Please cite this article in press as: Colaianni et al., A complex immune response to flagellin epitope variation in commensal communities, Cell Host & Microbe (2021), https://doi.org/10.1016/j.chom.2021.02.006

A complex immune response to flagellin epitope variation in commensal communities

Nicholas R. Colaianni,[1,2,3,6] Katarzyna Parys,[4,6,7] Ho-Seok Lee,[4,6] Jonathan M. Conway,[1,3] Nak Hyun Kim,[1,3] Natalie Edelbacher,[4] Tatiana S. Mucyn,[1,3] Mathias Madalinski,[4] Theresa F. Law,[1,3] Corbin D. Jones,[1,2,5,*] Youssef Belkhadir,[4,*] and Jeffery L. Dangl[1,2,3,8,*]

[1]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Howard Hughes Medical Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[4]Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna, Austria
[5]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[6]These authors contributed equally
[7]Present address: Faculty of Biology, Genetics, University of Munich (LMU), 82152 Martinsried, Germany
[8]Lead contact
*Correspondence: cdjones@email.unc.edu (C.D.J.), youssef.belkhadir@gmi.oeaw.ac.at (Y.B.), dangl@email.unc.edu (J.L.D.)
https://doi.org/10.1016/j.chom.2021.02.006

SUMMARY

Immune systems restrict microbial pathogens by identifying "non-self" molecules called microbe-associated molecular patterns (MAMPs). It is unclear how immune responses are tuned to or by MAMP diversity present in commensal microbiota. We systematically studied the variability of commensal peptide derivatives of flagellin (flg22), a MAMP detected by plants. We define substantial functional diversity. Most flg22 peptides evade recognition, while others contribute to evasion by manipulating immunity through antagonism and signal modulation. We establish a paradigm of signal integration, wherein the sequential signaling outputs of the flagellin receptor are separable and allow for reprogramming by commensal-derived flg22 epitope variants. Plant-associated communities are enriched for immune evading flg22 epitopes, but upon physiological stress that represses the immune system, immune-activating flg22 epitopes become enriched. The existence of immune-manipulating epitopes suggests that they evolved to either communicate or utilize the immune system for host colonization and thus can influence commensal microbiota community composition.

INTRODUCTION

Plants and animals are constantly surveilling their extracellular environment for the presence of external threats, environmental conditions, and cell-derived messages. One method of surveillance utilizes pattern recognition receptors (PRRs) to monitor the extracellular space for ligands that reveal microbiome conditions (Belkhadir et al., 2014; Steinbrenner, 2020; Ronald and Beutler, 2010). During microbial colonization, diverse microbial derived protein fragments called microbial associated molecular patterns (MAMPs) act as ligands for PRRs (Ronald and Beutler, 2010; Boutrot and Zipfel, 2017; Zhou and Zhang, 2020). In plants, MAMP recognition leads to an immune response capable of halting microbial growth, termed MAMP-triggered immunity (MTI) (Li et al., 2005; Hacquard et al., 2017). The analysis of bacterial genomes from plant-associated isolates indicate that commensal bacteria produce MAMPs that can be detected by cognate plant immune receptors (Garrido-Oter et al., 2018). However, commensals also produce diverse MAMPs (Teixeira et al., 2019). Thus, furthering our understanding of MAMP variability and functional consequences will improve our understanding of how plant-associated microbial communities persist in the face of an immune response adapted to restrict microbial growth.

One of the most well studied MAMPs is a peptide derived from the abundant bacterial flagellum protein, FliC (Felix et al., 1999; Fliegmann and Felix, 2016; Boutrot and Zipfel, 2017). In the reference plant *Arabidopsis thaliana* (hereafter Arabidopsis), a 22-amino-acid epitope of FliC, termed flg22 is required for bacterial motility, but is also sufficient to induce MTI (Naito et al., 2008; K.P. et al., unpublished data; Gómez-Gómez et al., 1999). Flg22 peptides can be generated by an unidentified protease that degrades deglycosylated FliC proteins (Buscaill et al., 2019). On the cell surface, the recognition of and response to flg22 in Arabidopsis is mediated by a two-receptor system. First, the flg22 peptide binds to FLAGELLIN SENSING 2 (FLS2) (Chinchilla et al., 2006). The C-terminal region of the flg22 peptide then induces the binding of FLS2 to a co-receptor BRI1-ASSOCIATED KINASE 1 (BAK1) (Chinchilla et al., 2007). This binding strategy defines the "address-message" concept of flg22 recognition: the "address" is the N-terminal portion of the peptide that defines its ability to bind FLS2, while the "message" is the C-terminal portion of the peptide that defines the formation of the FLS2-BAK1 complex. FLS2-BAK1 complex formation

drives subsequent signaling that leads to functional MTI (Meindl et al., 2000; Sun et al., 2013).

The formation of the flg22-FLS2-BAK1 complex triggers a phosphorylation cascade that results in a set of temporally distinct responses. Within minutes, downstream signaling components are phosphorylated, including BOTRYTIS-INDUCED KINASE 1 (BIK1), which directly activates the production of a reactive oxygen species (ROS) burst by phosphorylating the NADPH oxidase RbohD (Lu et al., 2009; Lin et al., 2014; Kadota et al., 2014). Flg22 also promotes a MAP kinase cascade that leads to rapid transcriptional reprogramming, indicative of MTI activation (Asai et al., 2002). Temporally later responses were identified from chronic exposure to immunogenic flg22. One of these chronic responses is seedling growth inhibition (SGI), illustrated by stunted seedling growth due to a growth-defense trade-off (Gómez-Gómez et al., 1999; Belkhadir et al., 2012; Albrecht et al., 2012).

FLS2 responses induced by flg22 in Arabidopsis roots are not only cell type specific but also require cellular damage for maximal output (Zhou et al., 2020; Emonet et al., 2020; Rich-Griffin et al., 2020). These distinct flg22-induced responses indicate that FLS2 activities are adaptable and carefully deployed. This is supported by the specificity of immune elicitation across MAMPs (Vetter et al., 2016). To date, studies on the immune responses produced by a particular MAMP have mostly focused on a handful of epitope variants derived from pathogens and beneficial bacteria. Thus, there is a major gap in our understanding of the spectrum of responses engendered by natural MAMP variation present in commensal communities, and the breadth of responses to MAMPs, like those engendered by flg22, remains undefined.

Here, we leveraged advancements in the genomic characterization of hundreds of Arabidopsis commensal bacterial strains to address these knowledge gaps. We identified and screened, at previously unprecedented scales, the natural flg22 diversity present in commensal communities. First, we generated a community accessible database of culturable root and leaf commensal bacteria all derived from Arabidopsis to query flg22 peptide diversity. Second, we screened >90 of these flg22 variants for the production of MTI responses. We found that commensal bacteria encode significant flg22 sequence diversity and that a majority of this diversity results in flg22 peptides that evade defense activation. Among the evading flg22 variants identified in our primary screen, we identified flg22 peptides that antagonize the formation of the FLS2-BAK1 complex induced by active flg22 peptides. In the process of characterizing FLS2 antagonists, we also identified flg22 variants that prime the early temporal responses to subsequent flg22 elicitation but alter temporally late responses. We demonstrate that antagonistic flg22 peptides are found in prevalent commensal community members and that a commensal strain functionally antagonizes the canonical flg22 response *in vivo*. We uncover flg22 variants that uncouple FLS2-BAK1-mediated signal transduction pathways. We also observe that evading variants are dominant in a large synthetic community experiment but that immune-activating flg22 variants are enriched under an abiotic stress condition that represses the immune system. Thus, our data demonstrate that there is pervasive evasion of immune activation through reducing FLS2 activation and through the manipulation of immune outputs by receptor antagonism or modulation. This work provides a framework to better understand how commensal microbiomes assemble in the interface of immune surveillance.

RESULTS

Arabidopsis commensal bacteria encode substantial variation in the flg22 epitope of FliC

To identify the flg22 diversity present in natural communities, we built a database of 627 genomes derived from bacteria isolated from healthy Arabidopsis plants (Table S1; Bai et al., 2015; Levy et al., 2017). We identified 779 FliC proteins from 61% of the microbial genomes in the Arabidopsis database (Figure 1A). We created a phylogenetic tree using primary sequence similarity and classified the FliCs into 3 clades (Figures 1A and S1A). We found that FliC clade assignment was indicative of taxonomic origin. Clade 1 FliCs are primarily found in Betaproteobacteria, Gammaproteobacteria, and Sphingomonadales; clade 2 FliCs are primarily found in Bacilli and Actinobacteria; and clade 3 FliCs are primarily found in Alphaproteobacteria, particularly Rhizobiales and Caulobacterales. The average primary sequence identity of FliCs within clades 1, 2, and 3 was 47%, 58%, and 41%, respectively (Figure 1A). Bacterial motility has been demonstrated for bacteria that encode all three of these FliC clades (Kühn et al., 2018; Clarke et al., 2013; Haiko and Westerlund-Wikström, 2013; Fujii et al., 2008; Iida et al., 2009), suggesting that adaptive levels of bacterial motility can be retained across the broad flagellin sequence divergence observed here.

To investigate whether the diversity of FliC proteins is reflected in flg22 sequence variation, we extracted 268 unique flg22 sequences from the FliC proteins in the database. The flg22 sequences from clades 1 and 2 resemble the canonical, MTI active flg22 variant from *Pseudomonas aeruginosa* (Pa22) (Figures 1C and 1D). The clade 3 flg22 sequences were the most divergent from Pa22 (Figure 1E). Importantly, residues interacting with FLS2, Pa22 Asp^{15} and Pa22 Gly^{18}, and BAK1, Pa22 Leu^{19} in the Pa22-BAK1-FLS2 crystal structure are commonly altered in clade 3 flg22 peptides (Sun et al., 2013, Figure 1E). This suggests that the natural diversification of flg22 sequences in clade 3 FliC proteins occurs on residues that are critical for detection by the plant immune system. We posit that flg22 sequence diversity in clade 3 has functional consequences for immunogenicity. This proposition is consistent with results from a large-scale synthetic analysis of flg22 variation detailed in K.P. et al., unpublished data.

Within the clade 3 FliCs, ~50% have a unique flg22 sequence. This is twice the amount of diversity found in clades 1 and 2, in which only ~25% FliC proteins have unique flg22 sequences. To understand how clade 3 FliCs became so diverse, we investigated *fliC* copy number (Figures 1B and S1B). We found a median of two *fliC* genes in genomes with clade 3 FliCs, while only a median of one for clades 1 and 2 (Figure 1B). Up to six *fliC* gene products can be incorporated into the flagellum (Iida et al., 2009). Within genomes containing multiple *fliC* genes, we identified operons where *fliC* genes occur within 10 kB of one another, suggesting that *fliC* expansion is the result of gene duplication (Figure S1B). Thus, diversification of the flg22 region in clade 3, especially the prominent changes of Asp^{15}, Gly^{18}, and Leu^{19},

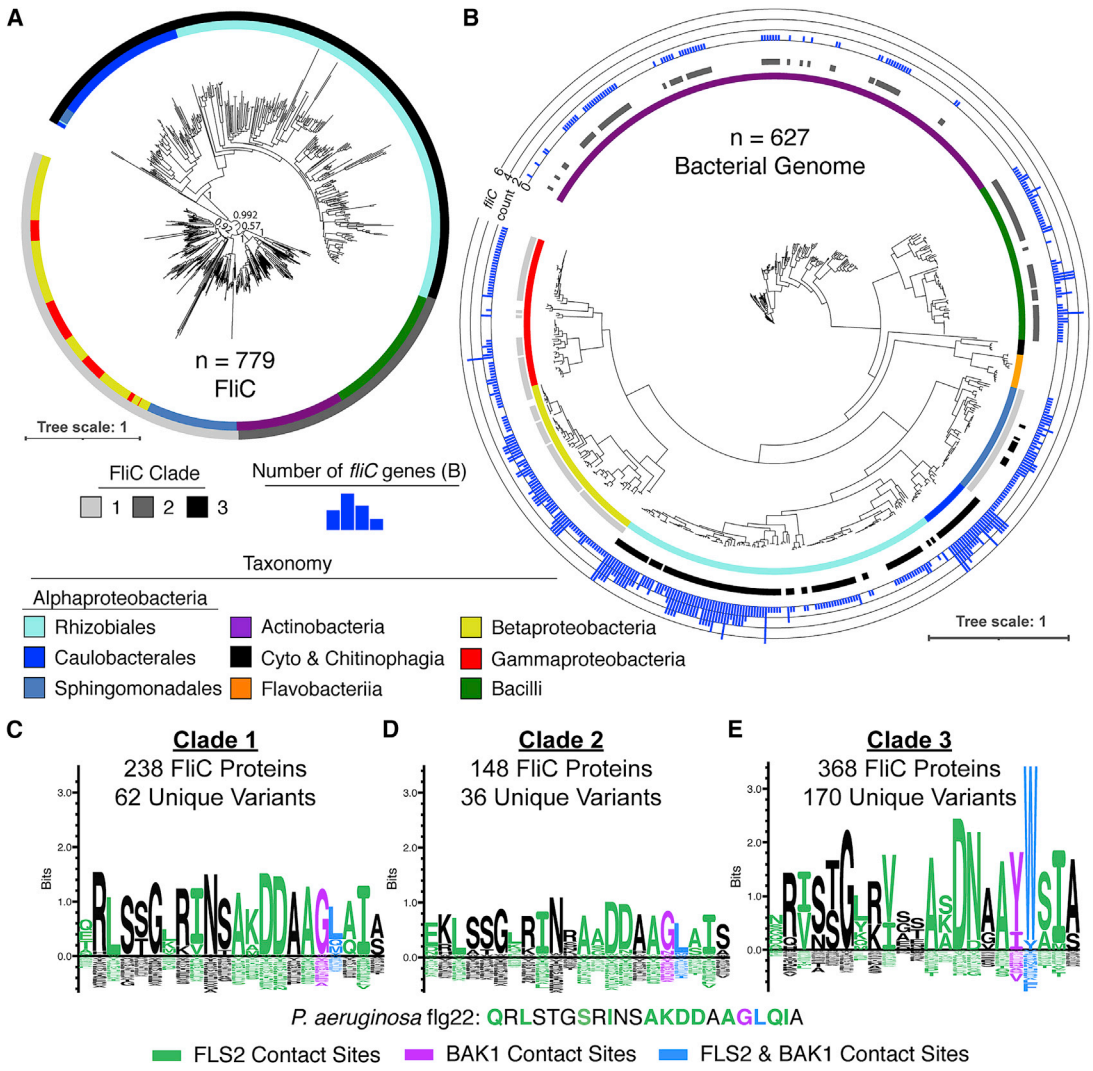

Figure 1. Taxonomically distinct FliC proteins contain three distinct clades of flg22 sequences

(A) An illustrative tree produced using 779 high-quality FliC proteins shows that the amino-acid sequence similarity of FliC proteins corresponds with bacterial taxonomy. From these relationships we defined three clades based on sequence similarity and taxonomic relationships (more in Figure S1A). Tree scale describes the branch lengths.

(B) Bacterial genomes may contain multiple copies of *fliC* genes. A bacterial species tree of bacterial strains isolated from healthy Arabidopsis plants (Levy et al., 2017) that was created from a concatenated alignment of 31 single-copy genes found in each microbe. The metadata from the inner to outer circles represent: taxonomy, FliC clade any of their flagellin genes come from, and the number of *fliC* genes within each genome. Tree scale describes the branch lengths.

(C–E) Sequence logos representing the unique flg22 sequences found in the FliC clades depicted in Figure 1A. Letters are colored based on contact sites inferred from the crystal structure of the flg22-FLS2-BAK1 complex (Sun et al., 2013). A previously studied immunogenic flg22 sequence from *Pseudomonas aeruginosa* (Pa22) is shown for comparison. (C) Sequence logo of the unique sequences from clade 1. (D) Sequence logo of the unique sequences from clade 2. (E) Sequence logo of the unique sequences from clade 3.

may have been aided by *fliC* gene duplication, the maintenance of multiple *fliC* genes and antagonistic pleiotropy as detailed in the accompanying manuscript (K.P. et al., unpublished data).

Pervasive evasion of FLS2 activation by commensal flg22 peptide variants

To characterize functionally the diverse flg22 variants found in Arabidopsis commensals, we synthesized 97 variants covering >30% of the flg22 diversity across all three FliC clades (clade 1: 23, clade 2: 17, clade 3: 57) along with four control flg22 peptides (Table S2). We first screened the peptides for ROS burst (early response) and SGI (late response) induction on Arabidopsis plants.

We used the active *Pseudomonas syringae* pv. *tabaci* immunogenic flg22 peptide (Pta22) as a positive control and a variant with an Asp^{14} to Ala substitution (PtaDA), which is inactive at 10 nM but weakly active at 100 nM, as our baseline (Naito et al., 2008; Figures S2A and S2B). This allowed us to quantify ROS activity across a dynamic range between the highly active Pta22 peptide variant and the much less active mutant PtaDA

peptide variant (Figures S2A and S2B). We also included the canonical immunogenic Pa22 variant (Figures S2A and S2B) and an inactive 2-amino-acid-truncated version of Pa22, hereafter called Pa20, as additional controls (Bauer et al., 2001). Strikingly, 64% (62/97) of the flg22 variants tested did not induce a ROS burst at either 10 or 100 nM (Figures 2A and S2D). Of these, 57/62 peptides belong to clade 3, while the remaining five were split between clade 1 and clade 2 (four and one flg22 sequences, respectively). The 35/97 active flg22 variants produced ROS at both 10 and 100 nM (Figures 2A and S2D). To ensure that the flg22 variants were activating via FLS2, we assayed the active variants on Col-0 mutant plants that lack the FLS2 receptor (*fls2 efr*, see STAR methods). All 35 flg22 variants were inactive on *fls2 efr* mutant plants (Figure S2C). This demonstrates that FLS2 is capable of mediating a gradient of ROS burst responses from complete inactivity to that observed with 100 nM of Pta22 or Pa22 (Figure 2A). Thus, in contrast to its singular all or none response to canonical epitopes, FLS2 allows a range of responses to low abundance immunogenic epitopes in a commensal landscape potentially dominated by non-immunogenic signals.

We next performed SGI assays with 10 nM of each flg22 peptide variant. In these conditions, Pta22 and Pa22 readily reduced seedling fresh weight after 9 days of exposure (Figure 2A). Nearly all ROS-inactive flg22 variants failed to induce significant SGI (Figure 2A). We found that 19/35 (54%) of our ROS-active flg22 peptide variants induced significant SGI responses (Figure 2A) that were not observed on *fls2 efr* seedlings (Figure S2E). The remaining flg22 variants (16/35; 46%) were able to induce ROS but had no effect on SGI and were termed "deviant" peptide variants (Figures 2A and 2B). At least three classes of flg22 variants emerged: immunogenic peptides (20% of our tested variants), evading peptides that do not induce ROS burst or SGI (64% of our tested variants), and deviant peptides that only induce a ROS response (16% of our tested variants) (summarized in Figure 6).

Next, we constructed a phylogenetic tree using amino acid sequences from the flg22 variants to determine if ROS burst and SGI responses cluster by sequence (Figure 2C). We found that immunogenic flg22 sequences are derived mainly from clade 1 FliCs (I, Figure 2C). In contrast, non-immunogenic flg22 variants are mainly from clade 3 FliCs (III, Figure 2C) confirming our hypothesis that the high sequence divergence of flg22 epitopes in this clade has functional consequences on FLS2 responses. These data indicate that evasion of FLS2 activation is pervasive in Rhizobiales and Caulobacterales bacteria harboring clade 3 FliCs, potentially as a requirement for optimal colonization of Arabidopsis. Interestingly, deviant flg22 variants are mainly derived from clade 2 FliC proteins (II, Figure 2C). Deviant flg22 variants share high similarity to Pa22 and are mainly found in Bacillus and Actinobacteria genomes (Figure 1D). Unexpectedly, the variation in these flg22 sequences occur in the "address" portion of the peptide (Figure 1D), where a change of Gln^{1} to Glu and various changes at Lys^{13} are prominent in deviant peptides. Using synthetic flg22 variants, we found that $Pa22^{Q1E}$ increases ROS even with decreased binding affinity to FLS2 (K.P. et al., unpublished data). By contrast, the $Pa22^{K13D}$ leads to pronounced reductions in ROS levels by affecting the stable association of FLS2 to BAK1 (K.P. et al., unpublished data). We propose that the diversity in clade 2 flg22 peptides evolved to disrupt the sequential signaling outputs of FLS2, perhaps by a mechanism that involves the inappropriate recruitment of other regulators to FLS2. Overall, the clustering of flg22 responses based on taxonomic origin suggests that these responses play a role in community establishment and maintenance (Figure 2C). Our unprecedented screening of commensal-derived MAMP diversity demonstrates that a majority of commensal flg22 peptide variants evade FLS2 activation (64%) and do not significantly affect plant growth (80%).

Manipulation of immunity through receptor antagonism and signal modulation

Most flg22 variants from commensal bacteria evade FLS2 activation, but there are still many immunogenic variants. Microbial mechanisms to suppress MTI are beginning to be elucidated (Teixeira et al., 2019). One of these is receptor antagonism. Flg22 peptide variants that antagonize Pa22 induced phenotypes have been identified in *Ralstonia solanacearum*, *Pseudomonas cannabina* pv. *alisalensis* (ES4326), and among synthetic flg22 variants (Mueller et al., 2012; Clarke et al., 2013; Bauer et al., 2001). It was found that specific mutations in the C-terminal region of flg22 antagonized Pa22 by competing for FLS2 binding and blocking the formation of a stable heterocomplex with BAK1 (K.P. et al., unpublished data). Considering the large number of flg22 peptide variants that evade FLS2 activation and carry variable C-terminal domains (Figures 1E and 2C), we hypothesized that commensal bacteria have evolved flg22 peptide variants that block FLS2 activation.

To answer this, we tested flg22 peptides that failed to induce SGI for their ability to antagonize Pa22-induced expression of the root defense marker gene *Cytochrome P450 71A12 (CYP71A12)* (Millet et al., 2010). We found that 10 nM of Pa22 induced *proCYP71A12:GUS* expression in the root elongation zone, and the FLS2 receptor antagonist, Pa20, inhibited this response at a previously identified 10,000× ratio (100 μM) (Bauer et al., 2001; Figures 3A and S3A). We tested 56 non-immunogenic flg22 variants and found 10 that consistently antagonized Pa22-triggered *proCYP71A12:GUS* expression at 100 μM (Figures 3A and S3A). Interestingly, all antagonistic variants had mutations that could interfere with BAK1 binding at Gly^{18} and/or Leu^{19}; however, two different N-terminal modification strategies emerged. One set increased the negative charge, while the other contained high sequence similarity to the Pa22 peptide (Figures 3L and S3A). We hypothesize that both strategies support binding to FLS2 even with mutated C-terminal regions since they are able to block flg22 mediated *proCYP71A12:GUS* expression.

To understand the molecular mechanism of antagonism for naturally derived flg22 peptides, we tested six variants that completely antagonized *proCYP71A12:GUS* expression and are representative of antagonist sequence diversity for their ability to induce FLS2-BAK1 complex formation in Arabidopsis (Figures 3A and 3L). None of the six variants tested induced FLS2-BAK1 interaction in co-immunoprecipitation (coIP) assays (Figure 3B). This is consistent with these peptides' inability to elicit ROS and SGI.

Next, we tested if our candidate antagonist flg22 peptides were able to compete with Pta22 for the formation of FLS2-BAK1 heterocomplex *in vitro* (Figures 3C–3I and S3B). We found that three of the six peptide variants (An-$flg22^{2009}$, An-$flg22^{1186}$,

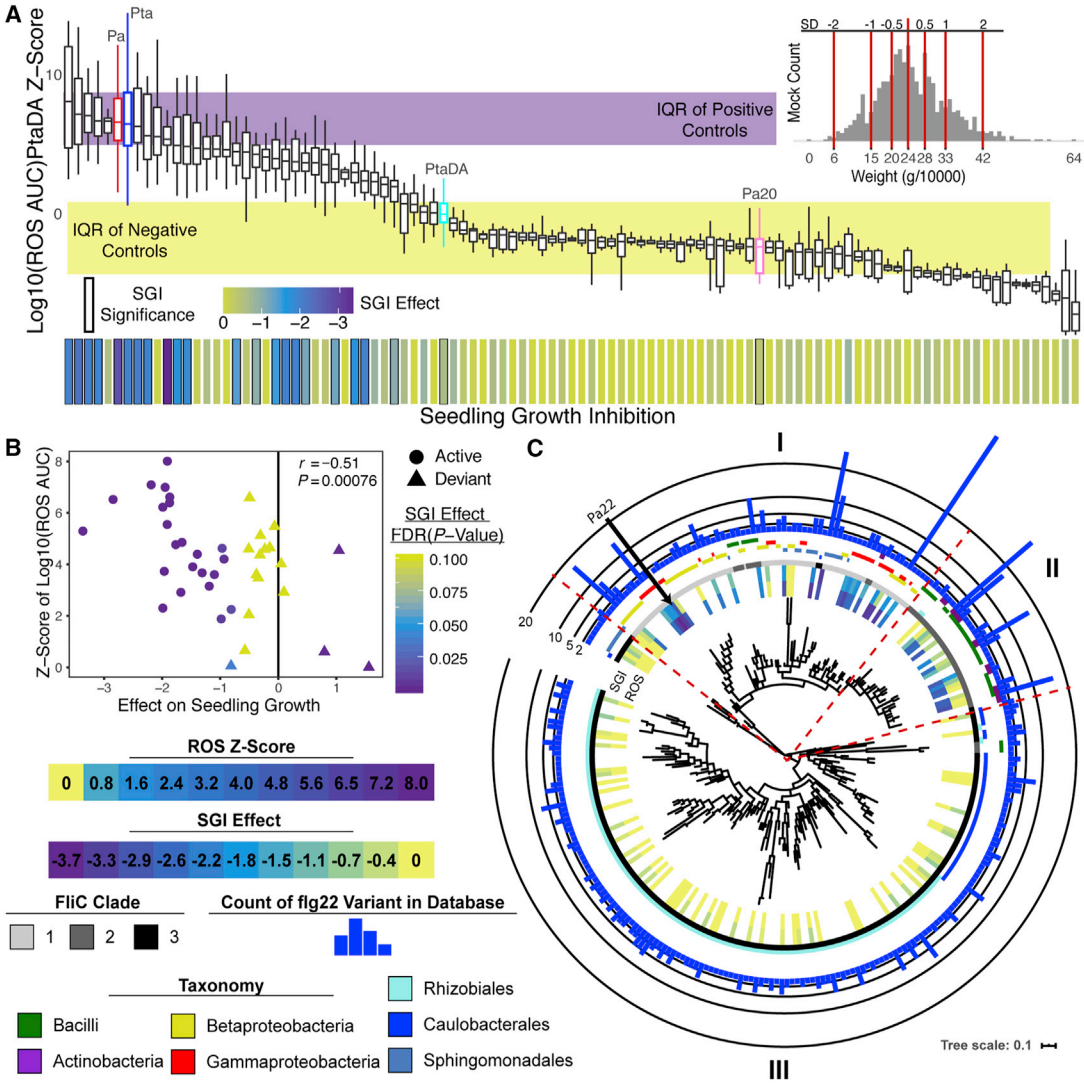

Figure 2. Evidence for pervasive evasion of flg22 induced MTI

(A) Commensal bacteria contain both immunogenic and non-immunogenic flg22 variants. Boxplot: ROS burst profiles after the addition of 100 nM of the flg22 peptide of interest to wild-type (WT) Col-0 leaf disks as measured by area under the curve (AUC) for at least 24 independent ROS burst profiles. Purple and yellow boxes indicate the combined interquartile range (IQR) of the positive (Pta22 and Pa22) and negative controls (PtaDA and Pa20), respectively. Barplot bottom: plate-based *Z* scores calculated for each flg22 peptide tested based on at least 16 independent SGI fresh weight measurements. Significance was calculated using a linear mixed model and is marked with black borders based on a peptide having an FDR-corrected p value of less than 0.01. The color scale bar above indicates the estimated SGI effect. Top right histogram: the SGI effect is interpreted as the fresh weight differences between mock (median of 24 g/1,000) and flg22-peptide-treated plants normalized to the standard deviation (SD) of mock-treated plants (~9 g/10,000).

(B) ROS burst and SGI are significantly correlated; however, there are exceptions where ROS burst does not correspond to SGI. All peptides that induce ROS burst greater than PtaDA are represented by ROS (*Z* score, y axis), SGI effect size (linear mixed model effects, x axis), and p values (SGI, color). The color scale corresponds to FDR-corrected p values calculated for the SGI linear mixed model effects.

(C) ROS burst and SGI discrepancies cluster based on flg22 sequence, FliC clade, and bacterial taxonomy. A phylogenetic tree from the amino acid sequences of 268 flg22 sequence variants. ROS burst (*Z* score) and SGI (linear mixed model effects) shown are derived from (A) (inner two circles). Taxonomy is colored based on classification of the genome in which that flg22 sequence resides. FliC clade is colored based on the clade of FliC from which that flg22 variant derives. Count of flg22 variants represents the number of times that flg22 variant was found in the Arabidopsis database. The dashed red lines define boundaries for three groupings of flg22 peptide variants based on phenotypes, ROS+ and SGI+ (I), ROS+ and SGI− (II), and ROS− and SGI− (III). The tree scale represents the branch lengths.

and An-flg22^{1410}) antagonized the FLS2-BAK1 interaction (Figures 3F, 3H, and 3I, respectively). The An-flg22^{1186} peptide produced significant antagonistic activity at 10× the concentration of Pta22 (Figure 3H). To extend these results *in vivo*, we tested all six peptides for their ability to antagonize Pta22-mediated ROS burst after 15 min of pre-exposure (Figures 3J–3K). We found that peptides An-flg22^{1186} and An-flg22^{1410} caused complete and persistent inhibition of ROS burst, while An-flg22^{2009}

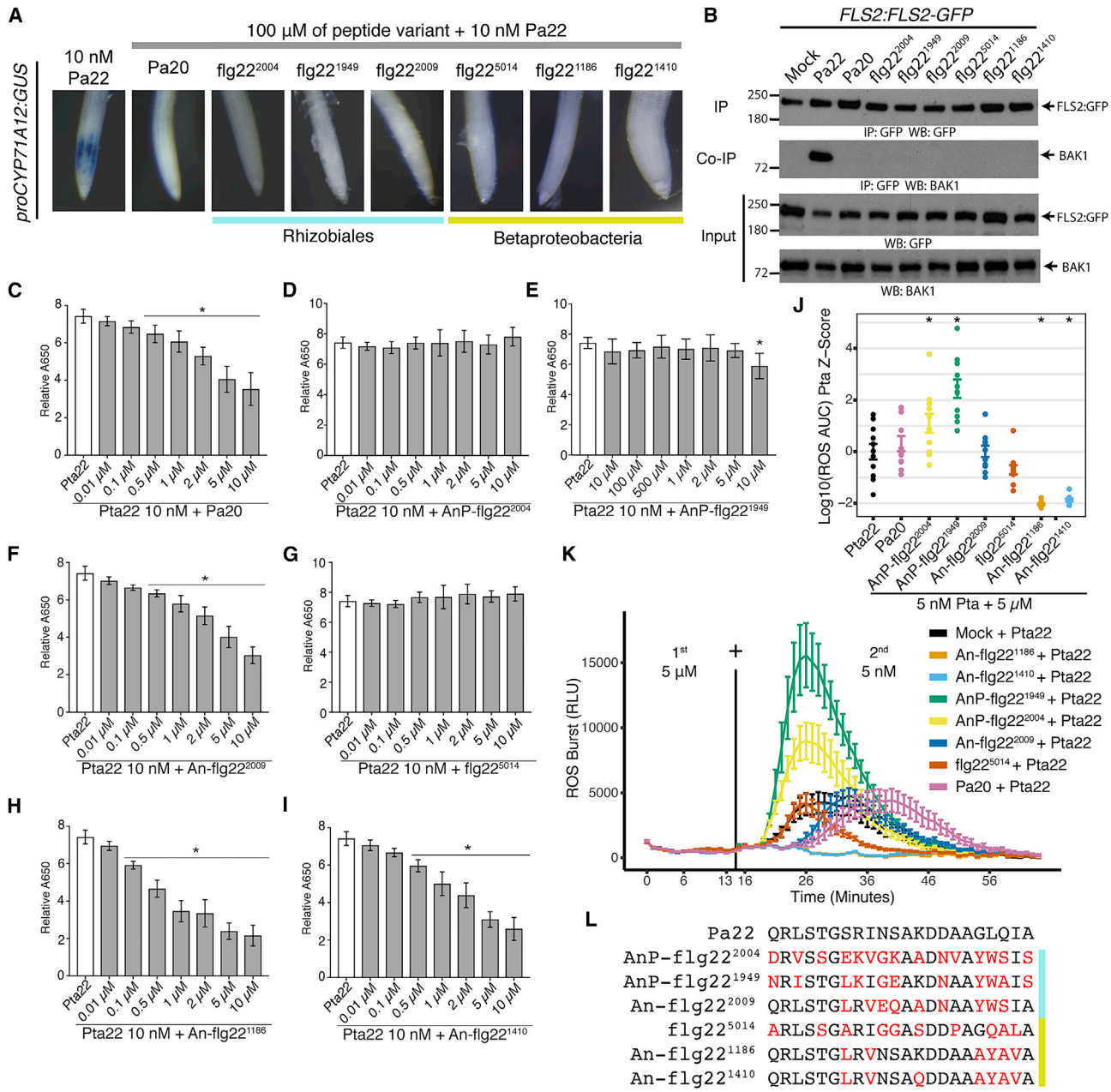

Figure 3. MTI altering flg22 variants are present in commensal Arabidopsis microbiomes

(A) Some natural flg22 variants can antagonize Pa22 mediated *proCYP71A12:GUS* expression. 10 nM of Pa22 and 100 µM of each flg22 peptide were added to 5–8 7-day-old *proCYP71A12:GUS* seedlings for 5 h before GUS staining. Each peptide shown suppressed GUS expression in all roots in two or more independent experiments. Colors associated with each flg22 peptide ID indicate the taxonomic group the variant is derived from. Positive *proCYP71A12:GUS* expression is displayed by the 10 nM Pa22 condition.

(B) Potential commensal encoding antagonists do not induce FLS2-BAK1 complex formation. CoIP using *proFLS2:FLS2-GFP* transgenic plants exposed to 1 µM of each flg22 variant for 15 min. Experiment was repeated three times with similar results.

(C–I) Flg22 peptides derived from commensal bacteria are able to antagonize FLS2-BAK1 heterocomplex formation measured *in vitro*. Shown is the relative absorbance (Abs 650 nm) over 2 h obtained in the presence of 10 nM Pta22 and at increasing concentrations of Pa20 and each potential antagonist. See Figure S3B for more information on the FLS2-BAK1 ectodomain interaction assay. Each barplot represents the mean and standard deviation from two independent biological experiments ($n \geq 5$). Statistical significance was assessed using a one-way ANOVA followed by Dunnett's multiple comparison test (*$p < 0.05$). (C) Pa20, (D) AnP-flg22^{2004} . (E) AnP-flg22^{1949}. (F) An-flg22^{2009}. (G) flg22^{5014}. (H) An-flg22^{1186}, (I) An-flg22^{1410}.

(legend continued on next page)

behaved similarly to Pa20 by delaying Pta22 mediated ROS burst (Figures 3J, 3K, and S3C). Interestingly, the other three peptides either slightly reduced the timing of maximum ROS (flg22^{5014}, Figure S3C) or, via priming, increased the ROS burst produced by Pta22 (AnP-flg22^{1949}, AnP-flg22^{2004}; AnP means antagonistic primer, Figures 3J and 3K). These data demonstrate that receptor antagonists are encoded in commensal communities and that some flg22 variants alter canonical MTI outputs in an incoherent manner. Thus, commensal organisms may have evolved specialized flg22 variants that exploit FLS2 signaling in order to facilitate plant colonization.

Motivated by the antagonistic oddities, we examined the ROS burst priming effect observed by the flg22 variants that antagonized Pa22-triggered *proCYP71A12:GUS* expression (Figure 3A) but did not inhibit Pta22-induced FLS2-BAK1 interaction (AnP-flg^{1949} and AnP-flg22^{2004}; Figures 3D and 3E). We found that Pta22-induced ROS burst scales with the concentration of AnP-flg22^{1949}; however, the priming effect could not surpass the maximum response observed for Pta22 (Figures S3D and S3E). We then determined that pre-exposure of AnP-flg22^{1949} primes BIK1 phosphorylation (Figure S3F). We controlled for specificity, by testing AnP-flg22^{1949} priming activity toward another peptide MAMP, elf18 (Kunze et al., 2004) and found that AnP-flg22^{1949} does not prime the ROS burst elicited by elf18 (Figure S3G). These data indicate that antagonistic priming variants specifically modulate the response to immunogenic flg22, leading to the elimination of temporally late flg22 signaling outputs. Thus, our data define at least two ways to alter flg22 signaling output. The first is via interference with co-receptor recruitment, and the second is by prevention of late output responses by an as yet unidentified mechanism that primes the canonical ROS response.

Active antagonistic peptides are encoded by prevalent community members

While commensals can produce antagonistic peptides, it is unclear if prevalent community members do so *in vivo*. Pathogenic *Ralstonia solanacearum* isolates were previously reported to encode a mix of antagonistic and evading flg22 variants (Mueller et al., 2012), prompting us to screen 9 flg22 variants found among 152 *Ralstonia* FliC proteins for antagonistic activity in Arabidopsis (Figure S4A; Table S4). Three of these (*Ra* An-flg22^{1186}, *Ra* An-flg22^{1410}, *Ra* An-flg22^{5001}) were inactive in ROS burst and SGI assays and were antagonists of all immune outputs analyzed (*Ra* An-flg22^{1186}, *Ra* An-flg22^{1410}: Figures 3A, 3B, and 3H–3K; *Ra* An-flg22^{5001}: Figures 4A–4C and S4B). We identified 2 additional variants (*Ra* E-flg22^{5004} and *Ra* E-flg22^{5005}) that were inactive for ROS and SGI but did not act as antagonists, thus representing true evaders (Figures 4A–4C and S4B). Notably, we identified one variant (*Ra* flg22^{5003}) that acted as a weak antagonist (Figures 4A–4C and S4B).

We used a root enriched *Ralstonia* isolate (*Ra* CL21) that expresses the An-flg22^{1186} antagonist to test if *Ra* CL21 produces functional An-flg22^{1186} peptides (Levy et al., 2017). We replaced its native *fliC* with alleles encoding flg22 epitopes changed to *Ra* E-flg22^{5005} or *Ra* flg22^{5003}. CL21Δ*fliC*::*Ra* flg22^{5003} and CL21Δ*fliC*::*Ra* flg22^{5005} mutant strains retained 55% of *Ra* CL21 motility capacity (Figure 4D). *Ra* CL21 can suppress flg22-mediated root growth inhibition (RGI), and we reasoned that the antagonistic An-flg22^{1186} variant is at least partially responsible for this phenotype (Teixeira et al., unpublished data). We found that *Ra* CL21 can partially suppress RGI induced by 100 nM of Pa22 in *UBQ10:FLS2-GFP* Arabidopsis plants that express FLS2 across all root cell types (Wyrsch et al., 2015) (Figure 4E). The *Ra* CL21 mutant expressing the weakly antagonistic *Ra* flg22^{5003} variant retained *Ra* CL21 levels of RGI suppression (Figure 4E). Strikingly, we found that swapping *Ra* CL21 flg22 epitope to the *Ra* E-flg22^{5005} variant significantly reduced RGI suppression. This demonstrates that functional antagonists can be produced by prevalent commensal bacteria and can alter flg22 perception *in vivo*.

We wondered whether antagonists were encoded in other prevalent commensal isolates, like the operational taxonomic unit (OTU) group defined in the phyllosphere by *Pseudomonas* strains *Pseudomonas OTU5* (Karasov et al., 2018). Genomes from this OTU encode 2 major flg22 epitopes, OTU5 An-flg22^{Pv1} and OTU5 flg22^{Pv2} found in 50% and 48% of the OTU5 genomes, respectively (Figure S4C; Karasov et al., 2018). The OTU5 An-flg22^{Pv1} variant has co-occurring mutations of Asn^{15} and Val^{18} that act as an antagonist in the Pa22 sequence background, and the OTU5 flg22^{Pv2} variant is similar to a previously identified atypical flg22 sequence encoded in *Pseudomonas cannabina* pv. *alisalensis* (ES4326) (K.P. et al., unpublished data; Clarke et al., 2013). Thus, we hypothesized that the major flg22 variants from this OTU group would act as antagonists. We found that both OTU5 variants suppressed Pa22-mediated induction of *CYP71A12:GUS*, whereas only the OTU5 An-flg22^{Pv1} variant suppressed flg22 mediated ROS burst (Figures S4C and S4D). This is consistent with findings from K.P. et al. (unpublished data) demonstrating antagonism of the FLS2-BAK1 complex *in vivo*. These data indicate that intimately associated phyllosphere and root communities contain commensals that produce active antagonistic peptide variants and that they may be produced in proportions that reduce immune system activities.

Natural flg22 variants drive separable MTI responses

Canonical flg22 peptides activate all known MTI responses. We defined deviant flg22 variants that induce ROS burst but not SGI

(J and K) Commensal flg22 peptides use different mechanisms to alter flg22 output responses in Arabidopsis. Flg22-induced ROS burst in WT Col-0 plants driven by 5 nM of Pta22 after pre-exposing leaves to 5 μM of one of the antagonistic flg22 peptides indicated. Distilled water was used as mock treatment. The experiment was repeated three times with similar results; n = 11. AnP = antagonistic primer; an, antagonist; see text for definitions.

(J) Integration of the area under the curve (AUC) measurements from ROS burst assays on independent leaf disks. These values are *Z* score normalized using the 5 nM of Pta22 and no pre-treatment condition. Bars represent the mean ± the standard error. Significant differences to the AUC of 5 nM Pta22 were determined using a one-way ANOVA followed by Dunnett's multiple comparison test (*< 0.05).

(K) The ROS burst kinetics for the same AUC data shown in (J). Each line represents the mean, and the bars represent the standard error of the measurements at each time point. 1st indicates the ROS burst induced by pre-exposure to the flg22 variants, and 2nd indicates the ROS induced after addition of the Pta22 peptide.

(L) Amino acid sequence for each flg22 peptide analyzed. Differences to the Pa22 sequence are highlighted in red. Light blue and yellow indicates the flg22 variants are derived from Rhizobiales and Betaproteobacteria, respectively.

Please cite this article in press as: Colaianni et al., A complex immune response to flagellin epitope variation in commensal communities, Cell Host & Microbe (2021), https://doi.org/10.1016/j.chom.2021.02.006

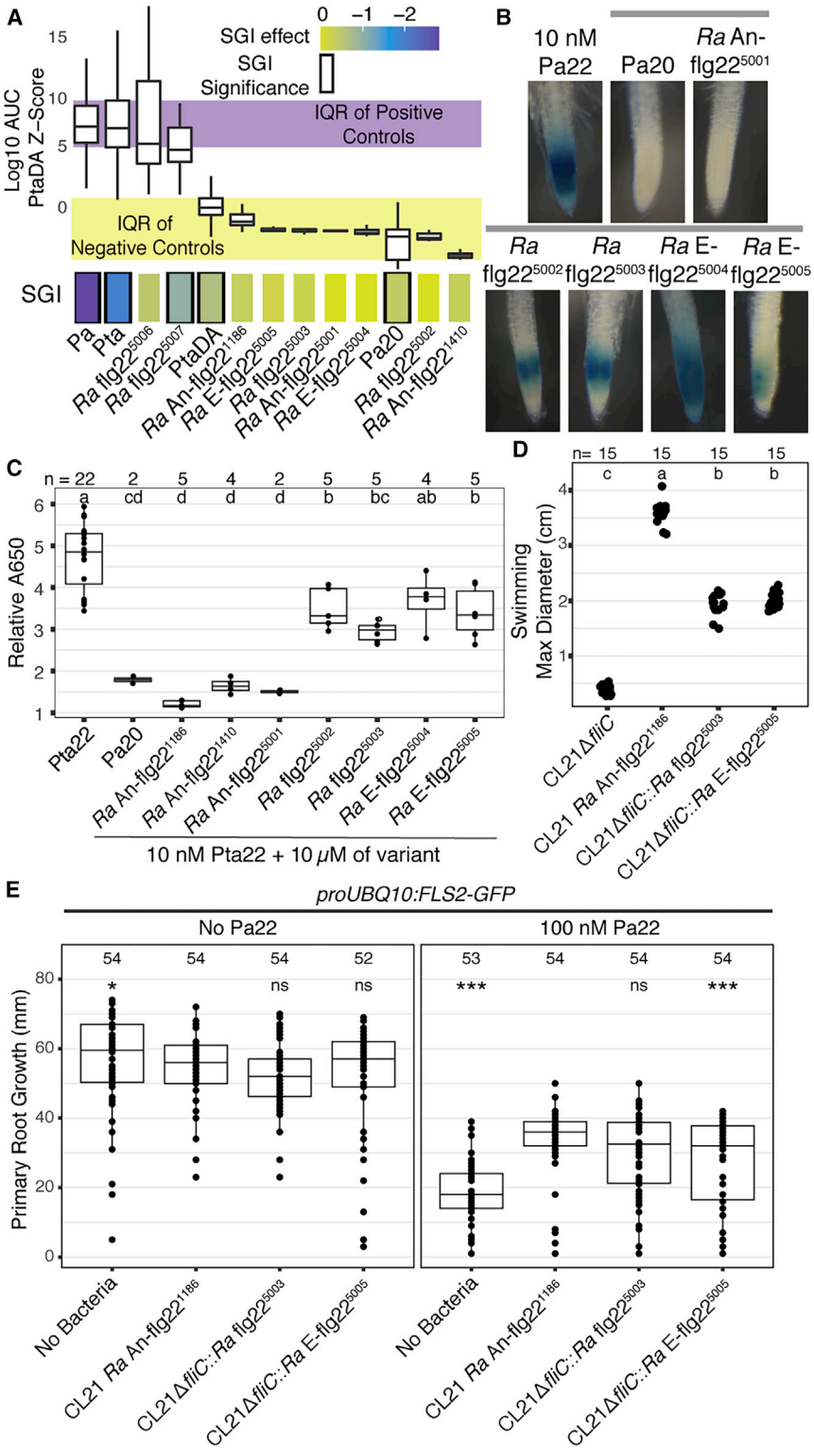

Figure 4. Prevalent commensal microbes produce flg22 antagonists

(A) *Ralstonia* produce flg22 variants that evade immune elicitation. Boxplot: area under the curve (AUC) was calculated for at least 24 independent ROS burst profiles after the addition of 100 nM of the flg22 peptide of interest to WT Col-0 leaf disks. Purple and yellow boxes indicate the combined interquartile range (IQR) of the positive (Pta22 and Pa22) and negative controls (PtaDA and Pa20), respectively. Barplot bottom: plate-based *Z* scores calculated for each flg22 peptide tested based on at least 16 independent SGI fresh weight measurements. Significance was assessed using a linear mixed model and is marked with black borders based on a peptide having an FDR-corrected p value of less than 0.01. The color scale bar on top indicates the SGI effect compared to mock-treated plants. In all panels E-flg22 indicates evading peptides and An-flg22 indicates antagonistic peptides that phenocopy those defined in Figure 3.

(B) *Ralstonia* flg22 variants that evade recognition can antagonize recognition of Pa22. Antagonism GUS assay where 10 nM of Pa22 and either 100 μM of Pa20 or a *Ralstonia* flg22 peptide were added to 7-day-old *proCYP71A12:GUS* seedlings for 5 h before GUS staining. Each peptide displayed as an antagonist suppressed GUS expression in all roots from at least two independent experiments. Positive *proCYP71A12:GUS* expression is displayed by the 10 nM Pa22 condition. All conditions with a gray bar indicate that 10 nM of Pa22 was added with 100 μM of the variant.

(C) *Ralstonia* flg22 peptides quantitatively differ in antagonism of the FLS2-BAK1 heterocomplex formation measured *in vitro*. Shown is the relative absorbance (Abs 650 nm) over 2 h obtained in the presence of 10 nM Pta22 and 10 μM of Pa20 and each *Ralstonia* flg22 peptide. Statistically different groups displayed with letters at the top were identified using a one-way ANOVA followed by a Tukey's test ($\alpha = 0.05$).

(D) *Ralstonia* CL21 derivatives with mutant *fliC* genes containing swapped flg22 epitopes retain motility function. The CL21 strains indicated were inoculated in the center of 15 independent motility plates. The resulting swimming phenotype was scanned and the location with the largest diameter was measured using ImageJ. Statistically different groups were identified using a one-way ANOVA followed by a Tukey's test ($\alpha = 0.05$).

(E) CL21 strains produce antagonistic variants that suppress flg22-mediated root growth inhibition. The CL21 strains indicated were spread on plates supplemented with or without 100 nM of Pa22 at an OD_{600} of 0.0002. 7-day-old *proUBQ10:FLS2-GFP* seedlings were transferred to these plates. Main root elongation was measured with a ruler seven days later. Main root elongation differences to WT CL21 within the Pa22 minus or plus conditions were evaluated with a linear mixed model (*$p < 0.05$ and ***$p < 0.001$), and numbers at the top indicate the number of roots measured.

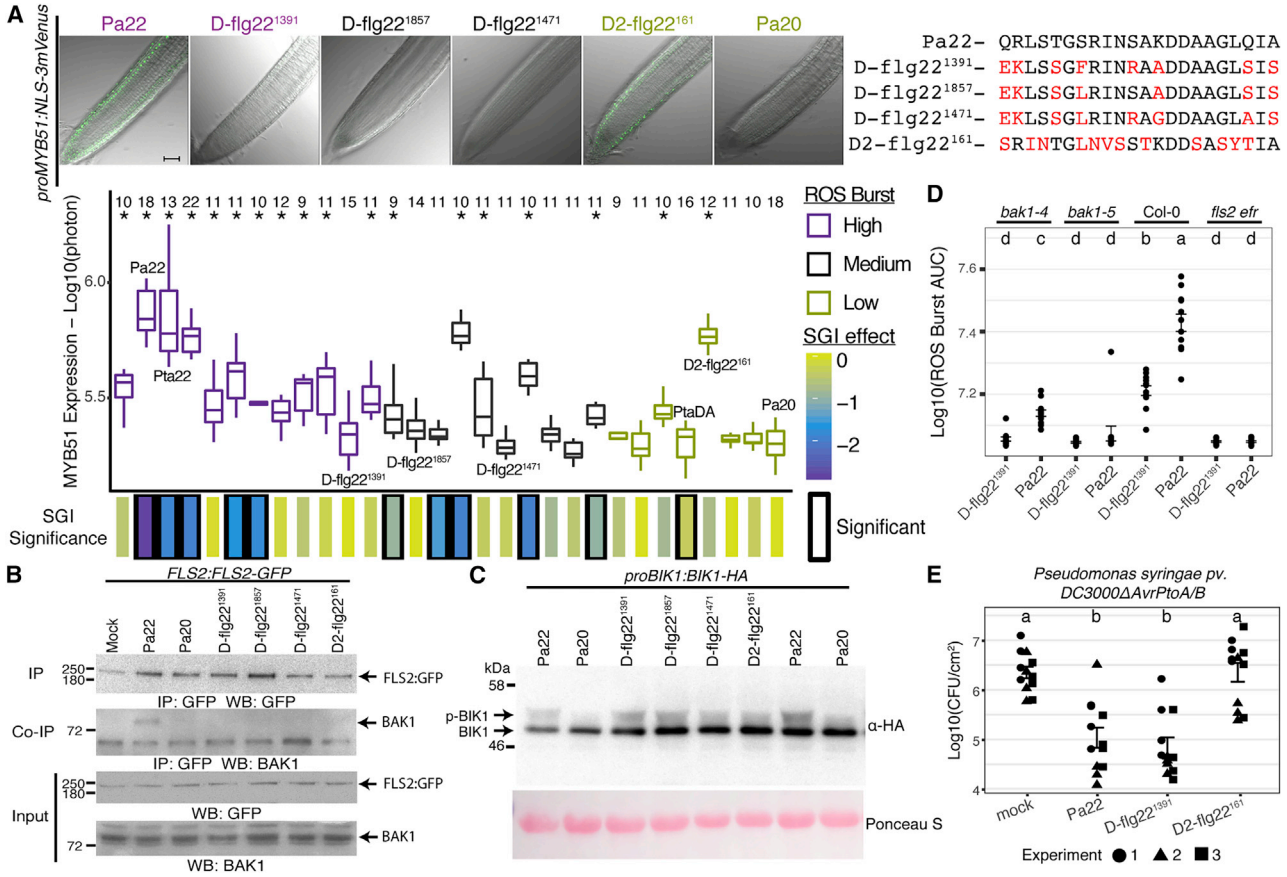

Figure 5. Deviant natural flg22 variants drive separable MTI responses

(A) Flg22 variation can result in non-canonical gene expression changes. Top: expression pattern of defense marker gene *MYB51* in Arabidopsis roots exposed to 10 nM of each four deviant (D; see text for definition) flg22 variants (numbered at top; color coded by ROS burst activity). Shown are merged fluorescent and brightfield confocal microscope images of representative roots. The black scale bar represents 50 μm. Bottom: fluorescence signal quantifications for microscopic observations of *proMYB51:NLS-3mVenus* expression after 10 nM exposure to each flg22 peptide for 24 h. Peptides are ordered and color coded by ROS burst activity (Figure 2A). The number above each boxplot represents the number of roots imaged for each variant. Significant *proMYB51:NLS-3mVenus* expression was determined by comparing each variant with the gene expression observed for Pa20 using an ANOVA and Dunnet's test (*$p < 0.05$). The amino acid sequences of deviant and deviant-2 peptides are shown to the right with differences to the Pa22 sequence highlighted in red.

(B) The deviant and deviant-2 peptides do not induce FLS2-BAK1 complex. CoIP using *proFLS2:FLS2-GFP* transgenic plants exposed to 100 nM of each numbered flg22 peptide for 15 min. Western blot analyses of lysates were performed using anti-BAK1 and anti-GFP antibodies. This experiment was repeated three times with similar results.

(C) Deviant and deviant-2 peptides activate BIK1 phosphorylation. Western blot analyses of BIK1 phosphorylation in *proBIK1:BIK1-HA* seedlings exposed to 100 nM of each peptide variant for 40 min. The experiment was repeated two times with similar results.

(D) ROS burst for D-flg22^{1391} is BAK1-dependent even though D-flg22^{1391} does not induce coIP between FLS2 and BAK1. Flg22-induced ROS burst of WT Col-0, *fls2 efr*, *bak1*-5, and *bak1*-4 leaf disks exposed to D-flg22^{1391} or Pa22 at 100 nM. Shown are AUC measurements, where each error bar represents the mean ± the standard error of 12 independent leaf disks at each time point measured. Statistically different groups indicated by letters at the top were identified with a one-way ANOVA and Tukey's test ($\alpha = 0.05$). Experiment was repeated three times with similar results.

(E) The D-flg22^{1391} variant activates effective MTI, while D2-flg22^{161} does not. WT Col-0 plants were hand infiltrated with 100 nM of peptide or distilled water (mock) 24 h before hand infiltration of *Pseudomonas syringae* pv. *tomato* DC3000Δ*AvrPtoA/B* at an OD600 of 0.0002. 12 data points from three independent experiments are displayed. Significantly different groups indicated at the top were identified using a two-way ANOVA controlling for batch effect and a Tukey test ($\alpha = 0.05$). The error bars indicate the mean ± standard error.

(Figures 2A and 2B). To investigate their inability to activate full classical MTI defense markers, and to address whether the response to flg22 peptide variants is organ-specific, we monitored the activation of the transcriptional reporters *proMYB51:NLS-3mVenus* and *proWRKY11:NLS-3Venus* in Arabidopsis roots and epidermal cells of cotyledons, respectively (Poncini et al., 2017). ROS-inducing deviant peptide variants D-flg22^{1391}, D-flg22^{1857}, and D-flg22^{1471} failed to induce the transcription of either reporter genes, as predicted from their inability to induce SGI (Figures 5A, S5A, and S5B). Unexpectedly, we identified a second deviant class, deviant-2, that induced transcriptional changes in our reporter lines similar to Pa22 but induced less ROS than PtaDA at 100 nM (Figures 5A, S5A, and S2C; D2-flg22^{161}). This very low-level ROS burst was FLS2 dependent

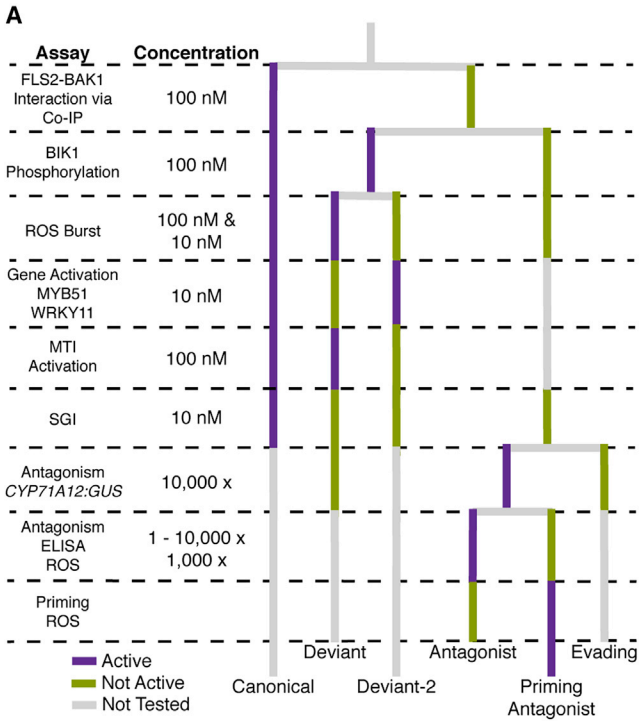

Figure 6. Commensal communities can produce at least 6 unique classes of flg22 variants, each of which results in a different response from the plant

Some of these may reshape the interaction between the plant and other community members. A summary of our experiments that collectively defines 6 unique classes of flg22 variants.

(Figure S2C). Previous work has reported the uncoupling of ROS burst and defense gene induction after MAMP recognition using chemical manipulation and mutant Arabidopsis lines (Lu et al., 2009; Shinya et al., 2014; Tran et al., 2020). We show that natural MAMP epitopes are also capable of decoupling MTI outputs. This indicates that the specificity of commensal-encoded flg22 peptide recognition by FLS2 may result in the recruitment of different signaling components, leading to heterogeneous intracellular signaling outputs.

To identify the machinery necessary for the signaling of D-flg22^{1391}, D-flg22^{1857}, and D-flg22^{1471}, we tested if they induced FLS2-BAK1 complex formation *in vivo* with 100 nM of each peptide variant and found that none did (Figure 5B). Strikingly, however, these flg22 variants retained the ability to induce BIK1 phosphorylation (Figure 5C). We hypothesized that these variants likely activate an FLS2 receptor complex that does not include BAK1. We therefore tested the requirement of BAK1 for the ROS burst induced by deviant variants D-flg22^{1391}, D-flg22^{1857}, and D-flg22^{1471} using two different BAK1 mutants, *bak1-4* (a null allele; Chinchilla et al., 2007) and *bak1-5* (a dominant negative allele; Schwessinger et al., 2011). We found that these three deviant flg22 variants required BAK1 for ROS burst production (Figures 5D, S5C, and S5D). This suggests that BAK1-FLS2 binding dynamics and/or complex-associated proteins are capable of fine-tuning flg22 signaling output.

To identify the functional significance of deviant and deviant-2 peptides in MTI, we performed flg22-driven protection assays. We hand inoculated leaves with 100 nM of Pa22, D-flg22^{1391}, or D2-flg22^{161}, 24 h before hand inoculation with *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B*, a weak pathogen lacking two type III effector genes required to dampen MTI (He et al., 2006). We found that the Pa22 peptide reduced the colonization of DC3000Δ*avrPtoA/B* significantly compared with plants inoculated with distilled water (Figure 5E). Interestingly, the deviant peptide variant, D-flg22^{1391}, limited the growth of DC3000Δ*avrPtoA/B* to the same level as Pa22, while the deviant-2 peptide variant, D2-flg22^{161}, did not reduce the growth of DC3000Δ*avrPtoA/B* (Figure 5E). This demonstrates that flg22 variants can induce distinct MTI responses that alter colonization outcomes. In sum, we define 6 classes of flg22 peptide variants based upon the immune outputs induced in mono-association or by how they affect the response to other flg22 peptides (Figure 6A).

Non-immunogenic variants dominate the flg22 functional repertoire in plant-associated communities

The composition of flg22 peptides in plant-associated communities may be actively shaped by plant immune surveillance. Alternatively, community homeostasis could be an epiphenomenal effect of flg22 variation on the immune system. In either case, the effects of flg22 variation on commensal community structure are unknown. We therefore leveraged existing synthetic community (SynCom) composition data from plant-associated microbiota assembled from a 185-member SynCom (Finkel et al., 2020) to analyze the flg22 functional repertoire (evading, antagonistic, and deviant, as defined above). It is important to note that both our flg22 database and this 185-member community are based on culturable isolates and contain approximately 65% of the class level diversity found in natural communities (Figure S6A, Finkel et al., 2019). However, they nonetheless provide a good representation of the isolated fraction of plant-associated communities (Figure S6A, Finkel et al., 2019). We categorized all FliC variation and functionally characterized approximately two-thirds of the total flg22 abundance within the 185-SynCom communities. This included the relative abundance of flg22 variation and functional classes within and across plant-associated root and shoot fractions (Figure S6B). Importantly, we could almost completely annotate the relative abundance of flg22 variants from clades 1 and 2, which contain the active flg22 variants identified in our functional screen (Figures S6C and 2C).

We first analyzed the FliC clade relative abundance in microbiota derived from the agar, root, and shoot fractions of seedlings colonized by SynCom185. We found a significant enrichment in bacteria encoding clade 3 FliCs in the root- and shoot-associated communities, while bacteria encoding clade 1 FliCs were depleted (Figure 7A). Bacteria containing no *fliC* genes were unchanged across the fractions, suggesting that FliC type affects colonization (Figure 7A). To understand if the colonization effect of FliC clades varies with flg22 activity, we assigned the relative abundance of each unique V3-V4 16S sequence variant (Useq) derived from SynCom185 to all unique flg22 variants represented by the Useq. We then aggregated flg22 relative abundance values across flg22 functional classes. We found that evading peptide variants were enriched while immune active flg22 variants were depleted

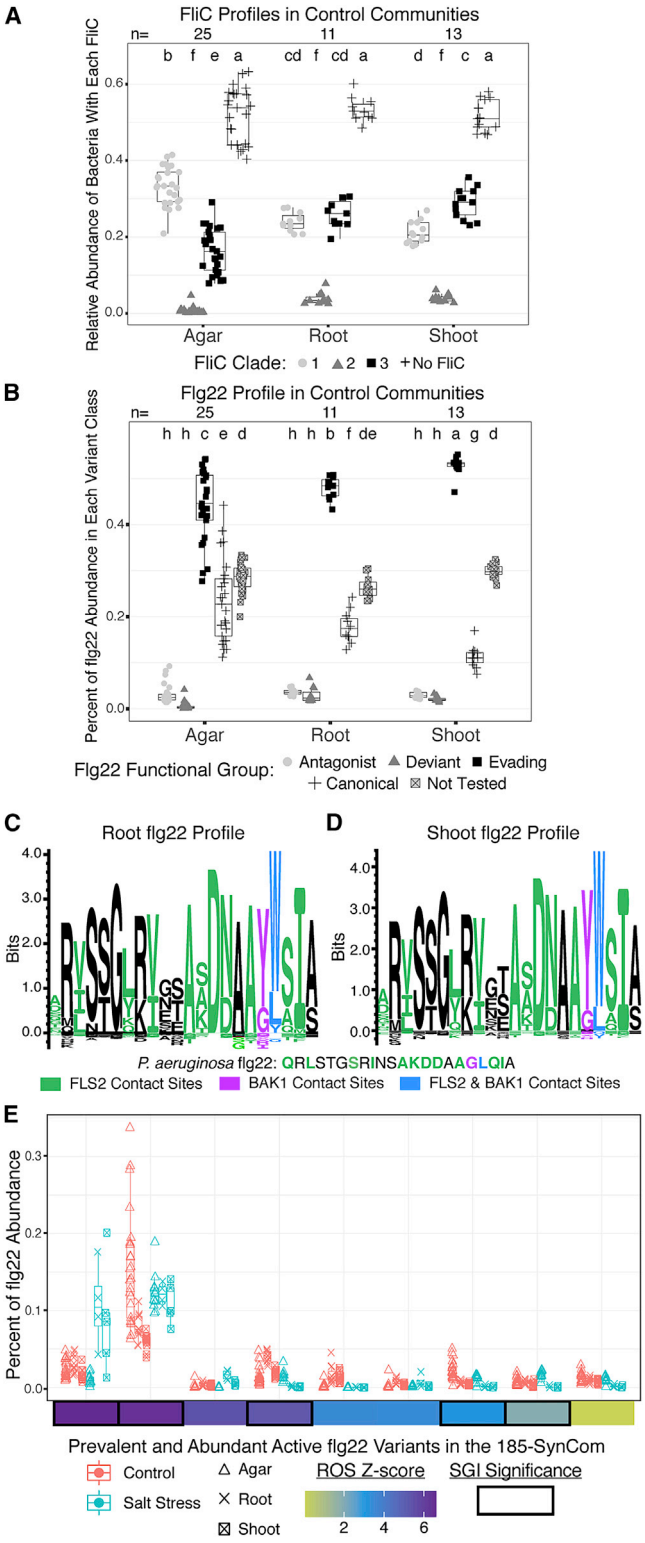

Figure 7. Non-immunogenic variants dominate the flg22 functional repertoire in plant-associated synthetic communities (SynCom)

(A) Plant-associated communities are enriched for bacteria encoding clade 3 FliCs, while being depleted in clade 1 FliCs. A 185-member SynCom was applied to Arabidopsis seedlings in normal growth conditions (21°C and 1,000 μM Pi) and were sequenced for the V3-V4 16S region (Finkel et al., 2020). The relative abundance values for each unique V3-V4 16S sequence variant in SynCom185 (Useq) were assigned to the FliC clade encoded in the various genomes. Statistically significant groups were identified using a two-way ANOVA and Tukey's test ($\alpha = 0.05$) and are indicated by the letters at the top.

(B) The flg22 functional repertoire in plant-associated communities is dominated by evading peptide variants, while immunogenic flg22 variants are depleted. Useq relative abundance values were assigned to the flg22 variants represented by the Useq. Percentages were then calculated using the total abundance over all flg22 variants defined in the community. Statistically significant groups were identified using a two-way ANOVA and Tukey test ($\alpha = 0.05$) and are indicated by the letters at the top.

(C and D) The flg22 composition in root and shoot communities resembles immune evading flg22 variants. Sequence logos created from representing each flg22 variant based on its median relative abundance in root (C) and shoot (D) communities. For comparison, the Pa22 amino acid sequence variant is shown. The colors represent contact sites inferred from the crystal structure of the Pa22-FLS2-BAK1 complex (Sun et al., 2013).

(E) Plant-associated communities assembled under salt stress are enriched in the most immunogenic flg22 variants. Prevalent and abundant flg22 sequences are defined as sequences having a flg22 percent abundance statistically different from 0. ROS (*Z* score) and SGI significance are taken from Figure 2A. Flg22 percent abundances are taken from the data displayed in Figures 7B and S6C.

in plant-associated communities (Figure 7B). Sequence logos representing the flg22 functional capacity in root and shoot communities, resemble evasive flg22 variants based on the mutated C-terminal regions (Figures 7C and 7D). These data suggest that evading flg22 variants are dominant and important in plant-associated communities.

To gain insight into whether plants' responses to abiotic stress alter flg22 composition, and hence community composition, we compared the communities assembled under optimal growth conditions with communities assembled under salt stress. Salt stress reduces plant defense through abscisic-acid-mediated immune suppression (Berens et al., 2019). We hypothesized that communities formed under high salt would not be depleted in active variants or enriched in evading variants, and confirmed this (Figure S6D). Additionally, bacteria harboring the most immunogenic variants were now enriched in plant-associated communities (Figures 7E and S6E). Thus, commensal MAMPs are likely to be actively monitored by plants, and, as a result, flg22 variants that can evade may be selected for. This interplay among immunogenic variants, evading variants, antagonizing variants, and the host immune system may ultimately promote community diversity and maintain microbial homeostasis.

DISCUSSION

Triggering of plant immune receptors by MAMPs classically results in an immune response that facilitates halting the growth of an invading microbe. Yet, plants stably associate with a compendium of microbes that can display collections of putative MAMPs closely related to those found expressed by pathogens. While mechanisms to suppress MTI are present in pathogens and beneficial microbes, it is unclear how common and effective they are in the commensal community context (Teixeira et al., 2019). Thus, to coexist with beneficial communities of microbes, without suffering the tissue damage associated with chronic immunity, plants must have evolved to unfold their response to MAMPs in ways that allow a balance between pathogen surveillance and commensal colonization. To address this, we focused on the immune responses

engendered by flagellin epitope variation encoded in genomes of strains derived from commensal communities.

The primitive organisms *Hydra* and *Euprymna scolopes* utilize PRRs to recognize commensal bacteria present in their environment, suggesting that this may be an ancient function of innate immunity (Chu and Mazmanian, 2013). We propose that FLS2 evolved to act as a communication device to allow plants and microbes to interact optimally. Thus, our findings challenge dogmas of a "strict" immune signaling function for FLS2 and rather indicate that FLS2 functions as a community sensor that can be utilized to influence the assembly of specific commensal microbiota.

Commensal bacteria produce substantial flg22 diversity that contributes to pervasive evasion of MTI

We found that commensals contain abundant flg22 sequence diversity and that most flg22 variants are non-immunogenic to the host. A majority of these flg22 peptides commonly introduce a mutation of Asp^{15} to Asn^{15}, which is not sufficient for immune evasion (K.P. et al., unpublished data). Rather, we attribute the "evading" activities of these peptides to co-occurring polymorphisms in the "message" domain of flg22. Collectively, our data support the notion that these variants evolved naturally to reduce their agonistic activity and prevent maximal FLS2-BAK1 interaction for full MTI activation. This is consistent with data using synthetic flg22 variants that identified evolutionary strategies by which *Pseudomonads* evolve evading flg22 variants in the face of motility loss (K.P. et al., unpublished data). As most of the evading flg22 variants come from bacterial genomes with multiple *fliC* genes, we propose that gene duplication facilitates the exploration of this evolutionary space by producing multiple functional gene copies (Ohno, 1979).

Flg22 response diversity may play a role in community composition and maintenance

We found that the responses induced by flg22 variants seem to reflect the colonization ability of the bacterial taxa in which they are found. Evasion of flg22 induced immunity is pervasive in Rhizobiales and Caulobacteriales isolates (Figure 2C), which are exemplary colonizers of Arabidopsis (Finkel et al., 2020; Carlström et al., 2019). Alternatively, Gammaproteobacteria, which predominantly contain canonical flg22 variants (Figure 2C), are poor colonizers of Arabidopsis roots (Finkel et al., 2020; Carlström et al., 2019). We find that active variants are depleted, while evading variants are enriched in plant-associated communities. This is dependent on both flg22 epitope functional class and plant defense (Figures 7B and S6C). These results are not complete, as we only study flg22 variants and community dynamics from culturable isolates. However, these results support an active role of flg22 in community structuring. We speculate that this concept will translate to other MAMP signals and other host-microbiota interaction systems.

Modulation of FLS2 signaling transduction

The effects of root tissue type and cell identity on FLS2 signaling demonstrate that responses to flg22 are compartmented and modular (Zhou et al., 2020; Emonet et al., 2020; Rich-Griffin et al., 2020). Here, we extend this response complexity by reporting that flg22 variation can differentially activate defense pathway outputs. We find that deviant variants and antagonistic priming variants have negatively charged amino acids in the N terminus of the peptide, suggesting that the charge of this region controls signal attenuation. Negatively charged mutations in the C terminus of the Pa22 peptide affect FLS2-BAK1 complex formation (K.P. et al., unpublished data). Therefore, we speculate that FLS2-BAK1-binding dynamics play a role in signal specificity and modulation. Taken together, regulating expression of FLS2 in specific cell types and sequestering the receptor from chronic stimulation by a subset of commensal-derived receptor antagonists might represent a passive mechanism to ensure appropriate communication with commensal microbiota. How FLS2 is able to mediate tissue-specific responses while selectively allowing subsets of flg22 variants to modulate signal transduction is an exciting open question.

FLS2 and signal transduction as a tabulation machine

The integration of flg22 signaling outputs in a dominant background of evasive peptides is fascinating. We propose that defense attenuation via receptor antagonism and signal modulation is layered on top of signal specificity. In this scenario, FLS2 integrates the ratios of immunogenic pathogen-derived danger signals (canonical peptides) and safety signals (evading and modulating peptides) by operating as a molecular tabulator. We envision that during colonization, bacteria producing danger signals that cannot be masked by suppression mechanisms are accommodated by bacteria producing safety signals. In a steady state balanced microbiota, the detection of pathogen overproliferation is monitored by increases in danger signals and/or decreases in safety signals across a specific threshold. The identification of signal imbalance could be triggered via the increased colonization of a pathogen producing danger signals and/or alternatively through the decrease in absolute abundance of bacteria that produce safety signals. To add to this complexity, plants can alter the threshold needed to produce defense by tuning FLS2 expression in different cell types and developmental states. In the context of damage, which increases FLS2 expression, we hypothesize that plants increase that amount of safety signals needed for colonization. This would allow plants the ability to differentiate between friend and foe.

Here, we show that commensal communities contain substantial flg22 diversity leading to pervasive evasion, specific defense activation, and differential immune system signal modulation. Plants are colonized by bacterial communities encoding a wide range of proteinaceous MAMPs that are presumably detected by distinct immune receptors. Thus, our study of the natural responses of FLS2 to flg22 diversity provides a conceptual framework to better understand how MAMPs and their cognate receptors impact the way commensal communities are built and maintained to allow for colonization while simultaneously surveilling for pathogens.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2021.02.006.

ACKNOWLEDGMENTS

This work was supported by NSF grant IOS-1917270 to C.D.J. and J.L.D.; J.L.D. is an investigator of the Howard Hughes Medical Institute, supported by the HHMI. N.R.C. was supported by the NIH training grant T32GM135123. The work was also supported by grants from the Austrian Academy of Sciences through the Gregor Mendel Institute, the Vienna Science and Technology Fund project (LS17-047) and by the FWF Austrian Science Fund (I 3654) to Y.B. We thank the GMI/IMBA/IMP Protein Chemistry Core facility for mass spectrometry analysis and the Vienna Biocenter Core Facilities (VBCF ProTech) for help with protein production. We thank members of the Dangl and Belkhadir labs for useful discussion, and Drs. Sarah Grant, Oliver Furzer, and Sarah Gilbert for critical reading of the manuscript. We thank Ji-Woo Kim for her help with confocal imaging. We also thank Dr. Niko Geldner for kindly providing us *proUBQ10:FLS2-GFP* seeds.

AUTHOR CONTRIBUTIONS

N.R.C., C.D.J., Y.B., and J.L.D. conceptualized the project. C.D.J., Y.B., and J.L.D. supervised the research. N.R.C., K.P., H.-S.L., J.M.C., N.H.K., N.E., T.S.M., and T.F.L. performed experiments. M.M. synthesized peptides and performed quality control. All authors contributed to data analysis. N.R.C. developed figures and wrote the main draft of the text. C.D.J., Y.B., and J.L.D. contributed significantly to that draft and all other authors contributed edits and comments to finalize the paper.

DECLARATION OF INTERESTS

J.L.D. is a co-founder of, and shareholder in, AgBiome LLC, a corporation whose goal is to use plant-associated microbes to improve plant productivity.

Received: October 8, 2020
Revised: January 5, 2021
Accepted: February 9, 2021
Published: April 14, 2021

REFERENCES

Albrecht, C., Boutrot, F., Segonzac, C., Schwessinger, B., Gimenez-Ibanez, S., Chinchilla, D., Rathjen, J.P., de Vries, S.C., and Zipfel, C. (2012). Brassinosteroids inhibit pathogen-associated molecular pattern-triggered immune signaling independent of the receptor kinase BAK1. Proc. Natl. Acad. Sci. USA *109*, 303–308.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389–3402.

Amin, I.M., Richmond, G.E., Sen, P., Koh, T.H., Piddock, L.J.V., and Chua, K.L. (2013). A Method for generating marker-less gene deletions in multidrug-resistant Acinetobacter baumannii. BMC Microbiol *13*, 158.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gómez-Gómez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. Nature *415*, 977–983.

Bai, Y., Müller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P.C., Spaepen, S., Remus-Emsermann, M., et al. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. Nature *528*, 364–369.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Software *67*, 1–48.

Bauer, Z., Gómez-Gómez, L., Boller, T., and Felix, G. (2001). Sensitivity of different ecotypes and mutants of Arabidopsis thaliana toward the bacterial elicitor flagellin correlates with the presence of receptor-binding sites. J. Biol. Chem. *276*, 45669–45676.

Belkhadir, Y., Jaillais, Y., Epple, P., Balsemão-Pires, E., Dangl, J.L., and Chory, J. (2012). Brassinosteroids modulate the efficiency of plant immune responses to microbe-associated molecular patterns. Proc. Natl. Acad. Sci. USA *109*, 297–302.

Belkhadir, Y., Yang, L., Hetzel, J., Dangl, J.L., and Chory, J. (2014). The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. Trends Biochem. Sci. *39*, 447–456.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B Methodol. *57*, 289–300.

Berens, M.L., Wolinska, K.W., Spaepen, S., Ziegler, J., Nobori, T., Nair, A., Krüler, V., Winkelmüller, T.M., Wang, Y., Mine, A., et al. (2019). Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. Proc. Natl. Acad. Sci. USA *116*, 2364–2373.

Bolstad, W.M. (2010). Understanding Computational Bayesian Statistics (Wiley).

Boutrot, F., and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. Annu. Rev. Phytopathol. *55*, 257–286.

Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E.L., Morimoto, K., Kaiser, M., Preston, G.M., Ichinose, Y., et al.

(2019). Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. Science *364*, 364–374.

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

Carlström, C.I., Field, C.M., Bortfeld-Miller, M., Müller, B., Sunagawa, S., and Vorholt, J.A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. Nat. Ecol. Evol. *3*, 1445–1454.

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell *18*, 465–476.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature *448*, 497–500.

Chu, H., and Mazmanian, S.K. (2013). Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nat. Immunol. *14*, 668–675.

Chung, E.H., El-Kasmi, F., He, Y., Loehr, A., and Dangl, J.L. (2014). A plant phosphoswitch platform repeatedly targeted by type III effector proteins regulates the output of both tiers of plant immune receptors. Cell Host Microbe *16*, 484–494.

Clarke, C.R., Chinchilla, D., Hind, S.R., Taguchi, F., Miki, R., Ichinose, Y., Martin, G.B., Leman, S., Felix, G., and Vinatzer, B.A. (2013). Allelic variation in two distinct Pseudomonas syringaeflagellin epitopes modulates the strength of plant immune responses but not bacterial motility. New Phytol *200*, 847–860.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Emonet, A., Zhou, F., Vacheron, J., Heiman, C.M., Tendon, V.D., Ma, K.W., Schulze-Lefert, P., Keel, C., and Geldner, N. (2020). Spatially restricted immune responses allow for root meristematic activity during bacterial colonisation. bioRxiv. https://doi.org/10.1101/2020.08.03.233817.

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. Plant J *18*, 265–276.

Finkel, O.M., Salas-González, I., Castrillo, G., Conway, J.M., Law, T.F., Teixeira, P.J.P.L., Wilson, E.D., Fitzpatrick, C.R., Jones, C.D., and Dangl, J.L. (2020). A single bacterial genus maintains root growth in a complex microbiome. Nature *587*, 103–108, https://doi.org/10.1038/s41586-020-2778-7.

Finkel, O.M., Salas-González, I., Castrillo, G., Spaepen, S., Law, T.F., Teixeira, P.J.P.L., Jones, C.D., and Dangl, J.L. (2019). The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. PLoS Biol *17*, e3000534, https://doi.org/10.1371/journal.pbio.3000534.

Fliegmann, J., and Felix, G. (2016). Immunity: flagellin seen from all sides. Nat. Plants *2*, 16136.

Fujii, M., Shibata, S., and Aizawa, S.-I. (2008). Polar, peritrichous, and lateral flagella belong to three distinguishable flagellar families. J. Mol. Biol. *379*, 273–283.

Garrido-Oter, R., Nakano, R.T., Dombrowski, N., Ma, K.W., AgBiome Team, McHardy, A.C., Schulze-Lefert, P., and Schulze-Lefert, P. (2018). Modular traits of the Rhizobiales Root microbiota and their evolutionary relationship with symbiotic rhizobia. Cell Host Microbe *24*, 155–167.e5.

Gómez-Gómez, L., Felix, G., and Boller, T. (1999). A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. Plant J *18*, 277–284.

Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P. (2017). Interplay Between innate immunity and the plant microbiota. Annu. Rev. Phytopathol. *55*, 565–589.

Haiko, J., and Westerlund-Wikström, B. (2013). The role of the bacterial flagellum in adhesion and virulence. Biology *2*, 1242–1267.

Hamad, M.A., Zajdowicz, S.L., Holmes, R.K., and Voskuil, M.I. (2009). An allelic exchange system for compliant genetic manipulation of the select agents Burkholderia pseudomallei and Burkholderia mallei. Gene *430*, 123–131.

He, P., Shan, L., Lin, N.C., Martin, G.B., Kemmerling, B., Nürnberger, T., and Sheen, J. (2006). Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in Arabidopsis innate immunity. Cell *125*, 563–575.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. Biom J *50*, 346–363.

Iida, Y., Hobley, L., Lambert, C., Fenton, A.K., Sockett, R.E., and Aizawa, S.-I. (2009). Roles of multiple flagellins in flagellar formation and flagellar growth post bdelloplast lysis in Bdellovibrio bacteriovorus. J. Mol. Biol. *394*, 1011–1021.

Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics *11*, 431.

Kadota, Y., Sklenar, J., Derbyshire, P., Stransfeld, L., Asai, S., Ntoukakis, V., Jones, J.D., Shirasu, K., Menke, F., Jones, A., and Zipfel, C. (2014). Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. Mol. Cell *54*, 43–55.

Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J., et al. (2018). Arabidopsis thaliana and Pseudomonas pathogens exhibit stable associations over evolutionary timescales. Cell Host Microbe *24*, 168–179.e4.

Kühn, M.J., Schmidt, F.K., Farthing, N.E., Rossmann, F.M., Helm, B., Wilson, L.G., Eckhardt, B., and Thormann, K.M. (2018). Spatial arrangement of several flagellins within bacterial flagella improves motility in different environments. Nat. Commun. *9*, 5369.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular Evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. Plant Cell *16*, 3496–3507.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest package: tests in linear mixed effects models. J. Stat. Soft. *82*, 1–26.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res *47*, W256–W259.

Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., et al. (2017). Genomic features of bacterial adaptation to plants. Nat. Genet. *50*, 138–150.

Li, X., Lin, H., Zhang, W., Zou, Y., Zhang, J., Tang, X., and Zhou, J.M. (2005). Flagellin induces innate immunity in nonhost interactions that is suppressed by Pseudomonas syringae effectors. Proc. Natl. Acad. Sci. USA *102*, 12990–12995.

Lin, W., Li, B., Lu, D., Chen, S., Zhu, N., He, P., and Shan, L. (2014). Tyrosine phosphorylation of protein kinase complex BAK1/BIK1 mediates Arabidopsis innate immunity. Proc. Natl. Acad. Sci. USA *111*, 3632–3637.

Lozano-Durán, R., and Belkhadir, Y. (2017). A technical framework for studying the signaling nexus of brassinosteroids and immunity. Methods. Mol. Biol. *1564*, 49–61.

Lu, X., Tintor, N., Mentzel, T., Kombrink, E., Boller, T., Robatzek, S., Schulze-Lefert, P., and Saijo, Y. (2009). Uncoupling of sustained MAMP receptor signaling from early outputs in an Arabidopsis endoplasmic reticulum glucosidase II allele. Proc. Natl. Acad. Sci. USA *106*, 22522–22527.

Meindl, T., Boller, T., and Felix, G. (2000). The bacterial elicitor flagellin activates its receptor in tomato cells according to the address–message concept. Plant Cell *12*, 1783–1794.

Millet, Y.A., Danna, C.H., Clay, N.K., Songnuan, W., Simon, M.D., Werck-Reichhart, D., and Ausubel, F.M. (2010). Innate immune responses activated in Arabidopsis Roots by microbe-associated molecular patterns. Plant Cell *22*, 973–990.

Mueller, K., Bittel, P., Chinchilla, D., Jehle, A.K., Albert, M., Boller, T., and Felix, G. (2012). Chimeric FLS2 receptors reveal the basis for differential flagellin perception in Arabidopsis and tomato. Plant Cell *24*, 2213–2224.

Naito, K., Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2008). Amino acid sequence of bacterial microbe-associated

molecular pattern flg22 is required for virulence. Mol Plant Microbe Interact *21*, 1165–1174.

Nekrasov, V., Li, J., Batoux, M., Roux, M., Chu, Z.H., Lacombe, S., Rougon, A., Bittel, P., Kiss-Papp, M., Chinchilla, D., et al. (2009). Control of the pattern-recognition receptor EFR by an ER protein complex in plant immunity. EMBO J *28*, 3428–3438.

Ohno, S. (1979). Evolution by Gene Duplication (George Allen & Unwin)).

Poncini, L., Wyrsch, I., Tendon, V.D., Vorley, T., Boller, T., Geldner, N., Métraux, J.-P., and Lehmann, S. (2017). In roots of Arabidopsis thaliana, the damage-associated molecular pattern AtPep1 is a stronger elicitor of immune signalling than flg22 or the chitin heptamer. PLoS One *12*, e0185808.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. PLOS One *5*, e9490.

Rich-Griffin, C., Eichmann, R., Reitz, M.U., Hermann, S., Woolley-Allen, K., Brown, P.E., Wiwatdirekkul, K., Esteban, E., Pasha, A., Kogel, K.H., et al. (2020). Regulation of cell type-specific immunity networks in Arabidopsis Roots. Plant Cell *32*, 2742–2762.

Ronald, P.C., and Beutler, B. (2010). Plant and animal sensors of conserved microbial signatures. Science *330*, 1061–1064.

Schwessinger, B., Roux, M., Kadota, Y., Ntoukakis, V., Sklenar, J., Jones, A., and Zipfel, C. (2011). Phosphorylation-dependent differential regulation of plant growth, cell death, and innate immunity by the regulatory receptor-like kinase BAK1. PLoS Genet *7*, e1002046.

Shinya, T., Yamaguchi, K., Desaki, Y., Yamada, K., Narisawa, T., Kobayashi, Y., Maeda, K., Suzuki, M., Tanimoto, T., Takeda, J., et al. (2014). Selective regulation of the chitin-induced defense response by the Arabidopsis receptor-like cytoplasmic kinase PBL27. Plant J *79*, 56–66.

Smakowska-Luzan, E., Mott, G.A., Parys, K., Stegmann, M., Howton, T.C., Layeghifard, M., Neuhold, J., Lehner, A., Kong, J., Grünwald, K., et al. (2018). An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature *553*, 342–346.

Steinbrenner, A.D. (2020). The evolving landscape of cell surface pattern recognition across plant immune networks. Curr. Opin. Plant Biol. *56*, 135–146.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.M., and Chai, J. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science *342*, 624–628.

Teixeira, P.J.P.L., Colaianni, N.R., Fitzpatrick, C.R., and Dangl, J.L. (2019). Beyond pathogens: microbiota interactions with the plant immune system. Curr. Opin. Microbiol. *49*, 7–17.

Thomsen, M.C.F., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Res *40*, W281–W287.

Tran, T.M., Ma, Z., Triebl, A., Nath, S., Cheng, Y., Gong, B.-Q., Han, X., Wang, J., Li, J.-F., Wenk, M.R., et al. (2020). The bacterial quorum sensing signal DSF hijacks Arabidopsis thaliana sterol biosynthesis to suppress plant innate immunity. bioRxiv. https://doi.org/10.1101/2020.01.30.927731.

Vetter, M., Karasov, T.L., and Bergelson, J. (2016). Differentiation between MAMP triggered defenses in Arabidopsis thaliana. PLoS Genet *12*, e1006068.

Wyrsch, I., Domínguez-Ferreras, A., Geldner, N., and Boller, T. (2015). Tissue-specific FLAGELLIN-SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. New Phytol *206*, 774–784.

Zhou, J.M., and Zhang, Y. (2020). Plant immunity: danger perception and signaling. Cell *181*, 978–989.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. Cell *125*, 749–760.

Zhou, F., Emonet, A., Tendon, V.D., Marhavy, P., Wu, D., Lahaye, T., and Geldner, N. (2020). Co-incidence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. Cell *180*.

Please cite this article in press as: Colaianni et al., A complex immune response to flagellin epitope variation in commensal communities, Cell Host & Microbe (2021), https://doi.org/10.1016/j.chom.2021.02.006

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-GFP-HRP	Thermo Fisher Scientific	**Cat**#A10260; RRID: AB_2534022
Rabbit polyclonal anti-BAK1	Agrisera	Cat#AS12 1858; RRID: AB_2884902
anti-HA	Roche	Cat#11867431001; RRID: AB_390919
anti-rat IgG-HRP	Abcam	Cat#ab97057; RRID: AB_10680316
Mouse monoclonal anti-V5-HRP	Sigma Aldrich	Cat# R961-25 RRID: AB_2556565
Mouse monoclonal anti-Flag-HRP	Sigma Aldrich	Cat#A8592 RRID: AB_439702
Bacterial and virus strains		
Pseudomonas syringae pv. *tomato* DC3000*ΔavrPtoA/B*	He et al., 2006	NCBI:txid223283
Ralstonia sp. UNC404CL21Col	Levy et al., 2017	NCBI:txid1380362
Ralstonia sp. CL21Δ*fliC*	This study	N/A
Ralstonia sp. CL21Δ*fliC::Ra* flg22^{5003}	This study	N/A
Ralstonia sp. CL21Δ*fliC::Ra* flg22^{5005}	This study	N/A
NEB 5-alpha *E. coli*	New England Biolabs	Cat#C2987U
E. coli WM3064	William Metcalf (University of Illinois)	N/A
Chemicals, peptides, and recombinant proteins		
flg22 peptides	See Table S2	N/A
Elf18 (SKEKFERTKPHVNVGTIG)	GenScript	N/A
GFP-Trap A beads	Chromotek	Cat#gta100; RRID:AB_2631357
Extracellular domain (ECD) FLS2^{LRR} – Fc-V5	This paper; Smakowska-Luzan et al., 2018	N/A
BAK1^{LRR} – AP-Flag ECD	This paper; Smakowska-Luzan et al., 2018	N/A
Human recombinant IgG1-Fc protein	ThermoFisher Scientific	Cat#A42558
Horseradish Peroxidase	Thermo Fisher	Cat#31490
Luminol	Sigma	Cat#A8511
Complete Mini EDTA-free protease inhibitor cocktail	Roche	Cat#11836170001
Alkaline phosphatase	Sera Care	Cat#5120-0059
HiFi DNA Assembly Mastermix	New England Biolabs	Cat#E2621S
Q5 DNA Polymerase	New England Biolabs	Cat#M0491S
DpnI	New England Biolabs	Cat#R0176S
ExpreS^2 Insect-TRx5 Transfection Reagent	Expression System	Cat#S2-55A-001
Deposited data		
Plant-Associated Database (3837 genome)	Levy et al., 2017	http://labs.bio.unc.edu/Dangl/Resources/gfobap_website/faa_trees_metadata.html
Arabidopsis-Associated Database	This study	https://github.com/ncolaian/NatVar_proj
Ralstonia protein database	This study	https://github.com/ncolaian/NatVar_proj
Experimental models: cell lines		
Drosophila melanogaster Schneider 2 (S2) cells	Protein Technologies Facility of the Vienna Biocenter Core Facilities	N/A
Experimental models: organisms/strains		
Arabidopsis thaliana: WT Col-0	Dangl lab stock	NCBI:txid3702
Arabidopsis: fls2 efr	Nekrasov et al., 2009	SAIL_691_C4 / Salk_044334
Arabidopsis: bak1-5	Schwessinger et al., 2011	*elfin* mutants

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Arabidopsis: proFLS2:FLS2-GFP	Chinchilla et al., 2007	*fls2* mutant
Arabidopsis: proMYB51:NLS-3mVenus	Poncini et al., 2017	Transgenic Col-0
Arabidopsis: proBIK1:BIK1-HA	Lin et al., 2014	Transgenic Col-0
Arabidopsis: proWRKY11:NLS-3mVenus	Poncini et al., 2017	Transgenic Col-0
Arabidopsis: proUBQ10:FLS2-GFP	Wyrsch et al., 2015	Transgenic Col-0
Arabidopsis: proCYP71A12:GUS	Millet et al., 2010	Transgenic Col-0
Arabidopsis: bak1-4	Chinchilla et al., 2007	Salk_116202
Oligonucleotides		
DNA primers	Eurofins	See Table S3
450 bp DNA synthesis *Ra* flg22^{5003} & *Ra* E-flg22^{5005}	Genewiz	See Table S3
Recombinant DNA		
pMo130-TelR	Amin et al., 2013, Addgene	Plasmid #50799
pLysS (isolated from BL21(DE3) pLysS *E. coli*)	Millipore Sigma	Cat#69451-3
pMo130-cmR	This study	N/A
pJMC168	This study	N/A
pJMC174	This study	N/A
pJMC176	This study	N/A
pJMC182	This study	N/A
pECIA-2-FLS2^{LRR} ECD	Smakowska-Luzan et al., 2018	Addgene: #115120
pECIA-14-BAK1^{LRR} ECD	Smakowska-Luzan et al., 2018	Addgene: #114775
Software and algorithms		
Psiblast	Altschul et al., 1997	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download
HMMER3	Johnson et al., 2010	http://hmmer.org/
MUSCLE	Edgar, 2004	http://www.drive5.com/muscle/
FastTree	Price et al., 2010	http://www.microbesonline.org/fasttree/#Install
ITOL	Letunic and Bork, 2019	https://itol.embl.de/
trimAl	Capella-Gutiérrez et al., 2009	http://trimal.cgenomics.org/
Seq2logo	Thomsen and Nielsen, 2012	http://www.cbs.dtu.dk/biotools/Seq2Logo/
Bolstad2	Bolstad, 2010	https://cran.r-project.org/web/packages/Bolstad2/index.html
lme4	Bates et al., 2015	https://cran.r-project.org/web/packages/lme4/index.html
lmerTest	Kuznetsova et al., 2017	https://cran.r-project.org/web/packages/lmerTest/index.html
MEGA7	Kumar et al., 2016	https://www.megasoftware.net/
Agricolae	N/A	https://cran.r-project.org/web/packages/agricolae/index.html
Multcomp	Hothorn et al., 2008	https://cran.r-project.org/web/packages/multcomp/index.html
Other		
Resources and code used for all analyses	This study	https://github.com/ncolaian/NatVar_proj

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeffery L. Dangl (dangl@email.unc.edu).

Materials availability

Bacterial mutants generated in this study are available upon request.

Data and code availability

The code and source data used for the figures and analyses in this study are available on github at https://github.com/ncolaian/NatVar_proj.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plants

Arabidopsis thaliana accession Columbia (Col-0) was used as wild type (WT) in all experiments performed. The mutant lines *fls2 efr* (SAIL_691_C4 / Salk_044334), *bak1-4, and bak1-5* were described elsewhere (Nekrasov et al., 2009, Chinchilla et al., 2007, Schwessinger et al., 2011). For all experiments testing the necessity of FLS2, *fls2 efr* mutant plants were used. The EFR and FLS2 receptors are genetically independent, and their respective immunogens are significantly different from one another. It is extremely unlikely that any flg22 variant would signal via EFR, and data shows this to be so (Nekrasov et al., 2009, Zipfel et al., 2006). The transgenic lines *proFLS2:FLS2-GFP* (Chinchilla et al., 2007), *proMYB51:NLS-3mVenus* (Poncini et al., 2017), *proWRKY11:NLS-3mVenus* (Poncini et al., 2017), *proCYP71A12:GUS* (Millet et al., 2010), *proBIK1:BIK1-HA* (Lin et al., 2014), and *proUBQ10:FLS2-GFP* (Wyrsch et al., 2015) were described previously. Plants were grown on soil in short-day light conditions (12 h light/12 h dark) or vertically in Petri dishes containing 1/2 Murashige and Skoog (MS) medium, 0.8% plant agar, and 1% sucrose in long-day light conditions (16 h light/8 h dark) unless otherwise specified.

Bacteria

The bacterial strains that were not generated for this study include: *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B* (He et al., 2006) and *Ralstonia sp.* UNC404CL21Col (Genome ID 2558309150, henceforth CL21, Levy et al., 2017). The CL21Δ*fliC*, CL21Δ*fliC::Ra* flg22^{5003}, and CL21Δ*fliC::Ra* flg22^{5005} strains generated in this study were created from the *Ralstonia_sp._UNC404CL21* strain. All bacteria were stored in 40% glycerol stocks at -80°C.

Genomic information

We downloaded all 3837 genomes of plant-associated bacteria from a previously built database (Levy et al., 2017). We used the metadata to identify genomes isolated from Arabidopsis, resulting in 627 genomes (Table S1). Of these, 390 isolates were collected from the root rhizoplane and root endophytic compartments, while the rest are derived from rhizosphere and leaf communities (Table S1). We downloaded 156 *Ralstonia* genomes from NCBI on 12/20/2018 (database on Github at https://github.com/ncolaian/NatVar_proj) and extracted 152 FliC genes that we used for subsequent analyses (Table S4).

METHOD DETAILS

FliC identification and phylogenetic tree creation

387 high quality FliC proteins were selected by first performing psiblast on the 3837 genome Levy et al. database (Altschul et al., 1997, Levy et al., 2017) with the FliC protein from *Pseudomonas aeruginosa*. These results were then filtered based on PFAM and KO annotations. Each protein used is annotated with at least one of the following: flagellin PFAM annotations of pfam00669 and pfam00700, and/or the flagellin KO annotation of K02406. This resulted in 387 FliC proteins (Data S1) which we then aligned with MUSCLE (Edgar, 2004). We used the resulting multiple sequence alignment to create a hidden Markov model (HMM) profile with hmmbuild and then scanned each genome in our database using hmmscan from the HMMER3 software package (Johnson et al., 2010). We only considered genes with an e-value equal to or greater than 10^{-50} as encoding FliC proteins in all subsequent analyses. This cutoff produced high specificity for FliC proteins; we found no other genes in our dataset at this threshold.

The resulting proteins were aligned using MUSCLE (Edgar, 2004). The alignment was trimmed to include columns containing amino acids from 90% of the sequences with trimAl (Capella-Gutiérrez et al., 2009), and the protein phylogeny was created using FastTree (Price et al., 2010). Branch length is estimated by FastTree and is interpreted as the substation rate between FliC proteins, and is normalized to the substitution rate of the FliC proteins to randomly generated proteins (Price et al., 2010). We observed, at a minimum, three distinct clades. It is possible that there could be four based on the node support calculated by a Shimodaira-Hasegawa test from FastTree (Price et al., 2010). For simplicity, we collapsed the Bacilli and Actinobacteria FliC clades together since one potential clade contained only 20 FliC proteins (Figure S1A). Visualization of the phylogenetic tree was performed with the online tool ITOL (Letunic and Bork, 2019). It is important to note that the node support around the clade separations was much stronger than the

support closer to the tips of the tree. For visualization purposes, we did not collapse weakly supported nodes, so please be aware that the relationships around the tips of the FliC tree in Figure 1A may not be fully resolved.

The median number of *fliC* genes in genomes encoding clade 3 FliCs is two, while the median number for clades 1 and 2 is one. Bacteria with clade 3 FliCs can encode as many as 7 *fliC* genes (Figure 1B). This is consistent with observations in *Bdellovibrio bacteriovorus* which encodes 6 *fliC* genes; all FliC proteins produced from these genes are incorporated into the flagellum (Iida et al., 2009). We found that 94% of genomes with more than one *fliC* gene contained FliC proteins from the same clade, indicating that horizontal gene transfer of *fliC* genes is rare (Figure S1B).

Identification of flg22 sequences

Flg22 epitope amino acid sequences were identified from the FliC proteins by first splitting the FliC proteins by clade and aligning each clade to the *P. aeruginosa* flg22 sequence using MUSCLE (Edgar, 2004). The FliC proteins identified in *Ralstonia* were also aligned to the *P. aeruginosa* flg22. Then, using a custom script, a continuous 22 amino acid stretch of sequence was obtained based on the beginning (N-terminus) of the alignment of each FliC sequence to flg22 in the multiple sequence alignment. In order to ensure proper alignment, position 14 was checked for aspartic acid (D), as this position is the most conserved residue in the flg22 region based on previously known flg22 sequences. Any flg22 sequence not containing aspartic acid (D) at position 14 was flagged for manual curation or discarded.

Position Weight Matrices (PWM)

The flg22 sequences were then used to create a position weight matrix (PWM) using a custom script. The PWM matrices were created using a pseudocount that was adjusted based on equal background amino acid frequencies and the number of flg22 sequences being analyzed. These PWMs were then visualized with the online tool Seq2logo (Thomsen and Nielsen, 2012).

fliC operon analysis

We defined an operon as units of *fliC* genes that were within 10 kb of each other. This means that an operon could span over 10 kb, but each *fliC* would be no more than 10 kb from another. They were identified using a custom script.

Peptide synthesis

All flg22 peptides listed in Table S2 were synthesized at >95% purity by an in-house protein chemistry facility (Gregor Mendel Institute, GMI), Shanghai Apeptide Co., and/or GenScript. All experiments were performed using peptide synthesized from GMI or Shanghai Apeptide Co, except for the leaf infection assay performed using peptides synthesized from GenScript. All synthesized peptides were dissolved in 200 μL of pure water and their concentration was quantified using the Direct Detect Spectrometer (Millipore Sigma). These stock solutions were used to prepare 100 μM aliquots of the peptides and stored at -20 °C. The purity of all peptides used were validated with MALDI-TOF MS before and after ROS, SGI, GUS, and gene expression assays. Elf18 peptide (SKEKFERTKPHVNVGTIG, Kunze et al., 2004) was ordered from GenScript.

Reactive Oxygen Species (ROS) burst

ROS burst assays were performed using a luminol based assay (Lozano-Durán and Belkhadir, 2017). 5mm leaf disks were taken from the leaves of healthy 6-week-old Arabidopsis plants and placed adaxial side up in a 96-well plate (Greiner bio-one, ref 655075) with 100μL of distilled water overnight. The distilled water was replaced with a solution of 200μM luminol (SIGMA Cat. No. A8511) and 10μg/mL of Horseradish Peroxidase (Fisher Cat. No. 31490) with or without a flg22 peptide. The plate was immediately placed into the luminometer (SpectraMax L, Molecular Devices) where luminescence was recorded each minute for 50 minutes at the target wavelength of 470nm using an integration time of 0.35 s. The peptide concentrations used in the experiments are specified in the figure legends.

The antagonistic ROS burst was performed as described above. The only difference is the timing of flg22 peptide addition to the plants. 50 μL of reaction solution supplemented with potential antagonist at double the final concentration was added to each well and placed in the luminometer for 15 minutes. After, 50 μL of the Pta flg22 peptide solution at double the final concentration was added to the mixture and placed into the luminometer for 50 minutes. The final concentrations of flg22 peptides are specified in each figure legend.

ROS data analysis

Area under the ROS kinetics curve (AUC) was calculated by integrating the area under the ROS burst curve from 0-50 minutes with the sintegral function from the Bolstad2 (1.0-28, Bolstad, 2010) package in R.

In order to compare ROS curves across multiple plates, 12 AUC measurements for loss of function flg22 peptide derived from *Pseudomonas syringae* pv. *tabaci* (PtaDA - TRLSSGLKINSAKADAAGLQIA) (Naito et al., 2008) were recorded on every plate. A z-score was calculated for each leaf disk using the AUC mean and standard deviation from the PtaDA replicates from the plate on which it was measured (Figures 2A–2C and 6B–6K).

For the ROS burst experiments comparing the response of each peptide in *fls2 efr* and Col-0 plants, we paired *fls2 efr* and Col-0 plants on the plate for each peptide. We adjusted the Log_{10}(AUC) measurements from each plate by subtracting all AUC measurements by the median *fls2 efr* value on the plate, n=48 (Figure S2B). 24 AUC measurements from 2 independent experiments

from *fls2 efr* and Col-0 plants exposed to each peptide variant were compared using a t-test with the stat_compare_means function from ggpubr (0.2.5).

For the 10nM ROS burst experiments (Figure S2C), we could not adjust the plate measurements with the PtaDA based Z-score method because it is inactive at 10nM. We still included PtaDA on every plate, and used a linear mixed model with the model formula:

$$AUC \sim Name + (1|plate)$$

Where *AUC* is the Log_{10} area under the curve values, *Name* is peptide ID, and *(1|plate)* models the random plate effect to compare all the peptides to the PtaDA peptide with a two-sided t-test using the lme4 (1.1-24, Bates et al., 2015) and lmerTest (3.1-1, Kuznetsova et al., 2017) packages in R. *P*-values were adjusted with the Benjamini & Hochberg method (FDR) in R (Benjamini and Hochberg, 1995).

Seedling growth inhibition (SGI) assay

Following a defined protocol (Gómez-Gómez et al., 1999), Col-0 or *fls2 efr* Arabidopsis seeds were surface sterilized using the standard bleach sterilization method and then stratified for two days at 4°C. Seeds were then germinated on ½ MS plates containing 0.8% plant agar and 1% sucrose for five days at 22°C. The seedlings were then transferred to a 48-well plate (Griner Bio one, Cat. No. 677180) containing 1 mL of liquid ½ MS medium supplemented with 10 nM of flg22 peptide. After nine days of growth, the seedlings were patted dry and weighed. Each experiment included 8 seedlings per treatment. A mock control and canonical active flg22 sequence from *Pseudomonas aeruginosa* (Pa22- QRLSTGSRINSAKDDAAGLQIA) were assayed on each plate.

SGI normalization and modeling

To normalize the data points across the experiments, a z-score was calculated for each weight by using the mean and standard deviation of the weights observed for the mock controls on each plate. Each flg22 peptide was tested on least 3 different plates for Col-0 experiments (Figures 2A–2C) and only one plate for *fls2 efr* experiments (Figure S2D). A linear-mixed model was implemented to identify variants that significantly reduced the growth Arabidopsis compared to the mock controls, while controlling for the still significant plate effect. The model is:

$$Z(weight) \sim ID + Exp + (1|plate\_exp)$$

Z(weight) is the z-score calculated for each weight measurement, *ID* is equal to the variant, *Exp* describes the batch the plate was performed in, and *plate_exp* is a unique ID given to each plate using the lme4 (1.1-24, Bates et al., 2015) package in R. A *P*-value was estimated with a two-sided t-test using the lmerTest (3.1-1, Kuznetsova et al., 2017) package in R, and then adjusted with the Benjamini & Hochberg method (FDR) in R (Benjamini and Hochberg, 1995).

Flg22 tree creation

A maximum likelihood tree was created with a WAG substitution matrix using the MEGA7 software package (Kumar et al., 2016) and the flg22 amino acid sequences. The tree was then visualized with the data layered on the tree using the online software ITOL (Letunic and Bork, 2019).

Antagonism of flg22 driven CYP71A12:GUS expression in Arabidopsis seedlings

5-8 proCYP71A12:GUS plants were germinated per well in a 48-well plate (Griner Bio one, Cat. No. 677180) in liquid MS (Murashige and Skoog basal medium with vitamins containing 0.5 g/L MES hydrate and 0.5% sucrose at pH 5.7; Millet et al., 2010). After 7 days, a solution of liquid MS and antagonist peptide at final concentration of 100 μM was added and pre-incubated with seedlings for 15 minutes. Next, the media was supplemented with 10 nM Pa22 and incubated for 5 hours at 21°C. The media was then removed, and each well was washed with 50 mM sodium phosphate (pH 7). The GUS substrate solution was added (50 mM Na3PO4, 10 mM EDTA, 0.5 mM K4[Fe(CN)6], 0.5 mM K3[Fe(CN)6], 0.5 mM X-Gluc, 0.01 % Silwet L-77) and incubated overnight in the dark at 37°C. Seedlings were then fixed in 3:1 EtOH:acetic acid at 4°C for 5 hours and stored in 95% EtOH. Root pictures were taken using a Leica M205FA stereoscope coupled to a Leica DFC310FX camera.

Protein extraction and Co-immunoprecipitation in Arabidopsis

proFLS2:FLS2-GFP Arabidopsis thaliana plants were grown for two weeks on ½ MS agar plates with 1% sucrose in long-day photoperiod. About 0.5 g of plant material was transferred into 6-well plates (Griner Bio one, Cat. No. 657185). 2mL of water supplemented with each flg22 variant was added to the wells and the plants were vacuum infiltrated for 15 minutes. Plants were then dried and ground to a fine powder in liquid nitrogen. The ground plant tissue was then incubated with ~ 2 mL of protein extraction buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10% glycerol, 5 mM EDTA, 1 mM Na_2MoO_4, 20 mM NaF, with fresh 1 mM DTT, and 1 protease inhibitor cocktail tablet/50 mL) and 1% IGEPAL for 30 minutes at 4°C with rotation. The samples were then centrifuged for 10 min at 4°C, 1300 rpm, and the supernatant was filtered through Miracloth. 100μL of the filtered supernatant was mixed with 100 μL of Laemmli sample buffer (2x) as input.

The supernatant was diluted 1:1 in extraction buffer without IGEPAL, and 15 μl of equilibrated GFP-Trap A beads (Chromotek, Cat. No. gta-100) were added to each sample. The samples were incubated for 4 hours at 4°C with rotation. The supernatant was discarded and the beads were washed three times in washing buffer (50 mM Tris pH 7.5 and 100 mM NaCl) for 5 minutes at 4°C

with rotation (800g). The elution of the protein complexes was performed in 100μL of Laemmli sample buffer (2x), boiled for 10 min at 95°C. This was subsequently analyzed by SDS-PAGE followed by immunoblotting using anti-GFP-HRP (Thermo Fisher Scientific, Cat. No. A10260) and anti-BAK1 (Agrisera, Cat. No. AS12 1858) antibodies. 45 μL of material from the bead extractions was loaded in the western blot analysis for the Co-IP image.

BIK phosphorylation assay

Six 2-week old *proBIK1:BIK1-HA* seedlings grown on ½ MS plates were transferred to 6-well plates (Griner Bio one, Cat. No. 657185) containing 5 mL of sterile water. After 30 min, the water was replaced by elicitor solution at the final concentration described in the figures and legends. For assays with the priming antagonist, a solution containing the priming antagonist was added 15 minutes before the addition of the elicitor. The seedlings were then vacuum infiltrated for 10 minutes and left to incubate for 30 minutes. Immediately afterwards, the seedlings were transferred to 2 mL tubes containing glass beads and frozen with liquid N_2. Tissues were homogenized with TissueLyser II (Qiagen) and 300 μL of protein extraction buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 10 % glycerol, 5 mM EDTA, 1 mM Na_2MoO_4, 20 mM NaF, 1 mM DTT, Complete Mini EDTA-free protease inhibitor cocktail (Roche) was added. After centrifugation (20 min, 20,000 *g*, 4°C), the supernatant mixed with Laemmli sample buffer and denatured for 5 min at 95°C. The proteins were resolved in 10% SDS-PAGE gels and transferred to nitrocellulose membrane (GE healthcare). HA-tagged BIK1 proteins were immunoblotted with primary anti-HA (Roche, 1:1000) and secondary anti-rat IgG-HRP (Abcam, 1:5000).

Western blot imaging

Western blots were imaged two separate ways. For Figures S3F, 5B, and 5C, proteins were illuminated using Amersham ECL Prime Western Blotting Detection Reagent (Cytiva, RPN2236). Images were taken by placing the membrane in a dark box and taking pictures with a digital camera using KwikQuant Imager (Kindle Bio). For Figure 3B, proteins were illuminated using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, REF 34579). Images were taken using KODAK X-OMAT 5000 RA Processor and CL-XPosure™ Film (12,5x17,5 cm) (Fisher Scientific).

In vitro FLS2-BAK1 ectodomain interaction studies

This assay was performed as described in (Parys et al., submitted, see Figure S3B). Briefly, Extracellular Domains (ECDs) of FLS2 cloned in pECIA2 (bait protein, plate bound) and BAK1 cloned in pECIA14 vector (prey protein, interacts with substrate) were transiently expressed in Drosophila melanogaster Schneider 2 (S2) cells. The protein expression was then confirmed by immunoblotting using anti-V5 (Invitrogen, Cat. No. R961-25) and anti-Flag (Sigma Aldrich, Cat. No. A8592) antibodies. Following, the $FLS2^{ECD}$ was diluted 20 times in 1xPBS containing 0.1 % Tween-20 (PBS-T), and mixed 1:1 with BAK1-ECD diluted 40 times. Accordingly, the Pta22 peptide was added to the protein solution to a final concentration of 10 nM along with the antagonistic flg22 peptide in final concentration ranging from 0.01 – 10 μM. The protein mix was pre-incubated 2 h in 4°C. Following, 100 μl of the protein solution was transferred to a protein-A coated 96 well plate (Thermo Fisher Scientific, ref 15132) and incubated overnight in 4°C. The next day the plate was washed two times with 100 μl of PBS-T and 100 μl of alkaline phosphatase (Sera care, Cat. No. 5120-0059), the substrate was added. Two hours upon addition of the substrate, the absorbance was measured at 650 nm using a Synergy H4 Multi-Mode plate reader (BioTek). Each plate contained BIR4-BAK1 positive control (Smakowska-Luzan et al., 2018). To control for the prey unspecific binding to protein A coated wells, the bait protein was substituted with a solution of human recombinant IgG1-Fc protein (Invitrogen™ Sino Biological™) at final concentration 625 pg/ul. For each flg22 peptide tested, relative A650 was quantified by dividing the raw A650 after 2h measurement to the signal of mock control (IgG1-Fc).

Microscopic observations of flg22 elicited expression of fluorescent reporter genes

5-days-old seedlings grown on ½ MS with 1% sucrose agar plates were transferred to liquid medium (½ MS + 1% sucrose) with each flg22 peptide variant at 10 nM. 24hr after treatment, confocal laser scanning microscopy (CLSM) experiments were performed by Zeiss LSM780. Excitation and detection wavelengths were set as 488 nm and 493-598 nm respectively. The adaxial side of leaves for *proWRKY11:NLS-3xmVenus* and meristem zone of roots for *proMYB51:NLS-3xmVenus* were examined by CLSM. Images were processed and analyzed using the Fiji software.

Vector construction

Vectors were constructed for the deletion of the *Ralstonia sp.* UNC404CL21Col (CL21) *fliC* (pJMC168) and subsequent insertion of various *fliC* alleles containing altered flg22 coding sequences (pJMC174, pJMC176, and pJMC182). All of these vectors were constructed by Gibson Assembly using HiFi DNA Assembly Mastermix (New England Biolabs) to assemble DNA fragments amplified using the Q5 DNA Polymerase including the optional GC enhancer (New England Biolabs). DNA fragments amplified from vector templates were treated with DpnI (New England Biolabs) to digest vector template DNA prior to inclusion in Gibson Assembly reactions, as appropriate. All DNA fragments were cleaned-up as necessary using the QIAquick PCR Purification Kit (Qiagen). All assembled vectors were transformed into NEB 5-alpha chemically competent *E. coli* (New England Biolabs) and selected on LB media with 30 μg/mL chloramphenicol. Vectors were miniprepped using the ZR Plasmid Miniprep Classic Kit (Zymo Research) and sequence confirmed by Sanger Sequencing (Genewiz). All primer sequences and vectors used in this study are available in Table S3.

Because CL21 is naturally resistant to kanamycin, the antibiotic resistance marker on suicide vector pMo130 (Hamad et al., 2009) was changed from kanamycin resistance to chloramphenicol resistance. Vector pMo130-cmR was assembled using two fragments:

the pMo130 vector amplified using primers JMC661-JMC662, and the chloramphenicol resistance gene and its promoter amplified from pLysS using primers JMC663-JMC664. This new vector, pMo130-cmR, was used as the base vector for subsequent vectors. For the knockout vector pJMC168, regions flanking *fliC* were amplified from CL21 genomic DNA using primers JMC635-JMC636 for the upstream flanking region and primers JMC637-JMC638 for the downstream flanking region. These were assembled with the pMo130-cmR backbone amplified with primers JMC203-JMC204. To construct a base vector for the *fliC* variant knock-in vectors (pJMC174), primers JMC635-JMC638 were used to amplify the region containing the wild type *fliC* and both flanking regions. This was assembled with the pMo130-cmR backbone amplified with primers JMC203-JMC204 resulting in pJMC174. The *fliC* variant knock-in vectors (pJMC176 and pJMC182) were constructed by assembling a 450bp DNA synthesis product (Genewiz), starting 3bp before the *fliC* start codon and containing an altered flg22 coding sequence in *fliC* (*Ra* E-flg22^{5005} in pJMC176 and *Ra* flg22^{5003} in pJMC182), with the vector backbone amplified from pJMC174 using primers JMC719-JMC720. The DNA synthesis product sequences for these vectors can be found in Table S3. To prepare for biparental mating with CL21, all vectors were transformed into biparental mating *E. coli* strain WM3064 and selected on LB agar plates containing 30 μg/mL chloramphenicol and 0.3 mM diaminopimelic acid (DAP) at 37°C.

CL21Δ*fliC* strain construction

The unmarked deletion of *fliC* (Gene ID 2558853875) in strain *Ralstonia sp.* UNC404CL21Col (CL21) was constructed using two-step allelic exchange based on a genetic system developed for *Burkholderia* spp. (Hamad et al., 2009). Biparental mating between *E. coli* strain WM3064 containing knockout vector pJMC168 and CL21 was performed by growing each strain overnight: *E. coli* containing pJMC168 in LB medium with 30 μg/mL chloramphenicol and 0.3 mM diaminopimelic acid (DAP) at 37°C and CL21 in 2xYT medium containing 100μg/mL ampicillin at 28°C. The strains were washed three times with 2xYT medium lacking antibiotics, resuspended in 1/10 the volume and mixed in equal proportion donor:recipient, and plated on LB agar containing 0.3 mM DAP and grown overnight at 28°C. Exconjugates were selected on LB agar plates containing 150 μg/mL chloramphenicol, 100μg/mL ampicillin, and lacking DAP at 28°C. First crossover strains were confirmed for the insertion of knockout vector pJMC168 at the correct genomic location in CL21 exconjugant strains using primers outside the regions of homologous recombination (JMC655, JMC656) and primers in the pMo130-cmR vector backbone (JMC321, JMC634). Confirmed first crossovers were resolved by passaging one time on 2xYT medium containing 100μg/mL ampicillin and 1mM IPTG. This was followed by one passage on media containing 10g/L tryptone, 5 g/L yeast extract, 100 g/L sucrose, 100μg/mL ampicillin, and 1mM IPTG, and then plating the strains on the same media containing 1.5% agar. The resulting strains were screened for the deletion of *fliC* using primers JMC655-JMC656. Deletion strains were plate-purified by streaking two additional times on LB agar plates containing 100μg/mL ampicillin. The resulting strain was confirmed by PCR using JMC655-JMC656 and by confirming the loss of motility using the motility assay described below. The final strain is designated CL21Δ*fliC*.

CL21 *fliC* allele knock-in

The same two-step allelic exchange procedure used to construct the CL21Δ*fliC* strain was used to insert different alleles of *fliC* containing altered flg22 coding sequences into the CL21Δ*fliC* strain background. Biparental matings of *E. coli* WM3064 strains containing vector pJMC176 or pJMC182 were performed with strain CL21Δ*fliC.* First crossovers exconjugants were selected on LB agar plates containing 150 μg/mL chloramphenicol, 100μg/mL ampicillin, and lacking DAP at 28°C. The vector insertion was resolved as described above and the resulting strains were screened by PCR with primers JMC655-JMC656. Strains with a *fliC* insertion were passaged two times on LB agar plates containing 100μg/mL ampicillin before performing PCR confirmation with primers JMC655-JMC656. This PCR product was sequenced to confirm the insertion of the correct *fliC* allele. The resulting strains using the pJMC176 and pJMC182 vectors were named CL21Δ*fliC::Ra* E-flg22^{5005} and CL21Δ*fliC::Ra* flg22^{5003} respectively based on the swapped in flg22 amino acid sequences displayed in Figure S6A.

Motility assay

To assess the motility of strains, LB medium plates solidified with 0.3% (w/v) agar were prepared. To prepare an inoculum for the motility assay, strains were grown overnight in 2xYT medium. The motility plates were inoculated by dipping a sterile toothpick into the overnight culture and stabbing it once into the center of the LB 0.3% agar plate. Plates were incubated at 28°C for four days and images taken on a document scanner. Motility distance was quantified as the farthest point of bacterial spread from the inoculation point using ImageJ.

Root Growth Inhibition (RGI) assay

proUBQ10:FLS2-GFP seeds were sterilized by vigorously shaking with a 60% bleach solution for 7 minutes followed by washing with sterilized water four times. Seeds were stratified in the dark at 4°C for 48 hr, and then placed onto sterile Murashige & Skoog (MS) agar plates (M404; Phytotech labs, Lenexa, Kansas) containing 0.5% sucrose. The seedlings were then grown vertically for seven days in a 16-hr light, 21°C / 8-hr dark, 18°C regime.

Bacteria were grown in the dark on Luria Broth (LB) agar supplemented with 100 μg/mL ampicillin for two days at 28°C. Single colonies were selected and grown overnight in 2xYT liquid media at 28°C while being shaken vigorously. The next day, bacteria were centrifuged, then washed three times, and diluted to OD_{600} 0.0002 using 10 mM $MgCl_2$. Bacteria were then spread onto MS plates containing no sucrose with or without 100 nM Pa22. Six 7-day old *proUBQ10:FLS2-GFP* seedlings were transferred onto

each plate, and the root length at time of transfer was demarcated. The plants were then grown vertically for seven more days at the conditions above. Plates were scanned with a document scanner, and root elongation measured using a ruler. Significant differences between bacterial treatments were identified using a linear mixed model with lme4 (1.1-24) that controlled for a significant random plate effect and a significant interaction between batch and bacterial treatment (Bates et al., 2015).

Leaf infection assays

We followed a previously described protocol (Chung et al., 2014). The day before bacterial inoculation, water or 100nM solution of the flg22 peptide variant indicated was infiltrated into leaves of 5-week-old Col-0 Arabidopsis plants using a needless syringe. *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B* was plated on King's B medium (KB) supplemented with 50 μg/mL rifampicin and 50 μg/mL kanamycin. The bacteria were then grown overnight in the dark at 28°C. The next day, 1 mL of 10 mM $MgCl_2$ was added to the bacterial plates and carefully scraped to collect bacteria. The optical density (OD) was immediately measured at OD^{600} and a solution of 1 x 10^5 cfu/mL (OD = 0.0002) of *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B* and 10 mM $MgCl_2$ was infiltrated into previously pretreated leaves using a needless syringe. Leaves were allowed to dry (~3 hours), and then placed in short day conditions (8h light, 16h dark) at 21°C for 3 days.

After, four 5 mm leaf disks were taken from four separate plants and placed into a sterile 2 mL Eppendorf tube containing three 4mm beads and 400 μL of distilled water. The plant tissue was then lysed for 45 seconds using the FastPrep-24™ (MP Biomedicals). We then added 600 μL of distilled water to each sample, and performed 10x serial dilutions (20 μL in 180 μL of distilled water). 5 μL from each dilution (0-6) were plated on KB plates supplemented with 50 μg/mL rifampicin and 50 μg/mL kanamycin. These plates were placed in the dark at 28°C for ~24-hr and then CFUs were counted under a microscope. We determined statistically different groups using a two-way ANOVA controlling for batch effect and a post-hoc Tukey test as implemented by the HSD.test function in the agricolae (1.3-2) package.

16S community analysis

We downloaded 16S data reported by a previous study that establishes a 185-member SynCom (Finkel et al., 2019). Briefly, soil was collected from the long-term Pi fertilization field ("Field D") trial at the Julius Kühn Experimental Station at Martin Luther University of Halle-Wittenberg (51°29′45.6″N, 11°59′33.3″E). The 16S community composition was determined by forming amplicon sequence variants (ASVs) and assigning taxonomy to these ASVs. The soil, root, and shoot samples are from the collected soil samples (Figure S6A). The analysis of the 185-SynCom inoculum sample was performed by mapping 16S reads to each unique 16S sequence variant (Useq) present in the 185-member SynCom (Finkel et al., 2020). For ease of comparison, the inoculum and Arabidopsis database stacked bar figures were normalized to the wild root abundance assigned to classes present in the Arabidopsis database (Figure S6A).

FliC and flg22 analysis of 185-member SynCom

We assessed how variation in FliC and flg22 could affect and be affected by community flg22 composition. We downloaded 16S sequence data generated from a study of bacterial colonization of the plant using an inoculation of a 185-member synthetic community (SynCom) on 7-day old Col-0 seedlings (Finkel et al., 2020). The community was allowed to colonize seedlings for 12 days under a 16-hr dark/8-hr light regime at 21°C light/18°C dark. We reanalyzed data from control plates used in the 1000 μM Pi or 21°C conditions and from plates supplemented with 200 mM of NaCl as the salt stressed condition. For all analyses we used the relative abundance values for each unique sequence variant (Useq) reported by the authors (Finkel et al., 2020). Each Useq is a unique V3-V4 16S variant that is associated with one or more strains in the 185 SynCom.

We first determined the approximate number of each clade of FliC in each community. To estimate the relative abundance of genomes containing each FliC clade (Figure 7A), we assigned the relative abundance value of each Useq to the FliC clades found within each Useq. If a Useq was associated with multiple FliC clades or had genomes with or without a *fliC* gene, the Useq relative abundance values were assigned to all the FliC clades present and/or to the no FliC group. Thus, a Useq could contribute to the relative abundance of multiple FliC clades. However, even if an Useq contained genomes with multiple FliCs from the same clade, the relative abundance value was counted once for each FliC clade. The relative abundance values for each FliC clade were then tabulated for each sample. Statistically significant differences in relative abundance values for FliC clades within and between plant fractions were identified using a two-way ANOVA, followed by a post-hoc Tukey test using the HSD.test function as implemented by the agricolae (1.3-2) package in R. Our model controlled for a significant interaction between the relative abundance of bacteria and the two different conditions/batches.

We next assessed the functional flg22 repertoire within the communities by assigning each unique flg22 variant within a Useq the relative abundance of that Useq. In principle, this means that a Useq could contribute its relative abundance to as many unique flg22 variants as are harbored within that Useq. To be conservative, the relative abundance values were only counted once for each flg22 variant found within the Useq even if it contained multiple copies of the same flg22 variant. The ratio of un-tested flg22 variants was determined by dividing the total abundance of the flg22 variants not in our screen by the total flg22 abundance we calculated.

$$f_{untested} = \frac{\sum x_{unobserved}}{\sum \hat{x}_{flg22}}$$

Where x and $\hat{x}$ represent relative abundance values for unobserved and all flg22 variants respectively. We calculated the percentage of the untested abundance from each FliC clade by assigning the relative abundance of each flg22 variant to the FliC clade it is found in (Figure S6B). The functional flg22 repertoire was summarized from each flg22 variant's relative abundance that fell into the five functional classes, Antagonist, Deviant, Evading, Canonical, and Not tested (definitions in this manuscript). Similarly to above, relative abundance values for all flg22 variants in a functional class was divided by the total flg22 relative abundance to get the percentage of the total flg22 relative abundance associated with each functional class.

Flg22 amino-acid sequence logos corresponding to each experimental condition were created by representing each flg22 sequence proportional to its median relative abundance value across all control conditions. The median relative abundance values calculated over all control samples analyzed were multiplied by 1000 and rounded to the nearest integer. This value reflects the number of times this flg22 amino-acid sequence was represented in the complete flg22 functional repertoire list. This list was then used to create PWM matrices and logos.

The analysis of individual flg22 variants present in each sample was performed on a subset of active flg22 variants. The flg22 percent relative abundance values were used to identify active flg22 variants that had percent relative abundance values significantly different from zero using a linear model within the agar, root, and shoot fractions. We plotted the percent of flg22 abundance over all samples for each of the flg22 variants that were significantly different to zero in any of the fractions (Figure 7E).

QUANTIFICATION AND STATISTICAL ANALYSIS

Programs used for statistical analysis and data visualization

GraphPad PRISM 8.0 and the R programming environment (version R 3.6.2) were used for data analysis and visualization. Unless specified otherwise, boxplots represent the 1st and 3rd quartiles, and the median is represented in the center. Whiskers indicate the lesser of either the maximum/minimum data points or 1.5*IQR from the 1st and 3rd quartiles.

All statistical tests were performed in R. Tukey tests were performed with the HSD.test function from the agricolae (1.3-2) package. Dunnets tests were performed with the glht() function from the multcomp (1.4-12) package (Hothorn et al., 2008). All linear mixed models were performed using lme4 (1.1-24, Bates et al., 2015). *P*-values from linear mixed models were calculated using Satterthwaite approximation implemented by lmerTest (3.1-, Kuznetsova et al., 2017). When many statistical tests were performed on the same data and many *P*-values were obtained we performed Benjamini and Hochbergs *P*-value correction method (FDR) to control for false discovery rate (Benjamini and Hochberg, 1995). An alpha value of 0.05 was used for all statistical tests to assign significance unless otherwise specified. All statistical analysis methods are described in the figure legends.