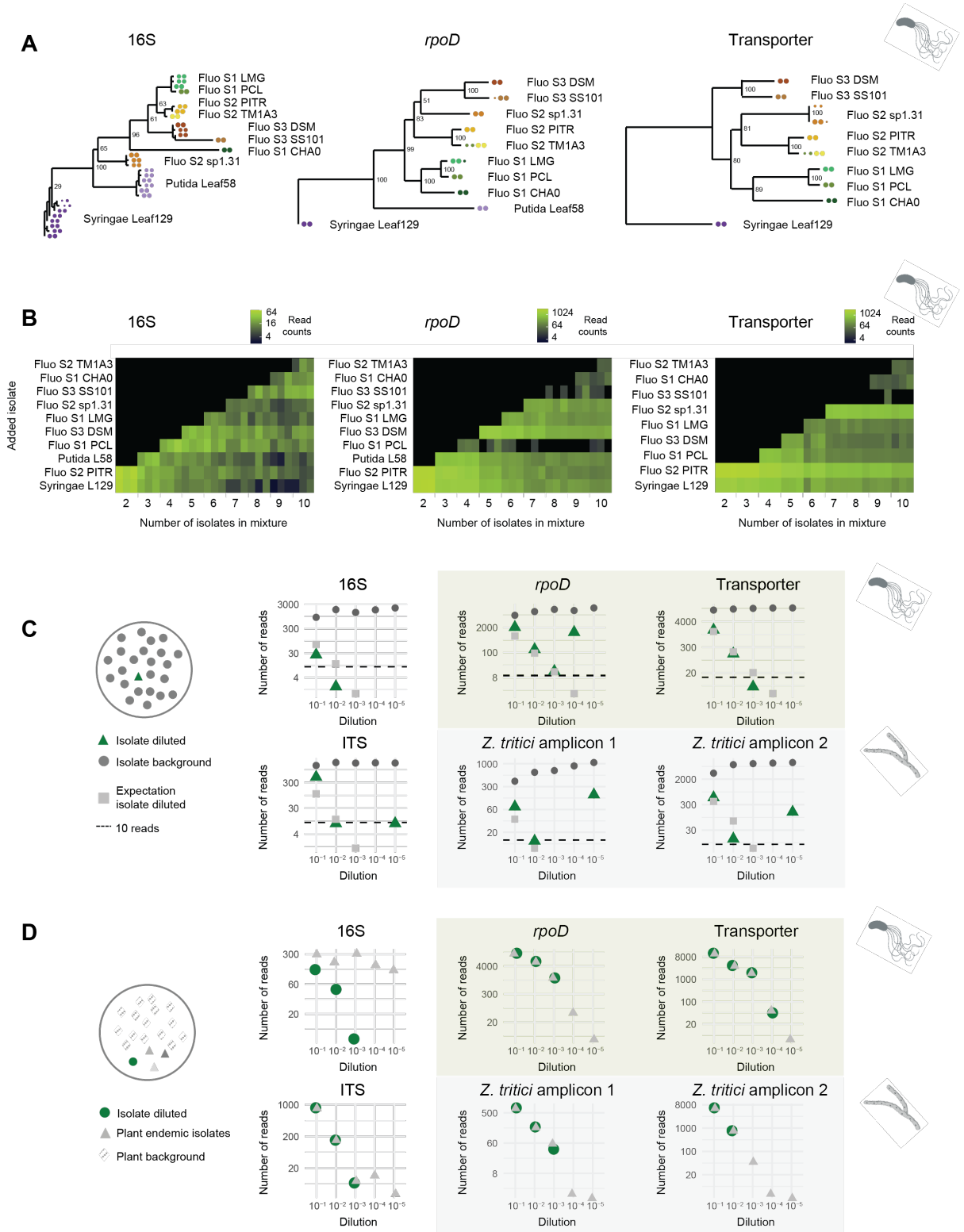


Figure 2: Detection limits of taxon-specific amplicons using mock communities. (A) Phylogenetic trees of amplified sequence variants (ASVs) from ten individual *Pseudomonas* cultures based on the full-length 16S amplicon, the *Pseudomonas*-specific *rpoD* and transporter amplicon. The *Pseudomonas* cultures originate from the following strains: *P. syringae* Leaf129 (Syringae L129), *P. thivervalensis* PITR2 (Fluo S2 PITR), *P. putida* Leaf58 (Putida L58), *P. chlororaphis ssP. piscium* PCL1391 (Fluo S1 PCL), *P. fluorescens* type strain DSM50090 (Fluo S3 DSM), *P. chlororaphis ssP. chlororaphis* type strain LMG5004 (Fluo S1 LMG), *Pseudomonas sP.* 1.31 (Fluo S2 sp1.31), *Pseudomonas sP.* SS101 (Fluo S3 SS101), *P. protegens* CHA0 (Fluo S1 CHA0) and *P. brassicacearum* TM1A3 (Fluo S2 TM1A3). For the isolates from the *P. fluorescens* group, the

subclades Fluo S1, S2 and S3 are indicated as defined by Flury et al., 2016. Colors indicate different cultures, and the dot size corresponds to the abundance of the ASVs for each of the two replicates. The transporter amplicon failed to amplify the *P. putida* Leaf58 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* PITR2 (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* PITR2. Bottom panel: Serial dilution series of *Z. tritici* ST99CH 1E4 (green) in *Z. tritici* ST01IR 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* ST99CH 1E4 isolate are indicated in light grey and calculated from the background read numbers of the *Z. tritici* ST01IR 48b isolate. (D) Top panel: Read numbers of a serial dilution series of *P. thivervalensis* PITR2 (green) mixed into a wheat leaf DNA extract background for the full-length 16S amplicon and the *Pseudomonas*-specific *rpoD* and transporter amplicons. In grey, the total read count of the wheat-endemic bacterial isolates is indicated for the 16S amplicon, respectively the total read count of the wheat-endemic *Pseudomonas* for the *Pseudomonas*-specific *rpoD* and transporter amplicons. Bottom panel: Read numbers of a serial dilution series of *Z. tritici* ST99CH 1E4 (green) mixed into a wheat leaf DNA extract background for the full-length ITS and the *Z. tritici*-specific amplicons. In grey, the total read count of the wheat-endemic fungal isolates is indicated for the ITS amplicon, respectively the total read count of the wheat-endemic *Z. tritici* for the *Z. tritici*-specific amplicons. For all serial dilutions of (C) and (D), dilutions are performed with an initial DNA input of 7.5 ng per isolate.

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. Analyzing 480 wheat leaf samples, Pseudomonas-specific amplicons revealed in total 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 [37,38].

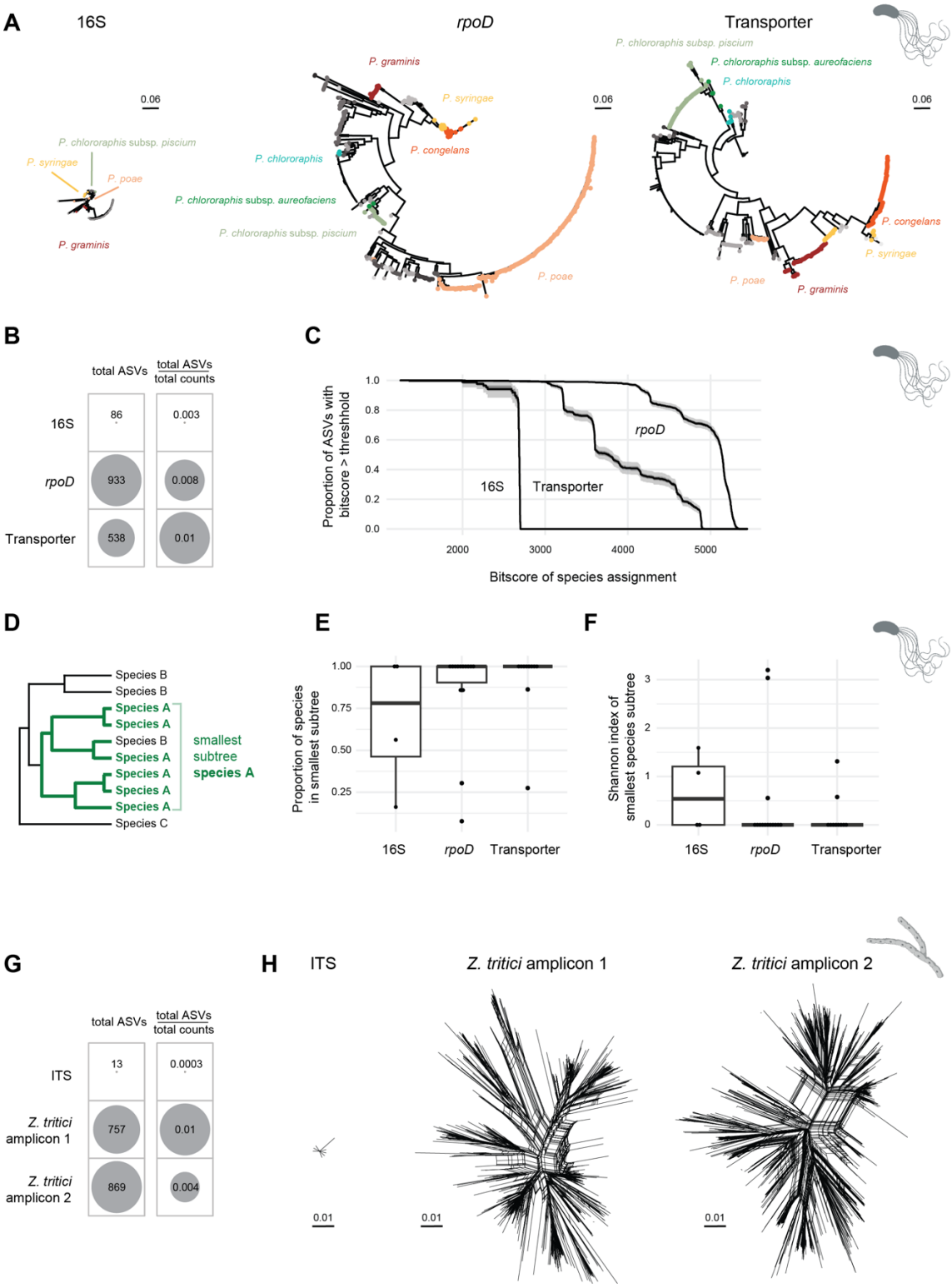


Figure 3: Phylogenetic resolution of taxon-specific and ribosomal DNA amplicons. Taxon-specific amplicon analysis based on 480 wheat leaf samples. (A) Phylogenetic trees of *Pseudomonas* amplified sequence variants (ASVs) of the full-length 16S amplicon as well as the *Pseudomonas*-specific *rpoD* and 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. (E) Proportion of species in the smallest subtree containing >90% of the species ASVs assigned to specific species. (F) 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 of 480 samples across space and time tracking expected microbiome shifts in a wheat field. Specifically, sampling was conducted at five different timepoints during the wheat growing season, from May (first node appearance) to July (prior to harvest). At each timepoint, leaves were sampled at three different canopy heights to capture developmental patterns of the plant.

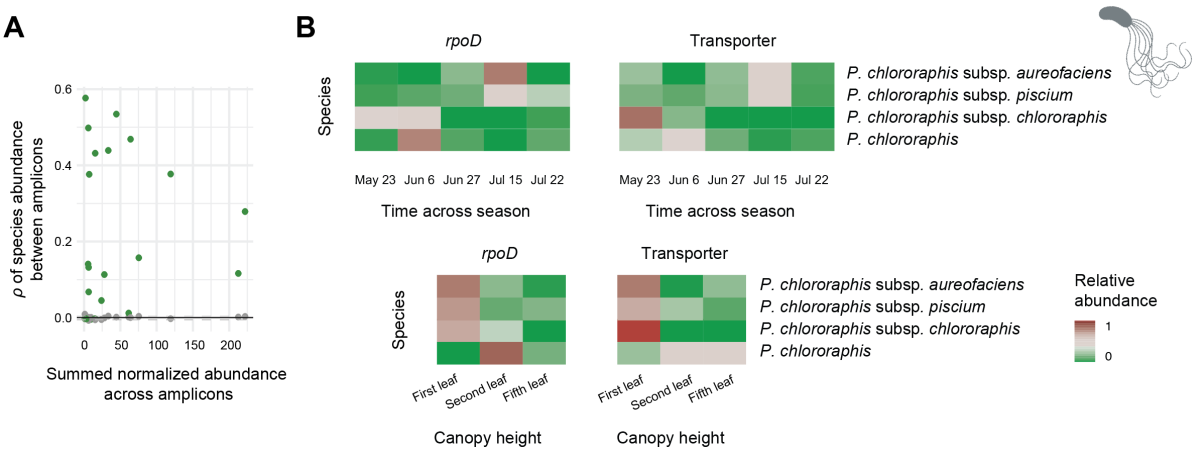


Figure 4: Differential abundance of wheat phyllosphere *Pseudomonas* species across season and canopy height. (A) Spearman correlation of *Pseudomonas* species abundance across samples between the *Pseudomonas*-specific *rpoD* and transporter amplicons. Correlation values were compared to null expectations 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

294 *Pseudomonas* species abundance across samples and between the two amplicons (Figure 4A and
295 Supplementary Figure S5). We compared correlation values to null expectations based on permutations.
296 We observed that 68% and 77% using Spearman and Pearson correlations, respectively, of the species
297 correlations were higher than the 95% confidence interval of the null expectation.

298 Next, we investigated species abundance changes across time and space. We identified different
299 subspecies of *P. chlororaphis* exhibited variable abundance patterns throughout the season and across
300 canopy heights in both amplicons (Figure 4B). *P. chlororaphis* subsP. *aurofaciens* and subsP. *piscium*
301 were most abundant in July, whereas *P. chlororaphis* and *P. chlororaphis* subsP. *chlororaphis* were
302 most abundant in May and June. *P. chlororaphis* subsP. *aurofaciens*, subsP. *piscium* and
303 subsP. *chlororaphis* were most prominent at the bottom of the canopy, whereas *P. chlororaphis* was
304 most abundant on upper leaves.

305

306 WITHIN-SPECIES DIVERSITY OF A MAJOR WHEAT PATHOGEN

307 Genetic diversity within pathogen species can underpin rapid breakdowns of fungicide efficacy or host
308 resistance [39,40]. How genotypic diversity changes over pathogen life cycles remains largely unknown
309 though [41,42]. Here, we examined changes in genotypic diversity across the epidemic phase for the fungal
310 pathogen *Z. tritici* based on species-specific amplicons. The two *Z. tritici*-specific amplicons showed
311 consistent numbers of ASVs per leaf throughout the season, indicating minimal turnover in pathogen
312 diversity within a single epidemic phase (Figure 5A). Similarly, the number of ASVs did not differ
313 significantly across different canopy heights (Figure 5A). Tracking individual ASVs across the season,
314 we observed two distinct groups based on their abundance patterns (Figure 5B). The first group
315 consisted of a small number of highly abundant strains persisting throughout the entire growing season.
316 In contrast, the second group consisted of numerous strains that are predominantly scarce and only
317 detected at specific timepoints.

318

319 PANGENOME-INFORMED AMPLICON TEMPLATES FOR ADDITIONAL TAXA

320 We assessed the potential to generate high-resolution amplicons for other ecologically relevant taxa.
321 We used the same principle as for *Pseudomonas* to define a representative set of high-quality genomes
322 to construct a pangenome to search for core regions conserved within the taxon. Such regions can then
323 be systematically examined for potential amplicon sequences. We generated pangenomes for *Rhizobia*
324 and *Streptomyces* following a systematic procedure to allow for comparison (Figure 6 and
325 Supplementary Figure S6). We also investigated the important environmental fungal species
326 *Aspergillus fumigatus* causing opportunistic human infections. The pangenome analyses defined 22

327 core regions for *Rhizobia*, 515 for *Streptomyces*, and 5,107 for *A. fumigatus*, respectively. These
328 findings indicate genomic regions that are suitable for primer design (Figure 6 and Supplementary
329 Figure S6). The relatively small number of candidate core regions for *Rhizobia* correlates with its
330 greater 16S rRNA gene diversity compared with *Streptomyces* and *Pseudomonas*. Additionally, all
331 pangenome genes were categorized into orthogroups based on protein homology. Our analysis reveals
332 that the proportion of orthogroups present in all isolates of the pangenome varies from 16% to 84%
333 among the analyzed taxon-specific pangenomes. This range reflects the extent of diversity within the
334 pangenome that our method can accommodate (Figure 6 and Supplementary Figure S6).

335

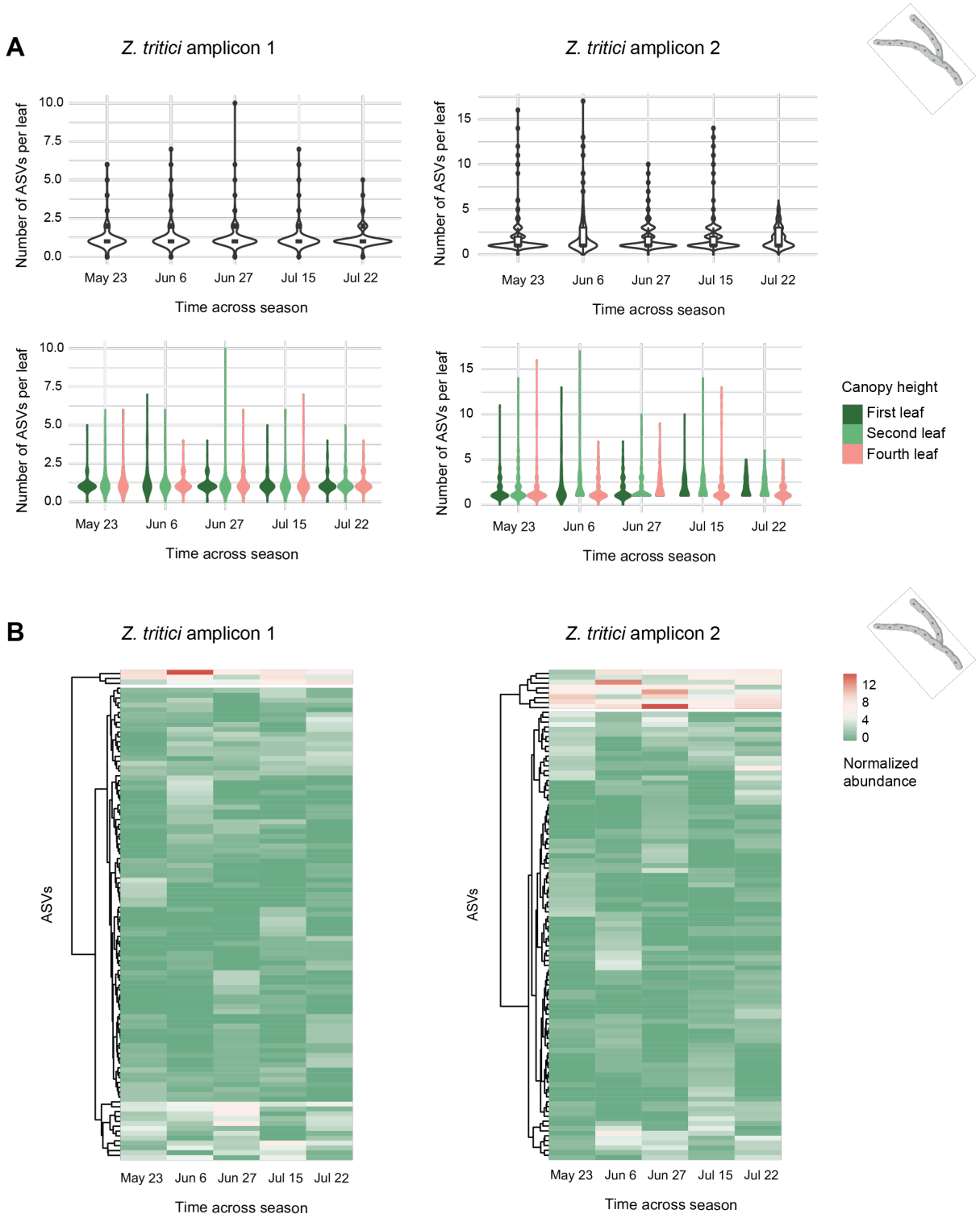


Figure 5: Genotypic diversity of the wheat phyllosphere pathogen *Z. tritici* across season. (A) Number of unique *Z. tritici* amplicon sequence variants (ASVs) per leaf across timepoints based on the *Z. tritici*-specific amplicons on chromosome 9 and 13. The same analysis is shown highlighting canopy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

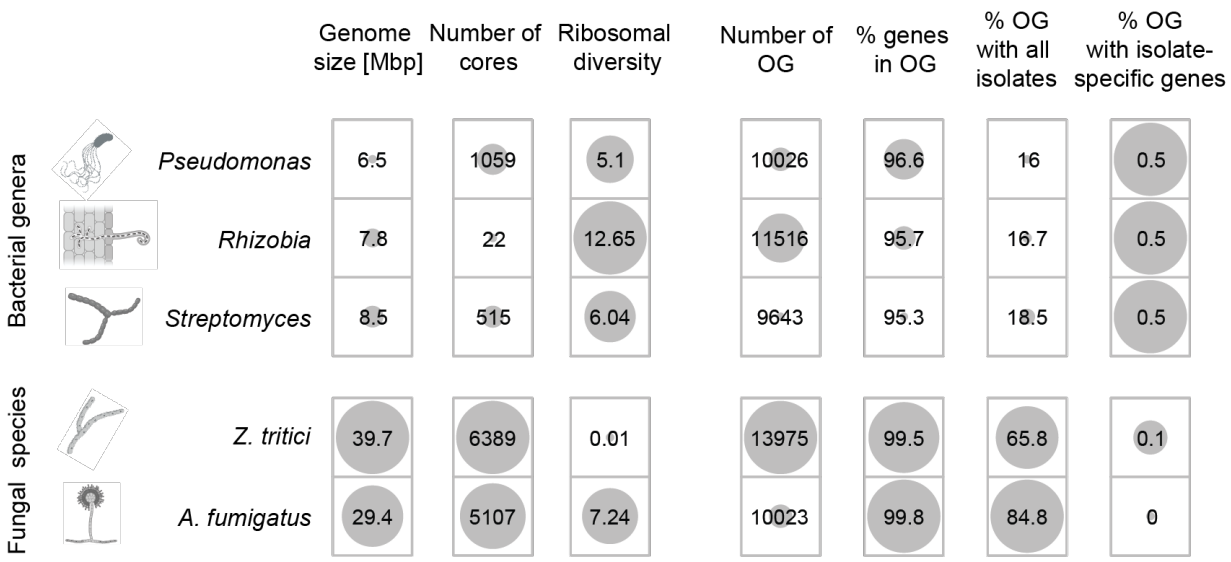


Figure 6: Pangenome and amplicon templates for the bacterial genera *Pseudomonas*, *Rhizobia*, and *Streptomyces*, as well as the fungal species *Z. tritici* and *A. fumigatus*. Pangenomes were each generated based on 19 representative genomes. For each pangenome, the average genome size in Mbp, the number of candidate core segments and the length-normalized ribosomal nucleotide diversity of the 16S or ITS regions is indicated. Genes were grouped into orthogroups based on protein homology for each pangenome. The total number of orthogroups, the percentage of genes included in orthogroups, the percentage of orthogroups that contain all isolates, and the percentage of orthogroups that contain isolate-specific genes are shown for each species group.

352 ## DISCUSSION

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

357 The pangenome-informed design of amplicons for pseudomonads was optimized to capture diversity
358 specifically and exclusively in this ubiquitous group of bacteria. Compared to full-length 16S
359 amplicons, the new amplicons provide consistently higher confidence in species assignments across
360 clades. The confidence in species assignments stems from the high degree of species resolved to
361 monophyletic clades for most *Pseudomonas* species. This enhanced resolution enabled us to
362 differentiate *Pseudomonas* at the subspecies level, which is not feasible using 16S alone. We also
363 focused on the dominant wheat phyllosphere pathogen *Z. tritici* and achieved a strain-level resolution
364 inaccessible with existing single-locus genotyping approaches. The recovered diversity reflects
365 genotyping resolution typically only achievable by whole-genome sequencing [37,38].

366 We evaluated the discriminant power of the new bacterial and fungal amplicons by tracking microbial
367 communities in the wheat phyllosphere. We contrasted the resolution to the universal barcoding loci
368 16S and ITS in samples covering different time points and canopy heights. The increased resolution
369 indeed revealed species and strain-specific changes across time and canopy height inaccessible by
370 ribosomal barcoding datasets. One noteworthy finding was the differential abundance patterns observed
371 for the *P. chlororaphis* subspecies. Subspecies of *P. chlororaphis* have garnered agricultural interest
372 due to their biocontrol potential, with several strains being utilized in commercial formulations [19,43].
373 This potential is attributed to the production of antimicrobial metabolites, notably phenazines, which
374 are not only linked to microbial antagonism but also plant defense upregulation [19,43]. However,
375 phenazine expression and regulation vary among *P. chlororaphis* subspecies. Therefore, monitoring
376 *P. chlororaphis* at the subspecies level is crucial for analyzing antagonistic behavior in the field and
377 prioritizing the most promising biocontrol candidates. The assay developed here is the first to facilitate
378 this. Additionally, we were able to resolve genotypic diversity in a highly diverse fungal pathogen and
379 track changes throughout a single epidemic season. The virulence of different *Z. tritici* isolates varies
380 greatly, making it interesting to track the genotypic diversity across the epidemic season[44,45]. Despite
381 the overall large genotypic diversity (>700 ASVs within a single field), we identified a small set of
382 genotypes that dominated the entire season, which can be informative for control strategies. Our data
383 would also allow us to infer strain-specific bacterial-fungal interactions based on abundance data, to
384 prioritize biocontrol candidates antagonizing specific pathogen isolates.

385 The replicated monitoring of *Pseudomonas* and plant pathogen diversity with two independent
386 amplicons each enabled us to assess the reproducibility of species and strain identification as well as

387 abundance estimates. Certain *Pseudomonas* species were over- or underrepresented in a particular
388 amplicon and across samples. Such discrepancies could stem from differences in primer binding
389 efficiency and variations in amplification success due to amplicon sequence variability [46]. Our findings
390 align with previous studies investigating reproducibility of different 16S amplicons [47]. This underscores
391 the importance of implementing controls in quantitative microbiome studies. The availability of two
392 independent 3-kb amplicons for *Pseudomonas* in addition to a control by the 16S provides a powerful
393 toolset to track the reproducibility of microbiome community assessments across samples and
394 taxonomic groups.

395 In our proof-of-principle demonstration, we multiplexed nearly 10,000 reactions in a single sequencing
396 run. This achievement was made possible by recent improvements in Sequel II. The quality of the
397 sequences was high enough to recover sequences at an impressive rate using asymmetric barcoding,
398 which allowed us to multiplex nearly 10,000 samples with the standard set of barcodes. Therefore, the
399 level of multiplexing in this workflow can be determined solely by the number of reads desired per
400 sample. With taxon-specific amplicons, the sample complexity is lower, making higher multiplexing
401 desirable. Additionally, we enhanced the efficiency and reproducibility of sample handling through
402 robotic liquid handling.

403 The high quality of PacBio HiFi reads enables detailed microbial community profiling using long
404 amplicons. With many high-quality genomes available, we can leverage pangenomes for more
405 integrative analyses of genetic diversity within taxonomic groups and identify the most informative
406 amplicons. Despite the increasing use of pangenomes to analyze genetic diversity, they have not been
407 utilized for amplicon design to the best of our knowledge [48–50]. With the rapidly growing number of
408 genomes, this approach has broad applicability. We demonstrated this by calculating the pangenome
409 for the bacterial groups *Rhizobia* and *Streptomyces*, as well as for the fungal species *A. fumigatus*, and
410 identified an array of candidate regions suitable for primer design. A recent study on *Streptomyces*
411 taxonomy found that 16S rRNA sequence variation does not reliably delineate *Streptomyces* species [51].
412 The study advocates for alternative markers and suggests a reclassification of *Streptomyces* taxonomy,
413 historically based on 16S data. *Rhizobia* is a highly diverse bacterial group characterized by its ability
414 to fix atmospheric nitrogen. Species delineation often involves multilocus sequence analysis using
415 housekeeping genes [52–54]. Longer, high-resolution amplicons could track community dynamics in soil
416 and root nodules, revealing species competitiveness in nodule occupancy. *A. fumigatus* is both an
417 opportunistic human fungal pathogen as well as an environmental saprobe. Despite extensive genomic
418 analysis of the species, the diversity in environmental versus clinical niches is still poorly understood
419 [55,56]. Similar to other fungal species, metagenomic analysis for *A. fumigatus* remains challenging due to
420 the small fraction of fungal DNA in samples and the significantly larger genome compared to bacteria
421 [57]. High-resolution amplicons could provide a suitable approach to gain strain resolution of

environmental and clinical samples, enhancing our understanding of strain competitiveness across different environments.

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). The wheat developed from the growth stage Feekes 7.0 until 11.4 [58]. 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 flag leaf (fourth 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

We established three sets of mock communities using 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 the DNA from two *Pseudomonas* isolates (*P. syringae* Leaf129 and *P. thivervalensis* PITR2) and two *Z. tritici* isolates (ST99CH 1E4 and ST01IR 48b), up to a dilution of 10^{-5}. The DNA input per sample was 15 ng, with 7.5 ng DNA each from *Pseudomonas* and *Z. tritci* for the undiluted sample. For

454 the 10^{-1} dilution, we used 0.75ng DNA for the *Pseudomonas* and *Z. tritici* isolate that was to be diluted,
455 together with 6.75 ng of the background *Pseudomonas* and *Z. tritici* isolate respectively. The second
456 set consisted of a ten-fold serial dilution series up to 10^{-5}, where *P. thivervalensis* PITR2 and ST99CH
457 1E4 were diluted in a leaf DNA extraction sample. We used two different leaf DNA extracts, Leaf 1
458 (Timepoint 2, Cultivar Zinal, Plot2, Plant 2, Leaf height: top) and Leaf 2 (Timepoint 3, Cultivar Titlis,
459 Plot 2, Plant 2, Leaf height: top). The DNA input per sample was 15 ng, starting with 1.5 ng of isolate
460 and 8.5 ng of leaf DNA for the 10^{-1} dilution. For the third set, we combined increasing numbers of
461 *Pseudomonas* isolate DNA extracts (from two to ten isolates) in equimolar concentrations. The total
462 DNA input per mixture was 15 ng, adding 7.5 ng per isolate for the two-strain mixture and 1.5 ng for
463 the 10-strain mixture. To estimate the number of cells per input DNA, we based our calculations on the
464 average genome size of *Pseudomonas* (6.5 Mbp) and *Z. tritici* (39.7 Mbp), respectively. We used a
465 conversion factor of 660 daltons per base pair for these calculations.

466 PANGENOME CONSTRUCTION

467 For pangenome construction, we used Panseq with the following settings: fragmentationSize = 5000,
468 minimumNovelRegionSize = 500, novelRegionFinderMode = no_duplicates, percentIdentityCutoff =
469 60, runMode = pan, storeAlleles = 1, allelesToKeep = 2, frameshift = 1, overwrite = 1,
470 maxNumberResultsInMemory = 500, blastWordSize = 11, nucB = 200, nucC = 65, nucD = 0.12, nucG
471 = 90, nucL = 20, cdhit = 0, sha1 = 0 [59]. To create the *Pseudomonas* pangenome, we selected 19 high-
472 quality genomes representing all subgroups of the genus [32,33] (Supplementary Table 5). For the *Z. tritici*
473 pangenome, we used 19 global reference isolates used for the pangenome analysis of Badet et al. 2020
474 (Supplementary Table 6). For the *A. fumigatus* pangenome creation, we used all publicly available
475 genomes of the pangenome study of Barber et al. 2021 (n=253, Supplementary Table 7). For the
476 *Rhizobia* pangenome we used the genomes of the pangenome analysis of Yang et al. 2020 (n= 84,
477 Supplementary Table 8). For the *Streptomyces* pangenome, we based the genome selection on Kieper
478 et al. 2023. They clustered all available Streptomyces genomes into 186 distinct species based on
479 average nucleotide identity. For our pangenome construction, we aimed to use one genome per species,
480 but filter for complete genomes or chromosome-level assemblies, leaving us with 97 genomes for
481 pangenome construction (Supplementary Table 9). Bacterial plasmid sequences were removed for all
482 pangenome analysis. Furthermore, to be able to better compare the *Rhizobium*, the *Streptomyces* and
483 the *A. fumigatus* pangenome characteristics to the *Pseudomonas* and *Z. tritici* pangenomes, we subset
484 the genome sets to 19 genomes each for a second pangenome construction based on the same number
485 of genomes (Supplementary Table 6-9). Regions that were present in >50% of the pangenome isolates
486 were selected as "core" for further analysis of the amplicon primer design. We used this relaxed
487 criterion to be able to scan divergent regions for amplicons potentially revealing the highest resolution
488 among strains. Supplementary Table 10 details the *Pseudomonas* pangenome core regions, and

489 Supplementary Table 11 details the *Z. tritici* pangenome core regions. Orthologues genes of all
490 pangenomes were identified Orthofinder v2.5.5 with default parameters (Emms and Kelly, 2019).

491 AMPLICON PRIMER DESIGN

492 For each pangenome core fragment of the *Pseudomonas* and the *Z. tritici* pangenome, we created a
493 multiple sequence alignment using muscle v.3 with default parameters [61]. From each core multiple
494 sequence alignment, we created a consensus sequence using the EMBOSS v.6 cons function, setting
495 the identity to ten and plurality to 0.8 [62]. We then ran Primer3 v. 2.5 on each core consensus sequence
496 [63]. We used the following Primer3 settings: PRIMER_PICK_LEFT_PRIMER=1,
497 PRIMER_PICK_INTERNAL_OLIGO=0, PRIMER_PICK_RIGHT_PRIMER=1,
498 PRIMER_OPT_SIZE=20, PRIMER_MIN_SIZE=18, PRIMER_MAX_SIZE=22,
499 PRIMER_PRODUCT_SIZE_RANGE=1500-3100, PRIMER_NUM_RETURN=8,
500 PRIMER_EXPLAIN_FLAG=1. For each primer pair, we cut out the respective amplicon from the
501 multiple sequence alignment using the EMBOSS v. 6 extractalign function [62]. To evaluate amplicon
502 diversity, we calculated the nucleotide diversity of each amplicon multiple sequence alignment within
503 the pangenome. For this, we first converted the multiple sequence alignment to vcf format using snp-
504 sites [64], and then vcftools v0.1.16 with the options –sites-pi –haploid to calculate the nucleotide diversity
505 [65]. For both *Pseudomonas* and *Z. tritici*, we selected the ten amplicons with the greatest summed
506 nucleotide diversity for further evaluation. Each *Pseudomonas* candidate primer pair was then blasted
507 against all available *Pseudomonas* genomes from the ncbi nucleotide collection using BLASTn to
508 ensure matches in all genomes [66]. All hits were aligned using MAFFT v7.427 with the option –auto [67,68].
509 For *Z. tritici*, we analyzed each candidate amplicon against the worldwide collection of 1109 *Z. tritici*
510 isolates [69]. Here, amplicon sequences were extracted from variant call files (i.e. vcf) using bedtools filter
511 [70]. For both *Pseudomonas* and *Z. tritici* amplicon evaluation, we assessed allele frequencies at every
512 position of the multiple sequence alignment using base R functions [71]. We manually adjusted primer
513 sequences to account for the specific combinations of alleles necessary to amplify all known
514 *Pseudomonas,* or *Z. tritici* sequences, respectively.

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

519 SAMPLE HOMOGENIZATION, DNA EXTRACTION, AMPLIFICATION, POOLING AND
520 CLEANUP

521 Leaves were lyophilized for 48 h and weighed. Then, the complete leaves were homogenized using
522 0.5 mm and 0.2 mm zirconium beads in the Bead Ruptor bead mill homogenizer (OMNI) using the

523 following settings: speed 5.00, number of cycles 2, time of cycle 1:00, time distance between cycles
524 1:00. DNA extraction was performed with automated magnetic-particle processing using the
525 KingFisher Flex Purification Systems (Thermo Scientific). To enhance the DNA extraction of fungal
526 and bacterial DNA, lyticase and lysozyme was added to the first lysis step with PVP buffer. Specifically,
527 for 10mg dry leaf mass 3.9 µl lyticase (200,000 U/mg, diluted to 6.5 mg/ml), 3.9 µl lysozyme
528 (22,500 U/mg, diluted to 10 mg/ml) and 98µl PVP lysis buffer was added, and samples were incubated
529 at 55°C for 30min. Then, 3.9µl proteinase K (30 U/mg, diluted to 10mg/ml) for 10mg dry mass was
530 added and incubated at 55°C for 30min. From each sample, 150µl of clear lysate was transferred to an
531 empty binding plate (KingFisher Flex, Thermo Scientific). For each sample, 360 µl PN binding buffer,
532 30 µl well suspended Sbeadex beads were added. Using the KingFisher Flex, the first washing step was
533 performed using 400 µl PN1 buffer per sample, then a second wash using 390 µl buffer PN1 with 10µl
534 RNase A (diluted to 10mg/µl in water) and a third wash using 400 µl PN2 buffer. Each sample was
535 eluted in 100 µl nuclease-free water. The DNA concentration was measured using the Spark Microplate
536 reader (Tecan). Then, DNA concentrations were diluted to 5 ng/µl using the Liquid Handling Station
537 (BRAND). PCR reactions were pipetted using the Mosquito HV liquid handling robot (SPT Labtech).
538 The first amplicon PCR reaction was performed in a 15 µl reaction volume. Specifically, 7.5 µl KAPA
539 HiFi HotStart ReadyMix (2x), 1.5µl forward primer (3 µM), 1.5ul reverse primer (3 µM), 3 ul DNA
540 (5 ng/µl) and 1.5 µl HPLC water were combined. Primer sequences and cycling protocols are
541 documented in Supplementary Table 13 and 14. All primers were synthesized by IDT (Integrated DNA
542 Technologies, Coralville, IA). *Pseudomonas* and *Z. tritici*-specific amplicon PCR products were diluted
543 1:5, PCR products from 16S and ITS 1:10. The second barcoding PCR reaction was performed in a
544 25 µl reaction volume. Specifically, 12.5 µl KAPA HiFi HotStart ReadyMix (2x), 2.5 µl M13 forward
545 barcode (3 µM), 1.5 µl M13 reverse barcode (3 µM), 2 µl diluted PCR product and 5.5 µl HPLC water
546 were combined. Barcode sequences are available in Supplementary Table 15. Samples were pooled by
547 amplicon taking 1.5 µl from each barcoded product. Each amplicon pool was cleaned using AMPure
548 XP beads using a bead ratio of 0.8x.

549 LIBRARY PREPARATION AND SEQUENCING

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

557 PACBIO RAW READ PROCESSING

558 CCS were extracted from raw reads using the ccs software from the bioconda package pbccs provided
559 by the manufacturer (Pacific Biosciences). For the first run pbccs v. 6.0.0 was used and for the second
560 run v. 6.3.0. We split the CCS reads by barcodes using the software lima 2.0.0 (Pacific Biosciences)
561 with the following parameters lima –log-level INFO –per-read –min-passes 0 –split-bam-named –ccs
562 –different -A 1 -B 3 -D 2 -I 2 -X 0. We assigned the reads to the respective amplicons using BLASTn
563 assignments to reference amplicon from the *P. fluorescens* CHA0 and *Z. tritici* isolate 1E4 [66]. All six
564 reference amplicons were blasted against each read and reads were then assigned to the reference hit
565 with the lowest e-value, the highest length and the highest identity. We removed primer sequences using
566 cutadapt v. 3.4 with the following parameters: cutadapt -a
567 FORWARD_PRIMER_SEQ...REVERSE_PRIMER_SEQ –discard-untrimmed –revcomp [72]. Primers
568 were treated as linked, i.e. reads without primers at both ends were discarded. The R package dada2 v.
569 1.28.0 was used to infer ASVs for each amplicon separately [73]. In the following, dada2 steps for each
570 amplicon are described. Reads were filtered and trimmed using the function filterAndTrim with the
571 parameters minLen=minLength, maxLen=maxLength, rm.phix=FALSE, maxEE=2, qualityType =
572 "FastqQuality", multithread=TRUE. Length ranges for each amplicon are described in Supplementary
573 Table 9. Reads were dereplicated using the function derepFastq with the parameters verbose=TRUE,
574 qualityType="FastqQuality". Error models were estimated with the function learnErrors with the
575 parameters errorEstimationFunction=dada2:::PacBioErrfun, BAND_SIZE=32, multithread=TRUE.
576 Reads were denoised using the function dada with the parameters BAND_SIZE=32,
577 multithread=TRUE, pool=FALSE. Importantly, we did not used the option pool=TRUE, as this
578 introduced spurious ASVs in the reference cultures of the mock community. ASV sequence tables were
579 generated using the function makeSequenceTable. To remove chimeras from the sequence table, the
580 function removeBimeraDenovo with parameters method="consensus",
581 minFoldParentOverAbundance=3.5, multithread=TRUE, verbose=TRUE was used. Supplementary
582 Figure S7 shows the read tracking through the dada2 pipeline for each amplicon.

583 TAXONOMIC CLASSIFICATION

584 We assigned 16S reads using the dada2-formatted Silva database v. 138 [74] and ITS reads to the UNITE
585 database v. 8.3 [75]. To perform taxonomy assignment, we used the function assignTaxonomy from the
586 dada2 package. As the UNITE database comprises mostly ITS1-ITS2 reference sequences and only few
587 full-length ITS-LSU sequences, we truncated all ITS reads to the ITS1-ITS2 fragment for assignment.
588 We blasted the reads against the ITS1-ITS2 fragment of *Z.tritici* strain S-46 (KT336200.1) and used
589 the hit coordinates to truncate reads with the seqkit software function subseq [76]. We assigned
590 *Pseudomonas* reads of the *rpoD*, transporter and 16S amplicons to *Pseudomonas* species using
591 BLASTn against all 1071 full-length *Pseudomonas* genomes available from the *Pseudomonas* db v.
592 21.1 (2022-11-20) (Supplementary Table 4) [66,77]. The best assignment was chosen according to BLASTn

593 bitscores. We assigned *Z. tritici* reads of the *Z. tritici* amplicons on chromosomes 9 and 13, and of the
594 ITS amplicon to a database of previously sequenced *Z. tritici* strains. For this, we used draft assemblies
595 of previously collected 177 genomes from the Eschikon Field Station of the ETH Zurich, Switzerland
596 [37], as well as genomes from the reference pangenome [34]. The best assignment was chosen according to
597 bitscores.

598 ALIGNMENT, PHYLOGENETIC TREE AND NETWORK CONSTRUCTION

599 We performed multiple sequence alignment using PASTA v.1.9.0 with the following MAFFT
600 arguments --leavegappyregion --6merpair --maxiterate 0 --adjustdirection --reorder and FastTree model
601 -gtr -gamma -fastest [78]. We built phylogenetic trees of leaf samples using FastTree v.2.0.0 [79], and of the
602 *Pseudomonas* mock samples using raxml-ng v. 1.2.0 with the options --all --model GTR+G --opt-model
603 on --threads 4 --seed 2 --outgroup "Main ASV of *P. syringae* Leaf129" [80]. To create unrooted
604 phylogenetic networks for the *Z. tritici* amplicon 1 on chromosome 13, amplicon 2 on chromosome 9
605 and the ITS, the software SplitsTree v. 4.19.1 was used using uncorrected p distances [81].

606 STATISTICS

607 Regression analyses were performed using the R function lm from the package stats v. 4.2.2 [82].
608 Correlations were calculated using the R function cor.test from the package stats v. 4.2.2 [82]. Shannon
609 diversity was calculated using the function diversity from the R package vegan v. 2.6.4 [83]. Permutations
610 were calculated for 1000 iterations using the base R function sample with replace=TRUE [71]. Subtrees
611 were identified using the function subtrees from the R package ape v. 5.7.1 [84].

612 VISUALIZATION

613 To visualize ASV abundances, the R package phyloseq v. 1.42.0 was used [85]. Counts were normalized
614 by sample using the phyloseq function transform_sample_counts(phyloseq_object, function (x) x /
615 sum(x)). The heatmap was created using the R package pheatmap v. 1.0.12 [86]. Table numbers were
616 visualized with the method = "circle" of the R package corrplot v. 0.92 [87]. Phylogenetic trees were
617 visualized using the R package ggtree v. 3.6.2 [88,89]. All other plots were created using the R package
618 ggplot2 v. 3.4.2 [90]. Organism and machine icons were created with BioRender.com.

619

620 DATA AVAILABILITY

621 Sequencing data generated for this work are available at the NCBI Sequence Read Archive (SRA).
622 Analyzed datasets are provided as supplementary information.

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.

FUNDING

MM and DC were supported by the Swiss National Science Foundation (grants 177052 and 201149).

REFERENCES

1. Song, S., Liu, Y., Wang, N. R. & Haney, C. H. Mechanisms in plant-microbiome interactions: lessons from model systems. *Curr Opin Plant Biol* **62**, (2021).

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

3. Xu, L. *et al.* Holo-omics for deciphering plant-microbiome interactions. *Microbiome* **9**, (2021).

4. Choi, K., Khan, R. & Lee, S. W. Dissection of plant microbiota and plant-microbiome interactions. *J Microbiol* **59**, 281–291 (2021).

5. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* **18**, 607–621 (2020).

6. Vorholt, J. A., Vogel, C., Carlström, C. I. & Müller, D. B. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host Microbe* **22**, 142–155 (2017).

7. Liu, Y. X. *et al.* A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* **12**, 315–330 (2021).

8. Earl, J. P. *et al.* Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome* **6**, (2018).

9. Karst, S. M. *et al.* High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods* **18**, 165–169 (2021).

10. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology 2019 37:10* **37**, 1155–1162 (2019).

657 11. Wagner, J. *et al.* Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene
658 classification. *BMC Microbiol* **16**, 1–17 (2016).

659 12. Callahan, B. J. *et al.* High-throughput amplicon sequencing of the full-length 16S rRNA gene
660 with single-nucleotide resolution. *Nucleic Acids Res* **47**, E103 (2019).

661 13. Escapa, I. F. *et al.* Construction of habitat-specific training sets to achieve species-level
662 assignment in 16S rRNA gene datasets. *Microbiome* **8**, 65 (2020).

663 14. Xin, X. F., Kvitko, B. & He, S. Y. Pseudomonas syringae: What it takes to be a pathogen. *Nat*
664 *Rev Microbiol* **16**, 316–328 (2018).

665 15. Großkinsky, D. K. *et al.* Cytokinin production by Pseudomonas fluorescens G20-18 determines
666 biocontrol activity against Pseudomonas syringae in Arabidopsis. *Sci Rep* **6**, (2016).

667 16. Gislason, A. S. & de Kievit, T. R. Friend or foe? Exploring the fine line between Pseudomonas
668 brassicacearum and phytopathogens. *J Med Microbiol* **69**, 347–360 (2020).

669 17. Zboralski, A. & Filion, M. Genetic factors involved in rhizosphere colonization by
670 phytobeneficial Pseudomonas spp. *Comput Struct Biotechnol J* **18**, 3539 (2020).

671 18. Costa-Gutierrez, S. B., Adler, C., Espinosa-Urgel, M. & de Cristóbal, R. E. Pseudomonas putida
672 and its close relatives: mixing and mastering the perfect tune for plants. *Appl Microbiol*
673 *Biotechnol* **106**, 3351–3367 (2022).

674 19. Raio, A. & Puopolo, G. Pseudomonas chlororaphis metabolites as biocontrol promoters of plant
675 health and improved crop yield. *World J Microbiol Biotechnol* **37**, (2021).

676 20. Weller, D. M. Pseudomonas biocontrol agents of soilborne pathogens: looking back over 30
677 years. *Phytopathology* **97**, 250–256 (2007).

678 21. Passera, A. *et al.* Not Just a Pathogen? Description of a Plant-Beneficial Pseudomonas syringae
679 Strain. *Front Microbiol* **10**, (2019).

680 22. Bailly, A. & Weisskopf, L. Mining the Volatilomes of Plant-Associated Microbiota for New
681 Biocontrol Solutions. *Front Microbiol* **8**, (2017).

682 23. Biessy, A. & Filion, M. Phenazines in plant-beneficial Pseudomonas spp.: biosynthesis,
683 regulation, function and genomics. *Environ Microbiol* **20**, 3905–3917 (2018).

684 24. Laino, P., Limonta, M., Gerna, D. & Vaccino, P. Morpho-physiolological and qualitative traits
685 of a bread wheat collection spanning a century of breeding in Italy. *Biodiversity Data Journal*
686 *3: e4760* **3**, e4760- (2015).

687 25. Barroso-Bergada, D. *et al.* NGS data reveals the phyllosphere microbiome of wheat plants
688 infected by the fungal pathogen Zymoseptoria tritici. *bioRxiv* (2022)
689 doi:10.1101/2022.02.07.479402.

690 26. Kerdraon, L., Barret, M., Laval, V. & Suffert, F. Differential dynamics of microbial community
691 networks help identify microorganisms interacting with residue-borne pathogens: The case of
692 Zymoseptoria tritici in wheat. *Microbiome* **7**, (2019).

693 27. Sapkota, R., Jørgensen, L. N. & Nicolaisen, M. Spatiotemporal variation and networks in the
694 mycobiome of the wheat canopy. *Front Plant Sci* **8**, (2017).

695 28. Kavamura, V. N., Mendes, R., Bargaz, A. & Mauchline, T. H. Defining the wheat microbiome:
696 Towards microbiome-facilitated crop production. *Comput Struct Biotechnol J* **19**, 1200 (2021).

697 29. Chen, J. *et al.* Wheat Microbiome: Structure, Dynamics, and Role in Improving Performance
698 Under Stress Environments. *Front Microbiol* **12**, (2021).

699 30. Seybold, H. *et al.* A fungal pathogen induces systemic susceptibility and systemic shifts in wheat
700 metabolome and microbiome composition. *Nature Communications 2020 11:1* **11**, 1–12 (2020).

701 31. Levy, E., Eyal, Z. & Chet, I. Suppression of septoria tritici blotch and leaf rust on wheat seedling
702 leaves by pseudomonads. *Plant Pathol* **37**, 551–557 (1988).

703 32. Gomila, M., Peña, A., Mulet, M., Lalucat, J. & García-Valdés, E. Phylogenomics and
704 systematics in Pseudomonas. *Front Microbiol* **6**, 1–13 (2015).

705 33. Peix, A., Ramírez-Bahena, M. H. & Velázquez, E. The current status on the taxonomy of
706 Pseudomonas revisited: An update. *Infection, Genetics and Evolution* **57**, 106–116 (2018).

707 34. Badet, T., Oggenfuss, U., Abraham, L., McDonald, B. A. & Croll, D. A 19-isolate reference-
708 quality global pangenome for the fungal wheat pathogen Zymoseptoria tritici. *BMC Biol* **18**, 1–
709 18 (2020).

710 35. Flury, P. *et al.* Insect pathogenicity in plant-beneficial pseudomonads: Phylogenetic distribution
711 and comparative genomics. *ISME Journal* **10**, 2527–2542 (2016).

712 36. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level
713 microbiome analysis. *Nature Communications 2019 10:1* **10**, 1–11 (2019).

714 37. Singh, N. K., Karisto, P. & Croll, D. Population-level deep sequencing reveals the interplay of
715 clonal and sexual reproduction in the fungal wheat pathogen Zymoseptoria tritici. *Microb*
716 *Genom* **7**, 678 (2021).

717 38. McDonald, B. A., Suffert, F., Bernasconi, A. & Mikaberidze, A. How large and diverse are field
718 populations of fungal plant pathogens? The case of Zymoseptoria tritici. *Evol Appl* **15**, 1360
719 (2022).

720 39. Ebert, D. & Fields, P. D. Host-parasite co-evolution and its genomic signature. *Nat Rev Genet*
721 (2020) doi:10.1038/s41576-020-0269-1.

722 40. Lucas, J. A., Hawkins, N. J. & Fraaije, B. A. *The Evolution of Fungicide Resistance*. *Advances*
723 *in Applied Microbiology* vol. 90 (Elsevier Ltd, 2015).

724 41. Tack, A. J. M., Thrall, P. H., Barrett, L. G., Burdon, J. J. & Laine, A. L. Variation in infectivity
725 and aggressiveness in space and time in wild host–pathogen systems: causes and consequences.
726 *J Evol Biol* **25**, 1918–1936 (2012).

727 42. Barrett, L. G., Thrall, P. H., Burdon, J. J. & Linde, C. C. Life history determines genetic structure
728 and evolutionary potential of host-parasite interactions. *Trends Ecol Evol* **23**, 678–685 (2008).

43. Anderson, A. J. & Kim, Y. C. Insights into plant-beneficial traits of probiotic Pseudomonas chlororaphis isolates. *J Med Microbiol* **69**, 361–371 (2020).

44. Sánchez-Vallet, A. *et al.* The Genome Biology of Effector Gene Evolution in Filamentous Plant Pathogens. *https://doi.org/10.1146/annurev-phyto-080516-035303* **56**, 21–40 (2018).

45. Plissonneau, C., Hartmann, F. E. & Croll, D. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol* **16**, 1–16 (2018).

46. Yeh, Y. C. *et al.* Comprehensive single-PCR 16S and 18S rRNA community analysis validated with mock communities, and estimation of sequencing bias against 18S. *Environ Microbiol* **23**, 3240–3250 (2021).

47. Wear, E. K., Wilbanks, E. G., Nelson, C. E. & Carlson, C. A. Primer selection impacts specific population abundances but not community dynamics in a monthly time-series 16S rRNA gene amplicon analysis of coastal marine bacterioplankton. *Environ Microbiol* **20**, 2709 (2018).

48. Brockhurst, M. A. *et al.* The Ecology and Evolution of Pangenomes. *Curr Biol* **29**, R1094–R1103 (2019).

49. Cummins, E. A., Hall, R. J., McInerney, J. O. & McNally, A. Prokaryote pangenomes are dynamic entities. *Curr Opin Microbiol* **66**, 73–78 (2022).

50. Tettelin, H. & Medini, D. The Pangenome. *The Pangenome: Diversity, Dynamics and Evolution of Genomes* 1–307 (2020) doi:10.1007/978-3-030-38281-0.

51. Kiepas, A. B., Hoskisson, P. A. & Pritchard, L. 16S rRNA phylogeny and clustering is not a reliable proxy for genome-based taxonomy in Streptomyces. *bioRxiv* (2023) doi:10.5281/zenodo.8223787.

52. Rajkumari, J., Katiyar, P., Dheeman, S., Pandey, P. & Maheshwari, D. K. The changing paradigm of rhizobial taxonomy and its systematic growth upto postgenomic technologies. *World Journal of Microbiology and Biotechnology 2022 38:11* **38**, 1–23 (2022).

53. Miranda-Sánchez, F., Rivera, J. & Vinuesa, P. Diversity patterns of Rhizobiaceae communities inhabiting soils, root surfaces and nodules reveal a strong selection of rhizobial partners by legumes. *Environ Microbiol* **18**, 2375–2391 (2016).

54. Yang, L.-L., Jiang, Z., Li, Y., Wang, E.-T. & Zhi, X.-Y. Plasmids Related to the Symbiotic Nitrogen Fixation Are Not Only Cooperated Functionally but Also May Have Evolved over a Time Span in Family Rhizobiaceae. *GBE* (2020) doi:10.1093/gbe/evaa152.

55. Barber, A. E. *et al.* Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. *Nat Microbiol* **6**, 1526–1536 (2021).

56. Kowalski, C. H. *et al.* Heterogeneity among Isolates Reveals that Fitness in Low Oxygen Correlates with Aspergillus fumigatus Virulence. *mBio* **7**, (2016).

57. Usyk, M. *et al.* Comprehensive evaluation of shotgun metagenomics, amplicon sequencing, and harmonization of these platforms for epidemiological studies. *Cell Reports Methods* **3**, 100391 (2023).

767 58. Large, E. C. Growth stages in cereals - illustration of the Feekes scale. *Plant Pathol* **3**, 128–129
768 (1954).

769 59. Laing, C. *et al.* Pan-genome sequence analysis using Panseq: An online tool for the rapid
770 analysis of core and accessory genomic regions. *BMC Bioinformatics* **11**, 1–14 (2010).

771 60. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative
772 genomics. *Genome Biol* **20**, (2019).

773 61. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
774 *Nucleic Acids Res* **32**, 1792 (2004).

775 62. Rice, P., Longden, L. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software
776 Suite. *Trends Genet* **16**, 276–277 (2000).

777 63. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res* **40**, e115
778 (2012).

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

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

782 66. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search
783 tool. *J Mol Biol* **215**, 403–410 (1990).

784 67. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
785 Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).

786 68. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple
787 sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).

788 69. Feurtey, A. *et al.* A thousand-genome panel retraces the global spread and adaptation of a major
789 fungal crop pathogen. *Nat Commun* **14**, (2023).

790 70. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic
791 features. *Bioinformatics* **26**, 841–842 (2010).

792 71. R Core Team. R: A Language and Environment for Statistical Computing. https://www.r-
793 project.org/ (2022).

794 72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
795 *EMBnet J* **17**, 10–12 (2011).

796 73. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data.
797 *Nature Methods 2016 13:7* **13**, 581–583 (2016).

798 74. Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic
799 frameworks. *Nucleic Acids Res* **42**, D643–D648 (2014).

800 75. Nilsson, R. H. *et al.* The UNITE database for molecular identification of fungi: handling dark
801 taxa and parallel taxonomic classifications. *Nucleic Acids Res* **47**, D259–D264 (2019).

802 76. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q
803 File Manipulation. *PLoS One* **11**, e0163962 (2016).

804 77. Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of
805 Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res* **44**, D646–
806 D653 (2016).

807 78. Mirarab, S. *et al.* PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and
808 Amino-Acid Sequences. *Journal of Computational Biology* **22**, 377 (2015).

809 79. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees
810 with Profiles instead of a Distance Matrix. *Mol Biol Evol* **26**, 1641 (2009).

811 80. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
812 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

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

815 82. Best, D. J. & Roberts, D. E. Algorithm AS 89: The Upper Tail Probabilities of Spearman's Rho.
816 *Appl Stat* **24**, 377 (1975).

817 83. Oksanen Jari. vegan: Community Ecology Package. Preprint at (2022).

818 84. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary
819 analyses in R. *Bioinformatics* **35**, 526–528 (2019).

820 85. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis
821 and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).

822 86. Kolde, R. pheatmap: Pretty Heatmaps. (2019).

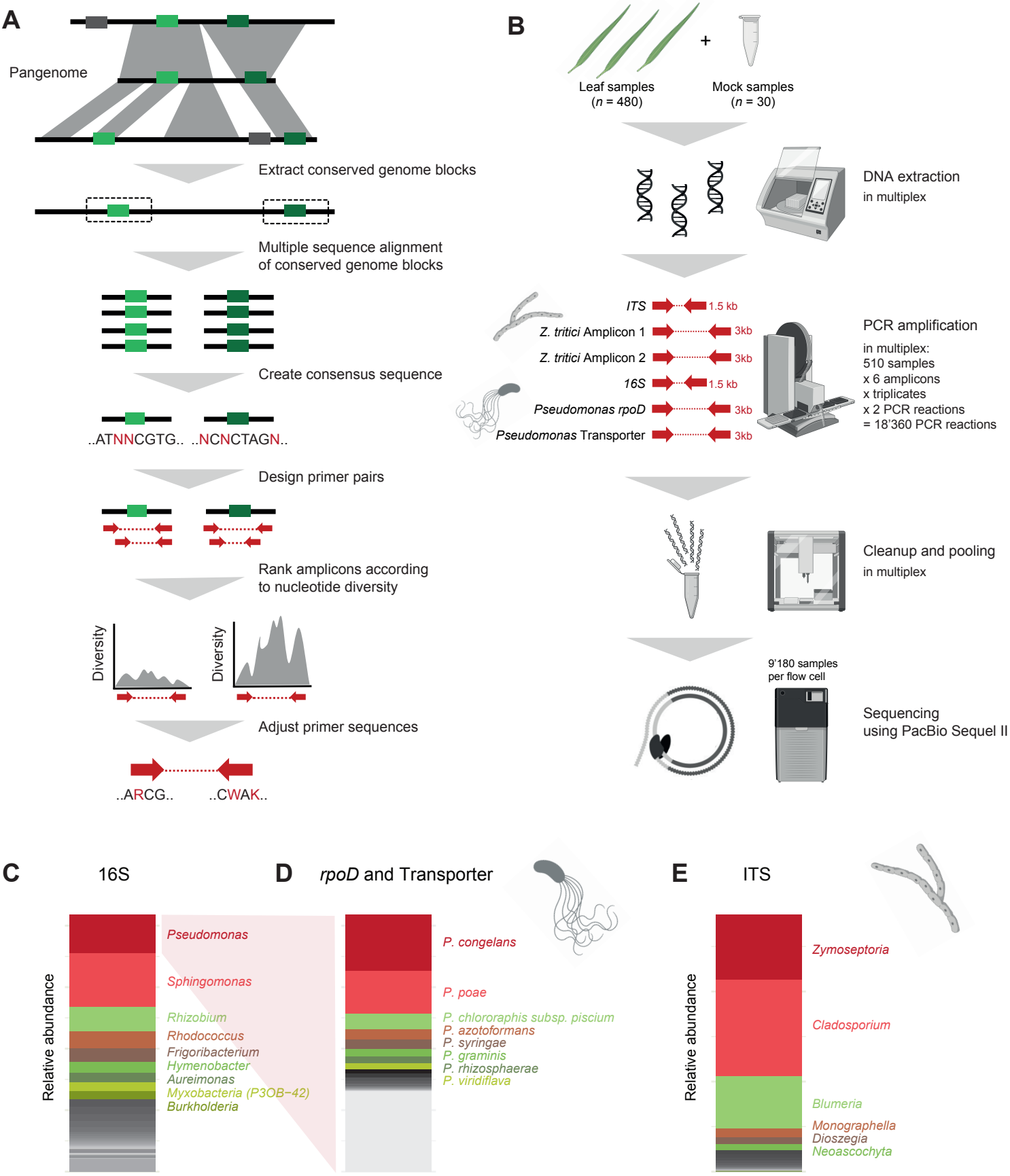
823 87. Wei, T. & Simko, V. R package 'corrplot': Visualization of a Correlation Matrix. (2021).

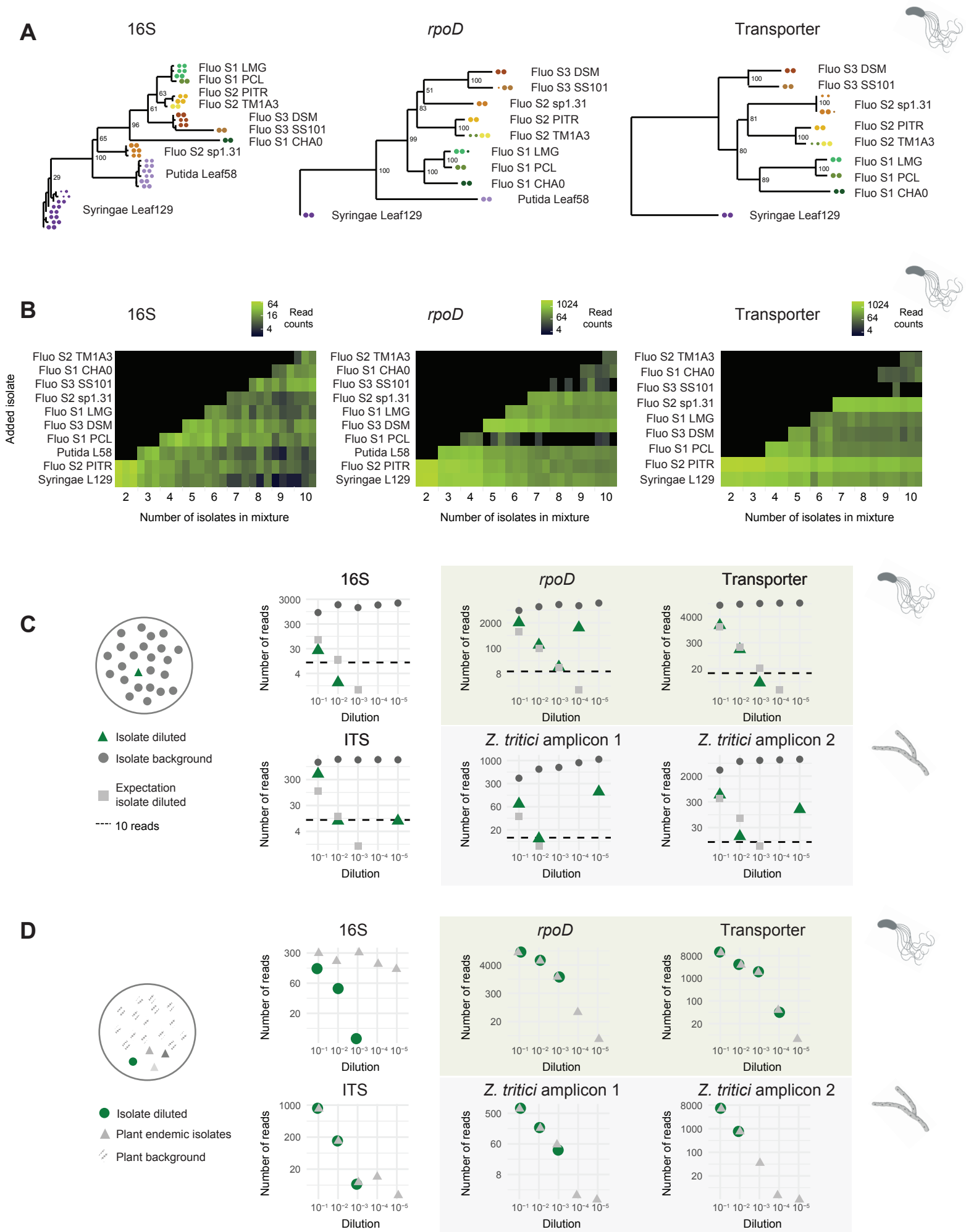
824 88. Curry, E. & Zhang, H. *Data Integration, Manipulation and Visualization of Phylogenetic Trees*.
825 *Data Integration, Manipulation and Visualization of Phylogenetic Trees* (Chapman and
826 Hall/CRC, 2022). doi:10.1201/9781003279242.

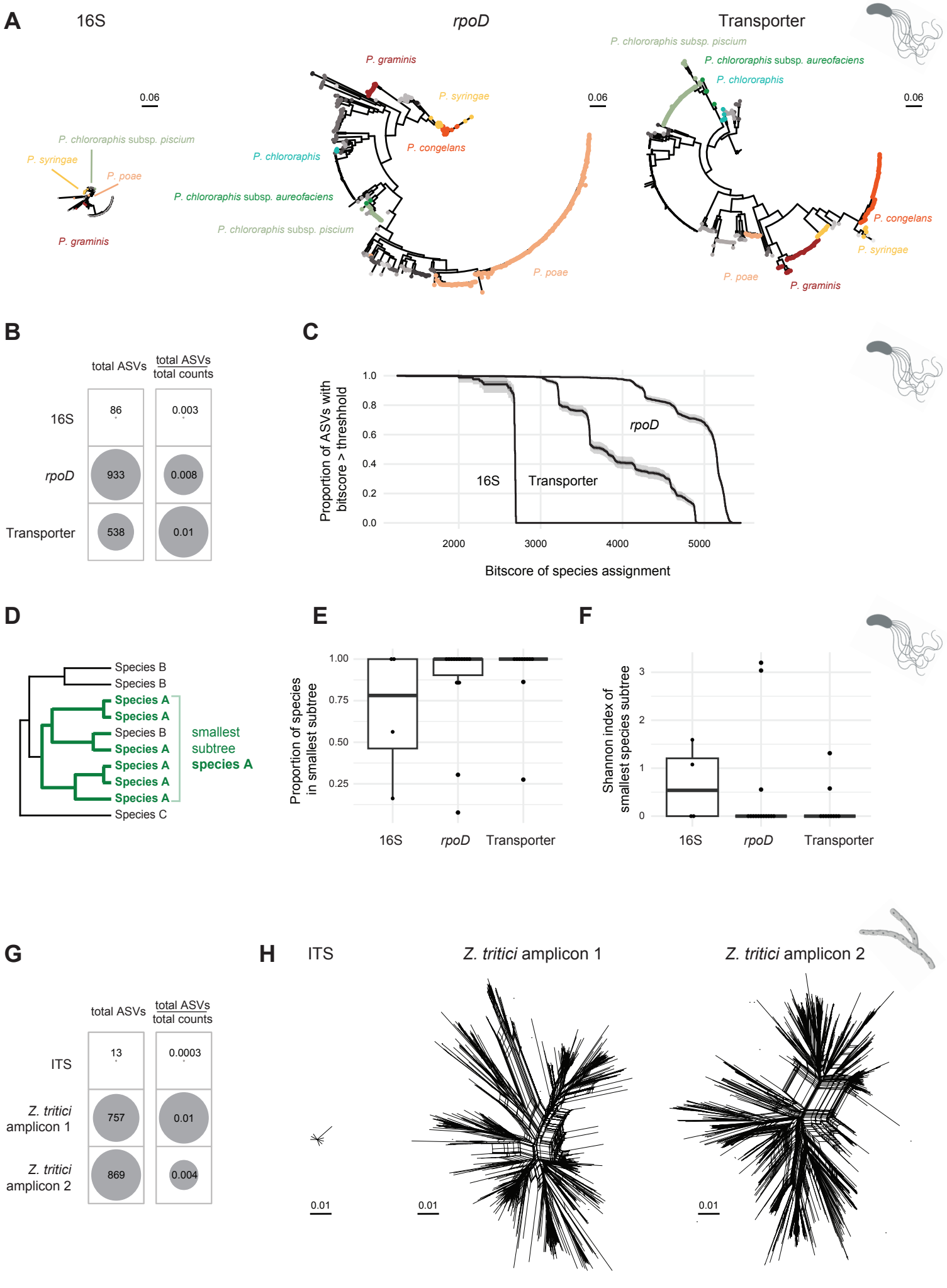
827 89. Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics* **69**,
828 (2020).

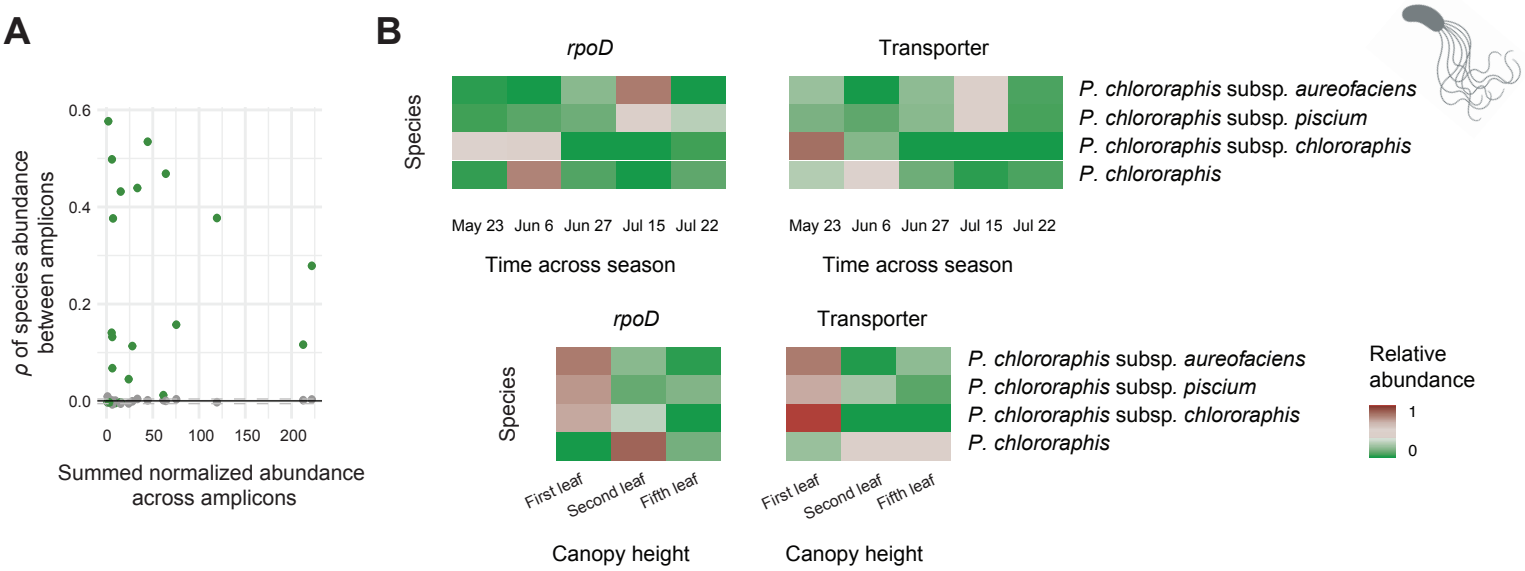
829 90. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer International Publishing,
830 Cham, 2016). doi:10.1007/978-3-319-24277-4.

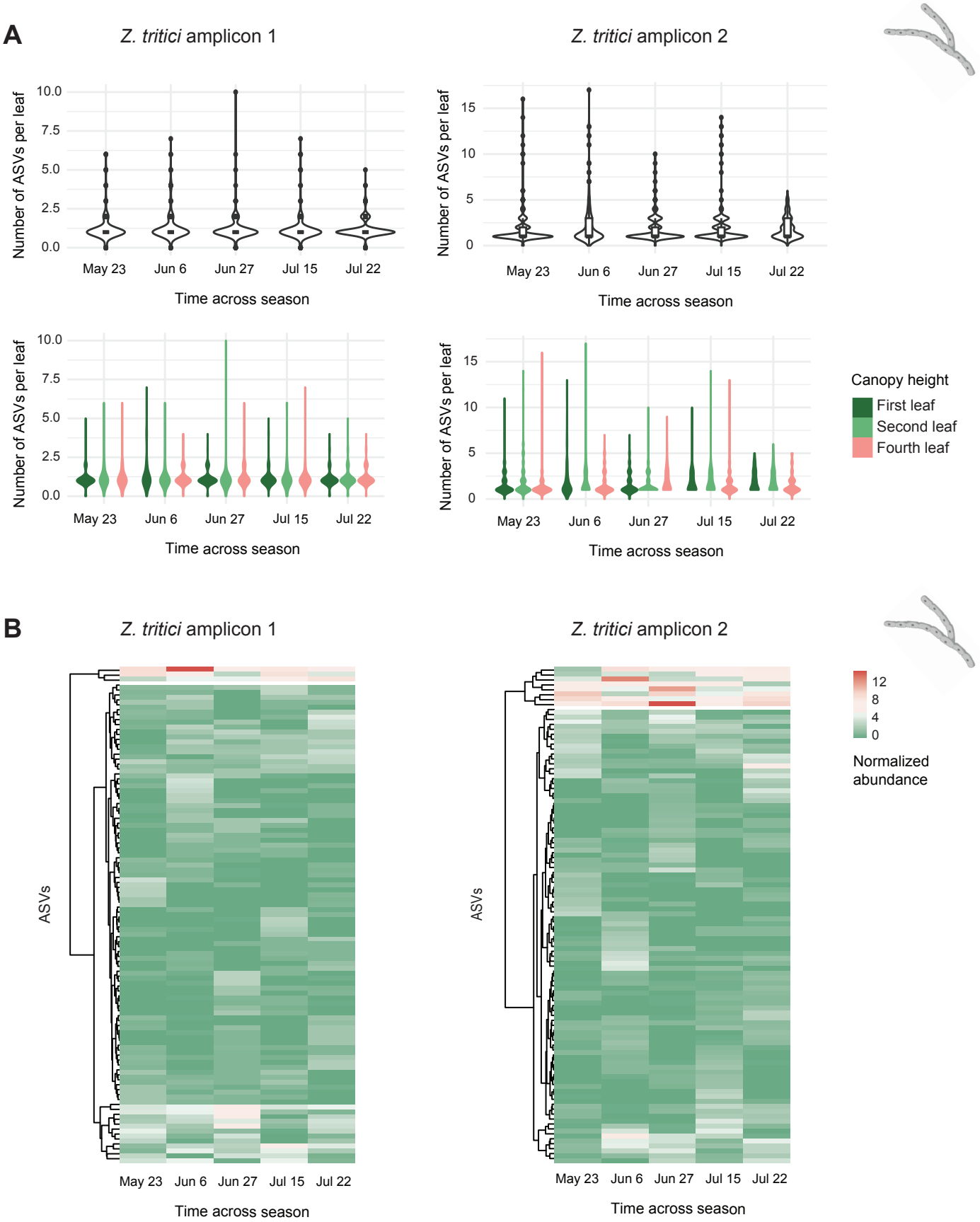
831


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
Leaf samples
(n = 480)
+
Mock samples
(n = 30)
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
B
16S
rpoD
Transporter
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 L58
Fluo S2 PITR
Syringae L129
Number of isolates in mixture
C
Isolate diluted
Isolate background
Expectation isolate diluted
10 reads
16S
rpoD
Transporter
ITS
Z. tritici amplicon 1
Z. tritici amplicon 2
Number of reads
Dilution
D
Isolate diluted
Plant endemic isolates
Plant background
16S
rpoD
Transporter
ITS
Z. tritici amplicon 1
Z. tritici amplicon 2
Number of reads
Dilution


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 ASVs 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
Fourth leaf
B
Z. tritici amplicon 1
Z. tritici amplicon 2
ASVs
12
8
4
0
Normalized abundance

		Genome size [Mbp]	Number of cores	Ribosomal diversity	Number of OG	% genes in OG	% OG with all isolates	% OG with isolate-specific genes
Bacterial genera	*Pseudomonas*	6.5	1059	5.1	10026	96.6	16	0.5
	Rhizobia	7.8	22	12.65	11516	95.7	16.7	0.5
	Streptomyces	8.5	515	6.04	9643	95.3	18.5	0.5
Fungal species	*Z. tritici*	39.7	6389	0.01	13975	99.5	65.8	0.1
	A. fumigatus	29.4	5107	7.24	10023	99.8	84.8	0