New Phytologist

Proteins released into the plant apoplast by the obligate parasitic protist *Albugo* selectively repress phyllosphere-associated bacteria

Daniel Gómez-Pérez[1] iD, Monja Schmid[1] iD, Vasvi Chaudhry[1] iD, Yiheng Hu[1] iD, Ana Velic[2], Boris Maček[2] iD, Jonas Ruhe[3], Ariane Kemen[1] iD and Eric Kemen[1] iD

[1]Microbial Interactions in Plant Ecosystems, Center for Plant Molecular Biology, University of Tübingen, 72076 Tübingen, Germany; [2]Department of Biology, Quantitative Proteomics Group, Interfaculty Institute of Cell Biology, University of Tübingen, 72076 Tübingen, Germany; [3]Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

Summary

- Biotic and abiotic interactions shape natural microbial communities. The mechanisms behind microbe–microbe interactions, particularly those protein based, are not well understood. We hypothesize that released proteins with antimicrobial activity are a powerful and highly specific toolset to shape and defend plant niches.
- We have studied *Albugo candida*, an obligate plant parasite from the protist Oomycota phylum, for its potential to modulate the growth of bacteria through release of antimicrobial proteins into the apoplast.
- Amplicon sequencing and network analysis of *Albugo*-infected and uninfected wild *Arabidopsis thaliana* samples revealed an abundance of negative correlations between *Albugo* and other phyllosphere microbes. Analysis of the apoplastic proteome of *Albugo*-colonized leaves combined with machine learning predictors enabled the selection of antimicrobial candidates for heterologous expression and study of their inhibitory function. We found for three candidate proteins selective antimicrobial activity against Gram-positive bacteria isolated from *A. thaliana* and demonstrate that these inhibited bacteria are precisely important for the stability of the community structure. We could ascribe the antibacterial activity of the candidates to intrinsically disordered regions and positively correlate it with their net charge.
- This is the first report of protist proteins with antimicrobial activity under apoplastic conditions that therefore are potential biocontrol tools for targeted manipulation of the microbiome.

Introduction

The plant leaf is a highly competitive habitat for microbes due not only to limited resources, but also to its instability as a result of rapidly changing conditions, for example, microbes triggering defense reactions or exploiting the habitat up to its destruction (Hassani *et al.*, 2018). As a consequence, mechanisms that enable microbes to fight off opponents, by, for example, outcompeting competitors for limiting resources or releasing antimicrobial compounds, are under strong selective pressure (Freilich *et al.*, 2011). Identification and characterization of such mechanisms could lead to breakthroughs in therapeutics and disease control (Bollenbach, 2015). In particular, studies on stable interactions in natural microbial communities have historically been considered an important resource in the discovery of new antimicrobial compounds (Molloy & Hertweck, 2017). Some of the best host-adapted microbes include obligate biotrophs, pathogens that can only survive on a living host (Ruhe *et al.*, 2016). They rely completely on intact plant niches where

host and microbes are in stable equilibrium. In microbial community network analyses of the phyllosphere, hub microbes emerge as highly interconnected microbes that play a central role in the management of the microbial composition (Agler *et al.*, 2016). The oomycete and obligate biotroph pathogen *Albugo* was shown to be such a microbe by reducing the growth of some microbes while increasing the growth of others, thereby significantly impacting the leaf microbial community (Agler *et al.*, 2016). However, the mechanisms that underlie inhibition or promotion of co-occurring microbes remain largely unexplored. Therefore, *Albugo* infection and its effect on the microbiome represents an ideal model system to identify and study antimicrobial strategies by obligate pathogens that need to defend their niche to keep the host alive.

Albugo is the causal agent of white blister rust on Brassicaceae plants. Taxonomically, it belongs to the Oomycota, a heterogeneous group of organisms comprising many highly adapted parasites of plants, animals, and humans. Oomycetes can be considered protists as they are eukaryotes that do not belong to

the animal, plant, or fungal kingdoms (Burki *et al.*, 2019). Following penetration into the plant host via the leaf stomata, *Albugo* develops intercellular hyphae to colonize the plant extracellular space, known as apoplast (Berlin & Bowen, 1964). Herein, it is in contact with other endophytic microbes and competes for nutrient and habitat dominance. As an obligate biotroph, *Albugo* relies on the living plant for nutrients and structural support and hence, for overall survival. As a consequence, *Albugo* is incapable of growing independently of its host and reduction in its genome has led to the loss of all of its secondary and most of its primary metabolic pathways (Kemen *et al.*, 2011). To shape its niche, *Albugo* releases proteins into the plant cytoplasm and the apoplast. Some of these so-called effector proteins modulate host immune responses, but for many of them the function remains unknown (Furzer *et al.*, 2022). As described only recently for a hemibiotrophic fungus, apoplastic-secreted proteins can also act as microbiome control agents since they selectively modify the endophytic bacterial community and can therefore be considered effectors governing microbe–microbe interactions (Snelders *et al.*, 2018, 2020, 2021).

Enrichment in long intrinsically disordered regions (IDRs) is a common feature of the secretome of plant pathogens (Marín *et al.*, 2013). IDRs are protein domains that, in general terms, lack a stable folding conformation (Oldfield *et al.*, 2019). This is due to their sequence, which is biased toward certain disorder-promoting amino acids and often shows hydrophilic tendencies (Dubreuil *et al.*, 2019). The function of IDRs in proteins can vary considerably depending on the environmental conditions, owing to their inherent structural pliability. In plant pathogenic eukaryotes, IDRs have been proposed to be relevant for extracellular effector protein delivery into host cells or the apoplast (Liu *et al.*, 2019). Effectors need to be flexible enough to evade host recognition and require a certain plasticity to bind host targets even with slight variations (Marín *et al.*, 2013). Recently, IDRs have been found to be responsible for antimicrobial activity, particularly in peptides with a positive net charge (Latendorf *et al.*, 2019). These cationic intrinsically disordered antimicrobial peptides (CIDAMPs) could be a novel source of highly specific antimicrobials, especially those from obligate biotrophs, as they do not harm the host but specifically shape the niche for the needs of the pathogen.

Here, we explore the antimicrobial activity associated with IDRs of apoplastic proteins from *Albugo candida* (Pers.) Kunze, and report for the first time an example of a protist and obligate biotrophic pathogen as a potential source for highly specific antimicrobials.

Materials and Methods

Interaction network inference

We inferred interaction correlations between operational taxonomic units (OTUs) on an amplicon sequencing dataset of wild *Arabidopsis thaliana*'s (L.) Heynh phyllosphere microbiome samples. The dataset combined analyses of 16S and 18S ribosomal RNA marker genes as well as of the internal transcribed spacer (ITS) regions with supporting metadata. We analyzed the raw OTU tables (M. Mahmoudi *et al.*, 2023, in preparation) for microbial correlations using the software FLASHWEAVE (Tackmann *et al.*, 2019; Supporting Information Methods S1).

Mass spectrometry and data analysis

We infected *A. thaliana* plants of ecotype WS-0 with *A. candida* strain Nc2 by spore suspension spray inoculation. We extracted apoplast samples from these as well as from uninfected plant leaves at 10 d post-infection (dpi) using vacuum pump infiltration (Methods S1).

For the mass spectrometry, we chemically denatured the apoplast proteins and reduced disulfide bonds followed by alkylation. We performed high pH fractionation using the Pierce High pH Reversed-Phase Peptide Fractionation Kit on the trypsinized peptides (Zittlau *et al.*, 2021) and labeled them with dimethyl on SepPak C18 Cartridges (Boersema *et al.*, 2009). We analyzed the samples on a Q Exactive HF mass spectrometer and on an Exploris mass spectrometer and processed raw data files with the MAXQUANT software suite v.1.6.7.0 (Cox & Mann, 2008). The MS/MS data were searched against the UniProt *A. thaliana* database (18 218 entries), the *A. candida* strain Nc2 predicted proteome, assembly accession: GCA_001078535.1 (Links *et al.*, 2011) and commonly observed contaminants (Methods S1). We considered an *Albugo* protein as present in the apoplast with high confidence when it had at least one predicted peptide match with an Andromeda score higher than 10 (Cox *et al.*, 2011). We assessed the relative abundance of proteins in the apoplast by the normalized ratio of the peptide intensities between infected and non-infected treatments.

Protein annotation and prediction

We performed functional annotation using INTERPROSCAN v.5 (Jones *et al.*, 2014) and analyzed the assigned gene ontology (GO) terms with the GOATOOLS software (Klopfenstein *et al.*, 2018). We used TPPRED 3.0 to predict mitochondrial localization (Savojardo *et al.*, 2015). For the antimicrobial assessment, we ran the predicted proteome of Nc2 through an antimicrobial prediction pipeline (https://github.com/danielzmbp/appred) consisting of several tools: ANTIMICROBIAL PEPTIDE SCANNER v.2, AMPGRAM, and AMPEPPY (Veltri *et al.*, 2018; Burdukiewicz *et al.*, 2020; Lawrence *et al.*, 2020). These machine learning-based models attempt *de novo* prediction of antimicrobial activity not by similarity but by the compound features of the amino acid sequence. We considered a prediction as positive for antimicrobial activity when the weighted average of the three methods was higher than the cutoff of 0.54 (Methods S1). To predict disordered regions, we used the tool FLDPNN with default settings and a cutoff of 0.3 (Hu *et al.*, 2021). We considered a protein to contain IDRs when it presented at least 15 predicted disordered residues in a consecutive order. To detect compositional bias in the protein sequences, we used FLPS 2.0 (Harrison, 2021). Further molecular prediction methods are described in Methods S1.

Heterologous expression, purification, and rebuffering

We synthesized complementary DNA (cDNA) from total RNA extracted from *A. candida*-infected *A. thaliana* plants. From this, we amplified candidate sequences and cloned them using the In-Fusion® cloning method (TaKaRa Bio, Kusatsu, Japan) into pET28b vectors for expression in *Escherichia coli* (Table S1). We cloned candidates that had a putative secretion signal as predicted by the slow model implemented in SIGNALP v.6, namely C14 and C15, without the secretion-signal encoding sequence (Teufel *et al.*, 2022). We overexpressed the candidate proteins in *E. coli* strain SHuffle® (New England Biolabs, Ipswich, MA, USA) at 30°C for candidates C14 and derivatives and strain Rosetta™ DE3 (Merck, Darmstadt, Germany) at 37°C for candidates C05, C15, C06, and derivatives (Methods S1).

We extracted inclusion bodies from cell pellets under denaturing conditions (Methods S1). The denatured proteins were purified with HisTrap™ excel (GE Healthcare, Chicago, IL, USA) in a one-step elution (elution buffer: 100 mM sodium phosphate, 10 mM Tris, 7 M urea, pH 4.5). This was followed by concentration and rebuffering in testing buffer (10 mM BisTris, pH 5.9) via dialysis (Methods S1).

We analyzed digestion in the native system by incubating the expressed proteins with extracted apoplastic fluid (Methods S1). We assessed the secondary structure of the purified protein domains by circular dichroism (CD) spectroscopy and application of a predictor to the spectra for classification into ordered or disordered (Micsonai *et al.*, 2022; Methods S1).

Antimicrobial testing

During the pairwise inhibition assays, we tested strains from a stock of plant-isolated bacteria collected during *A. thaliana* sampling (bacterial strains I01-I36, Table S2; strains I37-I46, Table S3; Methods S1). From these, we designed two synthetic communities (SynComs) for community testing. SynComA comprised 12 bacterial strains and three yeasts (Table S2) from the *A. thaliana* core community. SynComB was constructed as a smaller version of Syn-ComA and comprised five bacterial strains (I01, I03, I04, I13, and I28; Table S2). We grew overnight cultures from single colonies in nutrient broth medium at 22°C. We adjusted protein molarity and added the rebuffered proteins at a v/v of 1 : 1 to the cell cultures diluted to a starting 600 nm optical density (OD_{600}) of 0.1 for a final testing volume of 100 µl. We used the final buffer from the dialysis as control. We measured OD_{600} every 15 min over 19 h with shaking at 22°C (Methods S1).

We conducted in culture bacterial community experiments with two of the peptides (C06d and C06b) using SynComB. We mixed SynComB suspensions in a 1 : 1 v/v with adjusted protein solutions followed by a 24 h incubation at 22°C. We then added rifampicin-resistant *Pseudomonas syringae* pv tomato DC3000 (Pst) and continued incubation for another 24 h. We counted Pst colonies on rifampicin containing plates after incubating for 32 h at 22°C. We performed the following treatment combinations: (1) SynComB plus buffer, (2) SynComB without *Aeromicrobium fastidiosum* I01 plus buffer, (3) SynComB plus peptide C06d, (4) SynComB plus peptide C06b (Methods S1). For assessing interactions of *A. fastidiosum* I01 with other bacterial species from SymComA, we examined one-to-one inhibition by a co-cultivation setup and analyzed the inhibition zone (Methods S1).

In planta SynCom assays

To conduct the gnotobiotic SynCom experiments, we created a suspension of SynComA (12 bacteria and three yeast; Table S2) and sprayed it onto 4-wk-old sterile *A. thaliana* plants. We sprayed the plants with the Pst pathogen 1 wk after the SynCom inoculation (Methods S1). We sampled plants for amplicon sequencing immediately after SynCom inoculation, before Pst inoculation on day 7 and 3 d post Pst inoculation. Three samples were taken from each of the three treatments, whole SynComA, SynComA without *A. fastidiosum* I01 and mock control without SynCom. We extracted DNA from the samples using a PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc., Hilden, Germany). We created amplicon libraries of the bacterial 16S rRNA gene and ITS2 region using the AmpStop method (Agler *et al.*, 2016; Mayer *et al.*, 2021; Table S4) and performed Illumina sequencing (Methods S1). We processed and analyzed the raw sequencing reads using QIIME2, with which we calculated Pst relative abundances and principal coordinate analysis (PCA) based on Euclidean distances of the rarefied amplicon sequencing variant (ASV) frequencies to measure sample to sample variation (Bolyen *et al.*, 2019).

We used SynComB (I01, I03, I04, I13, I28; Table S2) for the *in planta* antimicrobial protein assay. We mixed SynComB with 2 µM purified peptides in a 1 : 1 ratio and applied it to non-sterile *A. thaliana* plants via dipping the individual leaf tips. In total, we applied two combinations: SynComB plus C06d and SynComB plus C06b. We combined three treated leaf tips per plant in one sample and used three plants for each treatment combination as biological replicates. We cut out leaf disks (diameter: 0.8 cm) from the treated leaf tips at sampling points 1 h (T0) and 3 d (T1) after application. We used *A. fastidiosum* I01 alone instead of SynComB in the same setup (Methods S1).

Nanopore sequencing and quantitative PCR

We used the phenol/chloroform method for extraction of DNA from the frozen leaf samples (derived from Sambrook & Russell, 2001). We amplified the 16S rRNA bacterial genes using the AmpStop method (Mayer *et al.*, 2021; Table S4). For the amplicons of each biological replicate, we constructed one Nanopore sequencing library using the Native Barcoding Kit (SQK-NBD112.24; Oxford Nanopore Technologies (ONT), Oxford, UK). Sequencing reactions were performed on a MinION flow-cell (R10.4; ONT) connected to a MK1B device (ONT). Sequences were assigned to the full 16S gene sequence from each of the five SynComB members and the *A. thaliana* chloroplast sequence through BLASTN (Camacho *et al.*, 2009). We calculated the relative reads of the microbial strains to the chloroplast reads from the plant in each respective sample (Methods S1).

To assess the absolute abundance of the *A. fastidiosum* I01, we performed quantitative PCR (qPCR) on the extracted DNA

using primers specific to the 16S rRNA gene of strain I01 and reference primers targeting the *A. thaliana* elongation factor 1-alpha (Table S4). We used the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and the CFX Connect™ system (Bio-Rad laboratories, Hercules, CA, USA). We assessed the abundance of *A. fastidiosum* I01 in reference to the elongation factor 1-alpha gene of the plant.

Results

Albugo is highly intercorrelated

We set out to predict robust and ecologically relevant correlations between *Albugo* and other co-occurring microbes, as well as assess the significance of the number of these correlations in the context of the phyllosphere. With these aims, we applied the software FLASHWEAVE to infer direct interactions between the OTUs of a large amplicon sequencing dataset of wild *A. thaliana* from six consecutive years in six locations (735 samples containing a total of 11 172 OTUs). In the resulting correlation network, consisting of 123 316 edges and 11 150 OTUs, we found the *Albugo* sp. OTU to be in the upper top 0.85 quantile of the total interactions with a degree of 33 including 21 positive interactions (Figs 1a, S1). Furthermore, this OTU was in the top 20 when ranked by negative interactions with a total of 12, including connections to bacteria, fungi, and one other non-fungal eukaryote (Figs 1b–d, S2). The negatively correlated bacteria included mostly Gram-negative strains of which the most abundant phylum was the Proteobacteria with four species members. The negatively correlated eukaryotes included an ascomycete fungus and two green algae (Fig. 1d). In summary, our analysis indicates that the protist pathogen *Albugo* is highly intercorrelated, relative to others in the phyllosphere, particularly through negative correlations when compared to the nature of interactions among other plant microbes.
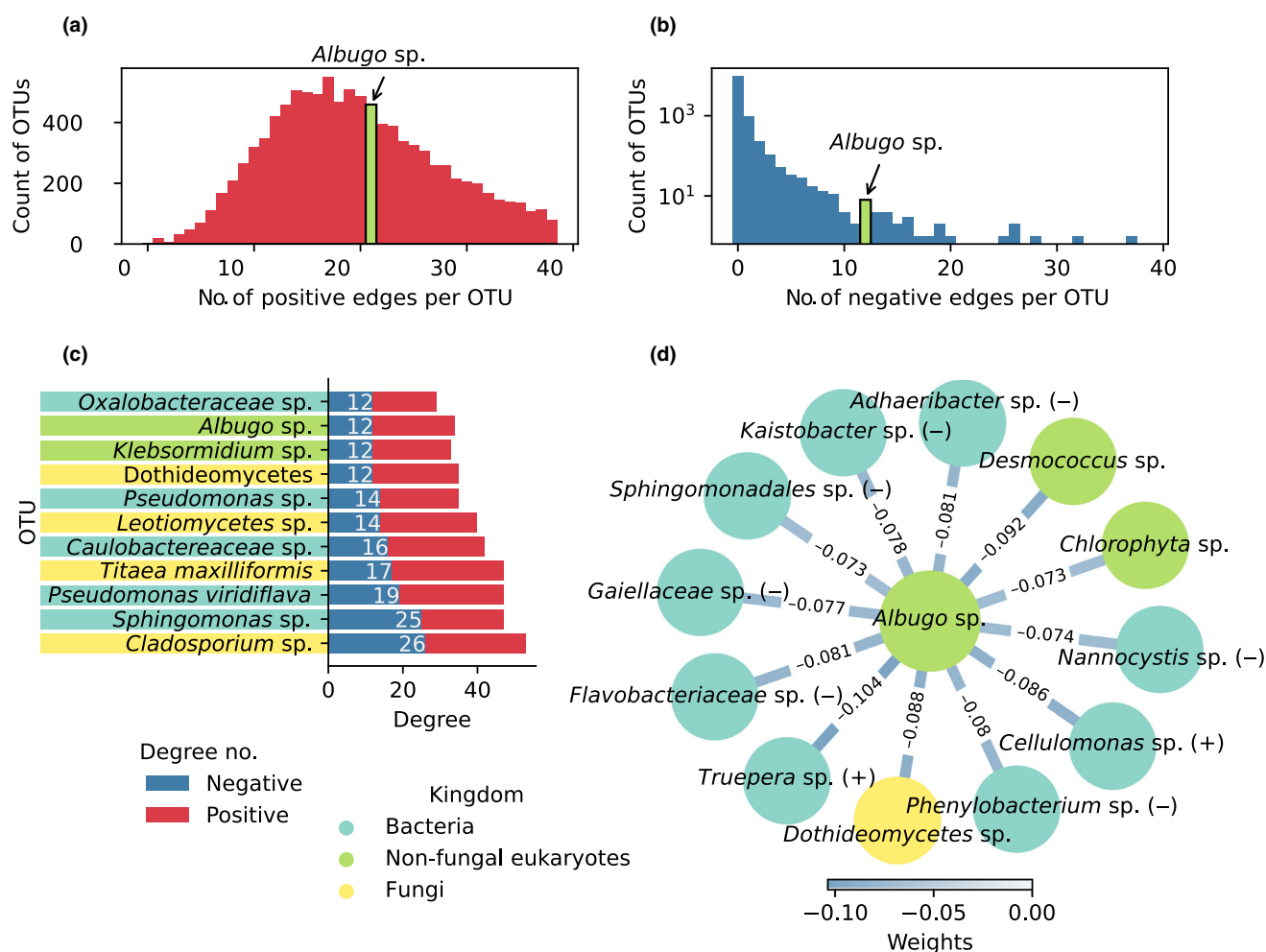

Fig. 1 Network of operational taxonomic unit (OTU) interactions in the phyllosphere of *Arabidopsis thaliana*. Histograms of the number of positive (a) and negative (b) correlations (degree) as predicted by FLASHWEAVE on the *A. thaliana* phyllosphere amplicon dataset. Highlighted is the location of the *Albugo* sp. OTU in the bins. (c) Degree distribution of OTUs in the interaction network of phyllosphere data, displaying nodes with 12 or more negative interactions and > 100 000 total reads in the dataset, where *Albugo* places tenth. (d) Inferred negative interactions for the *Albugo* sp. OTU, totaling 12 connections to other members in the network. Color of the edges and nodes represent the strength of the correlation and the phylogenetic kingdom to which the node belongs, respectively. Unless explicitly stated, taxonomy at the species level could not be resolved with confidence (bootstrap < 90). Gram stain of bacterial nodes is displayed in parentheses.

Antimicrobial proteins are enriched in the apoplast

To identify potential, protein-based causal agents of such negative correlations resulting in a reduction of microbial diversity, we studied *Albugo* proteins released into the plant apoplast (Fig. 2a). Through proteomic analysis, we identified 563 high confidence proteins from *A. candida* in the apoplast of infected *A. thaliana* leaves, representing 4.2% of the total predicted proteome of *A. candida* (Fig. 2b). Among these, 78 proteins (13.8%) carried a putative secretion signal, comparable to similar analysis in one other obligate biotrophic oomycete (Figueiredo *et al.*, 2022). Mitochondrial localization was assigned to 24 proteins (4.3%) with TPpred. Although this suggested minor contamination from broken hyphae, the intensity of mitochondrial-annotated protein was overall low, only seven surpassing the raw median intensity of *Albugo* proteins. In addition, it is not uncommon for fungi to release mitochondrial proteins in extracellular vesicles (Liebana-Jordan *et al.*, 2021), considering their similarity

in lifestyle, it is likely oomycetes use similar strategies. We found that 457 out of the 563 proteins (76.2%) resulted in significant hits for at least a GO term after annotation (Fig. S3). For these, we studied the enrichment of biological functions compared to the predicted intracellular proteome of *A. candida*. Carbohydrate, amino acid and nucleic acid, catabolism and biosynthesis featured prominently in the enriched terms (Fisher's exact test Holm-corrected $P < 0.001$; Fig. 2c). To find novel antimicrobial proteins, we employed a *de novo* antimicrobial activity prediction approach. In this analysis, we found two of the tools (AmpGram and amPEPpy) to be biased towards either a longer or shorter protein length in our dataset, therefore we adjusted the weight based on the *R*-squared value of the correlation (Fig. S4). Following the weighted score, a total of 131 apoplastic proteins were found to be positive for antimicrobial activity (23.3%, compared to 20.1.3% in the entire *A. candida* proteome; Table S5) with an average amino acid sequence length of 499. This corresponded to a significant enrichment for the presence of predicted
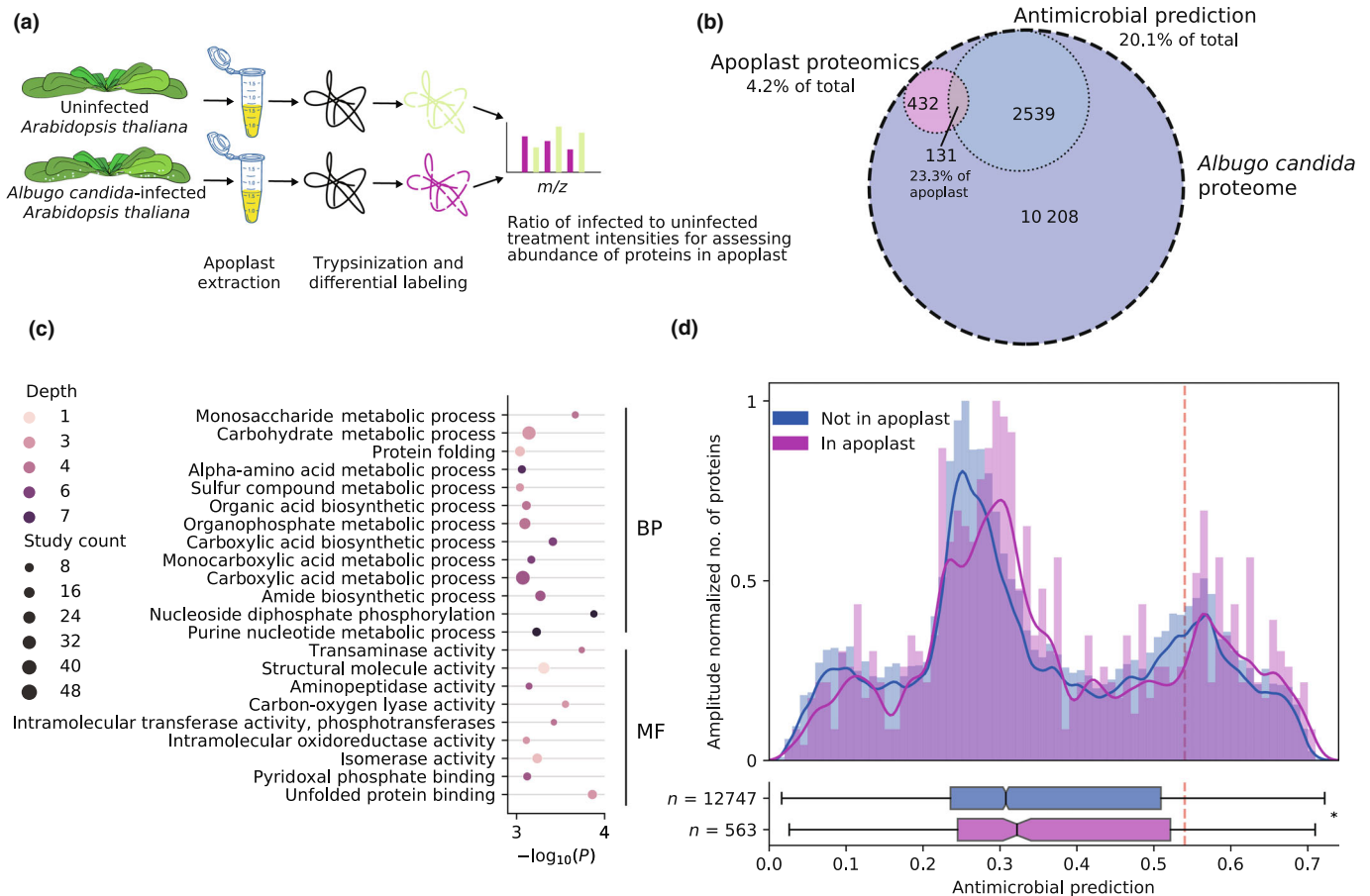
Fig. 2 Proteomics of *Albugo candida* apoplastic proteins from infected *Arabidopsis thaliana* leaves. (a) Workflow of the proteomics analyses on infected (10 d post-infection) and uninfected *A. thaliana*. (b) Venn diagram of *A. candida* proteome. A total of 563 proteins were identified in the apoplast, of which 131 have a predicted antimicrobial function. (c) Enrichment of gene ontology terms related to biological processes (BP) and molecular functions (MF) in apoplastic proteins compared to background proteome in *A. candida* with a significance of $P < 0.001$. Depth represents the position of the term in the gene ontology (GO) hierarchy tree (higher depth results in more specific terms) and study count the number of enriched genes which contain that term. (d) Normalized histogram of antimicrobial predictions in proteins found or not in the apoplastic proteomics for *A. candida* with kernel density estimate represented as a line. At the bottom, a box plot represents the distribution of proteins with antimicrobial prediction within the apoplastic and non-apoplastic subset (Mann–Whitney *U*-test, $P = 0.015$). A dashed red line indicates the cutoff of 0.54.
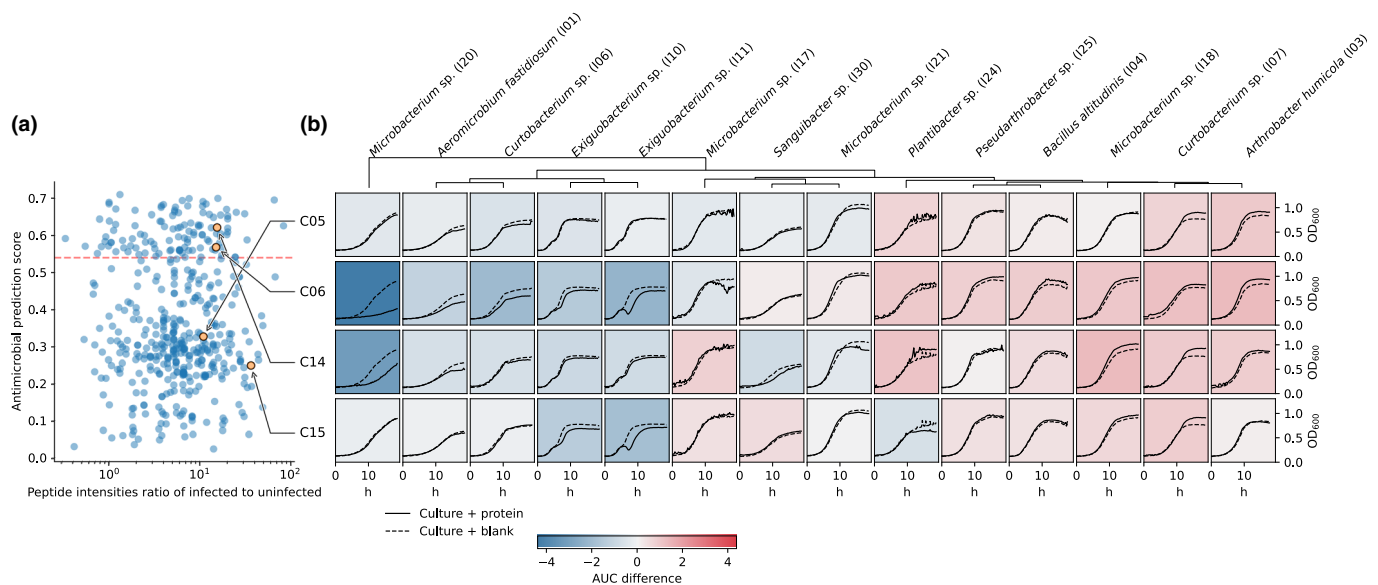
antimicrobial proteins in the apoplastic proteome (one-sided Fisher's exact test, $P = 0.031$; Fig. 2d).

To corroborate and study the potential antimicrobial properties of these proteins in more detail, we selected candidates for overexpression in a heterologous system. We considered the following properties when choosing candidate proteins: (1) Positive prediction for antimicrobial activity (cutoff higher than 0.54 for consensus prediction), (2) high abundance in the apoplast as measured by the relative peptide intensity compared to that of the uninfected treatment in the proteomics (above median of the normalized distribution), and (3) representative sequence length close to average length of candidate pool (499 amino acids). As negative controls, we selected proteins with a lower antimicrobial prediction that had a comparable size and abundance to the antimicrobial candidates. We were able to amplify representative candidates with a positive (C06 and C14) and a negative (C05 and C15) prediction for antimicrobial activity using as template a cDNA library of *A. candida*-infected *A. thaliana* (Fig. 3a; Table 1). All four proteins were found with a comparable abundance in the proteomics (Table 1) and had a similar molecular weight (49–60 kDa when expressed; Table 2). All except C06 had a functional annotation (Table S6). The candidates, as with the majority of proteins assigned to *A. candida* in the proteomic analysis, had no overlap with plant peptides (Fig. S5). Of note, the predicted peptide signal for classical secretion of C14 and C15 was removed during cloning. The candidates were subsequently heterologously expressed in *E. coli* to test *in vitro* the antimicrobial activity of the corresponding recombinant proteins as described below.

Heterologously expressed candidates show antimicrobial activity

During overexpression of the candidates in the *E. coli* system, we observed accumulation of C05 and C14 in inclusion bodies after IPTG induction under standard conditions. By systematically testing different expression settings, including lower temperature (15–37°C), lower inducing concentration (0.1–1 mM IPTG), and longer induction time (4–48 h), we could natively extract C06 and C15 as soluble proteins. Extraction under denaturing conditions using urea was successful for all proteins, regardless of whether they were synthesized into inclusion bodies or not (Fig. S6). Therefore, we used a denaturing extraction protocol as the standard purification method for comparison of all the expressed candidates (Fig. S7).

After purification, concentration and rebuffering of the candidates in a testing buffer resembling the apoplastic pH conditions (BisTris-based buffer at pH 5.9), we performed an antimicrobial screen on a selection of 24 strains from an in-house microbial strain collection of plant-isolated bacteria from *A. thaliana* samples that were cultured under standard conditions (Table S2). The microbial collection also includes strains that were detected as core members in the *A. thaliana* phyllosphere microbiome, meaning they are consistently present in natural communities and hence are likely to inhabit the plant before or during an *Albugo* infection (Almario *et al.*, 2022). Overall, we observed a variable effect on the growth of bacteria, with species within the same genus showing different responses. We found selective antimicrobial activity against five strains, including two *Exiguobacterium* (I10, I11), a *Curtobacterium* (I06), an

Fig. 3 Antibacterial prediction and activity of apoplastic candidate proteins from *Albugo candida*. (a) Antimicrobial prediction and relative intensity of the proteins found in the apoplast of *Arabidopsis thaliana*, highlighting the selected candidates. (b) Growth curves of Gram-positive bacterial strains from the phyllosphere microbiome with candidates C05, C06, C14, and C15 at a concentration per well of 0.75 μM compared to blank (dashed lines) during 19 h of growth. Background color represents inhibition (blue) or promotion (red) of growth based on the difference in the area under the curves (AUC). 95% confidence intervals for these tests are shown in Supporting Information Fig. S9. Information about tested strains is found in Table S2. Growth curves of all 24 tested strains are shown in Fig. S8.

14698137, 0, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.18995 by Daniel Gómez-Pérez - Cochrane Germany , Wiley Online Library on [06/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1 Apoplastic protein candidates from *Albugo candida* selected for heterologous expression and antimicrobial testing.

Candidate	Accession (NCBI)	Functional annotation (IP database)	Antimicrobial weighted score	Secretion signal (aa)	Length (aa)	Abundance	Predicted IDRs
C05	CCI44519.1	Myo-inositol-1-phosphate synthase	0.33	No (0)	516	10.9	No
C06	CCI46028.1	None	0.57	No (0)	505	15.1	Yes
C14	CCI45607.1	Glucanosyltransferase	0.62	Yes (22)	453	15.5	Yes
C15	CCI42480.1	Glucan-1,3-β-glucosidase	0.25	Yes (21)	449	36.5	Yes

Abundance in the apoplast is represented as the ratio of the intensity peptide values comparing infected and uninfected *Arabidopsis thaliana* of the three replicates after normalization. Detailed information regarding functional domains and orthogroups can be found in Supporting Information Table S6. aa, amino acid; IDR, intrinsically disordered region; IP, INTERPRO; NCBI, National Center for Biotechnology Information.

Table 2 Net charge at pH 5.9 and 7.2, theoretical isoelectric point (pI), and molecular weight in kDa of the expressed candidate proteins and domains.

| Candidate	Theoretical pI	Net charge pH 5.9	Net charge pH 7.2	Molecular weight (kDa)
C05	6.27	5.90	−7.92	60.2
C06	6.28	8.05	−9.96	57.1
C06d	6.86	13.55	−1.27	21.2
C06b	6.05	1.76	−7.92	39.4
C14	6.23	2.85	−5.27	48.9
C14d	8.79	8.03	4.00	15.1
C14b	5.60	−2.82	−9.91	35.0
C15	5.89	−0.36	−12.77	49.0

Aeromicrobium (*A. fastidiosum* I01), and a *Microbacterium* (I20). To a lesser extent, we observed antimicrobial activity on four other strains, a *Sanguibacter* (I30), a *Plantibacter* (I24), and two *Microbacterium* (I17, I21; Figs 3b, S8, S9).

Candidate C06 showed the highest inhibitory activity, while C14 and C15 showed minor inhibition at equal molarity (0.75 μM). Based on the *in silico* prediction, we expected C06 and C14 but not C15 to display antimicrobial activity (Fig. 3a). C05, instead and consistent with the antimicrobial prediction, had the least antimicrobial effect of all. We additionally found that all proteins displayed a variable growth promoting effect towards most other Gram-positive strains: a *Curtobacterium* (I07), a *Pseudarthrobacter* (I25), a *Microbacterium* (I18), a *Streptomyces* (I36), an *Exiguobacterium* (I09), two *Bacillus* (I04, I05), and two *Arthrobacter* (I02, I03). The five tested Gram-negative strains remained largely unaffected by incubation with the protein candidates during growth (Fig. S8). We also tested the antimicrobial activity at the higher pH of 7.2 to see whether apoplastic conditions were necessary for the inhibitory effect to take place or if cytoplasmic conditions are more likely to enhance antimicrobial function. We found a significant loss of antimicrobial activity for C06 at the pH of 7.2 (Fig. S10). Thus, in summary, we found a high correlation of the consensus prediction method with the experimental antimicrobial results exclusively at a pH reflecting apoplastic conditions.

Computational analysis of all predicted antimicrobial proteins that had been identified by proteomics of the apoplast revealed a significant enrichment for IDRs compared to non-apoplastic proteins (Fisher's exact one-tailed test, $P = 0.002$). This suggests the prevalence of disordered regions in the apoplastic proteome. Consistent with this, the candidates that showed antimicrobial activity had long predicted IDRs, notably in the C-terminus with a positive net charge (Fig. 4a; Table 2). C15, which was not predicted as an antimicrobial, did also present putative IDRs but they were slightly shorter in extension at the N-terminus (maximum stretch of 23 amino acids, vs 129 for C06 and 30 for C14). In addition, C06 had a compositional bias for alanine and glutamine, and C14 and C15 had a bias toward increased serine (Table S7), all of which are disorder-associated amino acids (Uversky, 2013). To test whether the positively charged IDRs were responsible for the antimicrobial activity, we separately cloned and purified the C-terminal regions of C06 and C14 (165 and 129 amino acids long, respectively), which were both predicted to contain IDRs and display high positive net charge (Tables 2, S8), as well as the remaining N-terminal region of each protein, which was for the most part not predicted to be disordered. We corroborated the disordered status of b and d domains through CD spectroscopy followed by application of a binary classifier. In the spectra comparison of both domains, a striking difference was apparent between the ordered and disordered domains of each protein, particularly of C06 (Fig. S11).

We found the strongest antimicrobial activity for both disordered C-terminal domains of C06 and C14 (C06d and C14d) when compared to the whole protein and the N-terminus domains (Figs 4b, S12). We observed a correlation between the predicted net charge of the peptides and proteins of C06 and C14 and their antimicrobial effect: the higher the net charge, the higher the antimicrobial activity (Fig. 4c). Furthermore, for C06d and C14d, we observed a concentration-dependent effect against *Exiguobacterium* strains I10 and I11, *A. fastidiosum* strain I01 and *Microbacterium* strain I20, with C06d reaching a complete inhibitory activity against *A. fastidiosum* I01 and *Microbacterium* I20 growth at 2.15 μM (Figs 4d, S13). In the proteomics dataset, we found not only coverage for a large part of the sequence of both positive candidates (Fig. 4a) but the identified peptides also indicated the presence of the disordered domains with antimicrobial activity in the apoplast, particularly for C06d (Table S9). To support the hypothesis that the proteins might get cleaved allowing the release of the IDR peptides in the apoplast, we incubated the full-length proteins with Nc2-infected apoplast extracts. We observed cleavage of these resulting in a decrease in the intensity of the band corresponding to the full-length protein
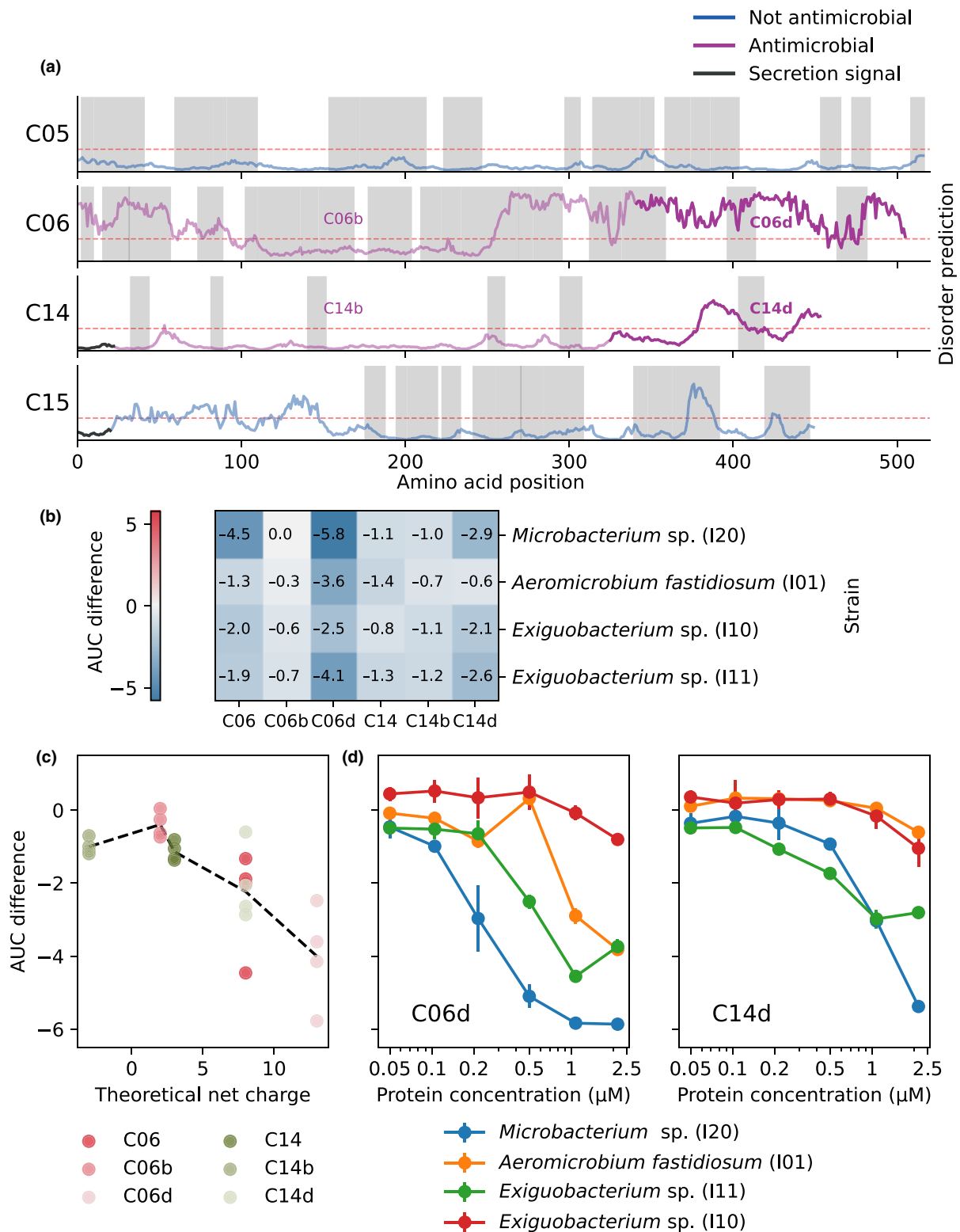
Fig. 4 Antimicrobial activity of protein candidates from *Albugo candida*, C06, C14, and their domains. (a) Individual candidate predictions of disorder per amino acid, emphasizing the annotated domains used for antimicrobial testing in bold (domain d) or faded color (domain b). The threshold for positive disorder (0.3) is marked in red. The secretion signal of C14 and C15 is depicted in gray. Shaded regions represent the peptide coverage as found in the proteomics analyses. (b) Comparison of the antimicrobial effect of the different domains of C06 and C14 on the growth of the bacterial strains at a final concentration per well of 1 μM shown as the differences in the area under the curve (AUC) between protein and buffer treatment during 19 h of growth. The values represent the mean of at least three biological replicates. Full curves are found in Supporting Information Fig. S12. (c) Antimicrobial activity as measured by the change in the AUC correlated to predicted net charge for all domain tests with LOWESS regression showing a downwards trend line (black dashed line). (d) Concentration-dependent inhibitory activity of C06 and C14 C-terminal disordered domains toward four sensitive strains. Vertical bars represent SD. Full curves are found in Fig. S13.

and appearance of smaller fragments (c. 20, c. 19, < 10 kDa for C06 and c. 45, c. 40, c. 34 and c. 30 kDa for C14; Fig. S14). In addition, when looking at the proteomics dataset that represents the apoplastic content derived from the host plant *A. thaliana*, we found evidence of an upregulation of proteases in the apoplast, particularly subtilisin-like serine proteases, in the presence of *A. candida* infection based on raw intensity peptide ratios between infected and non-infected plant samples (Table S10). Hence, we conclude that the observed antimicrobial function of C06 and C14 might be traced back to the IDRs in the C-terminus, which is in line with reports from human proteins that contain IDRs, the CIDAMPs.

To see whether C06d, as the strongest inhibitor, could in principle be used as an effective microbial control agent, we conducted antimicrobial assays with three Gram-positive strains isolated from *A. thaliana* that were closely related to known plant pathogens (Table S3). These included *Rhodococcus fascians* (Hjerde *et al.*, 2013; Dhaouadi *et al.*, 2020), *Clavibacter michiganensis* subsp. *tessellarius* (Carlson & Vidaver, 1982; Li & Yuan, 2017) and subsp. *capsici* (Oh *et al.*, 2016). As in the previous tests, we found the inhibition to be strain specific. While *R. fascians* (I45, I46) and *C. michiganensis* subsp. *tessellarius* (I43, I44) were unaffected by protein addition, all tested isolates from *C. michiganensis* subsp. *capsici* (I37–I42) showed a variable degree of growth inhibition (Fig. S15). Therefore, due to its ability to inhibit phytopathogenic *Clavibacter*, C06d would present a viable candidate for further investigation regarding its role as a potential pathogen-control agent.

Apart from pairwise inhibition assays, we were also interested in the community-level effects. To examine whether the peptides showed inhibitory activity within a microbial community context, we conducted SynCom experiments in culture media. SynComB, composed of *A. fastidiosum* I01, which was inhibited in the previous tests, and four additional strains (I03, I04, I13, and I28), all unaffected in their growth by protein addition, was constructed (Fig. S16). We added a rifampicin-resistant pathogenic strain from *P. syringae* (Pst) to use as readout. We found that the growth of Pst was increased when *A. fastidiosum* I01 was excluded from SynComB, as compared to the whole five-strain SynComB, when grown with testing buffer. To see whether C06d could effectively inhibit *A. fastidiosum* I01 in a multi-strain community, we incubated it with the protein solution. In this setting, the growth of Pst was comparable to the *A. fastidiosum* I01 dropout treatment, indicating inhibition by C06d. As a control, we also tested the SynCom with C06b, which showed no inhibitory activity. The Pst load, as measured by the colony-forming unit (CFU) count per ml was similar to the whole five-strain Syn-ComB treatment in buffer (Figs 5a, S17). Together, these results indicate that the inhibitory properties of C06d are persistent in an *in vitro* multi-strain bacterial setting.

With the aim to test whether these properties are maintained *in planta*, we conducted similar SynCom experiments on *A. thaliana*. Following incubation of *A. fastidiosum* I01 and Syn-ComB with either C06d or C06b on the surface of *A. thaliana* plants, we monitored the leaf microbiota at 1 h (T0) and 3 d (T1) after inoculation. When applying only strain I01 with

protein C06d or C06b, we found that the abundance of *A. fastidiosum* I01 decreased in the presence of C06d and increased with C06b between timepoints T0 and T1 (Fig. 5b,c). While individually not significant, we observed the same trend using two methods. On one hand, we found that *A. fastidiosum* I01 abundance relative to the plant measured in the qPCR was smaller at T1 than T0. On the other, we observed a decrease in the number of *A. fastidiosum* I01 reads relative to the plant during sequencing. Apart from analyzing the *in planta* single strain effect, we also assessed the community-level behavior. The abundance of *A. fastidiosum* I01 showed a similar decrease between timepoints with addition of C06d on both qPCR and sequencing. After adding C06b, we found a small decrease as well in the relative reads, however, much smaller than the decrease with C06d. For all the other strains except *Flavobacterium* I13, we observed an increase in their relative reads with the C06b domain while C06d treatment resulted in a stable or slight decrease in the reads for all (Fig. S18). Especially for *Flavobacterium*, but also for most other strains, these results do not exactly reflect the *in vitro* assays, likely due to unknown confounding variables playing a big role in the *in planta* experiments. However, taken together, these results support specific inhibition of C06d against *A. fastidiosum* I01 in the context of the plant.

Inhibited microbes are important for community stability

To test for the relevance of the inhibited microbes in a natural ecological context, we assessed their necessity for community stability *in planta*. We performed dropout experiments on a SynCom in *A. thaliana* consisting of 12 bacterial strains and three yeasts (Syn-ComA), all identified as core microbes, challenged with the model plant-pathogenic strain of *P. syringae*. In SynComA without *A. fastidiosum* I01, we observed a variable decrease in the resistance of the plant to infection (Fig. S19). We also observed an overall increase in the sample-to-sample variation of the bacterial abundances compared to the whole SynComA treatment, which were consistent for the three replicates. This was generalizable to the entire bacterial composition of the samples as measured by the lack of clustering in the PCA for the *A. fastidiosum* I01 dropout (Fig. 5d). In addition, through pairwise interaction assays with SynComA members, this *A. fastidiosum* strain I01 was shown to inhibit other bacterial strains. This highlighted its potential in stabilizing the microbiome by controlling the abundance of other core microbes in the phyllosphere of *A. thaliana* (Fig. S20). Further microbial lawn experiments of *A. fastidiosum* I01 showed its resilience against other SynCom members in one-to-one interactions (Fig. S21).

Discussion

The plant apoplast is a challenging habitat for the survival of microbes. On the one side, plant defenses, including oxidative bursts and proteases, and on the other, microbes defending their niche make it a difficult environment to conquer (Jashni *et al.*, 2015; Wang *et al.*, 2020). *Albugo* is a filamentous pathogen that has developed a niche-forming strategy, whereby it modifies
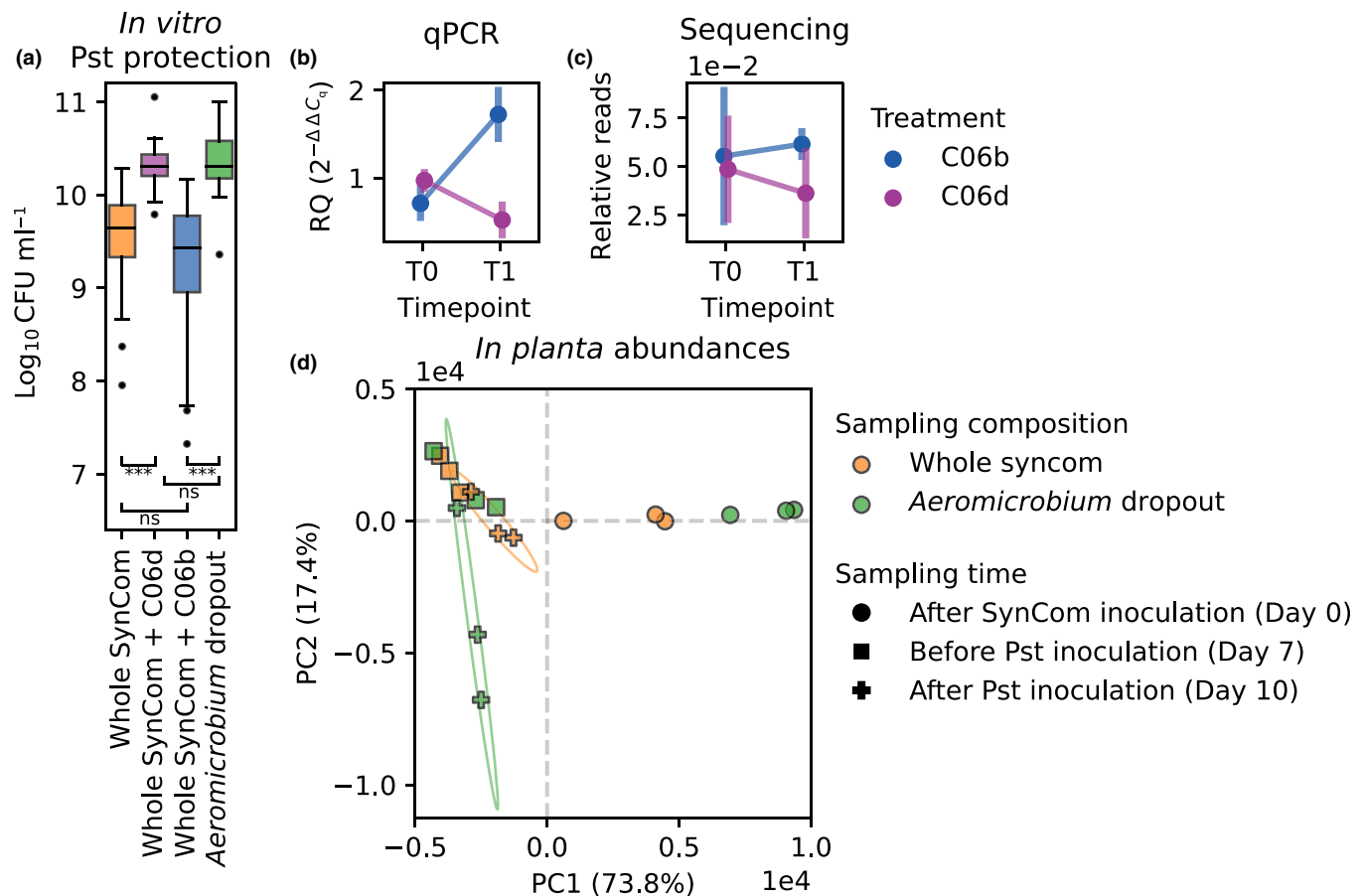
Fig. 5 Synthetic community (SynCom) experiments in liquid culture and *in planta* suggest relevance of *Aeromicrobium fastidiosum* I01. (a) C06d (purple) and C06b (blue) 1 μM protein treatments on a five-strain bacterial community (SynComB) in liquid culture result in variable community disruption as measured by the bacterial load of *P. syringae* pv tomato DC3000 (Pst) in colony forming units (CFU) ml^{-1}. Whole SynCom with buffer control (BisTris, pH 5.9) is shown in orange, *A. fastidiosum* I01 dropout community in green. Significance tests show a difference between whole SynCom treatment and C06d treatment, as well as C06b treatment and *A. fastidiosum* I01 dropout (Mann–Whitney *U*-test: ns, not significant; ***, $P < 0.001$). The box represents the distribution between the first and the third quartile and the whiskers extend one and a half times the interquartile range on each side. Outliers are denoted with a point. Detailed counts per replicate are found in Supporting Information Fig. S17. Change in abundance of *A. fastidiosum* I01 during single strain *in planta* application with domains C06d and C06b on *Arabidopsis thaliana* leaves as measured with qPCR (b) and sequencing (c). Time points T0 and T1 correspond to 1 h and 3 d after application, respectively. Confidence is shown as the SE based on three biological replicates. In (b), the relative abundance of *A. fastidiosum* I01 to *A. thaliana* as measured by quantitative PCR was calculated from the difference in quantification cycles (C_q) between this strain and *A. thaliana*. qPCR results from community *in planta* testing are found in Fig. S18a. In (c), the relative reads from *A. fastidiosum* I01 respective of chloroplast reads of *A. thaliana* were obtained from Nanopore sequencing. Sequencing results from community *in planta* testing are shown in Fig. S18b for strain I01 and the four other strains of SynComB. (d) Principal coordinate analysis of the amplicon sequencing variant (ASV) frequencies in the SynComA dropout experiment on *A. thaliana*. Calculated based on the Euclidean distances of bacterial 16S rRNA ASVs abundances. Ellipses drawn for the final time points based on two SDs show a large variation for the *A. fastidiosum* I01 dropout but not for the whole SynCom after Pst disruption.

the pre-existing plant microbiome to better fit its needs. Some mechanistic explanations for this phenomenon have been attributed to *Albugo*'s effect on the immune system of the plant and its indirect repercussions on other microbes (Cooper *et al.*, 2008; Ruhe *et al.*, 2016). Specifically, by reducing the plant's defense recognition systems, the pathogen allows the thriving of microbes that otherwise would not, or vice versa. Here, we provide evidence suggesting that *Albugo* directly contributes to shape the plant-associated microbial communities through the release of proteins and peptides with antimicrobial activity into the apoplast.

The OTU interaction analysis of the *A. thaliana* microbiome dataset showed a significant number of positive and negative interactions of *Albugo* with specific community members, supporting previous findings (Agler *et al.*, 2016). Overall, the normal

distribution of positive interactions in the entire network signifies that positive relationships are a common occurrence in the phyllosphere, likely due to sharing of metabolites and other nonspecific cooperation (Fig. 1a). Meanwhile, the power law distribution of negative degrees indicates specific interactions by members of the network, where only some nodes are highly negatively correlated (Fig. 1b). This suggests that these nodes might represent the source of inhibition, through, for example, release of antimicrobial compounds.

Particularly for *Albugo*, the adjacent positive interactions could imply either promotion of these organisms as potentially beneficial for *Albugo* or their hyperparasitic nature as they benefit from its presence (Fig. S1). Within the network of positive interactions, we found a correlation between *Albugo* and the bacterial

genus *Variovorax*, which, as a common plant growth-promoting bacterium, could promote survival of the host plant during infection, thus indirectly promoting *Albugo* (Chen *et al.*, 2013; Finkel *et al.*, 2020). Although mostly reported in the rhizosphere, it was shown to protect the plant when in the phyllosphere as well (Bruisson *et al.*, 2019; Perazzolli *et al.*, 2020). By contrast, the numerous direct negative interactions can be explained by *Albugo*'s need to outcompete microbes that may be detrimental to its own survival (Fig. 1d). Although the network is undirected and negative interactions could mean a repression by *Albugo* or the opposite, reduced microbial diversity in the phyllosphere upon *Albugo* infection points to a majority of negative connections being outgoing rather than ingoing (Agler *et al.*, 2016). In agreement with this, *Albugo*, as an obligate biotroph, also has a need for controlling the growth of microbes threatening survival of the host. By reducing the plant's defenses through the release of effectors, *Albugo* may leave the plant vulnerable to certain pathogens, which may have resulted in an adaptive antimicrobial response. For instance, *Cellulomonas*, which is shown in the network to be negatively correlated with *Albugo*, is a bacterial genus known for its plant cell wall degrading capabilities (Aydogan *et al.*, 2018; Carlos *et al.*, 2018). To keep the plant alive while suppressing its defense in the presence of numerous facultative pathogens is a general problem biotroph symbionts and pathogens face. In any case, they need to be able to control all those microbes, making them a currently unexploited resource for novel mechanisms of antimicrobial strategies.

One strategy to suppress competitors is the release of secondary metabolites. The genome of *Albugo*, however, does not encode key secondary metabolites potentially responsible for the observed negative correlations with other microbes (Kemen *et al.*, 2011). In fungi, protein effectors for microbe–microbe interactions were identified that have direct or indirect inhibitory effects on competitors (Snelders *et al.*, 2020; Eitzen *et al.*, 2021). For example, a fungal yeast has been shown to secrete a glycoside hydrolase responsible for inhibition of *Albugo* (Eitzen *et al.*, 2021). Functional annotation of *c.* 80% of the *A. candida*'s apoplastic proteins revealed a significant enrichment for metabolism-related processes (Fig. 2c). This might be explained by the sampling time point, that is, 10 d after infection, when a high metabolic turnover is required by the oomycete due to the active growth of hyphae. Furthermore, GO terms found to be enriched on our analysis (e.g. monosaccharide and carbohydrate metabolic processes, unfolded protein binding, oxidoreductase activity, protein folding) were also found to be enriched among released proteins by a hemibiotrophic phytopathogenic fungus (Bleackley *et al.*, 2020). Regarding glycosyl hydrolases with potential antimicrobial function, *Albugo* has a significantly lower number compared to hemibiotrophic and necrotrophic oomycetes and only a small fraction of those could be identified in the predicted secretome and in the apoplastic secretome (Kemen *et al.*, 2011). It has been shown that this is a common feature of obligate biotrophic pathogens, as lytic enzymes are potentially problematic since they may result in small molecular products that are recognized and trigger defense, eventually destroying the habitat (Zhang & Zhou, 2010). Thus, selection may have resulted in adaptation of proteins for this purpose. In this vein, C06 was unique in showing hints of positive selection as analyzed in a previous study, which may indicate recent selective adaptation (Gómez-Pérez & Kemen, 2021). We therefore hypothesized antimicrobial proteins and peptides as a mechanism to defend such a fragile niche within the leaf as these might be able to evade plant recognition but nevertheless restrict the growth of other microbes.

Machine learning has been used successfully to predict *de novo* antimicrobial activity during the screening and rational design of novel anti-infectives (Plisson *et al.*, 2020). With the approach applied in this study, we found a large percentage of proteins from *Albugo*'s predicted proteome to display putative antimicrobial activity, particularly those in the apoplastic dataset. Interestingly, all predicted antimicrobial proteins that were found in the apoplast displayed a significant enrichment in IDRs (Fig. S22). This is in line with previous findings from other phytopathogens that show enrichment of intrinsic disorder in extracellular effectors for the purpose of evading detection by the plant (Marín *et al.*, 2013). We hypothesize that apoplastic localization together with the presence of long IDRs and a positive net charge are good predictors for antimicrobial proteins and peptides. In our experiments, this was illustrated by C15, which despite the negative antimicrobial prediction, had IDRs and specific antimicrobial activity.

We found that C06 showed antimicrobial activity at pH 5.9 but not at pH 7.2, where it is expected to have a much more negative net charge (Table 2). In this manner, the pH could act as an external trigger of the antimicrobial effect exclusively in the more acidic conditions of the apoplast, thus preventing unwanted effects in the cytoplasm of the hyphae. Analogously, such a targeted mechanism has been reported for human antimicrobial peptides (AMPs), which are exclusively activated when they reach their site of action. This corresponds to the surface of the skin, which is an acidic environment with a comparable pH to that of the leaf apoplast (Malik *et al.*, 2016). Once released into the apoplast, *Albugo* proteins may be cleaved by proteases of plant or microbial origin, allowing for the release of AMPs and ensuring the full antimicrobial activity is only reached in proximity to the intended bacterial targets. IDRs may facilitate this process through the prevention of a globular conformation that restricts the access of proteases. Consistent with this hypothesis, we found evidence for an enrichment of *A. thaliana* proteases in the apoplast of *A. candida*-infected plants (Table S10) and the digestion of the candidates in the presence of infected apoplast (Fig. S14).

The antimicrobial effect of the proteins was much stronger when the IDR-rich domains were tested (C6d and C14d), while the less disordered domains (C06b and C14b) had the lowest activity at the same molarity (Fig. 4b). The antimicrobial activity of cationic peptides has been ascribed to their interaction with cell envelopes which is facilitated by positively charged residues such as arginine, lysine, and histidine (Cutrona *et al.*, 2015). Of note, candidate C06d had a significant compositional bias for the presence of histidine (Table S7). As it was the case for the inhibited microbes in this study, particularly susceptible to cationic peptides are Gram-positive bacteria due to the generally larger

presence of negatively charged phosphatidylglycerol in their membrane compared to gram negatives. Thus, the variable antimicrobial effects of cationic peptides can be explained by the charge of the target membrane (Malanovic & Lohner, 2016). This mechanism could translate into contexts other than the *Albugo–Arabidopsis* pathosystem, allowing for the possibility of employing these peptides as inhibitors against strains of interest, for example, phytopathogens like *C. michiganensis* subsp. *capsici*.

When looking at the bacteria inhibited by the *Albugo* protein candidates, we found them to have a large community-shaping potential. One of the highly inhibited strains, *A. fastidiosum* I01, was responsible for a large part of the community stability as measured by the relative bacterial abundances after dropout in a synthetic community of *A. thaliana* resulting in community disruption in the presence of bacterial phytopathogen Pst (Figs 5a,d, S19). In line with observations from Agler *et al.* (2016), *Albugo* infection results in a reduced alpha diversity in the phyllosphere community. As we also saw a decrease in abundance of *A. fastidiosum* I01 in the presence of C06d after 3 d on the plant during single strain and SynCom *in planta* testing (Figs 5b,c, S18), we suggest that inhibitory properties and selective activity may also apply in the native context. Specifically releasing protein effectors that target bacteria with an influence over a large part of the community may thus be a cost-effective way for *Albugo* to amplify its phyllosphere-shaping effect while having a reduced metabolic capability. The ability of the AMP C06d to inhibit bacteria in a complex *in vitro* community (Fig. 5a) as well as *in planta*, further adds to this hypothesis.

In summary, we have found three apoplastic proteins from the plant protist pathogen *A. candida* to be antimicrobial on several plant-isolated strains belonging to Gram-positive bacteria. These proteins were selected after apoplast proteomic analysis of leaf samples and *in silico* prediction and classification of the proteins in search for candidates with antimicrobial potential. Although their specific mechanism of action remains to be elucidated, we found a correlation of the antibacterial activity with the positive net charge in the IDRs of the C-terminal domains of two of these proteins. Given the large diversity of yet unexplored obligate biotrophs, this study opens the way for novel sources for the discovery of peptide-based antibiotics.

Acknowledgements

Competing interests

None declared.

Author contributions

ORCID

Vasvi Chaudhry https://orcid.org/0000-0001-9078-9292
Daniel Gómez-Pérez https://orcid.org/0000-0002-9938-9444
Yiheng Hu https://orcid.org/0000-0002-6513-9834
Ariane Kemen https://orcid.org/0000-0003-3533-1355
Eric Kemen https://orcid.org/0000-0002-7924-116X
Boris Maček https://orcid.org/0000-0002-1206-2458
Monja Schmid https://orcid.org/0000-0002-4741-6711

Data availability

All data discussed in this paper as well as the code to reproduce the analyses and figures are found at doi: 10.5281/zenodo.6325163. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD031981 (Perez-Riverol *et al.*, 2021). All quality filtered data from the Nanopore sequencing were uploaded to NCBI short read archive under the Bioproject PRJNA942319.

References

Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, Weigel D, Kemen EM. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology* 14: e1002352.

Almario J, Mahmoudi M, Kroll S, Agler M, Placzek A, Mari A, Kemen E. 2022. The leaf microbiome of Arabidopsis displays reproducible dynamics and patterns throughout the growing season. *mBio* 13: e02825-21.

Aydogan EL, Moser G, Müller C, Kämpfer P, Glaeser SP. 2018. Long-term warming shifts the composition of bacterial communities in the phyllosphere of *Galium album* in a permanent grassland field-experiment. *Frontiers in Microbiology* 9: 144.

Berlin JD, Bowen CC. 1964. The host-parasite interface of *Albugo candida* on *Raphanus sativus*. *American Journal of Botany* 51: 445–452.

Bleackley MR, Samuel M, Garcia-Ceron D, McKenna JA, Lowe RGT, Pathan M, Zhao K, Ang C-S, Mathivanan S, Anderson MA. 2020. Extracellular vesicles from the cotton pathogen *Fusarium oxysporum* f. sp. vasinfectum induce a phytotoxic response in plants. *Frontiers in Plant Science* 10: 1610.

Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR. 2009. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols* 4: 484–494.

Bollenbach T. 2015. Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Current Opinion in Microbiology* 27: 1–9.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F *et al.* 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37: 852–857.

Bruisson S, Zufferey M, L'Haridon F, Trutmann E, Anand A, Dutartre A, Vrieze MD, Weisskopf L. 2019. Endophytes and epiphytes from the grapevine leaf microbiome as potential biocontrol agents against phytopathogens. *Frontiers in Microbiology* 10: 2726.

Burdukiewicz M, Sidorczuk K, Rafacz D, Pietluch F, Chilimoniuk J, Rödiger S, Gagat P. 2020. Proteomic screening for prediction and design of antimicrobial peptides with AMPGRAM. *International Journal of Molecular Sciences* 21: 4310.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2019. The new tree of eukaryotes. *Trends in Ecology & Evolution* 35: 43–55.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Carlos C, Fan H, Currie CR. 2018. Substrate shift reveals roles for members of bacterial consortia in degradation of plant cell wall polymers. *Frontiers in Microbiology* 9: 364.

Carlson RR, Vidaver AK. 1982. Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *International Journal of Systematic and Evolutionary Microbiology* 32: 315–326.

Chen L, Dodd IC, Theobald JC, Belimov AA, Davies WJ. 2013. The rhizobacterium *Variovorax paradoxus* 5C-2, containing ACC deaminase, promotes growth and development of *Arabidopsis thaliana* via an ethylene-dependent pathway. *Journal of Experimental Botany* 64: 1565–1573.

Cooper AJ, Latunde-Dada AO, Woods-Tör A, Lynn J, Lucas JA, Crute IR, Holub EB. 2008. Basic compatibility of *Albugo candida* in *Arabidopsis thaliana* and *Brassica juncea* causes broad-spectrum suppression of innate immunity. *Molecular Plant–Microbe Interactions* 21: 745–756.

Cox J, Mann M. 2008. MAXQUANT enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 26: 1367–1372.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. 2011. Andromeda: a peptide search engine integrated into the MAXQUANT environment. *Journal of Proteome Research* 10: 1794–1805.

Cutrona KJ, Kaufman BA, Figueroa DM, Elmore DE. 2015. Role of arginine and lysine in the antimicrobial mechanism of histone-derived antimicrobial peptides. *FEBS Letters* 589: 3915–3920.

Dhaouadi S, Mougou AH, Rhouma A. 2020. The plant pathogen *Rhodococcus fascians*. History, disease symptomatology, host range, pathogenesis and plant–pathogen interaction. *The Annals of Applied Biology* 177: 4–15.

Dubreuil B, Matalon O, Levy ED. 2019. Protein abundance biases the amino acid composition of disordered regions to minimize non-functional interactions. *Journal of Molecular Biology* 431: 4978–4992.

Eitzen K, Sengupta P, Kroll S, Kemen E, Doehlemann G. 2021. A fungal member of the *Arabidopsis thaliana* phyllosphere antagonizes *Albugo laibachii* via a GH25 lysozyme. *eLife* 10: e65306.

Figueiredo J, Santos RB, Guerra-Guimarães L, Leclercq CC, Renaut J, Malhó R, Figueiredo A. 2022. An *in-planta* comparative study of *Plasmopara viticola* proteome reveals different infection strategies towards susceptible and Rpv3-mediated resistance hosts. *Scientific Reports* 12: 20794.

Finkel OM, Salas-González I, Castrillo G, Conway JM, Law TF, Teixeira PJPL, Wilson ED, Fitzpatrick CR, Jones CD, Dangl JL. 2020. A single bacterial genus maintains root growth in a complex microbiome. *Nature* 587: 103–108.

Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E. 2011. Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications* 2: 589.

Furzer OJ, Cevik V, Fairhead S, Bailey K, Redkar A, Schudoma C, MacLean D, Holub EB, Jones JDG. 2022. An improved assembly of the *Albugo candida* Ac2V genome reveals the expansion of the "CCG" class of effectors. *Molecular Plant–Microbe Interactions* 35: 39–48.

Gómez-Pérez D, Kemen E. 2021. Predicting lifestyle from positive selection data and genome properties in oomycetes. *Pathogens* 10: 807.

Harrison PM. 2021. FLPS 2.0: rapid annotation of compositionally-biased regions in biological sequences. *PeerJ* 9: e12363.

Hassani MA, Durán P, Hacquard S. 2018. Microbial interactions within the plant holobiont. *Microbiome* 6: 58.

Hjerde E, Pierechod MM, Williamson AK, Bjerga GEK, Willassen NP, Smalås AO, Altermark B. 2013. Draft genome sequence of the actinomycete *Rhodococcus* sp. strain AW25M09, isolated from the Hadsel Fjord, Northern Norway. *Genome Announcements* 1: e00055-13.

Hu G, Katuwawala A, Wang K, Wu Z, Ghadermarzi S, Gao J, Kurgan L. 2021. FLDPNN: accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications* 12: 4438.

Jashni MK, Mehrabi R, Collemare J, Mesarich CH, de Wit PJGM. 2015. The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant–pathogen interactions. *Frontiers in Plant Science* 6: 584.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. INTERPROSCAN 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240.

Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, 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.

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M *et al.* 2018. GOATOOLS: a PYTHON library for Gene Ontology analyses. *Scientific Reports* 8: 10872.

Latendorf T, Gerstel U, Wu Z, Bartels J, Becker A, Tholey A, Schröder J-M. 2019. Cationic intrinsically disordered antimicrobial peptides (CIDAMPs) represent a new paradigm of innate defense with a potential for novel anti-infectives. *Scientific Reports* 9: 3331.

Lawrence TJ, Carper DL, Spangler MK, Carrell AA, Rush TA, Minter SJ, Weston DJ, Labbé JL. 2020. AMPEPPY 1.0: a portable and accurate antimicrobial peptide prediction tool. *Bioinformatics* 37: btaa917.

Li XS, Yuan XK. 2017. Genome sequences for multiple clavibacter strains from different subspecies. *Genome Announcements* 5: e00721-17.

Liebana-Jordan M, Brotons B, Falcon-Perez JM, Gonzalez E. 2021. Extracellular vesicles in the fungi kingdom. *International Journal of Molecular Sciences* 22: 7221.

Links MG, Holub E, Jiang RH, Sharpe AG, Hegedus D, Beynon E, Sillito D, Clarke WE, Uzuhashi S, Borhan MH. 2011. *De novo* sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics* 12: 503.

Liu L, Xu L, Jia Q, Pan R, Oelmüller R, Zhang W, Wu C. 2019. Arms race: diverse effector proteins with conserved motifs. *Plant Signaling & Behavior* 14: 1557008.

Malanovic N, Lohner K. 2016. Antimicrobial peptides targeting Gram-positive bacteria. *Pharmaceuticals* 9: 59.

Malik E, Dennison SR, Harris F, Phoenix DA. 2016. pH dependent antimicrobial peptides and proteins, their mechanisms of action and potential as therapeutic agents. *Pharmaceuticals* 9: 67.

Marín M, Uversky VN, Ott T. 2013. Intrinsic disorder in pathogen effectors: protein flexibility as an evolutionary hallmark in a molecular arms race. *Plant Cell* 25: 3153–3157.

Mayer T, Mari A, Almario J, Murillo-Roos M, Abdullah HSM, Dombrowski N, Hacquard S, Kemen EM, Agler MT. 2021. Obtaining deeper insights into microbiome diversity using a simple method to block host and nontargets in amplicon sequencing. *Molecular Ecology Resources* 21: 1952–1965.

Micsonai A, Moussong É, Murvai N, Tantos Á, Tőke O, Réfrégiers M, Wien F, Kardos J. 2022. Disordered–ordered protein binary classification by circular dichroism spectroscopy. *Frontiers in Molecular Biosciences* 9: 863141.

Molloy EM, Hertweck C. 2017. Antimicrobial discovery inspired by ecological interactions. *Current Opinion in Microbiology* 39: 121–127.

Oh E-J, Bae C, Lee H-B, Hwang IS, Lee H-I, Yea MC, Yim K-O, Lee S, Heu S, Cha J-S *et al.* 2016. *Clavibacter michiganensis* subsp. *capsici* subsp. nov., causing bacterial canker disease in pepper. *International Journal of Systematic and Evolutionary Microbiology* 66: 4065–4070.

Oldfield CJ, Uversky VN, Dunker AK, Kurgan L. 2019. *Intrinsically disordered proteins.* Cambridge, MA, USA: Academic Press, 1–34. doi: 10.1016/b978-0-12-816348-1.00001-6.

Perazzolli M, Nesler A, Giovannini O, Antonielli L, Puopolo G, Pertot I. 2020. Ecological impact of a rare sugar on grapevine phyllosphere microbial communities. *Microbiological Research* 232: 126387.

Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M *et al.* 2021. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research* 50: D543–D552.

Plisson F, Ramírez-Sánchez O, Martínez-Hernández C. 2020. Machine learning-guided discovery and design of non-hemolytic peptides. *Scientific Reports* 10: 16581.

Ruhe J, Agler MT, Placzek A, Kramer K, Finkemeier I, Kemen EM. 2016. Obligate biotroph pathogens of the genus *Albugo* are better adapted to active host defense compared to niche competitors. *Frontiers in Plant Science* 7: 820.

Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual, vol. 1, 3rd edn.* Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Savojardo C, Martelli PL, Fariselli P, Casadio R. 2015. TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. *Bioinformatics* 31: 3269–3275.

Snelders NC, Kettles GJ, Rudd JJ, Thomma BPHJ. 2018. Plant pathogen effector proteins as manipulators of host microbiomes? *Molecular Plant Pathology* 19: 257–259.

Snelders NC, Petti GC, van den Berg GCM, Seidl MF, Thomma BPHJ. 2021. An ancient antimicrobial protein co-opted by a fungal plant pathogen for *in planta* mycobiome manipulation. *Proceedings of the National Academy of Sciences, USA* 118: e2110968118.

Snelders NC, Rovenich H, Petti GC, Rocafort M, van den Berg GCM, Vorholt JA, Mesters JR, Seidl MF, Nijland R, Thomma BPHJ. 2020. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nature Plants* 6: 1365–1374.

Tackmann J, Rodrigues JFM, von Mering C. 2019. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Systems* 9: 286–296.

Teufel F, Armenteros JJA, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology* 40: 1023–1025.

Uversky VN. 2013. The alphabet of intrinsic disorder. *Intrinsic Disordered Proteins* 1: e24684.

Veltri D, Kamath U, Shehu A. 2018. Deep learning improves antimicrobial peptide recognition. *Bioinformatics* 34: 2740–2747.

Wang Y, Wang Y, Wang Y. 2020. Apoplastic proteases: powerful weapons against pathogen infection in plants. *Plant Communications* 1: 100085.

Zhang J, Zhou J-M. 2010. Plant immunity triggered by microbial molecular signatures. *Molecular Plant* 3: 783–793.

Zittlau KI, Lechado-Terradas A, Nalpas N, Geisler S, Kahle PJ, Macek B. 2021. Temporal analysis of protein ubiquitylation and phosphorylation during Parkin-dependent mitophagy. *Molecular & Cellular Proteomics* 21: 100191.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Inferred positive interactions for the *Albugo* sp. operational taxonomic unit on the *Arabidopsis thaliana* phyllosphere amplicon dataset.

Fig. S2 Overview of negative correlation network inferred by FlashWeave on the *Arabidopsis thaliana* phyllosphere amplicon dataset.

Fig. S3 Venn diagram of annotated proteins in *Albugo candida*'s predicted proteome.

Fig. S4 Scatter plot comparing the three machine learning methods for antimicrobial prediction to protein length in the proteins found in the apoplast for *Albugo candida.*

Fig. S5 Venn diagram of assigned peptides during proteomic analysis of apoplast samples.

Fig. S6 Expression and purification analysis of the protein candidates by SDS-PAGE and Coomassie blue staining.

Fig. S7 Workflow of the denaturing purification process and subsequent *in vitro* activity assays.

Fig. S8 Bacterial growth curves with candidates C05, C06, C14, and C15 mean values.

Fig. S9 Bacterial growth curves with candidates C05, C06, C14, and C15 confidence intervals.

Fig. S10 Comparison of the antimicrobial activity of C06 at pH 5.9 and 7.2.

Fig. S11 Circular dichroism spectra of purified d and b domains from protein candidates C06 and C14.

Fig. S12 Inhibition curves of C06, C14, and their domains.

Fig. S13 Inhibition curves of six different concentrations of C06d and C14d domains toward four sensitive strains.

Fig. S14 Digestion of antimicrobial protein candidates in Nc2-infected apoplastic fluid.

Fig. S15 Inhibition curves of C06d on *Clavibacter* and *Rhodococcus* strains isolated from the phyllosphere of *Arabidopsis thaliana*.

Fig. S16 Growth curves from pairwise inhibition assays with the individual strains from the five-strain bacterial community SynComB.

Fig. S17 Colony-forming unit counts per replicate of *Pseudomonas syringae* pv Tomato DC3000 (Pst) in five-strain bacterial synthetic community (SynComB) experiments in liquid culture.

Fig. S18 Change in microbial abundances during *in planta* testing of SynComB with *Albugo candida* proteins C06d and C06b on *Arabidopsis thaliana* leaves.

Fig. S19 Protective effect of a 15-strain SynCom (SynComA) with or without *Aeromicrobium fastidiosum* I01 as measured *in planta* by relative abundances of *Pseudomonas syringae* pv Tomato DC3000 (Pst).

Fig. S20 Pairwise inhibition rings of *Aeromicrobium fastidiosum* I01 against other strains (lawn) in the synthetic bacterial community SynComA.

Fig. S21 *Aeromicrobium fastidiosum* I01 lawn for interaction testing with bacterial strains in the synthetic bacterial community SynComA.

Fig. S22 Antimicrobial prediction of apoplastic and non-apoplastic proteins grouped by the presence or absence of long intrinsically disordered regions (15 or more amino acids in a row) from *Albugo candida*.

Methods S1 Detailed information on all methods used in the main text.

Table S1 Primer sequences for amplification and subsequent cloning of candidate sequences from *Albugo candida* into vector pET28b.

Table S2 Strain collection of plant-isolated microbes from environmental sampling of *Arabidopsis thaliana*.

Table S3 Plant-isolated putative pathogenic Gram-positive microbes from environmental sampling of *Arabidopsis thaliana* employed in the testing of *Albugo candida* peptide C06d.

Table S4 Primer sequences for analysis of *in planta* microbial community abundances via sequencing and quantitative PCR.

Table S5 List of proteins from *Albugo candida* found in the apoplast predicted to be antimicrobial.

Table S6 Detailed description of the functional annotation of the four candidate proteins from *Albugo candida* based on domains and orthologs.

Table S7 Significant compositional bias toward specific amino acid residues of tested protein candidates from *Albugo candida*.

Table S8 Net charge at pH 5.9 and 7.2, theoretical isoelectric point (pI) and molecular weight in kDa of the protein domains of C06 and C14.

Table S9 Peptide coverage of the C-terminal disordered domains from candidate proteins in the proteomics analysis.

Table S10 Proteases from *Arabidopsis thaliana* found in the apoplast proteomics to be enriched in the *Albugo candida*-infected treatment.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.