Materials and Methods

Collecting and processing of field *A. thaliana* leaf samples

We shared most part of the sampling trip with Thiergart et al.[1]. We took samples from the same 12 sites across Sweden, Spain, France, and replaced three sites from Northen Germany to three sites near Tübingen, Germany (table S5). We harvested *Arabidopsis thaliana* from February to May at the same developmental stage (bolting/flowering stage) for one (one site), two (one site) and three (13 sites) consecutive years. Using sterilized tweezers, the whole leaf samples were prepared by roughly cleaning the leaves from dirt, inserting them into 2mL tubes, and storing the tubes immediately in dry ice. The endophytic samples were instead prepared by subjecting leaves to surface sterilization. The rosette was surface-sterilized by washing with 80% ethanol for 15 seconds(s), followed by 2% bleach (sodium hypochlorite) for 30s. Rosettes were rinsed three times with sterile autoclaved water for 10s, before placing them in a screw-cap tube and freezing them on dry ice as endophytic samples. All samples were shipped to the Max Plank Institute for Plant Breeding Research in Cologne (Germany) and stored at -80^{o}C until further processing. We took 6 to 10 plants from each site, with half processed as whole leave and half as endophytes. In total 347 samples were collected.

DNA extraction from field samples and amplicon library preparation

DNA extraction and library preparation of all field samples is carried out with the same method described previously[2,3]. Briefly, 0.2g acid-washed and sterilized zirconium beads at different sizes (0.1-0.5mm diameter) and 500µL DNA extraction buffer were added to plant samples. The buffer consisted of 0.5% SDS, 50mM TRIS buffer at pH 8, 200mM NaCl, 2mM EDTA, 1mg/mL lysozyme, and 100mg/mL proteinase K. The samples/beads were first incubated for 45 min at 37^{o}C, and were then beat using a Bertin Precellys 24 (Bertin Technologies, Inc) at 6300 rpm for 2 x 45 seconds with a 15 second pause. A second incubation was then carried out at 37^{o}C for 45 minutes in the presence of 10 µg/mL RNAse. The tubes were then centrifuged at 10000 rpm for 2 minutes and the liquid was transferred to a new tube. The nucleic acids were cleaned up with 600µL phenol/chloroform/isoamyl alcohol (25:24:1), and chloroform/isoamyl alcohol (24:1), then precipitated by adding 0.1x of the sample volume of 3M sodium acetate and 1500µL 100% ethanol and centrifuging at 4^{o}C at 15000 rpm for 40 minutes. The product was washed twice with 70% ethanol and eluted with 30µL 10mM Tris buffer pH8.0. To avoid potential problems with PCR inhibition, samples were

mixed 1:1 with 20% Chelex-100 and allowed a 30-minute contact time, after which the liquid fraction was recovered and stored at -20°C until further use.

PCR amplification was performed in two steps to allow the use of blocking primers to decrease plant plastid contamination that occurs with 16S amplification. The samples were diluted with 10mM Tris HCl to 50ng/µl, before being used in the first PCR step. In the first step, universal primers and blocking oligos (table S6) were used together to amplify targeted regions. For each target taxon group, two universal primer pairs were used. Each of the 20µL reactions contained 0.2µL Q5 high-fidelity DNA polymerase (New England Biolabs (NEB)), 4µL Q5 High-GC Buffer, 4µL Q5 5x reaction buffer, 0.16µL of each of forward and reverse primer (10µM), 0.5µL of each blocking primer, 0.5µL dNTPs (10µM), 1µL template DNA and the rest filled with nuclease free water. Triplicates were run in parallel on three independent thermocyclers (Bio-Rad Laboratories, Hercules, CA, USA); cycling conditions were 95°C for 40s, 10 cycles of 95°C for 35s, 55°C for 45s, 72°C for 15s, and a final elongation at 72°C for 3min. The three reactions were combined and 10µL were used for enzymatic cleanup with Antarctic phosphatase and Exonuclease I (NEB; 0.5µL of each enzyme with 1.22µL Antarctic phosphatase buffer at 37°C for 30min followed by 80°C for 15min). The second step PCR was carried in a single 50µL reaction. The reaction included: 0.5µL Q5 DNA polymerase (NEB), 10µL Q5 High-GC buffer, 10µL Q5 5x buffer, 1µL dNTPs(10µM), 0.83µL of each barcoded primer (10µM), and 26.34µL nuclease free water. Half µL of the cleaned-up 1st step PCR product was added to each reaction. The protocol used in the thermocycler was: 95°C for 40 sec, followed by 25 cycles of 95°C for 35 sec, 55°C for 45 sec, 72°C for 15 sec, and 72°C for 2 min. The barcoded amplicons were then cleaned using AmPure XP beads (Backman-Coulter, Brea, CA, USA) following the manufacturer's protocol. Forty µl of magnetic beads solution was added to the 50µl barcoding reaction, shaken 5min at 700rpm at room temperature, then washed twice with 200µl of 80% ethanol, dried, and resuspended in 25µl of 10mM Tris. Amplicons were quantified in duplicates with the PicoGreen system (Thermo Fisher Scientific, Waltham, MA, USA) and samples were combined in equimolar amounts into one library to reach 8pmol per library. The final libraries were also cleaned with 0.8x volume AmPure XP purification beads and eluted into 40µL. Libraries were prepared with the MiSeq Reagent Kit v3 for 2x300 bp paired-end reads (Illumina) with 3% PhiX control. Quality assessment and the eventual presence of contaminants were assessed using a bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The final concentration of the pooled libraries was finally assessed using a Quantus

Fluorometer (Promega) following the indications of the manufacturer. Since there is a maximum of 50 barcoded primers per locus available, we performed 8 runs in total to cover all the samples. The sequencing libraries were run by the Illumina MiSeq platform in house.

Amplicon data processing and OTU table filtering

The raw data were imported into QIIME2 (version 2023-07)[4] as EMP paired-end sequences. Following, the reads were demultiplexed using the QIIME2 demux plugin. In this step, the sequences were separated for each flow cell by the marker gene (see Amplicon_data_analysis/OTU_table_processing/Metadata/Mapfile_Euro.txt). Reads were processed with the DADA2 pipeline[5] for quality filtering, removal of chimeric and non-biological sequences and generation of representative sequences under the default parameter within the QIIME2 documentation. Reads were further truncated at specific lengths for forward and reverse reads for each taxonomic group. These threshold values were manually selected based on the demultiplexing summaries, ensuring the removal of low-quality and non-informative sequences. After denoising, the sequences from all flow cells were combined. Additionally, the sequences from marker genes coding for the same taxonomic group were summarized and further processed as one set. For that, the sequences were clustered to operational taxonomic units (OTUs) using the *de-novo* clustering option using the vsearch plugin[6] with a percent identity of 97%. The OTUs were further assigned with taxonomic information using a sklearn-based taxonomy classifier[7]. Specifically, for the bacterial group, the pre-trained SILVA classifier (Silva 138, 99% OTUs full-length sequences) provided in QIIME2 package was used[8,9]. For fungi and oomycete reads, the UNITE database (Version 9.0) was used[10]. For other eukaryotic OTUs, the PR2 database (Version 5.0.0) was used[11]. For both UNITE and PR2 databases, taxonomic classifiers were trained with the naive-Bayes classifier trainer[7] under the default settings before being applied to the respective OTU groups. The scripts including all parameters until this point are available under Amplicon_data_analysis/Amplicon_sequencing_processing.

The raw OTU tables for each taxonomic group output from the above steps were exported from QIIME and were handled with Python. We first chose the amplicon with the highest number of classified OTUs from each taxon group for further filtering. We next applied several modification steps sequentially: 1. filtering of OTUs that do not belong to the corresponding taxonomic group (e.g. Non-bacteria OTUs from BacV5 OTUs); 2. merging all OTU tables from each taxonomic group; 3. filtering of OTUs with lower than 50 counts and present in only one

sample; 4. separating OTUs tables by compartment – whole leaf and endophytes; 5. selection of the amplicon with higher number of classified taxa to proceed further for each taxonomic group; 6. normalizing the count of each OTU to the total count of their corresponding taxonomic group of each sample, and calculating the relative abundance; 7. selection of the samples containing OTUs from all four taxonomic groups. The detailed python script for this step is available under Amplicon_data_analysis/OTU_table_processing/filter_otu_and_tax_final.ipynb. The final OTU tables were deposited in Amplicon_data_analysis/OTU_table_processing/processed_OTU_tables.

Diversity analysis

Based on the final OTU table, we calculated the Shannon index using the R-package vegan[12] and grouped the OTUs by country of origin for comparison. To control the potential bias introduced by varying sample sizes across different countries, we sub-sampled the reads from each sample 100 times and conducted Kruskal-Wallis tests for each iteration. We only observed 10 from the 100 times of the sub-sampled reads showing significant differences in the Shannon index across the four countries, supporting the results without sub-sampling is unbiased.

For analysing beta diversity, distance matrices using Bray-Curtis distance were calculated for all samples and the subsets of whole leaf and endophytic samples using the R-packages vegan and ade4[13]. Plotting of the results and statistical testing using PERMANOVA were conducted with the first and second principal components using the ggplot2. In the statistical PERMANOVA analysis, all groups were compared at once for both endophyte and whole leaf plots using the function from the R-package vegan.

To identify the most occurring taxa from each country, we calculated the relative abundance of the 10 most abundant classes for bacteria, fungi and other eukaryotes. For oomycetes, we calculated the 10 most abundant genera since all samples only contain one (endophytes) or two (whole leaf) oomycete classes. We used phyloseq package[14] for this analysis. Samples were first merged by country and subsequently normalized to relative abundance, and OTUs with an undetermined class were grouped in the category ’unassigned’. All classes with lower abundances than those categorized in the top 10 were grouped into the category ’other.’ Subsequently, the average relative abundances were plotted in using the ggplot2[15].

All R scripts for analysis and figure plotting were summarized in Amplicon_data_analysis/diversity_analysis.

Network analysis

Before network construction, the OTU-table was filtered to only contain OTUs present in 10 or more samples to retain the most conserved taxa among all samples. The subsequent network construction was performed using Fastspar[16], using the recommended settings for bootstrapping and permutation counts. The resulting p-value and median correlation tables were transformed into an edge table using R, keeping all interactions with a p-value lower than 0.1 and a none-zero correlation value. We selected this relatively loose thresholds since we have applied strict filtering for the most conserved OTUs so we would like to retain most of their interactions including the weak interactions. We used the built-in function from the network analyser Cytoscape (Version 3.9.1)[17] to calculate degree, betweenness centrality, and closeness centrality, and the final visualizations of the networks were constructed utilizing Gephi (Version 0.10.1)[18]. The script for constructing the network is available at Amplicon_data_analysis/Network_analysis_and_Albugo-Dioszegia_interaction.

Microbial growth conditions

All microbial strains used in this study are summarized in table S2. All fungi used in this study were grown on potato dextrose agar (Carl Roth GmbH + Co. KG, Karlsruhe, Germany), and bacteria in nutrient broth agar (Carl Roth). The spore collection and inoculation of the *Albugo laibachii* was performed according to Kemen et al.[19]. Briefly, infected leaves with clear white pustules were collected and resuspended in Milli-Q water within 50mL tubes. The tubes were incubated on ice for 1 hour, and the spore suspension was filtered with 40µm filter (Greiner bio-one, Kremsmünster, Austria). The spore suspension was then sprayed on plants using a spray gun with an air pump, and plants were incubated at 8^{o}C in the dark overnight. Infected plants were kept under 8-h light and 16-h dark cycles with a 20°C day and 16°C night temperature, 60% humidity, for 14 days or until the infection symptoms appeared.

The gnotobiotic microbial inoculation on *A. thaliana*

The gnotobiotic microbial inoculation system is adapted from Eitzen et al.[20] and Ruhe et al.[21] with some modifications. Briefly, *A. thaliana* ws-0 (Wassilewskija) seeds were sterilized for 6-12 hours with chlorine gas. Sterilized seeds were immediately sown on 0.5 x Murashige and Skoog (MS) with 0.75% agar and incubated under 8-16-h light-dark cycles with a 20°C day and 16°C night temperature with 60% humidity. After 7 days, the germinated seedlings were

placed onto 12-well plates (Greiner bio-one) with each well filled with 2.5mL 0.5 x MS-agar. Plates with seedlings were further incubated for 3 weeks before spraying.

We primarily tested fungi (yeasts) and few bacteria in this study, either because isolation of the predicted *Albugo*-interacting bacteria was unsuccessful or these are slow growers (actinobacteria) and therefore not compatible with co-inoculation with *Albugo*. To spray different bacteria and fungal strains, we directly streaked fresh bacteria or yeast colonies (<5-day growth) from agar plates and resuspended them in 10mM $MgCl_2$ solution. We then diluted the strain resuspension to reach optical density of 600 nm (OD_{600}) at 0.5. To spray *Fusarium* and *Stagonosporopsis* strains, 10mL 10mM $MgCl_2$ was added into each plate that was fully covered with fungal mycelium and spores (>14-day growth). We then used 1mL pipette tip to streak the plate and release a maximum number of spores, which were filtered with 40µm filter (Greiner bio-one). We then count the spores with a haemocytometer and diluted the spores to reach 10^4-10^5 spores/mL. After dilution, all microbial solutions were transferred into the airbrush guns (Conrad electronics, Hirschau, Germany) and sprayed two times (~40µL each) with the air pump (Conrad electronics) at 0.5 bar pressure. After spraying buffer and different microbial strains, we waited 2-3 hours to allow the evaporation of excessive water before spraying *A. laibachii*. To minimize the effect of *A. laibachii* spore-associated microbes and ensure a gnotobiotic infection environment, we treat the *A. laibachii* spores with antibiotics. We first incubated the *Albugo*-infected leaves with autoclaved Milli-Q water on ice for 1h to facilitate the release and hatching of *A. laibachii* sporangia (one full 50mL tube of infected leaves filling up with water). The spore solution was then filtered with 40µm filter (Greiner bio-one) and centrifuged in 15mL tubes using 4000g for 10 mins at room temperature. We next remove the supernatant and resuspend the spore pellet gently with a 5mL antibiotic solution containing 2.5mg kanamycin, streptomycin, and rifampicin, plus 1.25mg geneticin. We incubate the spores with the antibiotic solution at room temperature for 25mins in the dark, before pelleting again by centrifuging the tube at 4000g for 10mins. The supernatant was removed and the antibiotic-treated spores were washed twice with autoclaved Milli-Q water and once with 10mM $MgCl_2$ using the same centrifugation conditions. The antibiotic-treated spores were resuspended with 10mM $MgCl_2$ and adjusted to 10^6-10^7 spores/mL before spraying on the sterile plants using the same airbrush setup as for other microbes. After spraying the *A. laibachii*, all 12-well plates were sealed with a breathable tap and bagged with transparent bags to keep the humidity at >90%. The bagged

plates were incubated at 8°C in the dark overnight before transferring to the growth chambers. The transparent bag was removed after another day (day2). The infected plants were scored and harvested for qPCR at 12 days after inoculation (dai).

Quantification of *Albugo in planta* with scoring and quantitative polymerase chain reaction (qPCR)

To score infected plants from gnotobiotic infection systems, we first removed all root tissue of each plant and recorded the number of infected leaves for each plant. We then divided the total number of infected leaves by the total number of plants from each treatment group to calculate the average number of infected leaves per plant (ANIL). Since the efficacy of *A. laibachii* infection varies significantly among individual plants, we tried to ensure the robustness of the experiment by only including results from experiments with 0.5 – 2 ANIL in the control group (Buffer and *A. laibachii*). The *A. laibachii* infection were considered unsuccessful if the ANIL was below 0.5. Additionally, the difference between buffer and treatment groups cannot be observed if the infection is too good (ANIL>2.5), since there were only 4-5 mature leaves in total when *A. thaliana* plants were sprayed. To control the impact of zeros in statistical analysis and balance results from experiments with varied infection levels, we also added a pseudo-count (0.05) to all scoring data before calculating the ANIL.

After scoring, the rosette was put into 50mL falcon tubes and divided into 3-6 groups as biological replicates with 3-5 plants per group for DNA extraction and qPCR. The rosette tissues were snap frozen in liquid nitrogen and grinded with mortars and pestles. We used the silica column based DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNA concentration was measured with NanoDrop (Thermo Fisher) and adjusted to 5ng/µL with Milli-Q water before processing for the qPCR experiment. We used SsoAdvanced Universal SYBR Green Supermix and CFX Connect real-time PCR detection system (Bio-Rad laboratories, Hercules, CA, USA) for qPCR reaction following the manufacturer's instruction. We used 20µL reactions and 5µL (25ng) DNA as template. For each sample, we performed two reactions for the *EF1*α genes of both *A. laibachii* and *A. thaliana*. Cq values obtained from *A. laibachii* DNA amplification were first normalized to *A. thaliana* DNA amplicon from the same sample using the formula $2^{-\Delta Cq}$ to calculate the relative *A. laibachii* quantity per sample. Then the relative *Albugo* quantities of each treatment sample were normalized to the average relative *Albugo* quantity of buffer control to compare among different experiments. Each data point in the figure represents three technical replicates.

***Dioszegia* whole genome sequencing, assembly and annotation**

The DNA of different *Dioszegia* strains was extracted using DNeasy Plant Mini Kit (Qiagen) for Illumina sequencing or the Puregene Yeast/Bact. Kit (Qiagen) for PacBio and Nanopore sequencing, following manufacturers' protocols. The DNA was sent on dry ice to BMKGene (Biomarker Technologies, Münster, Germany) for sequencing. For all Illumina data, raw fastq files were quality filtered using fastp tool[22] and only reads with Phred score > 30 were retained. The quality filtered fastq was error corrected using Musket[23] and removed PCR duplicates with fastuniq[24]. The quality controlled reads were assembled using SPAdes[25]. For PacBio data, we used Flye (v2.9.2)[26] with the --pacbio-raw flag to assemble the PacBio data. For nanopore data, we received fastq file basecalled with Guppy (v6, Oxford Nanopore Technologies (ONT), Oxford, UK). We then filtered out reads less than quality score 10 and less than 2000 base pairs with ProwlerTrimmer (ONT). Quality filtered nanopore reads were assembled using canu (v2.2)[27]. The reads used for assembly were then mapped back to the assembled contigs using minimap2[28] and polished with racon[29] for nine times. The racon polished assembly was again polished by medaka (ONT) once using the super high accuracy mode. All assemblies were assessed using BUSCO tool[30] to determine their completeness using the Tremellomycetes_odb10 database. We used the funannotate pipeline[31] for gene prediction using the *Dioszegia hungarica* PDD-24b-2 genome[32] as reference. The predicted genes were functionally annotated using both InterProScan[33] and eggnog-mapper (v2.1.8, *18*) under the Galaxy EU platform[35].

All sequenced strains and the quality assessment of assemblies are summarised in table S3. All raw data was submitted to NCBI short read archive (SRA) under the Bioproject PRJNA1190083. All genome assemblies were submitted to the GeneBank and the accession of each genome is recorded in table S3. A copy of the assemblies, gene predictions and functional annotation used in this project were posted at /Supplementary_files/Dioszegia_genomes_and_annotations/. Detailed scripts used for all above steps are summarized in /Scripts_for_data_analysis_and_figure_plotting/Genome_analysis.

Extraction of thiamine metabolism genes and cloning of *DhPER1* orthologues in different *Dioszegia* strains

We manually inspected the functional annotation result of *Dioszegia* EY and PT genomes and extracted the thiamine metabolism genes by searching the keyword 'thiamine' in the

functional annotation file. In total, 14 genes were identified from each genome. We paired these genes through alignment and validated them by Sanger sequencing in case the paired genes were predicted to be significantly different in gene size (>10%) or exon borders. After Sanger sequencing validation, all alleles ended up with a similar size and were translated to amino acid (AA) sequence for alignment. The CDS and AA sequence were then aligned to calculate the SNPs per kilobasepairs and variations per 100 amino acids.

To search for genes homologous to *DhPER1* from different *Dioszegia* strains, we used two strategies. Firstly, we searched different parts of *DhPER1* in all available *Dioszegia* genomes assemblies from both online databases and in house sequencing using Geneious Prime v2024.0.5 (https://www.geneious.com). We found six *DhPER1* homologous genes using this method. However, we did not find the *DhPER1* paralogue from the nanopore assembly of *Dioszegia* Y22 strain. Therefore, we performed a PCR test using the *DhPER1* primers and the PCR test resulted in the identification of the Y22 paralogue of *DhPER1*. This result also indicated that the nanopore assembly was largely inaccurate for this strain, and therefore we performed the Illumina sequencing and assembly again for Y22 (table S3).

RNAseq experiments and data analysis

We performed two RNAseq experiments. The first is the analysis of *Dioszegia* EY and PT transcriptome to validate the expression profile of *DhPER1 in vitro*. The second experiment is the RNAseq analysis of *A. laibachii* and *D. hungarica* inoculated on sterile *A. thaliana* leaves at 2 dai. For both experiments, the RNeasy Plant Mini Kit (Qiagen) was used for RNA extraction, and the Turbo DNA-free kit (Thermo Fischer) was used for the removal of the remaining DNA. The purified RNA was then quality controlled by agarose gel electrophoresis and bioanalyzer (Agilent Technologies) using the corresponding RNA kit before sending to BMKGene for sequencing.

We performed PacBio long-read RNA sequencing (IsoSeq) for the *D. hungarica* comparative transcriptomics experiment, with one sample for each strain. The circular consensus (CCS) reads from *Dioszegia* EY and PT were received and mapped to their corresponding *DhPER1* paralogues using Geneious RNA mapper and visualized within Geneious Prime (v2024.0.5). The raw CCS reads from EY and PT were deposited at NCBI SRA under SRR31480791 and SRR31480792.

For the RNAseq data from *Albugo* and *Dioszegia* inoculated sterile *A. thaliana* at 2 dai, we used Illumina sequencing and performed three biological replicates for each of the three

treatment groups (*Albugo*, *Albugo*+EY, *Albugo*+PT). We used a standard RNAseq analysis workflow[36] to analyse the data from each treatment group separately. Firstly, the raw Illumina fastq data were quality controlled with FastQC[37], before trimming with fastp (v0.24.0)[22]. We then concatenated the genome assembly, transcripts, and annotation files of *A. thaliana* (Tair10)[38] and *A. laibachii*[19], before mapping the quality-controlled fastq reads to it with HISAT2 (2.2.0) using the default parameters[39]. We assembled the mapped reads into transcripts and calculated the Fragments Per Kilobase of transcript per Million read pairs (FPKM) and Transcripts per million (TPM) for *A*. *thaliana* and *Albugo* genes using with StringTie (2.2.0) using the default parameters[40]. Principal component analysis (PCA) was performed on FPKM of each sample. Similarity among samples was displayed by reducing dimensionality into two or three principal components. Then, we used the ggplot2 R package[15] for visualizing the PCA of the two most related components. We next used DESeq2[41] to investigate the differentially expressed *Albugo* genes induced by *D. hungarica*. The list of differentially expressed *Albugo* genes were extracted and submitted to ShinyGO (v.0.82)[42] for the GO analysis with an FDR cut-off of 0.05 and 4 genes per group selection criteria. The Illumina RNAseq data was deposited at NCBI SRA under the following accessions: SRR31531712, SRR31531713, SRR31531714, SRR31531715, SRR31531716, SRR31531717, SRR31512421, SRR31512422 and SRR31512423. The intermediate HISAT2 and StringTie output files and main scripts used in this step are available at Supplementary_files/RNAseq_analysis/.

Structure modelling of *DhPER1* and molecular docking

Structure predictions were conducted with AlphaFold2 (v.2.3.0)[43] using the full databases for multiple sequence alignment (MSA) construction. We produced five models and selected the best model for visualisation purposes. The structure of the yeast thiamine permease *THI7* was downloaded from UniProt[44]. Structures of all permeases (.pdb files) and molecular docking files (.sdf) were posted at /Supplementary_files/Permease structure prediction/. We used PyMOL[45] for visualization of structures.

Gene cloning and plasmid construction

We used three plasmids in this study, and they were all amplified in *E. Coli* DH5α strain. The CRISPR plasmid (pHSP:CRISPR) contains a high-fidelity variant codon optimised Cas9 gene for *Ustilago maydis*[46]. The sgRNA was designed using Geneious Prime (v2024.0.5) and we chose the target close to the 5’ end of *DhPER1*. The plasmid pHSP:CRISPR was linearized with

restriction enzyme Acc65I and assembled with sgRNA oligo and "scaffold RNA" fragment with 3' downstream 20 bp overlap to the plasmid by Gibson Assembly (NEB) following the manufacturer's protocol.

The hygromycin resistance gene (HygR) and *DhPER1* expressing vectors were generated using a shuttle vector EC35 from the fungal Golden Gate system developed for *Ustilago maydis* as backbone[47]. To construct the HygR expressing vector pHSP:HygR, EC35 was first digested with BamHI and XbaI (NEB), and the HSP::HygR cassette was amplified from the pHSP-CRISPR vector using primer YH67 and YH55. The two fragments were assembled together using Gibson Assembly (NEB) following manufacturer's protocol. To generate the pHSP:PER plasmid, we first amplified the HSP70 promoter from the pHSP:CRISPR using YH56 and YH57, and the *DhPER1* transcript (including introns) and terminator from *Dioszegia* EY genomic DNA using YH58 and YH64. We then amplified the HSP::HygR cassette from the pHSP:CRISPR vector using YH54 and YH55. The three fragments were assembled into the same EC35 backbone (BamHI and XbaI digested) using Gibson Assembly (NEB) following manufacturer's protocol. We also cloned the *DhPER1* with its endogenous promoter from the *Dioszegia* EY genomic DNA into the same plasmid, but we could not identify any PT transformants showing a significant increase in *DhPER1* expression, indicating that in PT this promoter may not be recognized (fig. S4E). All primers used in this process are summarised in table S6. The plasmid maps were posted at /Supplementary_files/Plasmid map/.

Genetic modification of *D. hungarica* EY and PT strains

We adapted an electroporation method from *Cryptococcus* to transform *D. hungarica*[48]. Briefly, 100mL overnight culture of *D. hungarica* was centrifuged at 3500rcf for 10mins (All centrifugation in this experiment is at 4^{o}C) and washed 3 times with 10mL ice cold water using 3500rcf for 7mins. The pellet was then washed with 10mL electroporation buffer (EB buffer, 10mM Tris, 1mM $MgCl_2$, 270mM Sucrose, pH7.5) at 3500rcf for 7mins, and resuspended in 10mL EB buffer with 1mM dithiothreitol (DTT) on ice for 30mins. The solution was then centrifuged at 3500rcf for 7mins and resuspended in 1mL EB buffer to wash out DTT. The pellet was resuspended in 600µL of EB buffer and stored on ice, ready for electroporation.

For electroporation, 100µL of freshly prepared competent cells were gently mixed with 2µg plasmid DNA and loaded in a pre-cooled electroporation cuvette (MicroPulser, Bio-Rad) with a 0.2cm gap. We then used the GenePulser system (Bio-Rad) to perform the electroporation at 0.45kV, 125µE and 500Ω. The time constant value after electroporation should be 30–37ms.

After electroporation, the solution was streaked on fresh PDA plate with 100µg/L hygromycin for selection. The colonies were screened with colony PCR and further validated by qPCR on *DhPER1* expression (normalized to *DhEF1*α).

Visualization of physical interaction between *A. laibachii* zoospores with *D. hungarica*

We first hatched *A. laibachii* zoospores using cold treatment adapting a method from Phytophthora with minor modifications[49]. Briefly, 14-day old infected plants were placed at 8^{o}C with high humidity (covered in plastic bag) overnight to induce germination of sporangia. The leaves with clear *A. laibachii* pustules were then collected and soaked in Milli-Q water for 1 hour on ice to release sporangia and zoospores. The zoospore solution was filtered with a 40µm filter (Greiner bio-one) to remove contaminating mycelia. The filtered zoospore solution was kept at 8^{o}C and checked under light microscope every 15 mins to reach an optimal timepoint with the highest number of free zoospores. We normally found the optimal point around one hour after spore filtering.

During the waiting time, a small colony from a fresh-grown plate of *D. hungarica* was picked with a sterile pipette tip, and streaked onto a glass microscope slide. Tape was used to increase the gap between the cover slip and the slide holding *D. hungarica* to allow larger space for zoospore to swim. Around 50µL of the zoospore solution was pipetted on top of *D. hungarica* and observed under a light microscope. Microscopic pictures and recordings were obtained with a Zeiss AxioPhot (512 colour Axio camera, Carl Zeiss AG, Oberkochen, Germany). All pictures and videos were taken under either 40x or 64x magnification lens.

Evolutionary divergence of *DhPER1*

Firstly, an ortholog grouping across the predicted proteome of all *Dioszegia* strains (table S3) was performed using Proteinortho[50]. Only orthogroups of 4 or more members were used in downstream analyses. For each of the orthogroups, codon alignments were constructed by mapping the coding sequence onto its corresponding protein alignment using MAFFT with pal2nal[51,52]. On these, we applied the software PAML to estimate selective pressure[53], by using a combination of four codon substitution models, including null models 1 and 7, and models allowing for positive selection 2 and 8.

Phylogeny and selective pressures of permease across the fungal kingdom

The dataset of ortholog PER genes across the fungal kingdom was created by searching for homolog genes based on similarity with the ones from *Dioszegia* on the NCBI protein and EGGNOG ortholog databases[54]. After alignment with MAFFT and trimming with trimmAl[55], a

maximum likelihood tree was constructed using IQtree[56] with default codon model search and 1000 replicates for ultrafast bootstrap.

We applied the RELAX algorithm[57] from the hyphy software suite to test for relaxation pressures in various genera from the PER1 tree. As the testing branches, we used all those immediately leading to each gene, and as background all the branches in the complete tree. Scripts and output files for this part were summarized in Scripts_for_data_analysis_and_figure_plotting/Relax_test_tree/.

Assessment of translational optimization among *Dioszegia* genomes

The whole-genome phylogeny was calculated based on 2,461 single-copy orthologous protein sequences determined with orthofinder (v2.5.4)[58]. The protein sequences were aligned with MAFFT and the Maximum Likelihood phylogeny was calculated based on 500 bootstrap replicates with RAxML (v8.2.12)[59]. The tRNA genes were predicted with tRNAscan-SE (v2.0.9)[60] with default parameters for eukaryotic organisms. The prediction with a covariance score <30 or classified as pseudogenes were excluded from further analyses. The synonymous relative codon usage was calculated with BioKIT (v0.1.1)[61]. The tRNA adaptation index (tAI) and codon optimization (S) were calculated with the *tAI* (v0.2.1) R package[62]. All scripts and source files used in this part were deposited in Scripts_for_data_analysis_and_figure_plotting/Translational_optimization/.

Long term co-cultivation of *A. laibachii* and *D. hungarica* and *Dioszegia* quantifications

Both sterile and non-sterile long-term co-cultivation methods are largely similar to the standard 2-week infection experiments described above, with minor optimizations. For the long-term sterile infection experiment, we used 6-well plates instead of 12-well plates since the volume of media and the size of each well is not sufficient to support the growth of *A. thaliana* for the whole experimental period. Under non-sterile conditions, we sprayed 4–6-week-old *A. thaliana* ws-0 seedlings on 6-well seed germination trays (5x5cms) with different strains of *Dioszegia* (OD_{600}=0.5 in 10mM $MgCl_2$). Each *Dioszegia* strain was sprayed on a total of four 6-well trays, corresponding to four sampling timepoints. After 1 hour of *Dioszegia* spray or when the liquid evaporated and the *Dioszegia* solutions were settled, the 6-well trays were then sprayed with *A. laibachii* spore solution without antibiotic treatment. We adjusted the spore concentration to 10^4-10^5 spores/mL and sprayed two times (~40µl) per plant. After spraying *A. laibachii*, each 6-well tray was individually bagged with transparent cellophane bags and kept in the dark at 8°C overnight. The plants were moved to the growth chamber

and kept bagged for another 24 hours. Infected plants were kept under 8-h light and 16-h dark cycles with a 20°C Day and 16°C night temperature and 60% humidity.

For sampling of gnotobiotic experiments, the roots were removed and the whole rosette were collected and grinded in liquid nitrogen as one biological replicate. For each treatment group, we took six samples at each time point, so in total we collected 168 samples for each gnotobiotic experiment. We used 100mg tissue for DNA extraction and qPCR quantification. For plants from the holoxenic experiment, which grew on soil, three leaves were collected from one plant, and two plants combined as one biological replicate. We took three samples for each group at each timepoint, so we collected 96 samples for each holoxenic experiment. Samples were frozen in liquid nitrogen and grinded for DNA extraction. DNA was extracted using DNeasy Plant Mini Kit (Qiagen) and qPCR performed with *D. hungarica* EF1α-specific primer (table S6) and normalized to plant biomass using the same experimental setup as *A. laibachii* quantification described above. We performed this experiment twice and similar results were obtained.

Reference:

1. Thiergart, T., Durán, P., Ellis, T., Vannier, N., Garrido-Oter, R., Kemen, E., Roux, F., Alonso-Blanco, C., Ågren, J., Schulze-Lefert, P., et al. (2020). Root microbiota assembly and adaptive differentiation among European Arabidopsis populations. Nat Ecol Evol *4*, 122–131. https://doi.org/10.1038/s41559-019-1063-3.

2. Agler, M.T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D., and Kemen, E.M. (2016). Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. PLOS Biology *14*, e1002352. https://doi.org/10.1371/journal.pbio.1002352.

3. Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J., et al. (2018). Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host & Microbe *24*, 168-179.e4. https://doi.org/10.1016/j.chom.2018.06.011.

4. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

5. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods *13*, 581–583. https://doi.org/10.1038/nmeth.3869.

6. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584. https://doi.org/10.7717/peerj.2584.

7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

8. Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 90. https://doi.org/10.1186/s40168-018-0470-z.

9. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glöckner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research *42*, D643–D648. https://doi.org/10.1093/nar/gkt1209.

10. Abarenkov, K., Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., May, T.W., Frøslev, T.G., Pawlowska, J., Lindahl, B., Põldmaa, K., Truong, C., et al. (2024). The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. Nucleic Acids Research *52*, D791–D797. https://doi.org/10.1093/nar/gkad1039.

11. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research *41*, D597–D604. https://doi.org/10.1093/nar/gks1160.

12. vegan: Community Ecology Package (2012).

13. Dray, S., and Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software *22*, 1–20. https://doi.org/10.18637/jss.v022.i04.

14. McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

15. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer) https://doi.org/10.1007/978-0-387-98141-3.

16. Watts, S.C., Ritchie, S.C., Inouye, M., and Holt, K.E. (2018). FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics *35*, 1064. https://doi.org/10.1093/bioinformatics/bty734.

17. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

18. Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the International AAAI Conference on Web and Social Media *3*, 361–362. https://doi.org/10.1609/icwsm.v3i1.13937.

19. Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A.C., Balmuth, A.L., Robert-Seilaniantz, A., Bailey, K., Holub, E., Studholme, D.J., MacLean, D., et al. (2011). Gene Gain and Loss during

Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. PLOS Biology *9*, e1001094. https://doi.org/10.1371/journal.pbio.1001094.

20. Eitzen, K., Sengupta, P., Kroll, S., Kemen, E., and Doehlemann, G. (2021). A fungal member of the Arabidopsis thaliana phyllosphere antagonizes Albugo laibachii via a GH25 lysozyme. eLife *10*, e65306. https://doi.org/10.7554/eLife.65306.

21. Ruhe, J., Agler, M.T., Placzek, A., Kramer, K., Finkemeier, I., and Kemen, E.M. (2016). Obligate Biotroph Pathogens of the Genus Albugo Are Better Adapted to Active Host Defense Compared to Niche Competitors. Front. Plant Sci. *7*. https://doi.org/10.3389/fpls.2016.00820.

22. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

23. Liu, Y., Schröder, J., and Schmidt, B. (2013). Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics *29*, 308–315. https://doi.org/10.1093/bioinformatics/bts690.

24. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., and Chen, S. (2012). FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. PLOS ONE *7*, e52249. https://doi.org/10.1371/journal.pone.0052249.

25. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology *19*, 455. https://doi.org/10.1089/cmb.2012.0021.

26. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8.

27. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research *27*, 722. https://doi.org/10.1101/gr.215087.116.

28. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

29. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res *27*, 737–746. https://doi.org/10.1101/gr.214270.116.

30. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

31. Palmer, J.M., and Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation. Version v1.8.1 (Zenodo). https://doi.org/10.5281/zenodo.4054262 https://doi.org/10.5281/zenodo.4054262.

32. Jarrige, D., Haridas, S., Bleykasten-Grosshans, C., Joly, M., Nadalig, T., Sancelme, M., Vuilleumier, S., Grigoriev, I.V., Amato, P., and Bringel, F. (2022). High-quality genome of the basidiomycete

yeast Dioszegia hungarica PDD-24b-2 isolated from cloud water. G3 Genes|Genomes|Genetics *12*, jkac282. https://doi.org/10.1093/g3journal/jkac282.

33. Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. Nucleic Acids Res *49*, D344–D354. https://doi.org/10.1093/nar/gkaa977.

34. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

35. Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research *46*, W537. https://doi.org/10.1093/nar/gky379.

36. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc *11*, 1650–1667. https://doi.org/10.1038/nprot.2016.095.

37. Andrews, S. (2024). s-andrews/FastQC.

38. Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research *40*, D1202–D1210. https://doi.org/10.1093/nar/gkr1090.

39. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

40. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290–295. https://doi.org/10.1038/nbt.3122.

41. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

42. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629. https://doi.org/10.1093/bioinformatics/btz931.

43. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

44. The UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

45. Schrödinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

46. Zuo, W., Depotter, J.R., and Doehlemann, G. (2020). Cas9HF1 enhanced specificity in Ustilago maydis. Fungal Biol *124*, 228–234. https://doi.org/10.1016/j.funbio.2020.02.006.

47. Terfrüchte, M., Joehnk, B., Fajardo-Somera, R., Braus, G.H., Riquelme, M., Schipper, K., and Feldbrügge, M. (2014). Establishing a versatile Golden Gate cloning system for genetic engineering in fungi. Fungal Genetics and Biology *62*, 1–10. https://doi.org/10.1016/j.fgb.2013.10.012.

48. Lin, J., Fan, Y., and Lin, X. (2020). Transformation of Cryptococcus neoformans by Electroporation Using a Transient CRISPR-Cas9 Expression (TRACE) System. Fungal genetics and biology : FG & B *138*, 103364. https://doi.org/10.1016/j.fgb.2020.103364.

49. Fry, W.E., Patev, S.P., Myers, K.L., Bao, K., and Fei, Z. (2019). Phytophthora infestans Sporangia Produced in Culture and on Tomato Leaflet Lesions Show Marked Differences in Indirect Germination Rates, Aggressiveness, and Global Transcription Profiles. MPMI *32*, 515–526. https://doi.org/10.1094/MPMI-09-18-0255-TA.

50. Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., and Prohaska, S.J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics *12*, 124. https://doi.org/10.1186/1471-2105-12-124.

51. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics *9*, 286–298. https://doi.org/10.1093/bib/bbn013.

52. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research *34*, W609. https://doi.org/10.1093/nar/gkl315.

53. Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

54. Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Research *36*, D250–D254. https://doi.org/10.1093/nar/gkm796.

55. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

56. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

57. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Molecular Biology and Evolution *32*, 820–832. https://doi.org/10.1093/molbev/msu400.

58. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

59. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

60. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Research *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.

61. Steenwyk, J.L., Buida, T.J., III, Gonçalves, C., Goltz, D.C., Morales, G., Mead, M.E., LaBella, A.L., Chavez, C.M., Schmitz, J.E., Hadjifrangiskou, M., et al. (2022). BioKIT: a versatile toolkit for processing and analyzing diverse types of sequence data. Genetics *221*, iyac079. https://doi.org/10.1093/genetics/iyac079.

62. Reis, M. dos, Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res *32*, 5036–5044. https://doi.org/10.1093/nar/gkh834.