Protein-based interactions in microbial communities: the roles of amyloids and antimicrobials

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

M.Sc. Daniel Gómez Pérez

aus Ourense/Spanien

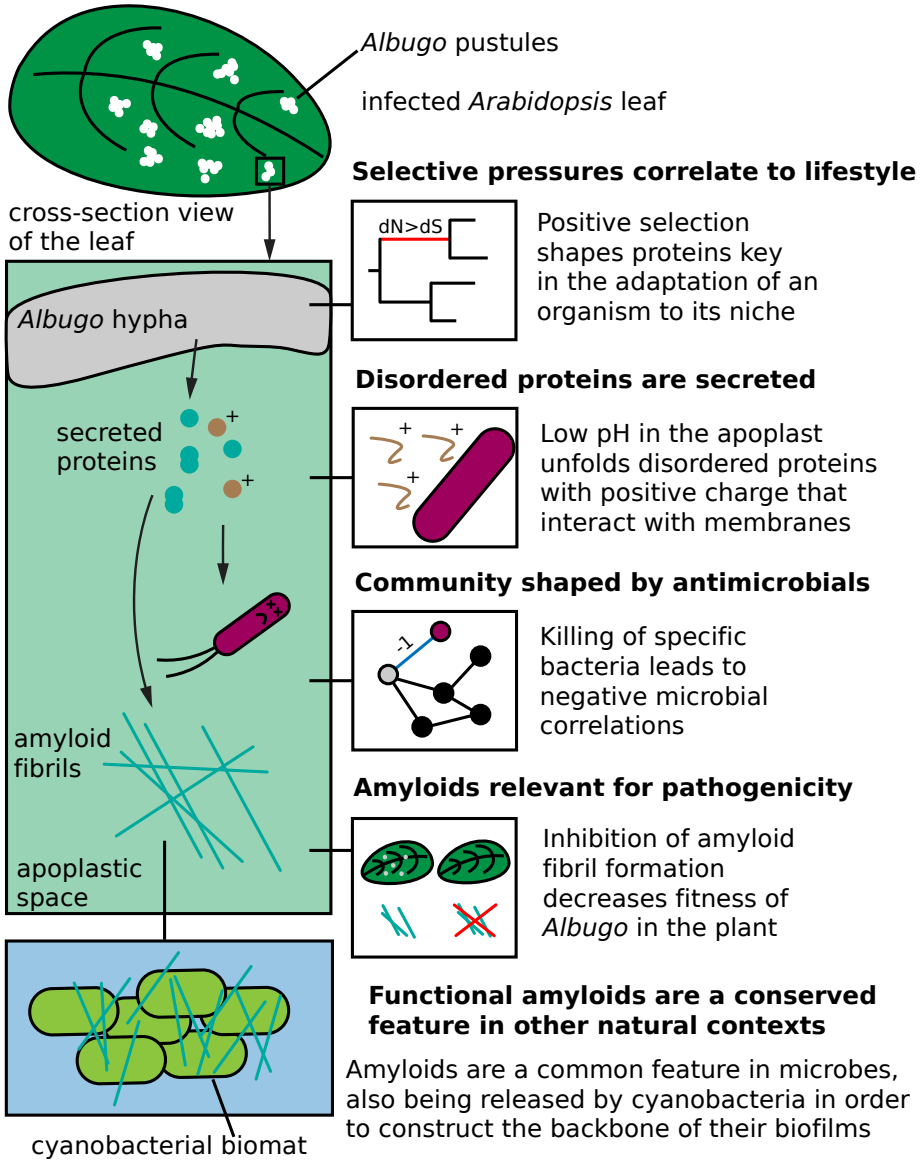
Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	22.03.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Eric Kemen
2. Berichterstatter:	Prof. Dr. Hannes Link
3. Berichterstatter:	Prof. Dr. Dominik Begerow

Graphical abstract



Abstract

Microbial communities ubiquitously inhabit the natural world. Members of these have evolved countless dynamic and complex relationships to ensure their survival. Whether through formation of protective enclosures, such as biofilms held together by amyloids, or actively killing competitors as antimicrobials, secreted proteins play key roles in microbial interactions. However, much remains to be understood about the specific mechanisms of activity behind many of these interactions.

The oomycetal pathogen *Albugo* is an obligate plant biotroph that strongly modifies its surrounding microbial community. It relies on specific proteins to exert its influence, as it has lost a large part of the biosynthetic power of free-living relatives, in part due to adaptation to an obligate biotrophic lifestyle. Firstly, we have put these proteins into an evolutionary context by studying the link between lifestyle and genome features in the oomycetes. This phylum comprises *Albugo* as well as other plant and animal pathogens with widely divergent lifestyles and hosts.

Furthermore, through a proteomics approach followed by heterologous expression, we have pinpointed as well as functionally and structurally characterized a number of proteins from *Albugo* that we found to be influential in controlling the surrounding microbial community. In particular, we have focused on those with antimicrobial potential as well as the ability to form amyloids. The former are interesting due to their direct role in antagonistic interactions and the high demand for novel and highly specific peptide-based antimicrobial compounds which could aid against the rise of multidrug resistant microbes. When studying antimicrobial proteins in *Albugo*, we could relate their effects to intrinsic disorder and high positive charge.

The amyloid fold, instead, is a prevalent and overlooked characteristic of many proteins relevant to microbial interactions and to survival in particular. Because of

their original discovery as etiological agents of human pathology, their study has been historically confined to the medical field. Based on current literature, however, the amyloid fold is now known to be omnipresent in the natural microbial world where it plays varied functional roles, including defense through antimicrobial activity. The protein candidates we have described in *Albugo* support the presence of this characteristic fold and functional relevance in protists as well, since we found amyloids to be important for pathogenicity. Finally, we have explored the amyloid-forming characteristics of proteins released by a cyanobacterium, *Synechococcus elongatus*, which are upregulated during the biofilm establishment stage.

On the whole, we have studied and described protein-based mechanisms relevant to complex microbial communities in natural ecosystems, focusing on amyloids and antimicrobials. These highlight the countless mechanisms that could be translated to biotechnological applications and the many that are yet to be discovered.

Kurzfassung

Mikrobielle Gemeinschaften sind ubiquitär vorkommende Bewohner der natürlichen Welt. Die einzelnen Mitglieder haben unzählige dynamische und komplexe Beziehungen entwickelt, um ihr eigenes Überleben in diesen Gemeinschaften und Ökosystemen sicher zu stellen. Ob durch Bildung schützender Zusammenschlüsse, wie Amyloid-gestützte Biofilme, oder durch aktive Inhibierung von konkurrierenden Organismen, sekretierte Proteine spielen eine Schlüsselrolle in mikrobiellen Interaktionen. Dennoch sind die spezifischen Mechanismen dieser Protein-basierten Interaktionen oftmals noch ungeklärt.

Albugo gehört zu den obligat biotrophen Pflanzenpathogenen aus dem Phylum der Eipilze und besitzt die Fähigkeit die mikrobielle Gemeinschaft in seiner Umgebung zu verändern. Da *Albugo* unter anderem durch Anpassungen an einen obligat biotrophischen Lebensstil einen großen Teil der biosynthetischen Aktivität, auf die freilebende verwandte Organismen angewiesen sind, aufgeben hat, beruht sein modifizierender Einfluss auf bestimmten Proteinen. Zunächst haben wir diese Proteine in einen evolutionären Kontext gesetzt, um den Zusammenhang zwischen Lebensstil und genomischen Merkmalen in Eipilzen zu untersuchen. Das Oomycota Phylum beinhaltet nicht nur *Albugo* sondern auch viele weitere Pathogene, divers in Lebensstil und Wirtsorganismus.

Mittels Proteomics gefolgt von heterologen Expression, haben wir darüber hinaus einige der Mikrobiota-modifizierenden von *Albugo* produzierten Proteine identifiziert und funktionell sowie strukturell näher bestimmt. Unser Fokus lag dabei auf den potentiell antimikrobiellen und amyloiden Eigenschaften dieser Proteine. Antimikrobielle Proteine sind interessant, da sie eine aktive Rolle in antagonistischen Interaktionen spielen können und derzeit ein großer Bedarf an neuen und spezifischen Protein-basierten antimikrobiellen Wirkstoffen besteht, die dem Aufstieg multi-resistenter Keime entgegenwirken können. Bei der Untersuchung möglicher Wirkungsweisen der Proteine von *Albugo* konnten wir einen Zusam-

menhang zwischen antimikrobieller Aktivität und intrinsischer Unordnung und positiver Ladung beobachten.

Während antimikrobielle Proteine schon lange positiv im Fokus der Wissenschaft stehen, gehört die Eigenschaft der amyloiden Faltung zu einer wichtigen dennoch unterschätzten Charakteristik vieler Proteine, die bedeutend für mikrobielle Interaktionen und relevant für das Überleben vieler Organismen ist. Da amyloide Proteine ursprünglich hauptsächlich mit menschlichen Krankheitsbildern in Verbindung gebracht wurden, ist ihre Erforschung historisch limitiert auf den medizinischen Bereich. Die aktuelle Literatur jedoch zeigt, dass amyloide Faltung omnipräsent in der mikrobiellen Welt verbreitet ist und zahlreiche funktionelle Rollen übernimmt, darunter Funktionen der Verteidugung durch antimikrobielle Wirkung. Da wir amyloide Proteine von *Albugo* als relevant für dessen Pathogenität identifizieren konnten, sprechen unsere Resultate für eine funktionelle Bedeutung dieser charakteristischen Faltung auch in dem bislang wenig untersuchten Phylum der Eipilze. Schlussendlich haben wir unsere Untersuchung der Amyloidformender Eigenschaften auf das Phylum der Cyanobakterien erweitert und Proteine von *Synechococcus elongatus* analysiert, die während der Etablierungsphase von Biofilmen vermehrt exprimiert werden.

Zusammengefasst haben wir Protein-basierte Mechanismen, insbesondere jene welche antimikrobielle und Amyloid-formende Proteine involvieren, untersucht und beschrieben. Diese Mechanismen sind funktionell relevant für komplexe mikrobielle Gemeinschaften in natürlichen Ökosystemen und betonen das Potential für mögliche biotechnologische Anwendungen und deuten zudem auf die vielen Mechanismen hin, die es noch zu entdecken gibt.

Acknowledgments

First, I would like to thank MS for her support, in the lab as well as outside of it. Without your help and motivation this could not have been possible. I also would like to thank my lab partners, VC in particular for always being willing to help, as well as the others, JA, YH, EK, MK, KL, AM, MM, VM, BÖ and PR.

Additionally, I am grateful to people involved one way or another with my work, which includes HBO, KF, BM, CG, SR, RS, YS and AV. A special mention goes to LLP for her interest on my research and her lightning-speed help on my manuscripts. I would also like to mention JR, who started working on this project back when the group was in Cologne.

Other people I would like to acknowledge include SM for helping me move to Tübingen; and TH, NHP, YMH, LV, KW and my family for their unrelenting support over the years. Finally, I would like to thank my supervisors AK and EK, for their thorough feedback on the thesis and extensive discussions during my PhD time.

I would like to acknowledge the support of my funding sources. These included the German Research Foundation (DFG) graduate school "Molecular principles of bacterial survival strategies", the European Research Council under the DeCoCt research program, the Open Access Publishing Fund of the University of Tübingen, as well as the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen.

The use of *we* instead of *I* is for the preservation of consistency throughout the dissertation. Also, it emphasizes the team effort that went into the last four and a half years. My individual contributions to the manuscripts that constitute this thesis are listed in the Appendices.

Contents

Abbreviations

aBSREL	adaptive Branch Site Random Effects Likelihood
A. candida	*Albugo candida*
A. laibachii	*Albugo laibachii*
AMP	Antimicrobial peptide
A. thaliana	*Arabidopsis thaliana*
AUC	Area under the curve
BeStSel	Beta Structure Selection
BSA	Bovine serum albumin
BUSCO	Benchmarking Universal Single-Copy Orthologous Groups
CAZyme	Carbohydrate-active enzyme
CD	Circular dichroism
C-DAG	Curli-dependent amyloid generator
CFU	Colony forming units
CR	Congo red
csg	curli-specific gene
$csgA_{ss}$	csgA secretion signal
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EbfG	Enabling biofilm formation with a GG motif
E. coli	*Escherichia coli*
FUBAR	Fast, Unconstrained Bayesian Approximation for Inferring Selection
GO	Gene Ontology
GUPS	Genes under positive selection
HGT	Horizontal gene transfer
IDR	Intrinsically disordered region

IPTG	Isopropyl β- d-1-thiogalactopyranoside
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography and tandem mass spectrometry
MAMP	Microbial-associated molecular pattern
NB	Nutrient broth
NCBI	National Center for Biotechnology Information
Ni-NTA	Nickel-nitrilotriacetic acid
OD	Optical density
OTU	Operational taxonomic unit
pI	Isoelectric point
plDDT	probability of local Distance Difference Test
Pst	*Pseudomonas syringae* strain DC3000
P. insidiosum	*Pythium insidiosum*
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
S. elongatus	*Synechococcus elongatus*
SynCom	Synthetic community
TEM	Transmission electron microscopy
ThT	Thioflavin T
UPGMA	Unweighted pair group method with arithmetic mean
WT	Wild type

Chapter 1

Introduction

1.1 Oomycete evolution

The study of evolution by natural selection helps make sense of biology, as all organisms have a shared history (Darwin, 1860; Dobzhansky, 1973). With the advent of mathematical theories of evolution and sequencing technologies in the 20th century, it has been possible to study adaptation of organisms from the point of view of their genomic material (Mendel, 1866; Fisher, 1919; Haldane, 1927; Sanger *et al.*, 1977). This, together with novel mechanisms of evolution discovered in microorganisms in the last decades, including horizontal gene transfer (HGT), has led to a more exhaustive picture of how evolution takes place (Moreno-Santillán and Ortega, 2021). However, many microorganisms remain to be thoroughly studied in this regard, particularly long overlooked protists such as the oomycetes, which until recently were still wrongly assigned to the fungal kingdom (Lévesque, 2011).

The phylum of oomycetes consists of microbial heterokonts from the Stramenopile clade (Beakes and Thines, 2017). They are eukaryotes with close phylogenetic connections to marine autotrophic organisms, brown algae and diatoms. About 400 million years ago, some of them left the oceans and evolved to have a wide range of parasitic lifestyles on land (Selosse *et al.*, 2015). Currently, they are known to infect animals, plants, fungi and even other oomycetes, causing economically relevant crop and veterinary diseases (Bebber and Gurr, 2015; Mendoza and Vilela, 2013). At multiple points in their evolution, oomycetes developed the ability to exclusively follow a free-living saprobe lifestyle (Misner *et al.*, 2015). Despite the large phylogenetic gap, oomycetes share morphological similarity with fungi due to convergent evolution, as they tend to occupy similar niches (Andersson, 2006).

However, there are major fundamental differences between oomycetes and fungi. For instance, they have a distinct cell wall composition, i.e., presence of cellulose in the former and chitin in the latter, as well as a different ploidy for the majority of their lifecycle, diploid in oomycetes and haploid in fungi (Latijnhouwers *et al.*, 2003).

As hinted above, there is a large diversity of lifestyles in the oomycetes phylum. Particular oomycetes, including members of the Albuginaceae and the downy mildews from the Peronosporaceae family, have adapted to an obligate biotrophic lifestyle (Ruhe *et al.*, 2016). Obligate biotrophs are exclusively bound to survival on a living plant, usually a narrow set of species within a plant family. Therefore, they cannot be propagated *in vitro*, which complicates their study. Additionally, there are other oomycetes that present a plant hemibiotrophic or a necrotrophic lifestyle, in either plant or animal hosts. The first case includes most members of the *Phytophthora* genus from the Peronosporaceae, and it consists of a biotrophic phase followed by a necrotrophic phase, in which the host tissues are attacked (Zuluaga *et al.*, 2016). The second category includes the necrotrophs, which which immediately kill their host after infection (Latijnhouwers *et al.*, 2003). These parasites often have facultative characteristics that enable them to live on their own from decomposing matter as saprobes. To this category belong all Saprolegniaceae and Pythiaceae as well as certain non-*Phytophthora* Peronosporaceae. This complex evolutionary history has contributed to tight intertwining of metabolism and lifestyle in the oomycete lineage (Rodenburg *et al.*, 2021).

1.1.1 Study of selective pressures from the genome

The impact on the genome of organisms' adaptation to different lifestyles and hosts has long been an important subject of study in evolutionary biology. As more and more genetic data are stored in publicly available databases, comparative genomic studies are becoming an essential asset. Several ways to measure selective pressures from the genome have been described (Booker *et al.*, 2017). One of the most widely employed is based on the determination of the rates of nucleotide changes by comparison of ortholog sequences in closely related organisms. In this case, the rates of nucleotide changes that alter the protein sequence (non-synonymous; dN) are compared to those that cause no amino acid change in the coding sequence

(synonymous; dS) such that

$$\omega = \frac{dN}{dS}.$$

As this ratio, ω, is expected to be ~ 1 when there are no selective pressures acting on the genome, $\omega > 1$ signifies positive selective pressure for this region. This means that nucleotides in a particular gene sequence have been subject to directional selection. In contrast, $\omega < 1$ is a sign of negative selective pressure. In this case, a higher rate of synonymous changes keeps the protein sequence from altering through genetic drift (Kimura, 1977). The pitfalls of studies using this method lie in the large number of false positives due to the low statistical power of its basic usage. To address this problem, *HyPhy*, a collection of software that implements this methodology within maximum likelihood and Bayesian statistical frameworks was developed (Pond *et al.*, 2005). We employed several packages from this collection to assess the selective pressures in oomycetes, which we describe in Section 2.1.2.

1.2 Microbial phyllosphere interactions

Plants make up the far majority of Earth's biomass and thus provide the largest living surface where microbes can settle and thrive on (Bar-On *et al.*, 2018). Because of their ubiquity, plants represent a relevant system to study host-microbe as well as microbe-microbe interactions. The phyllosphere, meaning the parts of the plant above ground as habitat for microbes, is a battleground for colonization and survival (Vorholt, 2012). Microbes that are able to live from the resources of the plant are highly adapted to this niche and compete with others to defend it. Some of these microbes, for example, the protist and oomycete *Albugo*, follow a strategy known as niche construction, whereby they reshape the existing microbial community of the plant at their own benefit (Mukhtar *et al.*, 2011). This results in such microbes being highly interconnected with others, both in a positive and a negative fashion (Agler *et al.*, 2016). Hub microbes, as they are called, may promote partners that are beneficial to them or inhibit others that pose a threat, either because of direct competition or indirectly because they compromise their niche. For example, a pathogen with a long-term biotrophic phase like *Albugo*, may be interested in the survival of the plant host and thus inhibit pathogens that

are threatening it.

The particular mechanisms that are the basis of these interactions have long been a focus of research in biology, particularly those relating to host interactions (Braga *et al.*, 2016). However, those that occur among the microbes in complex communities remain mostly unexplored. On one hand, a myriad positive interactions take place at any point in the phyllosphere. Some examples include construction of multispecies biofilms that protect several microbes other than the producing strain (Burmølle *et al.*, 2006). Or, microbes that seemingly selflessly release metabolites or *public goods* that are beneficial for other species as well (Levin, 2014). This presents a paradox where cheaters are prone to appear and take advantage of the situation (Friesen, 2020). However, modeling has shown that microbes following selfless strategies can achieve consistently higher fitness in complex bacterial communities (Smith and Schuster, 2019). This is due to division of labor, which in the long term results in higher prevalence, where microbes that share outcompete those that do not.

On the other hand, negative interactions are also abundant in shaping microbial communities. Examples include competition for particular metabolites and release of antimicrobial compounds, either general-acting or specific against certain competitor strains (Cycoń *et al.*, 2019; Machado *et al.*, 2021). The study of ecologically relevant microbe-microbe interactions involving inhibition could have impactful applications in, e.g., medicine as anti-infectives or in agriculture as biocontrol agents (Molloy and Hertweck, 2017). Therefore, it is of general interest to better predict them and understand their mechanisms of action. The increasing problem of multidrug-resistant microbes could be alleviated through discovery of new antimicrobial mechanisms (O'Connell *et al.*, 2013).

On top of this, the plant host also has a competing interest in shaping the composition and diversity of its associated microbiome. Microbes that are beneficial for their host in light of, e.g., growth promoting and/or protective effects are actively promoted and rewarded by the plant (Terauchi *et al.*, 2017). These interactions commonly happen through symbiotic relationships that profit both parties (Hayat *et al.*, 2010). Exemplary are nodule-forming rhizobia bacteria that fix atmospheric nitrogen while obtaining carbohydrates from the plant in return (Masson-Boivin and Sachs, 2018; Wang *et al.*, 2018). In contrast, detrimental microbes, such as disease-causing pathogens, are actively repressed by the plant. These

pathogen-prone microbes are recognized through the plant's sophisticated repertoire of pattern- and receptor-triggered defenses, followed by prompt suppression with oxidative bursts and lytic enzymes (Nishad *et al.*, 2020).

Two phyllosphere domains where these interactions can take place can be delineated: the epiphytic, which comprises the external surface of the stem and leaves, and the endophytic, which consists of the internal compartments of the plant (Whipps *et al.*, 2008). This is namely the apoplast, an extracellular chamber filled with liquid, where the endophytic microbes reside and thrive (Wang and Wang, 2018). The apoplast is slightly acidic (pH 5–7) but its pH can vary widely depending on different biotic or abiotic stresses (Geilfus, 2017). Additionally, it is plagued with proteases released by the plant as a defense response towards pathogens, which makes it a challenging growth environment for most microbes (Wang *et al.*, 2020; Jashni *et al.*, 2015). As the epiphytic region is more exposed to the environment and thus to abiotic factors, epiphytes are frequently more diverse and there is a higher turnover. Meanwhile, the endophytic community is more stable and less perturbed by abiotic factors (Bacon and White, 2016). Thus, lacking most of the exposure to these abiotic factors, endophytes represent an ideal model to study co-occurring microbe-microbe and host-microbe interactions that are taking place.

1.3 Protein-based interactions shape the phyllosphere

The mechanistic bases behind these interactions have undergone adaptation over millions of years of co-evolution and are thus linked to the specific evolutionary history of the organism. Some plant pathogens, including oomycetes and fungi, have undergone evolution towards a reduced genome due to adaptation and dependence on a particular host (Judelson, 2012). Reliance on the plant for nutrients has caused the microbes to confide mostly on salvage mechanisms (Yang *et al.*, 2020). The absence of many secondary metabolic pathways means that they lack the machinery to produce *de novo* metabolites to interact with the environment. This is the case for *Albugo*, where all the secondary metabolism, as well as a large part of the primary metabolism have been lost (Kemen *et al.*, 2011). As a hub microbe, *Albugo* has been reported to exert a large influence on the composition of its surrounding microbial community (Agler *et al.*, 2016). Some of these mecha-

nisms can be explained by the immune modulatory effects of *Albugo* on the host, which in turn modify the community (Ruhe *et al.*, 2016). However, there may be direct, protein-based microbial mechanisms not yet explored as evidenced in Eitzen *et al.* (2021), where a novel albugicidal glycoside hydrolase secreted by an apoplastic yeast was described.

Albugo as well as many other plant pathogens have a large arsenal of proteins called effectors that are released into the apoplast or directly into the plant cytoplasm to mediate a wide range of host-microbe interactions. These include facilitation of host entry and host immune evasion and suppression, among other functions (Wawra *et al.*, 2012; Toruño *et al.*, 2015). In recent years, effectors have been found to be relevant for functions beyond the interaction with the host. For example, one effector protein from a hemibiotrophic fungus, which was initially studied in the context of immune suppression, was found to exert a microbiome modifying effect (Snelders *et al.*, 2020). Based on this body of evidence, we hypothesize that effector proteins in *Albugo* pose an unexplored resource for the discovery of novel antimicrobial mechanisms that could be beneficial for industrial and medical applications.

1.3.1 Proteins as antimicrobial agents

The origins of a large part of the antibiotics in use today can be traced back to metabolites that were isolated from complex microbial communities, mainly located in soil. These metabolites are naturally secreted by microbes in order to safeguard the producer's niche (Chandra and Kumar, 2017). However, the increase in the prevalence of multi-drug resistant bacteria has led to a search for new mechanisms of antimicrobial activity, such as antimicrobial peptides (AMPs). AMPs are usually small peptides with a particular set of physical and structural properties that result in their growth-inhibiting effect on microbes. AMPs do not share a common mechanism of action with most known metabolite-based antibiotics. Their toxicity is rather a result of a compound effect of membrane disruption that ends up overwhelming the target microbe (Raheem and Straus, 2019). This can happen through binding and solubilization of the membrane or through formation of pore-like structures which permeabilize the membrane (Lee *et al.*, 2015). Such an unspecific set of mechanisms makes development of resistances less likely; how-

ever, some bacteria have been shown to overcome the inhibitory effects of AMPs to a certain extent (Shazely *et al.*, 2020).

Many novel AMPs have been characterized in recent years, particularly originating from humans and other higher eukaryotes, which use them as a first line of defense against infection (Rollins-Smith *et al.*, 2005; Hancock and Scott, 2000; Meister *et al.*, 1997). For example, fragments from the hornerin antimicrobial protein are released into the skin where they are cleaved into peptide fragments and activated due to the acidic pH (Gerstel *et al.*, 2018). Analogously, one could imagine such a mechanism where proteins released by the plant or endophytic microbes are cleaved and activated in the acidic pH of the apoplast in order to control the growth of pathogenic or competing microbes (Chandrasekaran *et al.*, 2016). Antimicrobial peptides that are released from plant cells into the apoplast have been described, however the evidence for eukaryotic pathogens releasing such proteins as a niche-shaping tool is scant (Tam *et al.*, 2015).

1.3.2 Intrinsically disordered proteins

Cationic intrinsically disordered peptides have been proposed to be relevant in antimicrobial interactions (Gerstel *et al.*, 2018). These contain intrinsically disordered regions (IDRs), which are protein domains that lack a stable structural conformation (Oldfield *et al.*, 2019). This is due to compositional bias towards an unusually increased presence of disorder-promoting amino acids (Theillet *et al.*, 2013). Outside domain-linking activity, IDRs can perform functions that are variable depending on their context (Babu *et al.*, 2011). Meaning, the same protein under different circumstances (pH, ligand, etc.) can have a variable functional role encoded within an identical amino acid sequence (Ittisoponpisan *et al.*, 2017). This is known as pleiotropy and is prevalent in the genome of many pathogens. Adaptive changes in pleiotropic proteins have a higher impact on fitness within shorter evolutionary time spans than those in proteins with a unique function, while helping maintain a reduced genome (Østman *et al.*, 2012).

IDR-containing proteins have been proposed to form an integral part of the effector arsenal from plant pathogens, including bacteria and fungi (Shen *et al.*, 2017; Liu *et al.*, 2019). Long IDRs may provide flexibility to the proteins, allowing for an easier translocation during secretion, immune evasion, and facilitation of the

binding to plant receptors, even without a perfect fit (Marín *et al.*, 2013). This last function is crucial to keep up with the pathogen-host co-evolution, as parasites unable to adapt to small changes in plant receptors may fall behind in the evolutionary race (Anderson *et al.*, 2010).

IDRs are biased towards a higher localized abundance of basic or acidic residues. This means that these regions have in general a high net charge when at a pH close to neutral, and that consequently, the residues within IDRs usually bear a charge (Theillet *et al.*, 2013). When the sequence is more abundant in basic residues, namely histidine, lysine or arginine, the IDRs carry a net positive charge at physiological pH. Peptide fragments of the IDR-containing proteins, when released by proteases, are able to form cationic AMPs, which have been associated with antimicrobial activity in the human skin (Latendorf *et al.*, 2019).

1.3.3 Functional amyloid proteins

Another structural pattern that has been reported in connection with IDRs is amyloid formation (Pan and Zhong, 2015; Gadhave and Giri, 2020; Korsak and Kozyreva, 2015). Amyloid proteins are being studied more and more extensively in their physiological and functional contexts (Appendix A.1 Figure 2), as compared to their initially established role as etiological agents of neurodegenerative disease (Balistreri *et al.*, 2020). These contexts include natural microbial communities such as the phyllosphere (Gómez-Pérez *et al.*, 2021). Functional amyloids, as they are commonly referred to in the literature, have a set of characteristics resulting from the formation of extremely stable and heterogeneous fibrils (Otzen and Riek, 2019). These fibrils are composed of β-sheet monomers that stack on top of each other resulting in a cross-β structure, which is highly stable due to the so-called steric zipper conformation. This consists of two separate phases formed respectively by hydrophobic and hydrophilic residues which weave cross-β sheets together through interdigitation (Sawaya *et al.*, 2007). Some amyloid proteins are highly prone to fibrillate spontaneously, these are said to have high amyloidogenicity and can act as nucleators, driving the amyloid formation of other proteins (Hammer *et al.*, 2007).

In the plant phyllosphere, amyloids are known to partake in a variety of functions related to survival and pathogenicity (Gómez-Pérez *et al.*, 2021). These in-

clude toxicity, signaling, biofilm and hydrophobic structure formation, both from the plant and the microbial side (Appendix A.1 Figure 1). The reason these proteins are able to excel at such a wide array of functions lies on several common physical properties of the amyloid fold. These include first, the templatability of cross-β monomers which despite variations in the sequence are able to grow through a nucleation/seeding mechanism (Xue *et al.*, 2008). And second, after fibrillation of the monomers, the mature fibrils are resilient to external conditions, including chemical digestion with proteases or denaturation through changes in temperature or mechanical stress (Choi *et al.*, 2014).

Traditionally, amyloid structures were identified by X-ray diffraction, which, when performed on mature fibrils, produces a characteristic pattern at 4.7 and 10 Å (Sunde and Blake, 1997). More recently, amyloid proteins have been identified by staining with dyes or by direct observation of the filaments using electron microscopy (Wu *et al.*, 2007b; Gras *et al.*, 2011). Congo Red (CR) and Thioflavin T (ThT) are commonly employed compounds to assess formation of amyloids. Both consist of freely rotating rings that become fixed when bound to amyloid fibrils. Thus, they shift their electromagnetic emission wavelength when transitioning from a soluble to an amyloid-bound state (Wu *et al.*, 2007a; Faverie *et al.*, 2014). Due to the unspecific binding of these dyes, more than one method is usually employed when reporting novel amyloid proteins.

As with the IDRs, of note is the recently reported relevance of small oligomeric amyloids as antimicrobial agents. The mechanism behind the antimicrobial activity of amyloids is not completely understood but the overlap with many known AMPs is striking (Kagan *et al.*, 2012). In fact, several described AMPs have been found to present an amyloid structure long after their discovery (Gour *et al.*, 2016). In this case, the long and stable amyloid fibrils act as a storage or reservoir containing the inert toxic agents, which correspond to monomers or short oligomers (Salinas *et al.*, 2021). These are activated and released only under certain conditions. A change in their location or environment, for example, a lower pH, can trigger the release. This also frequently results in a structural shift from β-sheet to an active α-helix-rich structure (Salinas *et al.*, 2021; Mandala *et al.*, 2016).

Amyloid proteins are an integral part of biofilms

One of the best characterized roles of amyloids, in part due to their importance for virulence, is in serving as backbone of biofilms originating from diverse bacterial species (Taglialegna *et al.*, 2016). The amyloid proteins create a strong mesh where the bacteria can settle and are protected from external biotic and abiotic pressures. Many biofilm-enabling amyloids have been described mainly in bacteria, where they are relevant for pathogenicity. However, many others originating from non-pathogens and found in environmental contexts remain largely understudied. In aquatic systems, biofilms are associated with biofouling, where layers of biological material accumulate in microbial mats, causing severe economic losses. However, recently, microalgal and cyanobacteria biofilms, which share a similar evolutionary ancestry of phototrophy in aquatic systems with oomycetes, have found uses in industry (Barros *et al.*, 2020).

One of the cyanobacteria that currently has wide-ranging biotechnological applications, including in biofuel and protein production, is *Synechococcus elongatus* (Azevedo *et al.*, 2019; Pathania and Srivastava, 2021). This autotrophic cyanobacterium is able to form biofilms, although this process is suppressed through a negative feedback loop in the wild type (Schatz *et al.*, 2013). When activated in a knockout strain that has the ability to construct biofilms, it abandons its usual planktonic growth and establishes itself in static communities. Currently, little is known about the mechanisms governing biofilm formation and maintenance in cyanobacteria. In Appendix B.2, we propose a role in biofilm formation for the putative amyloid proteins encoded in the *ebfG*-operon. EbfG stands for enabling biofilm formation with a GG motif, and the operon consists of four short proteins which are highly upregulated during the biofilm stage of *S. elongatus*.

1.4 Objectives and expected outcome of doctoral research

The main goal of my PhD was to better understand protein interactions among microbes against the backdrop of natural microbial communities. Foremost, investigating the pathosystem *Albugo-Arabidopsis* but also other organisms relevant for their applicability in biotechnology, such as cyanobacteria.

Firstly, the objective was to study proteins from *Albugo* in the context of its phylogenetic lineage, the oomycetes. Since this clade presents a large variety of lifestyles, we set out to study the link between lifestyle and the coding sequences of proteins as found in the genome. Effectively doing this could lead to the development of tools to predict lifestyle from genetic information, or vice versa. How genetic information relates to phenotype is a fundamental question in biology and is relevant for a myriad of applications (Orgogozo *et al.*, 2015). A better understanding of the lesser studied protist group of oomycetes could offer insights into this question.

One of the most studied categories of proteins related to microbe-microbe interactions due to their potential applications are those harboring antimicrobial activity. From previous studies, we know that *Albugo* is negatively correlated to many other microbes (Agler *et al.*, 2016). Using amplicon sequencing data of wild *Arabidopsis* populations and large libraries of microbes collected and made available to us by previous members of our group, we aimed to establish *in silico* and wet lab methods to study *Albugo* and its molecular mechanisms of inhibition. We expected to find antimicrobial proteins that would be active *in vitro*. In the long-term, we intended to study their activity also *in vivo* to assess their use as a tool to control specific microbes.

The overlap of antimicrobial activity with the amyloid fold is well reported for many proteins in the literature (Kagan *et al.*, 2012). Amyloids, apart from antimicrobial activity, are also known to have other microbe-microbe functions such as biofilm formation and signaling (Otzen and Riek, 2019). Using specific dyes and electron microscopy imaging techniques, we set out to identify some of the amyloid candidates in *Albugo*. Based on the reported prevalence of amyloids in survival and increased virulence, we also attempted to find out more about their contribution to *Albugo*'s infection (Gerven *et al.*, 2018).

As mentioned earlier, one of our concerns was the application of these amyloid

proteins in a biotechnological context. We investigated therefore potential amyloids from the cyanobacterium *Synechococcus elongatus*. Evidence for small proteins that are required and highly upregulated during biofilm formation, where they form part of the matrix, has been previously reported (Parnasa *et al.*, 2016). Our goal was to provide evidence in support of the relevance of the assemblage of these proteins into amyloids to the formation of biofilms, using the methods which we have established for *Albugo*.

All in all, with the work in this thesis we aimed to be a step closer to better understanding interactions involving proteins in natural microbial communities. This knowledge would be crucial in order to control them or apply them in other contexts.

Chapter 2

Results and discussion

2.1 Oomycete evolution

Proteins and their interactions underlie all the activities in biological systems. The study of protein evolution is thus key to put life science into perspective. In oomycetes, particularly, evolution is crucial to understand not only their repertoire of secreted proteins and how these have adapted, but also to discern how they relate to specific lifestyles. The latter could be useful in the study of, for example, emerging pathogens. Appearance of new pathogens through, for example, host jumps is a threat to humanity and crops worldwide (Bonneaud and Longdon, 2020; Thines, 2019). Therefore, a better grasp of evolutionary mechanisms in oomycetes could aid in prophylaxis. In this section, we describe a novel comparative genomic approach to the study of oomycete evolution through a combination of analyses of positive selective pressures and presence or absence of genome properties.

Initially, we collected a dataset of oomycete genome assemblies together with their predicted proteomes from the NCBI assembly database. We assessed their completeness and annotation quality as estimated by their BUSCO scores. BUSCO gives a measure of the prevalence of conserved orthologs, either single or duplicated, from phylogenetically close organisms (Seppey *et al.*, 2019). In public databases, such as NCBI, there is an increasing amount of sequencing data available, in part, due to the decreasing cost of next generation sequencing in recent years (Goodwin *et al.*, 2016). However, the data in such databases have disparate properties and origins, including the processing of samples and sequencing. For example, data may vary in the sequencing technology used to obtain the reads, which has a big impact on the final length of the reads, and therefore the overall

quality of the assembly (Pollard *et al.,* 2018). In oomycetes, this is particularly important due to the presence of long repetitive regions which are hard to assemble *de novo* with short reads (Judelson, 2012).

To test whether the latter would play a role in the presence or absence of annotations for selection analyses, we performed a BUSCO search using the stramenopile dataset and compared the scores. When looking at oomycete assemblies, we found no correlation between genome completeness and sequencing technology in the large majority of the species (Figure 2.1). For species that have been sequenced with both long and short read sequencing methods, the number of ortholog matches was comparable, despite a larger presence of duplicated orthologs in longer reads in some species, which does not perturb the presence/absence analysis. The only exception to this general trend was *Plasmopara viticola,* where a significantly lower number of ortholog matches was found for the short read assembly. Therefore, we selected and directly compared genomes independently of whether they originated from long or short reads with the goal to have a larger database (Appendix A.2 Figure S4).

In total, we chose 34 representative genome assemblies from oomycetes with a wide background of lifestyles among those that had an annotated proteome. Additionally, we selected eight further genomes from the closely related diatoms and brown algae as an outgroup for phylogenetic rooting and as a control group for downstream analyses (Appendix A.2 Table 1). The pipeline for these analyses was built in the *snakemake* framework and was made available at `https://github.com/danielzmbp/wsgups` (Köster and Rahmann, 2012). In the following sections, we describe our findings, including the relevance of positive selection for the prediction of lifestyle and its functional enrichment in oomycetes.

2.1.1 Metabolism follows lifestyle

In order to extract information from each of the genomes, we functionally annotated the predicted proteomes and classified their features using the genome-properties database (Richardson *et al.,* 2018). Clustering the taxa in the oomycete dataset by the presence or absence of these core annotations using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method, we found that similar lifestyles mostly grouped together, independently of the phylogenetic back-

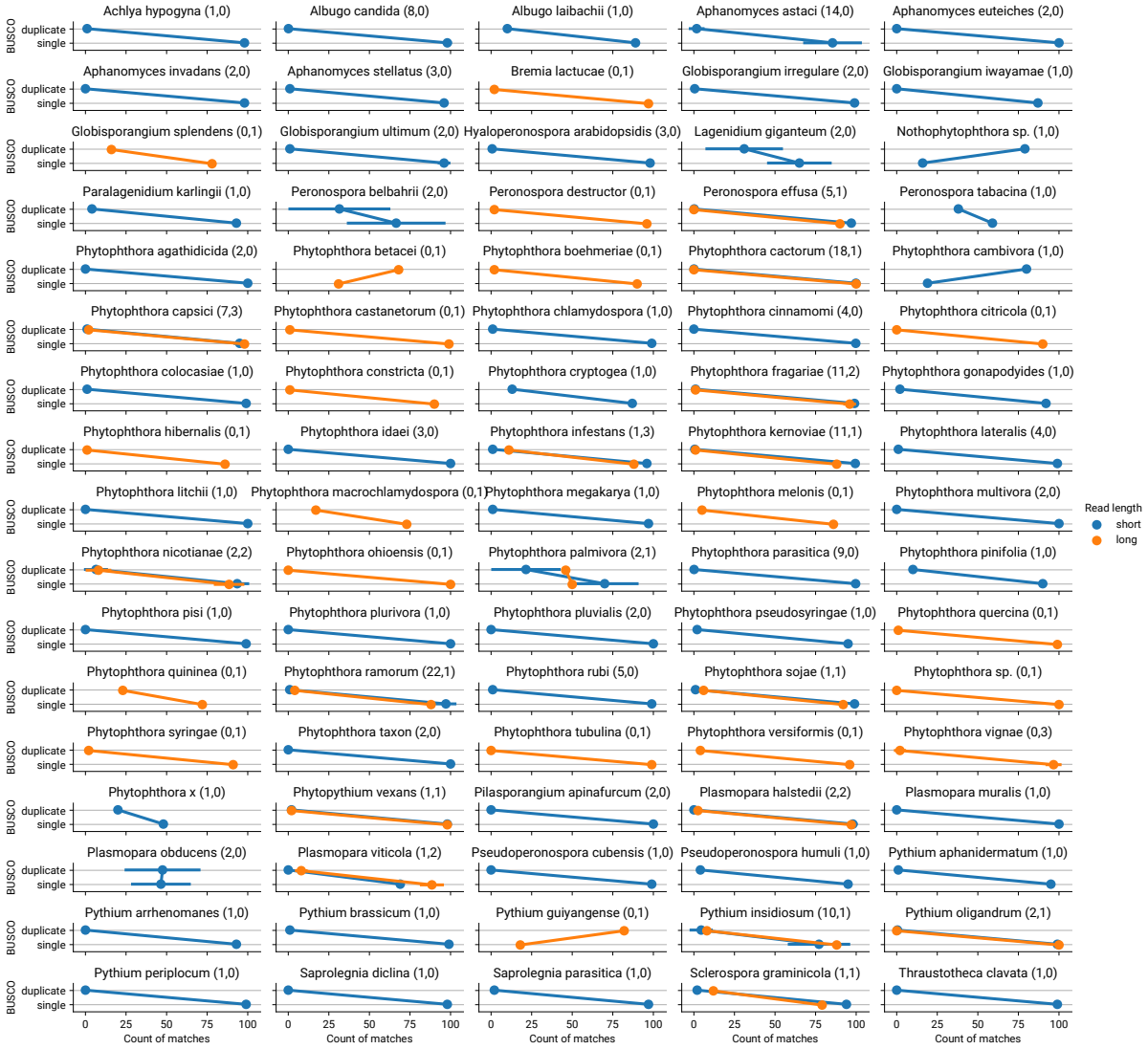


Figure 2.1: Single and duplicated number of ortholog matches for oomycete species available in NCBI using the BUSCO stramenopile dataset as of June 2021 (100 orthologs). The read sequencing length technology used to generate the assemblies is shown as the color of the dots and connecting lines. The numbers of genomes for each species are in the titles in parentheses as follows: number of short read genomes, number of long read genomes. The points show the mean for all genomes in the same category and horizontal bars the standard deviation.

ground. This was more evident for the obligate biotrophs, including downy mildews and Albuginaceae, but also to a lesser extent on the hemibiotrophs and necrotrophs of the dataset (Appendix A.2 Figure 1). The annotations relating to metabolite biosynthesis had the biggest impact on the clustering, followed by catabolism and other metabolic functions. In Rodenburg *et al.* (2020), a similar clustering was found in a smaller subset of pathway annotations and a comparable number of oomycetes. In this case, they used the percentage of pathway coverage, rather than presence or absence of enzymes.

Having similar lifestyles and inhabiting the same hosts requires similar strategies for survival, which are shaped by the same selective pressures. As comparable niches mold distant organisms through convergent evolution, their genomes grow alike from a functional point of view, even when they are phylogenetically distant. On a shorter timescale, HGT from coinhabiting microbes has been known to accelerate adaptation (Harrison and Brockhurst, 2012). Particularly, when an organism develops a strategy that confers a large advantage in fitness, which quickly spreads through the population (Mehrabi *et al.*, 2011). Thus, the pool of genes shared between similar-niche dwellers also increases over time, irrespectively of phylogeny. Exceptions to a convergence in genome properties were observed in some cases, such as for *Pythium insidiosum*. *P. insidiosum*, the only animal pathogen in the Pythiaceae and in our analyses, did not cluster with other animal pathogens. This may account for the different phylogenetic background or, more generally, nuances in lifestyle that do not fit the loosely defined boundaries between lifestyle categories. For example, many hemibiotrophs were thought to be necrotrophs in the past due to a partial understanding of their life cycle (Rajarammohan, 2021).

2.1.2 Positive selection pressures

For the study of selection pressures in the genome, we initially classified all proteins into orthogroups using a combined homology and synteny approach as implemented in the Proteinortho package (Lechner *et al.*, 2011). We considered protein families that had five or more members for downstream analysis in order to get a balance between high coverage of the dataset and a strong statistical power. The final number of analyzed proteins corresponded to about half of all proteins from the 42 genomes in the dataset and grouped into a majority (78.7%) of one-

to-one orthogroups (Appendix A.2 Figure 2). Therefore, we concluded that this dataset provided a good representation of the taxa analyzed.

Two packages from the *HyPhy* collection were used for the assessment of the selective pressures. The first one, "Fast, Unconstrained Bayesian Approximation for Inferring Selection" (FUBAR) was used to find codons in the alignment that had hints of positive selection (Murrell *et al.*, 2013). Because of its speed, it was used to perform an initial screen on all protein families. The positive families were tested with the second package, "adaptive Branch-Site Random Effects Likelihood" (aBSREL), pinpointing which of the branches in the phylogenetic tree were under positive selection (Smith *et al.*, 2015). Together, these methods were applied to infer the selective pressures in the clade of oomycetes. In summary, we found 8052 proteins from 6069 families, which out of a total of 29,122 makes 20.8% that had, according to FUBAR, at least a codon in their sequence under selection. Of these, 4291 (14.7%) protein families had at least a common ancestor node that resulted positive according to the aBSREL analysis. We considered the genes downstream of these significant nodes as genes under positive selection (GUPS). We implemented this as a standalone pipeline for analysis in *snakemake* available at `https://github.com/danielzmbp/wsgups` (Köster and Rahmann, 2012).

In the literature, the numbers reported for the prevalence of GUPS in the genome is highly variable, ranging from 1.9% to 44% of the tested protein families (Hill *et al.*, 2022; Cicconardi *et al.*, 2017). However, these percentages depend heavily on the organism of study as well as on the number of ortholog families tested. With nearly 15% of families showing evidence for positive selection out of all tested, the study in this dataset of stramenopiles falls within the lower end of this range. Due to the filtering of orthogroups with at least five members in the diverging group of oomycetes, 15% is likely an underestimate of the actual number. A better approximation would require a larger dataset with the addition of species in the families from which up to now few members have been sequenced, such as Albuginaceae.

Positive selection explains host range

When looking at the number of GUPS that were identified for each taxon and comparing members of the same clade that had a different lifestyle, we found that they inversely correlated to biotrophic potential (Appendix A.2 Figures 3 and S5), that

is, the dependence of the pathogen on their host for growth. This ranges from the highly reliant obligate biotrophs down to the less dependent facultative necrotrophs. As more complex environments consist of more degrees of freedom where more variables that affect an organism need to be addressed, a larger number of loci may be needed to keep up the evolutionary race. Additionally, more complex traits are usually influenced by more genes, that is, they are highly polygenic (Mitchell-Olds, 2013). Both these points, in turn, may translate into more genes that have positive selection signatures as was observed for the dataset of oomycetes.

This correlation between GUPS and lifestyle accounted for genome and proteome reduction, since genome size explained only but a fraction of the total variance in the dataset (Appendix A.2 Figure S4). Selective pressure to retain a small genome is commonly occurring in pathogens due to the significant maintenance and evolutionary costs a large genome entails (Knight *et al.*, 2005). The correlation also accounted for the phylogenetic relationships within the taxa in the dataset (Appendix A.2 Figure 2). However, since lifestyle also depends upon phylogeny, that is, phylogenetically closely related organisms tend to have the same lifestyle, it is hard to completely separate this confounder variable (Kuntner *et al.*, 2014; Diepeveen *et al.*, 2018).

Transport and metabolism are under selection in oomycetes

We studied the functional enrichment of the GO terms related to biological processes and cellular compartments for the GUPS belonging to all oomycetes in the dataset, and thus, independently of a specific lifestyle. The resulting biological functions related greatly to transport and metabolism of carbohydrates, lipids, and proteins (Appendix A.2 Figure 4a). The more enriched terms from the cellular component category related to the ribosome and transferase complex, including many associated to the plasma membrane (Appendix A.2 Figure 4b).

Transport at the plasma membrane interface is a crucial process for interacting with the habitat and adjusting to new environmental conditions in many organisms (Conde *et al.*, 2011; Konings, 2006). Therefore, it was not surprising to find functions related to transport in the most enriched GO terms of the GUPS in oomycetes. Proteins related to transport have played an important role in the adaptation

of oomycetes to new hosts to such an extent that they are often targeted by anti-oomycetal compounds for pest control (Chinchilla *et al.*, 2019). Transport-related terms, such as cation transport, featured prominently in the enriched terms (Appendix A.2 Figure 4a). As an example of their relevance to host adaptation, expansion of calcium channels occurred recently in the evolution of early oomycetes (Zheng and Mackrill, 2016).

Metabolic adaptation of the principal molecular building blocks has also been crucial for the evolution of the oomycete lineage (Gómez-Pérez *et al.*, 2021). Since many oomycetes are pathogens and thus depend to a varying level on their host for nutrients, their metabolic machinery has been acclimated to these corresponding pressures. Not only through a complete loss of non-essential genes, but also through positive selection on key enzymes. A special case is the biosynthesis of vitamins and cofactors. In general, they are expensive to produce in terms of the number of unique enzymatic steps required, and therefore their biosynthesis pathways are readily lost when a reliable source is available. This is known as auxotrophy (D'Souza *et al.*, 2014). For example, molybdopterin and folate are indispensable molecules that require many steps to be produced from raw materials (Zhou *et al.*, 2020; Vickers and Beverley, 2011). Therefore, obligate biotroph oomycetes, including *Albugo* and the downy mildews, have all lost these biosynthesis pathways, while they are maintained in other facultative oomycetes such as *Saprolegnia* (Kemen *et al.*, 2011; Duplessis *et al.*, 2011; Dahlin *et al.*, 2017).

Selection data improves lifestyle prediction

In order to detect if positive selection information improved lifestyle classification, we combined the positive selection data with the presence or absence of genome properties. We did this by calculating the ratio of GUPS to those sharing the same functional annotation and with no hints of positive selection. When performing hierarchical clustering on this data, we observed a grouping that differed from that of the mere presence of annotations and better represented the lifestyle of the taxa in the dataset (Appendix A.2 Table 1 and Figure 5). In this case, the plant necrotrophs from the separate families of the Peronosporaceae and Pythiaceae grouped together. We concluded therefore, that positive selection data extracted with the present method gives a better overview of the evolution of lifestyle in oo-

mycetes than merely the presence or absence of genomic features. However, the downsides relating to the need for a curated comparison dataset for the calculation of selective pressures hinder its large-scale application as a standalone predictor.

In plant pathogens, including fungi and oomycetes, the concept of two-speed genome has been discussed in recent years (Dong *et al.*, 2015). According to this model, the core genome evolves at a regular speed, while the accessory genome, which consists in large part of effector-encoding genes, rapidly evolves in response to co-evolution with plants. Therefore, both positive selective pressures and presence/absence of key domains likely played a big part in the adaptation of the plant-infecting oomycetes. Single nucleotide polymorphisms are expected to be more frequent in the core genome, while the deletion of key domains, due to, for example, nonsense mutations, would happen more often in the accessory. This effect as a whole may explain our observation that, in particular, plant pathogens cluster better when both metrics are considered.

2.1.3 Deep learning allows accurate lifestyle prediction

With the aim to study the genome and its features as a way to predict the lifestyle of the organism, we developed a machine learning prediction model. For the training, we collected all available genomes in the NCBI assembly database from eukaryotic plant pathogens to form the base dataset, including fungi and oomycetes. We annotated their lifestyle based on literature consensus and their genomic features by presence or absence of genes encoding key enzymes. To reduce redundancy, we removed non-informative and duplicated entries from the dataset. We then created a multi-layered deep neural network model after splitting the dataset into training, testing, and validation. We found that the constructed model was able to predict lifestyle with high reliance and speed on the unseen test set, consistently reaching higher than 90% accuracy (Appendix A.2 Figure 6). The model was built using the Keras framework and made available at `https://github.com/danielzmbp/lspred` (Chollet *et al.*, 2015).

In comparison to the tree classifier predictor that was described in Hane *et al.* (2020) and also attempts to predict lifestyle from, in this case, exclusively carbohydrate active enzymes (CAZymes), our model performed slightly better. Thus, we propose that using deep learning on curated datasets of genome features for the

prediction of lifestyle in plant eukaryotic pathogens can lead to highly generalizable models, including core annotations, and not just virulence-related enzymes, like CAZymes. With the advent of increasingly cheaper sequencing technologies, such tools will be key to identify potentially pathogenic microbes, for example from uncharacterized samples resulting from environmental surveys.

2.2 *Albugo* releases proteins into the apoplast

Among those proteins that have evolved to adapt to different lifestyles and environments, novel mechanisms for interaction with other microbes, be it positively or negatively, are yet to be thoroughly studied. The plant pathosystem consisting of *Albugo* and *Arabidopsis thaliana* presents an untapped resource for the discovery of novel inhibitory interactions that could be of potential interest in biotechnological applications. Thus, to identify and study proteins from *Albugo* that may be relevant in microbe-microbe interactions, we set out to first characterize those released by the former into the apoplast of *A. thaliana,* that is, the shared environment for endophytic microbes. In order to do this, we performed a proteomic analysis on apoplast samples from leaves of healthy and *Albugo*-infected plants (Figure 2.2). To obtain the latter we separately sprayed the plants with spore suspensions from two *Albugo* species: *Albugo candida* (Nc2) and *Albugo laibachii* (Nc14). We used these two different strains of *Albugo* for the proteomics, in order to compare and assess the success of the method. As described below, we observed a large overlap between the identified Nc2 and Nc14 proteins, since the majority (357 in Nc2 and 346 in Nc14) were direct orthologs, which reinforced our confidence in the apoplast extraction methods.

2.2.1 Apoplast infiltration and extraction

We infiltrated *A. thaliana* leaves by submerging them in MES buffer and applying vacuum in a chamber. This buffer had a pH of 5.5 to closely resemble conditions in the apoplast. After releasing the vacuum, we collected the infiltrate, obtaining a lightly yellow-colored liquid. The yellowish color of the infiltrate suggested the presence of chlorophyll due to broken plant cells. This was confirmed by the predominant presence of the major RuBisCo subunit in the extracted apoplast, as

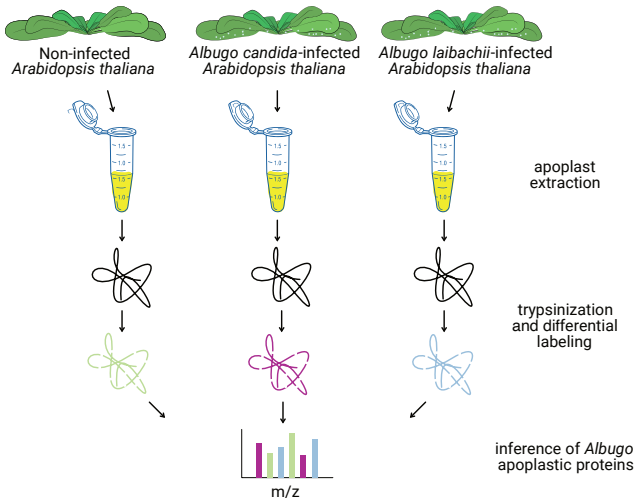


Figure 2.2: Graphical representation of the apoplast sampling and proteomics pipeline for the two *Albugo*-infected, *Albugo candida* (Nc2) and *Albugo laibachii* (Nc14), and uninfected treatments of *Arabidopsis thaliana* leaves.

observed with SDS-PAGE at around 55 kDa (Figure 2.3). The apoplast extraction was performed in three biological replicates for the three plant treatments: water-sprayed (uninfected), infected with *Albugo candida* (Nc2) and infected with *Albugo laibachii* (Nc14). The samples were sent to the proteomics facility of the University of Tübingen for analysis, as described in Appendix B.1 (Gómez-Pérez *et al.*, 2022).

2.2.2 Defense proteins are upregulated in infected plant apoplast

An average of 3893 proteins (standard deviation = 147.3) from *A. thaliana* were found in the three treatments. The number of the identified proteins was consistent, indicating a comparable amount of overall extracted proteins for the different treatments and replicates. When looking exclusively at the proteins from *A. thaliana*, the infected apoplast in both the Nc2 and Nc14 samples showed an enrichment in the number of proteins compared to the uninfected samples (Figure 2.4). Most of these proteins were known to be related to the plant defense response, which would match with the expected result (Table 2.1). The larger enrichment of such proteins in Nc2 is coherent with the more necrotic nature of its infection.

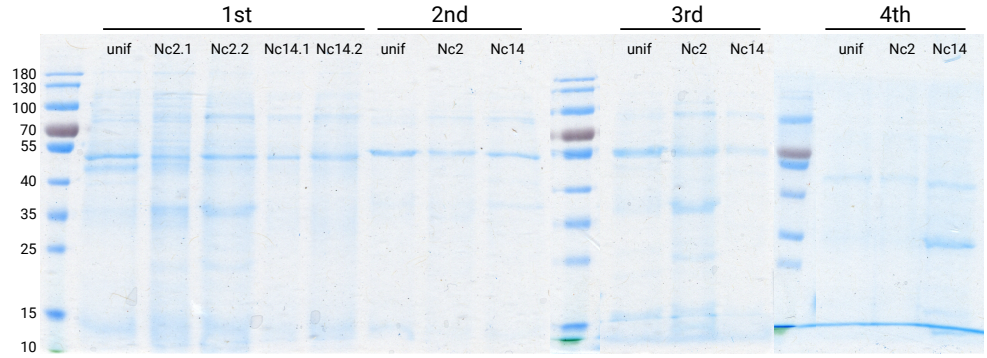


Figure 2.3: Prior to proteomic analysis, the apoplast infiltrate from the indicated plant samples was analyzed via SDS-PAGE. The numbers on top of the scan indicate the biological replicates (n=4), .1 and .2 representing two extractions for the same replicate which were combined before proteomic analysis. The last two biological replicates were also combined into one for analysis due to their lower concentration compared to the first two. The molecular ladder is given on the left in kDa. *Unif: uninfected, Nc2: Albugo candida, Nc14: Albugo laibachii*

2.2.3 *Albugo* proteins are present in the apoplast

As expected, due to the presence of *Albugo* and its secreted proteins, the total number of proteins was higher for the *Albugo*-infected plants. However, when looking at the total intensities before normalization, the uninfected *A. thaliana* treatment had the largest peptide intensity, followed by Nc2 and Nc14 in all cases (Figure 2.5). This may be related to the closing of stomata and rolling of leaves in the infected plants, which decreased the infiltration volume, and thus the amount of extracted proteins. Nc2 infection results in more apparent necrotic lesions on the plant leaf, which could have generated more cellular debris during apoplast extraction, explaining the higher peptide intensity compared to Nc14. For *Albugo*, we identified 1054 proteins for Nc2 and 1286 proteins for Nc14. Using a cutoff of ten for the Andromeda peptide score to identify high confidence proteins, the final numbers were 563 and 658 (4.23% and 4.77% of their corresponding predicted proteome), respectively.

Estimates of the number of secreted proteins, known as a whole as secretome, in fungi are between 100 and 1000 proteins, reaching between 2% and 7% of the total predicted proteome, and varying a lot depending on lifestyle (Lum and Min,

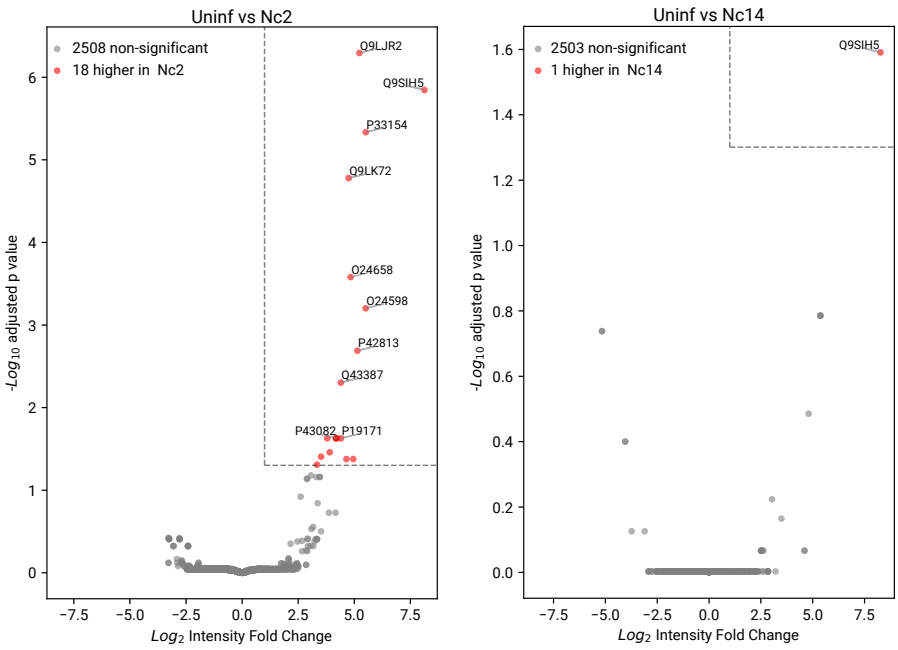


Figure 2.4: Volcano plot of the comparison of *Arabidopsis thaliana* proteins in the apoplast of uninfected (Uninf) against the two *Albugo* strain treatments, Nc2 and Nc14. Labeled are the ten topmost upregulated proteins among the significantly enriched proteins.

2011). The numbers we obtained for the two *Albugo* strains lay well within this range, suggesting good sensitivity of the proteomic analyses. However, since we performed proteomics at a single time point after infection, this is likely a small part of the full arsenal of secreted proteins from *Albugo*.

Apoplastic proteins are enriched in metabolism-related terms

To get an idea of the composition of the apoplastic proteome from *Albugo*, we studied the gene ontology (GO) enrichment in the apoplastic proteins of each *Albugo* strain. We did this by comparing the GO annotations from the high confidence apoplastic proteins to the corresponding predicted full proteome found in NCBI. The analysis of the GO terms relating to biological processes revealed a majority of these to overlap between both strains, particularly in their abundance of metabolism- and biosynthesis-related terms. However, Nc14 had more

Protein	Annotation	Function	Related to defense?
Q9LJR2	Lectin-like protein LEC	Plays a role in defense responses triggered by jasmonate, ethylene and chitin	Yes (Lyou *et al.*, 2009)
Q9SIH5	Probable F-box protein At2g36090	Unknown	Unknown
P33154	Pathogenesis-related protein 1	Partially responsible for acquired pathogen resistance	Yes (Brenya *et al.*, 2020)
Q9LK72	Lectin-like protein At3g16530	Unknown	Unknown
O24658, O24598	Endochitinases At2g43590 and At2g43580	Random endo-hydrolysis of N-acetyl-beta-D-glucosaminide (1->4)-beta-linkages in chitin and chitodextrins	Yes (Devoto *et al.*, 2005)
P42813	Ribonuclease 1	May remobilize phosphate, particularly when cells senesce or when phosphate becomes limiting	Yes (Kim *et al.*, 2019)
Q43387	Peroxidase 71	Removal of H_2O_2, oxidation of toxic reductants, biosynthesis and degradation of lignin, suberization, auxin catabolism, response to environmental stresses such as wounding, pathogen attack and oxidative stress	Yes (Schenk *et al.*, 2000)
P19171	Basic endochitinase B	Defense against chitin-containing fungal pathogens	Yes (Thomma *et al.*, 1998)
P43082	Hevein-like preproprotein	Fungal growth inhibitor	Yes (Bertini *et al.*, 2012)

Table 2.1: Annotations of significantly upregulated *Arabidopsis thaliana* proteins in the *Albugo*-infected apoplast proteomics. The names and functions correspond to those found in the UniProt database.

enriched terms than Nc2 due to the higher number of the identified apoplastic proteins. The proteins with higher enrichment in Nc14 than in Nc2 were related to catabolism, transmembrane transport of ions, and reactive oxygen production; those with lower enrichment were related to amino acid biosynthesis, as compared to Nc2 (Figure 2.6). The enrichment of molecular function GO terms resulted again in a substantial overlap between Nc2 and N14 (22 out of 43 terms were co-occurring). Common terms were related to peptidase, oxidoreductase, lyase,

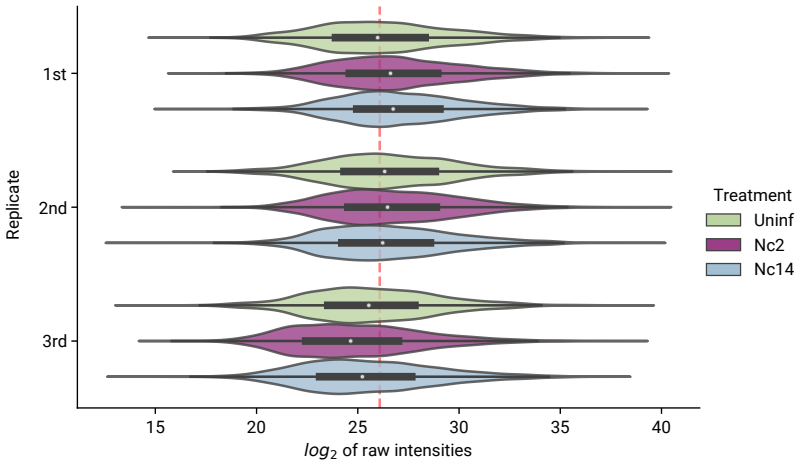


Figure 2.5: Violin plot of the total peptide intensity distributions before normalization of all three treatments (Uninf: uninfected, Nc2: *Albugo candida*, Nc14: *A. laibachii*) in all three biological replicates analyzed by proteomics. The median of all is represented as a dashed red line. The black square corresponds to the interquartile range and the white dot at the center the median for each distribution.

and hydrolase activities (Figure 2.7). Additionally, there was an enrichment in ion and carbohydrate binding functions, particularly for the Nc14 strain.

The reason for the abundance of metabolism-related terms in the apoplastic proteins might be twofold. First, the release of vesicles or exosomes containing enzymes that break down plant macromolecules for incorporation in the oomycete's metabolism as nutrients has been documented in the literature (Judelson, 2016; Presti and Kahmann, 2017). These exosomes were likely broken down and the containing proteins integrated into the samples during apoplast extraction. The time point of sampling, which corresponded to an intense growth stage of *Albugo*, would match with the observed enrichment. Second, due to the large phylogenetic difference with well-studied reference genomes there is a lower confidence in the annotations by homology. This translates into a successful match for only the highest conserved domains, which includes those related to primary metabolism. Particularly in the apoplast, there is a presumed enrichment of host-specific effectors, resulting in many proteins without assigned annotations, thus being excluded from the enrichment analysis (Koeck *et al.*, 2011). Within the high confi-

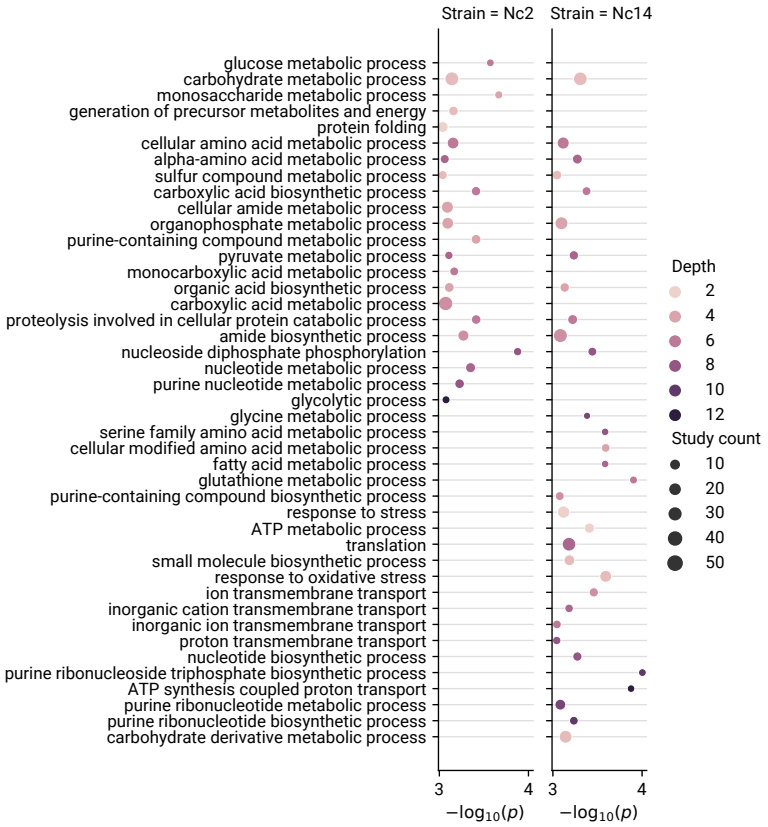


Figure 2.6: Gene ontology (GO) terms for biological processes under enrichment in the apoplast proteins of two species of *Albugo*, *candida* (Nc2) and *laibachii* (Nc14). The significance is given as the negative logarithm of the Holm-corrected p value calculated through Fisher's exact test. Shown in the graph are terms with a significance of 3 or more. Depth corresponds to the number of upstream terms in the GO hierarchy for a particular term. Study count is the number of terms within the category in the dataset.

dence apoplastic proteins, those without assigned GO terms corresponded to 385 (68.4%) and 396 (60.2%) in Nc2 and Nc14, respectively.

Minor cytoplasmic contamination from the hyphae is likely to have occurred, owing to the presence of mitochondria-localized annotated proteins in the resulting apoplast. However, similar applications of the same method have shown little to no cytoplasmic contamination, as measured through enzymatic assays (Zhang *et al.*, 2008). The presence of mitochondria-localized annotations in this dataset

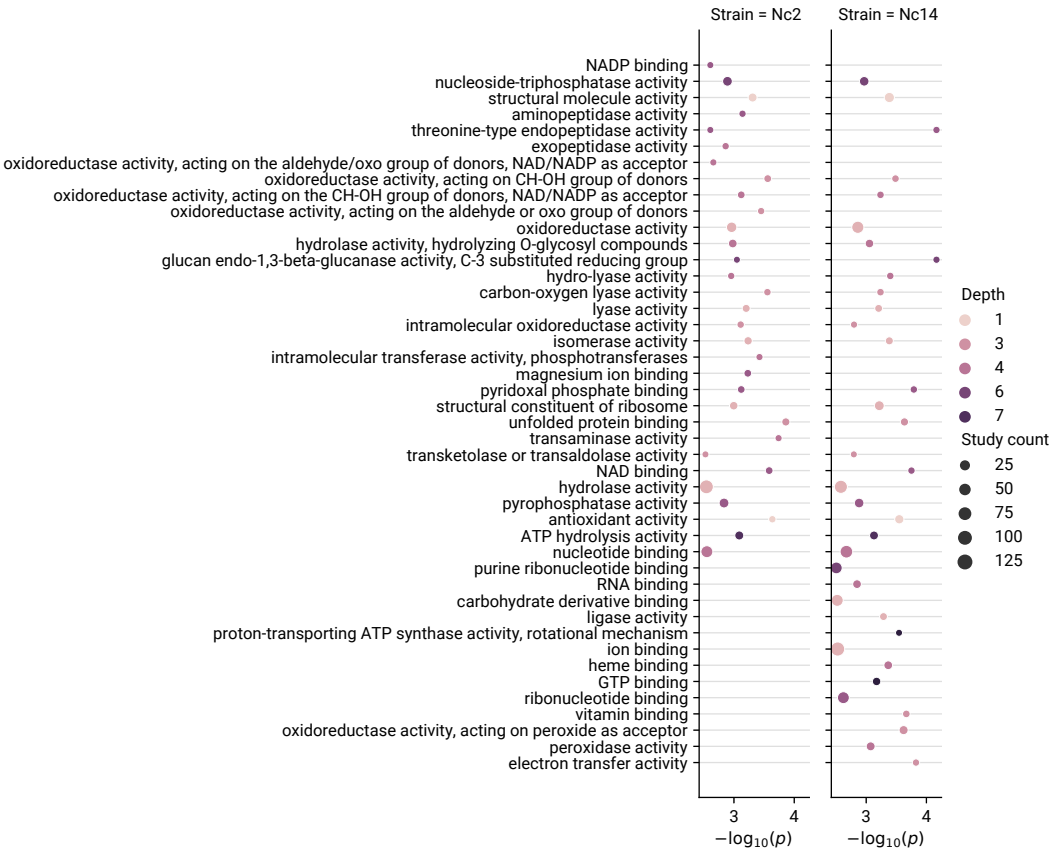


Figure 2.7: Gene ontology (GO) terms for molecular functions under enrichment in the apoplast proteins of two species of *Albugo, candida* (Nc2) and *laibachii* (Nc14). The significance is given as the negative logarithm of the Holm-corrected p value calculated through Fisher's exact test. Shown in the graph are terms with a significance of 2.5 or more. Depth corresponds to the number of upstream terms in the GO hierarchy for a particular term.

could therefore indicate not an issue with our implementation but a consequence of *Albugo* infection or an incorrect prediction. Currently, it is not clear to what extent release of cytoplasmic contents plays a role in *Albugo* infection. It can be hypothesized that when threatened, cytoplasmic proteins from the hyphae could be involved in microbial control, as known for plant cells (Ruano and Scheuring, 2020). Overall, enzymatic assays would need to be performed in the future to fully exclude cytoplasmic contamination, but the minor contamination observed could

be a feature of *Albugo* infection and not a problem with our methods.

Overall, via these proteomic analyses of infected apoplast from *A. thaliana*, we have identified proteins from the apoplastic secretome of *A. candida* and *laibachii*. Some of them include known virulence factors, such as the many identified CAZymes. Since *Albugo* lacks biosynthetic power, this dataset provides a starting point for the study of proteins relevant in microbe-microbe interactions through which *Albugo* may shape its environment and construct its niche. We have made available the raw proteomics data at a public repository, as it may provide useful insights to other researchers working on similar questions `https://www.ebi.ac.uk/pride/archive/projects/PXD031981/`.

2.3 *Albugo candida* releases antimicrobial proteins into the apoplast

As mentioned in the Introduction, *Albugo* is a plant pathogen that strongly modifies the surrounding microbial community in the apoplast, usually resulting in lower diversity and increased stability after infection (Agler *et al.*, 2016). Particularly interesting due to its translational potential is the inhibitory effect on the growth of bacteria. Based on the loss of its biosynthetic capacity we speculate that the antimicrobial effects of *Albugo* are mediated by proteins. In this section, we discuss the relevance of a subset of proteins secreted by *A. candida* into the apoplast of *A. thaliana* out of those that were identified in the apoplastic proteomics (Section 2.2) as a mechanism that could explain such antimicrobial activity.

2.3.1 *In silico* evidence suggests direct inhibition by *Albugo*

In order to identify inhibitory interactions in the phyllosphere originating from *Albugo*, we performed a series of *in silico* analyses. These included correlation network inference and protein predictions of antimicrobial activity on the *Albugo* apoplastic proteins. The goal was to establish potential candidate strains and proteins for *in vitro* testing. The correlation network was inferred from a large amplicon sequencing dataset from the phyllosphere of infected and uninfected wild *A. thaliana* plants (Mahmoudi *et al.*, in preparation).

Albugo interactions are abundant in the phyllosphere

To assess the relevance of the inhibitory effect on the phyllosphere and study the negative correlations of *Albugo*, we used FlashWeave to performed network interaction inference on a large dataset of environmental microbial abundances collected by members of our group over a period of six years (Tackmann *et al.*, 2019). In the dataset, *Albugo* amplicons were prominently present, as infected plants were deliberately sampled. FlashWeave applies a causal inference algorithm that attempts to find direct correlations between operational taxonomy units (OTUs), while controlling for spurious connections. The dataset originated from the amplicon sequencing sampling of *A. thaliana* plants in different wild populations. Taking into account this and other abiotic metavariables, as they can result in false connections, we found the OTU assigned to *Albugo* ranked in the top 20 when looking at the negative correlations of all OTUs (Appendix B.1 Figure 1a). Of note, many of the negative interactions from *Albugo* linked to bacterial OTUs (Appendix B.1 Figure 1b). Although the network was undirected, *Albugo* being the source of inhibition is consistent with an overall decrease in community complexity and diversity on the infected plants that was reported in Agler *et al.* (2016).

As described in Section 1.2, *Albugo* has been studied as a hub microbe. While the large number of interactions between *Albugo* and other microbes in the phyllosphere strongly supports this, many of the identified microbes are not culturable in standard isolating media, such as *Adhaeribacter* and *Truepera* (Albuquerque *et al.*, 2005; Rickard *et al.*, 2005). Therefore, these were likely not isolated during sampling and there were no samples of these genera available for us on which to test the protein. In soil, negative interactions have recently been shown to determine community assembling and functioning, thus highlighting the relevance of *Albugo* for the phyllosphere community (Romdhane *et al.*, 2022).

Antimicrobial prediction is enriched in apoplastic proteins

Using the proteomic work (Section 2.2) as a starting point, we selected one of the *Albugo* strains to dive more deeply into the mechanisms of proteins relevant for antimicrobial activity. Based on preliminary tests in which apoplast extracts were incubated with bacteria, we chose to focus on the Nc2 strain of *A. candida* rather than Nc14 (Jonas Ruhe dissertation, 2016). We performed a series of machine learning

predictions on the putative proteome to identify the most interesting candidates, including prediction of intrinsic disorder and antimicrobial activity. We predicted disorder using flDPnn, which was the best performing tool at the recent critical assessment of protein intrinsic disorder prediction experiment, where research teams competed for the most accurate software at predicting disorder (Necci *et al.*, 2021; Hu *et al.*, 2021).

For the antimicrobial activity prediction, we used the following published machine learning methods to create a consensus score, *Antimicrobial peptide scanner v2*, *amPEPpy* and *AmpGram* (Veltri *et al.*, 2018; Lawrence *et al.*, 2020; Burdukiewicz *et al.*, 2020). To leverage the correlation between protein length and prediction score that was evident for two of the predictors, we applied to each a weight proportional to the R^2 value of the best least-squares linear regression fit (Appendix B.1 Figure S5). The final score consisted of the weighted sum of the three scores. For each protein, we considered a positive prediction when its weighted score was higher than 0.5. Using this approach, although numerous positives were found in the predicted proteome of *A. candida*, we observed a significant enrichment of antimicrobial proteins in the apoplast subset (Appendix B.1 Figure 2d).

Such consensus estimation methods, in which multiple predictors trained on different databases are integrated to get a unique score, are useful to alleviate outlier and biased predictions. Although implemented in other areas of protein prediction such as amyloid or disorder prediction, this is the first pipeline consisting of state-of-the-art methods for high-throughput antimicrobial activity prediction based on amino acid sequences (Emenecker *et al.*, 2021; Tsolis *et al.*, 2013). We made it publicly available at `https://github.com/danielzmbp/appred`.

To sum up, *Albugo* is predicted to have numerous negative interactions predicted in the phyllosphere environment and potential antimicrobial activity is enriched in the apoplast. Both of these add support to the hypothesis that *Albugo* actively modifies the preexisting plant leaf microbiome in order to construct its niche (Laland and O'Brien, 2010).

2.3.2 Cloning and overexpression

To verify our prediction pipeline and study the antimicrobial potential of these proteins *in vitro*, we selected two predicted antimicrobial candidates with a high

relative abundance in the apoplastic proteome, C06 and C14, for heterologous overexpression. We additionally selected two predicted non-antimicrobial proteins, C05 and C15, which were comparable in their apoplast relative abundance and length to the previous two (Appendix B.1 Figure 3a and Table 1). All four except C06 had a predicted enzymatic domain. C14 and C15 had predicted secretion signal peptides, which were excluded during cloning.

After trying eukaryotic expression methods like *Pichia pastoris* which yielded low protein output, we decided to select a prokaryotic expression system in order to obtain enough protein to test its activity on a collection of strains. We employed the *E. coli* strains Rosetta™ DE3 and SHuffle® to improve the expression of eukaryotic and cysteine-rich proteins (Lobstein *et al.*, 2012). First, we extracted total RNA from *Albugo*-infected leaves and synthesized complementary DNA. From this template, we amplified the candidates. We cloned the candidates into the multiple cloning site of the pET28b (+) vector which contained a T7 promoter domain inducible by IPTG. Overall, we observed a significantly larger expression of the candidates in the *E. coli* expression system than in the eukaryotic one. Therefore, we concluded *E. coli* to be a suitable system for expression of eukaryotic proteins when a large concentration for testing is required.

For the purification process, the candidates had either one 6x histidine tag fused at the C-terminus or two flanking such tags. The latter was the case for candidates C05, C06 and domain derivatives of C06. They were so designed to improve binding of the tag to the HisTrap's Ni-NTA agarose matrix and thus increase yield of the purified protein (Appendix B.1 Figure S6). As a consequence, we successfully observed larger concentrations of these candidates when compared to C14, C15 and domain derivatives of C14 after the elution from the purification column (Table 2.2).

After overexpression of the proteins in *E. coli*, we found that C05 and C14 formed insoluble inclusion bodies. This was likely because of a larger number of cysteine residues than the other candidates (Table 2.2), and hence a higher number of disulfide bonds in the native protein fold (Fischer *et al.*, 1993). Testing different inducing concentrations and incubation conditions resulted in no yield of soluble protein. For comparability, therefore, we performed a denaturing purification of all candidates using high urea concentrations followed by a refolding procedure (Appendix B.1 Figure S7). The refolding was performed by dialysis in a buffer

Candidate	Concentration (mg/ml)	Number of 6x HisTags	Number of cysteines	Protein length (aa)	Molecular weight (kDA)
C05	0.244	2	6	544	60.24
C06	0.284	2	0	533	57.12
C06d	0.115	2	0	194	21.27
C06b	0.201	2	0	368	39.36
C14	0.024	1	15	440	48.94
C14d	0.064	1	6	138	15.15
C14b	0.054	1	9	311	35.04
C15	0.057	1	2	437	49.01

Table 2.2: Concentration of candidate proteins from *Albugo candida* after denaturing purification and elution in 50 ml of buffer from HisTrap column. Labeled for each candidate is whether they present one or two 6x histidine tags (HisTags) at the C-terminus, or C- and N-termini, respectively. The number of cysteines corresponds to the occurrences of this residue in the expressed sequence of the candidate.

resembling apoplastic conditions, in order to have a comparable environment to the native one for the antimicrobial tests, with four buffer changes over a period of two days. The refolding was performed using an air oxidation method, as the standard glutathione oxidation was found to interfere with the antimicrobial tests downstream (Fischer *et al.*, 1993). Circular dichroism (CD) spectroscopy analysis on C14 after this refolding treatment showed a positive spectrum for a secondary structure matching the AlphaFold prediction (Jumper *et al.*, 2021), which includes a core rich in β-sheets (Figure 2.8 and Section 2.3.5).

The last buffer change during dialysis was used as the blank negative control without protein in the antimicrobial tests described below. The concentration of the candidates was assessed by Bradford assay using Bovine Serum Albumin (BSA) as a standard. Finally, the purity of the candidates after dilution for testing was measured by SDS-PAGE, where at least 90% of the intensity was found in the bands corresponding to the molecular weight of the candidates (Appendix B.1 Figure S6).

2.3.3 Antimicrobial effect

In this section, we describe the results of the inhibitory activity from the tested candidates on the growth of plant-isolated bacterial strains and discuss poten-

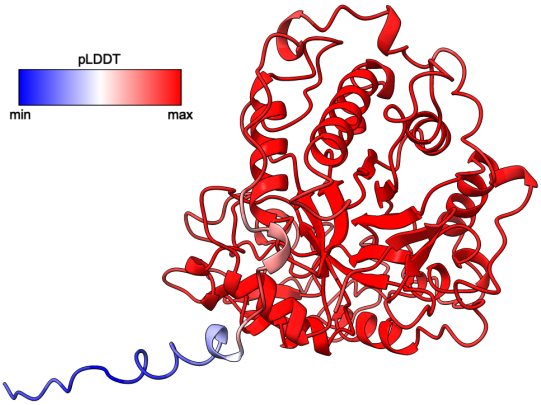


Figure 2.8: Alphafold's relaxed model best fit for protein C14. The pLDDP score per residue represents the accuracy of the prediction and the disorder. In red are the areas with high prediction and in blue, those with low. Note the C-terminal disordered tail on the bottom left, and the core rich in β-sheets.

tial mechanisms responsible for this effect. We performed antimicrobial tests on the full protein candidates C05, C06, C14 and C15 by monitoring the growth of microbes incubated with the protein and comparing the area under the curve (AUC) to negative controls. To better narrow down the source of antimicrobial activity, we also separately cloned arbitrarily delimited domains from C06 and C14, named *b* and *d*. The *b* domains corresponded to the less disordered N-terminal regions, C06b and C14b, while the *d* domains corresponded to the more disordered C-terminal regions, C06d and C14d (Appendix B.1 Figure 4a). Sometimes known as *divide-and-conquer*, this approach consists in the separate expression of domains in order to functionally study a protein.

Protein candidates selectively inhibit bacteria

We performed antimicrobial assays in 96-well microtiter plates using the purified proteins diluted to equal molarity on a selection of strains from a library of plant-isolated bacteria (Appendix B.1 Table S3). This library was obtained by previous members of our group in parallel to the sample collection of the amplicon sequencing analysis of microbial phyllosphere isolates of *A. thaliana* (Almario *et al.*, 2022). The strains were incubated together with the protein candidates at a con-

stant orbital shaking overnight and room temperature. In order to monitor the growth of bacteria, the optical density at 600 nm (OD_{600}) of samples with the protein was measured every 15 minutes and compared to that of those with added blank instead (rebuffering dialysis wash).

We found specific inhibitory activity on a number of Gram-positive bacterial strains mainly for C06 and C14, the two candidates predicted to be antimicrobial *in silico*, (Appendix B.1 Figures 3b and S8). To a lesser extent, we also found C15 to be antibacterial against certain strains, which, interestingly, did not always overlap with C06- and C14-sensitive strains. This suggests distinct mechanisms of action for the different candidates. On the majority of the other tested bacterial strains, there were slight growth promoting effects with all proteins. This was perhaps due to the ability of the microbes to digest the proteins and use the peptides as nutrients. In the remaining of the tested strains, including five Gram-negative bacteria, there were no growth differences between blank and protein treatments. Gram-positive and -negative bacteria are known to secrete different proteases to their environment, which could explain the discrepancies in the observed growth (Wandersman, 1989). Fitting our hypothesis, Gram-positive bacteria from soil have been shown to more rapidly integrate amino acids and peptides from the environment when compared to Gram-negatives (Broughton *et al.*, 2015).

We are still investigating the mechanisms responsible for the observed antimicrobial effects, but we discuss some of our observations below in Section 2.3.3. All in all, the inhibition on the sensitive strains correlated with the presence of IDRs in the candidates, as C06, C14 and C15 had intrinsic disorder to a varying degree, while C05 had not (Appendix B.1 Figure 4a).

All tests were performed at pH 5.9 to resemble the apoplast and to match with the literature on the conditions for higher effectivity of cationic AMPs (Malik *et al.*, 2016; Walkenhorst *et al.*, 2013). Since the lower pH causes the protonation of basic residues, inhibitory activity that is dependent on the net charge for binding to the membrane is increased in many cases (Malanovic and Lohner, 2016). Consistent with this, at the higher pH of 7.2, C06 showed less to no inhibitory activity against sensitive strains (Appendix B.1 Figure S8). Taking this data together, we hypothesize that the pH might act as a trigger for the antibacterial activity exclusively within the boundaries of the apoplast.

When looking at the domain tests, we found a correlation between the degree of

inhibition and the predicted net charge (Appendix B.1 Figures 4b and 4c). That is, the higher the theoretical net positive charge, the stronger the inhibitory effect. Additionally, the degree of inhibition also correlated with the presence of predicted IDRs in the candidates (Appendix B.1 Figure 4a), and enriched GO terms related to unfolded proteins, including protein folding and unfolded protein binding (Figures 2.6 and 2.7). The highly charged and disordered domains C06d and C14d had the strongest effect, while the less disordered and more negatively charged domains, C06b and C14b, had the weakest effect. A concentration dependent inhibition was observed for the *d* domains. Of note, C06d and C14d reached full inhibition at 2.15 µM towards the *Microbacterium* strain I20 and for the same concentration, C06d prevented growth of *Aeromicrobium fastidiosum* strain I01 (Appendix B.1 Figure 4d). For the domain C06d, we additionally found inhibition of plant pathogenic strains from the *Clavibacter* genus (Appendix B.1 Figure S13).

From the proteomic data, we detected evidence for the presence of all candidates in the apoplast along most of the entirety of their sequence (Appendix B.1 Figure 4a). The exception was the disordered N-terminal domain of C15, stretching a length of 150 amino acids, suggesting a misannotation or processing during release. The presence of C05 and C06 in the apoplast would indicate an unconventional secretion, as in contrast to C14 and C15, they do not contain a predicted secretion signal. Regarding the *d* domains from the positive candidates, which were responsible for most of the antimicrobial activity, four peptides from C06d and one from C14d were found (Appendix B.1 Table S6). These peptides encompass almost half of the intrinsically disordered residues as predicted *in silico*. All in all, these data support the presence of the candidates in the apoplast of infected plants, including their more active regions as indicated by the antimicrobial *in vitro* tests. In the next section, we describe potential mechanisms that could explain the observed inhibitory effects.

Mechanisms of inhibition

The *d* domains, apart from the largely disordered sequence prediction, have a high putative positive charge. They could therefore be considered cationic AMPs. Cationic AMPs are characteristic for having a net positive charge near physiological pH (Bradshaw, 2003). As mentioned above, this is due to an increased rela-

tive local abundance of certain basic residues, such as arginine, lysine or histidine, which bring the predicted isoelectric point (pI) of the individual proteins up. Cationic AMPs are known to target Gram-positive bacterial strains due to their more negatively charged membrane (Omardien *et al.*, 2016). The inhibitory activity occurring at pH 5.9 could suggest a specific inhibition exclusively in the apoplastic fluid due to a dependence on the higher net charge for effect, in the same way as antimicrobial peptides are activated in the acidic pH of the skin (Zheng *et al.*, 2020). We thus hypothesize the inhibitory effects to be related to the IDRs. These, apart from having a large compositional bias for the presence of basic residues, would not be tightly folded, allowing these residues to be susceptible to protonation, further increasing the positive charge.

The numerous plant and microbial proteases present in the apoplast could have produced the excision of peptides from the full proteins (Wang *et al.*, 2020). Consistent with this hypothesis, a number of plant proteases, namely subtilisin-like serine proteases, were found to be significantly upregulated in the apoplast of infected plants (Appendix B.1 Table 3). IDRs, which lack a stable folding, could in this regard be a more readily accessible target for proteases, resulting in free peptides from these regions in specific conditions (Suskiewicz *et al.*, 2011). For example, when pH in the apoplast changes due to biotic stress. The peptide AQSAVHEQTEPSKPGFGEK from the C-terminus of C06, which corresponded to the *d* region and overlapped with the IDR, had the highest absolute peak intensity of all found peptides matching to this candidate in the apoplast proteomics. Thus, we suggest that proteins may be secreted as a whole and later activated by proteases. This kind of mechanism has been studied in other AMPs, originating from bacteria and human skin, which are activated after proteolysis of the propeptide (Yamasaki and Gallo, 2007; Corvey *et al.*, 2003).

The lowest concentration we tested that had an effect on the growth was 0.1 µM (Appendix B.1 Figure 4d). Although not much is known about the native concentration of proteins in the apoplast, the tested protein concentrations were likely higher than what would be present in the natural system. It has been estimated that proteins inside cells are found within the nanomolar range (Komatsubara *et al.*, 2019); however, since the apoplast is a much larger volume, released apoplastic proteins would be expected to occur natively in a smaller concentration. We hypothesize thus that the effectivity of these antimicrobial proteins would have to

rely on localized release to be bacteriostatic.

2.3.4 Ecological relevance of the inhibitory activities

To study the ecological relevance of the bacteria inhibited by the heterologously-expressed proteins of *Albugo candida*, we used two synthetic communities (SynComs) of microbes (Appendix B.1 Table S3). These SynComs were composed of core *A. thaliana* microbes including the I01 strain of *Aeromicrobium fastidiosum* that showed high inhibition in the *in vitro* antimicrobial tests. As strain I01 played a role in stabilization of the community against the bacterial phytopathogen *Pseudomonas syringae* DC3000 (Pst), we used a Pst strain with rifampicin resistance to monitor its abundance in the community through variations in the composition of the SynComs, including exclusion of the I01 strain (Chaudhry *et al.*, in preparation).

First, we tested a SynCom comprised of five bacterial strains to investigate the consequences on removal of strain I01 in regard to resistance to added Pst. We found that the whole SynCom could inhibit somewhat the growth of Pst, while removing *Aeromicrobium* led to increased growth of Pst. This indicated that also in this *in vitro* setting, I01 supports resistance against pathogen invasion. We then added the peptides derived from C06: C06b, which had no effect on I01, and C06d, which was highly effective at inhibiting this strain. We found that adding C06b could indirectly inhibit growth of Pst, likely through protection by *Aeromicrobium*, which is not inhibited by this domain. Meanwhile, adding C06d, which had an inhibiting effect on *Aeromicrobium* in the previous *in vitro* tests, resulted in an increased Pst growth. Comparable to the SynCom without *Aeromicrobium*, Pst was not inhibited (Appendix B.1 Figure 5a). This showed that the proteins are active in a community context and are able to specifically inhibit sensitive bacteria.

We also performed experiments in planta with a larger SynCom of 12 bacterial and three yeast phyllosphere core members. From these, we observed that excluding *Aeromicrobium* had a large effect on the stability and variability of the relative abundance of Pst load readout (Appendix B.1 Figure 5b). Additionally, this was reflected in the relative bacterial abundances, as measured by the lack of clustering of the dropout treatment in the principal coordinate analysis (Appendix B.1 Figure 5c). Taken together, the SynCom experiments demonstrate the potential of

peptide C06d to specifically control the microbial community and pave the way for new biocontrol treatments based on peptides (Keymanesh *et al.*, 2009; Montesinos, 2007).

2.3.5 Secondary structure of C14 is enriched in β-sheets

To get an idea of the structure and the successful folding of the protein candidates, we performed CD spectroscopy on candidate C14 (Figure 2.9). In C14's spectrum, we observed a large component of antiparallel β-sheet. This could suggest a mechanism related to, for example, the formation of β-barrels, which some AMPs are known to assemble in, leading to cell permeabilization (Shafrir *et al.*, 2010). Supramolecular structures, like amyloids, are also implied by this finding. In the future, we plan to perform CD spectroscopy in the remaining candidates and assess their structure predictions and putative intrinsic disorder. In the next section, we discuss C14 and other protein candidates from *Albugo* in the context of functional amyloid formation.

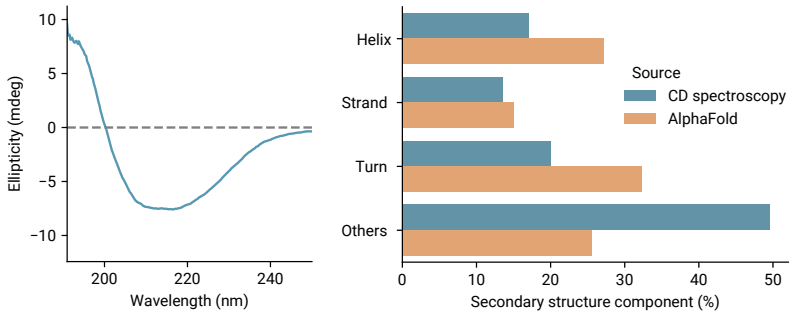


Figure 2.9: CD spectrum obtained from a solution of candidate C14 in testing buffer (BisTris 10mM pH 5.9). The spectrum of the buffer control was subtracted from the curve. On the right, the secondary structure component is represented as a percentage, as predicted from the CD spectrum with BeStSel, and compared to the AlphaFold prediction (Micsonai *et al.*, 2015).

2.3.6 Antimicrobial proteins in *Albugo*

In summary, building upon the apoplast proteomics described in Section 2.2, we have discovered two candidate proteins with antimicrobial properties released by *Albugo* during infection. In line with a similar study for a fungal effector, we

observe specific activity against certain plant-associated bacteria (Snelders *et al.*, 2020). However, in this case, and for the first time reported, we identified proteins originating from a protist. As we see a negative correlation with yeasts in the network analyses, it would be interesting in the future to test the antifungal properties of the *Albugo* proteins as well.

2.4 Amyloid proteins are functionally relevant for *Albugo candida*

Amyloids are an exciting class of proteins that encompass a large functional variety under the umbrella of a common tertiary protein structure. In the last 20 years, they have been discovered and widely studied in microbes including fungi and bacteria, serving key functions to the benefit of their survival (Gómez-Pérez *et al.*, 2021). However, they remain poorly characterized in the group of protists, which includes the phylum of oomycetes. Here, we have studied amyloids produced by *Albugo*, following evidence of Thioflavin T (ThT) staining in the apoplast of infected plants compared to non-infected, as well as binding of ThT in the hyphae (Appendix C.1 Figure 1; Jonas Ruhe dissertation, 2016). We used the secreted apoplast protein data from *A. candida* described in Section 2.2 to identify the most pertinent candidates for the study of functional amyloids. Additionally, through the use of synthetic amyloid structure inhibitor compounds, we established the importance of this structure for the virulence and successful infection of *Albugo* in its host, *A. thaliana*.

2.4.1 Amyloid inhibitor treatment improves plant resistance to *Albugo*

As amyloids play a role in the virulence of many pathogens, we decided to study the relevance for *Albugo*. To this end, we infiltrated four amyloid inhibitors in a 1% DMSO aqueous solution into *A. thaliana* leaves before infecting them with *Albugo* to check the relevance of amyloids for the successful infection of *Albugo*. These inhibitors are chemical compounds derived from curcumin, a natural preventer of amyloid fibrillation (Yang *et al.*, 2005). They were synthesized with different side chains in order to improve solubility and cell uptake. For one of the inhibitors tested, named B13, we observed a significant decrease in the infection phenotype.

The infectivity was evaluated based on the measured number of spore pustules per leaf area due to Nc2 or Nc14 infection after infiltration with amyloid inhibitors in comparison to DMSO control (Appendix C.1 Figures 3 and 4). This suggested that, regardless of the particular mechanism, amyloids produced by *Albugo* play an important role in successful plant infection. This could open new avenues for white rust prophylaxis in agricultural contexts.

As amyloids are also involved in the normal functioning of the plant, including signaling and microbial control, we monitored the growth of the *A. thaliana* plants for toxicity (Antonets and Nizhnikov, 2017). Other than inhibition of plant amyloids, it may be possible that *A. thaliana* detects the inhibitor through a defense pathway and triggers a response, indirectly reducing the number of *Albugo* spores unrelated to amyloid inhibition. To test this, we took images of *A. thaliana* rosettes and applied a custom phenotype imaging pipeline using PlantCV to measure their sizes (Gehan *et al.*, 2017). We found for the second replicate a decrease in the growth of Nc2-infected plants when looking at the area of the rosettes infiltrated with the inhibitor compared to control. However, when taking into account both replicates and strains, as well as another measure of size (longest path), this effect was not significant, but rather it increased slightly in the opposite direction (Figure 2.10). Overall, these data suggested that the impact of inhibitor B13 on plant growth may be negligible, green-lighting its possible consideration for application in the prevention of white rust infection.

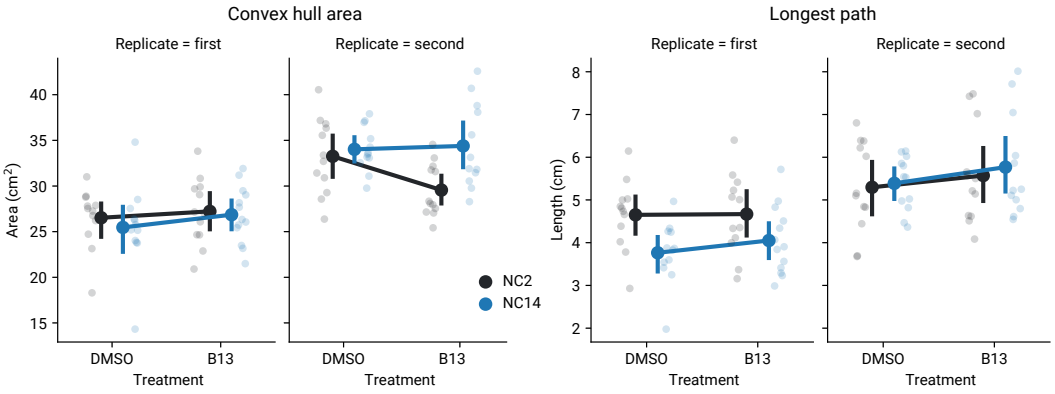


Figure 2.10: Size of *Arabidopsis thaliana* plants as measured by the convex hull area and longest path of the rosettes for DMSO and B13 inhibitor infiltrated treatments. Nc2 and Nc14 represent *Albugo candida* and *Albugo laibachii* strains, respectively.

2.4.2 Amyloid prediction results in apoplastic amyloid candidates

To narrow in on the proteins responsible for amyloid formation, we annotated the putative proteome of *A. candida* using two machine learning predictors for amyloid folding. These were *APPNN* and *AmyloGram* (Família *et al.*, 2015; Burdukiewicz *et al.*, 2017). As for the antimicrobial prediction described in Section 2.3, we found a high correlation of the scores with protein length, therefore, we applied a correction proportional to protein size. We used as cutoff the median of the distribution (0.849) to differentiate between a likely prediction of amyloid formation. The rationale was to get the same number of amyloid and non-amyloid candidates for the whole predicted proteome, since we had to establish the baseline for the biological meaning of such an artificial consensus score. The subset of the apoplastic proteome had an enrichment in proteins above the cutoff as evidenced by the normalized cumulative distribution function (Appendix C.1 Figure 6). As with the study of antimicrobial proteins in *A. candida,* we used the predictors and the abundance in the apoplast analysis from Section 2.2 as a basis on which to select candidates for testing in the laboratory (Appendix C.1 Table 1).

As mentioned earlier, and similar to the overlap between amyloids and antimicrobials which has been observed for some proteins in the literature (Kagan *et al.*, 2012), we found that many of the candidates had both, an antimicrobial and amyloid prediction (Appendix C.1 Figure 7). Nevertheless, the results in the prediction pipelines could have also been skewed due to bias. For example, due to intrinsic disorder, which antimicrobials and amyloids share to a certain extent and both class of predictors take into account (Figure 2.11). However, this was not always the case, as C06 and C15, which showed antimicrobial activity, lacked a high amyloid prediction. For the selection of candidates to test, we included the four proteins used in the antimicrobial assays, as well as other highly expressed proteins, some predicted amyloid and some not, to cover the broad span of the predictions. They ranged from C06 with the lowest prediction, to C16 with the highest (Appendix C.1 Table 1).

2.4.3 Amyloids form *in vivo* when heterologously expressed

We used the Curli-Dependent Amyloid Generator (C-DAG) system to express and study the potential amyloid fold of the candidates (Sivanathan and Hochschild,

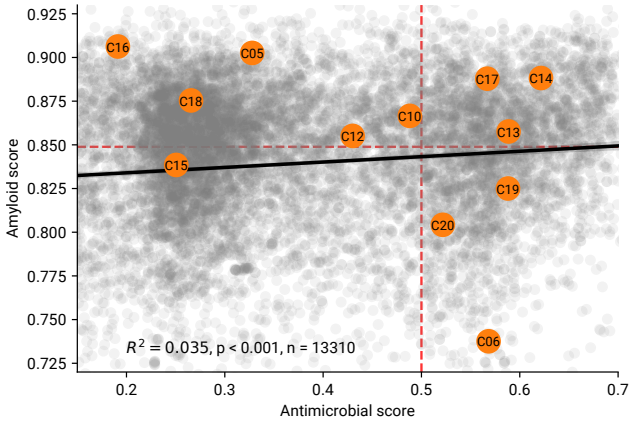


Figure 2.11: Correlation between amyloid score and antimicrobial score for all of the proteins in the putative proteome of *Albugo candida* strain Nc2. The black line represents the linear regression and gray blobs the proteins from the predicted proteome. Dashed red lines represent the cutoffs for binary classification into their respective category. Selected candidates for amyloid and antimicrobial tests are depicted in orange.

2012, 2013). This system takes advantage of the native curli expression machinery from *E. coli* to heterologously test for amyloid formation *in vivo* on a strain that lacks production of the major curli component protein, CsgA, and nucleator, CsgB. Instead, CsgA's secretion signal ($csgA_{ss}$) is fused with a three-alanine linking region to the N-terminus of the protein of interest, whose expression is induced by arabinose under regulation of a T7 promoter in the pExport plasmid. In addition, the *E. coli* strain contains the plasmid pVS76, which overexpresses CsgG after induction with IPTG. After intracellular expression of the candidate protein, $csgA_{ss}$ is recognized and cleaved off by CsgG, the remaining being natively processed as curli component proteins. Thereby, they are secreted in a way that exposes the candidate's amyloidogenic regions and triggers formation of amyloid fibrils outside the outer bacterial membrane. As controls, we used the yeast prion Sup35 with its aggregating domain as positive (Sup35[NM]) and without it as negative (Sup35[M]) control for the formation of amyloids.

The heterologously expressed candidates bind Congo Red

We analyzed the formation of amyloids by observing the color phenotype of C-DAG transformants in LB agar with CR. The shift towards a red color of the colonies indicates formation of amyloids *in vivo*. We selected candidates C05, C06, C10, C12, C13, C14, C15, C16, C17, C18, C19 and C20 from the apoplastic proteome of *Albugo* for testing. As mentioned previously, the choice was made in order to cover the spectrum of amyloid predictions and test those that were abundant in the apoplast from the proteomic analyses (higher than median raw intensity values). The labeling of the candidates reflects the initial choice selection of proteins, and the missing numbers represent the candidates that we were unable to express or amplify.

We observed for the colonies of all predicted amyloids a red hue after four-day growth at 22 °C. C06, C15, C19 and C20, which had an overall lower prediction and placed below the median cutoff, showed none or very little phenotypic coloration, similar to the negative control, Sup35[M] (Appendix C.1 Figure 8). The best candidates for amyloid formation based on this test were C10, C14, C16 and C17. Therefore, we concluded that the prediction methods we employed to study amyloid structure were reliable in the assessment of amyloid formation by apoplastic *Albugo* proteins (`https://github.com/danielzmbp/amypred`). By extension, the pipeline and associated scripts could be applied to other organisms, to get a first impression of their potential to form amyloids.

Fibril-like structures are visible in electron microscopy

In order to further study the positive candidates and directly observe their potential fibrils, we looked at C-DAG cell samples under the TEM. These were prepared as described in Section 2.5. Under the TEM, we observed that, for some candidates including C10, C13, C14 and C16, fibril-like structures were apparent in the extracellular space, consistent with their prediction as amyloids and CR colony color phenotype (Appendix C.1 Figure 10).

Since C14 is of particular interest for its relevance as antimicrobial and showed as well the strongest fibril formation under the TEM, we created a similar construct to pExport-C14 but containing a 6x histidine tail at the C-terminus of C14. With it, we performed immunostaining of the samples using a nano gold secondary antibody. This showed that the fibril-like structures for C14 contained the

tagged protein (Figure 2.12). As with the positive control and cyanobacterial proteins that we tested in this system (Section 2.5), the staining was not evenly distributed but increased in the more amorphous regions of the fibril-like structures, potentially related to the accessibility of the tag.

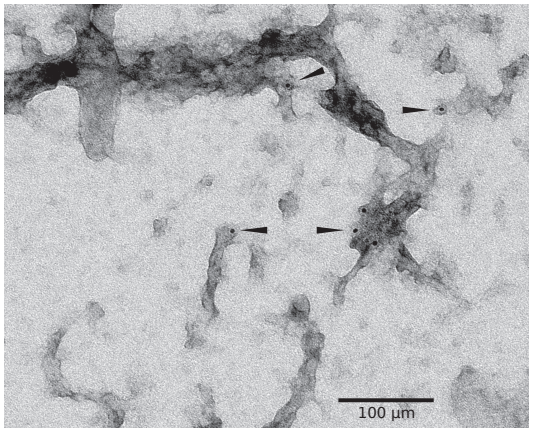


Figure 2.12: Transmission electron microscopy negative staining image of C14 construct in the C-DAG expression system labeled with 6 nm gold particles. Fibril-like and amorphous structures can be seen labeled with the probe. The arrow heads point to the colloidal gold particles in the picture. Captured by Sandra Richter.

2.4.4 Amyloid structure is conserved for one of the candidates

Encouraged by the promising results, we set out to study the homology and evolutionary relevance of C14. This candidate protein was annotated with a 1,3-beta-glucanosyltransferase domain which presented a C-terminal anchoring motif coupled by a disordered stretch (Figure 2.8). A homolog of this protein from *Saccharomyces cerevisiae* encoded by the GAS1 gene was previously found to form amyloids in the same C-DAG system (Ryzhova *et al.*, 2018). This conservation of amyloid-forming characteristics over such a large phylogenetic distance hints at a basic role for amyloid conformation that is key to the functional success of the protein. As suggested in Ryzhova *et al.* (2018), the formation of amyloids encoded by GAS1 could be a way to collect and direct these proteins through templating of their domains to a particular location in the cell wall. In *Albugo* in particular, this accumulation could help direct growth towards the extending pole of the hyphae as it

fills the apoplastic space and maintains the integrity of the cell wall by means of 1,3-beta-glucan bond rearrangements (Mouyna *et al.*, 2000).

In Ryzhova *et al.* (2018), they tested in the C-DAG system a 196 amino acid stretch at the C-terminal end of the protein encoded by the GAS1 gene from *S. cerevisiae*. This region coincides mostly with the domain C14d, which we described in Section 2.3.3 to present the highest antimicrobial activity. When expressing C14d in the C-DAG system, we saw a positive phenotype, however lower than for the full protein. This may indicate that the C-terminus region is partially responsible for the amyloid formation, despite having the least homology conservation to GAS1 from *S. cerevisiae* (Figure 2.13).

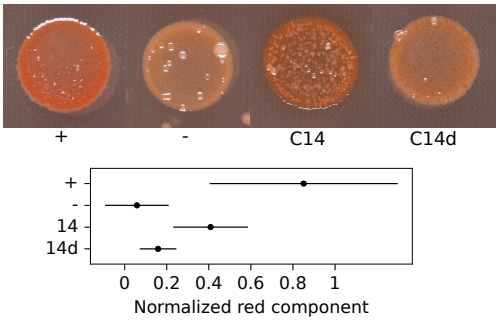


Figure 2.13: Comparison of the colony color phenotype of C14 and C14d in the C-DAG system. Depicted are the positive control (+), Sup35[NM], and the negative control (-), Sup35[M] for amyloid formation. On the bottom, the mean and 95% confidence interval for the normalized red component based on three samples is depicted. This was measured using image analysis software on the rim of each colony, where the red color was more intense.

We constructed a phylogenetic tree to study the presence of this enzyme in oomycetes and compare the phylogenetic distance to *S. cerevisiae* (Appendix C.1 Figure 11). We found that, while most other oomycetes had several copies, this protein family was reduced in the obligate biotrophs, including *Albugo*. This could be indicative of pleiotropy in this enzyme. Alternatively, it could also be related to the lower abundance of 1,3-beta-glucan in the cell wall of biotrophs, as it can act as a microbe-associated molecular pattern (MAMP) that triggers plant defense (Fesel and Zuccaro, 2016). Additionally, the already mentioned genome reduction underwent by *Albugo* could add to this explanation. Interestingly, GAS1 from *S. cerevisiae* has been shown to protect against bacterial infection in aquaculture fish, suggest-

ing also a conservation of its antimicrobial activity (Cornet *et al.*, 2021).

2.5 Amyloid proteins in *Synechococcus elongatus*

Albugo, as described in Section 1.1, cannot be cultured *in vitro* due to its obligate biotrophic lifestyle. It requires living plant leaf material and a constant state of infection to be maintained in a laboratory setting. This makes it a difficult system to study, as well as to develop new synthetic biology applications, for example, as a producer of amyloids or antimicrobials. Other systems, which share a similar evolutionary history of marine autotrophy, like cyanobacteria, have a long history of proven use cases in biotechnology. Therefore, when looking into potential applications of these proteins and translation into other contexts, we may study such systems.

A property that is associated both with increased pathogenicity and the production of amyloids is the formation of biofilms. Many bacterial genera, such as *Pseudomonas, Bacillus* and *Escherichia,* have dedicated systems for the production and secretion of amyloid proteins that form the essential backbone of their biofilms (Taglialegna *et al.*, 2016). Here, we studied putative matrix proteins released by the cyanobacterium *Synechococcus elongatus* when in biofilm-forming stage for their potential of amyloid formation. Interestingly, these proteins are distant relatives of microcins, which are also small proteins characterized by their double-glycine secretion motif (Parnasa *et al.*, 2019). Microcins are antimicrobial proteins that are secreted during competition with other bacteria, which deepens the connection between both amyloids and antimicrobials as seen in literature and also observed in some *Albugo* candidate proteins, e.g., C14 (Baquero *et al.*, 2019).

We used the *in silico* prediction pipeline and the C-DAG system mentioned in the previous section to study the amyloid-forming properties of the four proteins encoded by the operon *ebfG*. This operon is significantly upregulated during non-planktonic stage, and the corresponding secreted proteins have been shown to be necessary for this phenotype (Schatz *et al.*, 2013). The *ebfG* encodes four short proteins and each has a double glycine motif near its N-terminus, which is cleaved during secretion to release the mature protein to the extracellular space (Parnasa *et al.*, 2016).

2.5.1 Amyloid hotspots are conserved in EbfG proteins

For the assessment of the formation of amyloids in each of the four *S. elongatus* EbfG proteins, we used the same prediction pipeline as described previously in Section 2.4.3. However, this time, we included the aggregation prediction from the software TANGO, which looks at the physical properties of the proteins, to try to further reduce false positives which we obtained with the *Albugo* proteins (Fernandez-Escamilla *et al.*, 2004). We found EbfG2 and 3 to score above the 0.5 cutoff for amyloid formation. Following, we used a series of amyloid predictors with different predictive background, either based on statistical analysis, machine learning, structure prediction or a combination thereof (Oliveberg, 2010; Emily *et al.*, 2013; Walsh *et al.*, 2014; Ahmed *et al.*, 2015; Família *et al.*, 2015; Wojciechowski and Kotulska, 2020). These give a score per amino acid so that we were able to pinpoint the peptide core likely responsible for amyloid aggregation in both proteins. When applied on the EbfG proteins, the regions overlapped for the alignment of EbfG2 and 3, thus suggesting their homologous nature. However, they were distinct in their particular sequence, but similar in terms of the physical characteristics of the residues (Appendix B.2 Figure 8a). This could indicate cross-interaction when forming amyloids, as imperfect repeats are characteristic of a more stable fibril morphology (Christensen *et al.*, 2020).

2.5.2 EbfG proteins show amyloid formation when expressed heterologously

Again, we employed the C-DAG system for *in vivo* expression of the peptides in a heterologous system to substantiate the predictions. In this case, we separately cloned the mature EbfG proteins, that is, from the GG motif up to the C-terminus, which was fused to a 6x histidine tag for downstream identification. After expression of the peptides in this system, we found that, in agreement with their predictions, EbfG2 and 3 showed a red color colony phenotype due to the binding of CR from the media, indicative of their amyloid nature. EbfG1, although not predicted amyloid, also showed a similar, albeit weaker, positive phenotype. Finally, consistent with its prediction, EbfG4 showed no amyloidogenic phenotype in the CR plate tests, comparable to that of the negative control (Appendix B.2 Figures 8b and 8c). Such variable result for the formation of amyloids in the EbfG proteins is

striking since they all share a similar sequence and they all partially recover the phenotype for biofilm formation, however, only the four of them together do so completely, suggesting synergistic interaction (Parnasa *et al.*, 2016).

Fibril-like structures from EbfG2 bind colloidal gold

In order to observe the potential fibrils produced by the EbfG proteins, we used the candidate with the stronger amyloid phenotype for CR binding in inducible plates, EbfG2, as a target in TEM experiments. Under the microscope, we observed fibril-like structures in the extracellular space of EbfG2-producing *E. coli* cells, however, much shorter than those from the positive control of Sup35[NM] (Appendix B.2 Figure 9). Using an immunostaining approach, consisting in the use of golden nano beads bound to an antibody directed to the primary polyhistidine-tag antibody, we observed the colocalization of the golden probes to these fibril-like structures, verifying that they belonged to the protein candidate. The staining was sparse, which, together with increased labeling of amorphous regions, indicated a general unavailability of the tag for successful antigen binding when the protein was forming amyloids, as observed also in the positive control and previous experiments with *Albugo* proteins (Section 2.4.3). Our results are comparable to the level of staining from similar studies as well, indicating a general difficulty in antibody binding of the tag when the protein is in amyloid conformation (Antonets *et al.*, 2020).

2.5.3 EbfG proteins assemble into amyloids

Based on the evidence that all four EbfG proteins are relevant for the biofilm phenotype and that they all share homology of certain domains which are highly amyloidogenic in EbfG2 and 3, we hypothesize that there may be cross-seeding (Parnasa *et al.*, 2016). That is, formation of fibrils consisting of different cross-beta domains, which result in heterogeneous amyloid matrix formation. As is the case for other biofilm-forming bacteria, the amyloid backbone proteins are not homogeneous in their sequence. Rather, they are composed of domains that vary in their composition and play different roles during the nucleation and seeding processes which ultimately lead to mature fibrils. For example, in *E. coli*, CsgB performs the function of nucleator, triggering the polymerization of CsgA, which by itself

does not spontaneously assemble into curli fibrils (Shu *et al.*, 2012). This sort of system, which is widespread among amyloid-forming bacteria, has likely evolved in response to adaptation to a controlled production of amyloids, which can easily get out of control and overwhelm the folding and proteolytic machinery of the cell. Such mechanisms could also be generalized to the amyloids in *Albugo*, which, despite the distant phylogeny, share an ancestral lifestyle. It could explain why we did not find strong amyloids when testing candidates one at a time (Section 2.3.3).

2.6 Conclusions

In this dissertation, we have studied protein-based interactions among microbes, firstly, in the context of the phyllosphere with *Albugo*, and lastly, with *Synechococcus*. As a first step, we performed a comparative evolutionary study, from which we obtained a good overview of the adaptation of proteins from organisms and their relevance for lifestyle. Using the proteomics of Section 2.2 as an outset, we developed and applied two pipelines for the prediction of antimicrobial and amyloid activity. In Section 2.3.1, we described two antibacterial peptides that were capable of selectively inhibiting plant-isolated bacteria. Microbiome-modifying effectors have been studied in some pathogenic fungi, however, they were here reported for the first time in a protist. On the amyloid side discussed in Section 2.4.2, we were able to find evidence for amyloid folding in many of the putative candidates of *Albugo*. Therefore, we conclude that *in silico* methods for protein classification, such as the consensus predictors implemented here, are a promising tool to find novel antimicrobial and amyloid proteins in members of other understudied complex microbial communities. The abundance of amyloid-like proteins in the small number of tested proteins suggests a more general prevalence of amyloids than what has been thought up to now based on the literature for the phylogenetic group of protists. Finally, we studied native amyloids in the cyanobacteria *S. elongatus*, due to its potential uses in biotechnological tools. From this study, we reported for the first time amyloid proteins enabling biofilm formation in a cyanobacterium. Although the underlying mechanisms have yet to be fully elucidated, we have overall made great advances in gathering a better understanding of the roles of proteins in microbial interactions.

Bibliography

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**(1), 1–31.

Ahmed, A. B., Znassi, N., Château, M.-T., and Kajava, A. V. (01/06/2015). A structure-based approach to predict predisposition to amyloidosis. *Alzheimer's & Dementia*, **11**(6), 681–690.

Albuquerque, L., Simões, C., Nobre, M. F., Pino, N. M., Battista, J. R., Silva, M. T., Rainey, F. A., and Costa, M. S. (2005). Truepera radiovictrix gen. nov., sp. nov., a new radiation resistant species and the proposal of Trueperaceae fam. nov. *FEMS Microbiology Letters*, **247**(2), 161–169.

Almario, J., Mahmoudi, M., Kroll, S., Agler, M., Placzek, A., Mari, A., and Kemen, E. (2022). The Leaf Microbiome of Arabidopsis Displays Reproducible Dynamics and Patterns throughout the Growing Season. *mBio*, pages e02825–21.

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., and Singh, K. B. (2010). Plants versus pathogens: an evolutionary arms race. *Functional Plant Biology*, **37**(6), 499–512.

Andersson, J. O. (2006). Convergent Evolution: Gene Sharing by Eukaryotic Plant Pathogens. *Current Biology*, **16**(18), R804–R806.

Antonets, K. S. and Nizhnikov, A. A. (2017). Amyloids and prions in plants: Facts and perspectives. *Prion*, **11**(5), 300–312. PMID: 28960135.

Antonets, K. S., Belousov, M. V., Sulatskaya, A. I., Belousova, M. E., Kosolapova, A. O., Sulatsky, M. I., Andreeva, E. A., Zykin, P. A., Malovichko, Y. V., Shtark, O. Y., Lykholay, A. N., Volkov, K. V., Kuznetsova, I. M., Turoverov, K. K., Kochetkova, E. Y., Bobylev, A. G., Usachev, K. S., Demidov, O. N., Tikhonovich, I. A., and Nizhnikov, A. A. (2020). Accumulation of storage proteins in plant seeds is mediated by amyloid formation. *PLOS Biology*, **18**(7), e3000564.

Azevedo, R., Lopes, J. L., Souza, M. M. d., Quirino, B. F., Cançado, L. J., and Marins, L. F. (2019). Synechococcus elongatus as a model of photosynthetic bioreactor for expression of recombinant β-glucosidases. *Biotechnology for Biofuels*, **12**(1), 174.

Babu, M. M., Lee, R. v. d., Groot, N. S. d., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology*, **21**(3), 432–440.

Bacon, C. W. and White, J. F. (2016). Functions, mechanisms and regulation of endophytic and epiphytic microbial communities of plants. *Symbiosis*, **68**(1-3), 87–98.

Balistreri, A., Goetzler, E., and Chapman, M. (2020). Functional Amyloids Are the Rule Rather Than the Exception in Cellular Biology. *Microorganisms*, **8**(12), 1951.

Baquero, F., Lanza, V. F., Baquero, M.-R., Campo, R. d., and Bravo-Vázquez, D. A. (2019). Microcins in Enterobacteriaceae: Peptide Antimicrobials in the Eco-Active Intestinal Chemosphere. *Frontiers in Microbiology*, **10**, 2261.

Bar-On, Y. M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, **115**(25), 201711842.

Barros, A., Gonçalves, A., and Simões, M. (2020). Recent Trends in Biofilm Science and Technology. pages 127–156.

Beakes, G. W. and Thines, M. (2017). Handbook of the Protists. pages 435–505.

Bebber, D. P. and Gurr, S. J. (2015). Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genetics and Biology*, **74**, 62–64.

Bertini, L., Proietti, S., Aleandri, M. P., Mondello, F., Sandini, S., Caporale, C., and Caruso, C. (2012). Modular structure of HEL protein from Arabidopsis reveals new potential functions for PR-4 proteins. *bchm*, **393**(12), 1533–1546.

Bonneaud, C. and Longdon, B. (2020). Emerging pathogen evolution. *EMBO Reports*, **21**(9), e51374.

Booker, T. R., Jackson, B. C., and Keightley, P. D. (2017). Detecting positive selection in the genome. *BMC Biology*, **15**(1), 98.

Bradshaw, J. P. (2003). Cationic Antimicrobial Peptides. *BioDrugs*, **17**(4), 233–240.

Braga, R. M., Dourado, M. N., and Araújo, W. L. (2016). Microbial interactions: ecology in a molecular perspective. *Brazilian Journal of Microbiology*, **47**(Suppl 1), 86–98.

Brenya, E., Chen, Z.-H., Tissue, D., Papanicolaou, A., and Cazzonelli, C. I. (2020). Prior exposure of Arabidopsis seedlings to mechanical stress heightens jasmonic acid-mediated defense against necrotrophic pathogens. *BMC Plant Biology*, **20**(1), 548.

Broughton, R., Newsham, K., Hill, P., Stott, A., and Jones, D. (2015). Differential acquisition of amino acid and peptide enantiomers within the soil microbial community and its implications for carbon and nitrogen cycling in soil. *Soil Biology and Biochemistry*, **88**, 83–89.

Burdukiewicz, M., Sobczyk, P., Rödiger, S., Duda-Madej, A., Mackiewicz, P., and Kotulska, M. (01/12/2017). Amyloidogenic motifs revealed by n-gram analysis. *Scientific Reports*, **7**(1), 12961.

Burdukiewicz, M., Sidorczuk, K., Rafacz, D., Pietluch, F., Chilimoniuk, J., Rödiger, S., and Gagat, P. (2020). Proteomic Screening for Prediction and Design of Antimicrobial Peptides with AmpGram. *International Journal of Molecular Sciences*, **21**(12), 4310.

Burmølle, M., Webb, J. S., Rao, D., Hansen, L. H., Sørensen, S. J., and Kjelleberg, S. (2006). Enhanced Biofilm Formation and Increased Resistance to Antimicrobial Agents and Bacterial Invasion Are Caused by Synergistic Interactions in Multispecies Biofilms†. *Applied and Environmental Microbiology*, **72**(6), 3916–3923.

Chandra, N. and Kumar, S. (2017). Antibiotics and Antibiotics Resistance Genes in Soils, Monitoring, Toxicity, Risk Assessment and Management. *Soil Biology*, pages 1–18.

Chandrasekaran, M., Thangavelu, B., Chun, S. C., and Sathiyabama, M. (2016). Proteases from phytopathogenic fungi and their importance in phytopathogenicity. *Journal of General Plant Pathology*, **82**(5), 233–239.

Chinchilla, D., Bruisson, S., Meyer, S., Zühlke, D., Hirschfeld, C., Joller, C., L'Haridon, F., Mène-Saffrané, L., Riedel, K., and Weisskopf, L. (2019). A sulfur-containing volatile emitted by potato-associated bacteria confers protection against late blight through direct anti-oomycete activity. *Scientific Reports*, **9**(1), 18778.

Choi, B., Yoon, G., Lee, S. W., and Eom, K. (2014). Mechanical deformation mechanisms and properties of amyloid fibrils. *Physical Chemistry Chemical Physics*, **17**(2), 1379–1389.

Chollet, F. *et al.* (2015). Keras.

Christensen, L. F. B., Nowak, J. S., Sønderby, T. V., Frank, S. A., and Otzen, D. E. (2020). Quantitating denaturation by formic acid: Imperfect repeats are essential to the stability of the functional amyloid protein FapC. *bioRxiv*, page 2020.03.09.983882.

Cicconardi, F., Marcatili, P., Arthofer, W., Schlick-Steiner, B. C., and Steiner, F. M. (2017). Positive diversifying selection is a pervasive adaptive force throughout the Drosophila radiation. *Molecular Phylogenetics and Evolution*, **112**, 230–243.

Conde, A., Chaves, M. M., and Gerós, H. (2011). Membrane Transport, Sensing and Signaling in Plant Adaptation to Environmental Stress. *Plant and Cell Physiology*, **52**(9), 1583–1602.

Cornet, V., Khuyen, T. D., Mandiki, S. N. M., Betoulle, S., Bossier, P., Reyes-López, F. E., Tort, L., and Kestemont, P. (2021). GAS1: A New β-Glucan Immunostimulant Candidate to Increase Rainbow Trout (Oncorhynchus mykiss) Resistance to Bacterial Infections With Aeromonas salmonicida achromogenes. *Frontiers in Immunology*, **12**, 693613.

Corvey, C., Stein, T., Düsterhus, S., Karas, M., and Entian, K.-D. (2003). Activation of subtilin precursors by Bacillus subtilis extracellular serine proteases subtilisin (AprE), WprA, and Vpr. *Biochemical and Biophysical Research Communications*, **304**(1), 48–54.

Cycoń, M., Mrozik, A., and Piotrowska-Seget, Z. (2019). Antibiotics in the Soil Environment—Degradation and Their Impact on Microbial Activity and Diversity. *Frontiers in Microbiology*, **10**, 338.

Dahlin, P., Srivastava, V., Ekengren, S., McKee, L. S., and Bulone, V. (2017). Comparative analysis of sterol acquisition in the oomycetes Saprolegnia parasitica and Phytophthora infestans. *PLOS ONE*, **12**(2), e0170873.

Darwin, C. (1860). On the origin of species by means of natural selection: Or the preservation of the favoured races in the struggle for life.

Devoto, A., Ellis, C., Magusin, A., Chang, H.-S., Chilcott, C., Zhu, T., and Turner, J. G. (2005). Expression profiling reveals COI1 to be a key regulator of genes involved in wound- and methyl jasmonate-induced secondary metabolism, defence, and hormone interactions. *Plant Molecular Biology*, **58**(4), 497–513.

Diepeveen, E. T., Gehrmann, T., Pourquié, V., Abeel, T., and Laan, L. (2018). Patterns of Conservation and Diversification in the Fungal Polarization Network. *Genome Biology and Evolution*, **10**(7), 1765–1782.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher*, **35**(3), 125–129.

Dong, S., Raffaele, S., and Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, **35**, 57–65.

D'Souza, G., Waschina, S., Pande, S., Bohl, K., Kaleta, C., and Kost, C. (2014). Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution*, **68**(9), 2559–2570.

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U., Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearson, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Peer, Y. V. d., Rouzé, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J., and Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, **108**(22), 9166–9171.

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.

Emenecker, R. J., Griffith, D., and Holehouse, A. S. (2021). metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophysical Journal*, **120**(20), 4312–4319.

Emily, M., Talvas, A., and Delamarche, C. (2013). MetAmyl: A METa-Predictor for AMYLoid Proteins. *PLoS ONE*, **8**(11), e79722.

Família, C., Dennison, S. R., Quintas, A., and Phoenix, D. A. (04/08/2015). Prediction of Peptide and Protein Propensity for Amyloid Formation. *PLOS ONE*, **10**(8), e0134679.

Faverie, A. R. d. l., Guédin, A., Bedrat, A., Yatsunyk, L. A., and Mergny, J.-L. (2014). Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Research*, **42**(8), e65–e65.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (01/01/2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology*, **22**(10), 1302–1306.

Fesel, P. H. and Zuccaro, A. (2016). β-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants. *Fungal Genetics and Biology*, **90**, 53–60.

Fischer, B., Sumner, I., and Goodenough, P. (1993). Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies. *Biotechnology and Bioengineering*, **41**(1), 3–13.

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, **52**(2), 399–433.

Friesen, M. L. (2020). Social Evolution and Cheating in Plant Pathogens. *Annual Review of Phytopathology*, **58**(1), 1–21.

Gadhave, K. and Giri, R. (2020). Amyloid formation by intrinsically disordered trans-activation domain of cMyb. *Biochemical and Biophysical Research Communications*, **524**(2), 446–452.

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., Hoyer, J. S., Lin, A., Liu, S., Lizárraga, C., Lorence, A., Miller, M., Platon, E., Tessman, M., and Sax, T. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ*, **5**, e4088.

Geilfus, C.-M. (2017). The pH of the Apoplast: Dynamic Factor with Functional Impact Under Stress. *Molecular Plant*, **10**(11), 1371–1386.

Gerstel, U., Latendorf, T., Bartels, J., Becker, A., Tholey, A., and Schröder, J.-M. (2018). Hornerin contains a Linked Series of Ribosome-Targeting Peptide Antibiotics. *Scientific Reports*, **8**(1), 16158.

Gerven, N. V., Verren, S. E. V. d., Reiter, D. M., and Remaut, H. (01/10/2018). The Role of Functional Amyloids in Bacterial Virulence. *Journal of Molecular Biology*, **430**(20), 3657–3684.

Gómez-Pérez, D., Chaudhry, V., Kemen, A., and Kemen, E. (2021). Amyloid Proteins in Plant-Associated Microbial Communities. *Microbial Physiology*, pages 1–11.

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**(6), 333–351.

Gour, S., Kaushik, V., Kumar, V., Bhat, P., Yadav, S. C., and Yadav, J. K. (2016). Antimicrobial peptide (Cn-AMP2) from liquid endosperm of Cocos nucifera forms amyloid-like fibrillar structure. *Journal of Peptide Science*, **22**(4), 201–207.

Gras, S. L., Waddington, L. J., and Goldie, K. N. (2011). Protein Folding, Misfolding, and Disease, Methods and Protocols. *Methods in Molecular Biology*, **752**, 197–214.

Gómez-Pérez, D., Schmid, M., Chaudhry, V., Velic, A., cek, B. M., Kemen, A., and Kemen, E. (2022). Proteins released into the plant apoplast by the obligate parasitic protist Albugo selectively repress phyllosphere-associated bacteria. *bioRxiv*.

Haldane, J. B. S. (1927). A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **23**(7), 838–844.

Hammer, N. D., Schmidt, J. C., and Chapman, M. R. (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proceedings of the National Academy of Sciences*, **104**(30), 12494–12499.

Hancock, R. E. W. and Scott, M. G. (2000). The role of antimicrobial peptides in animal defenses. *Proceedings of the National Academy of Sciences*, **97**(16), 8856–8861.

Hane, J. K., Paxman, J., Jones, D. A. B., Oliver, R. P., and Wit, P. d. (2020). "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens – How Many Different Types of Filamentous Plant Pathogens Are There? *Frontiers in Microbiology*, **10**, 3088.

Harrison, E. and Brockhurst, M. A. (2012). Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology*, **20**(6), 262–267.

Hayat, R., Ali, S., Amara, U., Khalid, R., and Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology*, **60**(4), 579–598.

Hill, R., Buggs, R. J., Vu, D. T., and Gaya, E. (2022). Lifestyle Transitions in Fusarioid Fungi are Frequent and Lack Clear Genomic Signatures. *Molecular Biology and Evolution*, **39**(4), msac085–.

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., and Kurgan, L. (2021). flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications*, **12**(1), 4438.

Ittisoponpisan, S., Alhuzimi, E., Sternberg, M. J. E., and David, A. (2017). Landscape of Pleiotropic Proteins Causing Human Disease: Structural and System Biology Insights. *Human Mutation*, **38**(3), 289–296.

Jashni, M. K., Mehrabi, R., Collemare, J., Mesarich, C. H., and Wit, P. J. G. M. d. (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.

Judelson, H. S. (2012). Dynamics and Innovations within Oomycete Genomes: Insights into Biology, Pathology, and Evolution. *Eukaryotic Cell*, **11**(11), 1304–1312.

Judelson, H. S. (2016). Metabolic Diversity and Novelties in the Oomycetes. *Annual Review of Microbiology*, **71**(1), 1–19.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), 583–589.

Kagan, B. L., Jang, H., Capone, R., Arce, F. T., Ramachandran, S., Lal, R., and Nussinov, R. (01/04/2012). Antimicrobial Properties of Amyloid Peptides. *Molecular Pharmaceutics*, **9**(4), 708–717.

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., and Jones, J. D. G. (05/07/2011). Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. *PLoS Biology*, **9**(7), e1001094.

Keymanesh, K., Soltani, S., and Sardari, S. (2009). Application of antimicrobial peptides in agriculture and food industry. *World Journal of Microbiology and Biotechnology*, **25**(6), 933–944.

Kim, K., Yadav, D., and Cho, M. (2019). Multi-phased internalization of murine norovirus (MNV) in Arabidopsis seedlings and its potential correlation with plant defensive responses. *Microbial Pathogenesis*, **135**, 103648.

Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, **267**(5608), 275–276.

Knight, C. A., Molinari, N. A., and Petrov, D. A. (2005). The Large Genome Constraint Hypothesis: Evolution, Ecology and Phenotype. *Annals of Botany*, **95**(1), 177–190.

Koeck, M., Hardham, A. R., and Dodds, P. N. (2011). The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cellular Microbiology*, **13**(12), 1849–1857.

Komatsubara, A. T., Goto, Y., Kondo, Y., Matsuda, M., and Aoki, K. (2019). Single-cell quantification of the concentrations and dissociation constants of endogenous proteins. *Journal of Biological Chemistry*, **294**(15), 6062–6072.

Konings, W. N. (2006). Microbial transport: Adaptations to natural environments. *Antonie van Leeuwenhoek*, **90**(4), 325–342.

Korsak, M. and Kozyreva, T. (2015). Intrinsically Disordered Proteins Studied by NMR Spectroscopy. *Advances in Experimental Medicine and Biology*, **870**, 401–421.

Köster, J. and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, **28**(19), 2520–2522.

Kuntner, M., Năpăruş, M., Li, D., and Coddington, J. A. (2014). Phylogeny Predicts Future Habitat Shifts Due to Climate Change. *PLoS ONE*, **9**(6), e98907.

Laland, K. N. and O'Brien, M. J. (2010). Niche Construction Theory and Archaeology. *Journal of Archaeological Method and Theory*, **17**(4), 303–322.

Latendorf, T., Gerstel, U., Wu, Z., Bartels, J., Becker, A., Tholey, A., and 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**(1), 3331.

Latijnhouwers, M., Wit, P. J. d., and Govers, F. (2003). Oomycetes and fungi: similar weaponry to attack plants. *Trends in Microbiology*, **11**(10), 462–469.

Lawrence, T. J., Carper, D. L., Spangler, M. K., Carrell, A. A., Rush, T. A., Minter, S. J., Weston, D. J., and Labbé, J. L. (2020). amPEPpy 1.0: a portable and accurate antimicrobial peptide prediction tool. *Bioinformatics*, **37**(14), 2058–2060.

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**(1), 124.

Lee, T.-H., Hall, K. N., and Aguilar, M.-I. (2015). Antimicrobial Peptide Structure and Mechanism of Action: A Focus on the Role of Membrane Structure. *Current Topics in Medicinal Chemistry*, **16**(1), 25–39.

Levin, S. A. (2014). Public goods in relation to competition, cooperation, and spite. *Proceedings of the National Academy of Sciences*, **111**(Supplement 3), 10838–10845.

Liu, L., Xu, L., Jia, Q., Pan, R., Oelmüller, R., Zhang, W., and Wu, C. (2019). Arms race: diverse effector proteins with conserved motifs. *Plant Signaling & Behavior*, **14**(2), 1557008.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, **11**(1), 753.

Lum, G. and Min, X. J. (2011). FunSecKB: the Fungal Secretome KnowledgeBase. *Database*, **2011**(0), bar001–bar001.

Lyou, S. H., Park, H. J., Jung, C., Sohn, H. B., Lee, G., Kim, C. H., Kim, M., Choi, Y. D., and Cheong, J.-J. (2009). The Arabidopsis AtLEC gene encoding a lectin-like protein is up-regulated by multiple stimuli including developmental signal, wounding, jasmonate, ethylene, and chitin elicitor. *Molecules and Cells*, **27**(1), 75–81.

Lévesque, C. A. (2011). Fifty years of oomycetes—from consolidation to evolutionary and genomic exploration. *Fungal Diversity*, **50**(1), 35.

Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., and Patil, K. R. (2021). Polarization of microbial communities between competitive and cooperative metabolism. *Nature Ecology & Evolution*, **5**(2), 195–203.

Malanovic, N. and Lohner, K. (2016). Antimicrobial Peptides Targeting Gram-Positive Bacteria. *Pharmaceuticals*, **9**(3), 59.

Malik, E., Dennison, S. R., Harris, F., and Phoenix, D. A. (2016). pH Dependent Antimicrobial Peptides and Proteins, Their Mechanisms of Action and Potential as Therapeutic Agents. *Pharmaceuticals*, **9**(4), 67.

Mandala, V. S., Williams, J. K., and Hong, M. (2016). Structure and Dynamics of Membrane Proteins from Solid-State NMR. *Annual Review of Biophysics*, **47**(1), 1–22.

Marín, M., Uversky, V. N., and Ott, T. (2013). Intrinsic Disorder in Pathogen Effectors: Protein Flexibility as an Evolutionary Hallmark in a Molecular Arms Race. *The Plant Cell*, **25**(9), 3153–3157.

Masson-Boivin, C. and Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia—the roots of a success story. *Current Opinion in Plant Biology*, **44**, 7–15.

Mehrabi, R., Bahkali, A. H., Abd-Elsalam, K. A., Moslem, M., M'Barek, S. B., Gohari, A. M., Jashni, M. K., Stergiopoulos, I., Kema, G. H., and Wit, P. J. d. (2011). Horizontal gene and chromosome transfer in plant pathogenic fungi affecting host range. *FEMS Microbiology Reviews*, **35**(3), 542–554.

Meister, M., Lemaitre, B., and Hoffmann, J. A. (1997). Antimicrobial peptide defense in Drosophila. *BioEssays*, **19**(11), 1019–1026.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden.

Mendoza, L. and Vilela, R. (2013). The Mammalian Pathogenic Oomycetes. *Current Fungal Infection Reports*, **7**(3), 198–208.

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences*, **112**(24), E3095–E3103.

Misner, I., Blouin, N., Leonard, G., Richards, T. A., and Lane, C. E. (01/01/2015). The Secreted Proteins of Achlya hypogyna and Thraustotheca clavata Identify the Ancestral Oomycete Secretome and Reveal Gene Acquisitions by Horizontal Gene Transfer. *Genome Biology and Evolution*, **7**(1), 120–135.

Mitchell-Olds, T. (2013). Selection on QTL and complex traits in complex environments. *Molecular Ecology*, **22**(13), 3427–3429.

Molloy, E. M. and Hertweck, C. (2017). Antimicrobial discovery inspired by ecological interactions. *Current Opinion in Microbiology*, **39**, 121–127.

Montesinos, E. (2007). Antimicrobial peptides and plant disease control. *FEMS Microbiology Letters*, **270**(1), 1–11.

Moreno-Santillán, D. D. and Ortega, J. (2021). 50 Years of Bat Research, Foundations and New Frontiers. *Fascinating Life Sciences*, pages 273–287.

Mouyna, I., Fontaine, T., Vai, M., Monod, M., Fonzi, W. A., Diaquin, M., Popolo, L., Hartland, R. P., and Latgé, J.-P. (2000). Glycosylphosphatidylinositol-anchored Glucanosyltransferases Play an Active Role in the Biosynthesis of the Fungal Cell Wall*. *Journal of Biological Chemistry*, **275**(20), 14882–14889.

Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., Monachello, D., Boxem, M., Harbort, C. J., McDonald, N., Gai, L., Chen, H., He, Y., Consortium, E. U. E., Vandenhaute, J., Roth, F. P., Hill, D. E., Ecker, J. R., Vidal, M., Beynon, J., Braun, P., and Dangl, J. L. (2011). Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science*, **333**(6042), 596–601.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., and Scheffler, K. (2013). FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution*, **30**(5), 1196–1205.

Necci, M., Piovesan, D., Hoque, M. T., Walsh, I., Iqbal, S., Vendruscolo, M., Sormanni, P., Wang, C., Raimondi, D., Sharma, R., Zhou, Y., Litfin, T., Galzitskaya, O. V., Lobanov, M. Y., Vranken, W., Wallner, B., Mirabello, C., Malhis, N., Dosztányi, Z., Erdős, G., Mészáros, B., Gao, J., Wang, K., Hu, G., Wu, Z., Sharma, A., Hanson, J., Paliwal, K., Callebaut, I., Bitard-Feildel, T., Orlando, G., Peng, Z., Xu, J., Wang, S., Jones, D. T., Cozzetto, D., Meng, F., Yan, J., Gsponer, J., Cheng, J., Wu, T., Kurgan, L., Promponas, V. J., Tamana, S., Marino-Buslje, C., Martínez-Pérez, E., Chasapi, A., Ouzounis, C., Dunker, A. K., Kajava, A. V., Leclercq, J. Y., Aykac-Fas, B., Lambrughi, M., Maiani, E., Papaleo, E., Chemes, L. B., Álvarez, L., González-Foutel, N. S., Iglesias, V., Pujols, J., Ventura, S., Palopoli, N., Benítez, G. I., Parisi, G., Bassot, C., Elofsson, A., Govindarajan, S., Lamb, J., Salvatore, M., Hatos, A., Monzon, A. M., Bevilacqua, M., Mičetić, I., Minervini, G., Paladin, L., Quaglia, F., Leonardi, E., Davey, N., Horvath, T., Kovacs, O. P., Murvai, N., Pancsa, R., Schad, E., Szabo, B., Tantos, A., Macedo-Ribeiro, S., Manso, J. A., Pereira, P. J. B., Davidović, R., Veljkovic, N., Hajdu-Soltész, B., Pajkos, M., Szaniszló, T., Guharoy, M., Lazar, T., Macossay-Castillo, M., Tompa, P., and Tosatto, S. C. E. (2021). Critical assessment of protein intrinsic disorder prediction. *Nature Methods*, **18**(5), 472–481.

Nishad, R., Ahmed, T., Rahman, V. J., and Kareem, A. (2020). Modulation of Plant Defense System in Response to Microbial Interactions. *Frontiers in Microbiology*, **11**, 1298.

O'Connell, K. M. G., Hodgkinson, J. T., Sore, H. F., Welch, M., Salmond, G. P. C., and Spring, D. R. (2013). Combating Multidrug-Resistant Bacteria: Current Strategies for the Discovery of Novel Antibacterials. *Angewandte Chemie International Edition*, **52**(41), 10706–10733.

Oldfield, C. J., Uversky, V. N., Dunker, A. K., and Kurgan, L. (2019). Intrinsically Disordered Proteins. pages 1–34.

Oliveberg, M. (2010). Waltz, an exciting new move in amyloid prediction. *Nature Methods*, **7**(3), 187–188.

Omardien, S., Brul, S., and Zaat, S. A. J. (2016). Antimicrobial Activity of Cationic Antimicrobial Peptides against Gram-Positives: Current Progress Made in Understanding the Mode of Action and the Response of Bacteria. *Frontiers in Cell and Developmental Biology*, **4**, 111.

Orgogozo, V., Morizot, B., and Martin, A. (2015). The differential view of genotype–phenotype relationships. *Frontiers in Genetics*, **6**, 179.

Østman, B., Hintze, A., and Adami, C. (2012). Impact of epistasis and pleiotropy on evolutionary adaptation. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1727), 247–256.

Otzen, D. and Riek, R. (2019). Functional Amyloids. *Cold Spring Harbor Perspectives in Biology*, **11**(12), a033860.

Pan, K. and Zhong, Q. (2015). Amyloid-like fibrils formed from intrinsically disordered caseins: physicochemical and nanomechanical properties. *Soft Matter*, **11**(29), 5898–5904.

Parnasa, R., Nagar, E., Sendersky, E., Reich, Z., Simkovsky, R., Golden, S., and Schwarz, R. (2016). Small secreted proteins enable biofilm development in the cyanobacterium Synechococcus elongatus. *Scientific Reports*, **6**(1), 32209.

Parnasa, R., Sendersky, E., Simkovsky, R., Ben-Asher, H. W., Golden, S. S., and Schwarz, R. (2019). A microcin processing peptidase-like protein of the cyanobacterium Synechococcus elongatus is essential for secretion of biofilm-promoting proteins. *Environmental Microbiology Reports*, **11**(3), 456–463.

Pathania, R. and Srivastava, S. (2021). Synechococcus elongatus BDU 130192, an Attractive Cyanobacterium for Feedstock Applications: Response to Culture Conditions. *BioEnergy Research*, **14**(3), 954–963.

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., and Sandhu, M. S. (2018). Long reads: their purpose and place. *Human Molecular Genetics*, **27**(R2), R234–R241.

Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**(5), 676–679.

Presti, L. L. and Kahmann, R. (2017). How filamentous plant pathogen effectors are translocated to host cells. *Current Opinion in Plant Biology*, **38**, 19–24.

Raheem, N. and Straus, S. K. (2019). Mechanisms of Action for Antimicrobial Peptides With Antibacterial and Antibiofilm Functions. *Frontiers in Microbiology*, **10**, 2866.

Rajarammohan, S. (2021). Redefining Plant-Necrotroph Interactions: The Thin Line Between Hemibiotrophs and Necrotrophs. *Frontiers in Microbiology*, **12**, 673518.

Richardson, L. J., Rawlings, N. D., Salazar, G. A., Almeida, A., Haft, D. R., Ducq, G., Sutton, G. G., and Finn, R. D. (2018). Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. *Nucleic Acids Research*, **47**(D1), D564–D572.

Rickard, A. H., Stead, A. T., O'May, G. A., Lindsay, S., Banner, M., Handley, P. S., and Gilbert, P. (2005). Adhaeribacter aquaticus gen. nov., sp. nov., a Gram-negative isolate from a potable water biofilm. *International Journal of Systematic and Evolutionary Microbiology*, **55**(2), 821–829.

Rodenburg, S. Y., de Ridder, D., Govers, F., and Seidl, M. F. (2020). Oomycete metabolism is highly dynamic and reflects lifestyle adaptations. *bioRxiv*.

Rodenburg, S. Y. A., Seidl, M. F., Ridder, D. d., and Govers, F. (2021). Uncovering the Role of Metabolism in Oomycete–Host Interactions Using Genome-Scale Metabolic Models. *Frontiers in Microbiology*, **12**, 748178.

Rollins-Smith, L. A., Reinert, L. K., O'Leary, C. J., Houston, L. E., and Woodhams, D. C. (2005). Antimicrobial Peptide Defenses in Amphibian Skin. *Integrative and Comparative Biology*, **45**(1), 137–142.

Romdhane, S., Spor, A., Aubert, J., Bru, D., Breuil, M.-C., Hallin, S., Mounier, A., Ouadah, S., Tsiknia, M., and Philippot, L. (2022). Unraveling negative biotic interactions determining soil microbial community assembly and functioning. *The ISME Journal*, **16**(1), 296–306.

Ruano, G. and Scheuring, D. (2020). Plant Cells under Attack: Unconventional Endomembrane Trafficking during Plant Defense. *Plants*, **9**(3), 389.

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. *Frontiers in Plant Science*, **7**, 820.

Ryzhova, T. A., Sopova, J. V., Zadorsky, S. P., Siniukova, V. A., Sergeeva, A. V., Galkina, S. A., Nizhnikov, A. A., Shenfeld, A. A., Volkov, K. V., and Galkin, A. P. (2018). Screening for amyloid proteins in the yeast proteome. *Current Genetics*, **64**(2), 469–478.

Salinas, N., Tayeb-Fligelman, E., Sammito, M. D., Bloch, D., Jelinek, R., Noy, D., Usón, I., and Landau, M. (2021). The amphibian antimicrobial peptide uperin 3.5 is a cross-α/cross-β chameleon functional amyloid. *Proceedings of the National Academy of Sciences*, **118**(3), e2014442118.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**(12), 5463–7.

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., Thompson, M. J., Balbirnie, M., Wiltzius, J. J. W., McFarlane, H. T., Madsen, A. Ø., Riekel, C., and Eisenberg, D. (2007). Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature*, **447**(7143), 453–457.

Schatz, D., Nagar, E., Sendersky, E., Parnasa, R., Zilberman, S., Carmeli, S., Mastai, Y., Shimoni, E., Klein, E., Yeger, O., Reich, Z., and Schwarz, R. (2013). Self-suppression of biofilm formation in the cyanobacterium Synechococcus elongatus. *Environmental Microbiology*, **15**(6), 1786–1794.

Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C., and Manners, J. M. (2000). Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proceedings of the National Academy of Sciences*, **97**(21), 11655–11660.

Selosse, M., Strullu-Derrien, C., Martin, F. M., Kamoun, S., and Kenrick, P. (2015). Plants, fungi and oomycetes: a 400-million year affair that shapes the biosphere. *New Phytologist*, **206**(2), 501–506.

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). Gene Prediction, Methods and Protocols. *Methods in Molecular Biology*, **1962**, 227–245.

Shafrir, Y., Durell, S. R., Anishkin, A., and Guy, H. R. (2010). Beta-barrel models of soluble amyloid beta oligomers and annular protofibrils. *Proteins: Structure, Function, and Bioinformatics*, **78**(16), 3458–3472.

Shazely, B. E., Yu, G., Johnston, P. R., and Rolff, J. (2020). Resistance Evolution Against Antimicrobial Peptides in Staphylococcus aureus Alters Pharmacodynamics Beyond the MIC. *Frontiers in Microbiology*, **11**, 103.

Shen, D., Li, Q., Sun, P., Zhang, M., and Dou, D. (2017). Intrinsic disorder is a common structural characteristic of RxLR effectors in oomycete pathogens. *Fungal Biology*, **121**(11), 911–919.

Shu, Q., Crick, S. L., Pinkner, J. S., Ford, B., Hultgren, S. J., and Frieden, C. (2012). The E. coli CsgB nucleator of curli assembles to β-sheet oligomers that alter the CsgA fibrillization mechanism. *Proceedings of the National Academy of Sciences*, **109**(17), 6502–6507.

Sivanathan, V. and Hochschild, A. (2012). Generating extracellular amyloid aggregates using E. coli cells. *Genes & Development*, **26**(23), 2659–2667.

Sivanathan, V. and Hochschild, A. (2013). A bacterial export system for generating extracellular amyloid aggregates. *Nature Protocols*, **8**(7), 1381–1390.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Pond, S. L. K. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution*, **32**(5), 1342–1353.

Smith, P. and Schuster, M. (2019). Public goods and cheating in microbes. *Current Biology*, **29**(11), R442–R447.

Snelders, N. C., Rovenich, H., Petti, G. C., Rocafort, M., Berg, G. C. M. v. d., Vorholt, J. A., Mesters, J. R., Seidl, M. F., Nijland, R., and Thomma, B. P. H. J. (2020). Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nature Plants*, **6**(11), 1365–1374.

Sunde, M. and Blake, C. (1997). The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. *Advances in Protein Chemistry*, **50**, 123–159.

Suskiewicz, M. J., Sussman, J. L., Silman, I., and Shaul, Y. (2011). Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Science*, **20**(8), 1285–1297.

Tackmann, J., Rodrigues, J. F. M., and Mering, C. v. (2019). Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell Systems*, **9**(3), 286–296.e8.

Taglialegna, A., Lasa, I., and Valle, J. (2016). Amyloid Structures as Biofilm Matrix Scaffolds. *Journal of Bacteriology*, **198**(19), 2579–2588.

Tam, J. P., Wang, S., Wong, K. H., and Tan, W. L. (2015). Antimicrobial Peptides from Plants. *Pharmaceuticals*, **8**(4), 711–757.

Terauchi, M., Yamagishi, T., Hanyuda, T., and Kawai, H. (2017). Genome-wide computational analysis of the secretome of brown algae (Phaeophyceae). *Marine Genomics*, **32**, 49–59.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A. K., Daughdrill, G. W., and Uversky, V. N. (2013). The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins*, **1**(1), e24360.

Thines, M. (2019). An evolutionary framework for host shifts – jumping ships for survival. *New Phytologist*, **224**(2), 605–617.

Thomma, B. P. H. J., Eggermont, K., Penninckx, I. A. M. A., Mauch-Mani, B., Vogelsang, R., Cammue, B. P. A., and Broekaert, W. F. (1998). Separate jasmonate-dependent and salicylate-dependent defense-response pathways in Arabidopsis are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences*, **95**(25), 15107–15111.

Toruño, T. Y., Stergiopoulos, I., and Coaker, G. (2015). Plant Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in Spatial and Temporal Manners. *Annual Review of Phytopathology*, **54**(1), 1–23.

Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A., and Hamodrakas, S. J. (2013). A Consensus Method for the Prediction of 'Aggregation-Prone' Peptides in Globular Proteins. *PLoS ONE*, **8**(1), e54175.

Veltri, D., Kamath, U., and Shehu, A. (2018). Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, **34**(16), 2740–2747. Antimicrobial peptide scanner v2 citation.

Vickers, T. J. and Beverley, S. M. (2011). Folate metabolic pathways in Leishmania. *Essays in Biochemistry*, **51**, 63–80.

Vorholt, J. A. (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology*, **10**(12), 828–840.

Walkenhorst, W. F., Klein, J. W., Vo, P., and Wimley, W. C. (2013). pH Dependence of Microbe Sterilization by Cationic Antimicrobial Peptides. *Antimicrobial Agents and Chemotherapy*, **57**(7), 3312–3320.

Walsh, I., Seno, F., Tosatto, S. C., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Research*, **42**(W1), W301–W307.

Wandersman, C. (1989). Secretion, processing and activation of bacterial extracellular proteases. *Molecular Microbiology*, **3**(12), 1825–1831.

Wang, Q., Liu, J., and Zhu, H. (2018). Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions. *Frontiers in Plant Science*, **9**, 313.

Wang, Y. and Wang, Y. (2018). Trick or Treat: Microbial Pathogens Evolved Apoplastic Effectors Modulating Plant Susceptibility to Infection. *Molecular Plant-Microbe Interactions®*, **31**(1), 6–12.

Wang, Y., Wang, Y., and Wang, Y. (2020). Apoplastic Proteases - Powerful Weapons Against Pathogen Infection in Plants. *Plant Communications*, **1**(4), 100085.

Wawra, S., Belmonte, R., Löbach, L., Saraiva, M., Willems, A., and West, P. v. (2012). Secretion, delivery and function of oomycete effector proteins. *Current Opinion in Microbiology*, **15**(6), 685–691.

Whipps, J., Hand, P., Pink, D., and Bending, G. (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *Journal of Applied Microbiology*, **105**(6), 1744–1755.

Wojciechowski, J. W. and Kotulska, M. (2020). PATH - Prediction of Amyloidogenicity by Threading and Machine Learning. *Scientific Reports*, **10**(1), 7721.

Wu, C., Wang, Z., Lei, H., Zhang, W., and Duan, Y. (2007a). Dual Binding Modes of Congo Red to Amyloid Protofibril Surface Observed in Molecular Dynamics Simulations. *Journal of the American Chemical Society*, **129**(5), 1225–1232.

Wu, H., Min, J., Ikeguchi, Y., Zeng, H., Dong, A., Loppnau, P., Pegg, A. E., and Plotnikov, A. N. (01/07/2007b). Structure and Mechanism of Spermidine Synthases †. *Biochemistry*, **46**(28), 8331–8339.

Xue, W.-F., Homans, S. W., and Radford, S. E. (2008). Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proceedings of the National Academy of Sciences*, **105**(26), 8926–8931.

Yamasaki, K. and Gallo, R. L. (2007). Antimicrobial peptides in human skin disease. *European journal of dermatology : EJD*, **18**(1), 11–21.

Yang, C., Li, J., Huang, Z., Zhang, X., Gao, X., Zhu, C., Morris, P. F., and Zhang, X. (2020). Structural and catalytic analysis of two diverse uridine phosphorylases in Phytophthora capsici. *Scientific Reports*, **10**(1), 9051.

Yang, F., Lim, G. P., Begum, A. N., Ubeda, O. J., Simmons, M. R., Ambegaokar, S. S., Chen, P. P., Kayed, R., Glabe, C. G., Frautschy, S. A., and Cole, G. M. (2005). Curcumin Inhibits Formation of Amyloid β Oligomers and Fibrils, Binds Plaques, and Reduces Amyloid in Vivo. *Journal of Biological Chemistry*, **280**(7), 5892–5901.

Zhang, L., Tian, L.-H., Zhao, J.-F., Song, Y., Zhang, C.-J., and Guo, Y. (2008). Identification of an Apoplastic Protein Involved in the Initial Phase of Salt Stress Response in Rice Root by Two-Dimensional Electrophoresis. *Plant Physiology*, **149**(2), 916–928.

Zheng, H., Li, H., Zhang, J., Fan, H., Jia, L., Ma, W., Ma, S., Wang, S., You, H., Yin, Z., and Li, X. (2020). Serum amyloid A exhibits pH dependent antibacterial action and contributes to host defense against Staphylococcus aureus cutaneous infection. *Journal of Biological Chemistry*, **295**(9), 2570–2581.

Zheng, L. and Mackrill, J. J. (2016). Calcium Signaling in Oomycetes: An Evolutionary Perspective. *Frontiers in Physiology*, **7**, 123.

Zhou, Y.-y., Shao, W.-l., Liu, Y.-d., Li, X., Shan, X.-y., Jin, X.-b., Gao, J., and Li, W. (2020). Genome-based analysis to understanding rapid resuscitation of cryopreserved anammox consortia via sequential supernatant addition. *Science of The Total Environment*, **744**, 140785.

Zuluaga, A. P., Vega-Arreguín, J. C., Fei, Z., Matas, A. J., Patev, S., Fry, W. E., and Rose, J. K. C. (2016). Analysis of the tomato leaf transcriptome during successive hemibiotrophic stages of a compatible interaction with the oomycete pathogen Phytophthora infestans. *Molecular Plant Pathology*, **17**(1), 42–54.

Appendix A

Published manuscripts

A.1 Amyloid Proteins in Plant-Associated Microbial Communities

Contributions by DGP:

- Designed the initial manuscript.
- Wrote the initial draft, except for the biofilm section which was drafted by VC.
- Edited the final manuscript.
- Collected all the data.
- Created all the figures and tables.

Microbial Physiology

Review Article

Microb Physiol 2021;31:88–98
DOI: 10.1159/000516014

Received: December 31, 2020
Accepted: March 17, 2021
Published online: June 9, 2021

Amyloid Proteins in Plant-Associated Microbial Communities

Daniel Gómez-Pérez Vasvi Chaudhry Ariane Kemen Eric Kemen

ZMBP/IMIT, University of Tübingen, Tübingen, Germany

Keywords
Amyloid · Plant colonization · Microbial survival · Mechanism · Biofilm

Abstract
Amyloids have proven to be a widespread phenomenon rather than an exception. Many proteins presenting the hallmarks of this characteristic beta sheet-rich folding have been described to date. Particularly common are functional amyloids that play an important role in the promotion of survival and pathogenicity in prokaryotes. Here, we describe important developments in amyloid protein research that relate to microbe-microbe and microbe-host interactions in the plant microbiome. Starting with biofilms, which are a broad strategy for bacterial persistence that is extremely important for plant colonization. Microbes rely on amyloid-based mechanisms to adhere and create a protective coating that shelters them from external stresses and promotes cooperation. Another strategy generally carried out by amyloids is the formation of hydrophobic surface layers. Known as hydrophobins, these proteins coat the aerial hyphae and spores of plant pathogenic fungi, as well as certain bacterial biofilms. They contribute to plant virulence through promoting dissemination and infectivity. Furthermore, antimicrobial activity is an interesting outcome of the amyloid structure that has potential application in medicine and agriculture. There are many known antimicrobial amyloids released by animals and plants; however, those produced by bacteria or fungi remain still largely unknown. Finally, we discuss amyloid proteins with a more indirect mode of action in their host interactions. These include virulence-promoting harpins, signaling transduction that functions through amyloid templating, and root nodule bacteria proteins that promote plant-microbe symbiosis. In summary, amyloids are an interesting paradigm for their many functional mechanisms linked to bacterial survival in plant-associated microbial communities.

Introduction

Plants are an important system for the study of microbe-microbe and microbe-host interactions together with their mechanisms. Plants constitute approximately 80% of Earth's total biomass, which makes them the world's largest living surface area [Bar-On et al., 2018]. Furthermore, all plants are ubiquitously colonized by microbes, including bacteria, fungi, and oomycetes, to a variable extent [Beattie and Lindow, 1999; Kandel et al., 2017]. Plant-colonizing microbes thrive on primary and secondary plant-derived metabolites, which include nutrients and protective compounds. In addition, plants

karger@karger.com
www.karger.com/mip

Correspondence to:
Eric Kemen, eric.kemen@zmbp.uni-tuebingen.de

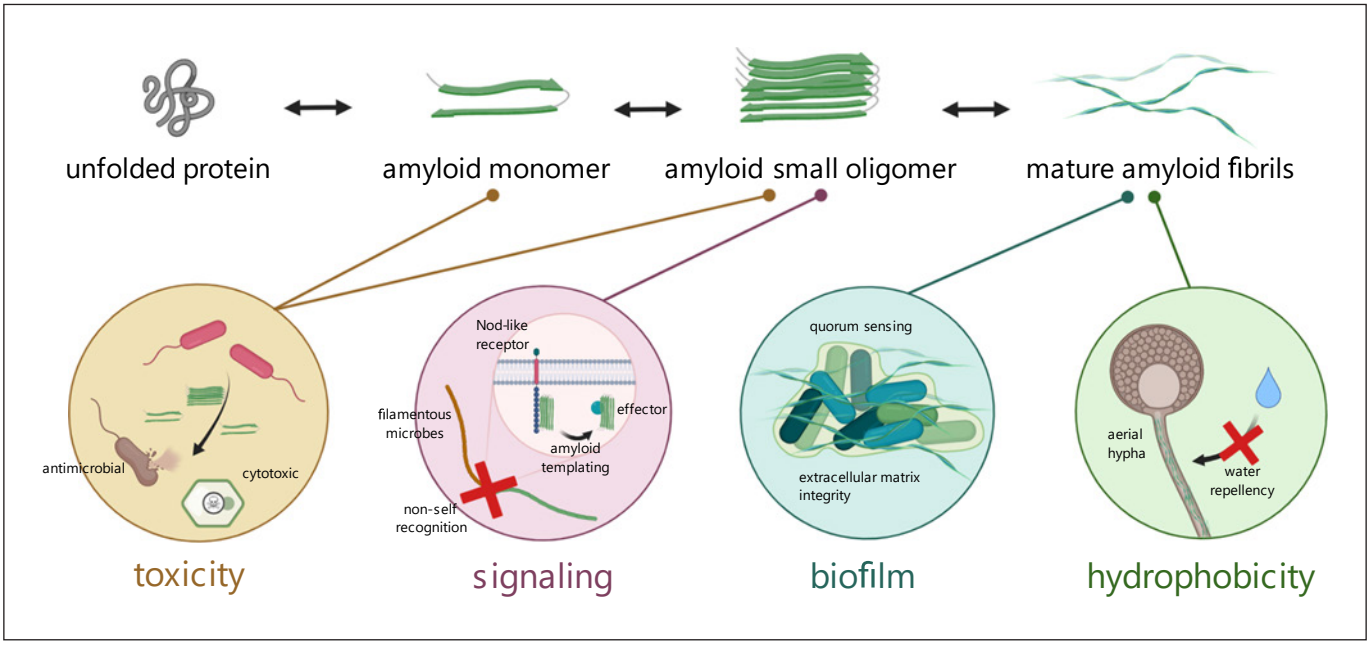


Fig. 1. Different functions of amyloid proteins related to plant microbial communities at different stages of amyloid structural conformations.

provide a niche that defends microbes against biotic and abiotic factors, in both the phyllosphere (above ground) and rhizosphere (below ground) parts of the plant [Mercado-Blanco, 2014]. As a consequence, complex interactions between prokaryotes and eukaryotes have evolved, where fungi and oomycetes are major determinants of the diversity and abundance of plant-associated bacteria [Agler et al., 2016]. This results in competition between microbes for access to the specific plant niches [Anderson et al., 2010; Chaudhry et al., 2021].

On one side of the spectrum, some microbes develop symbiotic relationships that benefit all parties by sharing common goods, which can be defined as secreted metabolites that benefit not just the producer but the whole community [Saikkonen et al., 2004]. On the other side, they can develop antagonistic relationships either toward other microbes or the host. In the first case, either a specific microbe or a number of them are inhibited through physical or chemical mechanisms. As a consequence, the competitors, either directly or indirectly, are denied access to the plant's resources. In the second, a pathogenic relationship develops, which benefits the microbe to the detriment of the host. Thus, pathogenic interactions primarily benefit the pathogen and depending on symptoms to the host can be fatal to the native community. The key to colonization of both beneficial and pathogenic microbes is therefore a robust interaction with the host that can resist perturbations. Crucial mechanisms include biofilm formation and the release of antimicrobial and cytotoxic peptides to enforce niche colonization.

One intriguing class of proteins that is increasingly linked with pathogenicity and microbial survival in plant-associated communities are amyloids. Amyloids are proteins diverse in nature that have a set of common structural properties, the most salient of which is the capacity to polymerize as long unbranched fibrils with convergent characteristics (Fig. 1) [Makin et al., 2005]. These fibrils show a consistent cross beta structure, which consists of two parallel or antiparallel beta-sheets held together on their perpendicular axis through intermolecular hydrogen bonds [Nelson et al., 2005]. Single protofibrils may associate laterally with other protofibrils and lead to mature amyloid fibrils, which are about 6–10 nm thick and up to several micrometers long [Khurana et al., 2003]. Amyloid fibrils result in the same X-ray refraction pattern and are detectable via binding of Congo Red and Thioflavin T dyes [Eanes and Glenner, 1968; Kuznetsova et al., 2012; Wu et al., 2012; Girych et al., 2016]. Other more general characteristics of amyloids include their resistance toward proteases and ionic detergents, and a nucleation-mediated growth, which is mostly homogeneous but can be heterogeneous at lower levels [Soto and Castaño, 1996; Šarić et al., 2014; Törnquist et al., 2018]. This latter feature makes it interesting for potential cross-in-

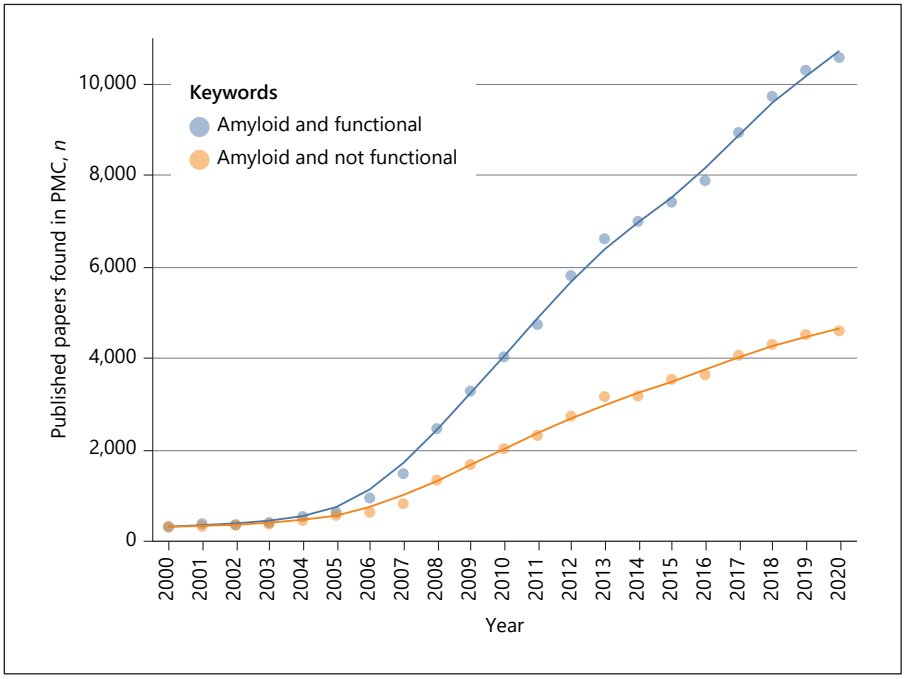


Fig. 2. Comparison of publications per year with the keyword "amyloid" and the presence/lack of the term "functional" as found in the PubMed Central (PMC) database as of December 30, 2020. Trend lines drawn over scatter plot with local regression based on the LOESS method.

teractions between amyloids of different species. Additionally, the study of antimicrobial properties of many of the known amyloids, including pathology-associated ones, has gained a lot of traction in recent years [Soscia et al., 2010; Kagan et al., 2011; Spitzer et al., 2016; Gosztyla et al., 2018; Martin et al., 2018]. Pore formation and general non-specific and irreversible interaction with phospholipid membranes have been proposed as mechanisms for antibiotic activity [Butterfield and Lashuel, 2010; Last and Miranker, 2013].

Initially, amyloids were investigated as the etiological agent of many neurodegenerative diseases [Muchowski, 2002]. However, over the last two decades, they have been increasingly studied in the context of their prevalence in many physiological processes in all three domains of life including bacteria, archaea, and eukarya [Levkovich et al., 2021]. Many amyloids that impact both virulence and survival in prokaryotes have been described [Antonets et al., 2020]. Thus, to distinguish pathogenic amyloids from the latter they are commonly referred to as functional amyloids in the literature [Badtke et al., 2009]. In Figure 2, the larger increase in recent years in publications containing the keyword "amyloid" related to functional as compared to those lacking the term "functional" is evident.

The characteristics of amyloids already mentioned, including resilience, heterogeneity, nucleation, and antimicrobial activity, underscore many of their physiological functions in plant-associated microbial communities. Here, we discuss in the context of plant colonization the history and recent developments of functional amyloids associated with bacterial survival strategies in host-associated and host-related communities. We classify these strategies into two main groups: structural modifications, which include biofilm and hydrophobic surface formation, and defense through antimicrobial activity. Lastly, miscellaneous plant niche-related functions are described, which include amyloids that regulate diverse aspects of their hosts survival and which do not fit within the other two categories. A visual summary of all of these functions is shown in Figure 1 and representative proteins mentioned in this mini-review are summarized in Table 1.

Structural Modifications of Microbial Amyloids

Biofilm Formation

Biofilms are complex microbial communities formed by the cooperation of single or multiple species that adhere to a surface and each other, secreting an extracellular matrix (ECM) [Dragoš and Kovács, 2017]. They represent one of the most widespread strategies for bacterial virulence and proliferation in the microbial world. Biofilms naturally exist in diverse niches of plants in both the

Table 1. Representative microbial, plant amyloids, and related proteins with a role in plant-associated microbial communities

Main function	Protein	Organism	Localization	Additional comments	References
Biofilm formation	CsgA CsgB	*Escherichia coli* *Salmonella enterica*	Extracellular	Virulence promotion Quorum sensing	[Barnhart and Chapman, 2006]
	TasA	*Bacillus* species	Extracellular	Antibacterial Membrane stabilization Quorum sensing	[Stöver and Driks, 1999; Romero et al., 2010]
	FapC	*Pseudomonas* species	Extracellular	Cell surface adhesion Quorum sensing	[Dueholm et al., 2013; Rouse et al., 2018a]
Hydrophobic layer formation	MPG1	*Magnaporthe oryzae*	Extracellular Cell wall	Surface detection Spore attachment	[Kershaw et al., 1998; Pham et al., 2016]
	BslA	*Bacillus subtilis*	Extracellular	Self-assembling protein that coats *Bacillus* biofilms	[Kovács et al., 2012; Hobley et al., 2013]
	Chaplins	*Streptomyces coelicolor*	Extracellular Cell wall	Surface attachment	[Elliot et al., 2003; Bokhove et al., 2013]
	Hum3 Rsp1	*Ustilago maydis*	Extracellular Cell wall	Shields fungus from the plant immune system	[Müller et al., 2008]
Antimicrobial activity	Microcin E492	*Klebsiella pneumoniae*	Extracellular	Amyloid fibrils act as a reservoir for antimicrobial peptides	[Bieler et al., 2005; Shahnawaz and Soto, 2012]
	Prohevein	*Hevea brasiliensis*	Extracellular	Latex tree antifungal peptide	[Berthelot et al., 2016]
	Cn-AMP2	*Cocos nucifera*	Extracellular	Antimicrobial peptide from coconut	[Gour et al., 2016]
	RsAFP-19	*Raphanus sativus*	Extracellular	Antifungal peptide from radish	[Garvey et al., 2013]
Plant virulence promotion	Harpins	*Xanthomonas* species *Erwinia amylovora* *Pseudomonas syringae*	Extracellular Cell wall	Effector translocation Plant hypersensitive response Plant cell toxicity	[Oh et al., 2007; Choi et al., 2013]
	RTP1p	*Uromyces* fabae	Extracellular Haustorium	Structural and stabilizing role	[Kemen et al., 2013]
Signaling	NLR amyloids	Filamentous fungi and bacteria	Membrane and cytosolic components	Non-self recognition	[Loquet and Saupe, 2017; Dyrka et al., 2020]
Plant symbiosis promotion	RopA and RopB	*Rhizobium leguminosarum*	Outer membrane Extracellular	Soluble forms are membrane proteins	[Kosolapova et al., 2019]

NLR, Nod-like receptor.

phyllosphere and rhizosphere and are accountable for a large part of all activity of bacteria in nature [Hall-Stoodley et al., 2004]. The ECM is composed of proteins, extracellular DNA, and polysaccharides, which in addition to maintaining microbial adhesion, protects the community and mediates interaction with the environment, including the host [Branda et al., 2005]. Within biofilms, common goods can be freely shared and bacteria are protected against harmful chemical and physical events, such as antimicrobials or displacement by rain [Patel, 2005; Arnaouteli et al., 2016]. The ECM components of biofilms are diverse in function and their composition varies for different microbial species [Flemming and Wingender, 2010]. However, the protein component is often consistently comprised of a single protein which forms a mesh of functional amyloid fibrils [Erskine et al., 2018a]. The main role of these fibrils is to build the scaffold on which the stationary cells and other ECM components rest. Additionally, it gives a greater degree of resilience to the structure, as mature fibrils are resistant to thermal and chemical denaturing conditions, including proteases. This helps to maintain the integrity of biofilms in a competitive environment, such as the extracellular compartments inside of plants, where these stresses are common

[Taglialegna et al., 2016]. Functional amyloids in association with biofilms have been studied in many different microbial species, the majority of them focusing on bacteria. Herein, we summarize the well-documented ones and their involvement in plant colonization.

Curli is the most studied biofilm-associated amyloid and was the first to be described [Olsen et al., 1989]. It is produced by, among others, the enteric bacteria *Escherichia coli* and *Salmonella enterica*, which are also found in the environment and are well prepared to form biofilms on plants [Danhorn and Fuqua, 2007; Carter et al., 2016; Pruteanu et al., 2020]. The major subunit protein of curli is CsgA, composed of five repeat units with conserved glutamine and asparagine residues important for amyloid formation [Wang et al., 2010]. Curli plays an important role in the adhesion and promotion of biofilm onto different phyllosphere surfaces, including various economically important crops [Jeter and Matthysse, 2005; Boyer et al., 2016]. Moreover, curli is involved in many aspects of the biology of its producer, including virulence promotion, and it is regulated through quorum sensing [Smith et al., 2017; Saxena et al., 2019]. Curli has been determined in addition as an important virulence factor in Shigatoxigenic *E. coli* on fresh produce of crops and therefore represents a serious risk to human health through ingestion of uncooked vegetables [Merget et al., 2019]. How far curli stabilizes biofilms on plants and how much it promotes resistance to biofilms of *E. coli* to mechanical stresses and removal by solvents is still under debate and an important topic in food security.

The formation of *Bacillus* biofilms in the rhizosphere and phyllosphere is associated with plant growth promotion [Hashem et al., 2019]. The major proteinaceous component of this biofilm is TasA, which forms amyloid fibrils and provides integrity to the ECM [Romero et al., 2010]. Despite claims to its non-amyloidogenic nature, it is still widely considered to be a functional amyloid [Erskine et al., 2018b]. TasA was also shown initially to be antibacterial [Stöver and Driks, 1999]. However, whether this function is associated with its capacity to form an amyloid structure is not known. Recently reported functions of this amyloid not related to biofilm formation include its potential contribution to membrane stabilization during the stationary phase of the cell and its role in community signaling [Steinberg et al., 2020; Cámara-Almirón et al., 2020]. Overall, TasA promotes many aspects of the fitness and survival of plant-associated *Bacillus* species.

Biofilm formation in the roots by some members of *Pseudomonas*, like *Pseudomonas fluorescens* and *Pseudomonas putida*, is also associated with plant growth promotion [Meliani et al., 2017]. *Pseudomonas* species secrete the functional amyloid protein (Fap), which contributes to stable and robust biofilm formation and renders protection against chemical and mechanical stresses [Ueda and Saneoka, 2015; Zeng et al., 2015; Rouse et al., 2018b]. The major amyloid fibril component in *Pseudomonas*' biofilm is FapC [Dueholm et al., 2010]. It presents three imperfect repeats of a glutamine- and asparagine-rich domain that are responsible for the formation of very stable amyloid fibrils [Rasmussen et al., 2019].

All of these biofilm amyloids require unique and intricate pathways with numerous intermediate enzymes and safety stops that keep aggregation under control [Balistreri et al., 2020], as its unintended trigger in the cytoplasm would overwhelm chaperones and lead to cell death [Landreh et al., 2015]. Additionally, there are usually two or more proteins that are directly responsible for amyloid formation: the major subunit protein that makes up most of the fibril's weight but is unable to polymerize on its own, or does so slowly, and the minor subunit that acts as a nucleator. This strategy is found, for example, in curli, where CsgA and CsgB fulfill those roles, respectively [Hammer et al., 2007; Yan et al., 2020].

The dependence on amyloid fibrils for biofilm construction makes them a central target for interference by plants in order to keep infection under control. Plant polyphenols and flavonoids, in particular, have been shown to inhibit the development of bacterial biofilms through the blocking of amyloid formation in several distinct bacterial species [Najarzadeh et al., 2019; Pruteanu et al., 2020]. Interestingly, rather than a broad anti-amyloidogenic effect, they target amyloids produced by specific bacterial species. This is in accordance with the reported benefits of some bacterial biofilms in the promotion of plant fitness, which therefore may be preferred by the plant over others.

The widespread occurrence of functional amyloids in their association with biofilm-forming pathogenic and beneficial microbes emphasize their significance in microbe-microbe and microbe-host interactions. Further studies are required to decipher their role in plant-associated microbes, which could lead to the development of novel strategies for plant disease management particularly through probiotics based on mixed cultures that could gain resilience under harsh natural conditions through amyloid-stabilized surface attachment.

Hydrophobic Surface Formation

Another class of structural modifications that are related to microbial survival in the plant holobiont includes hydrophobic layer formation by surface-active proteins. Surface-active proteins modify the properties of physical interfaces and are often linked to an amyloid structure [Sunde et al., 2016]. In fungi, these are called hydrophobins and are known to play an active role in plant-fungi interactions that favor virulence [Teertstra et al., 2009]. Proteins with a similar function as fungal hydrophobins have also been described in some filamentous bacteria. In *Streptomyces coelicolor*, the protein family of chaplins is composed of amyloids that play a role in the formation of aerial mycelia and attachment to surfaces [Elliot et al., 2003]. They respectively perform these roles in two distinct amyloid morphologies, the first formed at water-air interfaces, and the second formed in solution [Bokhove et al., 2013]. In *Bacillus subtilis*, the protein BslA is known to coat TasA-formed biofilm in a thin hydrophobic layer, which helps maintain impermeability and repellency [Kobayashi and Iwano, 2012; Bromley et al., 2015]. BslA is self-assembling, but in contrast to fungal hydrophobins, is not described as amyloidogenic in nature [Liu et al., 2017]. Both *Streptomyces* and *Bacillus* species are soil-dwelling bacteria associated with plants and have been shown to promote their growth and protect them against phytopathogens [Hashem et al., 2019; Kim et al., 2019].

Some Streptomyces species such as *S. scabies* have become serious crop pathogens where the ability to colonize plant niches including the secretion of plant hormones has become a virulence factor [Li et al., 2019]. The role of functional amyloids in such lifestyle shifts has been poorly studied and might in the future become an important target to study the transition from symbionts and facultative pathogens to obligate pathogens. In the obligate biotrophic plant pathogenic fungus *Ustilago maydis* two secreted candidate effectors Hum3 and Rsp1, a hydrophobin and a hydrophobic repeat-rich protein, are tightly bound to the cell wall and form amyloid-like fibrils that influence the surface hydrophobicity [Müller et al., 2008]. It was proposed that they play a role in shielding the fungal hyphae from the plant immune system [Lanver et al., 2017]. For obligate biotrophic fungi, the integrity of the host is crucial for their successful manifestation and the completion of their pathogen life cycle. The obligate rust fungus, *Uromyces fabae*, delivers the filament-forming protein RTP1p, via sub-compartments of the haustorium into the host cytoplasm where it plays a structural and stabilizing role [Kemen et al., 2005, 2013]. RTP1p has therefore been hypothesized to be a haustorial cell wall protein that extends the intracellular lifespan of the pathogen. Amyloid effector proteins may therefore represent a tool for extending the biotrophic phase and protecting the haustorium from the plant defenses even under conditions where cell death has been initiated by the host. If and how this is related to the green islands that can be observed when endophytes colonize plant leaves is a future topic of debate [Wemheuer et al., 2019].

In other filamentous phytopathogenic fungi, hydrophobin functions that promote virulence and pathogenicity include spore dispersal, attachment to hydrophobic surfaces, and immune evasion. Hydrophobins act as surfactants that break surface water tension and maintain a hydrophobic exterior to allow aerial hyphae to develop and prevent its desiccation [Linder et al., 2005]. This also helps with better dissemination as dry spores are lighter and carried farther away [Beever and Dempsey, 1978; Wessels, 1996]. Hydrophobins also contribute to surface detection and spore attachment to the hydrophobic leaf surface. Such is the case of the hydrophobin MPG1 from the pathogen *Magnaporthe oryzae* in rice [Kershaw et al., 1998], whose amyloid aggregation is triggered by a surface-driven mechanism [Pham et al., 2016]. Hydrophobins may also help mask spore epitopes recognized by the plant and thus evade immune detection [Aimanianda et al., 2009; Carrion et al., 2013; Marcos et al., 2016].

These hydrophobic coatings can also be understood as a way to prevent bacterial colonization from water droplets. They discourage accumulation and adsorption onto the surface, therefore effectively inhibiting bacterial adhesion and thus biofilm formation onto the hyphae, spores, or other biofilm surfaces [Wick et al., 2007; Artini et al., 2017]. In fact, new developments in antibacterial surfaces with application in, for example, medical devices, include the use of recombinant hydrophobins to prevent biofilm attachment [Wang et al., 2017; Berger and Sallada, 2019; Devine et al., 2019; Sorrentino et al., 2020].

Antimicrobial Properties of Amyloids

Many already known antimicrobials have been associated in their activity with their capacity to assemble amyloid structures, including mammalian Protegrin-1 and amphibian Uperin 3.5 [Jang et al., 2011; Martin et al., 2018; Salinas et al., 2020]. The bacterial microcin E492 produced by *Klebsiella pneumoniae,* a soil and plant dwelling bacterium, has been described as amyloid. This microcin is an antibacterial peptide that kills bacteria through the formation of channels that disrupt mem-

brane permeability and mannose metabolism [Biéler et al., 2010]. Another interesting aspect of this amyloid is that mature fibrils act as an inert reservoir for the toxic peptide. After triggering through external factors, such as low pH, small soluble oligomers are released, which are then responsible for its toxicity [Bieler et al., 2005; Shahnawaz and Soto, 2012].

On the plant side, certain defense-related peptides have been shown to exhibit amyloid-like properties in vitro. These include prohevein from *Hevea brasiliensis*, a wound-induced peptide whose C-terminus exhibits agglutination of pathogenic organisms [Berthelot et al., 2016]. Other antimicrobial amyloid peptides from plants include Cn-AMP2 from *Cocos nucifera*, an antimicrobial from coconut water effective against gram-positive and gram-negative bacteria, and RsAFP-19, an antifungal defensin from *Raphanus sativus* [Mandal et al., 2009; Gour et al., 2016]. Interestingly, the fungicidal activity of the latter is negatively correlated with its aggregation level. This seems to suggest that one of its roles is to act as a decoy for the inactivation of toxic oligomeric intermediates from competitors into non-active fibrils [Caughey and Lansbury, 2003; Bieler et al., 2005].

The antimicrobial nature of amyloids is a topic of research with many implications for their potential use in human health. Particularly interesting would be the applications of such antimicrobial peptides against multidrug-resistant bacteria, for which targeted antibiotic resistance is an increasing problem [Wise et al., 1998]. Since the mechanism of most antimicrobial peptides, including antimicrobial amyloids, is not linked to a specific target but rather to irreversible binding and disruption of membranes, mechanisms of resistance are less likely to evolve [Mwangi et al., 2019]. Additionally, there is a need for antimicrobial compounds with little environmental impact for their use in agriculture [Montesinos and Bardají, 2008]. However, their applicability in both of these branches is hindered by the lack of understanding of what makes some antimicrobial amyloids more cytotoxic than others [Voth et al., 2020].

How antimicrobial amyloid producers defend against their own peptides is not known and probably varies among specific amyloids. As already mentioned, a complex system of chaperones ensures that there is no aggregation in the cytoplasm and the protein is in a state ready for translocation across the membrane [Sugimoto et al., 2018]. Additionally, external conditions also trigger amyloid-dependent antimicrobial activity and therefore may help to direct its action through two main mechanisms. The first is conformational change into an amyloid structure that leads to a more toxic protein, for example, human Serum amyloid A, which is active only at the skin surface because of its sensitivity to lower pH [Zheng et al., 2020]. The second is due to the shedding of soluble oligomers from mature fibrils that may themselves be toxic, as is the case of microcin E492 [Shahnawaz and Soto, 2012]. Very little is known about toxic amyloid proteins in the plant microbial community. Such as microbes on the human skin, microbes in the plant apoplast face a low pH (healthy skin pH 5.4 to 5.9, plant apoplast pH 5 to 6) that can quickly get more alkaline upon stress [Geilfus, 2017]. These changes might have a severe impact on amyloid toxicity and functionality as described above for human Serum amyloid A and require a high degree of adaptation by the microbes [Zheng et al., 2020]. Identifying antimicrobial amyloids that react to pH shifts in plants might be key to identify novel antimicrobial compounds that do not harm the natural microbiota but do protect from specific pathogens.

Virulence, Signaling, and Symbiosis in Microbial Amyloids

Finally, we describe three classes of amyloids that are not directly related to structural or antimicrobial functions that have been described in plant-associated microbial communities. These include plant toxicity and hypersensitive response promotion by harpins, non-self recognition in filamentous fungi and bacteria, and root symbiosis promotion.

The harpins are a family of heat-stable proteins produced by the phytopathogenic bacteria *Xanthomonas* spp., *Erwinia amylovora*, and *Pseudomonas syringae* [Oh et al., 2007]. These proteins are associated with the promotion of virulence through several amyloid-related mechanisms: bacterial effector translocation, induction of plant hypersensitive response, and cytotoxicity against plant cells [Choi et al., 2013]. Harpin's ability to induce hypersensitive response was correlated to its capacity to form amyloid fibrils in vitro [Oh et al., 2007]. The cytotoxicity mechanism is believed to be due to the formation of beta sheet-rich pores that bind to membranes and cause depolarization in plant cells [Pike et al., 1998].

The role for non-self-recognition and programmed cell death of amyloids has been described in filamentous fungi [Glass and Dementhon, 2006]. Small amyloid motifs act as a signaling mechanism that works by linking receptor and activator protein domains through a templating fold, leading ultimately to cell death [Loquet and

Saupe, 2017]. Nod-like receptor-associated amyloid signaling motifs have been recently discovered in filamentous bacteria, termed BELL and BASS [Dyrka et al., 2020]. They are loosely homologous to the animal, plant, and fungi Nod-like receptors and are proposed to act through similar amyloid-templating mechanisms [Saupe, 2020]. Non-self-recognition plays a role in maintaining pathogen diversity and therefore promoting the exchange of pathogenic traits important for survival against an ever-evolving plant immune system [Ishikawa et al., 2012].

There are fewer reports of functional amyloids concerning symbiotic interactions, probably because research effort is biased toward pathogenic and virulence-promoting mechanisms. RopA and RopB are two recently described proteins from *Rhizobium leguminosarum,* which display amyloid formation linked to microbe-host symbiosis [Kosolapova et al., 2019]. These proteins show structural similarity and are predicted to be outer membrane porins in their soluble forms. Their expression correlates with the formation of capsules, extracellular structures associated with stationary growth, in this root nodule bacterium. Kosolapova et al. [2019] speculate on its role in the establishment of plant-microbial symbiosis through the observation of enhanced expression after the addition of a plant flavonoid.

Conclusion

Amyloid proteins have crucial properties that make them suitable to fill diverse roles in the bacterial and fungal survival of plant-associated communities. Their capacity to polymerize into very resistant fibrils helps them withstand the stresses associated with plant colonization. This is highlighted by the many amyloid biofilm-forming proteins, including curli, Fap, and TasA. Additionally, the tendency of small soluble oligomers to interact with membranes and depolarize them makes them a common structure among antimicrobial and cytotoxic peptides. In this mini-review, we have also discussed amyloids that take part in symbiosis, signaling, and virulence mechanisms. Such a plethora of functions, with what is at the core the same fold, hints at yet to be discovered interactions. Potential cross-seeding among different amyloids in microbial communities, like the plant microbiome, may have a big impact on bacterial survival and disease. Recent examples from human health about the involvement of bacterial amyloids in the seeding of pathogenic amyloids give us a hint of this untapped potential [Javed et al., 2020; Sampson et al., 2020]. All in all, understanding proteins in the context of their amyloid structure and cross-interactions will improve our understanding of the ecology of plant-associated microbial communities and help to develop new methods relevant to human medicine and pest biocontrol.

Acknowledgment

The authors are thankful to Paul Runge and Monja Schmid for their comments and suggestions. We further thank the anonymous reviewers for their time and input.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The authors are grateful to the German Research Foundation (DFG)-funded research training group RTG 1708 "Molecular principles of bacterial survival strategies" (grant # 174858087; Y.H) for providing the funding for the development of the manuscript.

Author Contributions

D.G.-P. conceived the idea for the review, drafted the manuscript, and designed the illustrations. V.C. contributed to the biofilm section. A.K. contributed to the hydrophobic surface section. E.K. contributed to the overall discussion. All authors reviewed and approved the final manuscript.

References

Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, et al. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. Plos Biol. 2016;14(1):e1002352.

Aimanianda V, Bayry J, Bozza S, Kniemeyer O, Perruccio K, Elluru SR, et al. Surface hydrophobin prevents immune recognition of airborne fungal spores. Nature. 2009;460(7259):1117–21.

Anderson JP, Gleason CA, Foley RC, Thrall PH, Burdon JB, Singh KB. Plants versus pathogens: an evolutionary arms race. Funct Plant Biol. 2010;37(6):499–512.

Antonets KS, Belousov MV, Sulatskaya AI, Belousova ME, Kosolapova AO, Sulatsky MI, et al. Accumulation of storage proteins in plant seeds is mediated by amyloid formation. Plos Biol. 2020;18(7):e3000564.

Arnaouteli S, MacPhee CE, Stanley-Wall NR. Just in case it rains: building a hydrophobic biofilm the Bacillus subtilis way. Curr Opin Microbiol. 2016;34:7–12.
Artini M, Cicatiello P, Ricciardelli A, Papa R, Selan L, Dardano P, et al. Hydrophobin coating prevents Staphylococcus epidermidis biofilm formation on different surfaces. Biofouling. 2017;33(7):601–11.
Badtke MP, Hammer ND, Chapman MR. Functional Amyloids Signal Their Arrival. Sci Signal. 2009;2(80):pe43–pe43.
Balistreri A, Goetzler E, Chapman M. Functional Amyloids Are the Rule Rather Than the Exception in Cellular Biology. Microorganisms. 2020;8(12):1951.
Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proc National Acad Sci. 2018;115:201711842.
Barnhart MM, Chapman MR. Curli Biogenesis and Function. Annu Rev Microbiol. 2006;60:131–47.
Beattie GA, Lindow SE. Bacterial Colonization of Leaves: A Spectrum of Strategies. Phytopathology. 1999;89(5):353–9.
Beever RE, Dempsey GP. Function of rodlets on the surface of fungal spores. Nature. 1978;272(5654):608–10.
Berger BW, Sallada ND. Hydrophobins: multifunctional biosurfactants for interface engineering. J Biol Eng. 2019;13:10.
Berthelot K, Lecomte S, Coulary-Salin B, Bentaleb A, Peruch F. Hevea brasiliensis prohevein possesses a conserved C-terminal domain with amyloid-like properties in vitro. Biochim Biophys Acta. 2016;1864(4):388–99.
Bieler S, Estrada L, Lagos R, Baeza M, Castilla J, Soto C. Amyloid Formation Modulates the Biological Activity of a Bacterial Protein. J Biol Chem. 2005;280(29):26880–5.
Biéler S, Silva F, Belin D. The polypeptide core of Microcin E492 stably associates with the mannose permease and interferes with mannose metabolism. Res Microbiol. 2010;161(8):706–10.
Bokhove M, Claessen D, de Jong W, Dijkhuizen L, Boekema EJ, Oostergetel GT. Chaplins of Streptomyces coelicolor self-assemble into two distinct functional amyloids. J Struct Biol. 2013;184(2):301–9.
Boyer RR, Sumner SS, Willia RC, Pierson MD, Poam DL, Kniel KE. Influence of Curli Expression by Escherichia coli O157:H7 on the Cell's Overall Hydrophobicity, Charge, and Ability To Attach to Lettuce. J Food Protect. 2016;70:1339–45.
Branda SS, Vik S, Friedman L, Kolter R. Biofilms: the matrix revisited. Trends Microbiol. 2005;13(1):20–6.
Bromley KM, Morris RJ, Hobley L, Brandani G, Gillespie RM, McCluskey M, et al. Interfacial self-assembly of a bacterial hydrophobin. Proc Natl Acad Sci USA. 2015;112(17):5419–24.
Butterfield SM, Lashuel HA. Amyloidogenic Protein–Membrane Interactions: Mechanistic Insight from Model Systems. Angewandte Chemie Int Ed. 2010;49:5628–54.
Cámara-Almirón J, Navarro Y, Díaz-Martínez L, Magno-Pérez-Bryan MC, Molina-Santiago C, Pearson JR, et al. Dual functionality of the amyloid protein TasA in Bacillus physiology and fitness on the phylloplane. Nat Commun. 2020;11(1):1859.
Carrion SJ, Leal SM, Ghannoum MA, Aimanianda V, Latgé JP, Pearlman E. The RodA hydrophobin on Aspergillus fumigatus spores masks dectin-1- and dectin-2-dependent responses and enhances fungal survival in vivo. J Immunol. 2013;191(5):2581–8.
Carter MQ, Louie JW, Feng D, Zhong W, Brandl MT. Curli fimbriae are conditionally required in Escherichia coli O157:H7 for initial attachment and biofilm formation. Food Microbiol. 2016;57:81–9.
Caughey B, Lansbury PT. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu Rev Neurosci. 2003;26:267–98.
Chaudhry V, Runge P, Sengupta P, Doehlemann G, Parker JE, Kemen E. Shaping the leaf microbiota: plant-microbe-microbe interactions. J Exp Bot. 2021 Jan 20;72(1):36–56.
Choi MS, Kim W, Lee C, Oh CS. Harpins, Multifunctional Proteins Secreted by Gram-Negative Plant-Pathogenic Bacteria. Mol Plant Microbe Interact. 2013;26(10):1115–22.
Danhorn T, Fuqua C. Biofilm Formation by Plant-Associated Bacteria. Annu Rev Microbiol. 2007;61:401–22.
Devine R, Singha P, Handa H. Versatile biomimetic medical device surface: hydrophobin coated, nitric oxide-releasing polymer for antimicrobial and hemocompatible applications. Biomater Sci. 2019;7(8):3438–49.
Dragoš A, Kovács ÁT. The Peculiar Functions of the Bacterial Extracellular Matrix. Trends Microbiol. 2017;25:257–66.
Dueholm MS, Petersen SV, Sønderkær M, Larsen P, Christiansen G, Hein KL, et al. Functional amyloid in Pseudomonas. Mol Microbiol. 2010;77(4):1009–20.
Dueholm MS, Søndergaard MT, Nilsson M, Christiansen G, Stensballe A, Overgaard MT, et al. Expression of Fap amyloids in Pseudomonas aeruginosa, P. fluorescens, and P. putida results in aggregation and increased biofilm formation. Microbiologyopen. 2013;2(3):365–82.
Dyrka W, Coustou V, Daskalov A, Lends A, Bardin T, Berbon M, et al. Identification of NLR-associated Amyloid Signaling Motifs in Bacterial Genomes. J Mol Biol. 2020;432(23):6005–27.
Eanes ED, Glenner GG. X-ray diffraction studies on amyloid filaments. J Histochem Cytochem. 1968;16(11):673–7.
Elliot MA, Karoonuthaisiri N, Huang J, Bibb MJ, Cohen SN, Kao CM, et al. The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in Streptomyces coelicolor. Genes Dev. 2003;17(14):1727–40.
Erskine E, MacPhee CE, Stanley-Wall NR. Functional Amyloid and Other Protein Fibers in the Biofilm Matrix. J Mol Biol. 2018a;430(20):3642–56.
Erskine E, Morris RJ, Schor M, Earl C, Gillespie RMC, Bromley KM, et al. Formation of functional, non-amyloidogenic fibres by recombinant Bacillus subtilis TasA. Mol Microbiol. 2018b;110(6):897–913.
Flemming HC, Wingender J. The biofilm matrix. Nat Rev Microbiol. 2010;8(9):623–33.
Garvey M, Meehan S, Gras SL, Schirra HJ, Craik DJ, Van der Weerden NL, et al. A radish seed antifungal peptide with a high amyloid fibril-forming propensity. Biochim Biophys Acta. 2013;1834(8):1615–23.
Geilfus C-M. The pH of the Apoplast: Dynamic Factor with Functional Impact Under Stress. Mol Plant. 2017;10(11):1371–86.
Girych M, Gorbenko G, Maliyov I, Trusova V, Mizuguchi C, Saito H, et al. Combined thioflavin T-Congo red fluorescence assay for amyloid fibril detection. Methods Appl Fluoresc. 2016;4(3):034010.
Glass NL, Dementhon K. Non-self recognition and programmed cell death in filamentous fungi. Curr Opin Microbiol. 2006;9(6):553–8.
Gosztyla ML, Brothers HM, Robinson SR. Alzheimer's Amyloid-β is an Antimicrobial Peptide: A Review of the Evidence. J Alzheimers Dis. 2018;62(4):1495–506.
Gour S, Kaushik V, Kumar V, Bhat P, Yadav SC, Yadav JK. Antimicrobial peptide (Cn-AMP2) from liquid endosperm of Cocos nucifera forms amyloid-like fibrillar structure. J Pept Sci. 2016;22:201–7.
Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the Natural environment to infectious diseases. Nat Rev Microbiol. 2004;2(2):95–108.
Hammer ND, Schmidt JC, Chapman MR. The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. Proc Natl Acad Sci USA. 2007;104(30):12494–9.
Hashem A, Tabassum B, Fathi Abd Allah E. Bacillus subtilis: A plant-growth promoting rhizobacterium that also impacts biotic stress. Saudi J Biol Sci. 2019;26(6):1291–7.
Hobley L, Ostrowski A, Rao FV, Bromley KM, Porter M, Prescott AR, et al. BslA is a self-assembling bacterial hydrophobin that coats the Bacillus subtilis biofilm. Proc Natl Acad Sci USA. 2013;110(33):13600–5.
Ishikawa FH, Souza EA, Shoji JY, Connolly L, Freitag M, Read ND, et al. Heterokaryon Incompatibility Is Suppressed Following Conidial Anastomosis Tube Fusion in a Fungal Plant Pathogen. Plos One. 2012;7(2):e31175.
Jang H, Arce FT, Mustata M, Ramachandran S, Capone R, Nussinov R, et al. Antimicrobial Protegrin-1 Forms Amyloid-Like Fibrils with Rapid Kinetics Suggesting a Functional Link. Biophys J. 2011;100(7):1775–83.

Javed I, Zhang Z, Adamcik J, Andrikopoulos N, Li Y, Otzen DE, et al. Accelerated Amyloid Beta Pathogenesis by Bacterial Amyloid FapC. Adv Sci (Weinh). 2020;7(18):2001299.
Jeter C, Matthysse AG. Characterization of the Binding of Diarrheagenic Strains of E. coli to Plant Surfaces and the Role of Curli in the Interaction of the Bacteria with Alfalfa Sprouts. Mol Plant Microbe Interact. 2005;18(11):1235–42.
Kagan BL, Jang H, Capone R, Teran Arce F, Ramachandran S, Lal R, et al. Antimicrobial Properties of Amyloid Peptides. Mol Pharm. 2011;9(4):708–17.
Kandel S, Joubert P, Doty S. Bacterial Endophyte Colonization and Distribution within Plants. Microorganisms. 2017;5(4):77.
Kemen E, Kemen AC, Rafiqi M, Hempel U, Mendgen K, Hahn M, et al. Identification of a Protein from Rust Fungi Transferred from Haustoria into Infected Plant Cells. Mol Plant Microbe Interact. 2005;18(11):1130–9.
Kemen E, Kemen A, Ehlers A, Voegele R, Mendgen K. A novel structural effector from rust fungi is capable of fibril formation. Plant J. 2013;75(5):767–80.
Kershaw MJ, Wakley G, Talbot NJ. Complementation of the Mpg1 mutant phenotype in Magnaporthe grisea reveals functional relationships between fungal hydrophobins. Embo J. 1998;17(14):3838–49.
Khurana R, Ionescu-Zanetti C, Pope M, Li J, Nielson L, Ramírez-Alvarado M, et al. A General Model for Amyloid Fibril Assembly Based on Morphological Studies Using Atomic Force Microscopy. Biophys J. 2003;85(2):1135–44.
Kim DR, Cho G, Jeon CW, Weller DM, Thomashow LS, Paulitz TC, et al. A mutualistic interaction between Streptomyces bacteria, strawberry plants and pollinating bees. Nat Commun. 2019;10(1):4802.
Kobayashi K, Iwano M. BslA(YuaB) forms a hydrophobic layer on the surface of Bacillus subtilis biofilms. Mol Microbiol. 2012;85(1):51–66.
Kosolapova AO, Belousov MV, Sulatskaya AI, Belousova ME, Sulatsky MI, Antonets KS, et al. Two Novel Amyloid Proteins, RopA and RopB, from the Root Nodule Bacterium Rhizobium leguminosarum. Biomolecules. 2019;9(11):694.
Kovács AT, van Gestel J, Kuipers OP. The protective layer of biofilm: a repellent function for a new class of amphiphilic proteins. Mol Microbiol. 2012;85(1):8–11.
Kuznetsova IM, Sulatskaya AI, Uversky VN, Turoverov KK. A New Trend in the Experimental Methodology for the Analysis of the Thioflavin T Binding to Amyloid Fibrils. Mol Neurobiol. 2012;45(3):488–98.
Landreh M, Rising A, Presto J, Jörnvall H, Johansson J. Specific Chaperones and Regulatory Domains in Control of Amyloid Formation. J Biol Chem. 2015;290(44):26430–6.
Lanver D, Tollot M, Schweizer G, Lo Presti L, Reissmann S, Ma LS, et al. Ustilago maydis effectors and their impact on virulence. Nat Rev Microbiol. 2017;15(7):409–21.
Last NB, Miranker AD. Common mechanism unites membrane poration by amyloid and antimicrobial peptides. Proc Natl Acad Sci USA. 2013;110(16):6382–7.
Levkovich SA, Gazit E, Bar-Yosef DL. Two Decades of Studying Functional Amyloids in Microorganisms. Trends Microbiol. 2021 Mar;29(3):251–65.
Li Y, Liu J, Díaz-Cruz G, Cheng Z, Bignell DRD. Virulence mechanisms of plant-pathogenic Streptomyces species: an updated review. Microbiology (Reading). 2019;165(10):1025–40.
Linder MB, Szilvay GR, Nakari-Setälä T, Penttilä ME. Hydrophobins: the protein-amphiphiles of filamentous fungi. Fems Microbiol Rev. 2005;29(5):877–96.
Liu W, Li S, Wang Z, Yan ECY, Leblanc RM. Characterization of Surface-Active Biofilm Protein BslA in Self-Assembling Langmuir Monolayer at the Air-Water Interface. Langmuir. 2017;33(30):7548–55.
Loquet A, Saupe SJ. Diversity of Amyloid Motifs in NLR Signaling in Fungi. Biomolecules. 2017;7(2):38.
Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular basis for amyloid fibril formation and stability. Proc Natl Acad Sci USA. 2005;102(2):315–20.
Mandal SM, Dey S, Mandal M, Sarkar S, Maria-Neto S, Franco OL. Identification and structural insights of three novel antimicrobial peptides isolated from green coconut water. Peptides. 2009;30(4):633–7.
Marcos CM, de Oliveira HC, de Melo WC, da Silva JF, Assato PA, Scorzoni L, et al. Anti-Immune Strategies of Pathogenic Fungi. Front Cell Infect Microbiol. 2016;6:142.
Martin LL, Kubeil C, Piantavigna S, Tikkoo T, Gray NP, John T, et al. Amyloid aggregation and membrane activity of the antimicrobial peptide uperin 3.5. Peptide Science. 2018;110(3):e24052.
Meliani A, Bensoltane A, Benidire L, Oufdou K. Plant growth-promotion and IAA secretion with Pseudomonas fluorescens and Pseudomonas putida. Res Rev J Bot Sci. 2017;6:16–24.
Mercado-Blanco J. Life of Microbes Inside the Plant. Microbes for Sustainable Agriculture. 2014:25–32.
Merget B, Forbes KJ, Brennan F, McAteer S, Shepherd T, Strachan NJC, et al. Relating growth potential and biofilm formation of Shigatoxigenic Escherichia coli to in planta colonisation and the metabolome of ready-to-eat crops. Biorxiv. 2019:523175.
Montesinos E, Bardají E. Synthetic antimicrobial peptides as agricultural pesticides for plant-disease control. Chem Biodivers. 2008;5(7):1225–37.
Muchowski PJ. Protein Misfolding, Amyloid Formation, and Neurodegeneration. Neuron. 2002;35(1):9–12.
Müller O, Schreier PH, Uhrig JF. Identification and characterization of secreted and pathogenesis-related proteins in Ustilago maydis. Mol Genet Genomics. 2008;279(1):27–39.
Mwangi J, Hao X, Lai R, Zhang ZY. Antimicrobial peptides: new hope in the war against multidrug resistance. Zool Res. 2019;40(6):488–505.
Najarzadeh Z, Mohammad-Beigi H, Nedergaard Pedersen J, Christiansen G, Sønderby TV, Shojaosadati SA, et al. Plant Polyphenols Inhibit Functional Amyloid and Biofilm Formation in Pseudomonas Strains by Directing Monomers to Off-Pathway Oligomers. Biomolecules. 2019;9(11):659.
Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, Grothe R, et al. Structure of the cross-beta spine of amyloid-like fibrils. Nature. 2005;435(7043):773–8.
Oh J, Kim JG, Jeon E, Yoo CH, Moon JS, Rhee S, et al. Amyloidogenesis of Type III-dependent Harpins from Plant Pathogenic Bacteria. J Biol Chem. 2007;282(18):13601–9.
Olsen A, Jonsson A, Normark S. Fibronectin binding mediated by a novel class of surface organelles on Escherichia coll. Nature. 1989;338:652–5.
Patel R. Biofilms and Antimicrobial Resistance. Clinical Orthopaedics and Related Research. 2005Aug;(437):41–7.
Pham CL, Rey A, Lo V, Soulès M, Ren Q, Meisl G, et al. Self-assembly of MPG1, a hydrophobin protein from the rice blast fungus that forms functional amyloid coatings, occurs by a surface-driven mechanism. Sci Rep. 2016;6:25288.
Pike SM, Ádám AL, Pu X-A, Hoyos ME, Laby R, Beer SV, et al. Effects ofErwinia amylovoraharpin on tobacco leaf cell membranes are related to leaf necrosis and electrolyte leakage and distinct from perturbations caused by inoculatedE. amylovora. Physiological and Molecular Plant Pathology. 1998;53(1):39–60.
Pruteanu M, Hernández Lobato JI, Stach T, Hengge R. Common plant flavonoids prevent the assembly of amyloid curli fibres and can interfere with bacterial biofilm formation. Environ Microbiol. 2020;22(12):5280–99.
Rasmussen CB, Christiansen G, Vad BS, Lynggaard C, Enghild JJ, Andreasen M, et al. Imperfect repeats in the functional amyloid protein FapC reduce the tendency to fragment during fibrillation. Protein Sci. 2019;28(3):633–42.
Romero D, Aguilar C, Losick R, Kolter R. Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. Proc Natl Acad Sci USA. 2010;107(5):2230–4.
Rouse SL, Matthews SJ, Dueholm MS. Ecology and Biogenesis of Functional Amyloids in Pseudomonas. J Mol Biol. 2018a;430(20):3685–95.
Rouse SL, Stylianou F, Wu HYG, Berry JL, Sewell L, Morgan RML, et al. The FapF amyloid secretion transporter possesses an atypical asymmetric coiled coil. J Mol Biol. 2018b;430(20):3863–71.
Saikkonen K, Wäli P, Helander M, Faeth SH. Evolution of endophyte-plant symbioses. Trends Plant Sci. 2004;9(6):275–80.

Salinas N, Povolotsky TL, Landau M, Kolodkin-Gal I. Emerging Roles of Functional Bacterial Amyloids in Gene Regulation, Toxicity, and Immunomodulation. Microbiol Mol Biol Rev. 2020;85(1).
Sampson TR, Challis C, Jain N, Moiseyenko A, Ladinsky MS, Shastri GG, et al. A gut bacterial amyloid promotes α-synuclein aggregation and motor impairment in mice. Elife. 2020;9:e53111.
Šarić A, Chebaro YC, Knowles TPJ, Frenkel D. Crucial role of nonspecific interactions in amyloid nucleation. Proc National Acad Sci. 2014;111:17869–74.
Saupe SJ. Amyloid Signaling in Filamentous Fungi and Bacteria. Annu Rev Microbiol. 2020;74:1–19.
Saxena P, Joshi Y, Rawat K, Bisht R. Biofilms: Architecture, Resistance, Quorum Sensing and Control Mechanisms. Indian J Microbiol. 2019;59(1):3–12.
Shahnawaz M, Soto C. Microcin Amyloid Fibrils A Are Reservoir of Toxic Oligomeric Species. J Biol Chem. 2012;287(15):11665–76.
Smith DR, Price JE, Burby PE, Blanco LP, Chamberlain J, Chapman MR. The Production of Curli Amyloid Fibers Is Deeply Integrated into the Biology of Escherichia coli. Biomolecules. 2017;7(4):75.
Sorrentino I, Gargano M, Ricciardelli A, Parrilli E, Buonocore C, de Pascale D, et al. Development of anti-bacterial surfaces using a hydrophobin chimeric protein. Int J Biol Macromol. 2020;164:2293–300.
Soscia SJ, Kirby JE, Washicosky KJ, Tucker SM, Ingelsson M, Hyman B, et al. The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide. Plos One. 2010;5(3):e9505.
Soto C, Castaño EM. The conformation of Alzheimer's beta peptide determines the rate of amyloid formation and its resistance to proteolysis. Biochem J. 1996;314(Pt 2):701–7.
Spitzer P, Condic M, Herrmann M, Oberstein TJ, Scharin-Mehlmann M, Gilbert DF, et al. Amyloidogenic amyloid-β-peptide variants induce microbial agglutination and exert antimicrobial activity. Sci Rep. 2016;6:32228.
Steinberg N, Keren-Paz A, Hou Q, Doron S, Yanuka-Golub K, Olender T, et al. The extracellular matrix protein TasA is a developmental cue that maintains a motile subpopulation within Bacillus subtilis biofilms. Sci Signal. 2020;13(632):eaaw8905.
Stöver AG, Driks A: Secretion, Localization, and antibacterial activity of TasA, a Bacillus subtilis spore-associated protein. J Bacteriol. 1999;181:1664–72.
Sugimoto S, Arita-Morioka KI, Terao A, Yamanaka K, Ogura T, Mizunoe Y. Multitasking of Hsp70 chaperone in the biogenesis of bacterial functional amyloids. Commun Biol. 2018;1:52.
Sunde M, Pham CLL, Kwan AH. Molecular Characteristics and Biological Functions of Surface-Active and Surfactant Proteins. Annu Rev Biochem. 2016;86:585–608.
Taglialegna A, Lasa I, Valle J. Amyloid Structures as Biofilm Matrix Scaffolds. J Bacteriol. 2016;198(19):2579–88.
Teertstra WR, van der Velden GJ, de Jong JF, Kruijtzer JA, Liskamp RM, Kroon-Batenburg LM, et al. The Filament-specific Rep1-1 Repellent of the Phytopathogen Ustilago maydis Forms Functional Surface-active Amyloid-like Fibrils. J Biol Chem. 2009;284(14):9153–9.
Törnquist M, Michaels TCT, Sanagavarapu K, Yang X, Meisl G, Cohen SIA, et al. Secondary nucleation in amyloid formation. Chem Commun (Camb). 2018;54(63):8667–84.
Ueda A, Saneoka H. Characterization of the Ability to Form Biofilms by Plant-Associated Pseudomonas Species. Curr Microbiol. 2015;70(4):506–13.
Voth S, Gwin M, Francis CM, Balczon R, Frank DW, Pittet JF, et al. Virulent Pseudomonas aeruginosa infection converts antimicrobial amyloids into cytotoxic prions. Faseb J. 2020;34(7):9156–79.
Wang X, Zhou Y, Ren JJ, Hammer ND, Chapman MR. Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis. Proc Natl Acad Sci USA. 2010;107(1):163–8.
Wang X, Mao J, Chen Y, Song D, Gao Z, Zhang X, et al. Design of antibacterial biointerfaces by surface modification of poly (ε-caprolactone) with fusion protein containing hydrophobin and PA-1. Colloids Surfaces B Biointerfaces. 2017;151:255–63.
Wemheuer F, Wemheuer B, Daniel R, Vidal S. Deciphering bacterial and fungal endophyte communities in leaves of two maple trees with green islands. Sci Rep. 2019;9(1):14183.
Wessels JGH. Hydrophobins: Proteins that Change the Nature of the Fungal Surface. Adv Microb Physiol. 1996;38:1–45.
Wick LY, Remer R, Würz B, Reichenbach J, Braun S, Schäfer F, et al. Effect of Fungal Hyphae on the Access of Bacteria to Phenanthrene in Soil. Environ Sci Technol. 2007;41(2):500–5.
Wise R, Hart T, Cars O, Streulens M, Helmuth R, Huovinen P, et al. Antimicrobial resistance. Is a major threat to public health. Bmj. 1998;317(7159):609.
Wu C, Scott J, Shea JE. Binding of Congo red to amyloid protofibrils of the Alzheimer Aβ(9-40) peptide probed by molecular dynamics simulations. Biophys J. 2012;103(3):550–7.
Yan Z, Yin M, Chen J, Li X. Assembly and substrate recognition of curli biogenesis system. Nat Commun. 2020;11(1):241.
Zeng G, Vad BS, Dueholm MS, Christiansen G, Nilsson M, Tolker-Nielsen T, et al. Functional bacterial amyloid increases Pseudomonas biofilm hydrophobicity and stiffness. Front Microbiol. 2015;6:1099.
Zheng H, Li H, Zhang J, Fan H, Jia L, Ma W, et al. Serum amyloid A exhibits pH dependent antibacterial action and contributes to host defense against Staphylococcus aureus cutaneous infection. J Biol Chem. 2020;295(9):2570–81.

A.2 Predicting Lifestyle from Positive Selection Data and Genome Properties in Oomycetes

Contributions by DGP:

- Designed the initial manuscript.
- Wrote the initial draft.
- Edited the final manuscript.
- Performed all computations and analyzed all the data.
- Created all the figures and tables.

pathogens

MDPI

Article

Predicting Lifestyle from Positive Selection Data and Genome Properties in Oomycetes

Daniel Gómez-Pérez and Eric Kemen *

Center for Plant Molecular Biology (ZMBP), University of Tübingen, 72074 Tübingen, Germany; daniel.gomez-perez@uni-tuebingen.de

* Correspondence: eric.kemen@uni-tuebingen.de; Tel.: +49-7071-29-78725

Abstract: As evidenced in parasitism, host and niche shifts are a source of genomic and phenotypic diversification. Exemplary is a reduction in the core metabolism as parasites adapt to a particular host, while the accessory genome often maintains a high degree of diversification. However, selective pressures acting on the genome of organisms that have undergone recent lifestyle or host changes have not been fully investigated. Here, we developed a comparative genomics approach to study underlying adaptive trends in oomycetes, a eukaryotic phylum with a wide and diverse range of economically important plant and animal parasitic lifestyles. Our analysis reveals converging evolution on biological processes for oomycetes that have similar lifestyles. Moreover, we find that certain functions, in particular carbohydrate metabolism, transport, and signaling, are important for host and environmental adaptation in oomycetes. Given the high correlation between lifestyle and genome properties in our oomycete dataset, together with the known convergent evolution of fungal and oomycete genomes, we developed a model that predicts plant pathogenic lifestyles with high accuracy based on functional annotations. These insights into how selective pressures correlate with lifestyle may be crucial to better understand host/lifestyle shifts and their impact on the genome.

Keywords: oomycetes; lifestyle; evolution

Citation: Gómez-Pérez, D.; Kemen, E. Predicting Lifestyle from Positive Selection Data and Genome Properties in Oomycetes. *Pathogens* **2021**, *10*, 807. https://doi.org/10.3390/pathogens10070807

Academic Editor: Paolo Gonthier

Received: 30 April 2021
Accepted: 21 June 2021
Published: 25 June 2021

1. Introduction

The adaptation of organisms as they evolve to occupy different niches or adopt different lifestyles is reflected in their genome. Expansion or contraction of gene families has been cited as a general mechanism for such adaptations [1,2]. Expansions arise mainly from gene duplication and, in some cases, from acquisition via horizontal gene transfer, whereas gene loss can happen by accumulation of loss-of-function mutations through genetic drift [3–5]. Fundamentally, both of these processes are driven by adaptive evolution, whereby beneficial mutations are selected for and deleterious ones removed from the gene pool, ultimately leading to phenotypic diversification [6]. More concretely, trends in the evolution of coding genes can be studied by measuring the ratio of non-synonymous (dN) to synonymous (dS) amino acid rates in the comparison to closely related sequences, usually represented as ω [7]. A ratio higher than one ($dN/dS = \omega > 1$) implies positive selection and thus functional diversification, while a ratio lower than one ($dN/dS = \omega < 1$) indicates the presence of purifying selection and thus a tighter constraint for the diversification of the gene sequence. Most genes in an organism are under strong purifying selection, as a change in a key amino acid of a protein would have a detrimental effect [8]. However, a small portion of them, those that have been subject to recent diversification, show signs of an increased nonsynonymous mutation rate. Codon models that take into account statistical rate variations are commonly used in comparative genomic studies [9]. When performed on related organisms that have different lifestyles and hosts, the study of positively selected genes together with their functional annotation illustrates which gene functions played important roles in the adaptation process.

Oomycetes are eukaryotic organisms belonging, together with diatoms and brown algae, to the group of stramenopiles [10,11]. Since their origin from a marine autotrophic common ancestor around 400 million years ago, oomycetes have adapted to multiple environments and lifestyles, and many of them are economically impactful plant and animal parasites [12–14]. Therefore, they represent a relevant and appropriate system to study the genetic impact of lifestyle and host adaptation on genetically close genomes. Four phylogenetic families, representative of oomycete's large diversity, are the target of most current research efforts: Albuginaceae, Peronosporaceae, Saprolegniaceae, and Pythiaceae. The Albuginaceae and most Peronosporaceae independently evolved the ability to survive exclusively on living host material, also known as obligate biotrophy [15]. However, some Peronosporaceae, particularly in the *Phytophthora* genus, are hemibiotrophs, i.e., they display an initial biotrophic phase followed by a necrotrophic one, during which they feed on the decaying living matter of their host [16]. Additionally in the Peronosporaceae, the early divergent clade of *Globisporangium* consists of plant necrotrophs previously classified as *Pythiaceae*. All Albuginaceae, Peronosporaceae, and most Pythiaceae are plant parasitic organisms [17]. On the contrary, most Saprolegniaceae are capable of infecting animals, with few exceptions including plant-causing root rot *Aphanomyces* and the free-living saprophyte *Thraustotheca clavata*, which does not need a host at any point in its life cycle [18–20].

Obligate biotrophs have a considerably reduced primary metabolism. Comparative genome studies have reported a significant and convergent loss of the enzymatic arsenal in independent lineages of the oomycetes following this lifestyle [21]. The picture is not so clear for the heterotrophs and their adaptation to different hosts. *Pythium insidiosum*, a mammal parasite responsible for pythiosis, shows a relatively recent divergence from *Pythium aphanidermatum* and *Pythium arrhenomanes*, both of which are plant pathogens [22]. There are many theories that explain how such drastic host shifts can occur in a small evolutionary timescale [23]. Particularly in oomycetes, large reservoirs of noncoding DNA material can readily evolve into small secreted proteins, known as effectors, facilitating new oomycete–host interactions [24]. Additionally, the readiness to take up genetic material through horizontal gene transfer from fungi and bacteria has been reported at multiple time points in the oomycete lineages [25–27]. However, the impact of host shifts on genomic selective pressures has not been extensively studied.

There is a high degree of convergent evolution between oomycetes and fungi [28]. Both share many of the niches mentioned, including pathogenic niches of animals and plants, as well as lifestyles, including saprotrophy, hemibiotrophy, and obligate biotrophy. Oomycetes and fungi have developed similar strategies to overcome the same challenges, including comparable filamentous and reproductive morphology, as well as akin infection strategies [29]. As mentioned above, convergence is probably promoted by genetic exchange, as the source of many oomycete genes with a role in host adaptation can be traced back to pathogenic fungi [30]. Because of the parallels between the adaptive strategies of these two eukaryotic phyla, we can infer underlying mechanistic principles in oomycetes on the basis of those further characterized in fungi.

How genome information relates to lifestyle and host adaptation is one of the big questions in ecology, and may be relevant to predict the appearance of new emerging diseases. Understanding the genome characteristics and selective pressures in organisms that have undergone host and niche shifts may offer insights into this question. In this study, we report the first whole-genome positive selection screening of the proteome of the oomycetes phylum, including 34 representative members and an outgroup of eight non-oomycete stramenopiles described in Appendix A. We compared the genes inferred as being under positive selection to the background annotated genes to identify enriched biological functions that may correlate to their adaption to different hosts and lifestyles. Additionally, we developed a method to predict plant pathogenic lifestyle with high accuracy from the genome of fungi and oomycetes, based on the presence or absence of key annotated functions.

2. Materials and Methods

2.1. Data Selection and Functional Annotation

We downloaded stramenopile genetic data from the NCBI and FungiDB databases setting as cutoff assemblies with reported gene annotation, resulting in a dataset of 42 total proteomes. We screened the genomes using BUSCO for high abundance of key orthologs in the stramenopile dataset as a form of quality control [31]. We performed functional annotation of the proteomes using InterProScan version 5.50–84.0 [32]. We validated proteins discussed in the manuscript through comparison with matches from the NCBI datatabase using BLAST [33]. We annotated the effectors in the stramenopile dataset by predicting the secretion signal using the tool SignalP 5.0b followed by an annotation with the model EffectorO [34,35]. We annotated the presence/absence of functional annotations from each genome with the Genome Properties database, performed the clustering with the Python library SciPy and visualized it with the package Seaborn [36,37]. We compared Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clusterings of the genome properties and genome properties with added positive selection information to the phylogenetic tree using the Robison–Foulds metric based on clusters with the application TreeCmp [38,39].

2.2. Phylogeny Inference

We constructed the concatenated stramenopile tree using IQ-TREE 2 with automated partitioned model selection on inferred one-to-one orthogroups present in at least 25 of the taxa in the dataset [40]. We assessed full branch support in all nodes of the phylogenetic tree with 1,000 ultrafast bootstrap repetitions using the IQ-TREE 2 software and displayed it rooted on the outgroup of non-oomycetal stramenopiles.

2.3. Orthogroup Classification and Positive Selection Analyses

We developed a pipeline for whole genome positive selection analysis in Python using the Snakemake modular workflow framework [41]. It uses as input the coding nucleotide sequences as well as their corresponding predicted proteins from each proteome. The code and documentation are available at https://github.com/danielzmbp/wsgups (accessed on 24 June 2021). The steps of the pipeline include: grouping of sequences into ortholog families, their alignment with MAFFT, phylogenetic tree inference using FastTree, codon alignment using PAL2NAL, and finally two positive selection algorithms from the HYPHY package [42–44]. The first step, consisting of the classification of these proteomes into ortholog groups was performed with the software Proteinortho version 6, using the synteny parameter and the Diamond algorithm for homology search [45]. The first HYPHY algorithm used in the pipeline is FUBAR, a site-based program that scans the alignment for pervasive positive selection [46]. Families with at least one codon position under positive selection were subsequently analyzed on all branches with the aBSREL algorithm to relate selective pressures to specific lineages [47]. Taxa downstream of nodes with a corrected p value of less than 0.05 were considered under positive selection for this particular gene.

2.4. Enrichment Analyses

We used the Gene Ontology (GO) released in 1 February 2021 [48,49]. We performed GO enrichment using the Python package Goatools based on the InterPro database annotations [50,51]. The background dataset corresponded to the sum of all proteome annotations for the corresponding taxa and the study dataset to the genes found to be under selection. Terms that did not have representative sequences in all analyzed taxa were filtered out. We used as a significance cutoff the negative base 10 logarithm of Holm–Bonferroni corrected p values that were higher than 1.3 (p value < 0.05). Broad and non-informative GO terms like biological or cellular processes were not included in the enrichment tables.

2.5. Machine Learning Model

The multilayered deep learning models were constructed using the Tensorflow version 2.3 library with the Keras application programming interface [52]. The dataset consisted

of 324 proteomes from fungi and oomycete plant pathogens as well as saprobes. We labeled each proteome as one of the four respective plant pathogenic classes based on the literature consensus: saprotroph, necrotroph, hemibiotroph and biotroph. For the genome properties model, we extracted the features of each genome and encoded them based on the presence or absence of all the identified pathways, which resulted in an array of 5024 binary features each. Removal of duplicated entries resulted in 319 unique samples. For the Carbohydrate-Active enZyme (CAZyme) model, we annotated the proteomes using the dbCAN2 database and encoded the absence or abundance of annotated CAZymes that were identified with the three implemented methods: HMMER, DIAMOND and Hotpep [53]. After removal of duplicates, 313 samples of 593 features were used for training and testing of the model. In both models, we performed a stratified split of the dataset into the training dataset, corresponding to 60% of the total, and the optimization and validation datasets, each corresponding to half of the remaining 40%. Hyperparameter optimization, namely of the learning rate, activating functions and dense layer units, was carried out using Keras Tuner and its implementation of the Hyperband algorithm [54,55].

3. Results

3.1. Proteome Annotation and Clustering

We downloaded the genomes of 34 oomycete species and eight non-oomycete stramenopiles from the NCBI and FungiDB databases and annotated their proteomes by the presence or absence of known functional signatures to gain insights into the divergence of the dataset (Figure 1) [56,57]. The UPGMA clustering based on the Euclidean distance along with midpoint rooting resulted in two main groups, one corresponding to the oomycetes and the other to the remaining stramenopiles. The main difference among them was the lack of photosynthesis-related annotations in the oomycetes, such as chlorophyll biosynthesis (Figure S1). In the oomycetes, we defined three clusters based on their distance (1–3 in Figure 1): obligate biotrophs, Saprolegniaceae, and a final one grouping most of the Perosporanaceae and Pythiaceae of the dataset. The obligate biotroph cluster consisted of the Albuginaceae and the downy mildews from the Peronosporaceae (*Bremia lactucae*, *Plasmopara halstedii*, *Peronospora effusa* and *Hyalopernospora arabidopsidis*). The most striking characteristic was an overall reduction in their metabolism, evident by the lack of many functional annotations in comparison with other oomycetes. This lack of core biosynthetic pathways, including vitamin and cofactor biosynthesis, makes them reliant on their host for growth and survival (Figure S1). The Saprolegniaceae group differed from other oomycetes mainly in the presence of steroid biosynthesis pathways (Figure S2). In the third cluster, we defined two subclusters, labeled as 3.1 and 3.2 in Figure 1. The first contained four of the *Pythium* and *Globisporangium* species of the dataset, and the second one included exclusively all *Phythophthora* in the dataset (except for *Phytophthora megakarya*). The *Pythium* and *Globisporangium* species in the dataset also had biosynthetic pathways that most other oomycetes lacked and that they often shared with the Saprolegniaceae, as a result most likely of their common facultative lifestyles. The hemibiotroph group, consisting of most of the *Phytophthora* species in the dataset, showed significant metabolic reduction, but not as extensive as in the obligate biotrophs [58].

These clusters and subclusters roughly reflected the lifestyles of the taxa in the dataset, mostly highlighted by the hemibiotrophs and obligate biotrophs. To a lesser extent, this was evident in the other two groups as most Saprolegniaceae in the dataset are facultative animal necrotrophs, and most *Pythium* and *Globisporangium* species facultative plant necrotrophs. Interestingly, *T. clavata*, the free-living organism in the dataset, clustered as an outgroup of the other phylogenetically close Saprolegniaceae, showing the greatest distance to its animal and plant-infecting neighbours. The most notable differences in the presence/absence of cellular pathways of this *T. clavata* assembly when compared to other Saprolegniaceae were the absence of the endopeptidase ClpXP complex and RuvB-like helicase I (Figure S2). However, there were some exceptions to this arrangement, with some taxa clustering with a different lifestyle or failing to cluster with their own lifestyle. For

example, the clustering of the two plant infecting necrotrophs of the Saprolegniaceae follows the phylogeny of the *Aphanomyces* genus. Additionally, *Globisporangium splendes* appears as an outgroup of group 3.1 despite having a similar plant necrotrophic lifestyle and being phylogenetically closely related to other members in this clade (Figure S3). An explanation for this could be its long-read sequencing-based assembly. Long-read sequencing technology has been shown to produce much larger genomes in oomycetes when compared to the classical short-read sequencing, as hard-to-assemble repeat-rich regions are a common feature of their genomes [59]. However, other long-read assemblies in the dataset (*B. lactucae*, *Saprolegnia parasitica*, *Phytopythium vexans* and *Phytophthora fragariae*) show no apparent influence in the clustering (Figure 1).

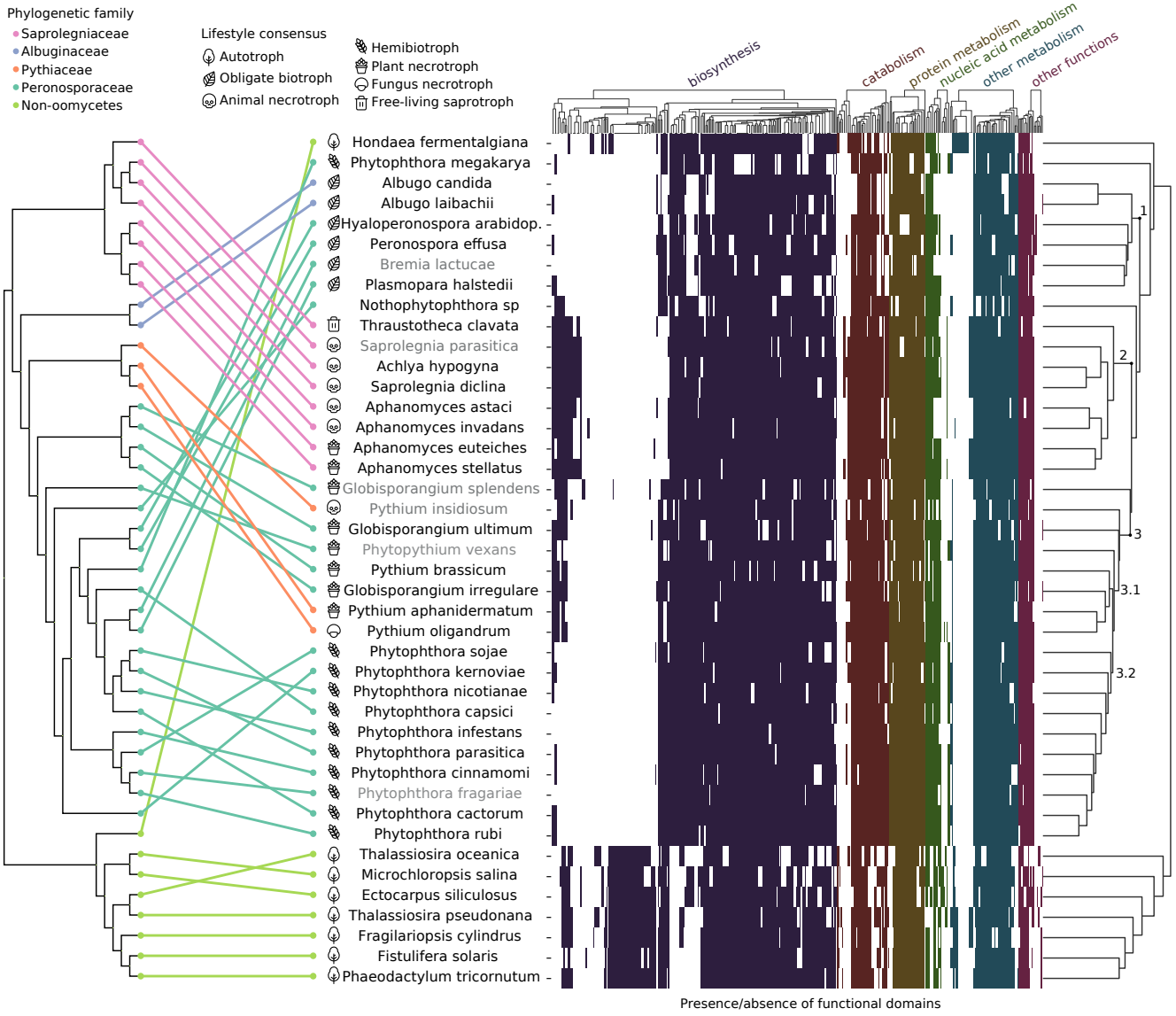


Figure 1. Presence/absence of functional attributes in the genomes of the stramenopile dataset correlated with phylogeny. Equal distance cladogram constructed from conserved families inferred by the maximum likelihood on the left and clustering by Unweighted Pair Group Method with Arithmetic Mean (UPGMA) of genome properties of the dataset on the right. In the equal distance phylogenetic tree, colored lines match phylogeny to the clustered taxa with annotated lifestyles. All nodes in the tree have a 100-bootstrap support. In the heatmap, different colors represent the presence or absence of particular functional groups belonging to the specified categories. Names in grey represent long-read sequencing assemblies in the dataset for that species.

3.2. Ortholog Group Classification

To infer positive selection from the stramenopile dataset of 42 genomes, we classified the proteomes into ortholog groups by taking sequence similarity and in addition gene order into account. We selected protein clusters that had at least five members from different taxa to obtain a good balance between a representative number of families and results that are statistically robust. This corresponded to 29,123 protein families, which cover about half (49.02%) of the total proteins in the dataset (Figure 2a). The orthogroups were mainly composed of one-to-one orthologs (78.70% of families); however, we detected a significant number of paralogs in some oomycetes, particularly for *Nothophytophthora* sp., as well as for *Phytophthora nicotianae*, *G. splendens* and *Phytophthora parasitica* (Figure 2b). This might be related to the reported whole genome duplications in *Phytopthora* species [60], as well as the recent hybridization event that gave rise to *Nothophytophthora* [61]. Additionally, the diatom *Fistulifera solaris*'s large presence of gene duplications highlights its recent whole genome duplication [62]. The larger presence of duplicates in *P. insidiosum* and *G. splendes* in comparison to their peers may originate from their long-read sequencing-based assembly, which is better able to resolve gene duplications.

The most abundant orthogroups had between five and nine members (Figure 2c). Orthogroups corresponding to all taxa were a minority. Instead, most orthogroups were present in closely related five- to ten-member clades. When looking at the number of genes not assigned to orthogroups in the oomycetes, the *Phytophthora* genus had the highest count (Figure 2a). This may be related to the large arsenal of unique effectors that lack conserved domains or homologs outside of their own species and play a large role in host adaptation. *Aphanomyces astaci* also had a high amount of genes outside of the orthogroups, most likely because of the recent expansions in its genome [63]. In summary, this highlights a patchy ortholog distribution in the dataset, with most protein families conserved only in phylogenetically close members of clades (Figure 2c). Despite this, a significant pool of ortholog protein families representative of the stramenopile genomes in the dataset could be inferred from the analysis as further discussed below.

3.3. Positive Selection Analyses

Positive selection screening for orthologous groups was performed by using first a site-specific codon model to detect families under selection. This was followed by a branch-site-specific codon model to detect the taxa experiencing positive selection on those genes. The number of genes under selection varied for the different phylogenetic clades. Members of the Saprolegniaceae and Pythiaceae, together with the necrotrophic *Globisporangium* had a higher count and therefore more genes under selection in orthogroups (mean = 1222, std = 152) than the remaining Peronosporaceae and the Albuginaceae (mean= 577, std = 245) (Figure S5). A special case was the hybrid *Nothophytophthora* sp., which had a comparable amount of positively selected genes to Pythiaceae and Saprolegniaceae, however, composed in great part by duplicated genes after speciation, 44.45% of the total (orange bar). When comparing necrotrophs, hemibiotrophs, and obligate biotrophs within the Peronosporaceae family (mean = 1344, 663, and 269, respectively), the trend was that of a decrease in the number of genes under positive selection with the increase in biotrophic potential (Figure 3).

To infer potential biases in our analyses, we tested for a correlation between the number of genes under positive selection and the amount of proteins classified into orthogroups for each taxa (Pearson's correlation, r = 0.50, p value < 0.01). A correlation of 0.5 suggested that there may be a larger number of positives because of more extensive testing in the oomycete species, as they have on average more members in the ortholog dataset. This bias is more evident in the non-oomycetes (Pearson's correlation, r = 0.52, p value = 0.18) than when considering just the oomycetes (Pearson's correlation, r = 0.15, p value = 0.39). As the proteomes of the non-oomycetes are overall smaller compared to oomycetes (Figure S4), we hypothesize that less extensive testing renders them more prone to this bias.

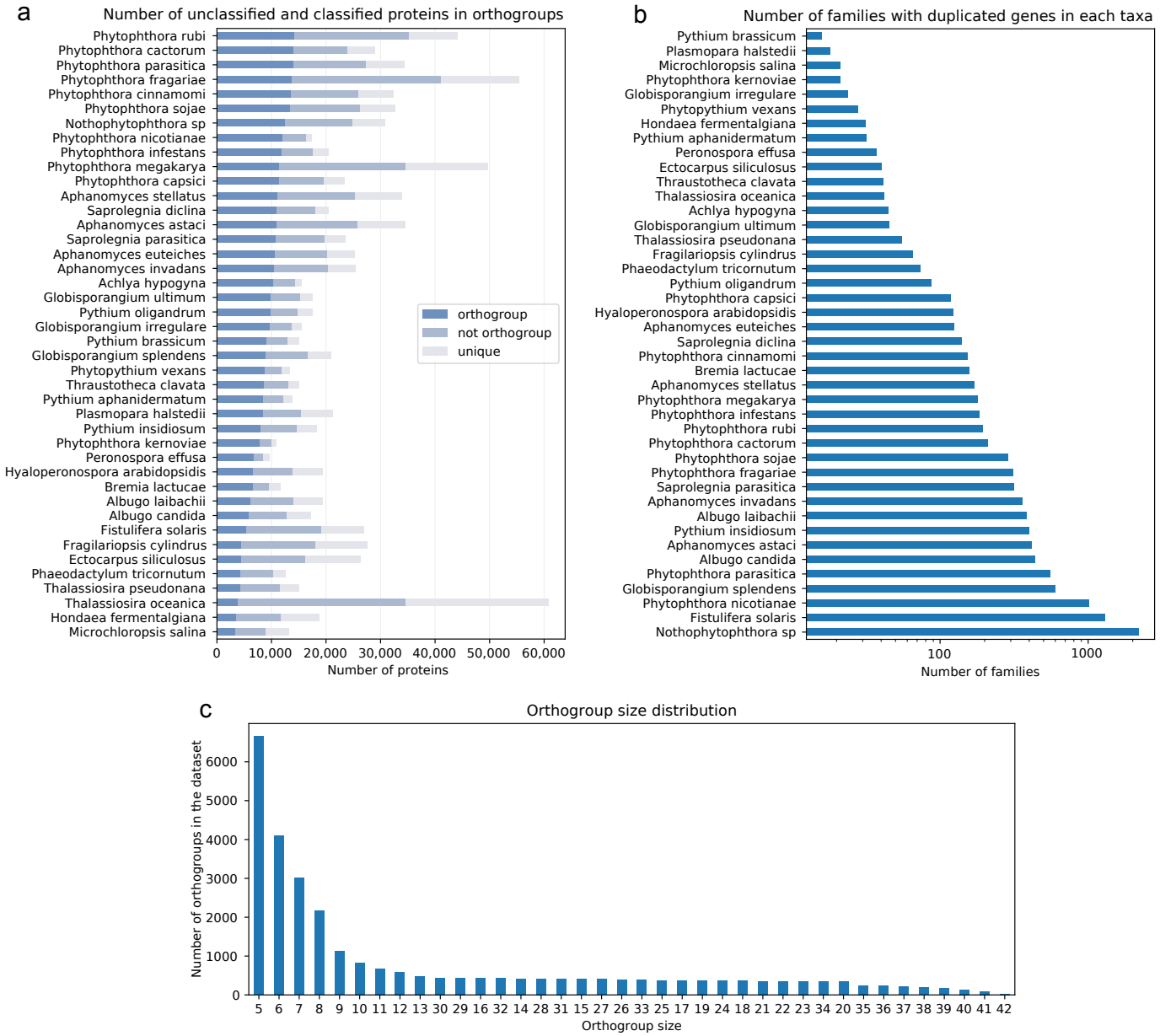


Figure 2. Orthogroup classification in the stramenopile dataset. (**a**) Protein-encoding genes from the dataset grouped into ortholog families. Number of genes classified into orthogroups (protein families of five or more members), not classified into orthogroups (protein families of less than five members), or unique (not in a protein family) are displayed per taxa. (**b**) Duplicates in protein families of the dataset. Number of ortholog families with five or more members from different taxa that contain paralogs (two or more genes from the same taxa). (**c**) Distribution of protein family size in the dataset. Number of families with the same member size are represented as a histogram.

Out of the 32,661 detected genes under positive selection, 21,247 were successfully annotated with at least a GO term (65%). We performed GO enrichment on the four main oomycete lifestyles in the stramenopile dataset. The results are discussed below. As a control for the reliability of the pipeline, we performed the same analyses in a subset of 26 plant pathogens from a dataset of 65 basidiomycete fungi (Table S1). Highly enriched terms included processes known to be associated with virulence in such pathogenic fungi, like fatty acid and certain amino acid biosynthesis, ion transport, and protein targeting and transport (Table S3) [64–66].

In summary, we could identify signatures of positive selection in 4.14% of all genes analyzed in the stramenopile dataset. A significant number could be functionally annotated and potential functions assigned.

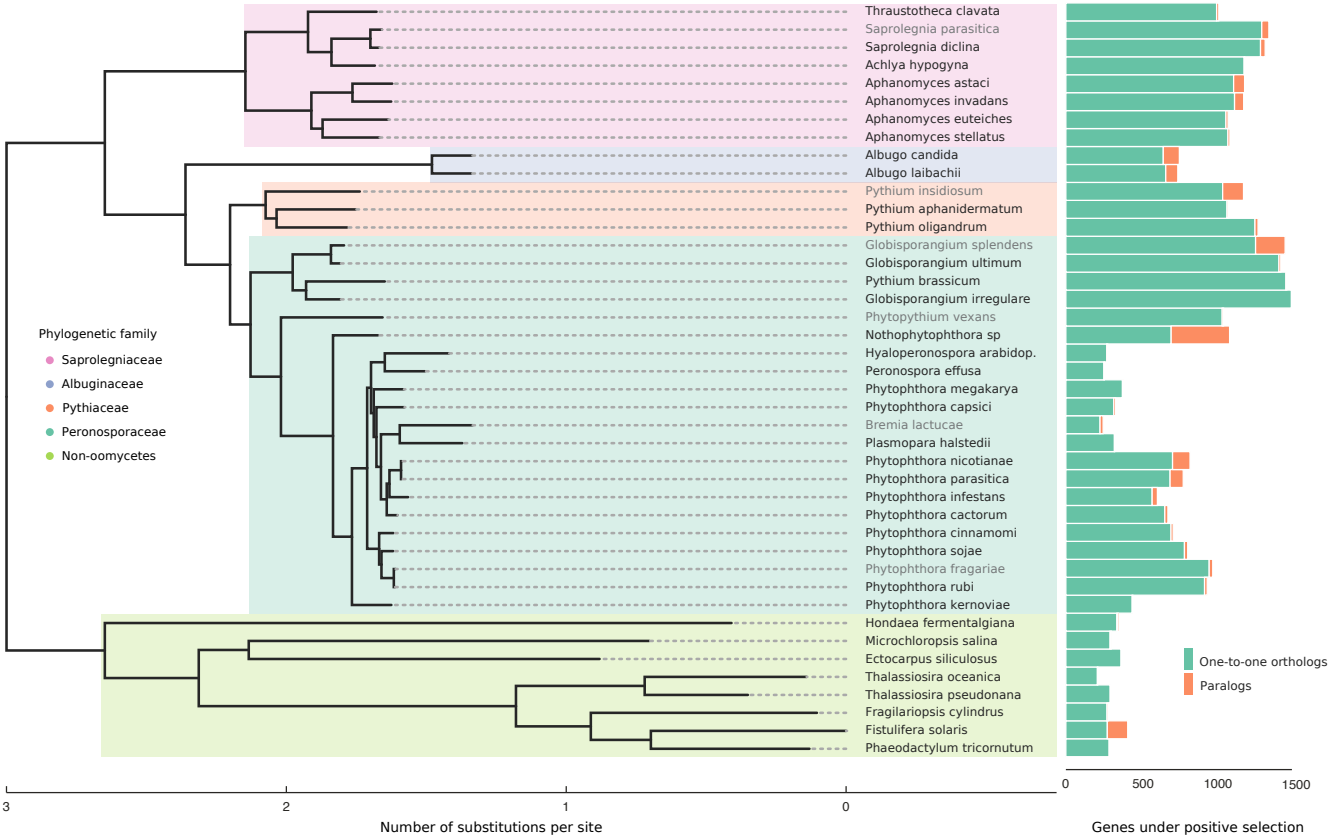


Figure 3. Number of genes under positive selection in the stramenopile dataset. The Maximum likelihood supertree with 100-bootstrap support in all nodes was constructed from inferred protein families in the stramenopile dataset that are conserved in at least 25 taxa, corresponding to 3013 families of orthologs. A number of positively selected genes are represented as bars. One-to-one orthologs are in green, duplicated genes inside the same family under positive selection in orange. Names in grey represent long-read sequencing assemblies in the dataset for that species.

3.4. Enriched Biological Functions under Selection

To gauge the selective pressures for adaptation to a parasitic lifestyle in the oomycetes, we explored the enriched GO terms that were pervasive in all oomycetes (Figure 4a). Highly enriched term categories related to response to stress, signal transduction, transmembrane transport, protein modification processes (phosphorylation, in particular), and localization, as well as numerous carbohydrate, lipid, nitrogen, and sulfur metabolism-related terms. Within the metabolism, abundant terms relating to biosynthesis are present. In the cellular compartment GO category, highly enriched terms include protein-containing complexes (for which transferase complexes show the larger significance), nucleus, intracellular organelles (for which ribosome shows the largest significance), and membranes (Figure 4b).

Additionally, we performed similar enrichments on the oomycete groups as defined by their lifestyle. We found the largest amount of unique GO terms to belong to the plant and animal necrotrophs (36 and 21, respectively). In the plant necrotrophs, these included terms related to ion transport, carbohydrate biosynthesis, protein modification, and gene expression regulation. In the animal necrotrophs, unique terms had to do with vitamin biosynthesis, cilium movement, and protein localization. There were three unique terms in the hemibiotrophs related to response against stress and transmembrane transport while no unique terms were identified in the obligate biotrophs. We observed the largest overlap between animal and plant facultative necrotroph groups (59 common terms). These terms

related to cell communication, glycolysis, organelle assembly, protein import, regulation of response to stimulus, translation, and numerous and diverse metabolic processes. This was followed by a smaller overlap of enriched functions in all four lifestyle groups, amounting to 33 terms (Figure 4c). The significant terms for each lifestyle are all listed in Tables S4–S9.

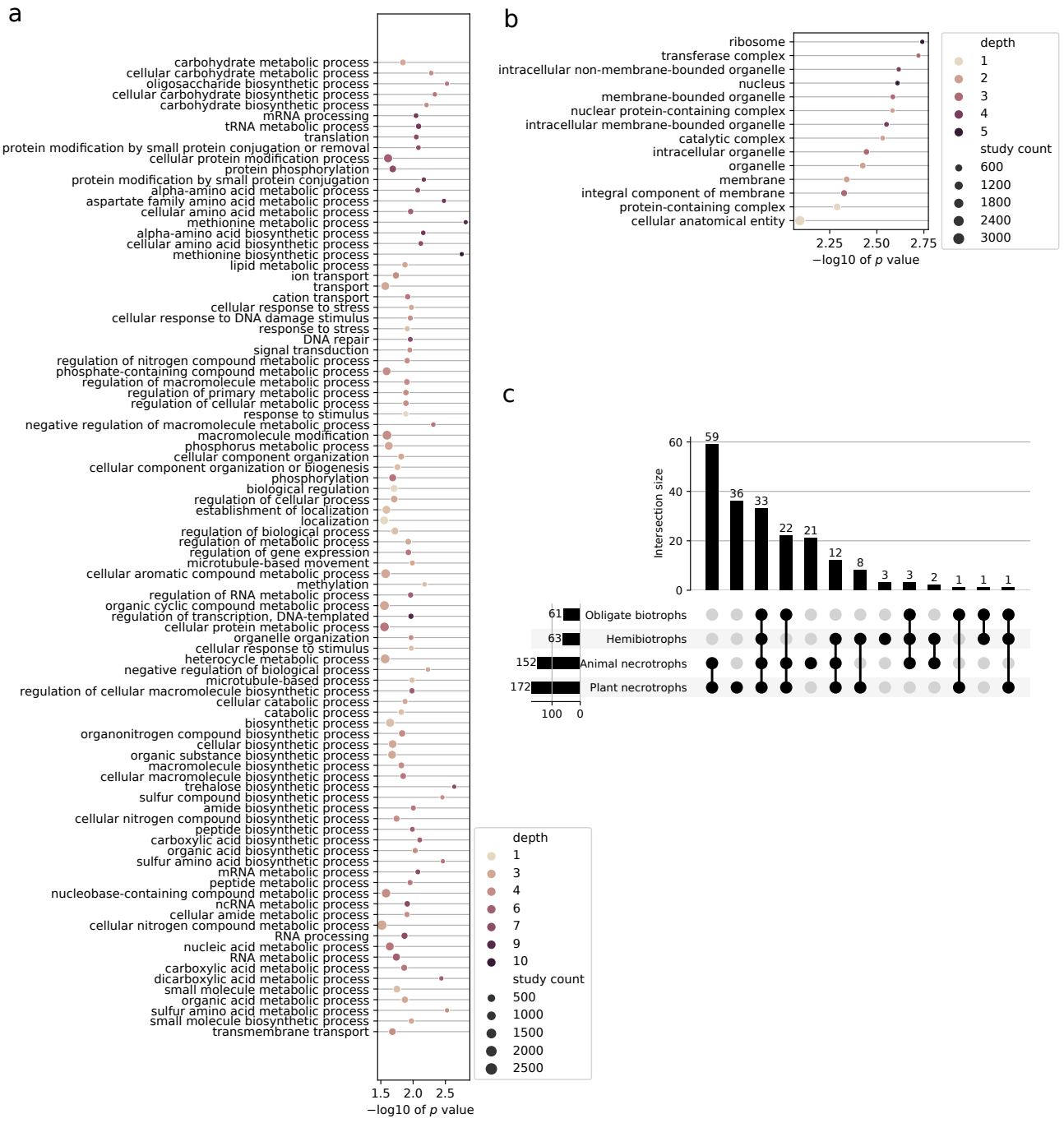


Figure 4. Functional and location enrichment of positively selected genes in the oomycetes. Significantly enriched (**a**) biological processes and (**b**) cellular compartments in all oomycetes that show signals of positive selection in the stramenopile dataset. Included are gene ontology (GO) terms with a corrected negative base 10 logarithm of the *p* value higher than 1.5 ordered by category using the GO slim database. The color represents the GO depth. GO depth is a measure of the number of parent nodes in the GO tree. That is, the more specific the GO term, the higher its depth. The size of the dots corresponds to the total number of proteins under selection in the stramenopile dataset that belong to said term. (**c**) Upset plot showing number of overlapping biological functions under selection in the oomycetes. The four groups correspond to the major lifestyles in the oomycetes of the stramenopile dataset.

We also studied the enrichment of biological functions in the expanded gene families of the dataset independently of whether the genes were under positive selection. In general, we found that it reflected positive selection enrichment; however, the terms were highly variable when comparing different species (Table S10). In the obligate biotrophs, these related to phospholipid metabolism, cell wall biosynthesis, protein modification, biological regulation, and transmembrane transport. In the hemibiotrophs, they related to lipid metabolism, signaling, protein modification, and again to biological regulation, and transmembrane transport. Finally, in the plant necrotrophs, they related to DNA integration and localization.

3.5. Lifestyle Prediction

We visualized in a heatmap all functional annotations with added information of positive selection by performing the same clustering as we did for the genome properties (Figure 5). We find that adding the positive selection data improves the clustering by lifestyle, particularly of the plant necrotrophs in the Pythiaceae and *Globisporangium*, which now form a single cluster that is closer to the other facultative necrotrophs of the dataset, the Saprolegniales, than to the obligate biotroph and hemibiotroph oomycetes in the dataset. Using the Robison–Foulds metric for clusters, we observed that there is a higher congruence between the phylogenetic tree and the genome properties clustering than to the positive selection one (Table 1).

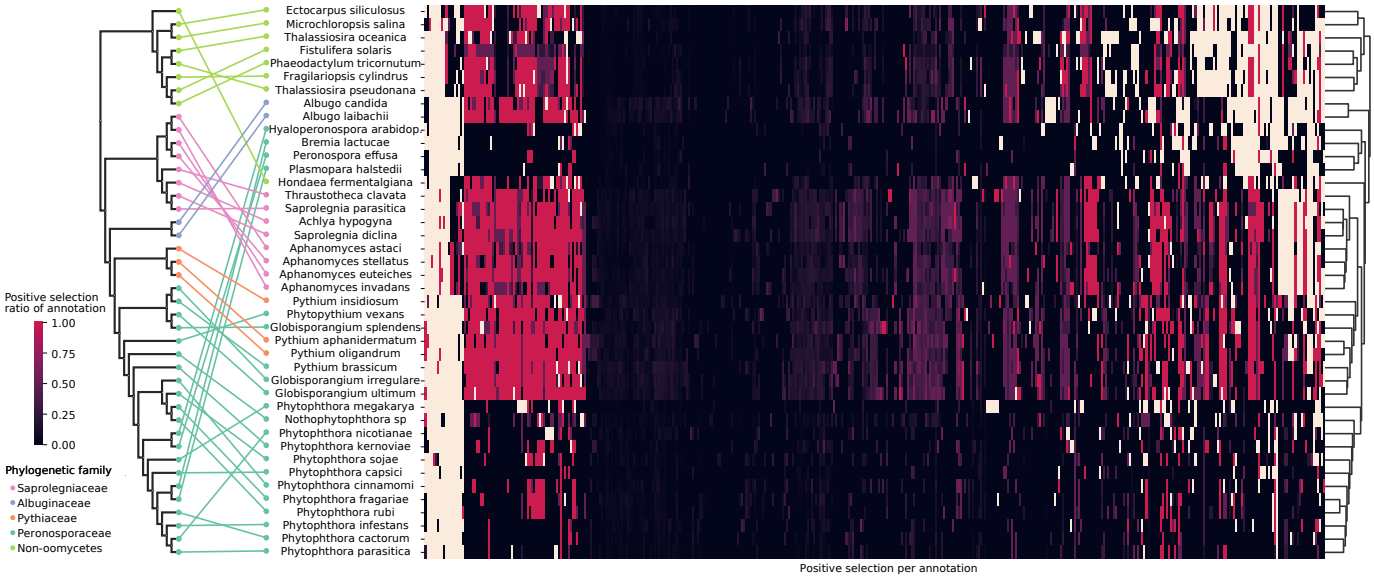


Figure 5. Heatmap of positive selection ratio of functional annotations in the stramenopile dataset. The color gradient from black to red in the heatmap represents the ratio of genes with a particular functional annotation that are under selection. Cream-colored cells represent the absence of the annotation for that species. The phylogenetic equal distance cladogram is represented on the left while UPGMA-based clustering of the distance between the taxa is represented on the right. Colored lines representing the phylogenetic family connect both clusterings.

Although positive selection information improved lifestyle clustering, we argue that it is impractical to implement as a prediction method because of its computationally demanding calculation and poor reproducibility when using different backgrounds for positive selection analyses. Therefore, we constructed a model to predict lifestyle in plant pathogenic fungi and oomycetes based on the genome properties alone. We assembled a dataset comprising 324 genomes from 115 plant pathogenic and saprotrophic fungal and oomycetal species (Table S2). We used the annoted genome properties as features to build a deep neural network classifier with four output classes corresponding to their lifestyle consensus in the literature: saprotroph, necrotroph, hemibiotroph and biotroph. We found a high accuracy on the validation dataset for the optimized model (loss = 0.11,

accuracy = 0.95), failing to predict two genomes in the hemibiotrophs and one in the biotrophs of the validation dataset (Figure 6). For comparison with published models, we additionally constructed a predictor based on Carbohydrate-Active enZymes (CAZymes), which also resulted in a high accuracy for the random validation dataset and performed better in the prediction of hemibiotrophs (loss = 0.14, accuracy = 0.97). Both models and the steps to reproduce them together with the entire dataset can be found at https://github.com/danielzmbp/lspred (accessed on 24 June 2021).

Table 1. Distance comparisons in the clusterings of the stramenopile dataset. Phylogenetic and genome properties clustering is shown in Figure 1 and positive selection clustering in Figure 5.

Clustering 1	Clustering 2	Robison–Foulds Distance Metric
Phylogenetic	Genome properties	28
Phylogenetic	Positive selection	30
Genome properties	Positive selection	24

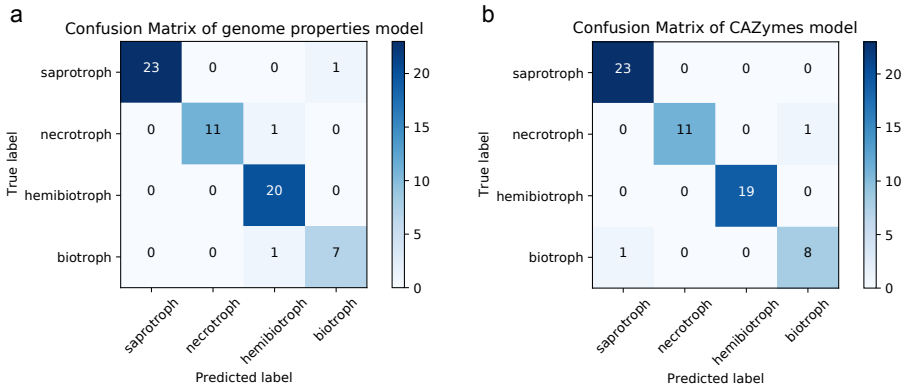


Figure 6. Confusion matrix of lifestyle predictor models. Prediction results in the random validation sets for the constructed models based on (**a**) genome properties and (**b**) Carbohydrate-Active enZymes (CAZymes), corresponding to 64 and 63 annotated proteomes, respectively (20% of the total dataset). True values are represented on the *x*-axis and predicted values on the *y*-axis.

4. Discussion

4.1. Functional Genome Annotations Largely Correlate with Lifestyle

Convergence of the presence/absence of key functional annotations in species that do not share the same phylogenetic history but have similar lifestyles has been shown before for different sets of organisms [67,68]. Distant species with the same lifestyle require similar functional biological processes, which results in similar selective pressures that analogously shape their genome, often leading to convergent evolution. Comparable to the study by Rodenburg et al. (2020) [69], we have shown the tight clustering of some oomycete groups with a similar lifestyle, most strikingly for the obligate biotrophs and hemibiotrophs. Conversely, we find a few exceptions in our dataset, such as the hemibiotroph *P. megakarya* and the necrotroph *G. splendens*, which do not clearly cluster with any of the other oomycetes. We hypothesize this may be partly due to the quality of their gene annotation. Both have a significantly lower number of complete key orthologs than the reference stramenopile BUSCO database as compared to other *Phytophthora* and *Globisporangium* species in the dataset (Table A1).

4.2. Generalists Have More Genes under Positive Selection

A higher number of genes under selection was found for the more generalist families of Saprolegniaceae, Pythiaceae, and necrotrophic Peronosporaceae, including the *Globisporangium* and *Phytopythium* clades, when compared to the higher number of specialists remaining in Peronosporaceae and Albuginaceae (Mann–Whitney test, $p < 0.01$). Within the Peronosporaceae, hemibiotrophs have a lower number of genes under selection than the facultative necrotrophs, and obligate biotrophs have in turn a lower number than hemibiotrophs (ANOVA one-tailed test, $p < 0.01$) (Figure S5). Thus, the number of genes under selection is inversely correlated to the biotrophic potential. With biotrophic potential, we refer to the capability of the pathogen to survive exclusively on a living host, such that no obligate biotroph can be cultured in vitro, while for some hemibiotrophs this is the case. On the opposite side of the spectrum, facultative plant necrotrophs thrive as saprotrophs without the need for a host. This correlation cannot be explained alone by the different sizes of the proteomes in the dataset or by their phylogenetic closeness (Figure S4). However, we hypothesize that both of these factors confound our results to a large extent. Smaller proteomes in the dataset, as is the case of the non-oomycetes, show a larger correlation of their size to the number of genes under positive selection. The phylogeny influence is highlighted by the similar number of genes under positive selection of taxa within the same genus as shown in Figure S4.

While all hemibiotrophs and biotrophs are obligate plant parasites, the necrotrophs in the Peronosporacea, Pythiaceae and Saprolegniaceae families show adaptation to a variety of lifestyles. They are facultative parasites of either animals, plants, or other fungi and oomycetes. Facultative parasites can live as saprotrophs on decaying matter but also as opportunistic necrotrophs on a suitable host [70]. The higher number of potential niches they are able to successfully occupy may drive a larger number of genes to be under positive diversifying selection. Additionally, when compared to the obligate biotrophs and hemibiotrophs, which are highly adapted to infect a particular species, e.g., lettuce for *B. lactucae* and soybean for *Phytophthora sojae*, most of the necrotrophs are able to infect a wide range of hosts. For instance, *A. astaci* is capable of infecting up to twelve genera of crayfish and is known for its ease of host jumping [71]. Having a higher number of genes under positive selection could be therefore correlated with this higher host flexibility.

4.3. Selective Pressures in the Ooomycetes Help Explain Host Adaptation

Biological functions under selection for all oomycetes in the stramenopile dataset, shown in Figure 4, give insight into which of these are important for the diversification in this clade. Many biosynthetic functions, particularly related to carbohydrates, are found to be enriched. Different cell wall carbohydrate composition has been found for different clades in the oomycetes, likely associated to different lifestyles [72]. Lipid metabolism, known also to be important for host adaptation in plant pathogenic fungi and oomycetes, is also enriched [73]. Transport-related proteins, and in particular cation transport, are also prominently enriched in these terms. As an example, the role of the expanded calcium transporter genes in the oomycetes has been extensively studied in the context of host interaction [74]. Overall, many of these terms allude to important virulence factors known for the oomycetes: transmembrane transport, effector protein processing and secretion, cell wall synthesis and remodeling, and lipid localization [75].

4.3.1. Selective Pressures Relate to Lifestyles in Oomycetes

The enriched terms common to the Albuginaceae and downy mildews greatly relate to known virulence factors for these plant pathogens, including carbohydrate metabolism, protein modification, transport, negative regulation of gene expression, and response to stimuli (Table S4). This suggests that these biological functions are under selection and played a big role in the adaptation of oomycetes to an obligate biotrophic lifestyle. Some of these, particularly carbohydrate metabolism, transport, and protein modification,

are common to the other plant pathogens in the hemibiotrophs and plant necrotrophs (Tables S5–S7), highlighting a broader mechanism of adaptation to a plant-parasitic lifestyle.

One of the most often found terms and among the most enriched in both the obligate biotrophs and the hemibiotrophs of the dataset corresponds to regulation of biosynthetic and metabolic processes, and particularly negative regulation. This may underscore the fitness advantage for rapid growth during the hyphal stage and its need for activation or deactivation according to the circumstances. When the hyphal stage takes place after colonization, the salvaging and biosynthesis of carbohydrates, nucleic acids, and lipids with the resources obtained from the plant host is key for a successful infection. Beta-glucan, for example, is an important component of the oomycete's cell wall and is also an elicitor of the plant immune response [76]. Its biosynthesis features prominently in the enriched terms for the hemibiotrophs. Following the hyphal stage and massive host colonization, there is a reproductive, mainly sporulating stage which needs to be tightly regulated. The latent period between infection and sporulation has been shown recently to correlate to different pathogenic lifestyles in leaf pathogens [77]. A widely diversified family of proteins in the oomycetes that take part in transcriptional activation and help regulate sporulation are Myb transcription factors [78,79]. The ortholog family of the gene *Myb3R7*, which was shown in *P. infestans* to be upregulated during sporulation, shows high rates of positive selection for the *Aphanomyces* and *Pythium* clades. Another gene associated to sporulation, classified as *CDC5*, was also found to be under selection in the oomycetes of the stramenopile dataset.

Secretion of small effector proteins, as in other fungal filamentous pathogens, is key for host adaptation in plant pathogenic oomycetes [80]. Many unique effector proteins have been characterized in the oomycetes that contribute to virulence by modulating the immune response of the plant [81]. Therefore, this dependence on the secretion machinery of the cell for successful infection and thus survival has led to high selective pressures on their genome. We observed significant enrichment of the effectors in the positively selected terms in all oomycetes of the dataset (hypergeometric test, $p < 0.01$). When looking at the enrichment per species; the majority of the *Phytophthora* and plant necrotrophs, which significantly depend on effector proteins for host infection, were also enriched (Figure 7a). The obligate biotrophs, which also depend greatly on secreted effectors, do not show enrichment in our analysis. This may be due to the lack of orthologs on host specific effectors and thus not them not being analyzed in the positive selection screen. There is a small correlation between the number of positively selected genes compared to those predicted to be effectors (Pearson's correlation, r = 0.22, $p = 0.22$), so these results may be slightly skewed due to testing bias (Figure 7b). In the GO enrichment of all oomycetes, there were several processes directly related to protein secretion under selection, including protein modification. Other secretion-related terms, although more general, also showed enrichment, including those relating to microtubule-based processes in the obligate biotrophs, and transmembrane transport in the hemibiotrophs.

Another interesting term indirectly related to effector proteins is sulfur amino acid biosynthesis. This term is highly enriched in the hemibiotrophs and the necrotrophs of the dataset. This may be associated with the abundance of cysteine-rich proteins in the effector arsenal of the plant pathogens with a necrotroph phase [82]. The disulfide bonds that link cysteine residues help maintain the structural integrity of the proteins released into the extracellular space called apoplast, a hostile environment that is slightly acidic and rich in plant proteases [83]. In general, elevated amino acid biosynthesis has been shown to be an important factor during infection of plant hemibiotrophs such as *P. infestans* [84]. We found enzymes related to amino acid biosynthesis, namely methionine synthase and ketol-acid reductoisomerase, to be under pervasive positive selective pressure in most oomycetes in the stramenopile dataset. These were previously found to be upregulated during early infection and the switch to necrotrophic phase in *P. infestans* [84].

When looking exclusively at the necrotroph groups, many terms in the plant pathogens overlap with the animal pathogens, most likely relating to their facultative saprobe lifestyle.

These include glycolysis, generation of energy, cell communication, as well as amino acid, tetrapyrrole, and amide biosynthetic processes. The latter group is most likely enriched as a result of their autotrophic and more developed secondary metabolism compared to that of other oomycetes, which makes them suited to a free-living lifestyle [85]. The highest enriched term under positive selection for both groups corresponds to the inosine salvage pathway, in particular the enzyme inosine 5′-monophosphate-specific 5′-nucleotidase. Nucleoside accumulation in the plant apoplast has been shown to increase susceptibility to the plant fungal necrotroph *Botrytis cinerea* [86]. A similar case in oomycete necrotrophs could have led to high selective pressure on this enzyme. Equally interesting is the term DNA ligation involved in DNA repair, which may be related to the defense against oxidative stress that is key to the immune response in plants and animals against such pathogens [87].

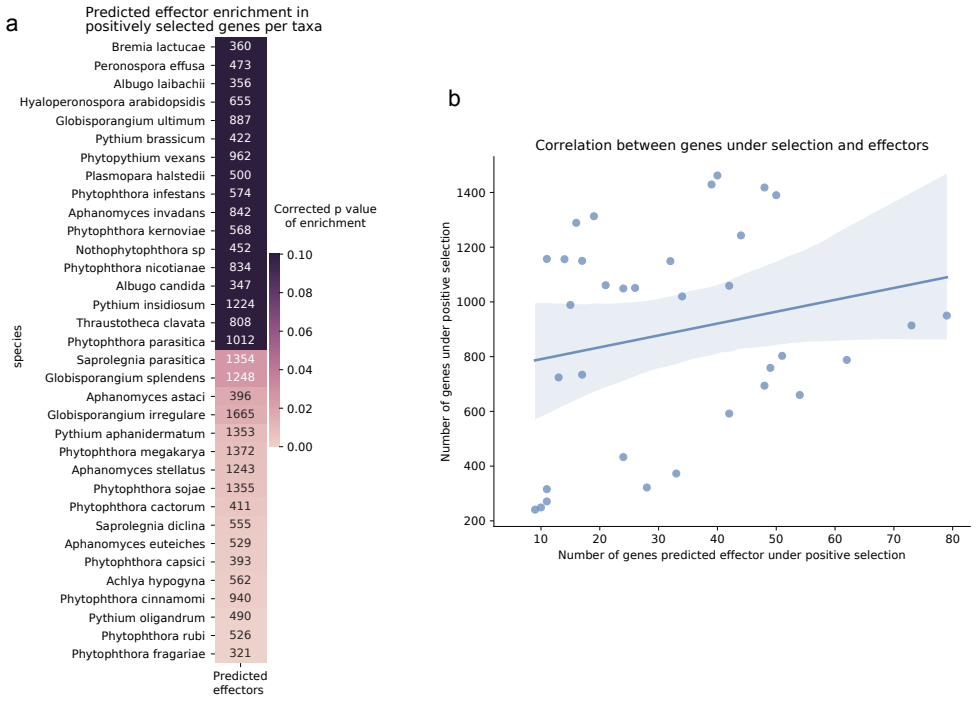


Figure 7. Effector proteins in the oomycetes of the stramenopile dataset. (**a**) Enrichment of genes coding for effector proteins under positive selection in oomycetes. The color gradient represents significant *p* values from hypergeometric tests per taxa corrected for multiple testing using Bonferroni. A lighter shade represents a more significant enrichment. The numbers within the cells represent the total effectors per proteome in the stramenopile dataset that were analyzed for positive selection. (**b**) Correlation between genes under selection and effectors in the oomycetes. Pearson correlation is represented as a straight line and the confidence interval is represented as a lighter shade (r = 0.22, *p* value = 0.22).

4.3.2. Biosynthetic Repertoire Is Important for Lifestyle Adaptation

As shown in Figure 1, the biosynthetic repertoire of each taxa plays a big role in defining the lifestyle of the organisms in the stramenopile dataset. Particularly interesting in oomycetes is the evolutionary history of sterol *de novo* biosynthesis. It is present in Saprolegniales and absent in other oomycete lineages due to their inability to synthesize oxidosqualene [88,89]. The squalene synthase shows hints of positive selection in *Aphanomyces* (Figure S6). Furthermore, positive selection is pervasive in the enzymes that take part in sterol biosynthesis in the stramenopile dataset.

Vitamin biosynthesis as well plays a big role in the evolution of pathogen adaptation to its host. Vitamins are expensive to produce and often require dedicated pathways. Heterotrophs that have adapted to obligate biotrophic lifestyles, such as *Albugo* and the downy mildews, circumvent this by losing their biosynthetic capabilities and developing ways of utilizing host vitamin supply, also known as auxotrophy [90]. Meanwhile, those that live without a host at any point in their lifecycle must maintain these pathways under strong purifying selection. In our dataset we found signatures of positive selection in several enzymes relating to tedrahydrofolate (THF) salvage and biosynthesis, namely dihydrofolate synthase and phosphoribosylglycinamide formyltransferase (Figure S7). As THF is a derivative of Vitamin B9 or folic acid, it is crucial for the synthesis of several amino acids such as serine and methionine as well as for purines and thiamine [91]. It is therefore likely that oomycetes that are not able to obtain THF from a living host have strong selection to maintain THF metabolism in order to ensure their own amino acid biosynthesis.

Molybdopterin cofactor is important for the production of certain detoxification enzymes [92]. In oomycete obligate biotrophs, molybdopterin-related biosynthetic pathways have been lost independently several times in the oomycetes' lineage due to host adaptation [15]. Molybdopterin metabolism was found under high selective pressure in the facultative necrotrophs and autotrophs of the stramenopile dataset, including *Saprolegniaceae* and *Pythiaceae* families, and *Phytophthora* genus (Figure S8). The biosynthesis of molybdopterin cofactor also features as an enriched GO term in the plant necrotrophs (Tables S6 and S7).

Proteins relating to the glycolysis pathway and amino acid biosynthesis have a special evolutionary history in the oomycetes [93]. Many of these enzymes originated from horizontal gene transfer from plants or bacteria. This might explain their high rate of positive selection, which is usually the case for genes recently acquired by horizontal transfer, as they need to be adapted to the new host. In the glycolysis pathway, we detected signatures of positive selection for most oomycetes in the stramenopile dataset, particularly in the enzymes glyceraldehyde-3-phosphate dehydrogenase and fructose-bisphosphate aldolase (Figure S9).

4.3.3. Protein Family Enrichment Reflects Lifestyle Selective Pressures

The large overrepresentation of paralogs as positively selected genes is evident in many of the taxa (Figure 3). After a gene duplication event occurs, there is usually an increase in the selective pressure on one of the copies that maintains the function. Meanwhile, in the other one, these constraints are relaxed, freeing it for potential divergent evolution [94]. Interestingly, many of the enriched functions in the paralogs correlated with terms under positive selection for their specific lifestyle (Table S11). In the *Phytophthora* lineages these include biological regulation, glycolipid biosynthesis, and transmembrane transport. In *Albugo* and other obligate biotrophs, protein modification, carbohydrate metabolism, biological regulation, and glutamine metabolism.

4.4. A Model Based on Genome Properties Accurately Predicts Lifestyle

The genome convergence of phylogenetically diverse fungi and oomycetes allowed us to create a model that can predict plant pathogenic lifestyle based on annotations from both eukaryotes. Assessment of lifestyle from genomic properties in plant pathogens has been traditionally performed by characterizing cell wall-degrading enzyme annotations [95]. To our knowledge, there is only one other published model that attempts to predict lifestyle from genomic features [96]. This model predicts trophic categories based on principal component analysis of CAZyme annotations. We find that our model, which in contrast is based on entire genome annotations, allows for better overall accuracy. However, its better performance may be partly related to its four output lifestyles compared to the seven classes in the model from Hane et al. (2020), since when comparing the genome properties model to a four-class model trained using similar parameters on CAZymes, the prediction is comparably precise (Figure 6). An advantage of the genome properties

model is that having trained it on a larger number of features per sample allows for a more accurate prediction of incompletely annotated specimens that may result from environmental sampling. Given the availability of increasing proteomic and transcriptomic data of unknown fungal and oomycetal origin, such prediction tools will become crucial to identify the pathogenic potential of facultative and obligate parasites.

5. Conclusions

The presence/absence of metabolism-related genes is known to converge for phylogenetically distant organisms that follow the same lifestyle [69,97]. Here, we report a similar case for our dataset of stramenopiles. We developed a pipeline for seamless throughput analysis of positive selective pressures using genome data as input and employed it to show that patterns of selective pressure also converge on hosts that cannot be explained by phylogeny alone. Clustering the taxa by their positive selection predictions, we showed improvement in the prediction of their lifestyle. Furthermore, we identified a number of biological functions that are commonly found under selection for all oomycetes independently of lifestyle. We explored and discussed lifestyle-specific adaptive genes that corresponded to biological regulation, transport, protein modification and metabolite biosynthesis. Finally, we described a model based on genome properties that is able to accurately predict the plant pathogenic lifestyle of filamentous fungi and oomycetes.

Overall, we believe our results highlight to date unexplored genes that could lead to further understanding of lifestyle evolution in oomycetes, and more broadly in filamentous pathogens. Experimental exploration on the impact these adaptations may have in the function and behaviour of the encoded proteins could improve prevention and prophylaxis against new emerging pandemic threats. Particularly, the enzymes reported in this manuscript, due to their likely adaptation as a consequence of a shift in host or lifestyle, may present interesting new targets which could be further explored for disease control.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/pathogens10070807/s1, Figure S1: Differences in annotated cellular pathways from the stramenopile dataset. Shown are pathways which have up to 36 repeated values per taxa. The clusters from Table 1 are encapsulated in a labeled square, Figure S2: Differences in annotated cellular pathways for the members of the Saprolegniaceae family in the stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa, Figure S3: Differences in annotated cellular pathways for the members of the Pythiaceae family and *Globisporangium* genus in the stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa, Figure S4: Correlation between genes under positive selection and proteome size in the stramenopile dataset. Oomycetes are in blue (Pearson's correlation, r = 0.15, p value = 0.39) and non-oomycetes in red (Pearson's correlation, r = 0.52, p value = 0.18). Pearson correlation is represented as a straight line and the confidence interval is represented as a lighter shade, Figure S5: Comparison of the distribution of positive genes under selection for different lifestyles. Significance between the different categories is $p < 0.01$ in both the upper graph (Mann–Whitney test) and the lower graph (ANOVA one-tailed test), Figure S6: Sterol biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to the sterol biosynthesis pathway in the stramenopiles. The yellow gradient represents the normalized ratio of the predicted positive selection in genes with this annotation, Figure S7: Tetrahydrofolate salvage and biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to tetrahydrofolate metabolism in the stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation, Figure S8: Molybdopterin biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to molybdopterin biosynthesis in the stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation, Figure S9: Glycolysis I, II and III-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to the glycolysis pathway in the stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation, Table S1: Summary of the basidiomycete dataset, Table S2: Summary of genomes used for the lifestyle model construction, Table S3: Significant GO terms with a depth higher than 7 found enriched in the

positively selected proteins in plant fungal pathogens, Table S4: Significantly enriched terms relating to biological processes in the positively selected obligate biotroph proteins, Table S5: Significantly enriched terms relating to biological processes in the positively selected hemibiotroph proteins, Table S6: Enriched terms relating to biological processes in the positively selected plant necrotrophs, Table S7: Enriched terms relating to biological processes in the positively selected plant necrotrophs (continued), Table S8: Enriched terms relating to biological processes in the positively selected animal necrotrophs, Table S9: Enriched terms relating to biological processes in the positively selected animal necrotrophs (continued), Table S10: Significant enriched terms relating to biological processes in the stramenopile dataset's paralogs.

Author Contributions: Conceptualization, D.G.-P. and E.K.; methodology, D.G.-P.; software, D.G.-P.; validation, D.G.-P.; formal analysis, D.G.-P.; investigation, D.G.-P.; resources, D.G.-P.; data curation, D.G.-P.; writing—original draft preparation, D.G.-P.; writing—review and editing, D.G.-P. and E.K.; visualization, D.G.-P.; supervision, E.K.; project administration, D.G.-P. and E.K.; funding acquisition, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the DFG-funded research training group RTG 1708 'Molecular principles of bacterial survival strategies' (grant number 174858087; Y.H.). We acknowledge support by the Open Access Publishing Fund of the University of Tübingen.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The stramenopile data presented in this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.4725040 (accessed on June 2021).

Acknowledgments: The authors would like to thank Libera Lo Presti and the anonymous reviewers for their helpful comments and suggestions. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CAZyme	Carbohydrate-Active enZyme
GO	Gene Ontology
THF	tetrahydrofolate
UPGMA	Unweighted Pair Group Method with Arithmetic Mean

Appendix A. Stramenopile Dataset

Table A1. Stramenopile genomes dataset used for positive selection analyses.

Phylogenetic Family	Species Name	Accession	Lifestyle	Complete BUSCOs	Complete and Single-Copy BUSCOs	Complete and Duplicated BUSCOs	Reference
Non-oomycete	Ectocarpus siliculosus	GCA_000310025.1	Autotroph	97	97	0	[98]
	Fistulifera solaris	GCA_002217885.1	Autotroph	97	14	83	[62]
	Fragilariopsis cylindrus	GCA_001750085.1	Autotroph	95	95	0	[99]
	Hondaea fermentalgiana	GCA_002897355.1	Autotroph	95	90	5	[100]
	Microchloropsis salina	GCA_004565275.1	Autotroph	92	90	2	[101]
	Phaeodactylum tricornutum	GCA_000150955.2	Autotroph	97	95	2	[102]
	Thalassiosira oceanica	GCA_000296195.2	Autotroph	90	90	0	[103]
	Thalassiosira pseudonana	GCA_000149405.2	Autotroph	97	95	2	[102]
Saprolegniaceae	Achlya hypogyna	GCA_002081595.1	Animal necrotroph [20]	99	98	1	[20]
	Aphanomyces astaci	GCA_000520075.1	Animal necrotroph [104]	100	82	18	
	Aphanomyces euteiches	GCA_009835175.1	Plant necrotroph [19]	99	99	0	
	Aphanomyces invadans	GCA_000520115.1	Animal necrotroph	100	83	17	
	Aphanomyces stellatus	GCA_009835185.1	Plant necrotroph [19]	97	96	1	
	Saprolegnia diclina	GCA_000281045.1	Animal necrotroph [105]	99	98	1	
	Saprolegnia parasitica	GCA_000151545.2	Animal necrotroph [105]	99	99	0	[106]
	Thraustotheca clavata	GCA_002081575.1	Free-living saprotroph [20]	99	98	1	[20]
Albuginaceae	Albugo candida	GCA_001078535.1	Obligate biotroph [107]	98	86	12	
	Albugo laibachii	PRJEA53219	Obligate biotroph [107]	95	82	13	[108]
Peronosporaceae	Bremia lactucae	GCA_004359215.1	Obligate biotroph [109]	96	90	6	[110]
	Globisporangium irregulare	GCA_000387425.2	Plant necrotroph [111]	98	96	2	[112]
	Globisporangium splendens	GCA_006386115.1	Plant necrotroph [113]	91	74	17	[114]
	Globisporangium ultimum	GCA_000143045.1	Plant necrotroph [115]	94	93	1	[112]
	Hyaloperonospora arabidopsidis	GCA_000173235.2	Obligate biotroph [116]	89	82	7	[116]
	Nothophytophthora sp.	GCA_001712635.2		90	28	62	[61]
	Peronospora effusa	GCA_003843895.1	Obligate biotroph [117]	94	93	1	
	Phytophthora cactorum	GCA_003287315.1	Hemibiotroph [118]	100	98	2	[119]
	Phytophthora capsici	GCA_000325885.1	Hemibiotroph [120]	98	97	1	[121]
	Phytophthora cinnamomi	GCA_001314365.1	Hemibiotroph [122]	96	94	2	[123]
	Phytophthora fragariae	GCA_009729455.1	Hemibiotroph	94	93	1	[124]
	Phytophthora infestans	GCA_000142945.1	Hemibiotroph [125]	100	99	1	
	Phytophthora kernoviae	GCA_001712645.2	Hemibiotroph [126]	96	96	0	[61]
	Phytophthora megakarya	GCA_002215365.1	Hemibiotroph	91	90	1	[127]
	Phytophthora nicotianae	GCA_001483015.1	Hemibiotroph [128]	99	86	13	[129]
	Phytophthora parasitica	GCA_000247585.2	Hemibiotroph [130]	98	87	11	
	Phytophthora rubi	GCA_009733145.1	Hemibiotroph	100	98	2	[124]
	Phytophthora sojae	GCA_000149755.2	Hemibiotroph [131]	99	98	1	[132]
	Phytopythium vexans	GCA_000387545.2	Plant necrotroph [17]	94	92	2	[17]
	Pythium brassicum	GCA_008271595.1	Plant necrotroph [133]	100	99	1	
	Plasmopara halstedii	GCA_900000015.1	Obligate biotroph [134]	100	100	0	
Pythiaceae	Pythium aphanidermatum	GCA_000387445.2	Plant necrotroph [135]	94	93	1	[17]
	Pythium insidiosum	GCA_001029375.1	Animal necrotroph [136]	99	87	12	[137]
	Pythium oligandrum	GCA_005966545.1	Fungal necrotroph [138]	100	100	0	[139]

References

1. Zhang, W.; Zhang, X.; Li, K.; Wang, C.; Cai, L.; Zhuang, W.; Xiang, M.; Liu, X. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. *Mycology* **2018**, *9*, 176–188. [CrossRef]
2. Props, R.; Monsieurs, P.; Vandamme, P.; Leys, N.; Denef, V.J.; Boon, N. Gene Expansion and Positive Selection as Bacterial Adaptations to Oligotrophic Conditions. *mSphere* **2019**, *4*. [CrossRef]
3. Scannell, D.R.; Byrne, K.P.; Gordon, J.L.; Wong, S.; Wolfe, K.H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **2006**, *440*, 341–345. [CrossRef] [PubMed]
4. Treangen, T.J.; Rocha, E.P.C. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. *PLoS Genet.* **2011**, *7*, e1001284. [CrossRef] [PubMed]
5. Sheridan, P.O.; Raguideau, S.; Quince, C.; Holden, J.; Zhang, L.; Gaze, W.H.; Holden, J.; Mead, A.; Raguideau, S.; Quince, C.; et al. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nat. Commun.* **2020**, *11*, 5494. [CrossRef]
6. Behe, M. Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". *Q. Rev. Biol.* **2010**, *85*, 419–445. [CrossRef]
7. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **1998**, *15*, 568–573. [CrossRef]
8. Hughes, A.L.; Packer, B.; Welch, R.; Bergen, A.W.; Chanock, S.J.; Yeager, M. Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15754–15757. [CrossRef] [PubMed]
9. Kosiol, C.; Vinař, T.; da Fonseca, R.R.; Hubisz, M.J.; Bustamante, C.D.; Nielsen, R.; Siepel, A. Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genet.* **2008**, *4*, e1000144. [CrossRef]
10. Cavalier-Smith, T.; Chao, E.E.Y. Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista). *J. Mol. Evol.* **2006**, *62*, 388–420. [CrossRef]
11. Beakes, G.W.; Thines, M. *Handbook of the Protists*; Springer: Cham, Switzerland, 2017; pp. 435–505._26. [CrossRef]
12. Matari, N.H.; Blair, J.E. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. *BMC Evol. Biol.* **2014**, *14*, 101. [CrossRef]
13. Bebber, D.P.; Gurr, S.J. Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet. Biol.* **2015**, *74*, 62–64. [CrossRef]
14. Derevnina, L.; Petre, B.; Kellner, R.; Dagdas, Y.F.; Sarowar, M.N.; Giannakopoulou, A.; De la Concepcion, J.C.; Chaparro-Garcia, A.; Pennington, H.G.; Van West, P.; et al. Emerging oomycete threats to plants and animals. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20150459. [CrossRef]
15. Kemen, E.; Jones, J.D. Obligate biotroph parasitism: Can we link genomes to lifestyles? *Trends Plant Sci.* **2012**, *17*, 448–457. [CrossRef]
16. Lee, S.J.; Rose, J.K.C. Mediation of the transition from biotrophy to necrotrophy in hemibiotrophic plant pathogens by secreted effector proteins. *Plant Signal. Behav.* **2010**, *5*, 769–772. [CrossRef]
17. Adhikari, B.N.; Hamilton, J.P.; Zerillo, M.M.; Tisserat, N.; Lévesque, C.A.; Buell, C.R. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. *PLoS ONE* **2013**, *8*, e75072. [CrossRef]
18. Steciow, M.M.; Lara, E.; Paul, C.; Pillonel, A.; Belbahri, L. Multiple barcode assessment within the Saprolegnia-Achlya clade (Saprolegniales, Oomycota, Straminipila) brings order in a neglected group of pathogens. *IMA Fungus* **2014**, *5*, 439–448. [CrossRef]
19. Gaulin, E.; Jacquet, C.; Bottin, A.; Dumas, B. Root rot disease of legumes caused by Aphanomyces euteiches. *Mol. Plant Pathol.* **2007**, *8*, 539–548. [CrossRef] [PubMed]
20. Misner, I.; Blouin, N.; Leonard, G.; Richards, T.A.; Lane, C.E. The Secreted Proteins of Achlya hypogyna and Thraustotheca clavata Identify the Ancestral Oomycete Secretome and Reveal Gene Acquisitions by Horizontal Gene Transfer. *Genome Biol. Evol.* **2015**, *7*, 120–135. [CrossRef] [PubMed]
21. McGowan, J.; O'Hanlon, R.; Owens, R.A.; Fitzpatrick, D.A. Comparative Genomic and Proteomic Analyses of Three Widespread Phytophthora Species: Phytophthora chlamydospora, Phytophthora gonapodyides and Phytophthora pseudosyringae. *Microorganisms* **2020**, *8*, 653. [CrossRef] [PubMed]
22. Weiblen, C.; Robe, L.J.; de Azevedo, M.I.; Ianiski, L.B.; Stibbe, P.C.; Ribeiro, T.C.; Zanette, R.A.; Pereira, D.I.B.; Santurio, J.M.; Botton, S.D.A. New insights on evolutionary aspects of *Pythium insidiosum* and other peronosporaleans. *Mycoses* **2020**. [CrossRef]
23. Thines, M. An evolutionary framework for host shifts—Jumping ships for survival. *New Phytol.* **2019**, *224*. [CrossRef]
24. Haas, B.J.; Kamoun, S.; Zody, M.C.; Jiang, R.H.Y.; Handsaker, R.E.; Cano, L.M.; Grabherr, M.; Kodira, C.D.; Raffaele, S.; Torto-Alalibo, T.; et al. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **2009**, *461*, 393–398. [CrossRef]
25. Richards, T.A.; Soanes, D.M.; Jones, M.D.M.; Vasieva, O.; Leonard, G.; Paszkiewicz, K.; Foster, P.G.; Hall, N.; Talbot, N.J. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15258–15263. [CrossRef]
26. Savory, F.; Leonard, G.; Richards, T.A. The Role of Horizontal Gene Transfer in the Evolution of the Oomycetes. *PLoS Pathog.* **2015**, *11*, e1004805. [CrossRef]

27. Savory, E.A.; Fuller, S.L.; Weisberg, A.J.; Thomas, W.J.; Gordon, M.I.; Stevens, D.M.; Creason, A.L.; Belcher, M.S.; Serdani, M.; Wiseman, M.S.; et al. Evolutionary transitions between beneficial and phytopathogenic Rhodococcus challenge disease management. *eLife* **2017**, *6*, e30925. [CrossRef] [PubMed]
28. Money, N.P.; Davis, C.M.; Ravishankar, J. Biomechanical evidence for convergent evolution of the invasive growth process among fungi and oomycete water molds. *Fungal Genet. Biol.* **2004**, *41*, 872–876. [CrossRef] [PubMed]
29. Latijnhouwers, M.; de Wit, P.J.; Govers, F. Oomycetes and fungi: Similar weaponry to attack plants. *Trends Microbiol.* **2003**, *11*, 462–469. [CrossRef]
30. Richards, T.A.; Dacks, J.B.; Jenkinson, J.M.; Thornton, C.R.; Talbot, N.J. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. *Curr. Biol.* **2006**, *16*, 1857–1864. [CrossRef] [PubMed]
31. Seppey, M.; Manni, M.; Zdobnov, E.M. Gene Prediction, Methods and Protocols. *Methods Mol. Biol.* **2019**, *1962*, 227–245. [CrossRef]
32. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef] [PubMed]
33. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
34. Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [CrossRef] [PubMed]
35. Nur, M.; Wood, K.; Michelmore, R. EffectorO: Motif-independent prediction of effectors in oomycete genomes using machine learning and lineage specificity. *BioRxiv* **2021**. [CrossRef]
36. Richardson, L.J.; Rawlings, N.D.; Salazar, G.A.; Almeida, A.; Haft, D.R.; Ducq, G.; Sutton, G.G.; Finn, R.D. Genome properties in 2019: A new companion database to InterPro for the inference of complete functional attributes. *Nucleic Acids Res.* **2018**, *47*, D564–D572. [CrossRef]
37. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
38. Robinson, D.; Foulds, L. Comparison of phylogenetic trees. *Math. Biosci.* **1981**, *53*, 131–147. [CrossRef]
39. Goluch, T.; Bogdanowicz, D.; Giaro, K. Visual TreeCmp: Comprehensive Comparison of Phylogenetic Trees on the Web. *Methods Ecol. Evol.* **2020**, *11*, 494–499. [CrossRef]
40. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A..; Lanfear, R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef]
41. Köster, J.; Rahmann, S. Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520–2522. [CrossRef]
42. Katoh, K.; Asimenos, G.; Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **2009**, *537*, 39–64. [CrossRef] [PubMed]
43. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]
44. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [CrossRef] [PubMed]
45. Lechner, M.; Findeiß, S.; Steiner, L.; Marz, M.; Stadler, P.F.; Prohaska, S.J. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinform.* **2011**, *12*, 124. [CrossRef] [PubMed]
46. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Pond, S.L.K. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [CrossRef] [PubMed]
47. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Pond, S.L.K.; Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. [CrossRef] [PubMed]
48. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
49. Carbon, S.; Douglass, E.; Dunn, N.; Good, B.; Harris, N.L.; Lewis, S.E.; Mungall, C.J.; Basu, S.; Chisholm, R.L.; Dodson, R.J.; et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2018**, *47*, gky1055. [CrossRef]
50. Klopfenstein, D.V.; Zhang, L.; Pedersen, B.S.; Ramírez, F.; Vesztrocy, A.W.; Naldi, A.; Mungall, C.J.; Yunes, J.M.; Botvinnik, O.; Weigel, M.; et al. GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **2018**, *8*, 10872. [CrossRef]
51. Hunter, S.; Apweiler, R.; Attwood, T.K.; Bairoch, A.; Bateman, A.; Binns, D.; Bork, P.; Das, U.; Daugherty, L.; Duquenne, L.; et al. InterPro: The integrative protein signature database. *Nucleic Acids Res.* **2009**, *37*, D211–D215. [CrossRef]
52. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 24 June 2021).
53. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, gky418. [CrossRef]
54. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; et al. Keras Tuner. 2019. Available online: https://github.com/keras-team/keras-tuner (accessed on 24 June 2021).
55. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *J. Mach. Learn. Res.* **2018**, *18*, 6765–6816.

56. Basenko, E.; Pulman, J.; Shanmugasundram, A.; Harb, O.; Crouch, K.; Starns, D.; Warrenfeltz, S.; Aurrecoechea, C.; Stoeckert, C.; Kissinger, J.; et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *J. Fungi* **2018**, *4*, 39. [CrossRef] [PubMed]
57. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef] [PubMed]
58. De Cock, A.; Lodhi, A.; Rintoul, T.; Bala, K.; Robideau, G.; Abad, Z.G.; Coffey, M.; Shahzad, S.; Lévesque, C. Phytopythium: Molecular phylogeny and systematics. *Persoonia Mol. Phylogeny Evol. Fungi* **2015**, *34*, 25–39. [CrossRef] [PubMed]
59. McGowan, J.; Fitzpatrick, D.A. Recent advances in oomycete genomics. *Adv. Genet.* **2020**, *105*, 175–228. [CrossRef]
60. Martens, C.; Van de Peer, Y. The hidden duplication past of the plant pathogen Phytophthora and its consequences for infection. *BMC Genom.* **2010**, *11*, 353. [CrossRef]
61. Studholme, D.J.; Panda, P.; Stowasser, E.S.V.; González, M.; Hill, R.; Sambles, C.; Grant, M.; Williams, N.M.; McDougal, R.L. Genome sequencing of oomycete isolates from Chile supports the New Zealand origin of *Phytophthora kernoviae* and makes available the first *Nothophytophthora* sp. genome: Comparative genomics of Chilean oomycete isolates. *Mol. Plant Pathol.* **2018**, *20*, 423–431. [CrossRef]
62. Tanaka, T.; Maeda, Y.; Veluchamy, A.; Tanaka, M.; Abida, H.; Maréchal, E.; Bowler, C.; Muto, M.; Sunaga, Y.; Tanaka, M.; et al. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* **2015**, *27*, 162–176. [CrossRef]
63. Gaulin, E.; Pel, M.J.C.; Camborde, L.; San-Clemente, H.; Courbier, S.; Dupouy, M.A.; Lengellé, J.; Veyssiere, M.; Ru, A.L.; Grandjean, F.; et al. Genomics analysis of *Aphanomyces* spp. identifies a new class of oomycete effector associated with host adaptation. *BMC Biol.* **2018**, *16*, 43. [CrossRef] [PubMed]
64. Stephenson, S.A.; Green, J.R.; Manners, J.M.; Maclean, D.J. Cloning and characterisation of glutamine synthetase from *Colletotrichum gloeosporioides* and demonstration of elevated expression during pathogenesis on *Stylosanthes guianensis*. *Curr. Genet.* **1997**, *31*, 447–454. [CrossRef]
65. Hallen, H.E.; Huebner, M.; Shiu, S.H.; Güldener, U.; Trail, F. Gene expression shifts during perithecium development in *Gibberella zeae* (anamorph *Fusarium graminearum*), with particular emphasis on ion transport proteins. *Fungal Genet. Biol.* **2007**, *44*, 1146–1156. [CrossRef]
66. Walley, J.W.; Kliebenstein, D.J.; Bostock, R.M.; Dehesh, K. Fatty acids and early detection of pathogens. *Curr. Opin. Plant Biol.* **2013**, *16*, 520–526. [CrossRef]
67. McCutcheon, J.P.; Moran, N.A. Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. *Genome Biol. Evol.* **2010**, *2*, 708–718. [CrossRef]
68. Shang, Y.; Xiao, G.; Zheng, P.; Cen, K.; Zhan, S.; Wang, C. Divergent and Convergent Evolution of Fungal Pathogenicity. *Genome Biol. Evol.* **2016**, *8*, 1374–1387. [CrossRef]
69. Rodenburg, S.Y.; De Ridder, D.; Govers, F.; Seidl, M.F. Oomycete metabolism is highly dynamic and reflects lifestyle adaptations. *BioRxiv* **2020**. [CrossRef]
70. Lewis, D.H. Concepts in Fungal Nutrition and the Origin of Biotrophy. *Biol. Rev.* **1973**, *48*, 261–277. [CrossRef]
71. Svoboda, J.; Mrugała, A.; Kozubíková-Balcarová, E.; Petrusek, A. Hosts and transmission of the crayfish plague pathogen *Aphanomyces astaci*: A review. *J. Fish Dis.* **2016**, *40*, 127–140. [CrossRef]
72. Mélida, H.; Sandoval-Sierra, J.V.; Diéguez-Uribeondo, J.; Bulone, V. Analyses of Extracellular Carbohydrates in Oomycetes Unveil the Existence of Three Different Cell Wall Types. *Eukaryot. Cell* **2013**, *12*, 194–203. [CrossRef]
73. Wang, E.; Schornack, S.; Marsh, J.; Gobbato, E.; Schwessinger, B.; Eastmond, P.; Schultze, M.; Kamoun, S.; Oldroyd, G. A Common Signaling Process that Promotes Mycorrhizal and Oomycete Colonization of Plants. *Curr. Biol.* **2012**, *22*, 2242–2246. [CrossRef]
74. Zheng, L.; Mackrill, J.J. Calcium Signaling in Oomycetes: An Evolutionary Perspective. *Front. Physiol.* **2016**, *7*, 123. [CrossRef]
75. Judelson, H.S. Dynamics and Innovations within Oomycete Genomes: Insights into Biology, Pathology, and Evolution. *Eukaryot. Cell* **2012**, *11*, 1304–1312. [CrossRef]
76. Klarzynski, O.; Plesse, B.; Joubert, J.M.; Yvin, J.C.; Kopp, M.; Kloareg, B.; Fritig, B. Linear β-1,3 Glucans Are Elicitors of Defense Responses in Tobacco. *Plant Physiol.* **2000**, *124*, 1027–1038. [CrossRef]
77. Précigout, P.A.; Claessen, D.; Makowski, D.; Robert, C. Does the Latent Period of Leaf Fungal Pathogens Reflect Their Trophic Type? A Meta-Analysis of Biotrophs, Hemibiotrophs, and Necrotrophs. *Phytopathology* **2020**, *110*, 345–361. [CrossRef]
78. Xiang, Q.; Judelson, H.S. Myb transcription factors in the oomycete *Phytophthora* with novel diversified DNA-binding domains and developmental stage-specific expression. *Gene* **2010**, *453*, 1–8. [CrossRef]
79. Xiang, Q.; Judelson, H.S. Myb Transcription Factors and Light Regulate Sporulation in the Oomycete *Phytophthora infestans*. *PLoS ONE* **2014**, *9*, e92086. [CrossRef]
80. Rocafort, M.; Fudal, I.; Mesarich, C.H. Apoplastic effector proteins of plant-associated fungi and oomycetes. *Curr. Opin. Plant Biol.* **2020**, *56*, 9–19. [CrossRef]
81. Wang, Y.; Tyler, B.M.; Wang, Y. Defense and Counterdefense during Plant-Pathogenic Oomycete Infection. *Annu. Rev. Microbiol.* **2019**, *73*, 667–696. [CrossRef]
82. Raffaele, S.; Win, J.; Cano, L.M.; Kamoun, S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of *Phytophthora infestans*. *BMC Genom.* **2010**, *11*, 637. [CrossRef] [PubMed]

83. Idänheimo, N.; Gauthier, A.; Salojärvi, J.; Siligato, R.; Brosché, M.; Kollist, H.; Mähönen, A.P.; Kangasjärvi, J.; Wrzaczek, M. The *Arabidopsis thaliana* cysteine-rich receptor-like kinases CRK6 and CRK7 protect against apoplastic oxidative stress. *Biochem. Biophys. Res. Commun.* **2014**, *445*, 457–462. [CrossRef]
84. Grenville-Briggs, L.J.; Avrova, A.O.; Bruce, C.R.; Williams, A.; Whisson, S.C.; Birch, P.R.; van West, P. Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet. Biol.* **2005**, *42*, 244–256. [CrossRef] [PubMed]
85. Dahlin, P.; Srivastava, V.; Ekengren, S.; McKee, L.S.; Bulone, V. Comparative analysis of sterol acquisition in the oomycetes *Saprolegnia parasitica* and *Phytophthora infestans*. *PLoS ONE* **2017**, *12*, e0170873. [CrossRef]
86. Daumann, M.; Fischer, M.; Niopek-Witz, S.; Girke, C.; Möhlmann, T. Apoplastic Nucleoside Accumulation in Arabidopsis Leads to Reduced Photosynthetic Performance and Increased Susceptibility against *Botrytis cinerea*. *Front. Plant Sci.* **2015**, *6*, 1158. [CrossRef]
87. Fones, H.; Preston, G.M. Reactive oxygen and oxidative stress tolerance in plant pathogenic *Pseudomonas*. *FEMS Microbiol. Lett.* **2012**, *327*, 1–8. [CrossRef] [PubMed]
88. Gottlieb, D. Differences in the Sterol Synthesizing Pathways of Sterol-Producing and Non-Sterol-Producing Fungi. *Phytopathology* **1978**, *68*, 1168. [CrossRef]
89. Gaulin, E.; Bottin, A.; Dumas, B. Sterol biosynthesis in oomycete pathogens. *Plant Signal. Behav.* **2010**, *5*, 258–260. [CrossRef] [PubMed]
90. Ah-Fong, A.M.V.; Kagda, M.S.; Abrahamian, M.; Judelson, H.S. Niche-specific metabolic adaptation in biotrophic and necrotrophic oomycetes is manifested in differential use of nutrients, variation in gene content, and enzyme evolution. *PLoS Pathog.* **2019**, *15*, e1007729. [CrossRef]
91. Huennekens, F. Folic Acid Coenzymes in the Biosynthesis of Purines and Pyrimidines. *Vitam. Horm.* **1969**, *26*, 375–394. [CrossRef]
92. Schwarz, G.; Mendel, R.R. Molybdenum Cofactor Biosynthesis and Molybdenum Enzymes. *Plant Biol.* **2006**, *57*, 623–647. [CrossRef] [PubMed]
93. Judelson, H.S. Metabolic Diversity and Novelties in the Oomycetes. *Annu. Rev. Microbiol.* **2016**, *71*, 21–39. [CrossRef]
94. Zhang, J.; Rosenberg, H.F.; Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3708–3713. [CrossRef]
95. King, B.C.; Waxman, K.D.; Nenni, N.V.; Walker, L.P.; Bergstrom, G.C.; Gibson, D.M. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnol. Biofuels* **2011**, *4*, 4. [CrossRef]
96. Hane, J.K.; Paxman, J.; Jones, D.A.B.; Oliver, R.P.; de Wit, P. "CATAStrophy", a Genome-Informed Trophic Classification of Filamentous Plant Pathogens—How Many Different Types of Filamentous Plant Pathogens Are There? *Front. Microbiol.* **2020**, *10*, 3088. [CrossRef]
97. Torruella, G.; de Mendoza, A.; Grau-Bové, X.; Antó, M.; Chaplin, M.; del Campo, J.; Eme, L.; Pérez-Cordón, G.; Whipps, C.; Nichols, K.; et al. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Curr. Biol.* **2015**, *25*, 2404–2410. [CrossRef] [PubMed]
98. Cock, J.M.; Sterck, L.; Rouzé, P.; Scornet, D.; Allen, A.E.; Amoutzias, G.; Anthouard, V.; Artiguenave, F.; Aury, J.M.; Badger, J.H.; et al. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* **2010**, *465*, 617–621. [CrossRef]
99. Mock, T. Extensive genetic diversity and differential bi-allelic expression in a Southern Ocean diatom. *Eur. J. Phycol.* **2015**, *50*, 75.
100. Seddiki, K.; Godart, F.; Cigliano, R.A.; Sanseverino, W.; Barakat, M.; Ortet, P.; Rébeillé, F.; Maréchal, E.; Cagnac, O.; Amato, A. Sequencing, De Novo Assembly, and Annotation of the Complete Genome of a New Thraustochytrid Species, Strain CCAP_4062/3. *Genome Announc.* **2018**, *6*, e01335-17. [CrossRef]
101. Ohan, J.A.; Hovde, B.T.; Zhang, X.L.; Davenport, K.W.; Chertkov, O.; Han, C.; Twary, S.N.; Starkenburg, S.R. Nuclear Genome Assembly of the Microalga Nannochloropsis salina CCMP1776. *Microbiol. Resour. Announc.* **2019**, *8*. [CrossRef]
102. Bowler, C.; Allen, A.E.; Badger, J.H.; Grimwood, J.; Jabbari, K.; Kuo, A.; Maheswari, U.; Martens, C.; Maumus, F.; Otillar, R.P.; et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **2008**, *456*, 239–244. [CrossRef]
103. Lommer, M.; Specht, M.; Roy, A.S.; Kraemer, L.; Andreson, R.; Gutowska, M.A.; Wolf, J.; Bergner, S.V.; Schilhabel, M.B.; Klostermeier, U.C.; et al. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **2012**, *13*, R66. [CrossRef]
104. Alderman, D.J.; Polglase, J.L.; Frayling, M. Aphanomyces astaci pathogenicity under laboratory and field conditions. *J. Fish Dis.* **1987**, *10*, 385–393. [CrossRef]
105. Willoughby, L.G. Saprolegnias of salmonid fish in Windermere: A critical analysis. *J. Fish Dis.* **1978**, *1*, 51–67. [CrossRef]
106. Jiang, R.H.Y.; de Bruijn, I.; Haas, B.J.; Belmonte, R.; Löbach, L.; Christie, J.; van den Ackerveken, G.; Bottin, A.; Bulone, V.; Díaz-Moreno, S.M.; et al. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen *Saprolegnia parasitica*. *PLoS Genet.* **2013**, *9*, e1003272. [CrossRef]
107. Ruhe, J.; Agler, M.T.; Placzek, A.; Kramer, K.; Finkemeier, I.; Kemen, E.M. Obligate Biotroph Pathogens of the Genus Albugo Are Better Adapted to Active Host Defense Compared to Niche Competitors. *Front. Plant Sci.* **2016**, *7*, 820. [CrossRef]
108. 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. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of *Arabidopsis thaliana*. *PLoS Biol.* **2011**, *9*, e1001094. [CrossRef]

109. Francis, D.M.; Hulbert, S.H.; Michelmore, R.W. Genome size and complexity of the obligate fungal pathogen, *Bremia lactucae*. *Exp. Mycol.* **1990**, *14*, 299–309. [CrossRef]
110. Fletcher, K.; Gil, J.; Bertier, L.D.; Kenefick, A.; Wood, K.J.; Zhang, L.; Reyes-Chin-Wo, S.; Cavanaugh, K.; Tsuchida, C.; Wong, J.; et al. Genomic signatures of somatic hybrid vigor due to heterokaryosis in the oomycete pathogen, *Bremia lactucae*. *BioRxiv* **2019**, 516526. [CrossRef]
111. Hancock, J.G. Seedling and Rootlet Diseases of Forage Alfalfa Caused by *Pythium irregulare*. *Plant Dis.* **1991**, *75*, 691. [CrossRef]
112. Lévesque, C.A.; Brouwer, H.; Cano, L.; Hamilton, J.P.; Holt, C.; Huitema, E.; Raffaele, S.; Robideau, G.P.; Thines, M.; Win, J.; et al. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **2010**, *11*, R73. [CrossRef] [PubMed]
113. Linde, C. Root and Root Collar Disease of *Eucalyptus grandis* Caused by *Pythium splendens*. *Plant Dis.* **1994**, *78*, 10061. [CrossRef]
114. Reghu, R.J.; Chellappan, B.V.; Beena, S.H.; Sasi, A.; Vasudevan, S.E.; Nair, A.S. Draft Genome Sequence of the Oomycete *Globisporangium splendens* Strain rgcb-1. *Microbiol. Resour. Announc.* **2020**, *9*. [CrossRef]
115. Howell, C.R. Suppression of *Pythium ultimum*-Induced Damping-Off of Cotton Seedlings by *Pseudomonas fluorescens* and its Antibiotic, Pyoluteorin. *Phytopathology* **1980**, *70*, 712. [CrossRef]
116. Baxter, L.; Tripathy, S.; Ishaque, N.; Boot, N.; Cabral, A.; Kemen, E.; Thines, M.; Ah-Fong, A.; Anderson, R.; Badejoko, W.; et al. Signatures of Adaptation to Obligate Biotrophy in the *Hyaloperonospora arabidopsidis* Genome. *Science* **2010**, *330*, 1549–1551. [CrossRef]
117. Lyon, R.; Correll, J.; Feng, C.; Bluhm, B.; Shrestha, S.; Shi, A.; Lamour, K. Population Structure of *Peronospora effusa* in the Southwestern United States. *PLoS ONE* **2016**, *11*, e0148385. [CrossRef]
118. Chen, X.R.; Zhang, B.Y.; Xing, Y.P.; Li, Q.Y.; Li, Y.P.; Tong, Y.H.; Xu, J.Y. Transcriptomic analysis of the phytopathogenic oomycete *Phytophthora cactorum* provides insights into infection-related effectors. *BMC Genom.* **2014**, *15*, 980. [CrossRef] [PubMed]
119. Armitage, A.D.; Lysøe, E.; Nellist, C.F.; Lewis, L.A.; Cano, L.M.; Harrison, R.J.; Brurberg, M.B. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. *PLoS ONE* **2018**, *13*, e0202305. [CrossRef]
120. Chen, X.R.; Huang, S.X.; Zhang, Y.; Sheng, G.L.; Li, Y.P.; Zhu, F. Identification and functional analysis of the NLP-encoding genes from the phytopathogenic oomycete *Phytophthora capsici*. *Mol. Genet. Genom.* **2018**, *293*, 931–943. [CrossRef] [PubMed]
121. Lamour, K.H.; Mudge, J.; Gobena, D.; Hurtado-Gonzales, O.P.; Schmutz, J.; Kuo, A.; Miller, N.A.; Rice, B.J.; Raffaele, S.; Cano, L.M.; et al. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. *Mol. Plant-Microbe Interact.* **2012**, *25*, 1350–1360. [CrossRef]
122. Santos, C.; Duarte, S.; Tedesco, S.; Fevereiro, P.; Costa, R.L. Expression Profiling of *Castanea* Genes during Resistant and Susceptible Interactions with the Oomycete Pathogen *Phytophthora cinnamomi* Reveal Possible Mechanisms of Immunity. *Front. Plant Sci.* **2017**, *8*, 515. [CrossRef]
123. Studholme, D.; McDougal, R.; Sambles, C.; Hansen, E.; Hardy, G.; Grant, M.; Ganley, R.; Williams, N. Genome sequences of six *Phytophthora* species associated with forests in New Zealand. *Genom. Data* **2016**, *7*, 54–56. [CrossRef]
124. Adams, T.M.; Armitage, A.D.; Sobczyk, M.K.; Bates, H.J.; Tabima, J.F.; Kronmiller, B.A.; Tyler, B.M.; Grünwald, N.J.; Dunwell, J.M.; Nellist, C.F.; et al. Genomic Investigation of the Strawberry Pathogen *Phytophthora fragariae* Indicates Pathogenicity Is Associated with Transcriptional Variation in Three Key Races. *Front. Microbiol.* **2020**, *11*, 490. [CrossRef] [PubMed]
125. Zuluaga, A.P.; Vega-Arreguín, J.C.; Fei, Z.; Ponnala, L.; Lee, S.J.; Matas, A.J.; Patev, S.; Fry, W.E.; Rose, J.K.C. Transcriptome of *P. infestans* in tomato. *Mol. Plant Pathol.* **2016**, *17*, 29–41. [CrossRef]
126. Denman, S.; Kirk, S.A.; Moralejo, E.; Webber, J.F. *Phytophthora ramorum* and *Phytophthora kernoviae* on naturally infected asymptomatic foliage. *EPPO Bull.* **2009**, *39*, 105–111. [CrossRef]
127. Ali, S.S.; Shao, J.; Lary, D.J.; Kronmiller, B.A.; Shen, D.; Strem, M.D.; Amoako-Attah, I.; Akrofi, A.Y.; Begoude, B.D.; Ten Hoopen, G.M.; et al. Phytophthora megakarya and Phytophthora palmivora, Closely Related Causal Agents of Cacao Black Pod Rot, Underwent Increases in Genome Sizes and Gene Numbers by Different Mechanisms. *Genome Biol. Evol.* **2017**, *9*, 536–557. [CrossRef]
128. Jimenez-Lopez, J.C.; Melser, S.; DeBoer, K.; Thatcher, L.F.; Kamphuis, L.G.; Foley, R.C.; Singh, K.B. Narrow-Leafed Lupin (*Lupinus angustifolius*) β1- and β6-Conglutin Proteins Exhibit Antifungal Activity, Protecting Plants against Necrotrophic Pathogen Induced Damage from *Sclerotinia sclerotiorum* and *Phytophthora nicotianae*. *Front. Plant Sci.* **2016**, *7*, 1856. [CrossRef]
129. Liu, H.; Ma, X.; Yu, H.; Fang, D.; Li, Y.; Wang, X.; Wang, W.; Dong, Y.; Xiao, B. Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*. *GigaScience* **2016**, *5*, 3. [CrossRef]
130. Cho, K.; Kim, Y.; Wi, S.J.; Seo, J.B.; Kwon, J.; Chung, J.H.; Park, K.Y.; Nam, M.H. Metabolic Survey of Defense Responses to a Compatible Hemibiotroph, *Phytophthora parasitica* var. *nicotianae*, in Ethylene Signaling-Impaired Tobacco. *J. Agric. Food Chem.* **2013**, *61*, 8477–8489. [CrossRef]
131. Moy, P.; Qutob, D.; Chapman, B.P.; Atkinson, I.; Gijzen, M. Patterns of Gene Expression Upon Infection of Soybean Plants by *Phytophthora sojae*. *Mol. Plant-Microbe Interact.* **2004**, *17*, 1051–1062. [CrossRef] [PubMed]
132. Tyler, B.M.; Tripathy, S.; Zhang, X.; Dehal, P.; Jiang, R.H.Y.; Aerts, A.; Arredondo, F.D.; Baxter, L.; Bensasson, D.; Beynon, J.L.; et al. *Phytophthora* Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science* **2006**, *313*, 1261–1266. [CrossRef]
133. Stanghellini, M.E.; Mohammadi, M.; Förster, H.; Adaskaveg, J.E. *Pythium brassicum* sp. nov.: A Novel Plant Family-Specific Root Pathogen. *Plant Dis.* **2014**, *98*, 1619–1625. [CrossRef] [PubMed]

134. Delmotte, F.; Giresse, X.; Richard-Cervera, S.; M'Baya, J.; Vear, F.; Tourvieille, J.; Walser, P.; de Labrouhe, D.T. Single nucleotide polymorphisms reveal multiple introductions into France of *Plasmopara halstedii*, the plant pathogen causing sunflower downy mildew. *Infect. Genet. Evol.* **2008**, *8*, 534–540. [CrossRef] [PubMed]
135. El-Tarabily, K.; Nassar, A.; Hardy, G.; Sivasithamparam, K. Plant growth promotion and biological control of *Pythium aphanidermatum*, a pathogen of cucumber, by endophytic actinomycetes. *J. Appl. Microbiol.* **2009**, *106*, 13–26. [CrossRef] [PubMed]
136. Gaastra, W.; Lipman, L.J.; Cock, A.W.D.; Exel, T.K.; Pegge, R.B.; Scheurwater, J.; Vilela, R.; Mendoza, L. *Pythium insidiosum*: An overview. *Vet. Microbiol.* **2010**, *146*, 1–16. [CrossRef]
137. Rujirawat, T.; Patumcharoenpol, P.; Lohnoo, T.; Yingyong, W.; Lerksuthirat, T.; Tangphatsornruang, S.; Suriyaphol, P.; Grenville-Briggs, L.J.; Garg, G.; Kittichotirat, W.; et al. Draft Genome Sequence of the Pathogenic Oomycete *Pythium insidiosum* Strain Pi-S, Isolated from a Patient with Pythiosis. *Genome Announc.* **2015**, *3*, e00574-15. [CrossRef]
138. Deacon, J. Studies on *Pythium oligandrum*, an aggressive parasite of other fungi. *Trans. Br. Mycol. Soc.* **1976**, *66*, 383–391. [CrossRef]
139. Faure, C.; Veyssière, M.; Boëlle, B.; San Clemente, H.; Bouchez, O.; Lopez-Roques, C.; Chaubet, A.; Martinez, Y.; Bezouška, K.; Suchánek, M.; et al. Long-Read Genome Sequence of the Sugar Beet Rhizosphere Mycoparasite *Pythium oligandrum*. *G3 Genes Genomes Genet.* **2020**, *10*, 431–436. [CrossRef]

Figure S1. Differences in annotated cellular pathways from the stramenopile dataset. Shown are pathways which have up to 36 repeated values per taxa. The clusters from Figure 1 are encapsulated in a labeled square.

Figure S2. Differences in annotated cellular pathways for the members of the Saprolegniaceae family in the stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa.

Figure S3. Differences in annotated cellular pathways for the members of the Pythiaceae family and *Globisporangium* genus in the stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa.

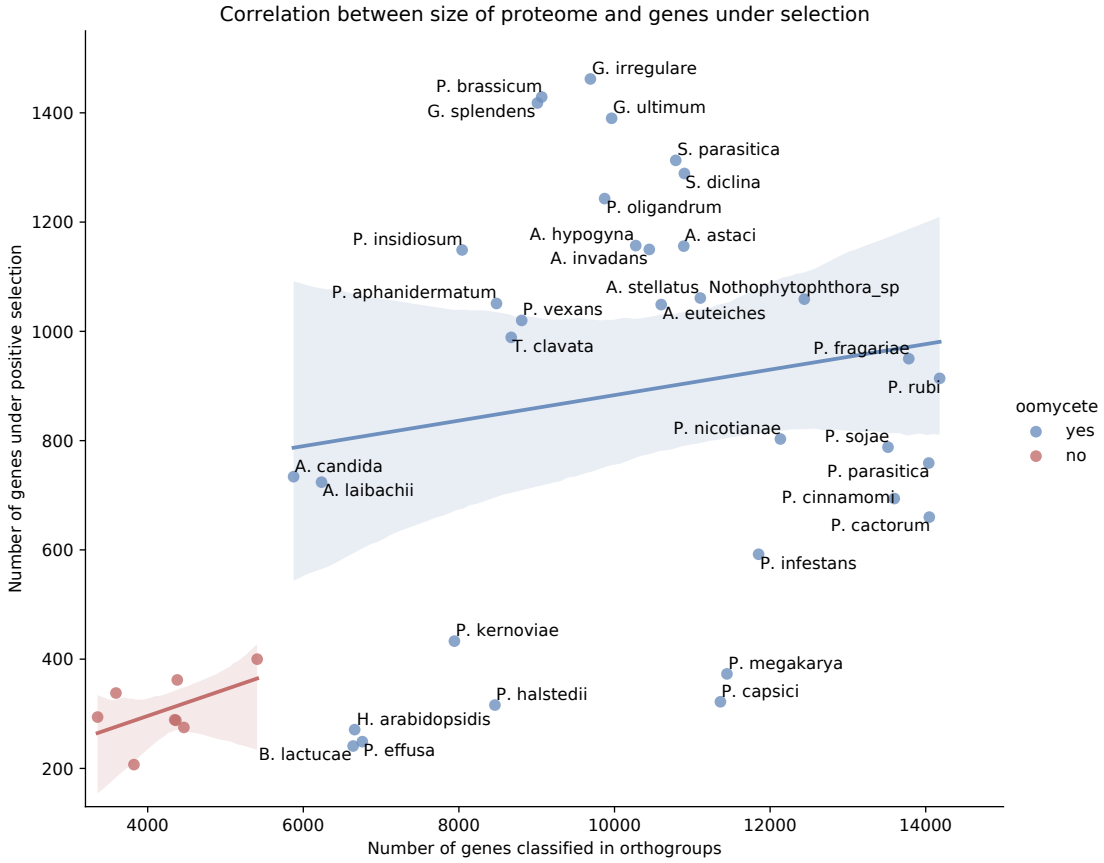


Figure S4. Correlation between genes under positive selection and proteome size in the stramenopile dataset. Oomycetes are in blue (Pearson's correlation, r = 0.15, *p* value = 0.39) and non-oomycetes in red (Pearson's correlation, r = 0.52, *p* value = 0.18). Pearson correlation represented as a straight line and the confidence interval represented as a lighter shade.

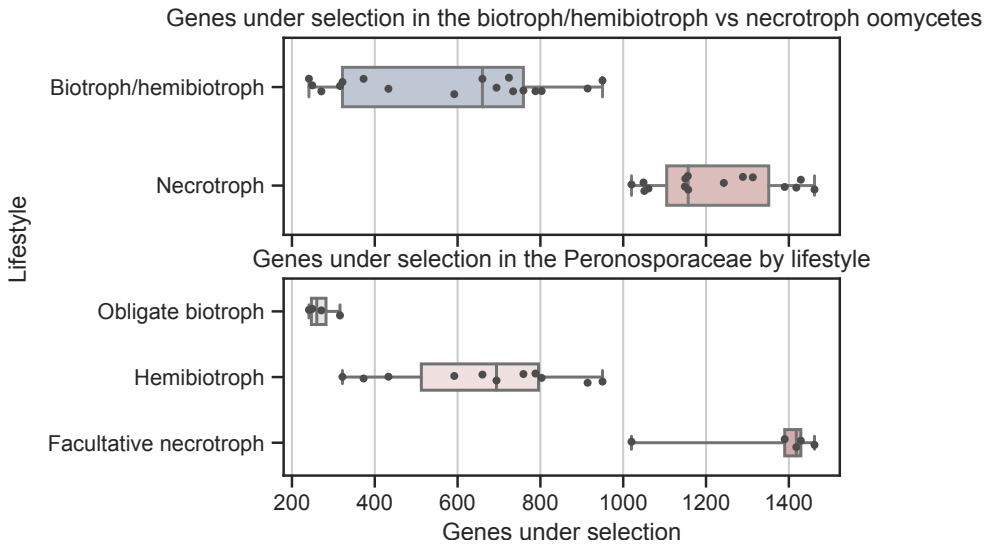


Figure S5. Comparison of the distribution of positive genes under selection for different lifestyles. Significance between the different categories is $p < 0.01$ in both the upper graph (Mann-Whitney test) and the lower graph (ANOVA one-tailed test).

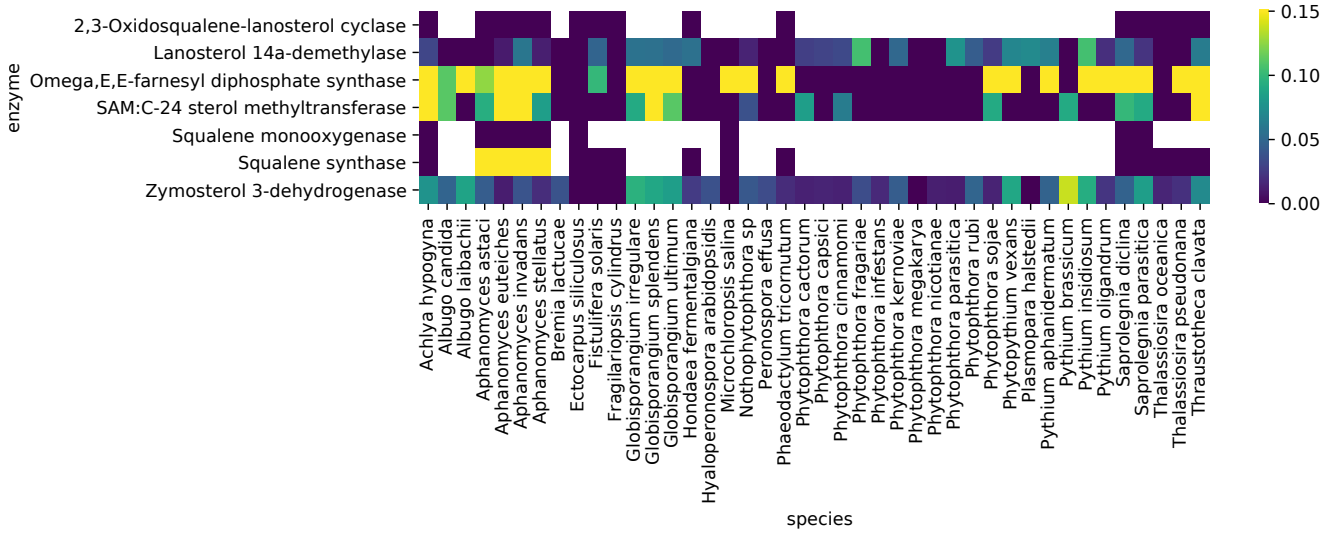


Figure S6. Sterol biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to sterol biosynthesis pathway in the stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.

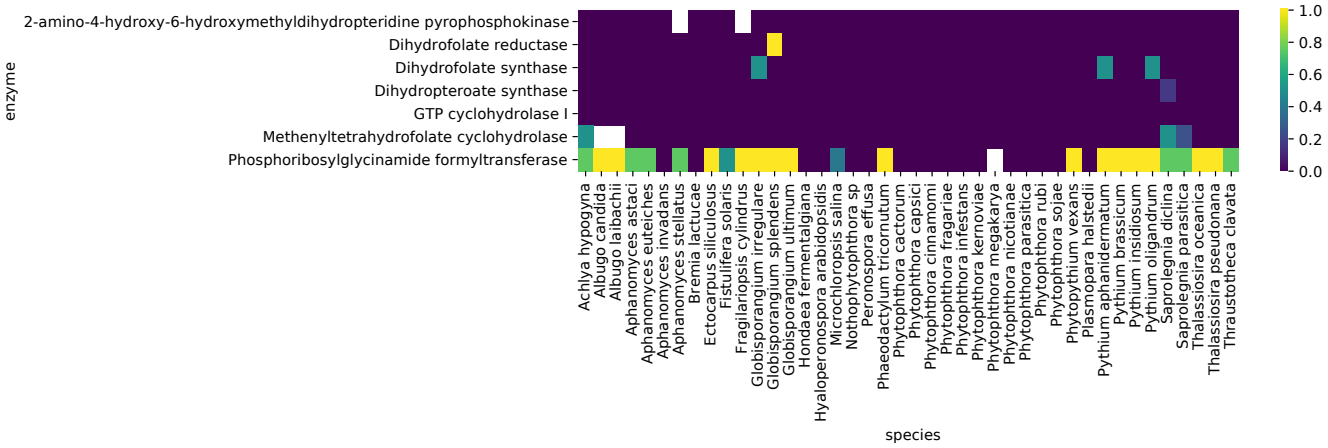


Figure S7. Tetrahydrofolate salvage and biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to tetrahydrofolate metabolism in the stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

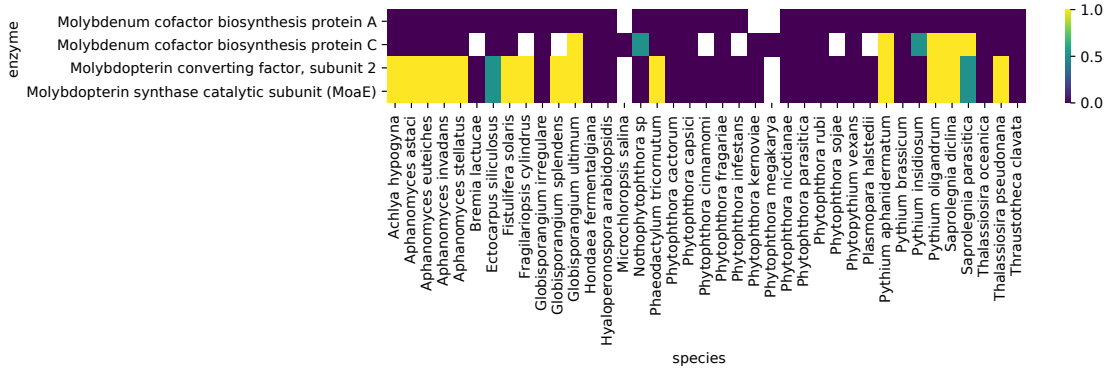


Figure S8. Molybdopterin biosynthesis-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to molybdopterin biosynthesis in the stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

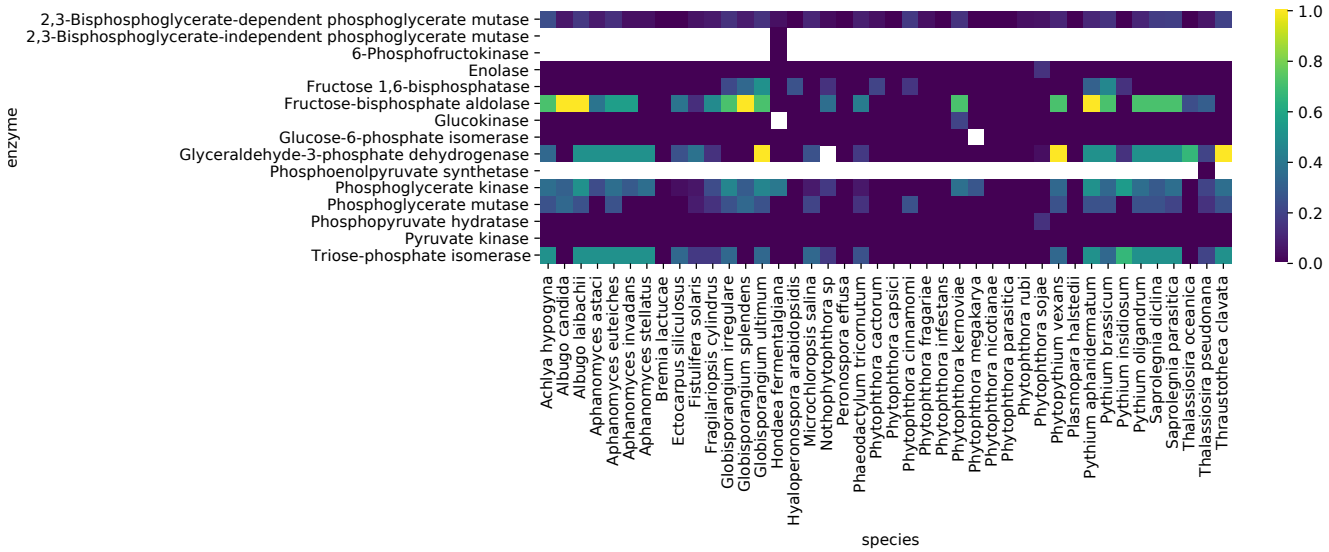


Figure S9. Glycolysis I, II and III-related enzymes in stramenopiles. Heatmap of the presence and absence of the enzymes relating to glycolysis pathway in the stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.

Table S1. Summary of basidiomycete dataset.

Species name	Plant pathogen	Accession
Acaromyces ingoldii	no	GCA_003144295.1
Anthracocystis flocculosa	no	GCA_000417875.1
Apiotrichum porosum	no	GCA_003942205.1
Ceraceosorus bombacis	yes	GCA_900000165.1
Ceraceosorus guamensis	no	GCA_003144195.1
Ceratobasidium theobromae	yes	GCA_009078325.1
Cryptococcus amylolentus	no	GCA_001720205.1
Cryptococcus gattii	no	GCA_000855695.1
Cryptococcus neoformans	no	GCA_000149245.3
Cryptococcus wingfieldii	no	GCA_001720155.1
Cutaneotrichosporon oleaginosum	no	GCA_001027345.1
Fomitiporia mediterranea	yes	GCA_000271605.1
Jaapia argillacea	no	GCA_000697665.1
Jaminaea rosea	no	GCA_003144245.1
Kalmanozyma brasiliensis	no	GCA_000497045.1
Kockovaella imperatae	no	GCA_002102565.1
Kwoniella bestiolae	no	GCA_000512585.2
Kwoniella dejecticola	no	GCA_000512565.2
Kwoniella pini	no	GCA_000512605.2
Leucosporidium creatinivorum	no	GCA_002105055.1
Malassezia globosa	no	GCA_000181695.1
Malassezia restricta	no	GCA_003290485.1
Malassezia sympodialis	no	GCA_000349305.2
Meira miltonrushii	no	GCA_003144205.1
Melampsora larici-populina	yes	GCA_000204055.1
Microbotryum lychnidis-dioicae	yes	GCA_000166175.1
Mixia osmundae	yes	GCA_000708205.1
Moesziomyces antarcticus	no	GCA_000747765.1
Moesziomyces aphidis	no	GCA_000517465.1
Moniliophthora roreri	yes	GCA_001466705.1
Paxillus involutus	no	GCA_000827475.1
Peniophora sp	no	GCA_900536885.1
Piloderma croceum	no	GCA_000827315.1
Pseudomicrostroma glucosiphilum	no	GCA_003144135.1
Pseudozyma hubeiensis	no	GCA_000403515.1
Puccinia coronata	yes	GCA_002873125.1
Puccinia graminis	yes	GCA_000149925.1
Puccinia sorghi	yes	GCA_001263375.1
Puccinia striiformis	yes	GCA_002920065.1
Puccinia triticina	yes	GCA_000151525.2
Rhizoctonia solani	yes	GCA_000524645.1
Rhodotorula graminis	no	GCA_001329695.1
Rhodotorula toruloides	no	GCA_000320785.2
Saitozyma podzolica	no	GCA_003942215.1
Serendipita indica	no	GCA_000313545.1
Serendipita vermifera	no	GCA_000827415.1
Sporisorium graminicola	no	GCA_005498985.1
Sporisorium reilianum	yes	GCA_900162835.1
Sporisorium scitamineum	yes	GCA_001243155.1
Testicularia cyperi	yes	GCA_003144125.1
Tilletia controversa	yes	GCA_001645045.2
Tilletia laevis	yes	GCA_009428275.1
Tilletia walkeri	yes	GCA_009428295.1
Tilletiaria anomala	yes	GCA_000711695.1
Tilletiopsis washingtonensis	yes	GCA_003144115.1
Trichosporon asahii	no	GCA_000293215.1
Ustilago bromivora	yes	GCA_900080155.1
Ustilago hordei	yes	GCA_000286035.1
Ustilago maydis	yes	GCA_000328475.2
Ustilago trichophora	yes	GCA_900323505.1
Violaceomyces palustris	no	GCA_003144235.1
Wallemia hederae	no	GCA_004918325.1
Wallemia ichthyophaga	no	GCA_000400465.1
Wallemia mellicola	no	GCA_000263375.1
Xanthophyllomyces dendrorhous	no	GCA_001007165.2

Table S2. Summary of genomes used for the lifestyle model construction.

Species name	Number of proteomes	Lifestyle
Agaricus bisporus	1	S
Albugo candida and laibachii	2	B
Alternaria alternata, arborescens, gaisen and tenuissima	14	N
Aphanomyces euteiches and stellatus	2	N
Ascochyta rabiei	1	N
Aspergillus fumigatus, nidulans, oryzae and niger	34	S
Bipolaris maydis, oryzae, victoriae and zeicola	5	N
Bipolaris sorokiniana	2	H
Blumeria graminis	4	B
Botrytis cinerea	3	N
Bremia lactucae	1	B
Colletotrichum fioriniae, gloeosporioides, graminicola, higginsianum, incanum, nymphaeae, orbiculare, simmondsii and sublineola	14	H
Debaryomyces hansenii	1	S
Dothistroma septosporum	1	H
Erysiphe necator	1	B
Eutypa lata	1	N
Fusarium culmorum and graminearum	6	H
Gigaspora margarita	1	B
Globisporangium irregulare, splendens and ultimum	3	N
Gloeophyllum trabeum	1	S
Hyaloperonospora arabidopsidis	1	B
Komagataella phaffii	5	S
Leptosphaeria maculans	1	H
Macrophomina phaseolina	1	H
Marssonina brunnea	1	H
Melampsora laricis-populina	1	B
Microbotryum violaceum	1	B
Monilinia laxa	1	N
Moniliophthora perniciosa and roreri	3	H
Neurospora crassa	2	S
Oidium neolycopersici	1	B
Parastagonospora nodorum	1	N
Peronospora effusa	2	B
Phytophthora cactorum, effusa, fragariae, infestans, kernoviae, megakarya, nicotianae, palmivora, parasitica, ramorum, rubi and sojae	38	H
Plasmodiophora brassicae	2	B
Plasmopara halstedii	1	B
Pleurotus ostreatu	1	S
Pseudocercospora fijiensis	1	H
Puccinia coronata, graminis, sorghi, striiformis and triticina	10	B
Pyrenophora teres f teres and tritici-repentis	18	N
Pyricularia oryzae	4	H
Pythium aphanidermatum and brassicum	2	N
Ramularia collo-cygni	1	H
Rhizoctonia solani	7	N
Rhizopus delemar	1	S
Saccharomyces cerevisiae	60	S
Schizosaccharomyces pombe	1	S
Sclerotinia borealis and sclerotiorum	3	N
Serpula lacrymans	2	S
Setosphaeria turcica	1	H
Sphaerobolus stellatus	1	S
Sporisorium reilianum	2	B
Stereum hirsutum	1	S
Synchytrium endobioticum	2	B
Taphrina deformans	1	B
Thraustotheca clavata	1	S
Tilletia indica	3	H
Tilletiaria anomala	1	B
Trametes versicolor	1	S
Tremella mesenterica	2	B
Trichoderma harzianum, reesei and virens	7	S
Uncinocarpus reesii	1	S
Ustilaginoidea virens	2	B
Ustilago bromivora, hordei and maydis	3	B
Venturia inaequalis	4	H
Verticillium dahliae	10	H
Yarrowia lipolytica	13	S
Zymoseptoria brevis and tritici	6	H

S: saprotroph, N: necrotroph, H: hemibiotroph, B: biotroph

Table S3. Significant GO terms with a depth higher than 7 found enriched in the positively selected proteins in plant fungal pathogens.

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:0009064	glutamine family amino acid metabolic process	140/13729	458/237259	8	57.33
GO:0006165	nucleoside diphosphate phosphorylation	99/13729	320/237259	8	40.37
GO:0006096	glycolytic process	80/13729	266/237259	12	31.2
GO:0006399	tRNA metabolic process	239/13729	1881/237259	8	25.41
GO:1901607	alpha-amino acid biosynthetic process	138/13729	830/237259	8	24.68
GO:0006525	arginine metabolic process	54/13729	157/237259	9	23.51
GO:0006546	glycine catabolic process	40/13729	86/237259	10	22.7
GO:0001510	RNA methylation	56/13729	211/237259	8	18.2
GO:0006750	glutathione biosynthetic process	29/13729	56/237259	8	17.46
GO:0034470	ncRNA processing	186/13729	1549/237259	8	16.55
GO:0008033	tRNA processing	130/13729	991/237259	9	13.87
GO:1901606	alpha-amino acid catabolic process	62/13729	359/237259	8	10.58
GO:0006418	tRNA aminoacylation for protein translation	109/13729	880/237259	10	9.59
GO:0009435	NAD biosynthetic process	34/13729	145/237259	11	8.46
GO:0016579	protein deubiquitination	59/13729	393/237259	9	7.36
GO:0009150	purine ribonucleotide metabolic process	119/13729	1092/237259	9	6.92
GO:0015693	magnesium ion transport	28/13729	123/237259	8	6.23
GO:0006633	fatty acid biosynthetic process	35/13729	219/237259	8	4.08
GO:0015031	protein transport	179/13729	2051/237259	8	3.95
GO:0006355	regulation of transcription, DNA-templated	337/13729	4376/237259	9	3.6
GO:0009165	nucleotide biosynthetic process	123/13729	1368/237259	8	2.48
GO:0006605	protein targeting	38/13729	288/237259	10	2.38
GO:0001522	pseudouridine synthesis	35/13729	260/237259	8	2.28
GO:0006364	rRNA processing	57/13729	515/237259	9	2.22
GO:0006511	ubiquitin-dependent protein catabolic process	86/13729	934/237259	8	1.37

Table S4. Significantly enriched terms relating to biological processes in the positively selected obligate biotroph proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	12/2535	54/75580	7	3.44
GO:0043648	dicarboxylic acid metabolic process	16/2535	77/75580	6	3.3
GO:0051253	negative regulation of RNA metabolic process	11/2535	47/75580	7	3.18
GO:0031324	negative regulation of cellular metabolic process	14/2535	70/75580	5	2.98
GO:0008033	tRNA processing	33/2535	289/75580	9	2.89
GO:0016053	organic acid biosynthetic process	38/2535	453/75580	4	2.69
GO:0018193	peptidyl-amino acid modification	29/2535	315/75580	7	2.63
GO:0010605	negative regulation of macromolecule metabolic process	21/2535	172/75580	5	2.52
GO:0006399	tRNA metabolic process	49/2535	597/75580	8	2.49
GO:0009064	glutamine family amino acid metabolic process	16/2535	118/75580	8	2.35
GO:0007018	microtubule-based movement	35/2535	407/75580	3	2.35
GO:0006082	organic acid metabolic process	79/2535	1249/75580	3	2.34
GO:0006396	RNA processing	80/2535	1217/75580	7	2.3
GO:0006468	protein phosphorylation	97/2535	1582/75580	7	2.26
GO:0034637	cellular carbohydrate biosynthetic process	16/2535	120/75580	5	2.26
GO:0016310	phosphorylation	110/2535	1819/75580	5	2.17
GO:0043412	macromolecule modification	206/2535	3434/75580	4	2.13
GO:0019752	carboxylic acid metabolic process	72/2535	1134/75580	5	2.13
GO:0060255	regulation of macromolecule metabolic process	67/2535	1054/75580	4	2.13
GO:0007017	microtubule-based process	41/2535	556/75580	2	2.1
GO:0006725	cellular aromatic compound metabolic process	212/2535	4200/75580	3	2.08
GO:1901360	organic cyclic compound metabolic process	220/2535	4285/75580	3	2.08
GO:0006464	cellular protein modification process	178/2535	3032/75580	6	2.08
GO:0009058	biosynthetic process	168/2535	3331/75580	2	2.07
GO:0044267	cellular protein metabolic process	208/2535	3721/75580	5	2.07
GO:0046483	heterocycle metabolic process	215/2535	4214/75580	3	2.07
GO:0034470	ncRNA processing	41/2535	560/75580	8	2.05
GO:0034660	ncRNA metabolic process	57/2535	880/75580	7	2.05
GO:0090304	nucleic acid metabolic process	168/2535	3199/75580	5	2.05
GO:0019538	protein metabolic process	245/2535	4869/75580	4	2.04
GO:0050789	regulation of biological process	111/2535	2041/75580	2	2.04
GO:0044249	cellular biosynthetic process	151/2535	3012/75580	3	2.01
GO:0006139	nucleobase-containing compound metabolic process	192/2535	3938/75580	4	2.0
GO:0016070	RNA metabolic process	122/2535	2310/75580	6	1.97
GO:0044260	cellular macromolecule metabolic process	302/2535	5672/75580	4	1.95
GO:1901566	organonitrogen compound biosynthetic process	91/2535	1648/75580	4	1.94
GO:0034641	cellular nitrogen compound metabolic process	245/2535	4998/75580	3	1.93
GO:1901564	organonitrogen compound metabolic process	340/2535	6708/75580	3	1.92
GO:0043170	macromolecule metabolic process	432/2535	8280/75580	3	1.89
GO:0006807	nitrogen compound metabolic process	500/2535	9762/75580	2	1.87
GO:0044237	cellular metabolic process	541/2535	10414/75580	2	1.86
GO:0071704	organic substance metabolic process	610/2535	11484/75580	2	1.84
GO:0044238	primary metabolic process	554/2535	10641/75580	2	1.84
GO:0008152	metabolic process	643/2535	12234/75580	1	1.83
GO:0046394	carboxylic acid biosynthetic process	31/2535	376/75580	6	1.83
GO:0031323	regulation of cellular metabolic process	58/2535	932/75580	4	1.77
GO:0045892	negative regulation of transcription, DNA-templated	9/2535	43/75580	10	1.69
GO:2000112	regulation of cellular macromolecule biosynthetic process	52/2535	816/75580	6	1.59
GO:0010468	regulation of gene expression	58/2535	936/75580	5	1.58
GO:0065007	biological regulation	113/2535	2204/75580	1	1.57
GO:1901576	organic substance biosynthetic process	150/2535	3124/75580	3	1.54
GO:0005975	carbohydrate metabolic process	49/2535	755/75580	3	1.54
GO:0050794	regulation of cellular process	99/2535	1887/75580	3	1.48
GO:0009086	methionine biosynthetic process	7/2535	26/75580	10	1.47
GO:0006520	cellular amino acid metabolic process	51/2535	803/75580	6	1.47
GO:0080090	regulation of primary metabolic process	57/2535	927/75580	4	1.44
GO:0044283	small molecule biosynthetic process	45/2535	674/75580	3	1.42
GO:0008652	cellular amino acid biosynthetic process	25/2535	291/75580	7	1.41
GO:1901605	alpha-amino acid metabolic process	29/2535	364/75580	7	1.35

Table S5. Significantly enriched terms relating to biological processes in the positively selected hemibiotroph proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:0009086	methionine biosynthetic process	13/6255	64/222540	10	3.38
GO:0051274	beta-glucan biosynthetic process	18/6255	104/222540	8	3.31
GO:0051253	negative regulation of RNA metabolic process	16/6255	91/222540	7	3.26
GO:0043648	dicarboxylic acid metabolic process	20/6255	150/222540	6	3.25
GO:0009082	branched-chain amino acid biosynthetic process	18/6255	81/222540	5	3.06
GO:0010605	negative regulation of macromolecule metabolic process	34/6255	300/222540	5	2.96
GO:0034637	cellular carbohydrate biosynthetic process	29/6255	226/222540	5	2.94
GO:0031324	negative regulation of cellular metabolic process	20/6255	129/222540	5	2.79
GO:0009312	oligosaccharide biosynthetic process	19/6255	146/222540	5	2.74
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	15/6255	101/222540	7	2.7
GO:0044262	cellular carbohydrate metabolic process	35/6255	455/222540	4	2.67
GO:0016051	carbohydrate biosynthetic process	37/6255	333/222540	4	2.61
GO:0000097	sulfur amino acid biosynthetic process	15/6255	121/222540	5	2.46
GO:0007017	microtubule-based process	72/6255	1355/222540	2	2.39
GO:0006355	regulation of transcription, DNA-templated	75/6255	1176/222540	9	2.39
GO:2000112	regulation of cellular macromolecule biosynthetic process	84/6255	1307/222540	6	2.39
GO:0051273	beta-glucan metabolic process	19/6255	190/222540	7	2.37
GO:0006396	RNA processing	95/6255	1887/222540	7	2.33
GO:0016071	mRNA metabolic process	52/6255	831/222540	7	2.32
GO:0051171	regulation of nitrogen compound metabolic process	95/6255	1522/222540	4	2.31
GO:0006974	cellular response to DNA damage stimulus	68/6255	1193/222540	4	2.29
GO:0010468	regulation of gene expression	104/6255	1526/222540	5	2.28
GO:0006950	response to stress	88/6255	1528/222540	2	2.27
GO:0006397	mRNA processing	42/6255	664/222540	8	2.27
GO:0019222	regulation of metabolic process	118/6255	1794/222540	3	2.27
GO:0051252	regulation of RNA metabolic process	81/6255	1206/222540	6	2.26
GO:0060255	regulation of macromolecule metabolic process	117/6255	1749/222540	4	2.25
GO:0033554	cellular response to stress	72/6255	1253/222540	3	2.24
GO:0051716	cellular response to stimulus	73/6255	1254/222540	2	2.24
GO:0050896	response to stimulus	96/6255	1605/222540	1	2.22
GO:0031323	regulation of cellular metabolic process	100/6255	1555/222540	4	2.21
GO:0005975	carbohydrate metabolic process	156/6255	2565/222540	3	2.19
GO:0050789	regulation of biological process	172/6255	3715/222540	2	2.12
GO:0048519	negative regulation of biological process	35/6255	527/222540	3	2.11
GO:0050794	regulation of cellular process	154/6255	3447/222540	3	2.05
GO:0044249	cellular biosynthetic process	208/6255	5071/222540	3	2.01
GO:0065007	biological regulation	178/6255	4153/222540	1	2.01
GO:0016070	RNA metabolic process	148/6255	3483/222540	6	2.01
GO:0006468	protein phosphorylation	215/6255	4156/222540	7	2.0
GO:0016310	phosphorylation	235/6255	4699/222540	5	2.0
GO:0009058	biosynthetic process	230/6255	5732/222540	2	1.96
GO:0006793	phosphorus metabolic process	292/6255	6981/222540	3	1.94
GO:0043412	macromolecule modification	366/6255	7831/222540	4	1.94
GO:0006464	cellular protein modification process	327/6255	7218/222540	6	1.94
GO:0044267	cellular protein metabolic process	372/6255	8250/222540	5	1.92
GO:0006796	phosphate-containing compound metabolic process	291/6255	6936/222540	4	1.9
GO:0019538	protein metabolic process	459/6255	11568/222540	4	1.84
GO:0034645	cellular macromolecule biosynthetic process	92/6255	1977/222540	5	1.82
GO:1901576	organic substance biosynthetic process	209/6255	5300/222540	3	1.82
GO:0044260	cellular macromolecule metabolic process	524/6255	13650/222540	4	1.81
GO:1901564	organonitrogen compound metabolic process	558/6255	15181/222540	3	1.79
GO:0043170	macromolecule metabolic process	713/6255	19213/222540	3	1.74
GO:0044237	cellular metabolic process	841/6255	22391/222540	2	1.71
GO:0006807	nitrogen compound metabolic process	788/6255	22412/222540	2	1.7
GO:0044238	primary metabolic process	941/6255	25043/222540	2	1.69
GO:0071704	organic substance metabolic process	1007/6255	26763/222540	2	1.67
GO:0008152	metabolic process	1069/6255	28228/222540	1	1.67
GO:0055085	transmembrane transport	202/6255	5180/222540	4	1.66
GO:0006813	potassium ion transport	25/6255	336/222540	6	1.6
GO:0045892	negative regulation of transcription, DNA-templated	11/6255	82/222540	10	1.43
GO:0009059	macromolecule biosynthetic process	99/6255	2252/222540	4	1.39

Table S6. Enriched terms relating to biological processes in the positively selected plant necrotrophs.

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:0006190	inosine salvage	8/9880	8/129511	11	3.63
GO:0009088	threonine biosynthetic process	8/9880	8/129511	10	3.63
GO:0006425	glutaminyl-tRNA aminoacylation	8/9880	8/129511	11	3.63
GO:1901031	regulation of response to reactive oxygen species	9/9880	11/129511	6	3.5
GO:0046168	glycerol-3-phosphate catabolic process	13/9880	18/129511	7	3.47
GO:0072350	tricarboxylic acid metabolic process	17/9880	32/129511	6	3.28
GO:0006537	glutamate biosynthetic process	14/9880	15/129511	10	3.23
GO:0006166	purine ribonucleoside salvage	14/9880	23/129511	10	3.22
GO:0009084	glutamine family amino acid biosynthetic process	27/9880	86/129511	9	3.14
GO:0005992	trehalose biosynthetic process	26/9880	73/129511	7	3.13
GO:0009082	branched-chain amino acid biosynthetic process	31/9880	67/129511	5	3.1
GO:0060271	cilium assembly	30/9880	122/129511	7	3.06
GO:0009064	glutamine family amino acid metabolic process	47/9880	153/129511	8	2.99
GO:0000096	sulfur amino acid metabolic process	28/9880	123/129511	4	2.98
GO:0006144	purine nucleobase metabolic process	18/9880	50/129511	7	2.97
GO:0006555	methionine metabolic process	19/9880	56/129511	9	2.96
GO:0006536	glutamate metabolic process	22/9880	39/129511	9	2.96
GO:0033014	tetrapyrrole biosynthetic process	29/9880	109/129511	5	2.95
GO:0051274	beta-glucan biosynthetic process	29/9880	75/129511	8	2.93
GO:0016573	histone acetylation	22/9880	75/129511	11	2.93
GO:0003341	cilium movement	25/9880	75/129511	4	2.93
GO:0001522	pseudouridine synthesis	45/9880	154/129511	8	2.93
GO:0071897	DNA biosynthetic process	18/9880	34/129511	7	2.92
GO:0006816	calcium ion transport	27/9880	124/129511	7	2.91
GO:0006102	isocitrate metabolic process	9/9880	16/129511	7	2.91
GO:0000097	sulfur amino acid biosynthetic process	21/9880	82/129511	5	2.89
GO:0032012	regulation of ARF protein signal transduction	16/9880	45/129511	9	2.88
GO:0030488	tRNA methylation	17/9880	51/129511	11	2.85
GO:0007154	cell communication	28/9880	76/129511	2	2.83
GO:0018205	peptidyl-lysine modification	32/9880	148/129511	8	2.82
GO:0017038	protein import	33/9880	133/129511	9	2.8
GO:0006606	protein import into nucleus	25/9880	83/129511	10	2.8
GO:0051056	regulation of small GTPase mediated signal transduction	28/9880	70/129511	7	2.78
GO:0006414	translational elongation	28/9880	90/129511	6	2.78
GO:0006075	(1->3)-beta-D-glucan biosynthetic process	22/9880	64/129511	9	2.75
GO:0009086	methionine biosynthetic process	18/9880	46/129511	10	2.74
GO:0034637	cellular carbohydrate biosynthetic process	56/9880	165/129511	5	2.73
GO:0006525	arginine metabolic process	19/9880	58/129511	9	2.72
GO:0016051	carbohydrate biosynthetic process	61/9880	238/129511	4	2.72
GO:0001510	RNA methylation	43/9880	205/129511	8	2.72
GO:0006400	tRNA modification	55/9880	228/129511	10	2.71
GO:0009250	glucan biosynthetic process	30/9880	84/129511	7	2.71
GO:0051273	beta-glucan metabolic process	30/9880	150/129511	7	2.69
GO:0006096	glycolytic process	41/9880	150/129511	12	2.69
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	20/9880	71/129511	7	2.69
GO:0070925	organelle assembly	41/9880	197/129511	5	2.68
GO:0043648	dicarboxylic acid metabolic process	34/9880	120/129511	6	2.67
GO:0043547	positive regulation of GTPase activity	15/9880	21/129511	6	2.67
GO:0010605	negative regulation of macromolecule metabolic process	52/9880	240/129511	5	2.64
GO:0009312	oligosaccharide biosynthetic process	31/9880	106/129511	5	2.63
GO:0006073	cellular glucan metabolic process	31/9880	159/129511	6	2.62
GO:0016570	histone modification	39/9880	199/129511	7	2.6
GO:1901615	organic hydroxy compound metabolic process	35/9880	190/129511	3	2.59
GO:0006091	generation of precursor metabolites and energy	57/9880	294/129511	3	2.55
GO:0006401	RNA catabolic process	38/9880	181/129511	7	2.53
GO:0072330	monocarboxylic acid biosynthetic process	28/9880	138/129511	7	2.48
GO:1901607	alpha-amino acid biosynthetic process	57/9880	356/129511	8	2.47
GO:0006418	tRNA aminoacylation for protein translation	56/9880	328/129511	10	2.46
GO:0048583	regulation of response to stimulus	51/9880	300/129511	3	2.46
GO:0016052	carbohydrate catabolic process	51/9880	261/129511	4	2.45
GO:0032259	methylation	72/9880	386/129511	2	2.45
GO:0008033	tRNA processing	96/9880	403/129511	9	2.45
GO:0034472	snRNA 3′-end processing	10/9880	22/129511	10	2.45
GO:0072594	establishment of protein localization to organelle	43/9880	251/129511	6	2.44
GO:0006457	protein folding	60/9880	404/129511	2	2.43
GO:0044262	cellular carbohydrate metabolic process	66/9880	330/129511	4	2.43
GO:0046034	ATP metabolic process	43/9880	227/129511	2	2.43
GO:0072521	purine-containing compound metabolic process	74/9880	488/129511	4	2.42
GO:0031324	negative regulation of cellular metabolic process	22/9880	95/129511	5	2.37
GO:0043414	macromolecule methylation	52/9880	320/129511	5	2.34
GO:0008652	cellular amino acid biosynthetic process	79/9880	427/129511	7	2.33
GO:0032787	monocarboxylic acid metabolic process	81/9880	463/129511	6	2.32
GO:0034470	ncRNA processing	140/9880	735/129511	8	2.3
GO:0006813	potassium ion transport	49/9880	309/129511	6	2.29
GO:0018193	peptidyl-amino acid modification	85/9880	518/129511	7	2.28
GO:0046394	carboxylic acid biosynthetic process	107/9880	573/129511	6	2.28
GO:1901605	alpha-amino acid metabolic process	87/9880	557/129511	7	2.26
GO:0009451	RNA modification	120/9880	540/129511	7	2.26
GO:0016053	organic acid biosynthetic process	144/9880	682/129511	4	2.23
GO:0006310	DNA recombination	40/9880	233/129511	7	2.22
GO:0034660	ncRNA metabolic process	201/9880	1090/129511	7	2.22
GO:0034613	cellular protein localization	44/9880	274/129511	4	2.22
GO:0006399	tRNA metabolic process	157/9880	748/129511	8	2.21
GO:0007018	microtubule-based movement	116/9880	839/129511	3	2.21
GO:0006814	sodium ion transport	17/9880	63/129511	6	2.19
GO:0006520	cellular amino acid metabolic process	167/9880	1097/129511	6	2.19
GO:0006355	regulation of transcription, DNA-templated	145/9880	1045/129511	9	2.18
GO:0044283	small molecule biosynthetic process	179/9880	1074/129511	3	2.18
GO:0006812	cation transport	162/9880	1214/129511	5	2.17
GO:0007017	microtubule-based process	141/9880	1103/129511	2	2.16
GO:0080090	regulation of primary metabolic process	174/9880	1307/129511	4	2.15
GO:0010468	regulation of gene expression	183/9880	1307/129511	5	2.15
GO:1901575	organic substance catabolic process	189/9880	1657/129511	3	2.15
GO:0051252	regulation of RNA metabolic process	147/9880	1056/129511	6	2.14
GO:0006629	lipid metabolic process	197/9880	1627/129511	3	2.14
GO:0000398	mRNA splicing, via spliceosome	50/9880	321/129511	11	2.12
GO:0005975	carbohydrate metabolic process	200/9880	1738/129511	3	2.11
GO:0006066	alcohol metabolic process	25/9880	121/129511	4	2.1
GO:0006082	organic acid metabolic process	321/9880	1814/129511	3	2.1
GO:0006396	RNA processing	263/9880	1641/129511	7	2.1
GO:0031323	regulation of cellular metabolic process	177/9880	1325/129511	4	2.09

Table S7. Enriched terms relating to biological processes in the positively selected plant necrotrophs (continued).

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:2000112	regulation of cellular macromolecule biosynthetic process	157/9880	1151/129511	6	2.08
GO:0005976	polysaccharide metabolic process	32/9880	175/129511	4	2.08
GO:0019222	regulation of metabolic process	215/9880	1513/129511	3	2.08
GO:0060255	regulation of macromolecule metabolic process	210/9880	1484/129511	4	2.07
GO:0031047	gene silencing by RNA	8/9880	15/129511	8	2.07
GO:0034645	cellular macromolecule biosynthetic process	217/9880	1621/129511	5	2.06
GO:0043604	amide biosynthetic process	114/9880	946/129511	5	2.04
GO:0016043	cellular component organization	226/9880	2030/129511	3	2.02
GO:0016070	RNA metabolic process	453/9880	2964/129511	6	2.01
GO:0009056	catabolic process	208/9880	1779/129511	2	2.0
GO:0050896	response to stimulus	141/9880	1228/129511	1	2.0
GO:0043087	regulation of GTPase activity	17/9880	65/129511	5	2.0
GO:0071840	cellular component organization or biogenesis	233/9880	2151/129511	2	2.0
GO:0009059	macromolecule biosynthetic process	231/9880	1856/129511	4	1.99
GO:0022607	cellular component assembly	110/9880	906/129511	4	1.99
GO:0019752	carboxylic acid metabolic process	270/9880	1645/129511	5	1.99
GO:1901566	organonitrogen compound biosynthetic process	319/9880	2320/129511	4	1.99
GO:0050789	regulation of biological process	385/9880	3249/129511	2	1.94
GO:0009112	nucleobase metabolic process	24/9880	116/129511	6	1.93
GO:0044281	small molecule metabolic process	405/9880	3119/129511	2	1.92
GO:0006885	regulation of pH	15/9880	53/129511	8	1.91
GO:0065007	biological regulation	409/9880	3527/129511	1	1.91
GO:0050794	regulation of cellular process	337/9880	3032/129511	3	1.91
GO:0046148	pigment biosynthetic process	17/9880	66/129511	3	1.91
GO:0050790	regulation of catalytic activity	38/9880	227/129511	3	1.88
GO:0006811	ion transport	328/9880	3267/129511	4	1.85
GO:0044249	cellular biosynthetic process	543/9880	4181/129511	3	1.85
GO:0090304	nucleic acid metabolic process	569/9880	5033/129511	5	1.84
GO:1901576	organic substance biosynthetic process	564/9880	4428/129511	3	1.83
GO:0006464	cellular protein modification process	551/9880	5511/129511	6	1.82
GO:0009058	biosynthetic process	613/9880	4761/129511	2	1.81
GO:0006793	phosphorus metabolic process	512/9880	5401/129511	3	1.8
GO:0006419	alanyl-tRNA aminoacylation	8/9880	16/129511	11	1.8
GO:0006101	citrate metabolic process	8/9880	16/129511	7	1.8
GO:0008612	peptidyl-lysine modification to peptidyl-hypusine	8/9880	16/129511	9	1.8
GO:0006423	cysteinyl-tRNA aminoacylation	8/9880	16/129511	11	1.8
GO:0044267	cellular protein metabolic process	655/9880	6393/129511	5	1.79
GO:0044271	cellular nitrogen compound biosynthetic process	265/9880	2604/129511	4	1.79
GO:0006796	phosphate-containing compound metabolic process	506/9880	5361/129511	4	1.79
GO:0043412	macromolecule modification	672/9880	6051/129511	4	1.78
GO:0006139	nucleobase-containing compound metabolic process	688/9880	6267/129511	4	1.78
GO:0006777	Mo-molybdopterin cofactor biosynthetic process	13/9880	42/129511	7	1.77
GO:1901360	organic cyclic compound metabolic process	786/9880	6882/129511	3	1.77
GO:0006725	cellular aromatic compound metabolic process	753/9880	6732/129511	3	1.77
GO:0034641	cellular nitrogen compound metabolic process	877/9880	7987/129511	3	1.74
GO:0046483	heterocycle metabolic process	759/9880	6710/129511	3	1.74
GO:0010629	negative regulation of gene expression	30/9880	166/129511	6	1.71
GO:0016071	mRNA metabolic process	84/9880	674/129511	7	1.69
GO:0044260	cellular macromolecule metabolic process	972/9880	9893/129511	4	1.69
GO:1901564	organonitrogen compound metabolic process	1213/9880	12929/129511	3	1.63
GO:0043170	macromolecule metabolic process	1472/9880	15069/129511	3	1.62
GO:0016310	phosphorylation	341/9880	3524/129511	5	1.6
GO:0006807	nitrogen compound metabolic process	1772/9880	17828/129511	2	1.59
GO:0044237	cellular metabolic process	1906/9880	17064/129511	2	1.57
GO:0009098	leucine biosynthetic process	8/9880	17/129511	8	1.57
GO:0006435	threonyl-tRNA aminoacylation	8/9880	17/129511	11	1.57
GO:0044238	primary metabolic process	1962/9880	19530/129511	2	1.57
GO:0008152	metabolic process	2337/9880	22313/129511	1	1.56
GO:0071704	organic substance metabolic process	2192/9880	21160/129511	2	1.54
GO:0048519	negative regulation of biological process	57/9880	414/129511	3	1.53
GO:0033554	cellular response to stress	108/9880	929/129511	3	1.51
GO:0006325	chromatin organization	47/9880	323/129511	4	1.42
GO:1901136	carbohydrate derivative catabolic process	18/9880	79/129511	4	1.39
GO:0006950	response to stress	129/9880	1162/129511	2	1.38
GO:0042364	water-soluble vitamin biosynthetic process	28/9880	156/129511	5	1.37
GO:0019243	methylglyoxal catabolic process to D-lactate via S-lactoyl-glutathione	8/9880	18/129511	9	1.35
GO:0019310	inositol catabolic process	8/9880	18/129511	7	1.35
GO:0018344	protein geranylgeranylation	8/9880	18/129511	8	1.35
GO:0006566	threonine metabolic process	11/9880	34/129511	9	1.3

Table S8. enriched terms relating to biological processes in the positively selected animal necrotrophs.

go number	name	ratio in study	ratio in population	depth	−log10 of *p* value
GO:0006190	inosine salvage	8/7214	8/114793	11	4.22
GO:0032955	regulation of division septum assembly	6/7214	6/114793	8	3.9
GO:0036159	inner dynein arm assembly	6/7214	6/114793	8	3.9
GO:0046168	glycerol-3-phosphate catabolic process	12/7214	20/114793	7	3.7
GO:0015940	pantothenate biosynthetic process	15/7214	20/114793	8	3.7
GO:0072350	tricarboxylic acid metabolic process	13/7214	26/114793	6	3.41
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	17/7214	50/114793	7	3.29
GO:0002098	tRNA wobble uridine modification	18/7214	50/114793	12	3.29
GO:0030488	tRNA methylation	15/7214	44/114793	11	3.24
GO:0051253	negative regulation of RNA metabolic process	17/7214	44/114793	7	3.24
GO:0003341	cilium movement	24/7214	57/114793	4	3.24
GO:0006536	glutamate metabolic process	15/7214	38/114793	9	3.23
GO:0009086	methionine biosynthetic process	14/7214	33/114793	10	3.11
GO:0006414	translational elongation	21/7214	80/114793	6	3.09
GO:0071897	DNA biosynthetic process	14/7214	39/114793	7	3.09
GO:1901031	regulation of response to reactive oxygen species	6/7214	7/114793	6	3.08
GO:0006425	glutaminyl-tRNA aminoacylation	6/7214	7/114793	11	3.08
GO:0051103	DNA ligation involved in DNA repair	6/7214	7/114793	8	3.08
GO:0032958	inositol phosphate biosynthetic process	13/7214	18/114793	7	3.05
GO:0001522	pseudouridine synthesis	29/7214	138/114793	8	3.03
GO:0006606	protein import into nucleus	24/7214	81/114793	10	3.0
GO:1901617	organic hydroxy compound biosynthetic process	19/7214	81/114793	4	3.0
GO:0017186	peptidyl-pyroglutamic acid biosynthetic process, using glutaminyl-peptide cyclotransferase	11/7214	14/114793	9	2.98
GO:0043648	dicarboxylic acid metabolic process	27/7214	105/114793	6	2.98
GO:0045892	negative regulation of transcription, DNA-templated	16/7214	40/114793	10	2.94
GO:0006525	arginine metabolic process	15/7214	53/114793	9	2.93
GO:0051056	regulation of small GTPase mediated signal transduction	27/7214	82/114793	7	2.92
GO:0009082	branched-chain amino acid biosynthetic process	24/7214	60/114793	5	2.91
GO:0016573	histone acetylation	21/7214	83/114793	11	2.83
GO:0019751	polyol metabolic process	19/7214	83/114793	5	2.83
GO:0009084	glutamine family amino acid biosynthetic process	22/7214	83/114793	9	2.83
GO:0007154	cell communication	18/7214	54/114793	2	2.82
GO:0006537	glutamate biosynthetic process	11/7214	19/114793	10	2.81
GO:0006166	purine ribonucleoside salvage	13/7214	19/114793	10	2.81
GO:0034637	cellular carbohydrate biosynthetic process	25/7214	133/114793	5	2.8
GO:0042398	cellular modified amino acid biosynthetic process	23/7214	116/114793	5	2.79
GO:0006066	alcohol metabolic process	26/7214	116/114793	4	2.79
GO:0017038	protein import	28/7214	116/114793	9	2.79
GO:0009064	glutamine family amino acid metabolic process	40/7214	142/114793	8	2.79
GO:0033014	tetrapyrrole biosynthetic process	21/7214	76/114793	5	2.78
GO:0001510	RNA methylation	40/7214	188/114793	8	2.78
GO:0009966	regulation of signal transduction	43/7214	245/114793	5	2.77
GO:0046034	ATP metabolic process	32/7214	198/114793	2	2.75
GO:0006096	glycolytic process	26/7214	134/114793	12	2.74
GO:0000398	mRNA splicing, via spliceosome	50/7214	286/114793	11	2.73
GO:0051274	beta-glucan biosynthetic process	17/7214	62/114793	8	2.71
GO:0016052	carbohydrate catabolic process	37/7214	236/114793	4	2.69
GO:0072330	monocarboxylic acid biosynthetic process	31/7214	101/114793	7	2.68
GO:0031324	negative regulation of cellular metabolic process	18/7214	70/114793	5	2.65
GO:0000097	sulfur amino acid biosynthetic process	20/7214	70/114793	5	2.65
GO:0032012	regulation of ARF protein signal transduction	16/7214	56/114793	9	2.6
GO:0060271	cilium assembly	27/7214	119/114793	7	2.59
GO:0008380	RNA splicing	53/7214	321/114793	8	2.59
GO:1901615	organic hydroxy compound metabolic process	37/7214	192/114793	3	2.58
GO:0046394	carboxylic acid biosynthetic process	96/7214	495/114793	6	2.53
GO:0006480	N-terminal protein amino acid methylation	6/7214	8/114793	9	2.52
GO:0070925	organelle assembly	35/7214	215/114793	5	2.52
GO:0018205	peptidyl-lysine modification	29/7214	165/114793	8	2.52
GO:0008033	tRNA processing	92/7214	407/114793	9	2.51
GO:0046129	purine ribonucleoside biosynthetic process	14/7214	50/114793	9	2.51
GO:0008652	cellular amino acid biosynthetic process	65/7214	388/114793	7	2.5
GO:0018193	peptidyl-amino acid modification	65/7214	497/114793	7	2.5
GO:0032787	monocarboxylic acid metabolic process	64/7214	357/114793	6	2.47
GO:0006400	tRNA modification	52/7214	220/114793	10	2.46
GO:0010605	negative regulation of macromolecule metabolic process	42/7214	220/114793	5	2.46
GO:0032259	methylation	64/7214	377/114793	2	2.46
GO:0016053	organic acid biosynthetic process	126/7214	602/114793	4	2.45
GO:0016071	mRNA metabolic process	80/7214	646/114793	7	2.45
GO:1901607	alpha-amino acid biosynthetic process	48/7214	329/114793	8	2.42
GO:0048583	regulation of response to stimulus	49/7214	258/114793	3	2.42
GO:1901605	alpha-amino acid metabolic process	71/7214	546/114793	7	2.4
GO:0042364	water-soluble vitamin biosynthetic process	24/7214	131/114793	5	2.39
GO:0043414	macromolecule methylation	48/7214	304/114793	5	2.38
GO:0009451	RNA modification	107/7214	512/114793	7	2.35
GO:0006165	nucleoside diphosphate phosphorylation	27/7214	159/114793	8	2.35
GO:0006397	mRNA processing	67/7214	528/114793	8	2.35
GO:0006399	tRNA metabolic process	132/7214	725/114793	8	2.35
GO:0034660	ncRNA metabolic process	167/7214	1014/114793	7	2.33
GO:0043604	amide biosynthetic process	110/7214	802/114793	5	2.32
GO:0007017	microtubule-based process	116/7214	1016/114793	2	2.32
GO:0006518	peptide metabolic process	111/7214	990/114793	5	2.31
GO:0034470	ncRNA processing	127/7214	690/114793	8	2.28
GO:0006412	translation	91/7214	714/114793	7	2.28
GO:0044283	small molecule biosynthetic process	158/7214	943/114793	3	2.27
GO:0072594	establishment of protein localization to organelle	35/7214	227/114793	6	2.27
GO:0006355	regulation of transcription, DNA-templated	112/7214	1004/114793	9	2.24
GO:0031323	regulation of cellular metabolic process	130/7214	1282/114793	4	2.21
GO:0010468	regulation of gene expression	143/7214	1284/114793	5	2.2
GO:0007018	microtubule-based movement	93/7214	774/114793	3	2.2
GO:1901575	organic substance catabolic process	147/7214	1452/114793	3	2.19
GO:0043603	cellular amide metabolic process	135/7214	1100/114793	4	2.19
GO:0034645	cellular macromolecule biosynthetic process	188/7214	1454/114793	5	2.19
GO:0006520	cellular amino acid metabolic process	132/7214	1045/114793	6	2.18
GO:0044262	cellular carbohydrate metabolic process	37/7214	258/114793	4	2.17
GO:0051171	regulation of nitrogen compound metabolic process	134/7214	1260/114793	4	2.17
GO:0006082	organic acid metabolic process	258/7214	1638/114793	3	2.17
GO:2000112	regulation of cellular macromolecule biosynthetic process	118/7214	1107/114793	6	2.16
GO:0051252	regulation of RNA metabolic process	113/7214	1021/114793	6	2.16
GO:0006396	RNA processing	240/7214	1505/114793	7	2.13
GO:0006091	generation of precursor metabolites and energy	38/7214	272/114793	3	2.12
GO:0009056	catabolic process	154/7214	1580/114793	2	2.12

Table S9. Enriched terms relating to biological processes in the positively selected animal necrotrophs (continued).

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 of *p* value
GO:0006629	lipid metabolic process	143/7214	1475/114793	3	2.12
GO:0060255	regulation of macromolecule metabolic process	160/7214	1442/114793	4	2.11
GO:0019752	carboxylic acid metabolic process	216/7214	1483/114793	5	2.09
GO:0034613	cellular protein localization	36/7214	248/114793	4	2.08
GO:1901566	organonitrogen compound biosynthetic process	277/7214	2065/114793	4	2.08
GO:0009059	macromolecule biosynthetic process	189/7214	1710/114793	4	2.07
GO:0006102	isocitrate metabolic process	7/7214	13/114793	7	2.05
GO:0050789	regulation of biological process	306/7214	3078/114793	2	2.03
GO:0016043	cellular component organization	184/7214	2008/114793	3	2.02
GO:0044281	small molecule metabolic process	317/7214	2782/114793	2	2.01
GO:0006401	RNA catabolic process	29/7214	185/114793	7	2.01
GO:0044271	cellular nitrogen compound biosynthetic process	215/7214	2275/114793	4	2.0
GO:0016070	RNA metabolic process	372/7214	2742/114793	6	2.0
GO:0050794	regulation of cellular process	270/7214	2861/114793	3	1.97
GO:0071840	cellular component organization or biogenesis	189/7214	2119/114793	2	1.96
GO:0090304	nucleic acid metabolic process	466/7214	3855/114793	5	1.96
GO:0044249	cellular biosynthetic process	430/7214	3741/114793	3	1.96
GO:1901576	organic substance biosynthetic process	464/7214	4002/114793	3	1.93
GO:0065007	biological regulation	330/7214	3366/114793	1	1.93
GO:0006139	nucleobase-containing compound metabolic process	548/7214	4921/114793	4	1.92
GO:0009058	biosynthetic process	506/7214	4268/114793	2	1.91
GO:0051273	beta-glucan metabolic process	22/7214	122/114793	7	1.88
GO:0006725	cellular aromatic compound metabolic process	604/7214	5329/114793	3	1.87
GO:1901360	organic cyclic compound metabolic process	632/7214	5489/114793	3	1.86
GO:0043412	macromolecule modification	580/7214	6499/114793	4	1.86
GO:0006555	methionine metabolic process	15/7214	64/114793	9	1.85
GO:0034641	cellular nitrogen compound metabolic process	717/7214	6287/114793	3	1.84
GO:0006464	cellular protein modification process	469/7214	5979/114793	6	1.83
GO:0046483	heterocycle metabolic process	600/7214	5319/114793	3	1.82
GO:0022607	cellular component assembly	90/7214	869/114793	4	1.81
GO:0044267	cellular protein metabolic process	566/7214	6741/114793	5	1.81
GO:0006793	phosphorus metabolic process	450/7214	5743/114793	3	1.78
GO:0019538	protein metabolic process	725/7214	9516/114793	4	1.77
GO:0044260	cellular macromolecule metabolic process	824/7214	9171/114793	4	1.74
GO:0006310	DNA recombination	31/7214	211/114793	7	1.7
GO:0044237	cellular metabolic process	1576/7214	15465/114793	2	1.67
GO:1901564	organonitrogen compound metabolic process	1025/7214	12079/114793	3	1.67
GO:0043170	macromolecule metabolic process	1243/7214	13653/114793	3	1.66
GO:0006541	glutamine metabolic process	10/7214	31/114793	9	1.65
GO:0006418	tRNA aminoacylation for protein translation	40/7214	302/114793	10	1.64
GO:0005975	carbohydrate metabolic process	121/7214	1279/114793	3	1.63
GO:0006796	phosphate-containing compound metabolic process	444/7214	5726/114793	4	1.63
GO:0006807	nitrogen compound metabolic process	1481/7214	15957/114793	2	1.63
GO:0044238	primary metabolic process	1583/7214	17428/114793	2	1.62
GO:0008152	metabolic process	1908/7214	19763/114793	1	1.61
GO:0071704	organic substance metabolic process	1793/7214	18791/114793	2	1.59
GO:0009987	cellular process	2251/7214	23958/114793	1	1.56
GO:0008150	biological_process	2760/7214	30756/114793	0	1.53
GO:0000096	sulfur amino acid metabolic process	22/7214	130/114793	4	1.43
GO:0002943	tRNA dihydrouridine synthesis	6/7214	11/114793	11	1.41
GO:0009072	aromatic amino acid family metabolic process	26/7214	169/114793	4	1.4

Table S10. Significant enriched terms relating to biological processes in the stramenopile dataset's paralogs.

GO number	Name	Ratio in study	Ratio in population	Depth	−log10 *p* value	Species
GO:0055085	transmembrane transport	16/62	557/11080	4	3.26	Achlya hypogyna
GO:0019637	organophosphate metabolic process	8/62	213/11080	4	1.36	Achlya hypogyna
GO:0043412	macromolecule modification	107/760	755/8600	4	2.26	Albugo candida
GO:0071704	organic substance metabolic process	271/760	2317/8600	2	2.12	Albugo candida
GO:0044237	cellular metabolic process	251/760	2142/8600	2	2.09	Albugo candida
GO:0044238	primary metabolic process	254/760	2154/8600	2	2.07	Albugo candida
GO:0006464	cellular protein modification process	95/760	671/8600	6	2.06	Albugo candida
GO:0008152	metabolic process	283/760	2451/8600	1	2.04	Albugo candida
GO:0044260	cellular macromolecule metabolic process	145/760	1177/8600	4	1.58	Albugo candida
GO:0043170	macromolecule metabolic process	196/760	1696/8600	3	1.45	Albugo candida
GO:0006807	nitrogen compound metabolic process	223/760	1970/8600	2	1.44	Albugo candida
GO:0006796	phosphate-containing compound metabolic process	90/760	660/8600	4	1.41	Albugo candida
GO:0044262	cellular carbohydrate metabolic process	12/504	37/8647	4	2.94	Albugo laibachii
GO:0034645	cellular macromolecule biosynthetic process	35/504	228/8647	5	2.47	Albugo laibachii
GO:0051560	mitochondrial calcium ion homeostasis	4/504	4/8647	10	1.7	Albugo laibachii
GO:0042592	homeostatic process	8/504	21/8647	3	1.64	Albugo laibachii
GO:0051274	beta-glucan biosynthetic process	6/504	11/8647	8	1.62	Albugo laibachii
GO:0098771	inorganic ion homeostasis	6/504	12/8647	6	1.34	Albugo laibachii
GO:0034637	cellular carbohydrate biosynthetic process	8/504	23/8647	5	1.31	Albugo laibachii
GO:0006821	chloride transport	6/477	22/17944	7	1.48	Aphanomyces astaci
GO:0045048	protein insertion into ER membrane	4/140	5/14252	8	4.08	Aphanomyces euteiches
GO:0046434	organophosphate catabolic process	4/140	16/14252	5	1.56	Aphanomyces euteiches
GO:0009084	glutamine family amino acid biosynthetic process	8/272	19/6501	9	3.14	Bremia lactucae
GO:0006561	proline biosynthetic process	6/272	10/6501	10	2.81	Bremia lactucae
GO:1901264	carbohydrate derivative transport	6/272	13/6501	7	1.94	Bremia lactucae
GO:0044271	cellular nitrogen compound biosynthetic process	0/272	274/6501	4	1.59	Bremia lactucae
GO:0006259	DNA metabolic process	18/843	803/12755	6	2.4	Globisporangium splendens
GO:0015074	DNA integration	2/843	666/12755	7	2.31	Globisporangium splendens
GO:0044260	cellular macromolecule metabolic process	81/843	1879/12755	4	1.68	Globisporangium splendens
GO:0034220	ion transmembrane transport	12/197	68/7213	5	3.39	Hyaloperonospora arabidopsidis
GO:0098656	anion transmembrane transport	8/197	33/7213	6	2.45	Hyaloperonospora arabidopsidis
GO:0055085	transmembrane transport	20/197	247/7213	4	1.69	Hyaloperonospora arabidopsidis
GO:0043933	protein-containing complex subunit organization	44/3329	134/20260	4	2.26	Nothophytophthora sp
GO:1901564	organonitrogen compound metabolic process	379/3329	2777/20260	3	1.4	Nothophytophthora sp
GO:0006665	sphingolipid metabolic process	4/123	20/13965	6	1.32	Phytophthora cinnamomi
GO:0007186	G protein-coupled receptor signaling pathway	4/201	12/19214	5	1.98	Phytophthora fragariae
GO:0098771	inorganic ion homeostasis	4/201	13/19214	6	1.83	Phytophthora fragariae
GO:0090304	nucleic acid metabolic process	2/201	1600/19214	5	1.81	Phytophthora fragariae
GO:1901360	organic cyclic compound metabolic process	4/201	1865/19214	3	1.4	Phytophthora fragariae
GO:0034219	carbohydrate transmembrane transport	2/41	2/8291	8	1.37	Phytophthora kernoviae
GO:0006643	membrane lipid metabolic process	6/103	37/18043	5	3.97	Phytophthora megakarya
GO:0009247	glycolipid biosynthetic process	4/103	18/18043	7	2.29	Phytophthora megakarya
GO:0006470	protein dephosphorylation	18/1465	55/12653	7	1.3	Phytophthora nicotianae
GO:0070536	protein K63-linked deubiquitination	4/748	5/17216	10	1.48	Phytophthora parasitica
GO:0006069	ethanol oxidation	2/43	2/9298	7	1.41	Pythium aphanidermatum

Appendix B

Submitted manuscripts

B.1 Proteins secreted by an obligate parasitic protist selectively repress phyllosphere-associated bacteria

Contributions by DGP:

- Designed the initial manuscript.
- Wrote the initial draft.
- Helped establish the experimental pipeline together with MS.
- Performed the protein purification.
- Edited the final manuscript.
- Performed all computations and analyzed all the data.
- Created all the figures and tables, except for Figures S16 and S17 which were captured by VC.

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

Daniel Gómez-Pérez [1], Monja Schmid [1], Vasvi Chaudhry [1], Ana Velic [2], Boris Maček [2], Ariane Kemen [1,✉], and Eric Kemen [1,✉]

[1] Microbial Interactions in Plant Ecosystems, Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany; [2] Department of Biology, Quantitative Proteomics Group, Interfaculty Institute of Cell Biology, University of Tübingen, Tübingen, Germany

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 secreted proteins 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 secretion of proteins into the apoplast. Amplicon sequencing and network analysis of *Albugo*-infected and uninfected 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 candidates for heterologous expression and study of their inhibitory activity. We found that three of the candidate proteins show selective antimicrobial activity on gram-positive bacteria isolated from *Arabidopsis thaliana* and that these inhibited bacteria are 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 net charge. This is the first report of protist proteins with antimicrobial activity under apoplastic conditions that therefore are potential biocontrol tools for targeted manipulations of the microbiome.

Correspondence: *ariane.kemen@uni-tuebingen, eric.kemen@uni-tuebingen.de*

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 e.g., 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, e.g., 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 and Hertweck, 2017). The best adapted microbes are obligate biotroph symbionts or 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 and 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 in 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 protists comprising many highly adapted parasites of plants, animals, and humans. Following penetration into the plant host via the leaf surface, *Albugo* develops intercellular hyphae to colonize the plant extracellular space, known as apoplast (Berlin and 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* secretes 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 potential microbiome control agents since they selectively modify the endophytic bacterial community and can therefore be considered effectors governing microbe-microbe interactions (Snelders et al., 2020).

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 towards certain disorder-promoting amino acids and often shows hydrophilic tendencies (Dubreuil et al., 2019). The function of IDRs in proteins can vary consid-

erably 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*, and report for the first time an example of a protist and obligate biotrophic pathogen as a potential source for highly specific antimicrobials.

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 operational taxonomic units (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 (Figure 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 eukaryote (Figure 1a, Figures S2 and S3). The negatively correlated bacteria included mostly gram-negative strains of which the most abundant phylum was the Proteobacteria with four members. The negatively correlated eukaryotes included an ascomycete fungus and two green algae (Figure 1b). In summary, our analysis indicates that the protist pathogen *Albugo* is highly intercorrelated in the phyllosphere, particularly through negative correlations when compared to other plant microbes.

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 as previously described by (Agler et al., 2016), we studied *Albugo* secreted proteins in the plant apoplast (Figure 2a). Through proteomic analysis, we identified 563 proteins from *A. candida* in the apoplast of infected *A. thaliana* leaves with at least one peptide at an Andromeda score higher than ten, representing 4.2% of the total predicted proteome of *A. candida* (Figure 2b). Among these, 70 proteins (12.4%) carried a putative secretion signal and 12 had a mitochondrial inner membrane localization according to the GO assignment compared to 28 in the predicted proteome of *A. candida*, suggesting minor contamination from broken hyphae. After performing annotation of biological processes through GO terms, we found that 429 out of the 563 proteins (76.2%) resulted in significant hits (Figure S4). 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 value < 0.001; Figure 2c). Additionally, we employed a de novo antimicrobial activity prediction approach in order to find antimicrobial proteins with no known conserved regions. We found two of the tools (AmpGram and amPEPpy) to be biased towards a longer or shorter protein length in our dataset, therefore we adjusted the weight based on the R-squared value of the correlation (Figure S5). Following the weighted score, a total of 154 apoplastic proteins were found to be positive for antimicrobial activity (27.3%, compared to 26.3% in the entire *A. candida* proteome). This corresponded to a significant enrichment for the presence of predicted antimicrobial proteins in the apoplastic proteome (Mann-Whitney U test, p value $= 0.013$; Figure 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 0.5 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. Short sequence length to match the upper bound of known effector proteins (less than 600 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* (Table S1). Of note, the predicted peptide signal for classical secretion of C14 and C15 was removed during cloning. The candidates were subsequently heterologously expressed in *Escherichia 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 °C to 37 °C), lower inducing concentration (0.1 mM to 1 mM IPTG) and longer induction time (4 h to 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 (Figure S6). Therefore, we used a denaturing extraction protocol as the standard purification method for comparison of all the expressed candidates (Figure S7). After purification and concentration of the candidates in a testing buffer resembling the apoplastic pH conditions

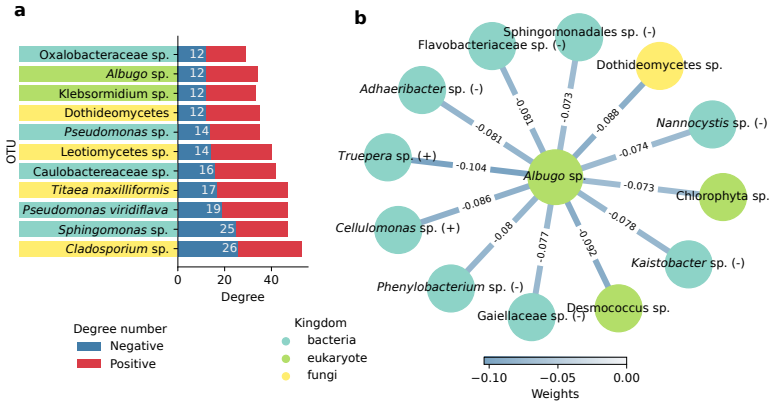


Figure 1. Network of operational taxonomic unit (OTU) interactions in the phyllosphere of *Arabidopsis thaliana*.
(**a**) Degree distribution of OTUs in the interaction network, showing nodes with 12 or more negative interactions and more than 100,000 total reads in the dataset, where *Albugo* places tenth. (**b**) Inferred negative interactions for the *Albugo* sp. OTU. 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.

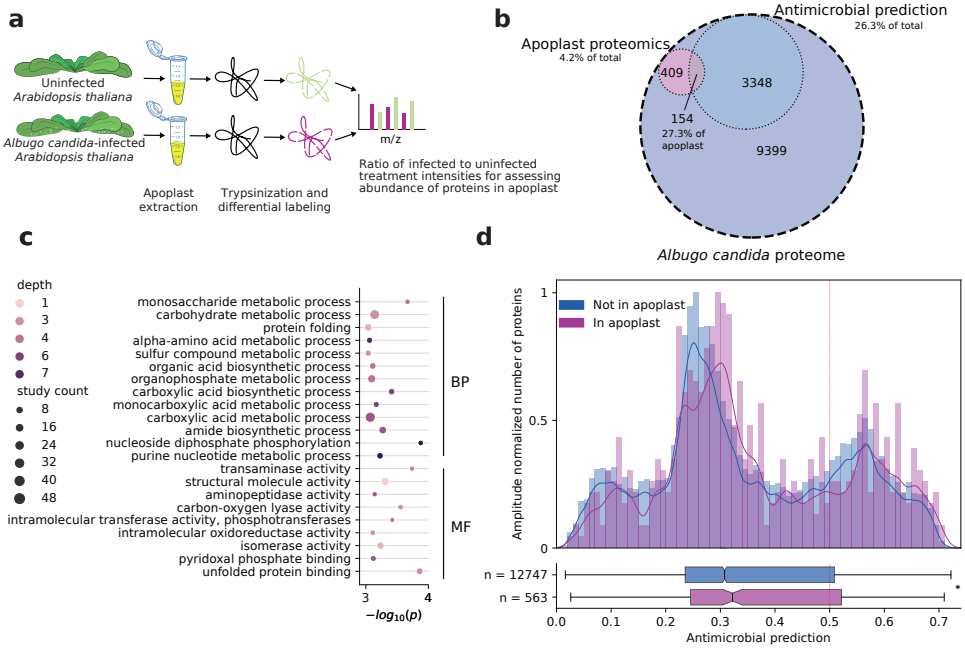


Figure 2. Proteomics of *Albugo candida* apoplastic proteins from infected *Arabidopsis thaliana* leaves.
(**a**) Workflow of the proteomics analyses on infected (10 days post infection) and uninfected *Arabidopsis thaliana*. (**b**) Venn diagram of *A. candida* proteome. A total of 563 proteins were identified in the apoplast, of which 154 have a predicted antimicrobial function. (**c**) Enrichment of gene ontology biological processes related to biological processes (BP) and molecular functions (MF) in apoplastic proteins compared to background proteome in *A. candida* with a significance of p value < 0.001. (**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 value $= 0.015$). A dashed red line indicates the cutoff of 0.5.

(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 that were cultured under standard conditions from *A. thaliana* samples (Tables S3). These strains were selected based on a pre-screen which revealed gram-positive bacteria to be more sensitive to the protein treatment than gram-negative. 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 *Aeromicrobium* (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; Figure 3b, S8 and 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 (Figure 3a). C05, instead and consistent with the antimicrobial prediction, had the least antimicrobial effect of all. We additionally found that all proteins displayed

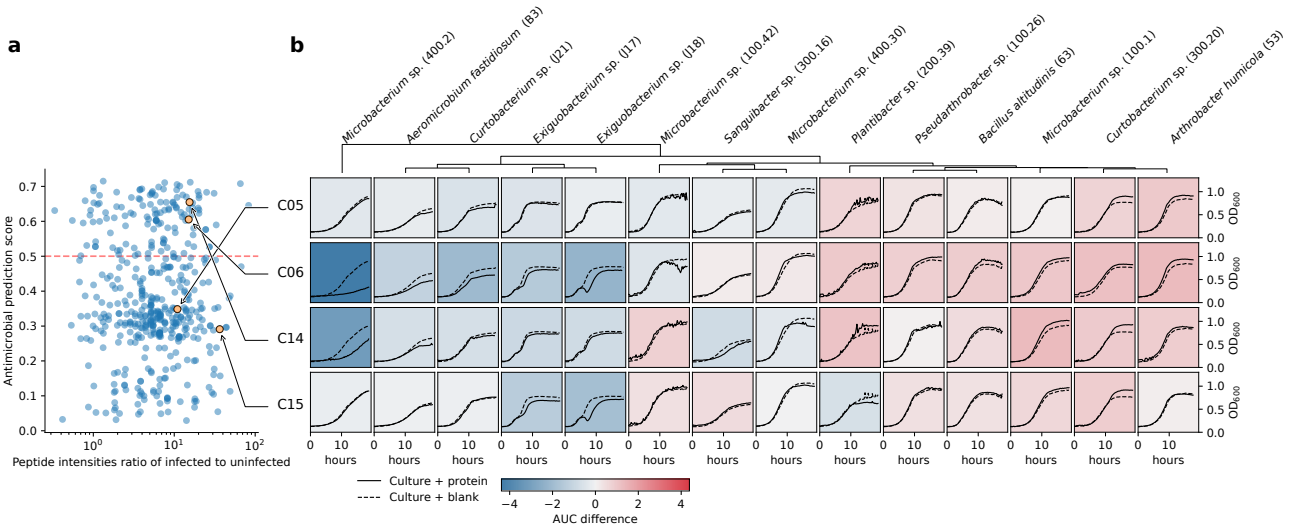


Figure 3. Antibacterial prediction and activity of apoplastic candidate proteins from *Albugo candida*.
(**a**) Antimicrobial prediction and relative intensity of the proteins found in the apoplast of *Arabidopsis thaliana*, highlighting the selected candidates. (**b**) Growth curves for candidates C05, 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. 95% confidence intervals for these tests are shown in Figure S9.

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 mostly unaffected by incubation with the protein candidates during growth (Figure 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 (Figure 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 in the apoplast revealed a significant enrichment for IDRs compared to non-apoplastic proteins (Fisher's exact one-tailed test, p value = 0.002). This, together with the highly enriched term for unfolded protein binding (Figure 2c), suggests the importance of disordered regions in the secreted proteome. Consistent with this, the candidates that showed antimicrobial activity had long predicted IDRs, notably in the C-terminus with a positive net charge (Figure 4a). 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). Additionally, C06 had a compositional bias for alanine and glutamine, and C14 and C15 had a bias towards increased serine (Table S5), 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 (Table S2), as well as the remaining N-terminal region of each protein, which was for the most part not predicted to be disordered.

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 (Figure 4b and Figure S11). 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 (Figure 4c and Figure S11). Furthermore, for C06d and C14d we observed a concentration dependent effect against *Exiguobacterium* strains I10 and I11, *Aeromicrobium fastiodiosum* strain I01 and *Microbacterium* strain I20, with C06d reaching a complete inhibitory activity against *Aeromicrobium* and *Microbacterium* growth at 2.15 µM (Figure 4d and Figure S12). In the proteomics dataset, we found not only coverage for a large part of the sequence of both positive candidates (Figure 4a) but the identified peptides also indicated the presence of the disordered domains with antimicrobial activity in the apoplast, particularly for C06d (Table S6). To support the hypothesis that the proteins might get cleaved allowing the release of the IDR peptides in the apoplast we used the proteomics dataset to study potential changes in *A. thaliana* protease activity upon infection. We found evidence of an upregulation of proteases in the apoplast, particularly subtilisin-like serine proteases, in the presence of *A. candida* infection (Table S7). 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 S6). These included *Rhodococcus fascians* (Dhaouadi et al., 2020; Hjerde et al., 2013), *Clavibacter michiganensis* subsp. *tessellarius* (Carlson and Vidaver, 1982; Li and 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. *tes-*

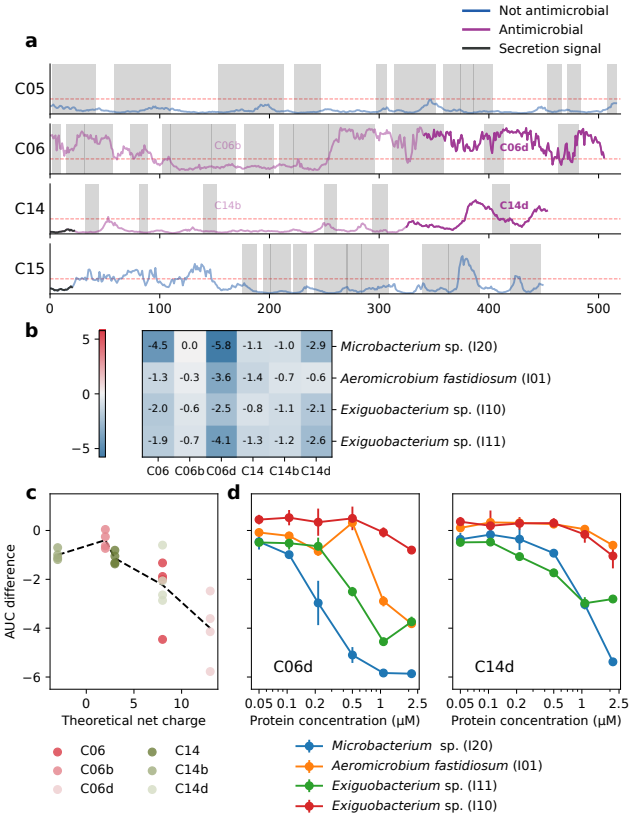


Figure 4. Antimicrobial activity of protein candidates from *Albugo candida*, C06, C14 and their domains.
(**a**) Individual candidate predictions of disorder per amino acid, emphasizing the annotated domains used for antimicrobial testing in bold (domain d) or faded color (domain b). The threshold for 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 activity of the different domains of C06 and C14 at a concentration of 1 µM shown as the differences in the area under the curve to blank during 19 h of growth. The values represent the mean of at least three biological replicates. Full curves are found in Figure S11. (**c**) Antimicrobial activity as measured by the change in the area under the curve (AUC) correlated to predicted net charge for all domain tests with LOWESS regression showing a downwards trend line. (**d**) Concentration dependent inhibitory activity of C06 and C14 C-terminal disordered domains towards four sensitive strains. Vertical bars represent standard deviation. Full curves are found in Figure S12.

sellarius (I43, I44) were unaffected by protein addition, all tested isolates from *C. michiganensis* subsp. *capsici* (I37-I42) showed a variable degree of growth inhibition (Figure S13). 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.

To examine whether the peptides showed inhibitory activity within a microbial community context, we conducted SynCom experiments in culture media. A SynCom composed of *Aeromicrobium* (I01), which was inhibited in the previous tests, and four additional strains (I03, I04, I13, I28), all unaffected in their growth by protein addition, was constructed (Figure S14). We added a rifampicin-resistant pathogen *Pseudomonas syringae* to use as readout. We found that the growth of Pst was increased when *Aeromicrobium* was excluded from the SynCom, as compared to the whole five-strain SynCom, when grown with testing buffer (Figure 5a and S15). To see if C06d could effectively inhibit *Aeromicrobium* in a multi-strain community, we added the protein solution. In this setting, the growth of Pst was comparable to the *Aeromicrobium* dropout treatment, indicating inhibition by C06d. As a control, we also tested the SynCom with C06b, which showed no inhibitory activity (Figure S14). The Pst load, as measured by the CFU count per mL was similar to the five-strain SynCom treatment in buffer (Figure Figure 5a and S15). Together, these results indicate that the inhibitory properties of C06d are persistent in a multi-strain bacterial setting.

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, all identified as core microbes, challenged with a model plant-pathogenic strain of *Pseudomonas syringae*. If in the SynCom, *Aeromicrobium fastidiosum* (I01) was not included, we observed a variable decrease in the resistance of the plant to infection (Figure 5b). We also observed an overall increase in the sample-to-sample variation of the bacterial abundances compared to the whole SynCom 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 principal coordinate analysis for the *Aeromicrobium* dropout (Figure 5c). Additionally, through pairwise interaction assays with SynCom members, this *Aeromicrobium* strain was shown to inhibit other bacterial strains. This highlighted *Aeromicrobium*'s potential in stabilizing the microbiome by controlling the abundance of other core microbes in the phyllosphere of *A. thaliana* (Figure S16). Further

microbial lawn experiments of *Aeromicrobium* showed its resilience against other SynCom members in one-to-one interactions (Figure S17).

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 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). The adjacent positive interactions imply a promotion of these organisms as potentially beneficial for *Albugo*. Within the network of positive interactions we found a correlation between *Albugo* and the bacterial genus *Variovorax*, which, as a common plant growth-promoting rhizobacterium, could promote survival of the host plant during infection, thus indirectly promoting *Albugo* (Chen et al., 2013; Finkel et al., 2020). In contrast, the numerous direct negative interactions can be explained by *Albugo*'s need to outcompete microbes that may be detrimental to its own survival (Figure 1b). 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 an interest in controlling the growth of microbes threatening survival of the host. By reducing the plant's defenses through secretion 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 contain key secondary metabolites potentially responsible for the observed broad spectrum of negative correlations with other microbes (Kemen et al., 2011). In fungi, potential protein effectors for microbe-microbe interactions were identified that have direct or indirect inhibitory effects on competitors (Eitzen et al., 2021; Snelders et al., 2020). 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 about 80% of the *A. candida*'s apoplastic proteins revealed a significant enrichment for metabolism-related processes. This might be explained by the sampling time point, that is, ten days after infection, when a high metabolic turn-over is required by the oomycete due to the active growth of hyphae. 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 and 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 and Kemen, 2021). We therefore hypothesized antimicrobial proteins or 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 high-throughput 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 (Figure S18). Additionally, in the GO enrichment of molecular functions we found the term unfolded protein binding (Figure 2c). 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 S2). 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 apoplast (Malik et al., 2016). Once secreted 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

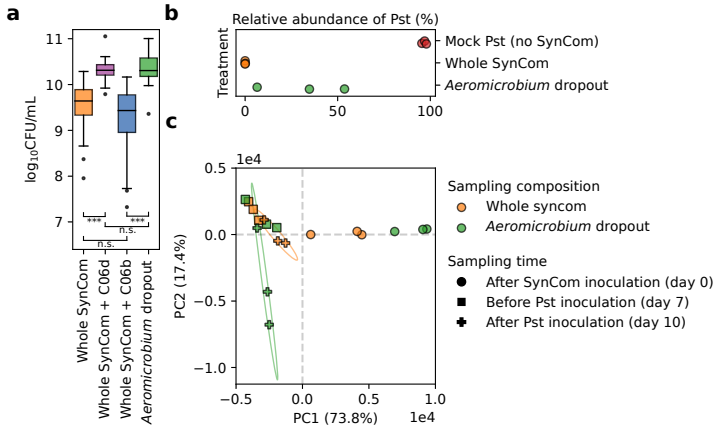


Figure 5. Synthetic community (SynCom) experiments in liquid culture and in planta suggest relevance of *Aeromicrobium fastidiosum* IO1.
(**a**) C06d (purple) and C06b (blue) 1 µM protein treatments on a five-strain bacterial community in liquid culture result in variable community disruption as measured by the bacterial load of *Pseudomonas syringae* pv. tomato DC3000 (Pst) in $CFUmL^{-1}$. Whole SynCom with buffer control (BisTris, pH 5.9) is shown in orange, *Aeromicrobium fastidiosum* (I01) dropout community in green. Significance tests show a difference between whole SynCom treatment and C06d treatment, as well as C06b treatment and *Aeromicrobium* dropout (Mann-Whitney U; n.s.: not significant, ***: p value < 0.001). Detailed counts per replicate can be found in Figure S15. (**b**) Protective effect of a 15-strain SynCom with or without *Aeromicrobium fastidiosum* (I01) as measured in planta by the relative abundances of Pst three days post infection. Mock treatment represents plants not treated with the SynCom. (**c**) Principal coordinate analysis of the amplicon sequencing variant (ASV) frequencies in the SynCom dropout experiment on *Arabidopsis thaliana*. Calculated based on the Euclidean distances of bacterial 16S rRNA ASVs abundances. Ellipses drawn for the final timepoints based on two standard deviations show a large variation for the *Aeromicrobium* dropout but not for the whole SynCom after Pst disruption.

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 S7).

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 (Figure 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 C06 had a significant compositional bias for the presence of histidine (Table S5). 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-negative. Thus, the variable antimicrobial effects of cationic peptides can be explained by the charge of the target membrane (Malanovic and 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, e.g., 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, *Aeromicrobium 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* following disruption with the bacterial phytopathogen Pst (Figure 5b and c). In line with observations from Agler et al. (2016), *Albugo* infection results in a reduced alpha diversity in the phyllosphere community. 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 antimicrobial peptide C06d to inhibit bacteria in a complex in vitro community, further adds to this hypothesis (Figure 5a).

In summary, we have found three apoplastic proteins from the plant protist pathogen *A. candida* to be antimicrobial on several plant-isolated strains of gram-positive bacteria. These proteins were selected after apoplast proteomic analysis of leaf samples and in silico 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.

Methods

Interaction network inference

We inferred interaction correlations between OTUs on an amplicon sequencing dataset of *Arabidopsis thaliana*'s phyllosphere sampled twice a year from 2014 to 2019 in several sites around the area of Tübingen, Germany. Microbiome samples were isolated from the endo- or epiphytic compartment separately. The dataset combined analyses of 16S and 18S ribosomal RNA phylogenetic marker genes as well as of the ITS regions. We analyzed the raw OTU tables (Mahmoudi et. al, 2022, in preparation) that were constructed after a 97% similarity OTU clustering, using the software FlashWeave with a 0.05 significance cutoff, controlling for confounding meta variables including season, year, sampling site and phyllosphere compartment (Tackmann et al., 2019). We considered OTUs with a taxonomy assignment bootstrap below 90 as unclassified for that taxonomy level.

Plant growth and infection

We infected *A. thaliana* plants of ecotype Ws-0, which had been grown under short day conditions for five weeks (8 h of light at 21 °C and 16 h of darkness at 16 °C), with *A. candida* strain Nc2 by spore suspension spray-inoculation. To obtain the latter, we submerged leaves with visibly sporulating *A. candida* pustules (at least 12 days post infection, dpi) in sterile ultrapure water for one hour and filtered the solution through miracloth (pore size of 25 µm). Following a cold treatment at 8 °C overnight and condensing humidity, we kept the inoculated plants under long day conditions (12 h of light and 12 h of darkness).

Sample preparation for mass spectrometry analysis

We extracted apoplast samples from *A. candida*-infected and uninfected *A. thaliana* leaves at 10 dpi using vacuum pump infiltration at 100 Pa with 180 mM MES buffer (pH 5.5). We collected the infiltrate through centrifugation in a swinging-bucket centrifuge at 1000 RCF for 20 min. We performed acetone-methanol precipitation overnight at −20 °C. We resuspended the protein pellets in a denaturation buffer (6 M urea, 2 M thiourea, 10 mM Tris, pH 8), and determined their concentration by Bradford assay. Disulfide bonds were reduced with 10 mM DTT before alkylation with 55 mM iodoacetamide, for one hour each. We used trypsin (Promega Corporation) for digestion (Zittlau et al., 2021), and purified the peptides on Sep-Pak C18 Cartridges (Waters).

High pH fractionation and dimethyl labeling

The Pierce High pH Reversed-Phase Peptide Fractionation Kit (kit no.: 84868; Thermo Fisher Scientific) was used to fractionate 150 µg of triple dimethyl-labeled and mixed proteome. Spin columns were conditioned twice with ACN and 0.1% TFA according to the vendor. We loaded the acidified peptides on columns and washed them once with water prior to elution. Peptides were eluted stepwise with 5%, 7.5%, 10%, 12.5%, 13.3%, 15%, 17.5%, 20% and final 50% ACN/ammonia. We acidified the fractions to pH 2 to 3 and desalted them on C18 StageTips before starting LC-MS/MS measurement. In total, we collected nine fractions per replicate. As previously described, we labeled the peptides with dimethyl on Sep-Pak C18 Cartridges (Boersema et al., 2009). We mixed triple dimethyl sets at an equal peptide ratio of 1:1:1. We inspected label efficiency and mixing in separate LC–MS/MS runs.

LC-MS analysis

We analyzed the samples on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific), and on an Exploris mass spectrometer (Thermo Fisher Scientific). An online-coupled Easy-nLC 1200 UHPLC (Thermo Fisher Scientific) was used to separate the peptides on a 20 cm analytical column, 75 µm ID PicoTip fused silica emitter (New Objective). This column was in-house packed with ReproSil-Pur C18-AQ 1.9 µm resin (Dr Maisch GmbH Ltd). We generated a gradient by solvent A (0.1% FA) and solvent B (0.1% FA in 80% ACN), at 40 °C and a 200 nL min^{-1} flow rate. We eluted the peptide fractions using a 90-min segmented linear gradient. We electrosprayed and analyzed eluted peptides in a positive ion, data-dependent acquisition mode. We selected the top 12 most intense peptides. Full mass spectrometry was acquired in a scan range of 300 to 1750 m/z at a resolution of 60,000.

MS data analysis and statistical analysis

We processed raw data files with the MaxQuant software suite version 1.6.7.0 (Cox and Mann, 2008). We searched the MS/MS data 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. We kept the search parameters to default values except for the dimethylation for light (28.03 Da), intermediate (32.06 Da), and heavy (36.08 Da) labels on lysine residues and peptide N-termini. We set oxidation of methionine and protein N-terminal acetylation as variable modifications and allowed carbamidomethylation of cysteine residues as fixed modification. All searches were performed in trypsin/P-specific digestion mode. Maximum of two missed cleavages were allowed. We considered an *Albugo* protein as present in the apoplast with a high confidence when it had at least one predicted peptide match with an Andromeda score higher than ten (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

For the functional annotation, we used InterProScan version 5 (Jones et al., 2014). We analyzed the assigned gene ontology (GO) terms using the GOATOOLS software (Klopfenstein et al., 2018). 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 vr. 2, AmpGram and amPEPpy (Burdukiewicz et al., 2020; Lawrence et al., 2020; Veltri et al., 2018). 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 added a weight to each method to account for their correlation to protein size, corresponding to one minus the absolute of the R-squared value when the least-squares linear regression was significant. We considered a prediction as positive for antimicrobial activity when the weighted average of the three methods was higher than 0.5. For the prediction of disordered regions, we used the tool flDPnn with default settings and 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. We calculated the molecular properties of the candidates, including theoretical isoelectric point and molecular weight using Expasy's ProtParam (Gasteiger et al., 2005) and net charge via the Henderson-Hasselbalch equation (Moore, 1985). To detect compositional bias in the protein sequences we used fLPS 2.0 (Harrison, 2021).

Cloning of constructs

We synthesized complementary DNA (cDNA) with SuperScript™ II Reverse Transcriptase (Invitrogen) from total RNA extracted from *A. candida*-infected *A. thaliana* leaves at 8 dpi using an RNeasy kit (QIAGEN). We amplified the candidate sequences for the genes of interest

with PCR using Phusion® High-Fidelity DNA polymerase (NEB) and primers designed for subsequent cloning (Table S4). We cloned the amplicons using the In-Fusion® cloning method (Takara Bio) into pET28b vectors for expression in *Escherichia coli* with two 6x Histidine tags (6xHis-tags) flanking the gene of interest for candidates C05, C06 and derivatives; and one 6xHis-tag at the C-terminus for the remaining candidates (C14 and derivatives, C15). We cloned candidates that had a putative secretion signal as predicted by SignalP version 5, namely C14 and C15, without the secretion-signal encoding sequence (Armenteros et al., 2019).

Protein expression

We overexpressed the candidate proteins in the *E. coli* strain SHuffle® (NEB) at 30 °C (for candidates C14 and derivatives) and *E. coli* strain Rosetta™ DE3 (Merck) at 37 °C (for candidates C05, C15, C06 and derivatives). We induced expression at a 600 nm optical density (OD_{600}) of 0.5-0.6 by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. We harvested 5 h post induction by collecting the pellets through centrifugation for 10 min at 8000 RCF, shock freezing and storing at −80 °C until further processing.

Denaturing purification

We extracted the inclusion bodies from the cell pellets under denaturing conditions via sonication for 10 min (13 kHz in 0.6 s long pulses followed by 0.4 s of rest) in lysis buffer (100 mM sodium phosphate, 10 mM Tris, 7 M urea, pH 8) followed by shaking at room temperature for 1 h and centrifugation for 1 h at 10,000 RCF and 4 °C. We purified the proteins with HisTrap™ excel (GE Healthcare) in a one-step elution (elution buffer: 100 mM sodium phosphate, 10 mM Tris, 7 M urea, pH 4.5). We concentrated the elution by centrifugation in a VivaSpin® 20 column (Sartorius). We performed refolding by dialysis in a 3,500-5,000 molecular weight cutoff Float-a-lyzer® G2 device (Spectrum Labs). We gradually exchanged the denaturing elution buffer with the testing buffer (10 mM BisTris, pH 5.9) in four steps over a period of 48 h at 7 °C. We kept the final rebuffering solution in contact with the dialysis tube as a negative control for the antimicrobial activity tests. We determined the final protein concentration by Bradford assay using a Bovine Serum Albumin (BSA) standard (Thermo Scientific) and assessed protein purity by SDS-PAGE.

Antimicrobial testing

We tested strains from a stock of plant-isolated bacteria collected during the *A. thaliana* sampling used to create the amplicon sequencing dataset mentioned above (Tables S3). The strain library comprises epi- and endophytic bacterial species isolated from either *Albugo*-infected or non-infected plants and includes the strains contained in the synthetic community (SynCom) mentioned in the next section. We applied a standard culturing approach to isolate strains from the *A. thaliana* samples. We used four different media for bacterial isolation (nutrient broth, Reasoner's 2A agar, tryptic soy broth and King's B agar) and cultivated at 22 °C. The strains used in this study were taxonomically assigned through similarity of their 16S region through blastn of the National Center for Biotechnology Information (NCBI) ribosomal RNA database. 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 volume ratio of 1:1 to the cell cultures diluted to a starting OD_{600} of 0.1 for a final testing volume of 100 µL. All tests were performed in at least three biological replicates in 96-well VWR transparent flat bottom plates. We used the TriStar2 plate reader (Berthold) with a program that measured OD_{600} every 15 min over 19 h with constant orbital medium strength shaking at room temperature (22 °C).
We conducted bacterial community experiments in culture with two of the peptides (C06d and C06b). For these, we constructed a five-strain SynCom (I01, I03, I04, I13, I28; Table S1). We grew overnight cultures of the individual strains at 22 °C, adjusted their OD_{600} to 0.2 in fresh NB medium and mixed them at equal volume to form the SynCom, with a final OD_{600} of 0.02 for each strain in the mix. We combined 45 µL of the SynCom with 50 µL of buffer or protein solution adjusted to 1 µM final well concentration in 96-well VWR transparent flat bottom plates, followed by incubation at 22 °C and 180 rpm shaking for 24 h. We added 5 µL of a rifampicin-resistant *Pseudomonas syringae* pv. tomato DC3000 (Pst) culture (OD_{600} = 0.2) to a total of 100 µL per well and continued incubation for another 24 h. We created dilution lines of each well and dropped 10 µl of the dilutions on NB agar plates containing 50 µg mL^{-1} rifampicin. We counted Pst colonies after incubating the plates for 32 h h at 22 °C. The experiments were repeated three times with nine biological replicates per treatment and round. We performed the following treatment combinations: 1. SynCom plus buffer, 2. SynCom without I01 plus buffer, 3. SynCom plus peptide C06d, 4. SynCom plus peptide C06b.
For the *Aeromicrobium fastidiosum* (I01) interactions with other bacterial species from the SymCom, we examined one-to-one inhibition by a co-cultivation setup. We grew overnight cultures in NB medium and adjusted OD_{600} to 1.0. We spread a lawn of test strains on NB agar medium plates and aseptically created a hole with a sterile cork borer to dispense 100 µL of culture. We incubated the plates at 22 °C for three days to visualize the zone of inhibition.

Synthetic microbial community experiment

For the SynCom experiment, we used *A. thaliana* plants of ecotype Ws-0. We sterilized the seeds with chlorine gas for 6 h (Lindsey et al., 2017). We further checked for bacterial and fungal seed-borne contaminants on NB and potato dextrose agar, respectively, by incubation for one week at 22 °C. After one week, seedlings were singularised and transferred to 1/2 Murashige and Skoog (MS) media in sterile 12-well plates (Greiner bio-one) and grown for three more weeks. The SynCom was assembled with core microbes from the wild *A. thaliana* population, consisting of 12 bacteria and three yeasts (Table S3). Members of the SynCom were grown and adjusted to an OD_{600} of 0.2 in 10 mM $MgCl_2$ with 0.04% Silwet. They were mixed in equal parts and sprayed onto the 4-week-old seedlings. One week after initial SynCom inoculation, we sprayed the plants with rifampicin-resistant model pathogen Pst. Pst was grown in NB at 22 °C with shaking at 180 rpm and processed in the same way as the SynCom members. We sampled plants for amplicon sequencing at three time points, after SynCom spraying on day 0, on day 7 be-

fore inoculation with Pst and on day 10, third day post Pst infection. In total, three biological replicates were taken from each of the three treatments, whole SynCom, SynCom without *Aeromicrobium* (I01) and mock control without SynCom. DNA extractions were performed using a PowerSoil DNA Isolation Kit (MO BIO Laboratories Inc.). Amplicon libraries for sequencing were prepared using V5-V7 regions from the small prokaryotic subunit (16S rRNA gene) and ITS2 regions (internal transcribed spacer) for targeting bacterial and fungal amplicons, respectively (Agler et al., 2016). Sequencing was performed with Illumina MiSeq platform using V3 kit (600 cycles). Finally, we processed and analyzed the raw sequencing reads using QIIME2, with which we calculated Pst relative abundances and performed principal coordinate analysis based on Euclidean distances of the rarefied amplicon sequencing variant (ASV) frequencies to measure sample to sample variation (Bolyen et al., 2019).

Bibliography

M. Amine Hassani, Paloma Durán, and Stéphane Hacquard. Microbial interactions within the plant holobiont. *Microbiome*, 6(1):58, 2018. doi: 10.1186/s40168-018-0445-0.

Shiri Freilich, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S. Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications*, 2(1):589, 2011. doi: 10.1038/ncomms1597.

Tobias Bollenbach. Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Current Opinion in Microbiology*, 27:1–9, 2015. ISSN 1369-5274. doi: 10.1016/j.mib.2015.05.008.

Evelyn M Molloy and Christian Hertweck. Antimicrobial discovery inspired by ecological interactions. *Current Opinion in Microbiology*, 39:121–127, 2017. ISSN 1369-5274. doi: 10.1016/j.mib.2017.09.006.

Jonas Ruhe, Matthew T. Agler, Aleksandra Placzek, Katharina Kramer, Iris Finkemeier, and Eric M. Kemen. Obligate Biotroph Pathogens of the Genus Albugo Are Better Adapted to Active Host Defense Compared to Niche Competitors. *Frontiers in Plant Science*, 7:820, 2016. doi: 10.3389/fpls.2016.00820.

Matthew T. Agler, Jonas Ruhe, Samuel Kroll, Constanze Morhenn, Sang-Tae Kim, Detlef Weigel, and Eric M. Kemen. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biology*, 14(1):e1002352, 2016. ISSN 1544-9173. doi: 10.1371/journal.pbio.1002352.

Jerry D. Berlin and C. C. Bowen. The Host-parasite Interface of Albugo candida on Raphanus sativus. *American Journal of Botany*, 51(4):445–452, 1964. ISSN 0002-9122. doi: 10.1002/j.1537-2197.1964.tb06655.x.

Eric Kemen, Anastasia Gardiner, Torsten Schultz-Larsen, Ariane C. Kemen, Alexi L. Balmuth, Alexandre Robert-Seilaniantz, Kate Bailey, Eric Holub, David J. Studholme, Dan MacLean, and Jonathan D. G. Jones. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. *PLoS Biology*, 9(7):e1001094, 2011. ISSN 1544-9173. doi: 10.1371/journal.pbio.1001094.

Oliver J Furzer, Volkan Cevik, Sebastian Fairhead, Kate Bailey, Amey Redkar, Christian Schudoma, Dan MacLean, Eric B Holub, and Jonathan D G Jones. An Improved Assembly of the Albugo candida Ac2V Genome Reveals the Expansion of the "CCG" Class of Effectors. *Molecular Plant-Microbe Interactions®*, 35(1):39–48, 2022. ISSN 0894-0282. doi: 10.1094/mpmi-04-21-0075-r.

Nick C. Snelders, Hanna Rovenich, Gabriella C. Petti, Mercedes Rocafort, Grardy C. M. van den Berg, Julia A. Vorholt, Jeroen R. Mesters, Michael F. Seidl, Reindert Nijland, and Bart P. H. J. Thomma. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nature Plants*, 6(11):1365–1374, 2020. doi: 10.1038/s41477-020-00799-5.

Macarena Marín, Vladimir N Uversky, and Thomas Ott. Intrinsic Disorder in Pathogen Effectors: Protein Flexibility as an Evolutionary Hallmark in a Molecular Arms Race. *The Plant Cell*, 25(9):3153–3157, 2013. ISSN 1040-4651. doi: 10.1105/tpc.113.116319.

Christopher J. Oldfield, Vladimir N. Uversky, A. Keith Dunker, and Lukasz Kurgan. Intrinsically Disordered Proteins. pages 1–34, 2019. doi: 10.1016/b978-0-12-816348-1.00001-6.

Benjamin Dubreuil, Or Matalon, and Emmanuel D. Levy. Protein Abundance Biases the Amino Acid Composition of Disordered Regions to Minimize Non-functional Interactions. *Journal of Molecular Biology*, 431(24):4978–4992, 2019. ISSN 0022-2836. doi: 10.1016/j.jmb.2019.08.008.

Liping Liu, Le Xu, Qie Jia, Rui Pan, Ralf Oelmüller, Wenying Zhang, and Chu Wu. Arms race: diverse effector proteins with conserved motifs. *Plant Signaling & Behavior*, 14(2):1557008, 2019. doi: 10.1080/15592324.2018.1557008.

Ties Latendorf, Ulrich Gerstel, Zhihong Wu, Joachim Bartels, Alexander Becker, Andreas Tholey, and Jens-Michael Schröder. Cationic Intrinsically Disordered Antimicrobial Peptides (CIDAMPs) Represent a New Paradigm of Innate Defense with a Potential for Novel Anti-Infectives. *Scientific Reports*, 9(1):3331, 2019. doi: 10.1038/s41598-019-39219-w.

Juliana Almario, Maryam Mahmoudi, Samuel Kroll, Mathew Agler, Aleksandra Placzek, Alfredo Mari, and Eric Kemen. The Leaf Microbiome of Arabidopsis Displays Reproducible Dynamics and Patterns throughout the Growing Season. *mBio*, pages e02825–21, 2022. doi: 10.1128/mbio.02825-21.

Vladimir N Uversky. The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins*, 1(1):e24684, 2013. ISSN 2169-0693. doi: 10.4161/idp.24684.

Sabrine Dhaouadi, Mougou A. H, and Ali Rhouma. The plant pathogen Rhodococcus fascians. History, disease symptomatology, host range, pathogenesis and plant–pathogen interaction. *Annals of Applied Biology*, 177(1):4–15, 2020. ISSN 0003-4746. doi: 10.1111/aab.12600.

Erik Hjerde, Marcin M. Pierechod, Adele K. Williamson, Gro E. K. Bjerga, Nils P. Willassen, Arne O. Smalås, and Bjørn Altermark. Draft Genome Sequence of the Actinomycete Rhodococcus sp. Strain AW25M09, Isolated from the Hadsel Fjord, Northern Norway. *Genome Announcements*, 1(2):e00055–13, 2013. ISSN 2169-8287. doi: 10.1128/genomea.00055-13.

Randall R. Carlson and Anne K. Vidaver. 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(3):315–326, 1982. ISSN 0020-7713. doi: 10.1099/00207713-32-3-315.

Xiang (Sean) Li and Xiaoli (Kat) Yuan. Genome Sequences for Multiple Clavibacter Strains from Different Subspecies. *Genome Announcements*, 5(38):e00721–17, 2017. ISSN 2169-8287. doi: 10.1128/genomea.00721-17.

Eom-Ji Oh, Chungyun Bae, Han-Beoyl Lee, In Sun Hwang, Hyok-In Lee, Mi Chi Yea, Kyu-Ock Yim, Seungdon Lee, Sunggi Heu, Jae-Soon Cha, and Chang-Sik Oh. Clavibacter michiganensis subsp. capsici subsp. nov., causing bacterial canker disease in pepper. *International Journal of Systematic and Evolutionary Microbiology*, 66(10):4065–4070, 2016. ISSN 1466-5026. doi: 10.1099/ijsem.0.001311.

Mansoor Karimi Jashni, Rahim Mehrabi, Jérôme Collemare, Carl H. Mesarich, and Pierre J. G. M. de Wit. 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, 2015. doi: 10.3389/fpls.2015.00584.

Yan Wang, Yuanchao Wang, and Yiming Wang. Apoplastic Proteases: Powerful Weapons against Pathogen Infection in Plants. *Plant Communications*, 1(4):100085, 2020. ISSN 2590-3462. doi: 10.1016/j.xplc.2020.100085.

A J Cooper, A O Latunde-Dada, A Woods-Tör, J Lynn, J A Lucas, I R Crute, and E B Holub. Basic Compatibility of Albugo candida in Arabidopsis thaliana and Brassica juncea Causes Broad-Spectrum Suppression of Innate Immunity. *Molecular Plant-Microbe Interactions®*, 21(6):745–756, 2008. ISSN 0894-0282. doi: 10.1094/mpmi-21-6-0745.

Lin Chen, Ian C. Dodd, Julian C. Theobald, Andrey A. Belimov, and William J. Davies. 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(6):1565–1573, 2013. ISSN 0022-0957. doi: 10.1093/jxb/ert031.

Omri M. Finkel, Isai Salas-González, Gabriel Castrillo, Jonathan M. Conway, Theresa F. Law, Paulo José Pereira Lima Teixeira, Ellie D. Wilson, Connor R. Fitzpatrick, Corbin D. Jones, and Jeffery L. Dangl. A single bacterial genus maintains root growth in a complex microbiome. *Nature*, 587(7832):103–108, 2020. ISSN 0028-0836. doi: 10.1038/s41586-020-2778-7.

Ebru L. Aydogan, Gerald Moser, Christoph Müller, Peter Kämpfer, and Stefanie P. Glaeser. 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, 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.00144.

Camila Carlos, Huan Fan, and Cameron R. Currie. Substrate Shift Reveals Roles for Members of Bacterial Consortia in Degradation of Plant Cell Wall Polymers. *Frontiers in Microbiology*, 9:364, 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.00364.

Katharina Eitzen, Priyamedha Sengupta, Samuel Kroll, Eric Kemen, and Gunther Doehlemann. A fungal member of the Arabidopsis thaliana phyllosphere antagonizes Albugo laibachii via a GH25 lysozyme. *eLife*, 10:e65306, 2021. doi: 10.7554/elife.65306.

Jie Zhang and Jian-Min Zhou. Plant Immunity Triggered by Microbial Molecular Signatures. *Molecular Plant*, 3(5):783–793, 2010. ISSN 1674-2052. doi: 10.1093/mp/ssq035.

Daniel Gómez-Pérez and Eric Kemen. Predicting Lifestyle from Positive Selection Data and Genome Properties in Oomycetes. *Pathogens*, 10(7):807, 2021. ISSN 2076-0817. doi: 10.3390/pathogens10070807.

Fabien Plisson, Obed Ramírez-Sánchez, and Cristina Martínez-Hernández. Machine learning-guided discovery and design of non-hemolytic peptides. *Scientific Reports*, 10(1):16581, 2020. doi: 10.1038/s41598-020-73644-6.

Erum Malik, Sarah R. Dennison, Frederick Harris, and David A. Phoenix. pH Dependent Antimicrobial Peptides and Proteins, Their Mechanisms of Action and Potential as Therapeutic Agents. *Pharmaceuticals*, 9(4):67, 2016. doi: 10.3390/ph9040067.

Kara J. Cutrona, Bethany A. Kaufman, Dania M. Figueroa, and Donald E. Elmore. Role of arginine and lysine in the antimicrobial mechanism of histone-derived antimicrobial peptides. *FEBS Letters*, 589(24PartB):3915–3920, 2015. ISSN 0014-5793. doi: 10.1016/j.febslet.2015.11.002.

Nermina Malanovic and Karl Lohner. Antimicrobial Peptides Targeting Gram-Positive Bacteria. *Pharmaceuticals*, 9(3):59, 2016. doi: 10.3390/ph9030059.

Janko Tackmann, João Frederico Matias Rodrigues, and Christian von Mering. Rapid Inference of Direct Interactions in Large-Scale Ecological Networks from Heterogeneous Microbial Sequencing Data. *Cell Systems*, 9(3):286–296.e8, 2019. ISSN 2405-4712. doi: 10.1016/j.cels.2019.08.002.

Katharina I. Zittlau, Anna Lechado-Terradas, Nicolas Nalpas, Sven Geisler, Philipp J. Kahle, and Boris Macek. Temporal Analysis of Protein Ubiquitylation and Phosphorylation During Parkin-Dependent Mitophagy. *Molecular & Cellular Proteomics : MCP*, 21(2):100191, 2021. ISSN 1535-9476. doi: 10.1016/j.mcpro.2021.100191.

Paul J Boersema, Reinout Raijmakers, Simone Lemeer, Shabaz Mohammed, and Albert J R Heck. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, 4(4):484–494, 2009. ISSN 1754-2189. doi: 10.1038/nprot.2009.21.

Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature*

Biotechnology, 26(12):1367–1372, 2008. ISSN 1087-0156. doi: 10.1038/nbt.1511.

Matthew G Links, Eric Holub, Rays HY Jiang, Andrew G Sharpe, Dwayne Hegedus, Elena Beynon, Dean Sillito, Wayne E Clarke, Shihomi Uzuhashi, and Mohammad H Borhan. De novo sequence assembly of Albugo candida reveals a small genome relative to other biotrophic oomycetes. *BMC Genomics*, 12(1):503–503, 2011. doi: 10.1186/1471-2164-12-503.

Jürgen Cox, Nadin Neuhauser, Annette Michalski, Richard A. Scheltema, Jesper V. Olsen, and Matthias Mann. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research*, 10(4):1794–1805, 2011. ISSN 1535-3893. doi: 10.1021/pr101065j.

Philip Jones, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, John Maslen, Alex Mitchell, Gift Nuka, Sebastien Pesseat, Antony F. Quinn, Amaia Sangrador-Vegas, Maxim Scheremetjew, Siew-Yit Yong, Rodrigo Lopez, and Sarah Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu031.

D. V. Klopfenstein, Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, Aurélien Naldi, Christopher J. Mungall, Jeffrey M. Yunes, Olga Botvinnik, Mark Weigel, Will Dampier, Christophe Dessimoz, Patrick Flick, and Haibao Tang. GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, 8(1):10872, 2018. doi: 10.1038/s41598-018-28948-z.

Michał Burdukiewicz, Katarzyna Sidorczuk, Dominik Rafacz, Filip Pietluch, Jarosław Chilimoniuk, Stefan Rödiger, and Przemysław Gagat. Proteomic Screening for Prediction and Design of Antimicrobial Peptides with AmpGram. *International Journal of Molecular Sciences*, 21(12):4310, 2020. doi: 10.3390/ijms21124310.

Travis J Lawrence, Dana L Carper, Margaret K Spangler, Alyssa A Carrell, Tomás A Rush, Stephen J Minter, David J Weston, and Jessy L Labbé. amPEPpy 1.0: A portable and accurate antimicrobial peptide prediction tool. *Bioinformatics*, 37(14):btaa917–, 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa917.

Daniel Veltri, Uday Kamath, and Amarda Shehu. Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, 34(16):2740–2747, 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty179. Antimicrobial peptide scanner v2 citation.

Gang Hu, Akila Katuwawala, Kui Wang, Zhonghua Wu, Sina Ghadermarzi, Jianzhao Gao, and Lukasz Kurgan. flDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications*, 12(1):4438, 2021. doi: 10.1038/s41467-021-24773-7.

Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. The Proteomics Protocols Handbook. pages 571–607, 2005. doi: 10.1385/1-59259-890-0:571.

Dexter S Moore. Amino acid and peptide net charges: A simple calculational procedure. *Biochemical Education*, 13(1):10–11, 1985. ISSN 1879-1468. doi: 10.1016/0307-4412(85)90114-1.

Paul M. Harrison. fLPS 2.0: rapid annotation of compositionally-biased regions in biological sequences. *PeerJ*, 9:e12363, 2021. ISSN 2167-8359. doi: 10.7717/peerj.12363.

José Juan Almagro Armenteros, Konstantinos D. Tsirigos, Casper Kaae Sønderby, Thomas Nordahl Petersen, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4):420–423, 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0036-z.

Benson E. Lindsey, Luz Rivero, Chistopher S. Calhoun, Erich Grotewold, and Jelena Brkljacic. Standardized Method for High-throughput Sterilization of Arabidopsis Seeds. *Journal of Visualized Experiments : JoVE*, (128):56587, 2017. doi: 10.3791/56587.

Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, Eric J. Alm, Manimozhiyan Arumugam, Francesco Asnicar, Yang Bai, Jordan E. Bisanz, Kyle Bittinger, Asker Brejnrod, Colin J. Brislawn, C. Titus Brown, Benjamin J. Callahan, Andrés Mauricio Caraballo-Rodríguez, John Chase, Emily K. Cope, Ricardo Da Silva, Christian Diener, Pieter C. Dorrestein, Gavin M. Douglas, Daniel M. Durall, Claire Duvallet, Christian F. Edwardson, Madeleine Ernst, Mehrbod Estaki, Jennifer Fouquier, Julia M. Gauglitz, Sean M. Gibbons, Deanna L. Gibson, Antonio Gonzalez, Kestrel Gorlick, Jiarong Guo, Benjamin Hillmann, Susan Holmes, Hannes Holste, Curtis Huttenhower, Gavin A. Huttley, Stefan Janssen, Alan K. Jarmusch, Lingjing Jiang, Benjamin D. Kaehler, Kyo Bin Kang, Christopher R. Keefe, Paul Keim, Scott T. Kelley, Dan Knights, Irina Koester, Tomasz Kosciolek, Jorden Kreps, Morgan G. I. Langille, Joslynn Lee, Ruth Ley, Yong-Xin Liu, Erikka Loftfield, Catherine Lozupone, Massoud Maher, Clarisse Marotz, Bryan D. Martin, Daniel McDonald, Lauren J. McIver, Alexey V. Melnik, Jessica L. Metcalf, Sydney C. Morgan, Jamie T. Morton, Ahmad Turan Naimey, Jose A. Navas-Molina, Louis Felix Nothias, Stephanie B. Orchanian, Talima Pearson, Samuel L. Peoples, Daniel Petras, Mary Lai Preuss, Elmar Pruesse, Lasse Buur Rasmussen, Adam Rivers, Michael S. Robeson, Patrick Rosenthal, Nicola Segata, Michael Shaffer, Arron Shiffer, Rashmi Sinha, Se Jin Song, John R. Spear, Austin D. Swafford, Luke R. Thompson, Pedro J. Torres, Pauline Trinh, Anupriya Tripathi, Peter J. Turnbaugh, Sabah Ul-Hasan, Justin J. J. van der Hooft, Fernando Vargas, Yoshiki Vázquez-Baeza, Emily Vogtmann, Max von Hippel, William Walters, Yunhu Wan, Mingxun Wang, Jonathan Warren, Kyle C. Weber, Charles H. D. Williamson, Amy D. Willis, Zhenjiang Zech Xu, Jesse R. Zaneveld, Yilong Zhang, Qiyun Zhu, Rob Knight, and J. Gregory Caporaso. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8):852–857, 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0209-9.

Yasset Perez-Riverol, Jingwen Bai, Chakradhar Bandla, David García-Seisdedos, Suresh Hewapathirana, Selvakumar Kamatchinathan, Deepti J Kundu, Ananth Prakash, Anika Frericks-Zipper, Martin Eisenacher, Mathias Walzer, Shengbo Wang, Alvis Brazma, and Juan Antonio Vizcaíno. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, 50(D1):D543–D552, 2021. ISSN 0305-1048. doi: 10.1093/nar/gkab1038.

Data availability

All data discussed in this paper as well as the code to reproduce the analyses and figures are found at https://doi.org/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).

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to acknowledge support from the graduate school GRK 1708 "Molecular principles of bacterial survival strategies" and the Open Access Publishing Fund of the University of Tübingen. We would like to thank Libera Lo Presti for her comments and suggestions on the manuscript. Furthermore, we would like to thank Maryam Mahmoudi for providing the raw OTU tables from the amplicon sequencing, Samuel Kroll, Paul Runge and Jonas Ruhe for isolating and sequencing some of the tested bacterial strains and Sophia Häußler for her help during the SynCom experiments.

Supplementary figures

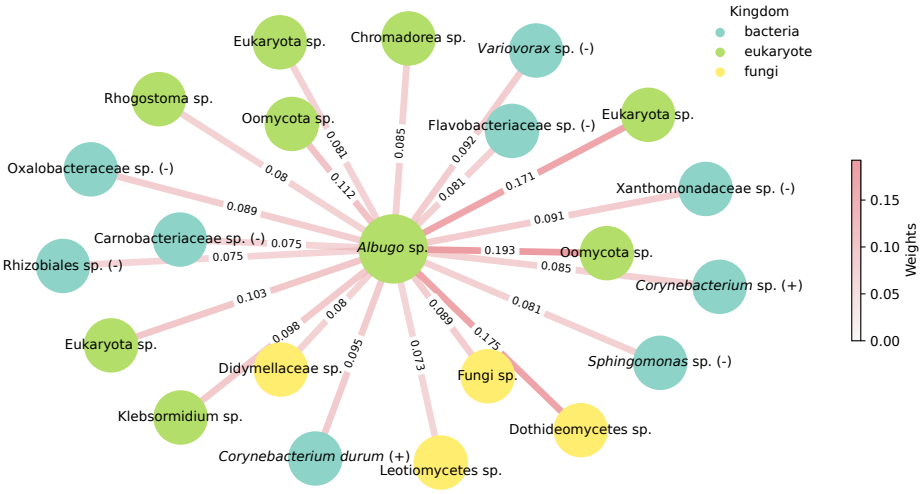


Figure S1. Inferred positive interactions for the *Albugo* sp. operational taxonomic unit on the *Arabidopsis thaliana* phyllosphere amplicon dataset.
Color of the edges represents the strength of the correlation and color of the nodes represents phylogenetic kingdom. Unless explicitly stated, taxonomy at the species level could not be resolved with confidence (bootstrap < 90).

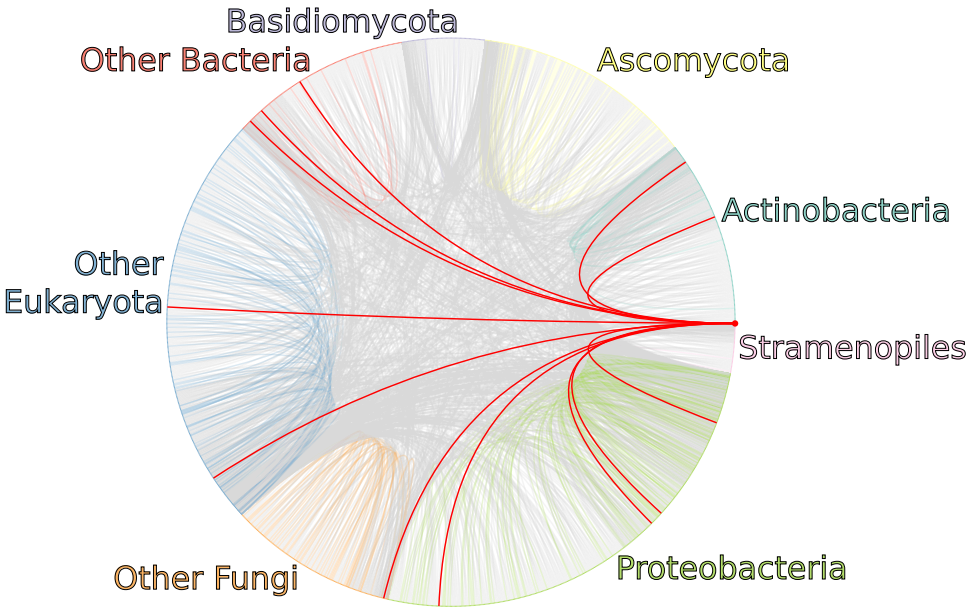


Figure S2. Overview of negative correlation network inferred by FlashWeave on the *Arabidopsis thaliana* phyllosphere amplicon dataset.
Operational taxonomic unit (OTU) nodes are grouped based on phylogenetic similarity and ordered by edge count. Highlighted are the *Albugo* sp. OTU and its connecting edges. Intergroup edges are colored gray while intragroup edges have the same color as the corresponding nodes.

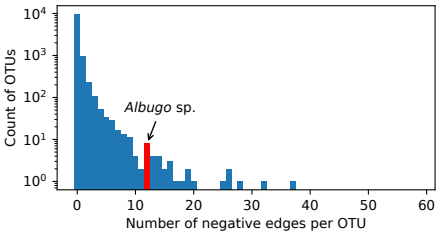


Figure S3. Histogram of the number of negative correlations as predicted by FlashWeave on the *Arabidopsis thaliana* phyllosphere amplicon dataset.
Highlighted is the location of the *Albugo* sp. operational taxonomic unit (OTU).

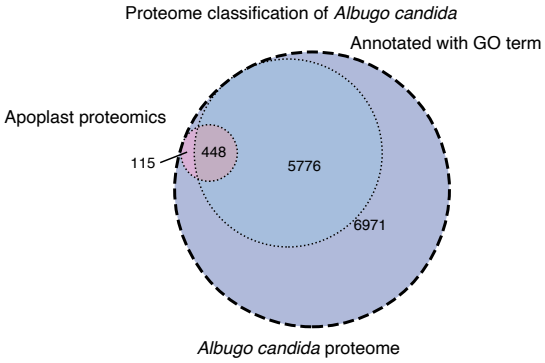


Figure S4. Venn diagram of annotated proteins in *Albugo candida*'s predicted proteome.
Proteins found in the apopolast are pink and proteins annotated with at least a gene ontology (GO) term are in blue.

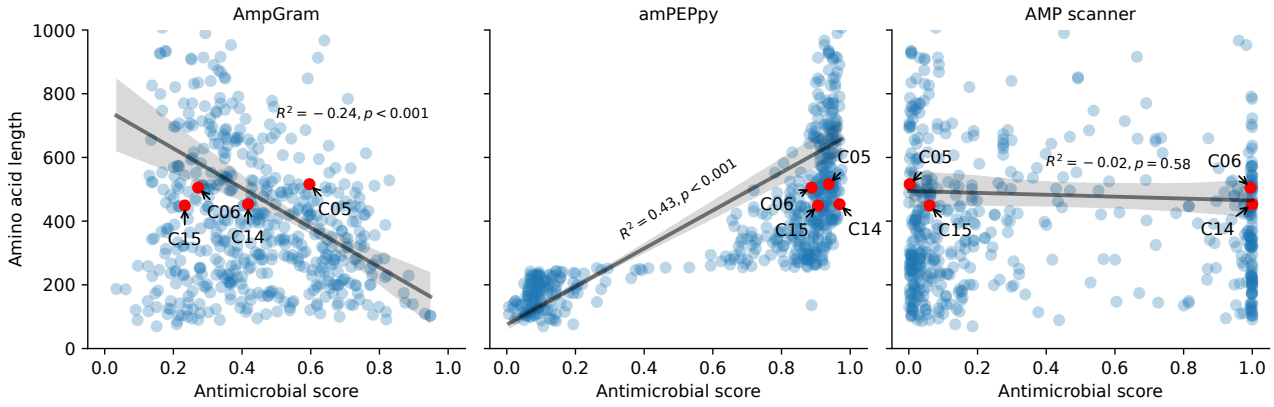


Figure S5. Scatter plot comparing the three machine learning methods for antimicrobial prediction to protein length in the proteins found in the apoplast for *Albugo candida*.
Highlighted are the selected candidate proteins. The y-axis is displayed up to a limit of 1000 amino acids. The black line represents the best least-squares fit for linear regression. Shaded is the 95% confidence interval for each regression. Depicted are the R-squared parameter and the p value for Wald test with t-distribution using as null hypothesis a zero-slope regression. Note the significant correlation of AmpGram and amPEPpy to protein length.

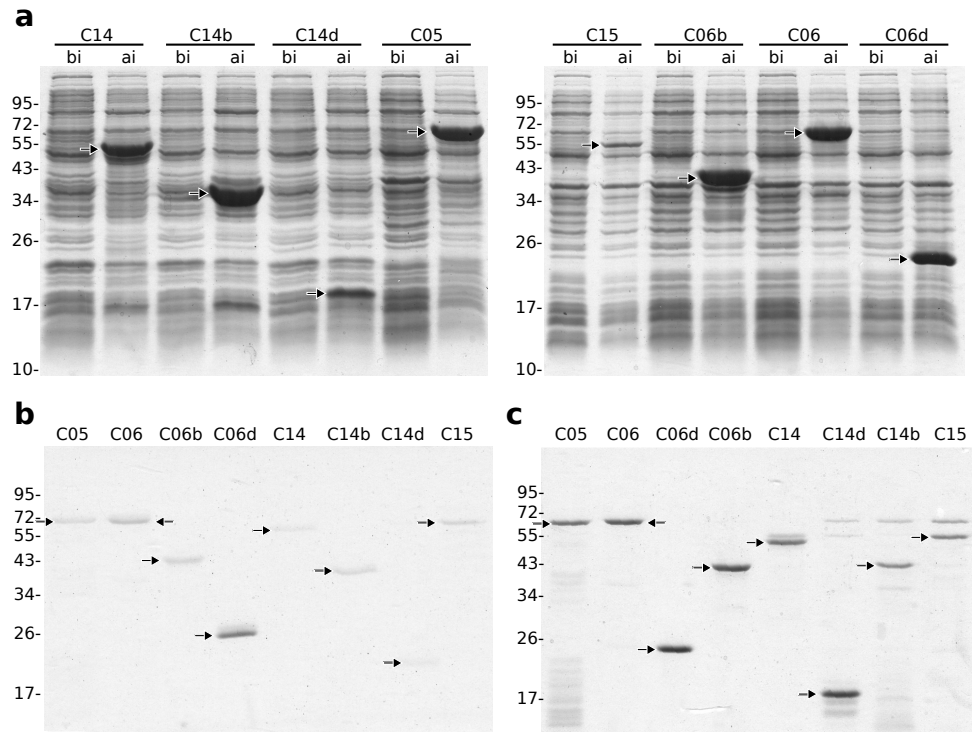


Figure S6. Expression and purification analysis of the protein candidates by SDS-PAGE and Coomassie blue staining. Scale on the left side indicates molecular weight in kDa. Black arrows highlight relevant bands. (**a**) Comparison of protein expression before (bi) and after (ai) IPTG induction of *Escherichia coli*. (**b**) Analysis of candidate protein elution fractions after purification in the denaturing buffer. (**c**) Analysis of protein candidates after concentration and rebuffering in the testing buffer.

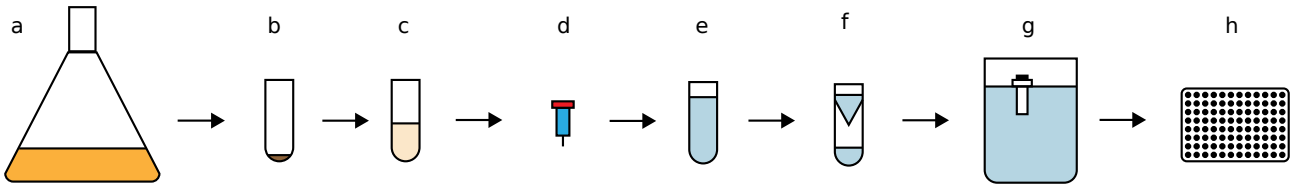


Figure S7. Workflow of the denaturing purification process and subsequent in vitro activity assays. (**a**) Overexpression of the candidate proteins in *Escherichia coli*. (**b**)Freezing of harvested pellets and storage at -80 ºC. (**c**) Resuspension in denaturing buffer and sonication to dissolve inclusion bodies. (**d**) Denaturing purification using a HisTrap™. (**e**) One-step elution using a low pH buffer. (**f**) Concentration of denatured protein in a VivaSpin® 20 column. (**g**) Dialysis of protein over two days in increasing concentrations of refolding buffer. (**h**) Antimicrobial assays on bacterial strains.

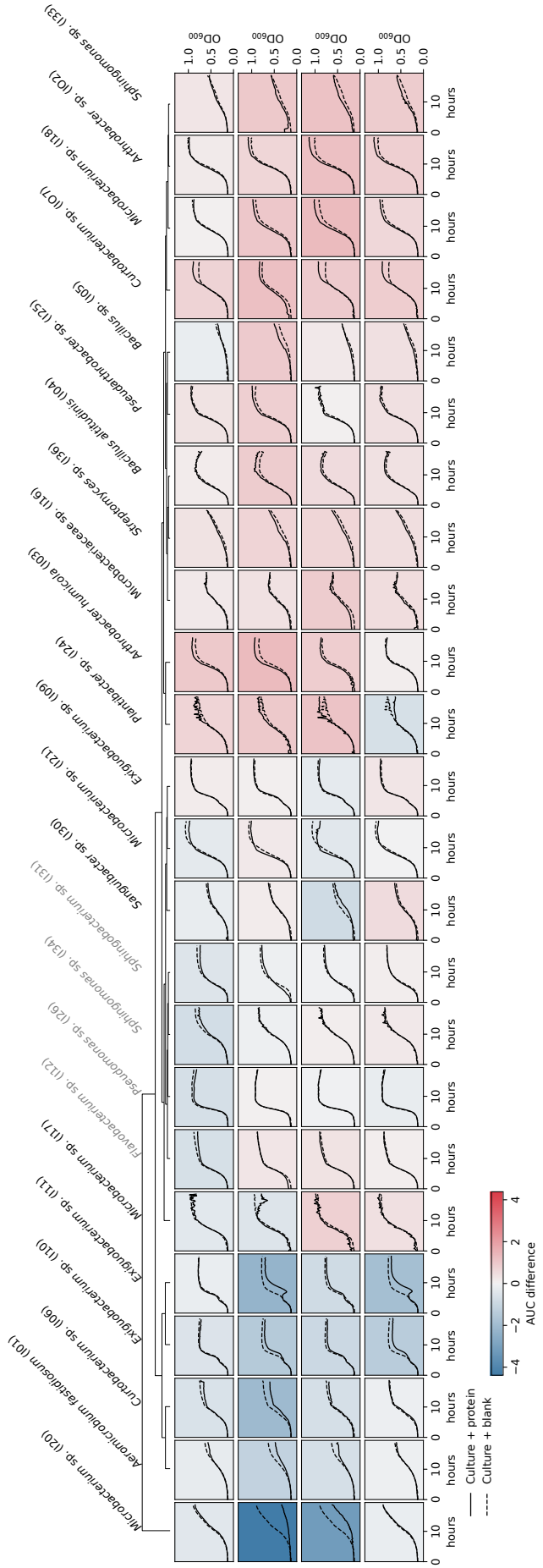


Figure S8. Growth curves for candidates C05, C06, C14 and C15 (from top to bottom row) at a concentration of 0.75 mM compared to blank (dashed lines). Background color represents inhibition (blue) or promotion (red) of growth based on the difference in the area under the curves of the means of three biological replicates. Names in gray represent gram-negative bacterial strains, while the rest are gram-positive.

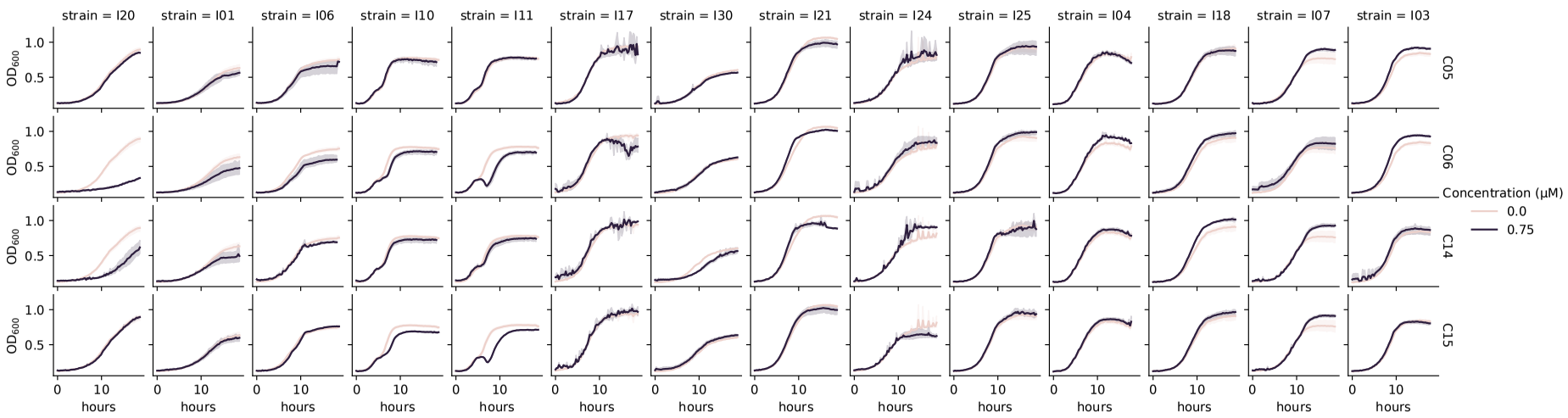


Figure S9. Growth curves for candidates C06, C06, C14 and C15 (from top to bottom row) in purple at a concentration per well of 0.75 µM compared to blank (beige). Background color represents inhibition (blue) or promotion (red) of growth based on the difference in the area under the curves of the means of three biological replicates. Confidence intervals of 95% from at least three biological replicates shown as the colored area around the mean.

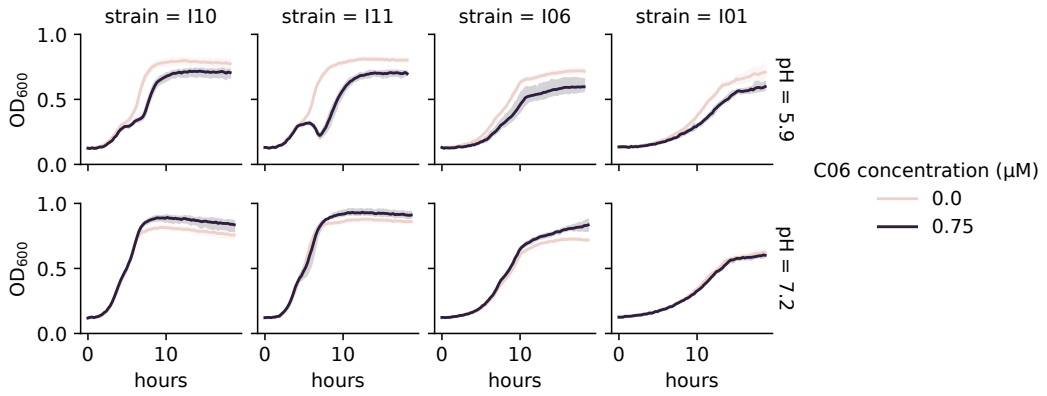


Figure S10. Comparison of the antimicrobial activity of C06 at pH 5.9 and 7.2 at molarity 0.75 µM.
Tested strains included *Exiguobacterium* I10 and I11, *Curtobacterium* I06 and *Aeromicrobium fastidiosum* I01. Confidence intervals of 95% from at least six biological replicates shown as the colored area around the mean. Blank treatments are represented as beige lines and protein treatments in purple.

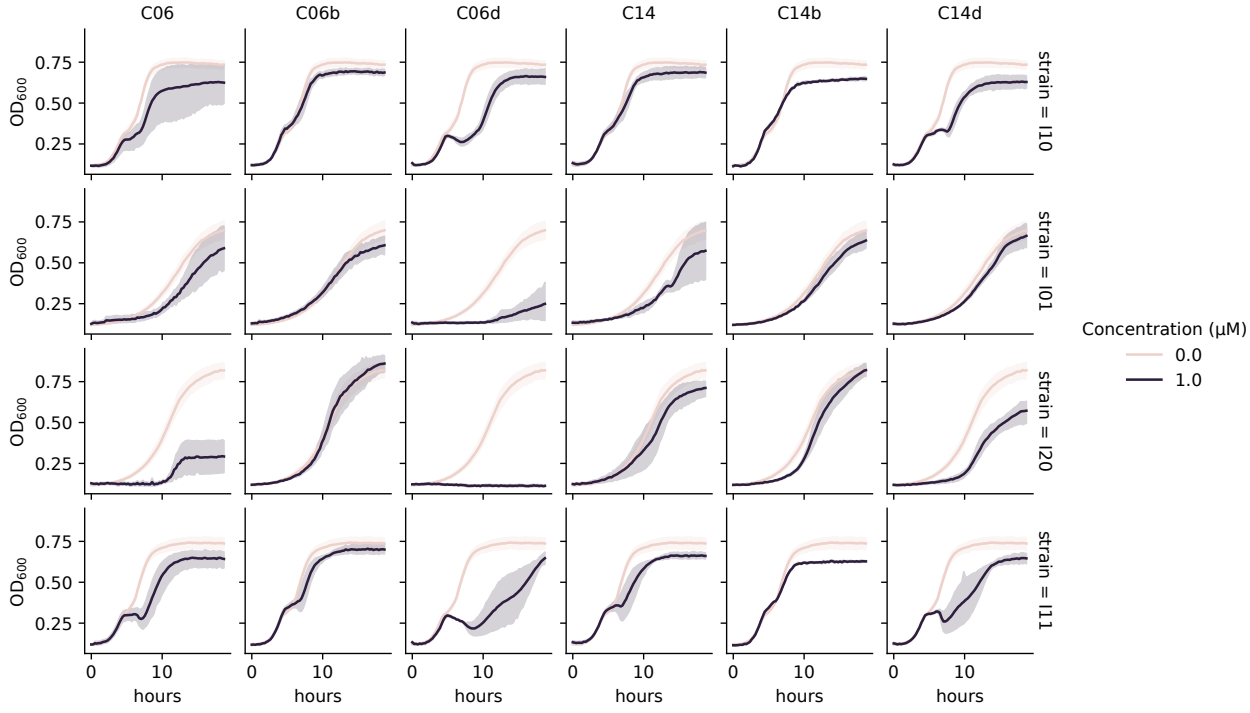


Figure S11. Inhibition curves of C06, C14 and their domains at a concentration of 1 µM.
The tested strains are represented in different rows as follows (from top to bottom): *Exiguobacterium* sp. I10, *Aeromicrobium fastidiosum* I01, *Microbacterium* sp. I20 and *Exiguobacterium* sp. I11. Confidence intervals of 95% from at least six biological replicates shown as the colored area around the mean. Blank treatments are represented as beige lines and protein treatments in purple.

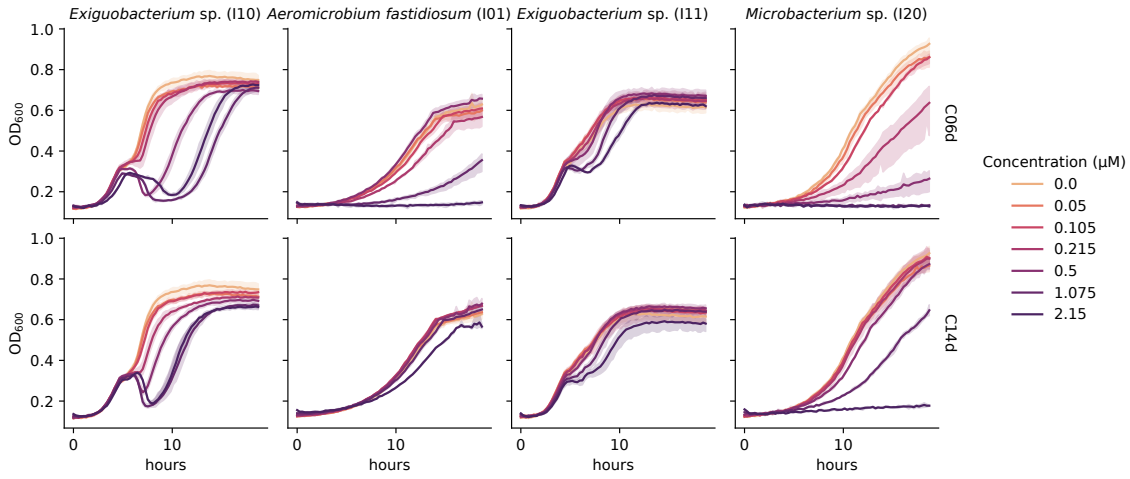


Figure S12. Inhibition curves of six different concentrations of C06d and C14d domains (first and second row, respectively) towards four sensitive strains.
Confidence intervals of 95% from three biological replicates shown as the area around the mean. The color gradient represents the concentration of proteins while beige is the blank.

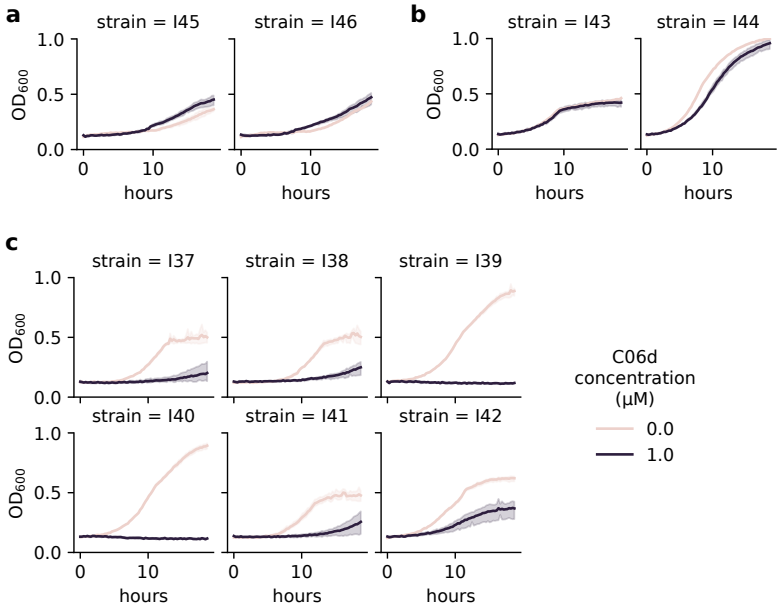


Figure S13. Inhibition curves of C06d on *Clavibacter* and *Rhodococcus* strains at a concentration of 1 µM (purple) and 0 µM (beige).
Confidence intervals of 95% from three biological replicates shown as the shaded area around the mean. (**a**) Growth curves of isolates from *Clavibacter michiganensis* subsp. *tessellarius*. (**b**) Growth curves of isolates from *Rhodococcus fascians*. (c) Growth curves of isolates from *C. michiganensis* subsp. *capsici*.

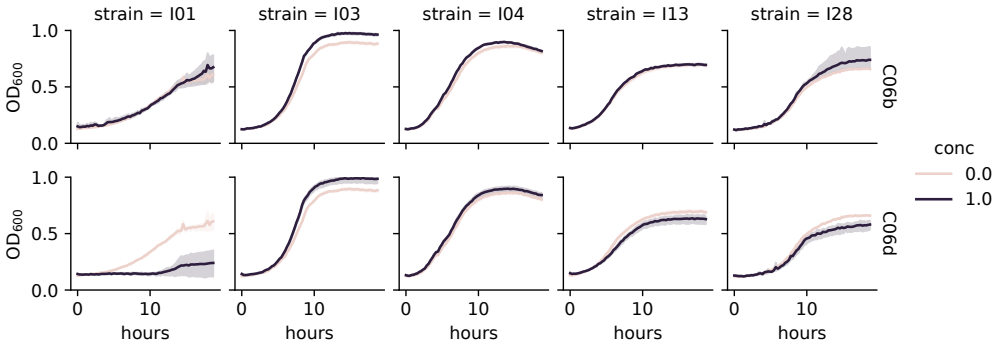


Figure S14. Growth curves from single-strain antimicrobial activity assays with the individual strains from the five-strain bacterial community and added C06b (first row) and C06d (second row) at a concentration of 1 µM (purple) and 0 µM (beige).

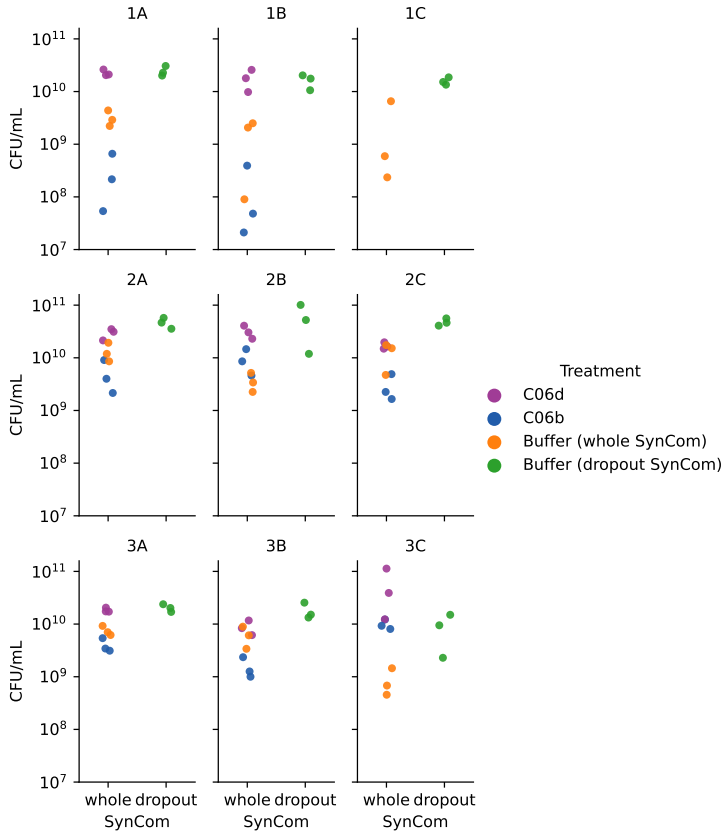


Figure S15. CFU counts per replicate of *Pseudomonas syringae* pv. Tomato DC3000 (Pst) in five-strain bacterial synthetic community (SynCom) experiments in liquid culture.
Effects of C06d (purple) and C06b (blue) candidates at a concentration of 1 µM on whole SynCom treatments are displayed. Effects of the buffer (BisTris, pH 5.9) on the whole community are shown in orange and on the dropout community in green. The x-axis displays community status ('whole' for whole five-strain community, 'dropout' for community without *Aeromicrobium fastidiosum* I01). Rows represent time points of replicates, and columns, different biological replicates for the same time point.

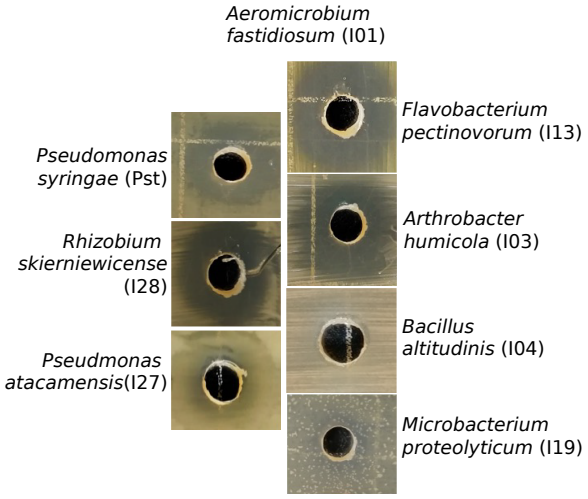


Figure S16. Pairwise inhibition rings of *Aeromicrobium fastidiosum* (I01) against other strains in the synthetic bacterial community.

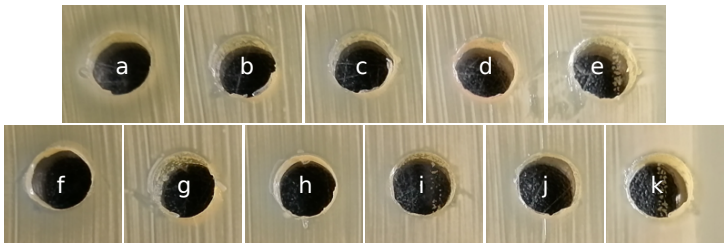


Figure S17. *Aeromicrobium fastidiosum* I01 lawn for interaction testing with bacterial strains in the synthetic bacterial community.
(**a**): *Pseudomonas atacamensis* (I27), (**b**): *Flavobacterium pectinovorum* (I13), (**c**): *Rhizobium skierniewicense* (I28), (**d**): *Microbacterium proteolyticum* (I19), (**e**): *Arhtrobacter humicola* (53), (**f**): *Bacillus altitudinis* (I04), (**g**): *Frigoribacterium faeni* (I14), (**h**): *Sphingomonas faeni* (I32), (**i**): *Nocardioides cavernae* (I22), (**j**): *Paenibacillus amylolyticus* (I23), (**k**): *Methylobacterium goesingense* (I15).

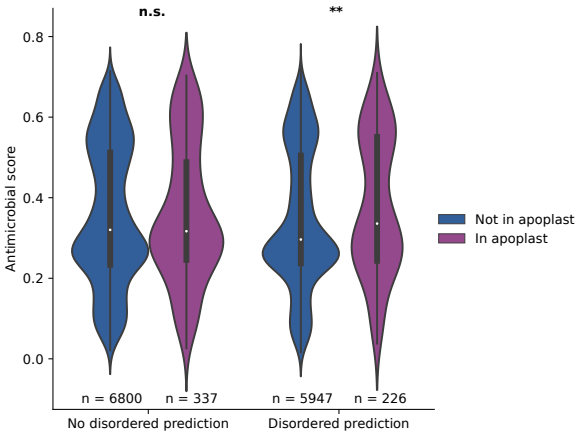


Figure S18. Antimicrobial prediction of apoplastic and non-apoplastic proteins grouped by the presence or absence of intrinsically disordered regions *Albugo candida*.
Significance was tested using Holm-corrected two-tailed Mann-Whitney U tests (p value $= 0.008$ for disordered prediction). *n.s.: not significant, **: p value < 0.01.*

Supplementary tables

Table S1. Apoplastic protein candidates from *Albugo candida* selected for antimicrobial testing. 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. *NCBI: National Center for Biotechnology Information, IP: InterPro, PTHR: Panther, aa: amino acid, IDR: intrinsically disordered region.*

Candidate	Accession (NCBI)	Functional annotation (IP/PTHR)	Antimicrobial weighted score	Secretion signal (aa)	Length (aa)	Abundance	Predicted IDRs
C05	CCI44519.1	Myo-inositol-1-phosphate synthase	0.43	No (0)	544	10.9	No
C06	CCI46028.1	None	0.68	No (0)	533	15.1	Yes
C14	CCI45607.1	Glucanosyltransferase	0.76	Yes (22)	440	15.5	Yes
C15	CCI42480.1	Glucan-1,3-beta-glucosidase	0.30	Yes (21)	391	36.5	Yes

Table S2. Net charge at pH 5.9 and 7.2, theoretical isoelectric point (pI) and molecular weight in kilodalton (kDa) of 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

Table S3. Strain collection of plant-isolated microbes from environmental sampling of Arabidopsis thaliana. It includes strains from the *A. thaliana* core synthetic community (SynCom); I03: GCF_022803015.1, I04: GCF_022803025.1, I13: GCF_022802975.1, I14: GCF_022805035.1, I15: GCF_022803065.1, I19: GCF_022803055.1, I27: GCF_022802935.1, I28: GCF_022802995.1, I29: GCA_022817825.1, I35: GCA_022817875.1. Taxonomic assignment was based on blastn matches on the 16S database from the National Center for Biotechnology (NCBI), as of February 2022. Taxonomy assignment of SynCom strains was based on their whole genome sequencing data. In parentheses, the closest matching species name is depicted. Fungal sequences were blasted against the non-redundant nucleotide database from NCBI. *Nc2: Albugo candida strain Nc2, Nc14: Albugo laibachii strain Nc14.*

Isolate	Closest matching species	Gram stain	Source
I01	*Aeromicrobium fastidiosum*	Gram(+)	Epiphytic isolate from *Arabidopsis thaliana*
I02	*Arthrobacter* sp. (*halodurans*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I03	*Arthrobacter humicola*	Gram(+)	Endophytic isolate from *A. thaliana*
I04	*Bacillus altitudinis*	Gram(+)	Endophytic/epiphytic isolate from *A. thaliana*
I05	*Bacillus* sp.(*bataviensis*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I06	*Curtobacterium* sp. (*ammoniigenes/herbarum*)	Gram(+)	Endophytic isolate from Nc2-infected *A. thaliana*
I07	*Curtobacterium* sp. (*herbarum*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I08	*Dioszegia hungarica*	Fungus	Endophytic isolate from *A. thaliana*
I09	*Exiguobacterium* sp. (*acetylicum*)	Gram(+)	Epiphytic isolate from Nc14-infected *A. thaliana*
I10	*Exiguobacterium* sp. (*indicum*)	Gram(+)	Epiphytic isolate from Nc14-infected *A. thaliana*
I11	*Exiguobacterium* sp. (*indicum*)	Gram(+)	Endophytic isolate from Nc14-infected *A. thaliana*
I12	*Flavobacterium* sp.	Gram(−)	Epiphytic isolate from Nc2-infected *A. thaliana*
I13	*Flavobacterium pectinovorum*	Gram(−)	Endophytic isolate from *A. thaliana*
I14	*Frigoribacterium faeni*	Gram(+)	Endophytic isolate from *A. thaliana*
I15	*Methylobacterium goesingense*	Gram(−)	Epiphytic/endophytic isolate from *A. thaliana*
I16	*Microbacteriaceae* sp.	Gram(+)	Epiphytic isolate from *A. thaliana*
I17	*Microbacterium* sp.	Gram(+)	Epiphytic isolate from *A. thaliana*
I18	*Microbacterium* sp.	Gram(+)	Epiphytic isolate from *A. thaliana*
I19	*Microbacterium proteolyticum* sp.	Gram(+)	Endophytic isolate from *A. thaliana*
I20	*Microbacterium* sp. (*saccharophilum*)	Gram(+)	Endophytic isolate from *A. thaliana*
I21	*Microbacterium* sp. (*saccharophilum*)	Gram(+)	Endophytic isolate from *A. thaliana*
I22	*Nocardioides cavernae*	Gram(+)	Endophytic isolate from *A. thaliana*
I23	*Paenibacillus amylolyticus*	Gram(+)	Endophytic isolate from *A. thaliana*
I24	*Plantibacter* sp. (*auratus/flavus*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I25	*Pseudarthrobacter* sp.(*polychromogenes*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I26	*Pseudomonas* sp. (*fluorescens*)	Gram(−)	Endophytic isolate from Nc2-infected *A. thaliana*
I27	*Pseudomonas atacamensis*	Gram(−)	Endophytic isolate from *A. thaliana*
I28	*Rhizobium skierniewicense*	Gram(−)	Endophytic isolate from *A. thaliana*
I29	*Rhodotorula kratochvilovae*	Fungus	Epiphytic/endophytic isolate from *A. thaliana*
I30	*Sanguibacter* sp. (*keddieii*)	Gram(+)	Epiphytic isolate from *A. thaliana*
I31	*Sphingobacterium* sp.	Gram(−)	Epiphytic isolate from Nc14-infected *A. thaliana*
I32	*Sphingomonas faeni*	Gram(−)	Epiphytic/endophytic isolate from *A. thaliana*
I33	*Sphingomonas* sp. (*faeni*)	Gram(−)	Epiphytic isolate from *A. thaliana*
I34	*Sphingomonas* sp. (*faeni*)	Gram(−)	Epiphytic isolate from *A. thaliana*
I35	*Sporobolomyces roseus* sp. (*faeni*)	Fungus	Epiphytic/endophytic isolate from *A. thaliana*
I36	*Streptomyces* sp.(*subrutilus*)	Gram(+)	Epiphytic isolate from *A. thaliana*

Table S4. Primer sequences for amplification and subsequent cloning of candidate sequences from *Albugo candida* into vector pET28b.

Candidate	Forward primer 5' to 3'	Reverse primer 5' to 3'	6x His-tag
C05	CGCGCGGCAGCCATATGACGACA GCATCGTTCC	GGTGGTGGTGCTCGAAATGGTGGCTT TCAATTTCACTCGC	N- and C-terminal
C06	CGCGCGGCAGCCATATGTCTGAT CGACAAGAACAAGTGA	GGTGGTGGTGCTCGAAATTGGAAGTG TCTTTGTGACGAGA	N- and C-terminal
C06d	CGCGCGGCAGCCATATGACGGA AGGTGAATCAACTTCTCA	GGTGGTGGTGCTCGAAATTGGAAGTG TCTTTGTGACGAGA	N- and C-terminal
C06b	CGCGCGGCAGCCATATGTCTGA TCGACAAGAACAAGTGA	GGTGGTGGTGCTCGAACACTCCTGCA TAATCCATTGCG	N- and C-terminal
C14	AGGAGATATACCATGAGCCCGA TTACAATTCAAGGC	GGTGGTGGTGCTCGACGCAATTTGACT TTTTGTTGGG	C-terminal
C14d	AGGAGATATACCATGTCACAATC GACATCCAATTTCCCATTCACA	GGTGGTGGTGCTCGACGCAATTTGACT TTTTGTTGGG	C-terminal
C14b	AGGAGATATACCATGAGCCCGAT TACAATTCAAGGC	GGTGGTGGTGCTCGACATTTGCTTTCTC CGTAGAGAACTC	C-terminal
C15	AGGAGATATACCATGTCGGCTCA ATCACCAATGG	GGTGGTGGTGCTCGACAAAAACTCCTT TTTCTTCATTTTC	C-terminal

Table S5. Significant compositional bias of tested protein candidates. Significance given as the negative logarithm of the binomial p values.

Candidates	Residues	Significance
C05	N	7.0
C06	A, Q	15.8, 12.4
C06b	Q, A	11.7, 9.3
C06d	H, A	13.5, 9.0
C14	S, C	10.8, 7.2
C14d	S	14.8
C15	S	10.3

Table S6. Peptide coverage of the N-terminal disordered domains from candidate proteins in the proteomics analysis. Underlined are residues predicted to be disordered.

Protein domain	Peptide
C06d	STSQQAQQAAGTLR
	NAADYTSDRTHEAADVTR
	AVHEQTEPSKPGFGEK
	AQSAVHEQTEPSKPGFGEK
C14d	SPSLSSVSWNGVEQVK

Table S7. Proteases from *Arabidopsis thaliana* enriched in the *Albugo candida*-infected treatment.

Gene	Description	Intensity ratio
SBT3.5	Subtilisin-like protease SBT3.5	3.99
SBT3.3	Subtilisin-like protease SBT3.3	3.34
SBT4.14	Subtilisin-like protease SBT4.14	1.80
LON2	Lon protease homolog 2, peroxisomal	1.43
SBT1.1	Subtilisin-like protease SBT1.1	1.23
SBT1.6	Subtilisin-like protease SBT1.6	1.21

Table S8. Plant-isolated putative pathogenic gram-positive microbes from environmental sampling of Arabidopsis thaliana employed in the testing of *Albugo candida* peptide C06d. Taxonomic assignment based on results of blastn on the 16S database from the National Center for Biotechnology, as of March 2022.

Isolates	Closest matching species	Pathogenic phenotype	Reference
I37,I38,I39, I40,I41,I42	*Clavibacter michiganensis* ssp. *capsici*	Pepper pathogen causing bacterial canker disease	Type strain PF008 (Oh et al., 2016)
I43,I44	*Clavibacter michiganensis* ssp. *tessellarius*	Wheat pathogen causing leaf freckles and spots	Strain 78181/ DSM 20741/ATCC 33566 (Carlson and Vidaver, 1982; Li and Yuan, 2017)
I45,I46	*Rhodococcus fascians*	Broad host range pathogen causing leafy gall disease	Type strain DSM 20669/CF17 (Dhaouadi et al., 2020; Hjerde et al., 2013)

B.2 Cell speciation in cyanobacterial biofilm development

Full title:

- "Cell specialization in cyanobacterial biofilm development revealed by expression of a cell-surface and extracellular matrix protein"

Contributions by DGP:

- Performed the experiments related to the amyloid part, except the electron microscopy imaging, which was performed by YDS.
- Wrote the parts of the draft related to amyloids.
- Edited the final manuscript.
- Created all the figures and tables related to the amyloid part (Figures 8, 9 and S3, and bottom third part of Table S1), except Figure 9 which was captured by YDS.

Cell specialization in cyanobacterial biofilm development revealed by expression of a cell-surface and extracellular matrix protein

Alona Frenkel[1#], Eli Zecharia[1#], Daniel Gómez-Pérez[2], Eleonora Sendersky[1], Yevgeni Yegorov[1], Avi Jacobs[1], Jennifer Benichou[1], York-Dieter Stierhof[2] Rami Parnasa[1], Susan S Golden[3,4], Eric Kemen[2] and Rakefet Schwarz[1*]

[1]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel, 5290002.

[2]Center for Plant Molecular Biology (ZMBP), University of Tübingen, 72074 Tübingen, Germany

[3]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

[4]Center for Circadian Biology, University of California, San Diego, La Jolla, CA 92093, USA

[#]These authors contributed equally to the study

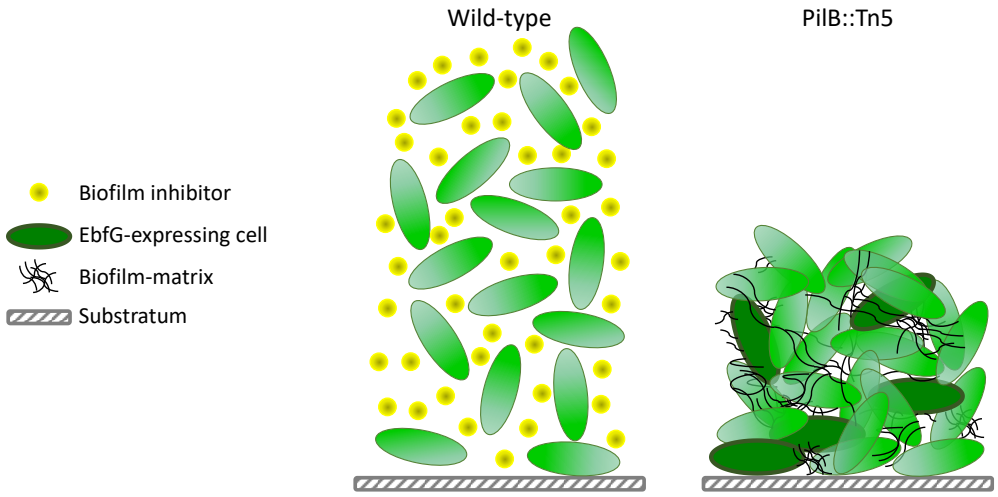
*Corresponding author

Running Title: Cell specialization in cyanobacterial biofilm development

Abstract

Cyanobacterial biofilms are ubiquitous and play important roles in diverse environments, yet, understanding of the processes underlying development of these aggregates is just emerging. Here we report cell specialization in formation of *Synechococcus elongatus* PCC 7942 biofilms - a hitherto unknown characteristic of cyanobacterial multicellularity. We show that only a quarter of the cell population expresses at high levels the four-gene *ebfG*-operon that is required for biofilm formation. Almost all cells, however, are assembled in the biofilm. Detailed characterization of EbfG4 encoded by this operon revealed cell-surface localization as well as its presence in the biofilm matrix. Moreover, EbfG1-3 were shown to form amyloid structures such as fibrils and are thus likely to contribute to the matrix structure. These data suggest a beneficial 'division of labour' during biofilm formation where only some of the cells allocate resources to produce matrix proteins – 'public goods' that support robust biofilm development by the majority of the cells. Additionally, previous studies revealed the operation of a self-suppression mechanism that depends on an extracellular inhibitor, which supresses transcription of the *ebfG*-operon. Here we revealed inhibitor activity at an early growth stage and its gradual accumulation along the exponential growth phase in correlation with cell density. Data, however, do not support a threshold-like phenomenon known for quorum-sensing in heterotrophs. Together, data presented here demonstrate cell specialization and imply density-dependent regulation thereby providing novel insights into cyanobacterial communal behaviour.

42 **Graphical Abstract**



43

44

Introduction

Cyanobacteria are highly abundant in the environment and are responsible for ~25% of the global primary production [1, 2]. Frequently, these photosynthetic prokaryotes are found in microbial assemblages known as biofilms or part of laminated biofilms, dubbed microbial mats [3-5]. Phototrophic biofilms are often associated with industrial problems [6-8]; in contrast, such microbial consortia are beneficial e.g., for effective biomass accumulation for the biofuel industry and for harvesting of secondary metabolites [9-12]. In-depth understanding of cyanobacterial biofilm development paves the way for inhibition of deleterious biofilms and promotion of beneficial ones.

The mechanisms involved in cyanobacterial aggregation or biofilm formation started emerging only in recent years. For example, similarly to heterotrophic bacteria, cyanobacteria use the second messenger cyclic-di-GMP for regulating aggregated vs planktonic mode of growth [13]. Furthermore, the thermophilic cyanobacterium *Thermosynechococcus vulcanus* employs cyanobacteriochrome photoreceptors to mediate light-colour input for controlling cell aggregation via c-di-GMP signaling [14-16].

Microbial cells within biofilms are encased in a self-produced matrix of hydrated extracellular polymeric substances (EPS) that allows multilayering of cells and structural stability and provides a protected environment. Numerous studies in diverse heterotrophic bacteria identified particular sugar polymers and protein filaments as matrix components [17, 18], however, little is known about the cyanobacterial biofilm matrix. Yet, extracellular polysaccharides were implicated in cyanobacterial biofilm formation, for example, studies of *Synechocystis* support involvement of extracellular polysaccharides in surface adhesion [19] and cell sedimentation [20] and cellulose accumulation is responsible for cell aggregation in *T. vulcanus* RKN [21]. The exo-protein HesF of *Anabaena* sp. PCC 7120 is required for aggregation and it was proposed that it interacts with polysaccharides [22], however, its detailed role in aggregation is still unknown.

Our previous studies revealed a biofilm self-suppression mechanism in *S. elongatus* that dictates planktonic growth of this strain (Fig. 1). Inactivation of gene Synpcc7942_2071, abrogates the biofilm inhibitory process and results in a biofilm-proficient strain in contrast to the planktonic nature of WT. This gene encodes a

homologue of ATPases of type 2 secretion (T2S) systems or type 4 pilus (T4P) assembly complexes, thus, the mutant was initially designated T2EΩ but recently renamed PilB::Tn5 [23-25]. The RNA chaperone Hfq and a conserved cyanobacterial protein (EbsA), which are part of the T4P complex, are also essential for the biofilm suppression mechanism [25]. Additionally, we identified four small proteins, each with a double glycine secretion motif, that enable biofilm formation (enable biofilm formation with a GG motif EbfG1-4).

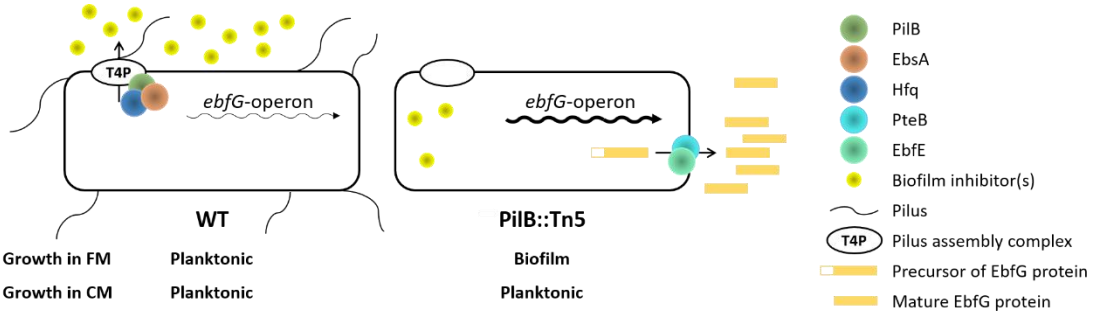


Figure 1: Biofilm regulation in *S. elongatus* by an extracellular inhibitor that dictates transcription of the *ebfG*-operon.

The type 4 pilus (T4P) assembly complex is involved in deposition to the extracellular milieu of biofilm inhibitor(s), which dictate transcription of the *ebfG*-operon. Low and high abundance of transcripts of this operon are indicated by thin and thick arrows, respectively. FM – fresh medium; CM – conditioned medium.

The T4P complex is involved in deposition to the extracellular milieu of a biofilm inhibitor that regulates expression of the *ebfG*-operon [23, 26]. Mass-spectrometry (MS) analyses revealed the presence of EbfG1-4 in conditioned medium (CM) of PilB::Tn5 [26, 27]. Furthermore, we demonstrated that proteins PteB (peptidase transporter essential for biofilm formation), which belongs to the C39 peptidases family, and EbfE (enable biofilm formation enzyme), a homolog of microcin processing peptidases, are implicated in secretion of EbfG1-4 to the extracellular milieu [26, 28]. The role of EbfG1-4 in biofilm formation, however, was unknown. Here, using a reporter construct we demonstrate that high expression of the *ebfG*-operon is limited to a small subpopulation of cells of PilB::Tn5. Further characterization indicates cell-surface and biofilm-matrix localization of EbfG4 and strongly supports amyloid nature of EbfG2. Together, the data indicate cell specialization and imply microbial cooperation for production of extracellular components beneficial for the whole population, known as “public-goods”. Additionally, the response of the reporter strain to conditioned media harvested at different stages of logarithmic growth of the wild-

type strain implies a density dependent mechanism in regulation of *S. elongatus* biofilm development.

Materials Methods

Strains and culture conditions and harvesting of conditioned medium (CM): Cultures of *S. elongatus* PCC 7942 and all derived strains were grown in Pyrex tubes under bubbling with air enriched with CO_2 as described previously [29, 30]. Construction of mutants and details of molecular manipulations are provided in Table S1. For harvesting of CM, wild-type cultures were initiated from liquid starters at OD_{750}=0.2. CM harvesting was performed as described [23].

Flow cytometry: 50 ml culture at exponential phase was centrifuged (6000g, room temperature), resuspended with 4 ml fresh BG11 to obtain a concentrated culture for inoculation into fresh medium or CM at an OD_{750} of 0.5. Aliquots of 0.5 ml were taken from each culture tube following 6 days of growth and then, in case of biofilm-forming strains, planktonic cells were removed. 1.5 ml BG11 were used to resuspend the biofilmed cells by rigorous pipettation and 0.13 ml were transferred to a 1.5 ml Eppendorf tube for homogenization with a pellet pestle (Sigma-Aldrich, Z359971-1EA). The homogenized samples were filtered through a mesh (pore size 52 µm), supplemented with formaldehyde to a final concentration of 1%, diluted with PBS to OD_{750} of ~0.0001 and measured using BD *FACSAria* (excitation 488nm, emission 530 ±30nm).

All statistical analyses were conducted in the statistical program R, version 3.3.2 [31]. FCS files obtained from *FlowJo* were analyzed with *flowcore* package [32]. Mean, median, and robust coefficient of variation (CV) of the intensity distribution for each sample were calculated. Robust CV was calculated as defined in the *FlowJo* documentation [33]. Intensity values were log-transformed. Significant difference between biofilm and planktonic cells of a particular culture was tested using Paired t-tests on several intensity distribution parameters (mean, median and robust CV). Initial analysis did not reveal significant differences between biofilm and planktonic cells within a particular culture, therefore, these data were combined for further analysis. Effect of growth medium or genetic background on intensity distribution parameters (mean, median and robust CV) was tested with 2-way repeated measures ANOVA. Specifically, mixed linear effect models were fitted with medium or genetic background

as fixed effects and biological replicates as random effect, (using *lmerTest* package [34]), and the ANOVA was performed on the resulting models. Post hoc pairwise comparisons were performed by testing linear contrasts (using emmeans R package [35]), and FDR correction was applied to control for multiple testing. Normality of residuals homogeneity of variances assumptions were checked graphically.

Dot-blot analysis: Cell extracts from 6 days old cultures were prepared as described previously [25] and 2.3 µl from diluted extracts were spotted onto TransBlot Turbo nitrocellulose membrane (Bio-Rad) and air dried for 5 min. All following procedures were performed at room temperature: Blocking was done for 1 h in 0.1% bovine serum albumin in TBST (20 mM Tris-HCl (pH 8.0) and 0.05% Tween20). Incubation with anti-FLAG (ab1162, Abcam; 1:2000 diluted in blocking solution) was performed for 1h following three washes in TBST for 5 min each. Incubation with secondary antibodies (goat anti rabbit IgG, 170-6515, Bio-Rad; 1:5000 diluted in blocking solution) was done for 1h following washes as above with extension of last wash to 15 min and signal detection using SuperSignal West Pico kit (Thermo Scientific, 34080).

Fluorescence microscopy: Cultures were initiated and grown as described for biofilm quantification. Cultures (30 ml) were centrifuged (5 min, 6000 g, room temp) and resuspended in 1 ml phosphate-buffered saline (PBS). In case of biofilm-forming strains, the planktonic cells were removed with a pipette and the biofilmed cells were gently resuspended using 1 ml PBS. The concentrated cultures were precipitated by centrifugation in Eppendorf tubes as above, resuspended in 1 ml PBS and formaldehyde, from 16% stock solution prepared as described in Cold Spring Harbor Protocols (http://cshprotocols.cshlp.org/content/2010/1/pdb.rec12102.full), was added to a final concentration of 2%. Cells were incubated in the dark (30 min at room temperature in a tube rotator followed by 30 min incubation on ice agitation), washed once in PBS, resuspended in 1 ml PBS and the mixture was equally divided into two Eppendorf tubes. These tubes were centrifuged - cells for imaging without permeabilization were resuspended in 1 ml PBS and saved in the dark at room temperature. For imaging following permeabilization, cells were resuspended in 500 µl 0.1% triton in PBS, incubated at room temperature in a tube-rotator for 15 min and centrifuged. Cell pellet was resuspended with lysozyme solution (0.2 mg/ml dissolved in 50 mM Tris-HCl, pH 7.5 and 10 mM EDTA) and incubated for 30 min at 37°C. Cells were washed twice with 1 ml PBS. An aliquot of 200 µl was treated with an equal

volume of freshly prepared blocking solution (5% BSA in PBS) in a tube-rotator for 1 h at room temperature. Cells were pelleted, resuspended in 100 µl anti-FLAG antibody (Abcam, 1:400 diluted with blocking solution), incubated for 40 min at room temperature and then 40 min at 30°C. Cells were washed twice with 100-200 µl PBS buffer, resuspended in 20 µl secondary antibody (Alexa Fluor® 488 Abcam) diluted 1:100 in blocking solution and incubated for 1h at 30°C. Pellet was washed once with 50 µl PBS and resuspended in 20 µl PBS. 3-5 µl were spread on microscopy slides prepared as follows. 10 µl of L-polylysine (Sigma) diluted 1:10 was spread on a microscope slide (approximately on a 1 cm x 1 cm region). Slides were air dried, washed by dipping them twice in double distilled water and air dried. Cells were layered on the coated area, air dried and slides were centrifuged in 50 ml falcon tubes to attach cells to the polylysine layer (300g 10 min, room temperature). 3 µl antifade [36] was spotted and covered with a coverslip. Images were recruited using Leica SP8 confocal microscope (autofluorescence: excitation - 630 nm; emission - 641 to 657 nm and detection of Alexa Fluor® 488: excitation - 488 nm; emission - 495 to 515 nm). ***Amyloid analysis:*** We employed the TANGO algorithm and the machine learning programs APPNN and AmyloGram for in silico amyloid prediction over the mature peptide sequence of EbfG1-4 [37-39]. The pipeline can be found at https://github.com/danielzmbp/amypred. After max-min normalization of the scores between 0 and 1, the cutoff for amyloid prediction was set at 0.5. For the annotation of amyloidogenic hotspots, we employed software with a diverse predictive background, including statistical sequence analysis (WALTZ), structural information analysis (ArchCandy and Pasta 2.0), machine-learning-based (APPNN and PATH), and metamyl, a consensus predictor [39-44]. We predicted the cross-beta three-dimensional structure from the amyloid peptide domains using Cordax and visualized it using ChimeraX [45, 46].

We used the Curli-Dependent Amyloid Generator (C-DAG) system to study amyloid formation *in vivo* [47, 48]. This system uses the built-in curli processing system from *Escherichia coli* to express recombinant proteins in order to test for their amyloid aggregation. Positive and negative controls for amyloid formation employed, the *Saccharomyces cerevisiae* prion Sup35 with aggregating domain (Sup35[NM]) and without (Sup35[M]), were encoded by pVS72 and pVS105 plasmids, respectively. EbfG proteins equipped with a CsgA secretion signal in place of the native secretion

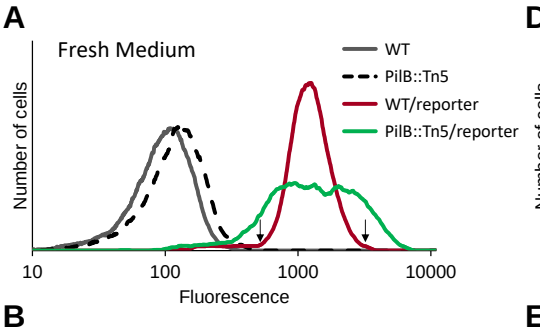
signal and fused to a 6x Histidine tag at the C-terminus were separately cloned in pExport plasmids and expressed in the *E. coli* strain VS45 (Table S1). For colony color phenotype analysis in inducing Congo Red plates (LB agar, 100 mg/l carbenicillin, 25 mg/l chloramphenicol, 0.2% w/v L-arabinose, 1mM IPTG and 10mg/l Congo Red), colonies were grown for four days at 22ºC in the dark.

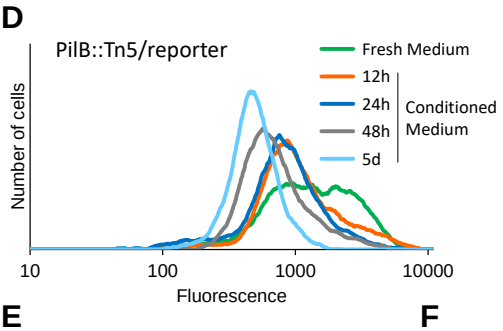
In order to perform transmission electron microscopy (TEM) on the samples, we deposited cells from the inducing Congo Red plates grown for four days on copper mesh grids. After drying, we incubated with anti-Histidine primary antibody for 1 h (mouse, Sigma Aldrich) followed by incubation with secondary anti-mouse IgG conjugated to 6 nm gold beads for 1 h (goat, dianova) for the immunostained samples. All samples were negatively stained for 30 seconds in aqueous uranyl acetate, before visualization in a JSM 1400 plus transmission electron microscope (JEOL).

Results

ebfG-operon expression in individual cells

Previous quantitative RT-PCR analyses indicated that the *ebfG*-operon is highly transcribed in PilB::Tn5 compared to WT [26]. These data reflect the averaged transcription level, thus, to gain insight into variation within the population, here we employ a reporter construct to follow expression of this operon in individual cells. To this end, a DNA fragment bearing the putative *ebfG*-operon promoter along with the 5' untranslated region was attached to a yellow fluorescence protein (*yfp*) gene, yielding the construct P-*ebfG*::YFP (Table S1). This fusion product was inserted in a neutral site 1 [49] in WT and PilB::Tn5 cells yielding WT/reporter and PilB::Tn5/reporter strains, respectively, which were analysed by flow cytometry (Fig. 2A&D).





B

Strains compared		Adjusted P value		
		Mean	Median	Robust CV
WT/reporter	PilB::Tn5/reporter	0.6624	0.364	<0.0001
	PilB::Tn5	<0.0001	<0.0001	<0.0001
	WT	<0.0001	<0.0001	<0.0001
PilB::Tn5/reporter	PilB::Tn5	<0.0001	<0.0001	0.0314
	WT	<0.0001	<0.0001	0.0001
PilB::Tn5	WT	0.4301	0.3896	0.0862

C

% cells of PilB::Tn5	Aver	Stdev	
Lower than WT	10.5	1.2	≤0.005
Higher than WT	24.2	6.6	≤0.0001

E

	Growth Medium compared		Adjusted P value		
			Mean	Median	Robust CV
Section I	Fresh Medium	12h-CM	0.3521	0.2015	0.0022
		24h-CM	0.0009	0.001	<0.0001
		48h-CM	<0.0001	<0.0001	<0.0001
		5d-CM	<0.0001	<0.0001	<0.0001
Section II	12h-CM	24h-CM	0.0289	0.0755	0.0214
		48h-CM	0.0009	0.0032	0.0001
		5d-CM	0.0013	0.0041	<0.0001
Section III	24h-CM	48h-CM	0.3492	0.3592	0.1095
		5d-CM	0.4108	0.3896	0.0591
	48h-CM	5d-CM	0.9126	0.9564	0.8748

F

	BF formation		
FM	√	√	√
12h-CM	√	√	√
24h-CM	√	√	–
48h-CM	–	–	–
5d-CM	–	–	–

Figure 2: Analysis of expression of the *ebfG*-operon by flow cytometry using reporter strains.

A and B. Number of cells as a function of fluorescence in cultures grown in fresh medium (FM) for 6 days (**A**). Strains analyzed: WT, PilB::Tn5 and their cognate reporter strains that bear a fusion of the regulatory region of the *ebfG*-operon with a yellow fluorescence protein (YFP). Arrows indicate fluorescence cutoff for calculating mutant cells with lower or higher expression of *ebfG*-operon compared to WT and summary of statistical analyses of reporter expression in FM (**B**). **C.** Fraction of PilB::Tn5/reporter cells with lower or higher expression of *ebfG*-operon compared to WT/reporter. Shown are averages and standard deviations from three independent experiments. **D and E.** Number of cells as a function of fluorescence (**D**) and summary of statistical analyses (**E**) of reporter expression in PilB::Tn5/reporter cells grown in FM and conditioned medium (CM). **B and E** show adjusted p-values of the mean, median and robust coefficient variation (rCV) from three independent experiments. **F.** Biofilm (BF) formation by PilB::Tn5/reporter cells grown in FM or CM harvested at different time points of WT culture.

Mean and median values of YFP fluorescence level is different between reporter strains and strains lacking the reporter construct, whereas these parameters are similar in WT/reporter and PilB::Tn5/reporter strains grown in fresh medium (Fig. 2B). In contrast, data dispersion is larger in PilB::Tn5/reporter compared to WT/reporter (Fig. 2A), a feature that is manifested in the significantly different robust coefficient variation (rCV, Fig. 2B). Moreover, data analysis revealed that on average, ~10% of PilB::Tn5/reporter cells are characterized by lower - and ~24% by higher - expression of the *ebfG*-operon (Fig. 2C). This observation suggests cell specialization in *S. elongatus* biofilm development.

Our previous studies demonstrated higher transcription of the *ebfG*-operon in PilB::Tn5 compared to WT when strains are cultured in fresh medium. Inoculation, however, of the mutant into conditioned medium (CM) from WT culture, strongly suppresses transcription [23, 26]. These data suggest involvement of intercellular communication in *S. elongatus* biofilm development possibly via a density dependent mechanism. To monitor the presence of the biofilm inhibitor along culture growth, CM was harvested from WT cultures grown for 12h, 24h, 48h and 5 days (Fig. S1) and the impact on YFP expression by PilB::Tn5/reporter was assessed. Note, conditioned media were supplemented with nutrients to negate possible nutrient limitation.

Individual flow cytometry experiments demonstrated a decrease in fluorescence intensity with CM age (e.g. Fig. 2D) in accordance with accumulation of the biofilm inhibitor during culture growth. Statistical analysis of three independent experiments indicated significant difference between rCV of cells grown in fresh medium and those grown in CM harvested at 12 h following culture initiation (12h-CM; Fig. 2E see section I of table). Additionally, mean, median and rCV were significantly different between cells grown in fresh medium and those inoculated into CM harvested at 24h, 48h and 5 days (24h-CM, 48h-CM and 5d-CM; Fig. 2E, section I of table). Moreover, mean and rCV were significantly different between cells exposed to 12h-CM and those inoculated into either 24h-CM, 48h-CM or 5d-CM (Fig. 2E, section II of table). Comparisons of the impact of CM from older cultures (Fig. 2E, section III; 24h-CM vs. 48h-CM and 5d-CM and 48h-CM vs. 5d-CM) did not reveal significant changes. Because individual experiments demonstrate accumulation of biofilm inhibitor in CM with growth time (e.g. Fig. 2D), we propose that the inhibitor is accumulated with culture age, which corresponds with culture density. Lack of significance, however, between data summarizing three biological repeats at 24h or longer culture growth, indicates variable kinetics of inhibitor accumulation in independent experiments.

Interestingly, 12h-CM had a significant impact on fluorescence rCV (Fig. 2E) and apparently, less cells expressed the *ebfG*-operon at high levels compared with fresh medium (Fig. 2D). Yet, biofilm development by these cultures (Fig. 2F) suggests that the small fraction of *ebfG*-operon highly expressing cells is sufficient to drive biofilm formation. 24h-CM significantly affected the mean, median and rCV compared with fresh medium (Fig. 2E, section I), however, in two of the three biological repeats biofilms were formed (Fig. 2F), in agreement with suggested variability in kinetics of accumulation of the biofilm inhibitor between individual experiments. 48h-CM and 5d-

CM consistently inhibited biofilm formation in accordance with substantial repression of *ebfG*-operon expression under these conditions (Fig. 2E section I and 2F). Together, data indicate presence of the inhibitor at early culture stages upon initiation of the logarithmic growth (Fig. S1 and Fig. 2D and E), and suggest further accumulation with time and culture density.

Localization of EbfG4

EbfG proteins do not share primary sequence similarity or domains with proteins of known function. To get insight into the role of these proteins in biofilm development, we selected EbfG4 for further characterization. This particular EbfG protein was chosen because a mutational approach impairing the secretion motif of either one of the EbfG proteins revealed that the mutation in EbfG4 had the most prominent impact on biofilm development compared with EbfG1-3 [26]. To follow EbfG4 localization in biofilm-forming and planktonic strains we introduced a FLAG-epitope tagged EbfG4 (EbfG4::FLAG) to the double mutant having inactivations in both *pilB* and Synpcc7942_1134 (PilB::Tn5/EbfG4Ω). The double mutant grows planktonically – similarly to WT, 100% of the chlorophyll is in planktonic cells as assessed by measurement of the relative amount of chlorophyll in suspension of total chlorophyll in the culture (Fig. 3). Insertion of a DNA fragment bearing either the native *ebfG*-operon or one encoding EbfG4::FLAG into the double mutant (PilB::Tn5/EbfG4Ω/Comp and PilB::Tn5/EbfG4Ω/EbfG4::FLAG, respectively), restored biofilm development; similarly to PilB::Tn5, less than 5% of the chlorophyll is in suspended cells (Fig. 3). This analysis validated functionality of the tagged EbfG4.

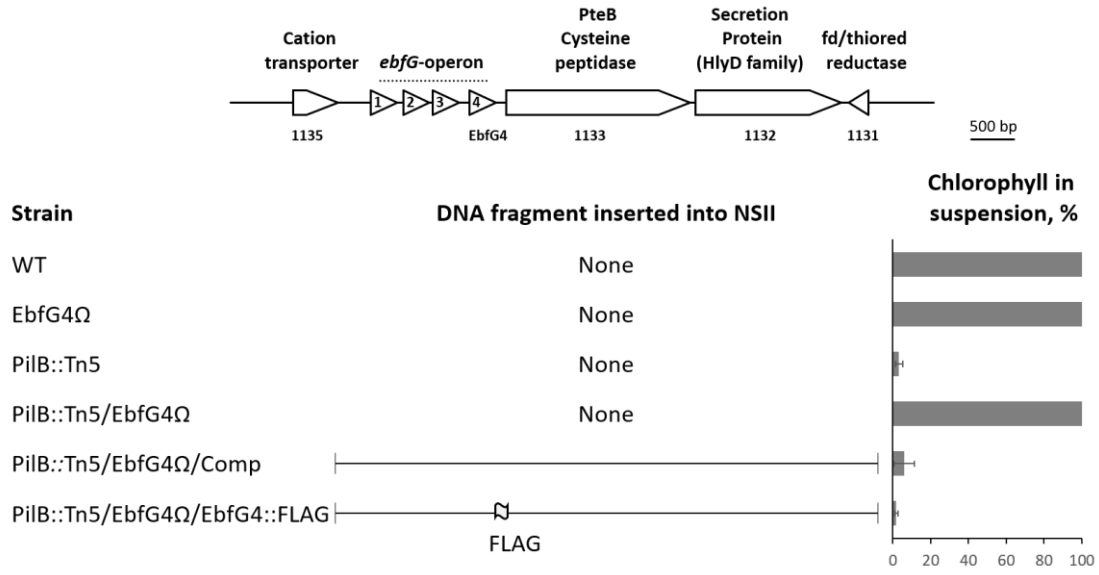


Figure 3: FLAG-tagged EbfG4 is functional in biofilm development.

Genomic region of the *ebfG*-operon. Bar graph presents percentage of total chlorophyll in the suspended cells (average of three independent biological repeats ± standard deviation). Strains analyzed include: WT, a strain in which *ebfG*4 was insertionally inactivated (EbfG4Ω), PilB::Tn5, the double mutant PilB::Tn5/EbfG4Ω, double mutants complemented with the indicated fragments encoding native or FLAG-tagged EbfG4 (PilB::Tn5/EbfG4Ω/Comp and PilB::Tn5/EbfG4Ω/EbfG4::FLAG, respectively.

A first indication that EbfG4 is localized to the cell surface was obtained by dot-blot analysis. Whole cells and cell extracts were spotted onto a nitrocellulose-membrane and probed with anti-FLAG antibodies (Fig. 4). This analysis suggested association of EbfG4 with the cell surface, as revealed by the signal obtained from whole cells (Fig. 4, PilB::Tn5/EbfG4Ω/EbfG4::FLAG). In contrast, the ATPase of T4P complex known to be localized cytoplasmically, was not detected in whole cells but only in cell extracts (Fig. 4, PilB::Tn5/PilB::FLAG), thus confirming availability of internal epitopes for detection only in cell extracts. EbfG4 was not detected in whole cells or in cell extracts of the planktonic strain EbfG4Ω/EbfG4::FLAG (Fig. 4), thus, EbfG4 is neither secreted nor accumulated internally when the biofilm suppression mechanism is active.

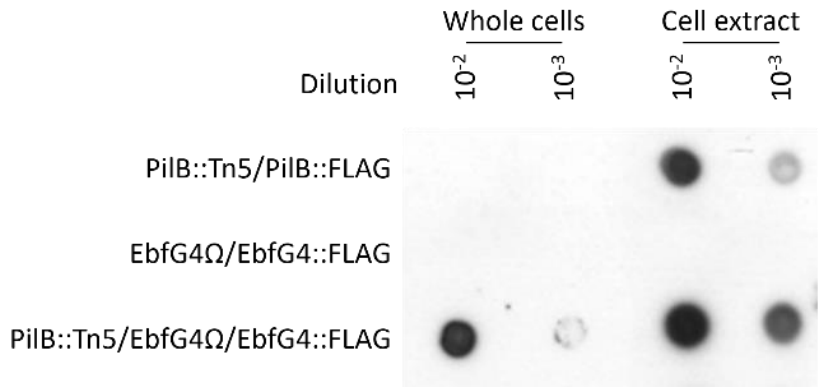


Figure 4: Dot-blot analysis using anti-FLAG antibodies.

Whole cells and cell extracts of the following strains were analyzed: pilB-inactivated strain complemented with FLAG-tagged PilB (PilB::Tn5/PilB::FLAG), ebfG4-inactivated strain complemented with FLAG-tagged EbfG4 (EbfG4Ω/EbfG4::FLAG) and the latter strain that also harbors pilB inactivation (PilB::Tn5/EbfG4Ω/EbfG4::FLAG). Strains PilB::Tn5/PilB::FLAG and EbfG4Ω/EbfG4::FLAG are planktonic and PilB::Tn5/EbfG4Ω/EbfG4::FLAG forms biofilm.

To follow up on the observation suggesting cell surface association of EbfG4 (Fig. 4), we examined by immunocytochemistry non-permeabilized cells that were subjected to anti-FLAG antibodies, thus allowing detection of only extracellular EbfG4. Green signal representing EbfG4 was detected only in strain PilB::Tn5/EbfG4Ω/EbfG4::FLAG but was absent from the cognate control strain PilB::Tn5/EbfG4Ω/Comp confirming specific detection of the FLAG epitope (Fig. 5). Moreover, the green signal is associated with clustered cells (Fig. 5, middle and right columns), while unclustered cells lack green signal and are visualized only by autofluorescence (Fig. 5).

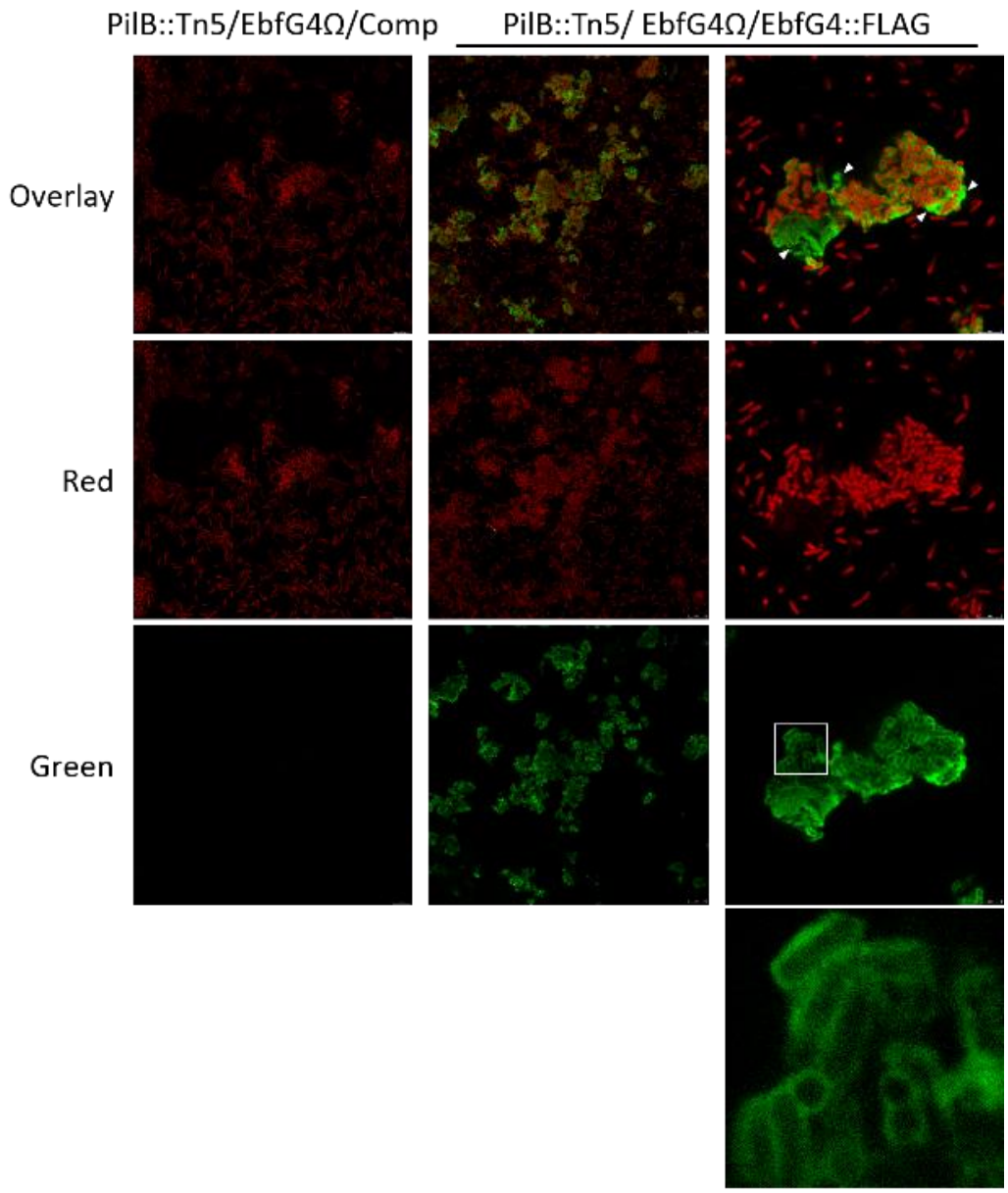


Figure 5: Visualization of EbfG4::FLAG in non-permeabilized cells by immunocytochemistry of confocal fluorescence microscopy.

Strains analyzed: PilB::Tn5/EbfG4Ω/Comp and PilB::Tn5/EbfG4Ω/EbfG4::FLAG. Red (excitation 630nm; emission 636-695nm) represents autofluorescence whereas green (excitation 490nm; emission: 497-539nm) indicates presence of EbfG4::FLAG. White square indicates the enlarged area shown in the panel below. Arrowheads point at patches of green fluorescence in areas void of cells.

Closer examination revealed green signal encasing many of the clustered cells indicating EbfG4 localization throughout the cell surface (Fig. 5, right column). Additionally, patches of green fluorescence are observed in areas void of cells (Fig. 5, right column, arrowheads), and 3D-imaging clearly indicates the presence of EbfG4 in between cells (Fig. 6) supporting its role as a matrix protein.

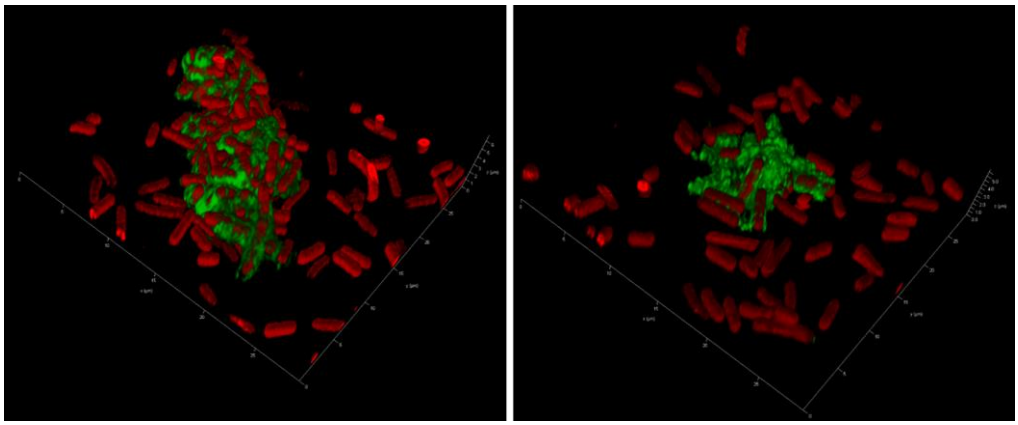


Figure 6: 3D-imaging of strain PilB::Tn5/EbfG4Ω/EbfG4::FLAG.
See details in legend to Fig. 5.

Next, we visualized permeabilized cells to test whether EbfG4 is also detected intracellularly. Images similar to those observed without permeabilization emerged from these analyses (Fig. 7). Close examination of an area mostly void of extracellular green fluorescence did not reveal green signal within the cells (Fig. 7, right column).

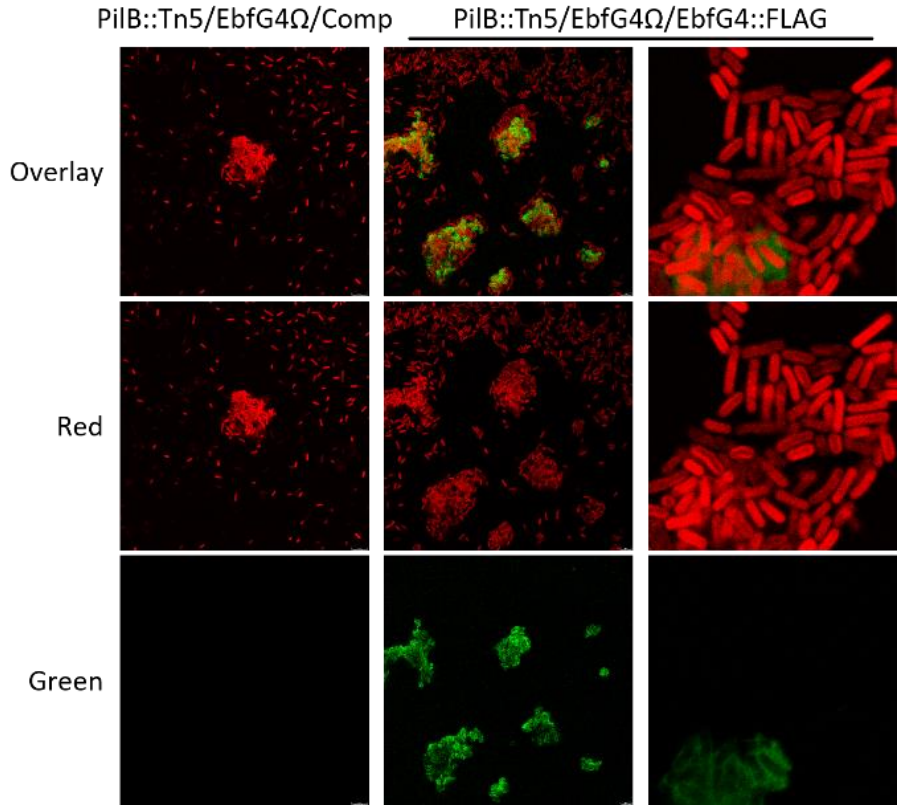


Figure 7: Visualization of EbfG4::FLAG in permeabilized cells by immunocytochemistry and confocal fluorescence microscopy.

Strains analyzed: PilB::Tn5/EbfG4Ω/Comp and PilB::Tn5/EbfG4Ω/EbfG4::FLAG. Red (excitation 630nm; emission 636-695nm) represents autofluorescence whereas green (excitation 490nm; emission: 497-539nm) indicates presence of EbfG4::FLAG.

Successful visualization of internal PilB::FLAG, exclusively in permeabilized cells, validate penetration of the antibodies used for detection (Fig. S2). Together, these data indicate that EbfG4 does not accumulate intracellularly to detectable levels and suggest efficient secretion of this protein.

Examination of amyloid formation by products of the ebfG-operon

To test whether EbfG proteins might contribute to biofilm matrix formation through insoluble aggregates, we initially performed an *in silico* characterization of their tendency to fold as amyloids using a pipeline that consists of several prediction software to build consensus models. We found that EbfG2 and 3 had the highest prediction score, which was above the 0.5 cut-off for classification into amyloid-forming proteins. Using separate tools with diverse prediction methods for the identification of amyloid hotspots within the amino acid sequence, we found that the highly predicted regions occurred at the same location in the alignment of EbfG2 and EbfG3. Despite variations in the individual sequences, the core hotspot followed the motif АΦNIП, where Φ represents a hydrophobic residue, F or I, and П, a small side chain residue, G or S (Fig. 8A). Such motif also occurs in EbfG1, however, in EbfG4 a polar uncharged amino acid (glutamine) is present instead of a hydrophobic residue (AQNIG). We modelled the amyloidogenic LAFNIG peptide from EbfG2 for its ability to form cross-beta structures. We found that the hexapeptide arranged in a steric zipper of antiparallel fashion (Figure S3), characteristic of amyloid proteins [50].

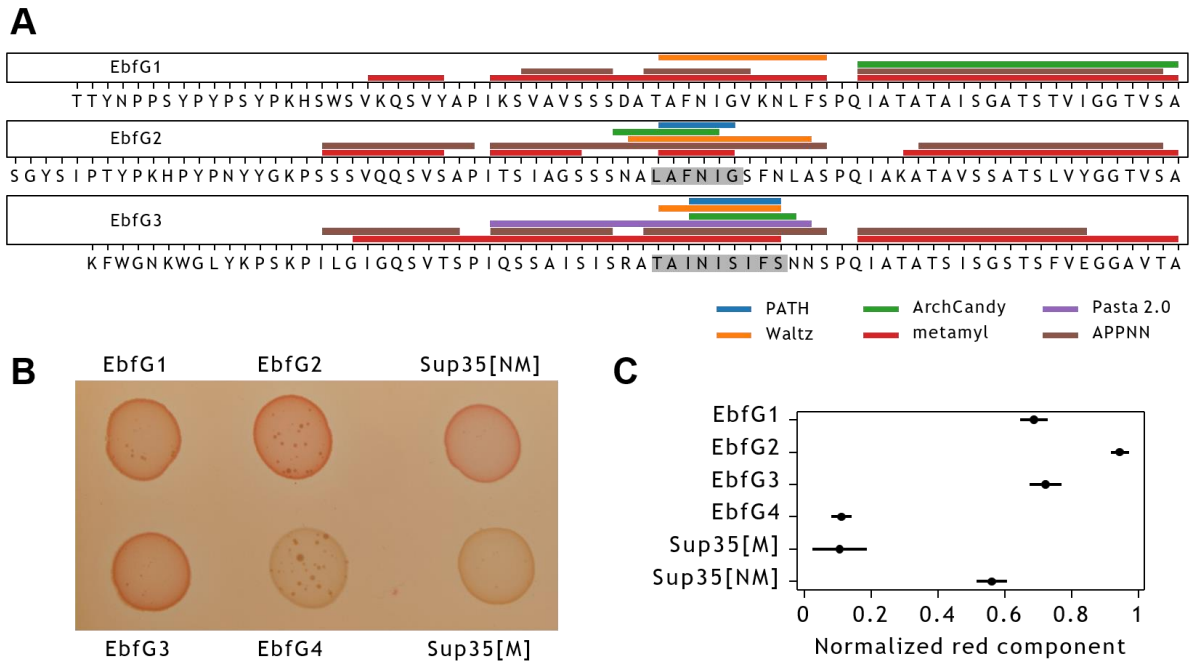


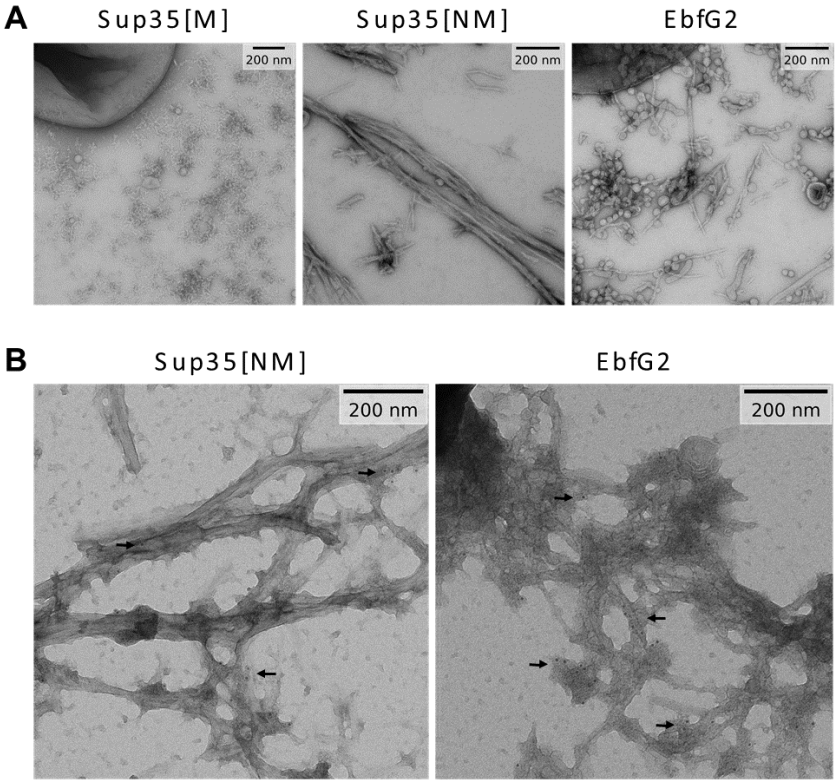
Figure 8: Amyloid prediction and heterologous expression of EbfG1-4 proteins.

(A) Prediction of amyloidogenic hotspots in the sequences of EbfG1-3 using six software prediction tools. Shaded fragments correspond to consensus positive prediction in at least four of the predictors. **(B)** Colony color phenotype of EbfG-expressing *Escherichia coli* compared to negative (Sup35[M]) and positive (Sup35[M]) controls for amyloid formation in inducing plates. (C) Normalized red component of the brim of each colony as analyzed in ImageJ, based on five biological replicates and displayed as the mean and 95% confidence interval (t distribution).

To support these predictions, we heterologously expressed the EbfG proteins to test for the formation of amyloids *in vivo*. To this end, we cloned the mature proteins upstream of the GG motif in the Curli-Dependent Amyloid Generator (C-DAG) system fused to a 6x Histidine Tag at the C-terminus (Table S1). After induction with arabinose, we found a phenotype for amyloid formation in EbfG1, 2 and 3, evident by the colour shift of the colonies due to the binding of Congo Red, similarly to the positive control (Fig. 8B). Consistent with the *in silico* predictions, EbfG2 showed the strongest phenotype while EbfG4 showed no sign of amyloid formation, comparable to the negative control staining (Fig. 8C).

When looking at the induced EbfG2-producing C-DAG cells under the transmission electron microscope, we detected fibril-like structures in the extracellular space. The fibrils resemble those produced by the positive control, however shorter in length and associated with vesicle-like structures (Fig. 9A). Using immunostaining methods, we corroborated the identity of the fibrils as containing EbfG2 protein, based on the

411 colocalization of gold nano-beads directed to the 6x Histidine tag from the protein. We
412 observed unevenness in the distribution of the labelling which could be due to variable
413 antibody accessibility of the tag, both for the positive control and EbfG2 (Fig. 9B).



414

415 **Figure 9: Transmission electron microscopy images of EbfG2 expressed in the C-DAG**
416 **system.**

417 **(A)** Fibril-like structures are found in the extracellular space for the positive control
418 (Sup35[NM]) and EbfG2 but not for the negative control (Sup35[M]) when expressed in the C-
419 DAG system. **(B)** Immunostained samples where 6 nm gold particles localize to the fibril-like
420 structures of the positive control and EbfG2.

421 ## Discussion

422 ***Cell specialization and density-dependent regulation of public goods***
423 ***production in S. elongatus biofilm development***: Previous genetic and point
424 mutation analyses demonstrated the requirement of the EbfG proteins for biofilm
425 formation [23, 26]. As manifested by analysis in individual cells, expression of the
426 operon encoding these genes varied substantially more in the biofilm-forming mutant
427 PilB::Tn5 than among WT cells (Fig. 2A-D). Although ~75% of the mutant cells express
428 the *ebfG*-operon at similar or lower level compared to WT (Fig. 2C), ~90% of the
429 mutant cells are found in the biofilm (Fig. 3, [23, 26, 28]). These findings are consistent
430 with cell specialization during *S. elongatus* biofilm development. Given that EbfG1-3

are prone to form amyloids (Fig. 8) and TEM analysis revealed that EbfG2 forms fibrillar structures similar to amyloids, we propose that a relatively small subpopulation of the mutant culture produces matrix components that support robust biofilm development by the majority of the cells. Taken together, we propose cell specialization in production of so-called public goods. Such a phenomenon may confer selective advantage because only a minority of the cells invest resources for the benefit of the population, which gains protection within the biofilm [51]. Cell specialization in microbes have been documented in heterotrophic bacteria, for example, during matrix formation in *Bacillus subtilis* [52]. Such differentiation, however, was not previously reported for cyanobacterial biofilm development, thus, this study, which uncovers cell differentiation suggests division of labor in communities of these ecologically important photosynthetic prokaryotes.

Comparisons of expression of the *ebfG*-operon in fresh medium and under CM harvested at different time points along growth of WT culture indicate production and secretion of the inhibitor at early culture stages and suggest further accumulation with time and cell density. The amount of biofilm inhibitor present following 12h growth is sufficient to affect reporter expression (Fig. 2D and E), yet is insufficient for biofilm inhibition (Fig. 2F). Up to 24h the inhibitor gradually accumulates (Fig. 2E) and at 48h the inhibitor reaches levels that consistently inhibit biofilm development (Fig. 2F). Together, data imply a density-dependent mechanism, however, a gradual impact of the inhibitor is observed rather than a “threshold-like” effect typical of quorum-sensing mechanisms known in heterotrophic bacteria, e.g. induction of the lux-operon [53-56]. Cyanobacterial quorum-sensing is largely unknown, although N-octanoyl homoserine lactone was suggested to be involved in quorum-sensing in *Gloeothece* PCC6909 [57]. Biosynthetic LuxI-like proteins, which are responsible for production of acylated homoserine lactones in numerous heterotrophs, however, are not encoded in the majority of cyanobacterial genomes [58], therefore, such molecules are not likely to represent a general mechanism for cyanobacterial intercellular communication. An additional study demonstrated a governing role of extracellular signals produced at high density on transcription of particular genes in low-density cultures of *N. punctiforme* PCC 73102. These data support the existence of a quorum-sensing-like mechanism(s), however, the nature of the signal(s) and the regulatory network have yet to be identified [58].

Dual function of EbfG4 – cell surface and matrix protein: Microscopic analysis revealed that some of the cells of PilB::Tn5/EbfG4Ω/EbfG4::FLAG present the EbfG4 protein on their cell surface (Figs. 5&7). Interestingly, EbfG4 is only observed on the surface of clustered cells and cells that lack EbfG4 labelling are dissociated from the clusters (Figs. 5&7). Together, these observations are in accordance with an adhisin function of EbfG4. Additionally, EbfG4 was observed in the intercellular space (Figs. 5-7), consistent with the hypothesis that it serves as a matrix component. It is possible that similarly to the adhisin SasG of *Staphylococcus aureus* [59], EbfG4 is initially deposited to the cell surface and later is shed to the extracellular matrix. The role of EbfG4 in the matrix is unknown, however, although it does not form amyloids by itself (Fig. 8) it may be associated with amyloid structures formed by EbfG1-3.

Amyloids as part of S. elongatus biofilm matrix: When establishing biofilms, microbes require a resilient scaffold on which the cells can settle. Bacteria from diverse ecosystems have solved this problem by producing and releasing functional amyloids into their environment [60, 61]. Amyloid proteins are able to assemble into long and strong fibrils, which can withstand chemical and physical stresses [62]. However, the production of amyloids is a process that can easily get out of control, therefore, it requires a complex and dedicated machinery for appropriate manufacturing. Here, we have investigated the amyloid forming capabilities of the *ebfG*-operon proteins and found strong evidence supporting amyloid formation in EbfG1-3, and most manifestly in EbfG2. EbfG4, which has a prominent role in biofilm formation, however, did not spontaneously form amyloid fibrils. Consistent with homology in the amyloid hotspots, we hypothesize this could be related to a mechanism intended to control aggregation. By separating the amyloid nucleators, in this case EbfG1-3, from other components of the fibril, e.g. EbfG4, better control over the synthesis of amyloids could be achieved. This would be analogous to, for example, the functioning of CsgB and CsgA in the production of curli, the biofilm backbone in *E. coli* [63]. As observed in the TEM pictures, the EbfG2 fibrils were much shorter than the positive control and did not bundle together. Heterogeneous fibrils formed of several EbfG proteins could result in more stable fibrils, as seeding of amyloids composed of perfect repeats has been shown to cause fragmentation [64].

Given the amyloid nature of EbfG1-3 proteins one may speculate that these matrix components of *S. elongatus* biofilms assist in recruitment of other microbes for establishment of multispecies biofilms. Additionally, our findings pave the way for controlling formation of unwanted biomats, by using amyloid disrupting compounds, as already shown for other bacteria [65, 66]. On the other hand, intentional use of protein seeds could facilitate stronger amyloids and hence elicit formation of beneficial biofilms.

Acknowledgments

We thank Ryan Simkovsky for providing the vector for EbfG4-tagging. Studies in the laboratories of Rakefet Schwarz and Susan Golden were supported by the program of the National Science Foundation and the US-Israel Binational Science Foundation (NSF-BSF 2012823). This study was also supported by grants from the Israel Science Foundation (ISF 1406/14 and 2494/19) to Rakefet Schwarz. Studies in the laboratory of Eric Kemen were supported by the graduate school GRK 1708 “Molecular principles of bacterial survival strategies” and the European Research Council (ERC) under the DeCoCt research program (grant agreement: ERC-2018-COG 820124).

References

1. Falkowski, P.G., *The role of phytoplankton photosynthesis in global biogeochemical cycles.* Photosynth. Res., 1994. **39**(3): p. 235-258.
2. Flombaum, P., et al., *Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus.* Proc Natl Acad Sci U S A, 2013. **110**(24): p. 9824-9.
3. Bolhuis, H., M.S. Cretoiu, and L.J. Stal, *Molecular ecology of microbial mats.* FEMS Microbiol Ecol, 2014. **90**(2): p. 335-50.
4. Rossi, F. and R. De Philippis, *Role of cyanobacterial exopolysaccharides in phototrophic biofilms and in complex microbial mats.* Life (Basel), 2015. **5**(2): p. 1218-38.
5. Veach, A.M. and N.A. Griffiths, *Testing the light:nutrient hypothesis: Insights into biofilm structure and function using metatranscriptomics.* Molecular Ecology, 2018. **27**(14): p. 2909-2912.
6. Wagner, M. and A. Loy, *Bacterial community composition and function in sewage treatment systems.* Curr. Opin. Biotechnol., 2002. **13**(3): p. 218-27.
7. Ivnitsky, H., et al., *Bacterial community composition and structure of biofilms developing on nanofiltration membranes applied to wastewater treatment.* Water Res., 2007. **41**(17): p. 3924-3935.

8. Belila, A., et al., *Bacterial community structure and variation in a full-scale seawater desalination plant for drinking water production.* Water Res, 2016. **94**: p. 62-72.
9. Heimann, K., *Novel approaches to microalgal and cyanobacterial cultivation for bioenergy and biofuel production.* Curr Opin Biotechnol, 2016. **38**: p. 183-189.
10. Strieth, D., R. Ulber, and K. Muffler, *Application of phototrophic biofilms: from fundamentals to processes.* Bioprocess Biosyst Eng, 2018. **41**(3): p. 295-312.
11. Bruno, L., et al., *Characterization of biofilm-forming cyanobacteria for biomass and lipid production.* J Appl Microbiol, 2012. **113**(5): p. 1052-64.
12. Egan, S., T. Thomas, and S. Kjelleberg, *Unlocking the diversity and biotechnological potential of marine surface associated microbial communities.* Curr. Opin. Microbiol., 2008. **11**(3): p. 219-25.
13. Agostoni, M., C.M. Waters, and B.L. Montgomery, *Regulation of biofilm formation and cellular buoyancy through modulating intracellular cyclic di-GMP levels in engineered cyanobacteria.* Biotechnol Bioeng, 2016. **113**(2): p. 311-319.
14. Enomoto, G., et al., *Cyanobacteriochrome SesA Is a Diguanylate Cyclase That Induces Cell Aggregation in Thermosynechococcus.* Journal of Biological Chemistry, 2014. **289**(36): p. 24801-24809.
15. Enomoto, G., et al., *Three cyanobacteriochromes work together to form a light color-sensitive input system for c-di-GMP signaling of cell aggregation.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(26): p. 8082-8087.
16. Enomoto, G., Y. Okuda, and M. Ikeuchi, *Tlr1612 is the major repressor of cell aggregation in the light-color-dependent c-di-GMP signaling network of Thermosynechococcus vulcanus.* Sci Rep, 2018. **8**(1): p. 5338.
17. Branda, S.S., et al., *Biofilms: the matrix revisited.* Trends Microbiol., 2005. **13**(1): p. 20-6.
18. Flemming, H.C. and J. Wingender, *The biofilm matrix.* Nature Reviews Microbiology, 2010. **8**(9): p. 623-633.
19. Fisher, M.L., et al., *Export of extracellular polysaccharides modulates adherence of the Cyanobacterium Synechocystis.* PLoS One, 2013. **8**(9): p. e74514.
20. Jittawuttipoka, T., et al., *Multidisciplinary evidences that Synechocystis PCC6803 exopolysaccharides operate in cell sedimentation and protection against salt and metal stresses.* PLoS One, 2013. **8**(2): p. e55564.
21. Kawano, Y., et al., *Cellulose accumulation and a cellulose synthase gene are responsible for cell aggregation in the cyanobacterium Thermosynechococcus vulcanus RKN.* Plant Cell Physiol, 2011. **52**(6): p. 957-66.
22. Oliveira, P., et al., *HesF, an exoprotein required for filament adhesion and aggregation in Anabaena sp PCC 7120.* Environmental Microbiology, 2015. **17**(5): p. 1631-1648.
23. Schatz, D., et al., *Self-suppression of biofilm formation in the cyanobacterium Synechococcus elongatus.* Environ Microbiol, 2013. **15**(6): p. 1786-94.
24. Nagar, E. and R. Schwarz, *To be or not to be planktonic? Self-inhibition of biofilm development.* Environmental Microbiology, 2015. **17**(5): p. 1477-1486.
25. Yegorov, Y., et al., *A Cyanobacterial Component Required for Pilus Biogenesis Affects the Exoproteome.* mBio, 2021. **12**(2).
26. Parnasa, R., et al., *Small secreted proteins enable biofilm development in the cyanobacterium Synechococcus elongatus.* Sci Rep, 2016. **6**: p. 32209.

27. Nagar, E., et al., *Type 4 pili are dispensable for biofilm development in the cyanobacterium Synechococcus elongatus.* Environ Microbiol, 2017. **19**(7): p. 2862-2872.
28. Parnasa, R., et al., *A microcin processing peptidase-like protein of the cyanobacterium Synechococcus elongatus is essential for secretion of biofilm-promoting proteins.* Environ Microbiol Rep, 2019. **11**(3): p. 456-463.
29. Sendersky, E., et al., *Quantification of Chlorophyll as a Proxy for Biofilm Formation in the Cyanobacterium Synechococcus elongatus.* Bio-protocol, 2017. **7**(14): p. www.bio-protocol.org/e2406 DOI:10.21769/BioProtoc.2406.
30. Suban, S., et al., *Impairment of a cyanobacterial glycosyltransferase that modifies a pilin results in biofilm development.* Environ Microbiol Rep, 2022. **14**(2): p. 218-229.
31. Team, R.C. *R: A language and environment for statistical computing*. https://www.R-project.org/ 2021.
32. Ellis B, H.P., Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, Jiang M, Finak G *flowCore: Basic structures for flow cytometry data*. R package version 2.0.1. 2020.
33. *https://docs.flowjo.com/flowjo/workspaces-and-samples/ws-statistics/ws-statdefinitions/*.
34. Kuznetsova, A., Per B. Brockhoff, and Rune HB Christensen, *lmerTest package: tests in linear mixed effects models.* Journal of statistical software 2017. **82**(13): p. 1-26 R package version 3.1-0.
35. Lenth, R.V. *emmeans: Estimated Marginal Means, aka Least-Squares Means.* . R package version 1.6.2-1 2021.
36. Johnson, G.D. and G.M. Nogueira Araujo, *A simple method of reducing the fading of immunofluorescence during microscopy.* J Immunol Methods, 1981. **43**(3): p. 349-50.
37. Fernandez-Escamilla, A.M., et al., *Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.* Nat Biotechnol, 2004. **22**(10): p. 1302-6.
38. Burdukiewicz, M., et al., *Amyloidogenic motifs revealed by n-gram analysis.* Sci Rep, 2017. **7**(1): p. 12961.
39. Familia, C., et al., *Prediction of Peptide and Protein Propensity for Amyloid Formation.* PLoS One, 2015. **10**(8): p. e0134679.
40. Oliveberg, M., *Waltz, an exciting new move in amyloid prediction.* Nat Methods, 2010. **7**(3): p. 187-8.
41. Ahmed, A.B., et al., *A structure-based approach to predict predisposition to amyloidosis.* Alzheimers Dement, 2015. **11**(6): p. 681-90.
42. Walsh, I., et al., *PASTA 2.0: an improved server for protein aggregation prediction.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W301-7.
43. Wojciechowski, J.W. and M. Kotulska, *PATH - Prediction of Amyloidogenicity by Threading and Machine Learning.* Sci Rep, 2020. **10**(1): p. 7721.
44. Emily, M., A. Talvas, and C. Delamarche, *MetAmyl: a METa-predictor for AMYLoid proteins.* PLoS One, 2013. **8**(11): p. e79722.
45. Louros, N., et al., *Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities.* Nat Commun, 2020. **11**(1): p. 3314.
46. Pettersen, E.F., et al., *UCSF ChimeraX: Structure visualization for researchers, educators, and developers.* Protein Sci, 2021. **30**(1): p. 70-82.
47. Sivanathan, V. and A. Hochschild, *A bacterial export system for generating extracellular amyloid aggregates.* Nat Protoc, 2013. **8**(7): p. 1381-90.

48. Sivanathan, V. and A. Hochschild, *Generating extracellular amyloid aggregates using E. coli cells.* Genes & Development, 2012. **26**(23): p. 2659-2667.
49. Taton, A., et al., *Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria.* Nucleic Acids Res, 2014. **42**(17): p. e136.
50. Sawaya, M.R., et al., *Atomic structures of amyloid cross-beta spines reveal varied steric zippers.* Nature, 2007. **447**(7143): p. 453-7.
51. Yin, W., et al., *Biofilms: The Microbial "Protective Clothing" in Extreme Environments.* International Journal of Molecular Sciences, 2019. **20**(14).
52. Dragos, A., et al., *Division of Labor during Biofilm Matrix Production.* Curr Biol, 2018. **28**(12): p. 1903-1913 e5.
53. Fuqua, C. and E.P. Greenberg, *Listening in on bacteria: acyl-homoserine lactone signalling.* Nat. Rev. Mol. Cell Biol., 2002. **3**(9): p. 685-695.
54. Parsek, M.R. and E. Greenberg, *Sociomicrobiology: the connections between quorum sensing and biofilms.* Trends Microbiol., 2005. **13**(1): p. 27-33.
55. Kolter, R. and E.P. Greenberg, *Microbial sciences - The superficial life of microbes.* Nature, 2006. **441**(7091): p. 300-302.
56. Bassler, B.L., *Small talk. Cell-to-cell communication in bacteria.* Cell, 2002. **109**(4): p. 421-424.
57. Sharif, D.I., et al., *Quorum sensing in Cyanobacteria: N-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium Gloeothece PCC6909.* ISME J, 2008. **2**(12): p. 1171-82.
58. Guljamow, A., et al., *High-Density Cultivation of Terrestrial Nostoc Strains Leads to Reprogramming of Secondary Metabolome.* Appl Environ Microbiol, 2017. **83**(23).
59. Geoghegan, J.A., et al., *Role of surface protein SasG in biofilm formation by Staphylococcus aureus.* J Bacteriol, 2010. **192**(21): p. 5663-73.
60. Gomez-Perez, D., et al., *Amyloid Proteins in Plant-Associated Microbial Communities.* Microb Physiol, 2021. **31**(2): p. 88-98.
61. Taglialegna, A., I. Lasa, and J. Valle, *Amyloid Structures as Biofilm Matrix Scaffolds.* Journal of Bacteriology, 2016. **198**(19): p. 2579-2588.
62. Rambaran, R.N. and L.C. Serpell, *Amyloid fibrils: abnormal protein assembly.* Prion, 2008. **2**(3): p. 112-7.
63. Hammer, N.D., J.C. Schmidt, and M.R. Chapman, *The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization.* Proceedings of the National Academy of Sciences, 2007. **104**(30): p. 12494-12499.
64. Rasmussen, C.B., et al., *Imperfect repeats in the functional amyloid protein FapC reduce the tendency to fragment during fibrillation.* Protein Sci, 2019. **28**(3): p. 633-642.
65. Romero, D., et al., *Biofilm Inhibitors that Target Amyloid Proteins.* Chemistry & Biology, 2013. **20**(1): p. 102-110.
66. Jain, N., et al., *Inhibition of curli assembly and Escherichia coli biofilm formation by the human systemic amyloid precursor transthyretin.* Proc Natl Acad Sci U S A, 2017. **114**(46): p. 12184-12189.

Supplementary Information

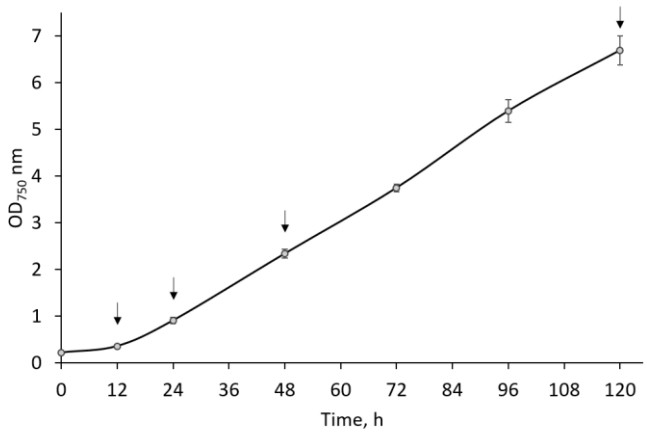


Figure S1: Growth of WT cultures as measured by OD_{750} nm as a function of time. Data represent averages and standard deviations from 3 technical repeats. Arrows indicate time points at which CM was harvested.

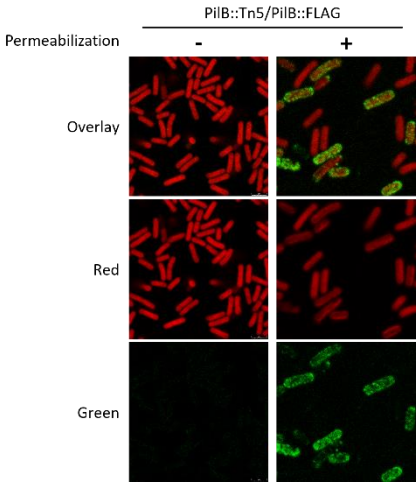


Figure S2: Immunocytochemistry with or without permeabilization of strain PilB::Tn5/PilB::FLAG.

Red (excitation 630nm; emission 636-695nm) represents autofluorescence whereas green (excitation 490nm; emission: 497-539nm) indicates presence of EbfG4::FLAG.

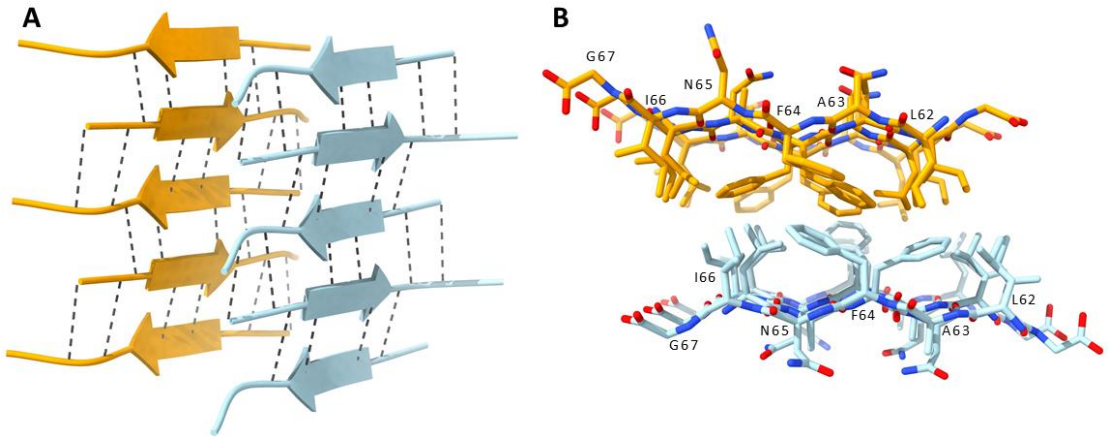


Figure S3: Cross-beta structure modelling of amyloidogenic LAFNIG peptide from EbfG2.

(A) Three-dimensional model of the LAFNIG peptide displaying a cross-beta antiparallel structure. In dashed black lines, hydrogen bonds holding together each cross-beta sheet are depicted. (B) Upper view of the structure highlighting the steric zipper, where hydrophobic residues (L, A, F and I) are concentrated in between the beta-sheets and hydrophilic residues (N and G) directed towards the outside.

Cloning of FLAG-tagged EbfG4*		
Primer Name	**Primer sequence (upper forward, lower reverse)**	**Purpose**
RP10-swaI-F	CGGCCAATAACCCAGGGATTTTGTACTGCGACTCGACCAAG	*Cloning FLAG Fragment #1*
RP10-swaI-R	CTGCCGGGGAGCTCCTTCATTTGTTATGAACGTTGGCGATCG	*Cloning Fragment #3*
EbfG4-FLAG-R	CTTATCGTCGTCATCCTTGTAATCAAGATTGACAGCTACGTTAATGG	*Cloning FLAG Fragment #1*
EbfG4-FLAG-F	GATTACAAGGATGACGACGATAAGTAGCCAGTGCGATCGCAAG	*Cloning FLAG Fragment #2*
NS2-2F	CTCTTGGTGCTGTTCAGTC	*Sequencing*
Cloning of YFP-reporter **		
Primer name	**Primer sequence (upper forward, lower reverse)**	**Purpose**
P-EbfG::YFP Forw	AGTCGGCCAATAACCCAGGGATTTTGTACTGCGACTCGACCAAG	P-EbfG::YFP, *Fragment #1*
P-EbfG_RBS::YFP tail Rev	CACCATCTTAAGACCTCCTTTATTTTTTTTGATTCGCCCTTGACTC	
P-EbfG:: YFP_Forw	AGTCGGCCAATAACCCAGGGATTTTGTACTGCGACTCGACCAAG	P-EbfG::YFP, *Fragment #2*
YFP_genomic 3' UTR Rev	CATCCGTCAGGATGGCCTTCTCCTGCAGGGCCGAGTTTGTACAAGAAAGC	
NSI- pRP22 ins F	ACCGGTTAATAGTACCTGTG	P-EbfG::YFP, PCR analysis on *E. coli* cells
NSI- pRP22 ins R	ATGCCCGCGACATCTTCC	
NSI and pRP22 F	AGACGAGGCAAGCATTGAGC	P-EbfG::YFP, PCR analysis on cyano cells
NSI- pRP22 ins R	ATGCCCGCGACATCTTCC	
Cloning of fragments encoding EbfG proteins into C-DAG system#		
Primer name	**Primer sequence (upper forward, lower reverse)**	**Purpose**
pExport:ebfg1	TATAGCGGCCGCAAAATTCTGGGGG	PCR amplification
	TATATCTAGATTAGTGATGATGGTGATGGTGGGCAGATACAGTTC	
pExport:ebfg2	TATAGCGGCCGCAAGTGGTTATTCCATT	PCR amplification
	TATATCTAGATTAGTGATGATGGTGATGGTGAGCTGAAACAGTC	
pExport:ebfg3	TATAGCGGCCGCAAAATTCTGGGGG	PCR amplification
	TATATCTAGATTAGTGATGATGGTGATGGTGGGCTGTAACTGC	
pExport:ebfg4	TATAGCGGCCGCAACTAATTGCAATCC	PCR amplification
	TATATCTAGATTAGTGATGATGGTGATGGTGTTGCGATTATATTTCTT	
pExport	CCTGACGCTTTTTATCG	Sequencing
	GCTGAAAATCTTCTCTC	

Table S1: Summary of cloning information

***** EbfG4::FLAG tagged vectors were generated using the GeneArt® Seamless Cloning and Assembly Kit (Life Technologies) and Top10 cells (Life Technologies) on three PCR-generated fragments per plasmid and a SwaI-digested CYANO-VECTOR cloning plasmid, pCV0049 or pAM4937, that encodes for kanamycin (Km) resistance, Neutral Site II (NS2) integration, and

a *ccdB*-suicide gene that is removed upon SwaI digestion [1]. PCR fragments were amplified from the complementation plasmid pAM4997 (pRP10, [2]) using Q5® High-Fidelity DNA polymerase (New England Biolabs) and primers designed to add overlaps for cloning into the SwaI cut site or to add the flag tags to the C-terminus of the encoded EbfG4 protein. Clones were screened by PCR and the entire complementation region of the newly generated vectors were verified by Sanger sequencing.

****** The genomic region bearing the promoter of the *ebfG*-opreon and YFP were PCR amplified and then served as primers for each other to PCR amplify a fusion product.

PCR products were amplified from pRP10 [2] and cloned into the vector pExport [3, 4]. Fragments were inserted at the NotI (5') and XbaI (3') sites between the CsgA secretion signal and stop codon of pExport. All cloning products were validated by PCR analyses and sequencing.

References

1. Taton, A., et al., *Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria.* Nucleic Acids Res, 2014. **42**(17): p. e136.
2. Parnasa, R., et al., *Small secreted proteins enable biofilm development in the cyanobacterium Synechococcus elongatus.* Sci Rep, 2016. **6**: p. 32209.
3. Sivanathan, V. and A. Hochschild, *A bacterial export system for generating extracellular amyloid aggregates.* Nat Protoc, 2013. **8**(7): p. 1381-90.
4. Sivanathan, V. and A. Hochschild, *Generating extracellular amyloid aggregates using E. coli cells.* Genes & Development, 2012. **26**(23): p. 2659-2667.

Appendix C

Completed manuscript draft

C.1 Plant-associated protist *Albugo candida* produces amyloid-like proteins relevant for pathogenicity

Contributions by DGP:

- Designed and wrote the initial draft together with MS.
- Helped establish, together with MS, the experimental pipeline.
- Edited the final manuscript.
- Performed all computations and analyzed all data.
- Created all the figures and tables except for Figures 1 and 2, which were created by JR, and the electron microscopy images which were captured by SR and YDS.

Plant-associated protist *Albugo candida* produces amyloid-like proteins relevant for pathogenicity

Daniel Gómez-Pérez [1,†], **Monja Schmid** [1,†], **Ariane Kemen**[1], **Sandra Richter**[2], **York-Dieter Stierhof**[2], **Jonas Ruhe**[3], **Elke Klenk**[1], **Sergey Ryazanov**[4], **Christian Griesinger**[4], **and Eric Kemen** [1,✉]

[1]Microbial Interactions in Plant Ecosystems, Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany

[2]Microscopy, Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany

[3]Max Planck Institute for Plant Breeding Research, Cologne, Germany

[4]Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany

[†]These authors contributed equally to this work

In many plant ecosystems, amyloid proteins are prevalent and functionally relevant mechanistic assets. Their characteristic folding enables them to function, among others, as structural support when in fibril stage, or as toxic compounds when in monomeric form. While many of their functions have already been studied in a medical context, much remains to be understood about their role in microbes from natural ecosystems. In this study, we explore amyloid proteins from an obligate biotroph and their role within the model plant pathosystem of *Albugo candida* and *Arabidopsis thaliana*. Firstly, we provided evidence of the production of amyloids by *Albugo*, particularly in the hyphal compartment within the plant apoplast. Secondly, we investigated if amyloids were relevant for *Albugo*'s virulence. After application of amyloid inhibitors on *Albugo*-infected plants, we observed a decrease in the infection efficiency. Thirdly, we identified amyloid-like proteins released by *A. candida*. By using in silico amyloid predictors on the apoplastic proteome of *A. candida*, we selected potential candidates. We expressed these candidates in vivo in a heterologous prokaryotic system. Consistent with the predictions, we observed amyloid formation of the candidates using Congo Red staining and TEM imaging. Taken together these results suggest the relevance of amyloid proteins for *Albugo*'s interactions and pathogenicity.

Correspondence: *eric.kemen@uni-tuebingen.de*

Introduction

Functional amyloids are proteins capable of folding in an amyloid structure that is relevant for their activity in biological systems (Otzen and Riek, 2019). The molecular structure of amyloids consists of up to several hundred nm long, 5 to 20 nm thick unbranched fibrils that are made up of monomeric cross-beta motifs stacking on their perpendicular axis (Jahn et al., 2010). The monomers can be variable in composition to a certain degree, which often results in the cross-seeding of heterogeneous domains (Chaudhuri et al., 2019). This, together with the growth-triggering effect or seeding activity of short oligomeric amyloids, is relevant for their pathogenic aspect (Ivanova et al., 2021). When fully mature, the fibrils are strong and resilient against denaturing stressors, which makes them useful for structural functions, such as forming the scaffolding of biofilms or creating hydrophobic structures (Macindoe et al., 2012; Erskine et al., 2018). In their intermediate oligomeric conformation, they may present other functions owing to their higher reactivity, such as signaling and antimicrobial activity (Kagan et al., 2012; Li et al., 2012). The latter is effective through non-specific binding of the membrane and is a topic of broad and current interest in research due to the increasing prevalence of antibiotic resistant bacteria and the need to find new antimicrobial agents (Park et al., 2011).

The most common methods for confirming the presence of amyloid structures are through staining with dyes, namely Congo Red (CR) and Thioflavin T (ThT), as well as the use of transmission electron microscopy (TEM) to directly observe the fibrils (Wu et al., 2007; Gras et al., 2011). Specially designed expression systems can be used for the production of proteins of interest to analyze their capability to form amyloid fibrils in vivo in a heterologous context. The Curli-Dependent Amyloid Generator (C-DAG) system based on *Escherichia coli* makes use of the curli biosynthesis pathway. By fusing the secretion signal of the major curli component protein ($csgA_{ss}$) to a potential amyloid, it is possible to exploit the native pathway to trigger amyloid formation (Sivanathan and Hochschild, 2013).

The study of functional amyloids has long been of interest for its link to microbial survival in medical and environmental conditions, for example, by promoting the prevalence of microbes through biofilm development (Gerven et al., 2018). In particular, amyloids are a ubiquitous folding pattern in proteins related to host-microbe and microbe-microbe interactions, including those in plant ecosystems, originating both from the plant and the microbes (Gómez-Pérez et al., 2021). However, most of what is currently known about functional amyloids originates from in vitro studies of isolated proteins, therefore much remains to be understood about their behavior in complex communities. Certain characteristics of amyloids, namely their heterogeneous templatability, growth by seeding, and nucleation mechanisms as well as their pervasiveness in nature, suggest that cross interactions from proteins of different origins may play an important role in the ecological interplay of microbial communities (Zhou et al., 2012). Since amyloid proteins have been associated with functions related to increased microbial pathogenicity in many organisms, they could be a target of new anti-infectives preventing microbial invasion and colonization (Shanmugam et al., 2019). In the medical field, amyloid inhibitors have already been applied as potential treatments for amyloid-associated neurodegenerative diseases and as inhibitors of biofilms (Mathew et al., 2022; Romero et al., 2013). As phytopathogens have been shown to employ amyloids, specific inhibitors against these pathogens could also be an option to fight infection in agricultural settings (Oh et al., 2007; Müller et al., 2008; Gómez-Pérez et al., 2021).

The plant pathosystem comprising *Albugo* and *Arabidopsis thaliana* is an established model to study microbial interactions within a host-pathogen context, since the former is a microbe crucial in shaping the microbial community of the latter (Agler et al., 2016). *Albugo* is a eukaryotic oomycete that follows an obligate biotrophic lifestyle and causes white rust disease in plants (Kemen et al., 2011). It lives within the leaves and stems of the host where it develops elongated cellular structures known as hyphae that branch throughout the space outside the plant plasma membrane, known as apoplast. Especially, obligate biotrophic pathogens within plant ecosystems strongly rely on proteins for survival and pathogenicity. Being closely adapted to their host organism has led to a loss of primary and secondary metabolic pathways, which highlights the relevance of proteins as the main mechanism for interaction with its host and surrounding microbial community (Kemen et al., 2011). Many aspects of the pathogenicity of oomycetes are homologous to those in filamentous fungi. Contrary to the scant evidence in protists, filamentous pathogens from the fungal kingdom have been shown to secrete functional amyloids related to virulence and pathogenicity (Talbot et al., 1996; Shanmugam et al., 2019). Examples include the rust fungus *Uromyces fabae*, from which the apoplastic effector RTP1 was identified as amyloid-forming (Kemen et al., 2013).

In this report, we study proteins secreted by *Albugo candida* for their capacity to form amyloids, and discuss the potential physiological relevance of these structures in such a plant protist pathogen.

Results

Amyloid-like proteins accumulate along *Albugo* hyphae

Amyloids have been shown to be functionally relevant when localized on the surface of apoplast-inhabiting filamentous fungi (Kemen et al., 2013; Tanaka and Kahmann, 2021). When investigating the presence of amyloids in *Albugo*, we conducted fluorescence microscopy experiments to investigate whether that could also be the case for the biotrophic oomycete. By imaging infected *A. thaliana* leaves, we showed significant staining with an amyloid-specific dye, ThT, of the *A. candida* hyphae growing within the plant apoplast as compared to the plant cellular background (Figure 1). While most of the hyphae was heavily stained with the dye, hot spots with increased fluorescence were visible in the hyphal structure. Since some amyloids have been shown to exhibit antimicrobial activity and partake in microbe-microbe interactions (Shahnawaz and Soto, 2012), we investigated the microbial landscape in the surroundings of *Albugo* hyphae within the apoplast. To this end, we performed staining of hyphae and bacteria and visualized it on multiple z-sections with depth coloring. These fluorescence microscopy images showed that bacteria colonized the apoplastic compartment around the grown hyphae, and that some of these bacteria were present closely adjacent to the hyphal surface (Figure 2).

Amyloid inhibitors decrease *Albugo* pathogenicity

In order to gauge the reliance of *Albugo* on amyloids for successful infection and hence pathogenicity, we performed infiltration experiments with amyloid inhibitors. By using the number of pustules per leaf area as a measure of infection efficiency, we observed a decrease in the pathogenicity of *Albugo* in the samples treated with amyloid inhibitors, as compared to the DMSO negative controls. In an initial screen with four amyloid inhibitors (B13, B14, B15, B16; Table 2) infiltrated prior to *A. candida* strain Nc2 or *A. laibachii* Nc14 infection, we found B13, which also had the largest polar surface of all four, to have the highest protection against infection. The three others (B14, B15 and B16) promoted infection compared to DMSO, but resulted in less infection compared to non-infiltrated plants (Figure 3). Upon infiltration with inhibitor B13, we found a decrease in the infection severity for both *Albugo* strains in two separate biological replicates of 12 plants each (Figure 4).

Amyloid prediction reveals enrichment of *Albugo* amyloids in apoplastic proteome

To assess the relevance of amyloid-forming proteins from *Albugo* in the apoplast, we performed machine-learning-based predictions of amyloid structure in the proteome of *A. candida*. We compared the scores of the apoplastic proteome, derived from our previous study (Gómez-Pérez et al., 2022), to those of the total predicted proteome of *A. candida* Nc2, as available in the GeneBank® database. A positive prediction score was defined as being above the median score of the whole proteome. Out of the 563 identified apoplastic proteins, 313 placed above the cutoff (Figure 5). This corresponded to a significant enrichment in the number of proteins with a positive amyloid prediction in the apoplast subset (Kolmogorov–Smirnov test, p value < 0.001), also evidenced by the larger component of the normalized cumulative function that is over the cutoff (Figure 6). We selected individual proteins for further study of their amyloid properties following certain conditions to minimize potential false positives: high prediction scores and large relative abundances in the apoplast as represented by their corresponding peptide intensity from the apoplast proteomic analyses (higher than the median of total raw intensities) while maintaining a protein length below 700 (Figure 7). The length criteria was chosen to be within the range of most known amyloid proteins, as assessed in the AmyPro database, a collection of amyloid-forming proteins from different organisms (Varadi et al., 2017). We chose candidates with a highly positive prediction in the fourth quartile of the prediction distribution (C16, C05, C14, C17 and C18), candidates with a prediction score closer to the cut-off in the interquartile range (C10, C13, C12, C15 and C19) and two candidates with a low amyloid prediction in the first quartile (C20 and C06) as negative controls. All selected candidates and their characteristics are found in Table 1.

Albugo protein candidates form amyloids in C-DAG system

We cloned the selection of candidates into the C-DAG system and expressed the proteins via *E. coli*'s native curli export pathway to see if they could adopt the amyloid fold upon secretion. We incubated the *E. coli* transformants containing the candidate constructs at 22 °C on inducing CR plates and observed differences in the color phenotype of the colonies. In general, we found that highest scoring amyloid proteins as predicted in silico, namely C14, C16 and C17 had a red phenotype closer in color to the positive control, yeast prion Sup35[NM], indicating the formation of amyloids in vivo. Candidates not expected to construct amyloids according to the weak in silico prediction (C06, C15, C19, C20) showed a white or faint red color phenotype matching the negative control, the same yeast prion without the aggregating domain. For the candidates with prediction scores in the cut-off area, we found red phenotypes of varying intensity. C10 presented a strong red

phenotype while C12, C13 and C18 showed a light red phenotype (Figure 8). For C05, the candidate with the second highest amyloid prediction, we observed poor growth after induction. To assess if the inhibited growth was correlated to amyloid production, we grew construct C05 on CR plates with variable arabinose concentration for induction. We found complete growth inhibition on the higher arabinose plates (0.8%), and normal growth with a strong red phenotype, comparable to the other positive candidates, on lower arabinose plates (0.02%; Figure 9).

TEM images show fibril formation

To examine the morphology of the fibrils and add further support to the hypothesis of amyloid production by *Albugo*, we imaged the cells expressing the candidate proteins using a TEM. We found distinct fibril-like structures surrounding the *E. coli* cells for proteins C05, C10, C13, C14 and C16. While C05 only showed scattered red spots, due to weak growth, most of the other four candidates were among the ones with the strongest red colony color phenotype . The fibrils from C10 and C16 were short and bundled together. On the other hand, for C05, C13 and C14, they were very long, thin and did not group together. For the candidates matching in color with the negative control, C06, C15 and C19, and for candidate C17, we did not find any fibrils. The candidates with a light red phenotype (C12, C18, C20) displayed sparse thin fibrils (Figure 10). For the majority, the appearance of the fibrils was matching with the observed colony color phenotype.

Glucosyltransferase family is contracted in *Albugo*

Among the candidates, we chose C14 for a more in depth analysis of its evolutionary background. It showed on the one hand, evidence of fibril formation and on the other, antimicrobial potential as described in a previous study (Gómez-Pérez et al., 2022), we aimed to understand its evolutionary history. To investigate this, we constructed a phylogenetic tree including homologs of the C14 gene from the oomycetes (Figure 11). As an outgroup, the four copies from *Saccharomyces cerevisiae* were chosen. The tree showed an expansion in the majority of plant parasitic oomycetes for this gene, including *Phytophthora* and *Pythium* species, and the presence of a unique gene in the two *Albugo* species, Nc2 and Nc14.

Discussion

Functional amyloid proteins have been widely reported and characterized in fungi (Macindoe et al., 2012; Loquet and Saupe, 2017). Due to the high morphology and lifestyle similarity of fungi and oomycetes which likely stems from convergent evolution, amyloids are potentially an important mechanistic asset as well in this lesser studied phylum. Some of their reported functions in fungi include: as an epigenetic mechanism in yeast that supports adaptation to new extreme environments, and as hydrophobins that help break the water-air interface for the release of spores and their hydrophobic coating for better dispersal (Nizhnikov et al., 2016; Puspitasari and Lee, 2021). The relevance of amyloid proteins has also been shown in obligate biotrophic rust fungi, where their function on the hyphal surface was related to protection from degradation, increasing the duration of the biotrophic phase (Kemen et al., 2013). Since some oomycetes follow a similar obligate biotrophic lifestyle, and due to the documented correlation between lifestyle and protein strategies (Gómez-Pérez and Kemen, 2021), we hypothesized that amyloid functionality may be overlapping. In this report, we have done the first screening of amyloid proteins in the secretome of a biotrophic protist, the oomycetal plant pathogen *Albugo candida*, and evaluated their role for pathogenicity.

We found evidence of staining of the *Albugo* hyphae with an amyloid dye, ThT, indicating production of amyloids (Figure 1). The presence and distribution of amyloids along the hyphal structure has been associated in fungi with adhesion (Lipke et al., 2012; Teertstra et al., 2009) or reinforcement of the cell wall (Kemen et al., 2013). However, the uneven distribution apparent in the picture could indicate a localized effect rather than ubiquitous presence. As predictions on the secretome of *Albugo* indicated amyloid-forming potential of many of the secreted proteins and as we found evidence of amyloid formation for some selected candidate proteins, secreted amyloids could be interesting contenders for potential contact points with surrounding microbes or host cells. We found bacteria growing in close proximity to the hyphal structure, which, as reported in literature, further emphasizes the potential for microbial interactions at this interface (Scheublin et al., 2010; Steffan et al., 2020).

The importance of amyloids for the pathogenicity of *Albugo* is also indicated by the decrease in infection efficiency upon addition of an amyloid inhibitor. Due to the involvement of amyloids in neurodegenerative diseases, the search for amyloid inhibitors as potential treatment is an important topic in medical research (Ankarcrona et al., 2016). Since certain phythopathogens have been shown to rely on amyloids for virulence (Oh et al., 2007; Shanmugam et al., 2019), we put forward the idea of using inhibitors as potential tools to fight disease in the agricultural sector as well. Based on the prevalence of putative amyloids in *Albugo*, especially in the secreted portion, we tested a set of amyloid inhibitors with different physical characteristics in order to find one that could successfully repress *Albugo* infection. We showed that infiltration of amyloid inhibitor B13, which had the highest polar surface area of all tested inhibitors, can decrease the *Albugo* infection level to a certain degree (Figures 3 and 4). As B13 was designed towards a general inhibition of amyloids without a particular target, there might also be nonspecific inhibition of the plant amyloids (Nasi et al., 2021; Swarupini and Bhuyan, 2018; Antonets and Nizhnikov, 2017). The latter could promote plant defenses and thereby make it harder for *Albugo* to thrive. As we observed slightly decreased infection when DMSO was infiltrated as control compared to no infiltration treatment, this procedure might induce stress on the plant. Since we observed fluorescence in the apoplastic background surrounding the heavily fluorescent *Albugo* hyphae (Figure 1), ThT was likely bound to amyloid fibrils originating from plant cells. Still, these results could be the first step for future research regarding amyloid inhibitors as a potential tool in the defense against crop pathogens.

After establishing that amyloid inhibitors decrease *Albugo*'s ability to infect its host, we looked at the specific proteins of the secretome of *Albugo* for their potential amyloid folding. We found evidence of amyloid formation within the subset of apoplastic proteins, on the one hand by positive amyloid prediction of a large part of the secretome and on the other by experimental evidence of individual protein candidates. During the analysis of these in the C-DAG system, C05, C10 and C14 stood out. We found that growth of cells expressing C05, a highly positive predicted amyloid, was inhibited at the recommended inducing concentration for the color phenotype analysis assay, suggesting a failure of the protein expression mechanism leading to overwhelming of the cell machinery. Lowering inducing concentration of arabinose led to normal growth with a strong red phenotype and clear fibrillar structures in the TEM, indicating that amyloids could successfully form under low induction conditions. C10 showed strong evidence for amyloid formation in the tests performed. Blastp search on the non-redundant protein NCBI database for this protein resulted in no homologs from close species, which together with its short length (178 amino acids) and abundance of cysteine residues (12) potentially forming disulfide bonds suggests of its nature as effector. Effectors are proteins involved in host mediation, therefore, they are usually relevant in pathogenicity.

Further evidence implying relevance of secreted proteins for pathogenicity can be found in C14. In Gómez-Pérez et al. (2022), we reported it to be highly antimicrobial against certain types of plant-associated bacteria, particularly in its disordered C-terminal domain. In this study, we found the candidate C14, a cell wall-associated glucosyltransferase, to additionally present characteristics of amyloid folding through evidence of its expression

in the C-DAG system. These results are in accordance with previous studies of the ortholog glucosyltransferase GAS1 from the yeast *Saccharomyces cerevisiae* (Kalebina et al., 2014; Sergeeva et al., 2018; Ryzhova et al., 2018). Enzymes from this carbohydrate-active family are thought to anchor to the cell wall and reshape it in response to external stimuli, as well as take part in other transcription-related functions (Ragni et al., 2007; Koch and Pillus, 2009; Zhang et al., 2011; Imai et al., 2019). The amyloid fold is suggested in Ryzhova et al. (2018) to be relevant to its function due to templating and agglutination of the enzyme at a certain point in the cell wall, for example, as a way to direct growth of the cell wall as the hyphae extends. The clade of albuginales, as opposed to most other oomycetes and fungi that have several genes encoding glycosyltransferases, presents a unique copy of the GAS1 gene (Figure 11). This may be related to genome reduction in obligate parasites, wherein pleiotropy or the development of new functions from the same protein unit may counteract the fitness loss from a reduction in the overall enzyme repertoire. The capacity to form amyloids under certain conditions could be related to the expansion of functions contained within the same sequence to save valuable genomic real estate. However, in this case the conservation of amyloid folding through such a long evolutionary distance resulting in a high sequence diverge poses an evolutionary riddle. It suggests that the necessity of maintaining an amyloid folding constricted the evolutionary pressures of this protein.

In this study, we present the first evidence of amyloid formation in proteins of an obligate biotrophic oomycetal plant-pathogen *A. candida*. Our results indicate that *Albugo* forms amyloid structures while inhabiting the apoplast and that these amyloids are relevant for pathogenicity, since infiltrating amyloid inhibitors decreases infection efficiency. Taken together, we add to the increasing evidence that suggests the relevance of the functional amyloid fold in natural ecosystems and propose the use of amyloid inhibitors as a way to fight or prevent infection of plants.

Methods

Thioflavin T in planta staining

To examine the amyloid production of the hyphal structure, we sectioned an *A. thaliana* leaf infected with the *A. candida* strain Nc2 in thin slices and floated them in a solution of 10 µM Tris buffer (pH 7.2) and 10 µM ThT. After incubation for 30 min in the dark, we analyzed the fluorescence staining under an LSM780 laser confocal microscope (Carl Zeiss, Germany). We took pictures using a 40x oil objective (laser: 458 nm, filter: 455-552 and 650-740). The fluorescence spectrum of amyloid-bound ThT was measured by performing lambda spectral stacks. We used this spectrum for lambda unmixing and compilation of a z-stack of the *Albugo* hyphae.

Uvitex 2B and Sytox^{TM} green staining

We used fluorescence staining to visualize apoplastic colonization during *A. candida* infection of *A. thaliana* (12 days post infection). We chose Uvitex 2B for hyphal staining and Sytox^{TM} green for bacteria. We prepared a 4% paraformaldehyde solution (0.4 g paraformaldehyde, 8 ml ddH_2O, 5 µl 1 M NaOH). The solution was heated to 70 °C and cooled to room temperature before adjusting volume to 9 mL and adding 1 ml 10x microtubule stabilizing buffer (MTSB, 50 mM PIPES, 5 mM EGTA, 5 mM $MgSO_4$). We incubated infected leaf slices in the prepared solution for 1 h. We washed the samples once for 15 min, once for 45 min in 1x MTSB prior to adding 489 µl 1x MTSB, 5 µl Sytox^{TM} green, 1 µl Uvitex 2B and 5 µl Triton X100 (10%). Incubation was continued for 1 h in the dark. We washed the samples again as described above and transferred them to a microscopy slide with 50% glycerol. We took the pictures using an LSM700 fluorescence laser confocal microscope (Carl Zeiss, Germany) with a 63x oil objective (lasers: 405 nm, 488 nm, filters: BP 420-480, BP 505-600) and the ZEN 2012 software for analysis.

Amyloid inhibitor tests

We infiltrated leaves from 6 to 8 week-old *A. thaliana* ecotype Ws-0 seedlings using a syringe with amyloid inhibitors (Table 2) at a final molarity of 100 µM. The inhibitors were dissolved in 1% DMSO/water (v/v) and as negative control we used 1% DMSO without inhibitor. For the infection of *Albugo* (*A. candida* and *A. laibachii*), we collected spores by submersion of visibly infected leaves in ice-cold sterile water for 1 h. One to two hours after infiltration of the inhibitors, we sprayed this suspension filtered with miracloth using compressed air. We maintained the plants under condensing humidity for two days, of which the first one was at low temperature (7 °C). The plants were kept at short-day conditions afterwards (14 h of darkness at 16 °C and 10 h of light at 21 °C). After 10 days, the rosettes with removed roots were imaged from below. From the visible regions of these, the white pustules were counted, and the leaf area measured using ImageJ.

Protein prediction

We used a pipeline described previously in Frenkel et al. (2022) consisting of two machine-learning programs based on their amino acid sequence: APPNN and AmyloGram (Família et al., 2015; Burdukiewicz et al., 2017). We applied it to *A. candida* strain Nc2 (GeneBank® database assembly accession: GCA_001078535.1). Due to the strong bias of the predictors towards higher length proteins, we applied a penalty score proportional to size. As it was challenging to define a cutoff, we used the median score for the proteome (0.849) as an arbitrary cutoff to classify proteins into positively predicted or not. We labeled the apoplast subset as the high confidence set of proteins reported in Gómez-Pérez et al. (2022) that was obtained by proteomics in *Albugo*-infected and non-infected apoplast extracted from *A. thaliana* plant leaves.

C-DAG expression

We extracted RNA using the RNeasy kit (QIAGEN) from *A. thaliana* leaves infected with *A. candida* strain Nc2 at eight days post infection. Plant growth and infection were performed as described in Gómez-Pérez et al. (2022). We synthesized complementary DNA with SuperScriptTM II Reverse Transcriptase (Invitrogen) from the extracted RNA to use as template for amplification of the candidate sequences. For testing of amyloid formation, we chose the Curli-Dependent Amyloid Generator (C-DAG) system, consisting of the *E. coli* strain VS45 transformed with pVS76 plasmid for overexpression of csgG, the transport system of amyloids, and the pExport plasmid, for overexpression of candidate amyloids (Sivanathan and Hochschild, 2013). The candidate sequences were amplified with Phusion® High-Fidelity DNA polymerase (NEB) and primers designed according to the manufacturer's recommendation (Table 3). We predicted the secretion signal peptide using SignalP version 5 and omitted it from the final construct (Armenteros et al., 2019). For cloning candidates into pExport and transformation into VS45 we followed the protocols suggested in Sivanathan and Hochschild (2013).

We performed colony color phenotype analysis in arabinose-induced plates as described previously, but with 30 µg/mL CR instead of 10 (Sivanathan and Hochschild, 2013). We used the Sup35 prion from yeast with the aggregating domain (Sup35[NM]) as positive control and without (Sup35[M]) as negative control, encoded by the plasmids pVS72 and pVS105, respectively (Sivanathan and Hochschild, 2012). We grew C-DAG-construct expressing C05 on CR plates with 10 µg/mL CR and varying inducing arabinose concentrations (0.02%, 0.2% and 0.8%).

Phylogenetic analysis

We searched for orthologs of the GAS1 gene from *Albugo candida*, C14, in the EGGNOG database, which contains curated gene families of orthologs (Huerta-Cepas et al., 2019). We selected all belonging to oomycetes plus those from *Saccharomyces cerevisiae* as an outgroup. We aligned the proteins using MAFFT (Katoh and Standley, 2013). After, we inferred the phylogenetic tree by maximum likelihood using IQTREE version 2, with default model selection and 1000 fast-bootstrap to get the branch support (Minh et al., 2020).

TEM imaging

The C-DAG constructs were grown in the same way as for the color phenotype assay. We prepared the constructs for observation under the TEM by light touch of the colonies from the inducing CR plates with a mesh copper grid followed by brief drying. We performed a negative stain of the sample with a 30-second aqueous uranyl acetate incubation. The microscope employed for visualization was a JEM-1400 Flash (JEOL, Japan).

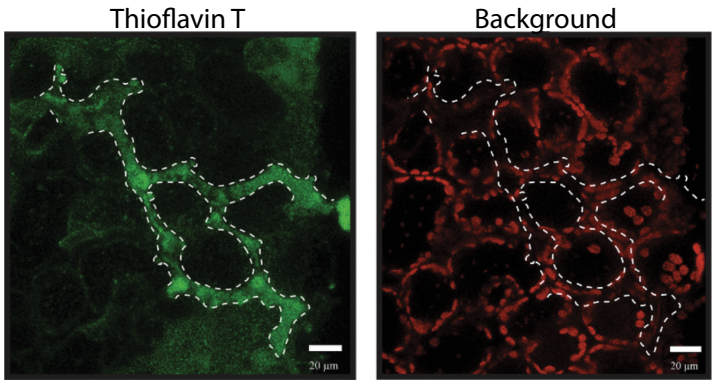


Figure 1. Staining of hyphae with ThT indicates production of amyloids by *Albugo*.
Albugo candida Nc2-infected *Arabidopsis thaliana* leaf stained with amyloid-specific dye Thioflavin T (ThT) shows green fluorescence associated with the hyphal structure. Measured ThT emission maximum at 486 nm (expected emission maximum: 482 nm bound, 438 nm unbound). The white bar corresponds to 20 µm. A dashed white line outlines the border of the hyphae.

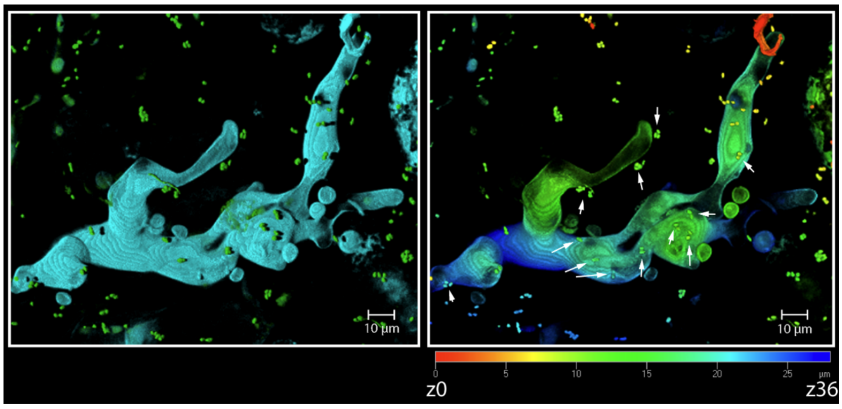


Figure 2. Fluorescence staining of *Albugo* hyphae and surrounding microbial environment.
Fluorescence z-stack projection visualizing the apoplast of *Arabidopsis thaliana* infected with *Albugo candida* strain Nc2. Left image shows *Albugo* hyphae in cyan, stained with Uvitex 2b, and bacterial background in green, stained with Sytox^{TM} green. Right image shows depth coloring, red: z0 (top) to blue: z36 (bottom). Spots of close proximity between hyphae and bacteria are highlighted with white arrows. The white scale bars represent 10 µm.

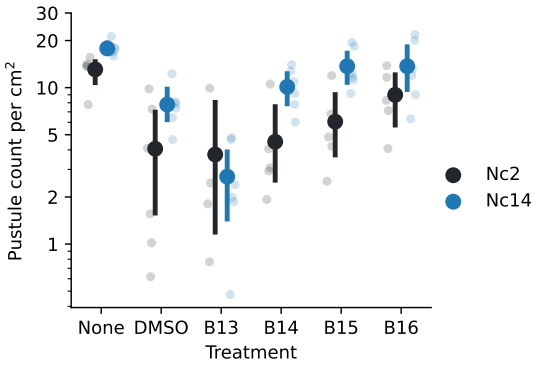


Figure 3. Virulence of *Albugo laibachii* and *Albugo candida* as measured by the number of pustules after amyloid inhibitor treatments.
Count of *A. laibachii* and *A. candida* spore pustules on leaves of *Arabidopsis thaliana* after treatment with a screen of amyloid inhibitors (B13, B14, B15 and B16), and subsequent *A. laibachii* strain Nc14 or *A. candida* strain Nc2 infection. Each dot represents the average of all leaves in a plant. The lines in bold colors represent the 95% confidence interval over six biological replicates per treatment (separate plants), and the middle points, the mean. Water with 1% DMSO (v/v) was used as control treatment. In the treatment labeled none, the plants were not infiltrated prior to infection.

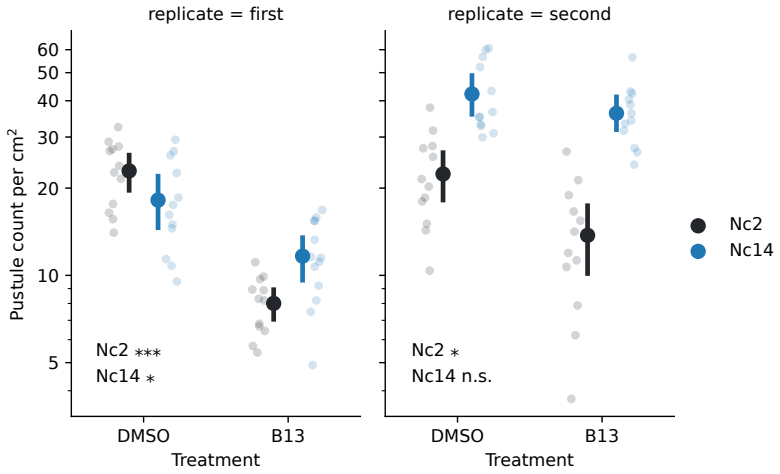


Figure 4. Virulence of *Albugo* strains as measured by the number of pustules after treatment with amyloid inhibitor B13. Count of *Albugo* spore pustules on leaves of *Arabidopsis thaliana* after treatment with amyloid inhibitor B13, followed by *Albugo* infection. We used 12 biological replicates (separate plants) per treatment (inhibitor B13 and control DMSO) and time point (first replicate, second replicate). *A. candida* strain Nc2 is depicted in black and *A. laibachii* strain Nc14 in blue. Each dot represents the average of all leaves in a plant. The lines in bold colors represent the 95% confidence interval over the 12 biological replicates per treatment and the middle points, the mean. Significance was measured using t-test. *n.s.:not significant, *: p value < 0.01, ***: p value < 0.0001*.

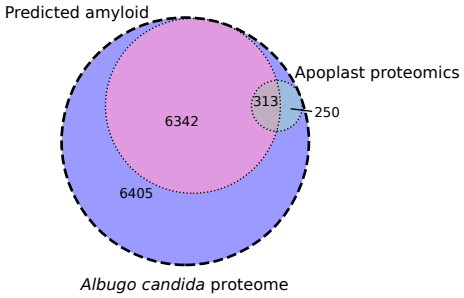


Figure 5. Venn diagram showing apoplast subset and positive amyloid prediction in the putative proteome of *Albugo candida*.
In pink are the in silico predicted amyloids, in blue the proteins that were found in the proteomics analysis (Gómez-Pérez et al., 2022) and in purple the remaining proteins from the proteome of *A. candida* (GeneBank® database assembly accession: GCA_001078535.1).

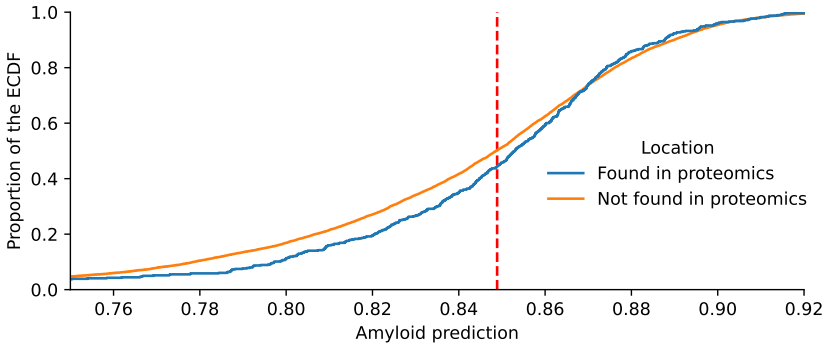


Figure 6. Empirical cumulative distribution function (ECDF) of the amyloid predictions in *Albugo candida*.
The apoplast fraction derived from the proteomics analyses from Gómez-Pérez et al. (2022) is shown as a blue line and the remaining proteome (GeneBank® database assembly accession: GCA_001078535.1) is represented by an orange line. Note the amyloid enrichment scores above the cut-off for the apoplastic proteins. Cut-off, shown as a dashed red line, was defined as the median score of the prediction for the entire putative *A. candida* proteome.

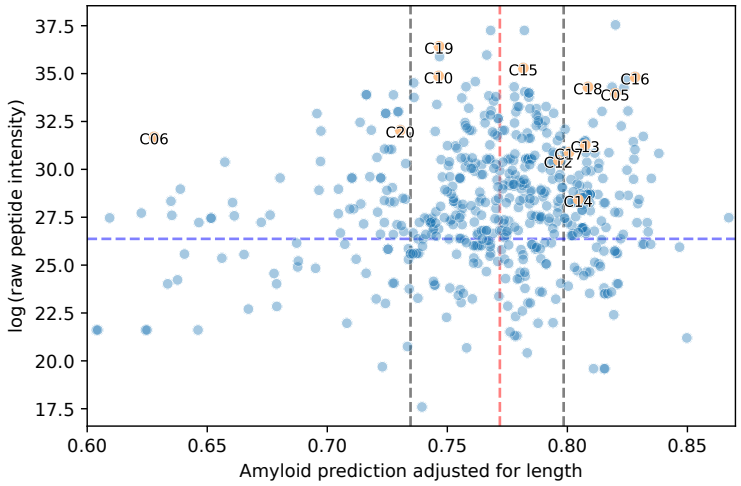


Figure 7. Amyloid prediction of the *Albugo candida* apoplastic proteome together with absolute peptide intensity. Selected candidates for the amyloid screening are labeled. Peptide intensity as measured in the apoplast proteomics analyses (Gómez-Pérez et al., 2022). A red line marks the cut-off for positive amyloid prediction, defined as the median score of the prediction of the entire *A. candida* proteome. On each side, two gray lines mark the first and third quartile for the prediction distributions. The median raw intensity in the dataset is marked with a dashed blue horizontal line.

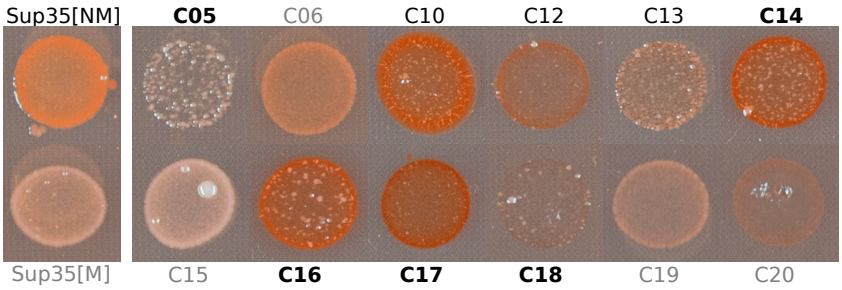


Figure 8. Colony color phenotype of C-DAG strains expressing candidate proteins from *Albugo candida*. The colonies were grown for 4 days with 30 µg/mL Congo Red in the media. The colors of the titles are representing the in silico prediction of amyloid propensity. Light gray represents the negative candidates, dark gray the candidates with scores close to the cut-off, and bold, the candidates with high positive prediction. Positive control strain consists of Sup35 prion from yeast (Sup35[NM]) and negative control of Sup35 without the aggregating domain (Sup35[M]).

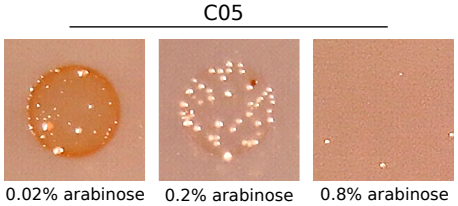


Figure 9. Colony color phenotype of C-DAG construct expressing *Albugo candida* candidate C05. Growth of C05-expressing C-DAG construct at varying concentrations of arabinose. From left to right: 0.02%, 0.2% and 0.8%. Red phenotype is increased at lower inducing concentrations. No growth was detected at higher arabinose concentrations.

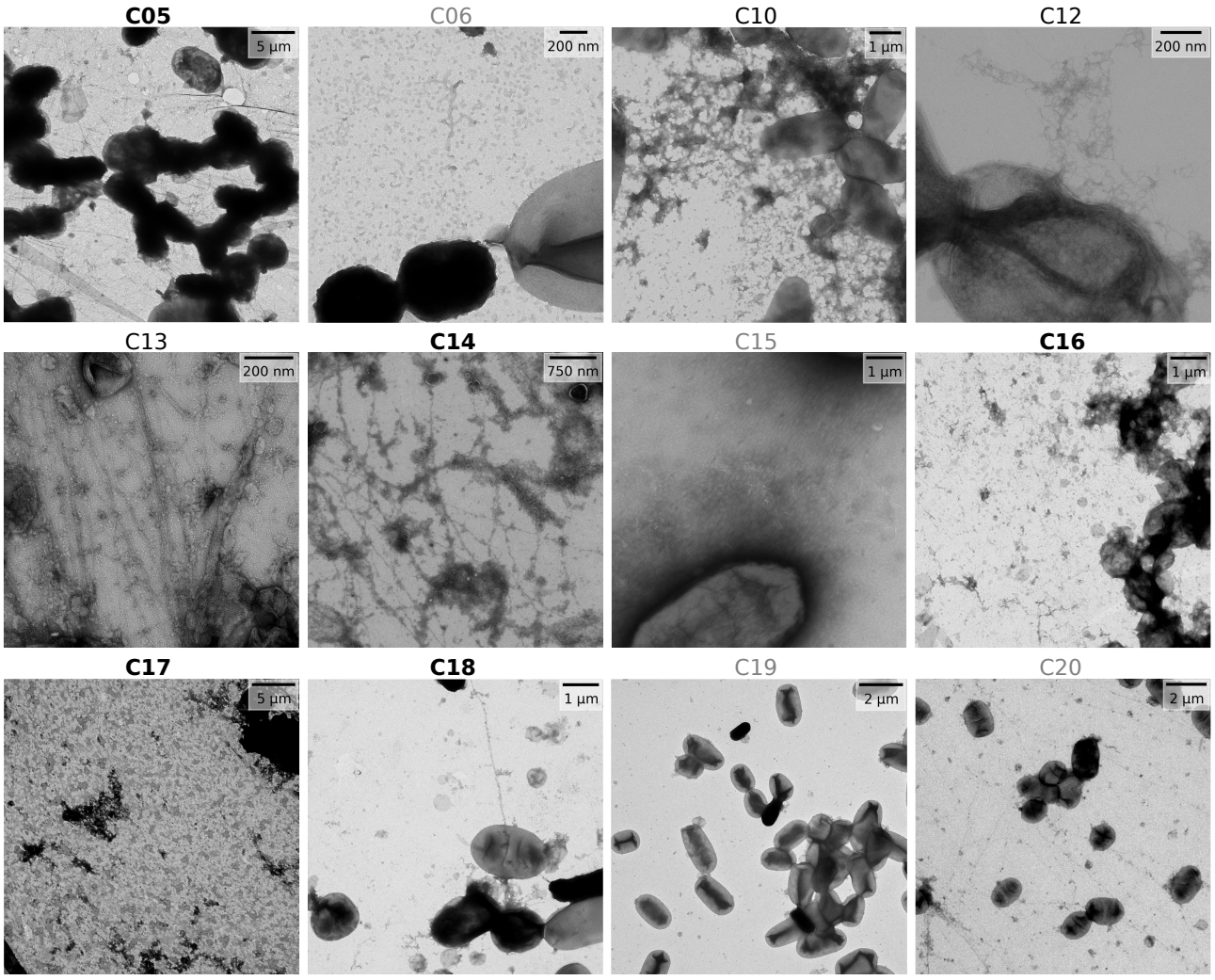


Figure 10. Transmission electron microscopy images (negative staining) of amyloid fibrils from *Albugo candida* candidates expressed in the C-DAG system.

Fibril-like structures appear in the extracellular space of the majority of the positively predicted constructs. Candidate names are displayed on top of each image. The shade of gray represents amyloid prediction score. Light gray shows the negative candidates, dark gray marks candidates with scores close to the cut-off and black marks candidates with high positive prediction. Black bar in the top right corner of each images marks the scale.

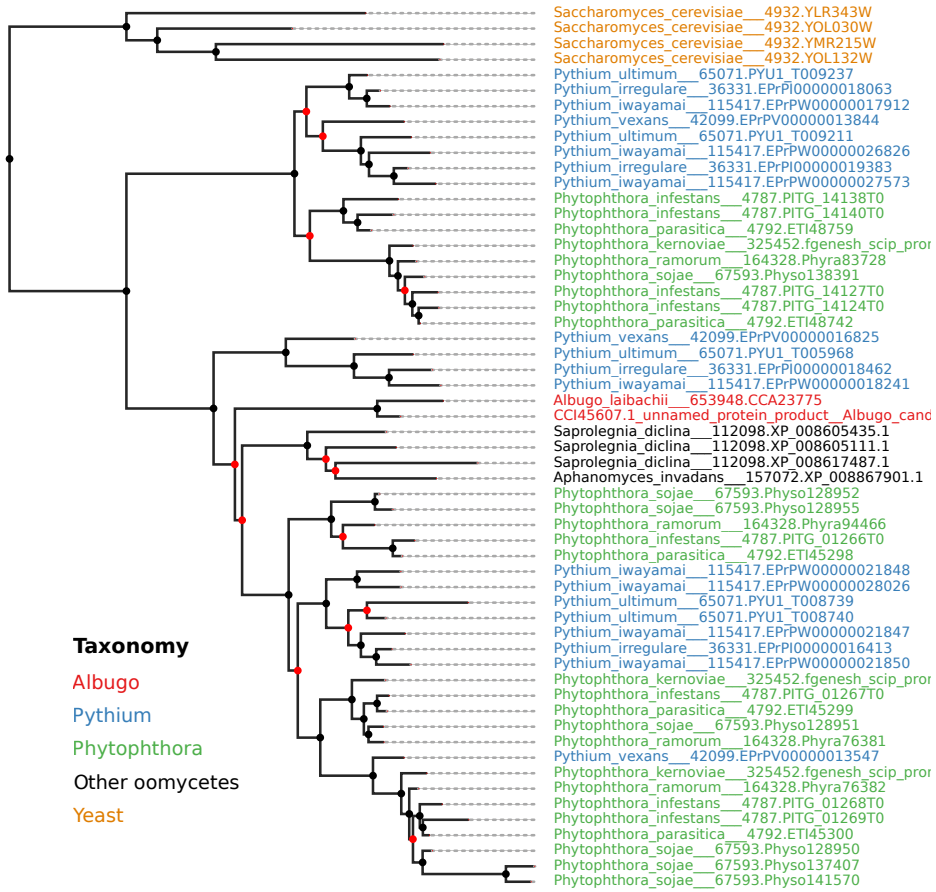


Figure 11. Maximum likelihood tree of beta-1,3-glucosyltranseferases in oomycetes.
The outgroup is composed of GAS1 homologs from *Saccharomyces cerevisiae*. Colored labels represent different genera, *Albugo* in red, *Pythium* in blue and *Phytophthora* in green, other oomycetes are shown in black and the yeast outgroup in orange. Node colors represent bootstrap values, black is higher than 90 and red lower. CC145607.1 unnamed protein product from *Albugo candida* corresponds to C14.

Table 1. Protein candidates from *Albugo candida* selected for their potential amyloid formation based on the in silico predictions. Apoplast abundance obtained from comparison of relative peptide intensity to uninfected apoplast. *NCBI: National Center for Biotechnology Information, IP: InterPro, PTHR: Panther, AA: amino acid.*

Name	Accession (NCBI)	Annotation (according to IP or PTHR)	Amyloid score	Abundance in apoplast	Length (No. AA)	Secretion signal
C05	CCI44519.1	Myo-inositol-1-phosphate synthase	0.82	23.55	516	no
C06	CCI46028.1	-	0.63	21.96	505	no
C10	CCI44717.1	-	0.75	24.17	175	no
C12	CCI46414.1	Elicitin	0.80	21.12	179	yes
C13	CCI47610.1	Glycoside hydrolase, family 5	0.81	21.68	662	yes
C14	CCI45607.1	Glucanosyltransferase	0.80	19.70	463	yes
C15	CCI42480.1	Glucan-1,3-beta-glucosidase	0.83	24.46	459	yes
C16	CCI41093.1	Glycoside hydrolase 131, catalytic N-terminal	0.80	24.13	239	yes
C17	CCI50402.1	Cellulose binding elicitor lectin	0.81	21.41	203	yes
C18	CCI39708.1	Glutathione transferase activity	0.75	23.78	202	no
C19	CCI43949.1	Pectate lyase	0.73	25.25	246	no
C20	CCI46204.1	Xenon atom binding	0.78	22.21	174	no

Table 2. Name and structure of amyloid inhibitors used to infiltrate *Arabidopsis thaliana* prior to infection with *Albugo candida* or *Albugo laibachii*. *tPSA: topological polar surface area*

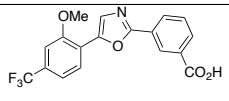
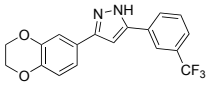
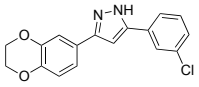
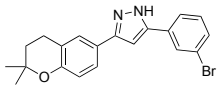
Inhibitor name	Code	Structure	tPSA (Å^2)
sery588	B13		68.12
sery598	B14		42.85
sery599	B15		42.85
sery603	B16		33.62

Table 3. Primer sequences for amplification and cloning via ligation of candidate sequences from *Albugo candida* into vector pExport.

Name	Forward primer (5' to 3')	Reverse primer (5' to 3')
C05	TATAGCGGCCGCAATGACGACAGCATC	TATATCTAGATCAATGGTGGCTTTC
C06	TATAGCGGCCGCAATGTCTGATCGAC	TATATCTAGATCAATTGGAAGTG
C10	TATAGCGGCCGCAATGTCTTATTACGAAC	TATATCTAGACTACTTGAACCAAG
C12	TATAGCGGCCGCAAATTATGCGTGCG	TATATCTAGACTACACCGTTGCG
C13	TATAGCGGCCGCACCTGCCGTTC	TATATCTAGATTAATTTCGACCGATAG
C14	TATAGCGGCCGCAAGCCCGATTACAATTC	TATATCTAGATTAGCAATTTGACTTTTTG
C15	TATAGCGGCCGCATCGGCTCAATCACC	TATATCTAGATTAAAAAACTCCTTTTTC
C16	TATAGCGGCCGCACCCAATTTCCACATGCC	TATATCTAGATTACCCTTCCGCACTTG
C17	TATAGCGGCCGCACGAGAGCCTCAATG	TATATCTAGACTACTGCAAACTCGG
C18	TATAGCGGCCGCAATGTCACCACAACCC	TATATCTAGATCAGCCTTTTCTGC
C19	TATAGCGGCCGCAATGCTGATGAAACAAC	TATATCTAGATCAGACATATGTGACGG
C20	TATAGCGGCCGCAATGCTGAAAGCGTTTTC	TATATCTAGACTATTCTACCTCAACG

Bibliography

Daniel Otzen and Roland Riek. Functional Amyloids. *Cold Spring Harbor Perspectives in Biology*, 11(12):a033860, 2019. ISSN 1943-0264. doi: 10.1101/cshperspect.a033860.

Thomas R. Jahn, O. Sumner Makin, Kyle L. Morris, Karen E. Marshall, Pei Tian, Pawel Sikorski, and Louise C. Serpell. The Common Architecture of Cross-β Amyloid. *Journal of Molecular Biology*, 395(4):717–727, 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2009.09.039.

Paramita Chaudhuri, Kailash P. Prajapati, Bibin G. Anand, Kriti Dubey, and Karunakar Kar. Amyloid cross-seeding raises new dimensions to understanding of amyloidogenesis mechanism. *Ageing Research Reviews*, 56:100937, 2019. ISSN 1568-1637. doi: 10.1016/j.arr.2019.100937.

Magdalena I. Ivanova, Yuxi Lin, Young-Ho Lee, Jie Zheng, and Ayyalusamy Ramamoorthy. Biophysical processes underlying cross-seeding in amyloid aggregation and implications in amyloid pathology. *Biophysical Chemistry*, 269:106507, 2021. ISSN 0301-4622. doi: 10.1016/j.bpc.2020.106507.

Ingrid Macindoe, Ann H. Kwan, Qin Ren, Vanessa K. Morris, Wenrong Yang, Joel P. Mackay, and Margaret Sunde. Self-assembly of functional, amphipathic amyloid monolayers by the fungal hydrophobin EAS. *Proceedings of the National Academy of Sciences*, 109(14):E804–E811, 2012. ISSN 0027-8424. doi: 10.1073/pnas.1114052109.

Elliot Erskine, Cait E. MacPhee, and Nicola R. Stanley-Wall. Functional Amyloid and Other Protein Fibers in the Biofilm Matrix. *Journal of Molecular Biology*, 430(20):3642–3656, 2018. ISSN 0022-2836. doi: 10.1016/j.jmb.2018.07.026.

Bruce L. Kagan, Hyunbum Jang, Ricardo Capone, Fernando Teran Arce, Srinivasan Ramachandran, Ratnesh Lal, and Ruth Nussinov. Antimicrobial Properties of Amyloid Peptides. *Molecular Pharmaceutics*, 9(4):708–717, 2012. ISSN 1543-8384. doi: 10.1021/mp200419b.

Jixi Li, Thomas McQuade, Ansgar B. Siemer, Johanna Napetschnig, Kenta Moriwaki, Yu-Shan Hsiao, Ermelinda Damko, David Moquin, Thomas Walz, Ann McDermott, Francis Ka-Ming Chan, and Hao Wu. The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis. *Cell*, 150(2):339–350, 2012. ISSN 0092-8674. doi: 10.1016/j.cell.2012.06.019.

Seong-Cheol Park, Yoonkyung Park, and Kyung-Soo Hahm. The Role of Antimicrobial Peptides in Preventing Multidrug-Resistant Bacterial Infections and Biofilm Formation. *International Journal of Molecular Sciences*, 12(9):5971–5992, 2011. doi: 10.3390/ijms12095971.

Chun Wu, Zhixiang Wang, Hongxing Lei, Wei Zhang, and Yong Duan. Dual Binding Modes of Congo Red to Amyloid Protofibril Surface Observed in Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 129(5):1225–1232, 2007. ISSN 0002-7863. doi: 10.1021/ja0662772.

Sally L. Gras, Lynne J. Waddington, and Kenneth N. Goldie. Protein Folding, Misfolding, and Disease, Methods and Protocols. *Methods in Molecular Biology*, 752:197–214, 2011. ISSN 1064-3745. doi: 10.1007/978-1-60327-223-0_13.

Viknesh Sivanathan and Ann Hochschild. A bacterial export system for generating extracellular amyloid aggregates. *Nature Protocols*, 8(7):1381–1390, 2013. ISSN 1754-2189. doi: 10.1038/nprot.2013.081.

Nani Van Gerven, Sander E. Van der Verren, Dirk M. Reiter, and Han Remaut. The Role of Functional Amyloids in Bacterial Virulence. *Journal of Molecular Biology*, 430(20):3657–3684, 2018. ISSN 0022-2836. doi: 10.1016/j.jmb.2018.07.010.

Daniel Gómez-Pérez, Vasvi Chaudhry, Ariane Kemen, and Eric Kemen. Amyloid Proteins in Plant-Associated Microbial Communities. *Microbial Physiology*, 31(2):88–98, 2021. ISSN 2673-1665. doi: 10.1159/000516014.

Yizhou Zhou, Daniel Smith, Bryan J. Leong, Kristoffer Brännström, Fredrik Almqvist, and Matthew R. Chapman. Promiscuous Cross-seeding between Bacterial Amyloids Promotes Interspecies Biofilms. *Journal of Biological Chemistry*, 287(42):35092–35103, 2012. ISSN 0021-9258. doi: 10.1074/jbc.m112.383737.

Nirukshan Shanmugam, Max O. D. G. Baker, Sarah R. Ball, Megan Steain, Chi L. L. Pham, and Margaret Sunde. Microbial functional amyloids serve diverse purposes for structure, adhesion and defence. *Biophysical Reviews*, 11(3):287–302, 2019. ISSN 1867-2450. doi: 10.1007/s12551-019-00526-1.

Anna Mathew, Vignesh Balaji E, Sreedhara Ranganath K Pai, Anoop Kishore, Vasudev Pai, Ramadevi Pemmireddy, and Chandrashekar K S. Current Drug Targets in Alzheimer's Associated Memory Impairment: A Comprehensive Review. *CNS & Neurological Disorders - Drug Targets*, 21, 2022. ISSN 1871-5273. doi: 10.2174/1871527321666220401124719.

Diego Romero, Edgardo Sanabria-Valentín, Hera Vlamakis, and Roberto Kolter. Biofilm Inhibitors that Target Amyloid Proteins. *Chemistry & Biology*, 20(1):102–110, 2013. ISSN 1074-5521. doi: 10.1016/j.chembiol.2012.10.021.

Jonghee Oh, Jung-Gun Kim, Eunkyung Jeon, Chang-Hyuk Yoo, Jae Sun Moon, Sangkee Rhee, and Ingyu Hwang. Amyloidogenesis of Type III-dependent Harpins from Plant Pathogenic Bacteria*. *Journal of Biological Chemistry*, 282(18):13601–13609, 5 2007. ISSN 0021-9258. doi: 10.1074/jbc.m602576200.

Olaf Müller, Peter H. Schreier, and Joachim F. Uhrig. Identification and characterization of secreted and pathogenesis-related proteins in Ustilago maydis. *Molecular Genetics and Genomics*, 279(1):27–39, 2008. ISSN 1617-4615. doi: 10.1007/s00438-007-0291-4.

Matthew T. Agler, Jonas Ruhe, Samuel Kroll, Constanze Morhenn, Sang-Tae Kim, Detlef Weigel, and Eric M. Kemen. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biology*, 14(1):e1002352, 2016. ISSN 1544-9173. doi: 10.1371/journal.pbio.1002352.

Eric Kemen, Anastasia Gardiner, Torsten Schultz-Larsen, Ariane C. Kemen, Alexi L. Balmuth, Alexandre Robert-Seilaniantz, Kate Bailey, Eric Holub, David J. Studholme, Dan MacLean, and Jonathan D. G. Jones. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. *PLoS Biology*, 9(7):e1001094, 2011. ISSN 1544-9173. doi: 10.1371/journal.pbio.1001094.

N J Talbot, M J Kershaw, G E Wakley, OMH De Vries, JGH Wessels, and J E Hamer. MPG1 Encodes a Fungal Hydrophobin Involved in Surface Interactions during Infection-Related Development of Magnaporthe grisea. *The Plant Cell*, 8(6):985–999, 1996. ISSN 1040-4651. doi: 10.1105/tpc.8.6.985.

Eric Kemen, Ariane Kemen, Andreas Ehlers, Ralf Voegele, and Kurt Mendgen. A novel structural effector from rust fungi is capable of fibril formation. *The Plant Journal*, 75(5):767–780, 2013. ISSN 1365-313X. doi: 10.1111/tpj.12237.

Shigeyuki Tanaka and Regine Kahmann. Cell wall–associated effectors of plant-colonizing fungi. *Mycologia*, 113(2):1–14, 2021. ISSN 0027-5514. doi: 10.1080/00275514.2020.1831293.

Mohammad Shahnawaz and Claudio Soto. Microcin Amyloid Fibrils A Are Reservoir of Toxic Oligomeric Species. *Journal of Biological Chemistry*, 287(15):11665–11676, 2012. ISSN 0021-9258. doi: 10.1074/jbc.m111.282533.

Daniel Gómez-Pérez, Monja Schmid, Vasvi Chaudhry, Ana Velic, Boris Maček, Ariane Kemen, and Eric Kemen. Proteins released into the plant apoplast by the obligate parasitic protist albugo selectively repress phyllosphere-associated bacteria. *bioRxiv*, 2022. doi: 10.1101/2022.05.16.492175.

Mihaly Varadi, Greet De Baets, Wim F. Vranken, Peter Tompa, and Rita Pancsa. AmyPro: a database of proteins with validated amyloidogenic regions. *Nucleic Acids Research*, 46(Database issue):gkx950–, 2017. ISSN 1362-4962. doi: 10.1093/nar/gkx950.

Antoine Loquet and Sven J. Saupe. Diversity of Amyloid Motifs in NLR Signaling in Fungi. *Biomolecules*, 7(2):38, 2017. doi: 10.3390/biom7020038.

Anton A. Nizhnikov, Tatyana A. Ryzhova, Kirill V. Volkov, Sergey P. Zadorsky, Julia V. Sopova, Sergey G. Inge-Vechtomov, and Alexey P. Galkin. Interaction of Prions Causes Heritable Traits in Saccharomyces cerevisiae. *PLoS Genetics*, 12(12):e1006504, 2016. ISSN 1553-7390. doi: 10.1371/journal.pgen.1006504.

Nathania Puspitasari and Cheng-Kang Lee. Class I hydrophobin fusion with cellulose binding domain for its soluble expression and facile purification. *International Journal of Biological Macromolecules*, 193(Pt A):38–43, 2021. ISSN 0141-8130. doi: 10.1016/j.ijbiomac.2021.10.089.

Daniel Gómez-Pérez and Eric Kemen. Predicting lifestyle and host from positive selection data and genome properties in oomycetes. *bioRxiv*, page 2021.01.12.426341, 2021. doi: 10.1101/2021.01.12.426341.

Peter N. Lipke, Melissa C. Garcia, David Alsteens, Caleen B. Ramsook, Stephen A. Klotz, and Yves F. Dufrêne. Strengthening relationships: amyloids create adhesion nanodomains in yeasts. *Trends in Microbiology*, 20(2):59–65, 2012. ISSN 0966-842X. doi: 10.1016/j.tim.2011.10.002.

Wieke R. Teertstra, Gisela J. van der Velden, Jan F. de Jong, John A.W. Kruijtzer, Rob M.J. Liskamp, Loes M.J. Kroon-Batenburg, Wally H. Müller, Martijn F.B.G. Gebbink, and Han A.B. Wösten. The Filament-specific Rep1-1 Repellent of the Phytopathogen Ustilago maydis Forms Functional Surface-active Amyloid-like Fibrils*. *Journal of Biological Chemistry*, 284(14):9153–9159, 2009. ISSN 0021-9258. doi: 10.1074/jbc.m900095200.

Tanja R Scheublin, Ian R Sanders, Christoph Keel, and Jan Roelof van der Meer. Characterisation of microbial communities colonising the hyphal surfaces of arbuscular mycorrhizal fungi. *The ISME Journal*, 4(6):752–763, 2010. ISSN 1751-7362. doi: 10.1038/ismej.2010.5.

Breanne N. Steffan, Nandhitha Venkatesh, and Nancy P. Keller. Let's Get Physical: Bacterial-Fungal Interactions and Their Consequences in Agriculture and Health. *Journal of Fungi*, 6(4):243, 2020. doi: 10.3390/jof6040243.

M. Ankarcrona, B. Winblad, C. Monteiro, C. Fearns, E. T. Powers, J. Johansson, G. T. Westermark, J. Presto, B.-G. Ericzon, and J. W. Kelly. Current and future treatment of amyloid diseases. *Journal of Internal Medicine*, 280(2):177–202, 2016. ISSN 1365-2796. doi: 10.1111/joim.12506.

Georgia I. Nasi, Foteini D. Aktypi, Panagiotis M. Spatharas, Nikolaos N. Louros, Paraskevi L. Tsiolaki, Vassiliki Magafa, Ioannis P. Trougakos, and Vassiliki A. Iconomidou. Arabidopsis thaliana Plant Natriuretic Peptide Active Domain Forms Amyloid-like Fibrils in a pH-Dependent Manner. *Plants*, 11(1):9, 2021. ISSN 2223-7747. doi: 10.3390/plants11010009.

D. Santi Swarupini and Abani K. Bhuyan. Amyloid fibrillation of an intrinsically disordered plant phloem protein AtPP16-1 under acidic condition. *Biophysical Chemistry*, 237:1–8, 2018. ISSN 0301-4622. doi: 10.1016/j.bpc.2018.03.004.

K. S. Antonets and A. A. Nizhnikov. Amyloids and prions in plants: Facts and perspectives. *Prion*, 11(5):300–312, 2017. ISSN 1933-6896. doi: 10.1080/19336896.2017.1377875.

Tatyana S. Kalebina, Tatyana A. Plotnikova, Anton A. Gorkovskii, Irina O. Selyakh, Oxana V. Galzitskaya, Evgeniy E. Bezsonov, Gerd Gellissen, and Igor S. Kulaev. Amyloid-like properties of Saccharomyces cerevisiae cell wall glucantransferase Bgl2p. *Prion*, 2(2):91–96, 2014. ISSN 1933-6896. doi: 10.4161/pri.2.2.6645.

A.V. Sergeeva, J.V. Sopova, T.A. Belashova, V.A. Siniukova, A.V. Chirinskaite, A.P. Galkin, and S.P. Zadorsky. Amyloid properties of the yeast cell wall protein Toh1 and its interaction with prion proteins Rnq1 and Sup35. *Prion*, 13(1):21–32, 2018. ISSN 1933-6896. doi: 10.1080/19336896.2018.1558763.

Tatyana A. Ryzhova, Julia V. Sopova, Sergey P. Zadorsky, Vera A. Siniukova, Aleksandra V. Sergeeva, Svetlana A. Galkina, Anton A. Nizhnikov, Aleksandr A. Shenfeld, Kirill V. Volkov, and Alexey P. Galkin. Screening for amyloid proteins in the yeast proteome. *Current Genetics*, 64(2):469–478, 2018. ISSN 0172-8083. doi: 10.1007/s00294-017-0759-7.

Enrico Ragni, Thierry Fontaine, Carmela Gissi, Jean Paul Latgè, and Laura Popolo. The Gas family of proteins of Saccharomyces cerevisiae: characterization and evolutionary analysis. *Yeast*, 24(4):297–308, 2007. ISSN 1097-0061. doi: 10.1002/yea.1473.

Melissa R. Koch and Lorraine Pillus. The glucanosyltransferase Gas1 functions in transcriptional silencing. *Proceedings of the National Academy of Sciences*, 106(27):11224–11229, 2009. ISSN 0027-8424. doi: 10.1073/pnas.0900809106.

Shizhu Zhang, Yuxian Xia, and Nemat O. Keyhani. Contribution of the gas1 Gene of the Entomopathogenic Fungus Beauveria bassiana, Encoding a Putative Glycosylphosphatidylinositol-Anchored β-1,3-Glucanosyltransferase, to Conidial Thermotolerance and Virulence. *Applied and Environmental Microbiology*, 77(8):2676–2684, 2011. ISSN 0099-2240. doi: 10.1128/aem.02747-10.

Yuki Imai, Takafumi Shimasaki, Chihiro Enokimura, Hokuto Ohtsuka, Satoshi Tsubouchi, Kunio Ihara, and Hirofumi Aiba. gas1 mutation extends chronological lifespan via Pmk1 and Sty1 MAPKs in Schizosaccharomyces pombe. *Bioscience, Biotechnology, and Biochemistry*, 84(2):1–8, 2019. ISSN 0916-8451. doi: 10.1080/09168451.2019.1676695.

Alona Frenkel, Eli Zecharia, Daniel Gomez-Perez, Yevgeni Yegorov, Eleonora Sendersky, Avi Jacobs, Jennifer Benichou, York-Dieter Stierhof, Rami Parnasa, Susan S. Golden, Eric Kemen, and Rakefet Schwarz. Cell specialization in cyanobacterial biofilm development revealed by expression of a cell-surface and extracellular matrix protein. *bioRxiv*, 2022. doi: 10.1101/2022.07.13.498973.

Carlos Família, Sarah R. Dennison, Alexandre Quintas, and David A. Phoenix. Prediction of Peptide and Protein Propensity for Amyloid Formation. *PLoS ONE*, 10(8):e0134679, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0134679. APPNN software paper.

Michał Burdukiewicz, Piotr Sobczyk, Stefan Rödiger, Anna Duda-Madej, Paweł Mackiewicz, and Małgorzata Kotulska. Amyloidogenic motifs revealed by n-gram analysis. *Scientific Reports*, 7(1):12961, 2017. doi: 10.1038/s41598-017-13210-9. Amylogram software paper.

José Juan Almagro Armenteros, Konstantinos D. Tsirigos, Casper Kaae Sønderby, Thomas Nordahl Petersen, Ole Winther, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4):420–423, 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0036-z.

Viknesh Sivanathan and Ann Hochschild. Generating extracellular amyloid aggregates using E. coli cells. *Genes & Development*, 26(23):2659–2667, 2012. ISSN 0890-9369. doi: 10.1101/gad.205310.112.

Jaime Huerta-Cepas, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, Ivica Letunic, Thomas Rattei, Lars J Jensen, Christian von Mering, and Peer Bork. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47 (Database issue):D309–D314, 1 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1085.

Kazutaka Katoh and Daron M Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–80, 1 2013. ISSN 0737-4038. doi: 10.1093/molbev/mst010.

Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.

Data availability

All data discussed in this paper as well as the code to reproduce the analyses and figures is provided at reasonable request.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to acknowledge support from the graduate school GRK 1708 “Molecular principles of bacterial survival strategies” and the European Research Council under the DeCoCt research program. Additionally, we would like to thank Ekaterina Batagova for her help during the inhibitor experiments.