Convergent gain and loss of genomic islands drives lifestyle changes in plant-associated bacteria

Ryan A. Melnyk[1,2], Sarzana S. Hossain[1,2], and Cara H. Haney[1,2*]

Affiliations: [1]Department of Microbiology and Immunology, The University of British Columbia, Vancouver, Canada; [2]Michael Smith Laboratories, The University of British Columbia, Vancouver, Canada; *Corresponding author – e-mail: cara.haney@msl.ubc.ca

Abstract

Host-associated bacteria can have both beneficial and detrimental effects on host health, but little is known about the evolution of these distinct outcomes. Using the model plant *Arabidopsis,* we found that closely related strains within the *Pseudomonas fluorescens* species complex (Pfl) promote plant growth and occasionally cause disease. To elucidate the genetic basis of the transition between commensalism and pathogenesis, we developed a novel computational pipeline and identified genomic islands that correlate with outcomes for plant health. One island containing lipopeptide biosynthesis and quorum sensing genes is required for pathogenesis and allows distantly related pathogens to cooperate in the environment. We found that genomic loci associated with both pathogenic and commensal lifestyles were convergently gained and lost in multiple lineages through homologous recombination, constituting an early step in the differentiation of pathogenic and beneficial lifestyles and providing insights into the evolution of host-associated bacteria.

Introduction

Host-adapted bacterial lifestyles range from mutualistic to commensal to pathogenic resulting in positive, neutral, or negative effects on host fitness, respectively (*1*). Many of these intimate associations are the product of millions of years of co-evolution, resulting in a complex molecular dialogue between host and bacteria (*2*, *3*). In contrast, horizontal gene transfer (HGT) can lead to rapid lifestyle transitions in host-associated bacteria through the gain and loss of virulence genes (*4*–*6*). For example, the acquisition and loss of pathogenicity islands plays a key role in the emergence of enteropathogenic *E. coli* strains from commensal lineages and vice versa (*5*). Similarly, a virulence plasmid transforms beneficial plant-associated *Rhodococcus* strains into pathogens, while strains without the plasmid revert to commensalism (*4*). In cases where bacterial lifestyle transitions are mediated by a recently acquired plasmid, there is little effect on the evolution of the core bacterial genome allowing for rapid reversibility between pathogenic and commensal lifestyles (*4*, *5*). It is unclear if reversibility of lifestyles is common in other bacteria, or if acquisition of pathogenicity genes drives loss of genomic features associated with commensalism (or vice versa).

To further examine how recent lifestyle transitions in plant-associated bacteria impact genome evolution, we focused on the *Pseudomonas fluorescens* (Pfl) species complex which contains both beneficial strains and pathogens (*7*–*13*). Pfl strains are enriched in close proximity to plant roots (the "rhizosphere") relative to surrounding soil in diverse plants including the model plant *Arabidopsis thaliana* (*9*, *14*, *15*). Single Pfl strains have been shown to have beneficial effects on *Arabidopsis* health by promoting lateral root formation, protecting against pathogens, and modulating plant immunity (*9*, *16*), while others cause diseases such as tomato pith necrosis (*17*) and rice sheath rot (*18*). Thus, we used the Pfl species complex in association with *Arabidopsis* to understand how strains shift along the symbiosis spectrum from pathogenic to beneficial and how that might influence genome evolution.

Here we show that within a single *Pseudomonas* operational taxonomic unit (OTU; >97% identity by 16S rRNA), gain and loss of multiple genomic islands through homologous recombination can drive the transition from pathogenesis to commensalism. Moreover, we found that island-mediated lifestyle transitions affect loci throughout the genome. Using a novel high-throughput comparative genomics pipeline followed by reverse genetics, we found two unique sets of genomic features associated with predicted pathogenic and beneficial strains. Evolutionary reconstruction indicated that gain and loss of these genomic features occurred multiple times, and that these changes were mediated by homologous recombination. Collectively this work implicates interactions between homologous recombination and horizontal gene transfer as primary drivers of bacterial adaptation associated with lifestyle transitions in the rhizosphere.

Results

A pathogen within a plant growth-promoting clade

To understand the emergence and phylogenetic distribution of plant-associated lifestyles, we focused on a well-characterized beneficial strain and asked whether its closest cultured relatives also were beneficial. *Pseudomonas* sp. WCS365 robustly colonizes plant roots (*19*, *20*), promotes growth (*9*), and protects plants from soil-borne fungal pathogens (*21*). Close relatives of WCS365 include an isolate from *Arabidopsis* (*Pseudomonas brassicacearum* NFM421) (*22*), and isolates from a nitrate-reducing enrichment of groundwater (N2E2 and N2C3) (*23*, *24*) (Figure 1A). Together, these 4 strains share nearly identical 16S rRNA sequences (>99.4% identity) and would be grouped into a single OTU in a community profile; however, it is unknown whether all members of this OTU share the beneficial antifungal and plant growth promotion abilities of WCS365.

We tested whether these 4 closely related strains isolated from different environments could promote plant growth. We found that in a gnotobiotic seedling assay where WCS365, NFM421 and N2E2 increased lateral root density (Figure 1B-C), N2C3 caused significant stunting of root and rosette development (Figure 1B) and a significant reduction in fresh weight relative to mock-inoculated seedlings (Figure 1C). Additionally, we found that N2C3 killed or stunted plants from the families *Brassicaceae* (kale, broccoli, and radish) and *Papaveroideae* (poppy), but has little to no effect on the *Solanaceae* (tomato and *Nicotiana benthamiana*) (Figure S1). Thus, unlike its close relatives that promote plant growth, N2C3 is a broad host range pathogen.

A pathogenicity island turns *P. fluorescens* into a pathogen

We reasoned that by comparing the genomic content of N2C3 to closely related commensal *P. fluorescens* strains and pathogenic *P. syringae* strains, we could identify the genetic mechanisms underlying pathogenicity or commensalism within this clade. The large number of sequenced genomes within the genus *Pseudomonas* made existing homolog detection methods (which scale exponentially) untenable for surveying the pangenome of the entire genus (*25*). Therefore, we sought to develop a method that coupled fast but robust homolog identification of a reference pangenome with a heuristic approach that generated binary homolog presence-absence data of the genes composing the reference pangenome for an arbitrarily large dataset.

In order to identify the genomic features associated with commensalism and pathogenesis, we built a machine learning-inspired bioinformatics pipeline called PyParanoid to generate the *Pseudomonas* reference pangenome. The first phase of PyParanoid uses conventional similarity clustering methods to identify the pangenome of a training dataset that includes phylogenetically diverse reference genomes and strains of experimental interest. The diversity of the training pangenome is then represented as a finite set of amino acid hidden Markov models (HMMs) which are then used in the second phase to catalog the pangenome content using computational resources that scale linearly (not exponentially) with the size of the dataset. The result of this pipeline is presence-absence data for a genome dataset that is not constrained by sampling density or phylogenetic diversity. This heuristic-driven approach enabled us to rapidly assign presence and absence of 24,066 discrete homology groups to 3,894 diverse genomes from the diverse *Pseudomonas* genus, assigning homology group membership to 94.2% of the 22.6 million protein sequences in our combined database (details in Materials and Methods). To our knowledge, this is the largest and most diverse genomic dataset ever used to generate a homology database and it was accomplished using reasonable computational resources (roughly ~230 core-hours on a single workstation).

Using the *Pseudomonas* reference pangenome, we searched for genes that were unique to the pathogenic N2C3 or its 3 closely related beneficial relatives. We found that N2C3 contains a conspicuously large 143-kb island comprising 2.0% of the N2C3 genome that is not present in the beneficial strains. The predicted functions of the genes are also consistent with a role in pathogenesis; the island features two adjacent large clusters of non-ribosomal peptide synthetase (NRPS) genes, as well as genes homologous to the acyl-homoserine lactone (AHL) quorum sensing system prevalent in the Proteobacteria, which can play a role in virulence (*26*) (Figure 2A). We designated this putative pathogenicity island the LPQ island (**l**ipo**p**eptide/**q**uorum-sensing). These clusters are very similar to genes involved in the production of cyclic lipopeptide pore-forming phytotoxins in *Pseudomonas syringae* spp. (syringopeptin and syringomycin), which have roles in virulence in many pathovars of *P. syringae* (*27–29*). The regions flanking the island are adjacent in other genomes, suggesting a possible insertion or deletion event (Figure 2B).

In order to determine if the LPQ island is necessary for pathogenesis in the Pfl clade, we used reverse genetics to disrupt portions of the LPQ island in N2C3. We deleted gene clusters predicted to encode syringopeptin (ΔSYP - 73 kb), syringomycin (ΔSYR - 39 kb), or both (ΔSYRΔSYP). We found that deletion of either cluster eliminated the N2C3 pathogenesis phenotype in our seedling assay (Figure 2C). This is consistent with observations that syringomycin and syringopeptin contribute to virulence in *P. syringae* B301D (*27*). We also generated knockouts of both the AHL synthase ($\mathrm{LuxI_{LPQ}}$) as well as the AHL-binding transcriptional regulator ($\mathrm{LuxR_{LPQ}}$). Both the $\Delta luxI_{LPQ}$ and $\Delta luxR_{LPQ}$ mutations abrogated the pathogenic phenotype (Figure 2D). These genetic results indicate that both lipopeptide biosynthesis and quorum sensing within the LPQ island are required for the pathogenicity of N2C3.

Because the LPQ island is necessary for pathogenesis in N2C3, we speculated that it may serve as a marker for pathogenic behavior in other *Pseudomonas* strains allowing us to find other genomic features that correlate with pathogenesis or commensalism. We searched the PyParanoid database for other strains with genes contained within the 15 homology groups unique to the lipopeptide island. While many of the lipopeptide biosynthesis-associated genes (10-12 groups) were found in a subset of *P. syringae* strains, the entire set of 15 genes including the quorum sensing system were uniquely found in three other groups that contain bona fide plant pathogens (*P. corrugata, P. mediterranea* and *P. fuscovaginae* sensu lato) within the *P. fluorescens* clade (Figure 2E and Data S1). Genomic and genetic evidence from these three pathogenic species support a role for the LPQ island in pathogenesis in a variety of hosts, suggesting that the mechanism used by N2C3 to kill *Arabidopsis* may be conserved in divergent strains throughout the *P. fluorescens* clade (*7*, *8*, *30–33*). Additionally, it was previously shown that the LPQ island is the source of antifungal cyclic lipopeptides in two other strains (DF41 and in5) (*34*, *35*). Collectively these data indicate that the LPQ island serves as a marker for plant pathogenic behavior, and possibly antifungal activity, in diverse *Pseudomonas* spp.

To determine if the presence of the island predicted pathogenesis, we acquired representative isolates from the three pathogenic groups as well as the antifungal strain *P. brassicacearum* DF41. The pathogenic isolates (*P. mediterranea*, *P. corrugata,* and *P. fuscovaginae*-like) all inhibited *Arabidopsis* to a similar degree as N2C3 (Figures 2E, S2). On the other hand, DF41 did not inhibit *Arabidopsis* growth, suggesting that the cyclic lipopeptides from different strains are regulated differently or have different activities (Figures 2E, S2). This identifies the lipopeptide island as a broadly conserved mechanism of pathogenesis throughout the *P. fluorescens* clade that can serve as a genetic marker for predicted pathogenic (LPQ+) or commensal (LPQ-) lifestyles.

Commensalism and pathogenesis are associated with multiple genomic features

Because N2C3 is closely related to growth-promoting strains, we considered whether presence or absence of the LPQ island alone determines lifestyle or if there are additional genomic loci associated with the transition from pathogenesis to commensalism. To answer this question, we identified a broader monophyletic group of 85 genomes encompassing the *P. brassicacearum* clade, as well as the sister group containing the LPQ+ pathogens *P. corrugata* and *P. mediterranea* (hereafter the "*bcm* clade"). Together the *bcm* clade corresponds to the '*P. corrugata'* subgroup identified in other *Pseudomonas* phylogenomic studies and shares >97% 16S identity despite containing 8 different named species (*25*, *36*). Constraining our analysis to a phylogenetically narrow clade containing both pathogenic and beneficial bacteria allowed us to examine lifestyle transitions over a short evolutionary time.

To test if other elements of the variable genome were associated with the pathogenicity island, we performed a genome-wide association study (GWAS) in order to link the presence and absence of specific genes (based on PyParanoid data) with the predicted pathogenic phenotype (i.e. presence of the LPQ island). We utilized treeWAS, which is designed to account for the strong effect of population structure in bacterial datasets (*37*). Using treeWAS, we identified 41 genes outside of the LPQ island which were significantly ($p < 0.01$) associated with the presence of the island based on three independent statistical tests (Data S2). 407 additional genes passed one or two significance tests, demonstrating that many genetic loci in the *bcm* clade are influenced by the presence of the LPQ island in the genome.

We explored the physical locations and annotations of the loci with significant associations with the LPQ island to identify clusters of genes with cohesive functional roles in plant-microbe interactions. Beyond the LPQ island we found 5 additional genomic loci: two were positively correlated with the LPQ island and 3 were negatively correlated with the LPQ island (Figure 3, Data S2). A subset of the genes significantly associated with the LPQ island were found in two small (<10kb) genetic clusters with unknown functions (putative pathogenicity islets I and II – PPI1 and PPI2) which were correlated with the presence of the LPQ island in validated and predicted pathogenic strains. Many significant genes were correlated with the absence of the LPQ island in validated and predicted commensal strains (Data S2). One genomic locus containing 28 of these genes encodes a type III secretion system (T3SS) and effectors. This T3SS island is part of the Hrp family of T3SSs important for *P. syringae* virulence (*2*) and suppression of pathogen- and effector-triggered immunity by beneficial rhizosphere *bcm* clade strains Q8r1-96 and Pf29Arp (*38*, *39*). Many commensal strains also have a single "orphaned" T3SS effector (T3SE) homologous to the *P. syringae hopAA* gene (*38*, *40*). Commensal strains were also highly likely to contain a gene cluster for biosynthesis of diacetylphloroglucinol (DAPG), a well-studied and potent antifungal compound important for biocontrol of phytopathogens (*41*). All 6 genetic loci (LPQ, PPI1, PPI2, T3SS, DAPG, and *hopAA* – Table S3) were polyphyletic, revealing a complex evolutionary history of lifestyle transitions within the *bcm* clade (Figure 3). Collectively, this indicates that acquisition or loss of a pathogenicity island is associated with reciprocal gain and loss of genes associated with commensalism.

Transitions between pathogenesis and commensalism arise from homologous recombination-driven genomic variation

To further understand the evolutionary history of the *bcm* clade, we searched for artifacts of the horizontal gene transfer (HGT) events that might cause the polyphyletic distribution of the 6 lifestyle-associated loci. For example, we might expect to see evidence of HGT such as genomic islands integrated at multiple distinct genomic locations or islands with a phylogenetic history very distinct from the core genome phylogeny. Additionally, we might find evidence of specific HGT mechanisms such as tRNA insertion sites, transposons, and plasmid- or prophage-associated genes (*42*, *43*). We used the PyParanoid database to examine the flanking regions of each of the five islands and the *hopAA* gene. We detected each locus only in a single genomic context, with flanking regions conserved in all *bcm* genomes (Figures 4A and Figures S3-S8). These loci are not physically linked in any of the *bcm* genomes, suggesting that linkage disequilibrium of these loci is driven by ecological selection ("eco-LD"), not physical genetic linkage (Figure 4B and Figures S3-S8) (*44*). Finally, there were no obvious genomic signatures of transposition, conjugation, transduction, or site-specific integration; all of which are commonly associated with horizontal gene transfer (HGT) of genomic islands (*43*, *45*). Together, the absence of HGT signals and the conservation of the flanking regions signify homologous recombination of flanking regions as the primary mechanism driving gain or loss of the lifestyle-associated loci.

Recombination events between distantly related strains can lead to incongruencies between gene and species phylogenies. To identify recombination events leading to island gain, we built phylogenies of the LPQ and T3SS islands and compared them to the species phylogeny. While the LPQ island phylogeny was largely congruent with the species phylogeny (Figure S9), the T3SS island had several incongruencies with the species tree (Figure S10). This indicates that recombination events leading to gain of the LPQ island were between closely related strains and are phylogenetically indistinguishable from clonal inheritance. In contrast, the history of the T3SS island shows evidence that the island was occasionally acquired from divergent donors.

Since the T3SS island's history included several instances of recombination between distantly related donors and recipients, we reasoned that there might be signatures of such events in regions flanking the island. To test this hypothesis, we built phylogenies of conserved genes flanking the T3SS. For one gene downstream of the T3SS island integration site (annotated as '*trx*-like', due to annotation as a thioredoxin-domain containing protein), we found that the gene tree was largely incongruent with the species tree, indicating horizontal gene transfer was prevalent in the history of the *trx*-like gene despite its conservation in all extant members of the *bcm* clade (Figure S11). By integrating the T3SS presence-absence data with the *trx*-like phylogeny and the species tree, we developed a model based on phylogenetic evidence that explains the origins of the T3SS island in extant *bcm* strains (Figures 4C, 4D and S11). Our model implicates homologous recombination between regions flanking genomic islands as the means of gain and loss of lifestyle-associated loci (Figure 4D). This provides an evolutionary mechanism underpinning the polyphyletic distribution of commensal and pathogenic islands and behavior within closely related strains of *P. fluorescens.*

Quorum interactions drive lipopeptide production and cooperative pathogenesis

Our phylogenomic analysis suggests that emergence of a pathogenic strain from a beneficial lineage could be triggered by the gain of the LPQ island followed by gain or loss of additional loci through homologous recombination. These additional events must occur before the nascent pathogen loses the LPQ island through recombination with closely related members of the original beneficial lineage, with whom homologous recombination is more efficient. Thus, lifestyle transitions in the *bcm* clade might be facilitated through an ecological mechanism that promotes gene flow from the more distantly related pathogenic donor lineage.

We hypothesized that quorum sensing could serve as such an ecological mechanism. Specifically, the *luxI*$_{LPQ}$/*luxR*$_{LPQ}$ quorum sensing mechanism in the LPQ island could act to promote gene flow by allowing LPQ+ strains to cooperate in the environment. If this cooperation proves mutually beneficial, this would place LPQ+ strains in increased proximity and promote genetic exchange at other genomic loci. This would be an example of a social behavior recognizing and cooperating with other strains that have the same allele, but not necessarily close relatives missing the allele (*46*, *47*). The term "kind selection" has been coined to describe this behavior and has been observed in multiple bacterial systems (*48–50*), but to our knowledge have not been observed in the context of AHL quorum sensing in a natural system (*51*). Such a mechanism could allow distantly related Pfl strains with the LPQ island to coordinate lipopeptide production and rhizosphere pathogenesis, thus increasing gene flow between LPQ+ strains.

If the *luxI*$_{LPQ}$/*luxR*$_{LPQ}$ system allows cooperation among distantly related LPQ+ strains, we would expect the system to be phylogenetically distinct from other AHL synthases and specifically associated with lipopeptide-producing strains within *Pseudomonas*. We found that LuxI$_{LPQ}$ represented a monophyletic clade of *Pseudomonas* LuxI sequences as delineated using our *Pseudomonas* reference pangenome (Figure 5A). Furthermore, the presence of LuxI$_{LPQ}$ had a positive correlation with all of the 14 other lipopeptide genes across the entire *Pseudomonas* clade (Figure 5B). While there are many lipopeptide-producing strains that lack LuxI$_{LPQ}$ (mostly *P. syringae*), every strain that has LuxI$_{LPQ}$ also has the entire LPQ island (Data S1). These *in silico* results conclude that LuxI$_{LPQ}$ is specifically associated with cyclic lipopeptide-producing *Pseudomonas* spp. across the entire genus.

To test if the LuxI$_{LPQ}$ homologs share the same signaling molecule, we co-inoculated *Arabidopsis* seedlings with DF41 (a non-pathogenic LPQ+ strain) and N2C3 Δ*luxI*$_{LPQ}$ and Δ*luxR*$_{LPQ}$ mutants, deficient in production of the AHL signal and signal perception, respectively. We found that DF41 restored pathogenicity of the non-pathogenic Δ*luxI*$_{LPQ}$ AHL synthase mutant, indicating that it can provide an activating AHL signal *in trans*. However, DF41 did not restore pathogenicity of the Δ*luxR*$_{LPQ}$ regulatory mutant (Figure 5C). Additionally, using an AHL biosensor that produces the purple pigment violacein in response to short-chain AHL molecules, we found that almost all of the strains containing the LPQ island elicited pigment production (5 out of 7), while none of the strains without the island (0 out of 20) or the N2C3 Δ*luxI*$_{LPQ}$ strain resulted in pigment production (Figure 5D, 5E) (*52*). Reports from *P. corrugata, P. mediterranea* and *P. brassicacearum* DF41 specifically implicate a C6-AHL molecule as the lipopeptide-associated signal (*32*, *33*, *53*), which is a strong inducer of the violacein-producing biosensor. Thus, the LPQ island has the capability to allow distantly related *Pseudomonas* spp. to coordinate lipopeptide production through community C6-AHL levels with other LPQ+ strains, providing a potential ecological mechanism for the gene flow patterns observed in the *bcm* clade.

Discussion

Here we provide evidence for a novel evolutionary mechanism that drives the transition between commensal and pathogenic lifestyles in plant-associated *Pseudomonas*. We have discovered that recombination mediates large differences in gene content that determine how *Pseudomonas* interacts with a plant host. Similar mechanisms of gene content variation through homologous recombination were reported to play a role in the propagation of a pathogenicity island within the *E. coli* species (*54*) and variation in siderophore production in a natural *Vibrio* population (*55*). While other studies have shown transitions from pathogenic to beneficial lifestyles in plant-associated microbes, these have been driven by plasmid transfer or experimental evolution (*4*, *56*). To our knowledge, this is the first direct description of homologous recombination in a population of closely related plant-associated strains leading to gain and loss of large pathogenic and beneficial genomic islands. Interestingly, the same island-based adaptations appear in multiple independent lineages, providing a compelling example of convergent evolution of pathogenic and beneficial lifestyles through gene gain and loss.

This study highlights the complexity inherent in studying the rhizosphere microbiome, particular when trying to link particular 16S sequences with functions in single strains. We found that labels like "beneficial" and "pathogen" break down over short evolutionary distances within a well-studied clade of *Pseudomonas* spp. Moreover, we found that one strain (DF41) may function as a beneficial strain in isolation but could potentially exacerbate the effects of bad actors through inter-strain quorum sensing. DF41 may also function as a genetic reservoir of deleterious alleles (e.g. the LPQ island) necessary for transitions to pathogenesis in the *bcm* clade. These strains are thus best designated as "tritagonists": a term that has been recently proposed to describe commensal community members that indirectly influence a host through modulating the activity of another organism (*57*).

The ecologically-driven linkage disequilibrium ("eco-LD") of the beneficial and pathogenic loci implies selection for one lifestyle or another. However, it is unclear how exactly microbe-mediated effects on the host translate to microbial fitness in the rhizosphere. For example, do pathogenic strains outcompete beneficial strains in a diseased plant? Furthermore, do recently diverging clades of pathogenic or beneficial *bcm* strains even inhabit the same ecological niche? Our work implicates gain of the LPQ island as a "niche-defining" evolutionary event that separates an incipient pathogen from its beneficial predecessors, leading to further divergence (*58*). Since the *bcm* clade contains pathogen to beneficial transitions as well, the T3SS island likely has a similar niche-defining role, possibly manipulating immune responses of the host plant to favor other T3SS+ strains. Our work elucidates the evolution of loci determining host-associated phenotypes, providing insight into specific mechanisms underlying early steps of differentiation between pathogenic and beneficial lifestyles. More generally, this study extends models of niche-driven speciation to host-associated bacteria, identifying effects on host health as a factor driving evolutionary divergence.

Acknowledgements

We thank Dr. Clay Fuqua for providing the CV026 biosensor strain, Dr. Teresa de Kievit and Dr. Ricardo Oliva for providing *Pseudomonas* isolates, and Dr. Adam Steinbrenner and Dr. Justin Meyer for critical reading of the manuscript. **Funding:** R.A.M. is a Simons Foundation Fellow of the Life Sciences Research Foundation. This work was also supported by an NSERC Discovery Grant (NSERC-RGPIN-2016-04121), Canada Foundation for Innovation, and Canada Research Chair grants awarded to C.H.H. The computational research was carried out with support provided by WestGrid and Compute Canada. **Authors' contributions:** R.A.M. and C.H.H. designed research and discussed results. R.A.M and S.S.H. performed experiments. R.A.M. wrote code and performed all bioinformatics analyses. R.A.M wrote the manuscript with input from C.H.H. **Competing interests:** The authors have no competing interests to declare. **Data and materials availability:** All data is available in the manuscript or the Supplementary Materials. Source code for PyParanoid is available at https://github.com/ryanmelnyk/PyParanoid.

Supplementary Materials

Materials and Methods
Figures S1 to S13
Tables S1 and S2
Data S1-S3

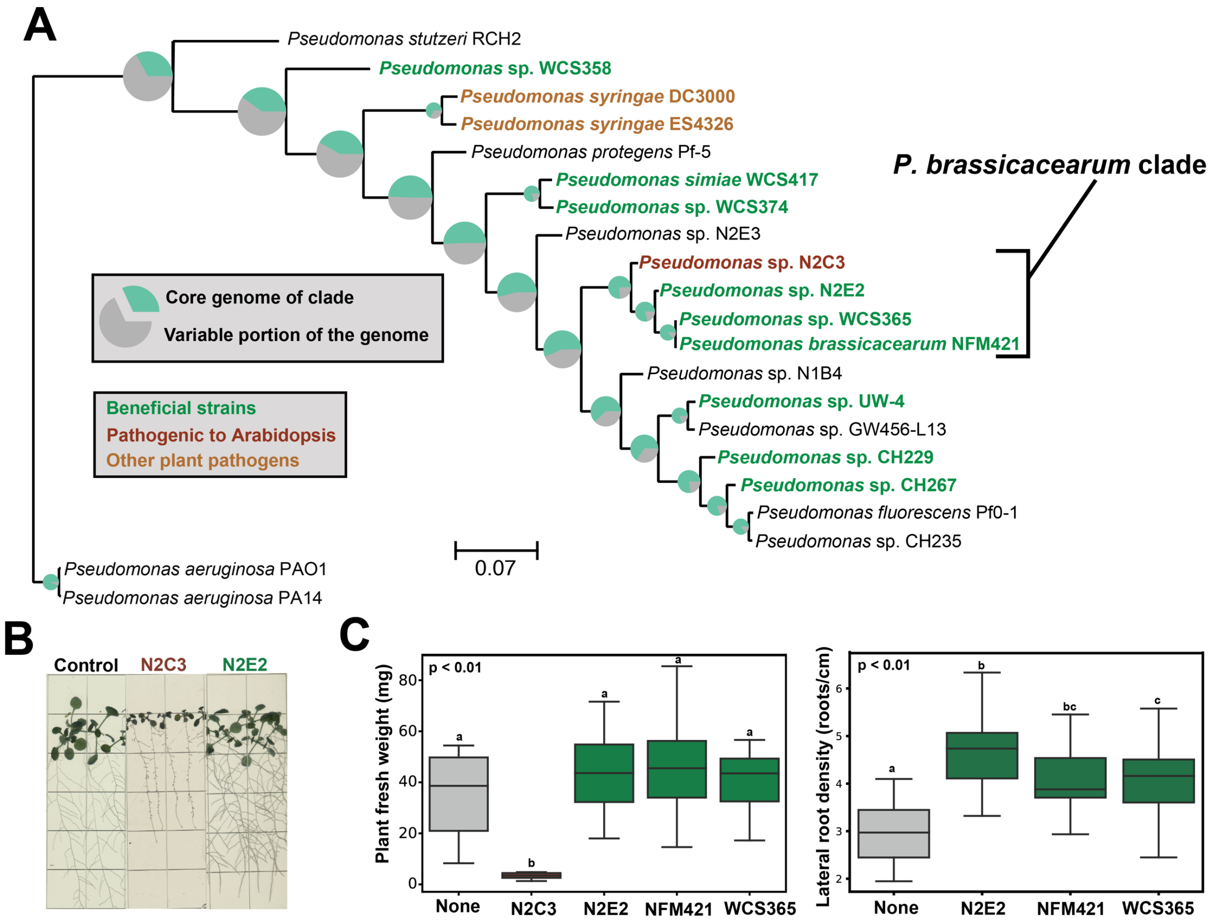

***Figure 1.* A pathogen within a clade of beneficial *Pseudomonas* spp.** (**A**) A species tree of characterized *Pseudomonas* strains was generated using 217 conserved housekeeping genes. Strain names are color-coded to reflect plant-associated lifestyles based on prior reports in the literature (*2*, *9*, *59*). Pie charts at each node reflect the proportion of the genome that encodes "core" genes that are present in each member of the clade, showing that there can be large differences in gene content over short evolutionary distances. (**B**) Gnotobiotic *Arabidopsis* seedlings inoculated with N2C3, N2E2, or a buffer control. (**C**) Fresh weight and lateral root density of seedlings. Lower-case letters indicate statistically significantly (p < 0.01) means as determined by an unpaired Student's T-test.

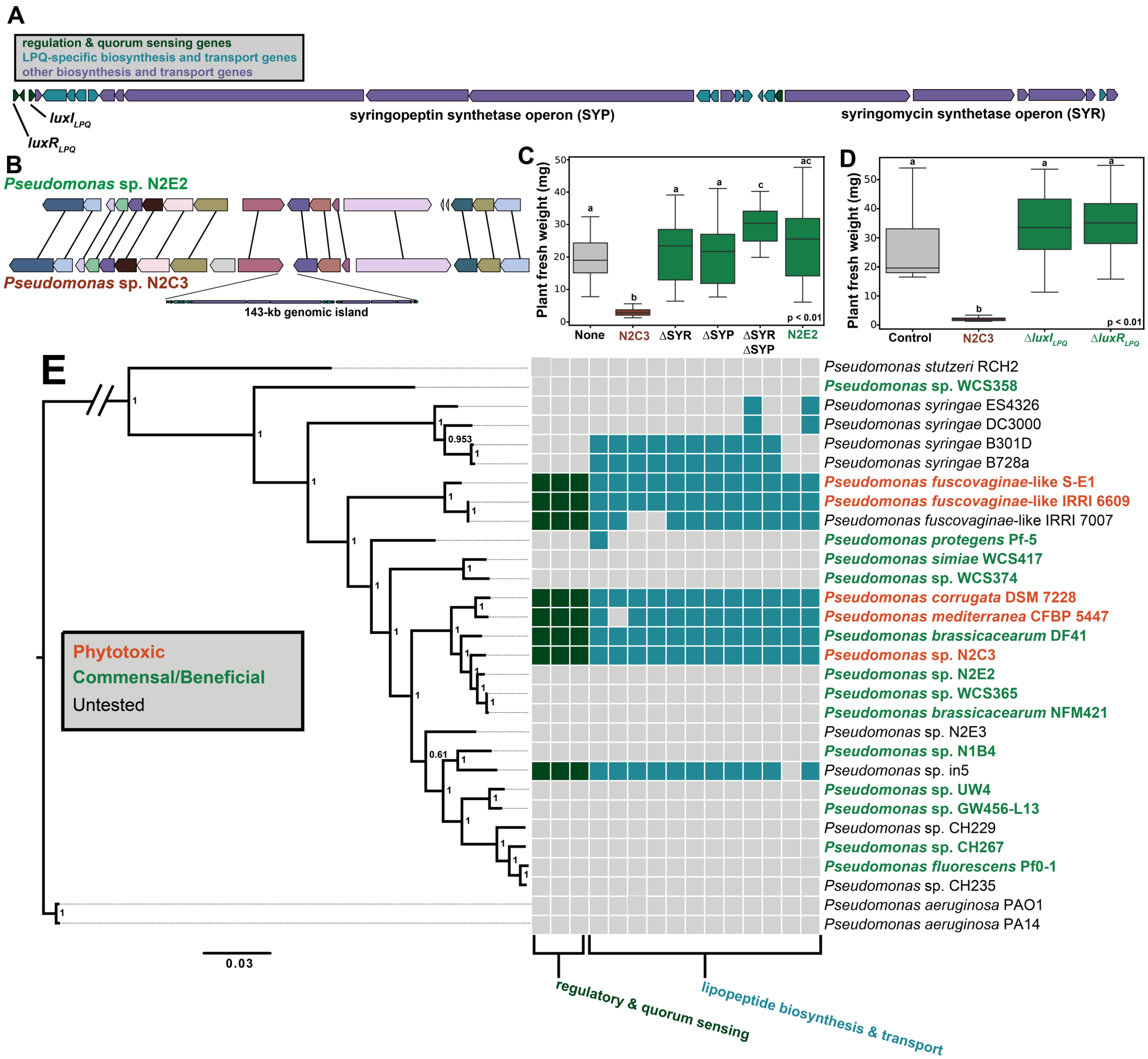

Figure 2. **A genomic island necessary for pathogenesis.** (**A**) A large genomic island encoding lipopeptide biosynthesis and quorum sensing genes (the LPQ island) is present in N2C3 but not in beneficial strains such as N2E2. (**B**) Diagram of the genomic island showing putative functions and locations of quorum sensing genes (*luxI*$_{LPQ}$ – AHL synthase, *luxR*$_{LPQ}$ – AHL-binding transcriptional regulator) and the lipopeptide biosynthesis clusters (syringomycin and syringopeptin). (**C, D**) Fresh weight of *Arabidopsis* treated with wild-type strains (N2C3 and N2E2) as well as N2C3 mutants from the LPQ island (ΔSYR – syringomycin cluster deletion, ΔSYP - syringopeptin cluster deletion, ΔSYRΔSYP – syringomycin and syringopeptin double deletion, Δ*luxI*$_{LPQ}$ – AHL synthase deletion, Δ*luxR*$_{LPQ}$ – AHL-binding transcriptional regulator deletion). Lower-case letters indicate statistically significant ($p < 0.01$) differences as determined by an unpaired Student's T-test. (**E**) Species tree showing the distribution of the island among divergent members of the *Pseudomonas* clade. Color coding of taxon labels summarizes the phenotypic data from Figs. 2A and S2. Colored squares

indicate that an individual gene (as classified by PyParanoid) is present in a given taxa. Only 15 of the 28 genes in **B** (shown in green and teal in D) that are unique to the pathogenic strains are shown in **E**. The other 13 genes (shown in purple in **B**) are part of larger homolog groups that are not specific to the LPQ island.

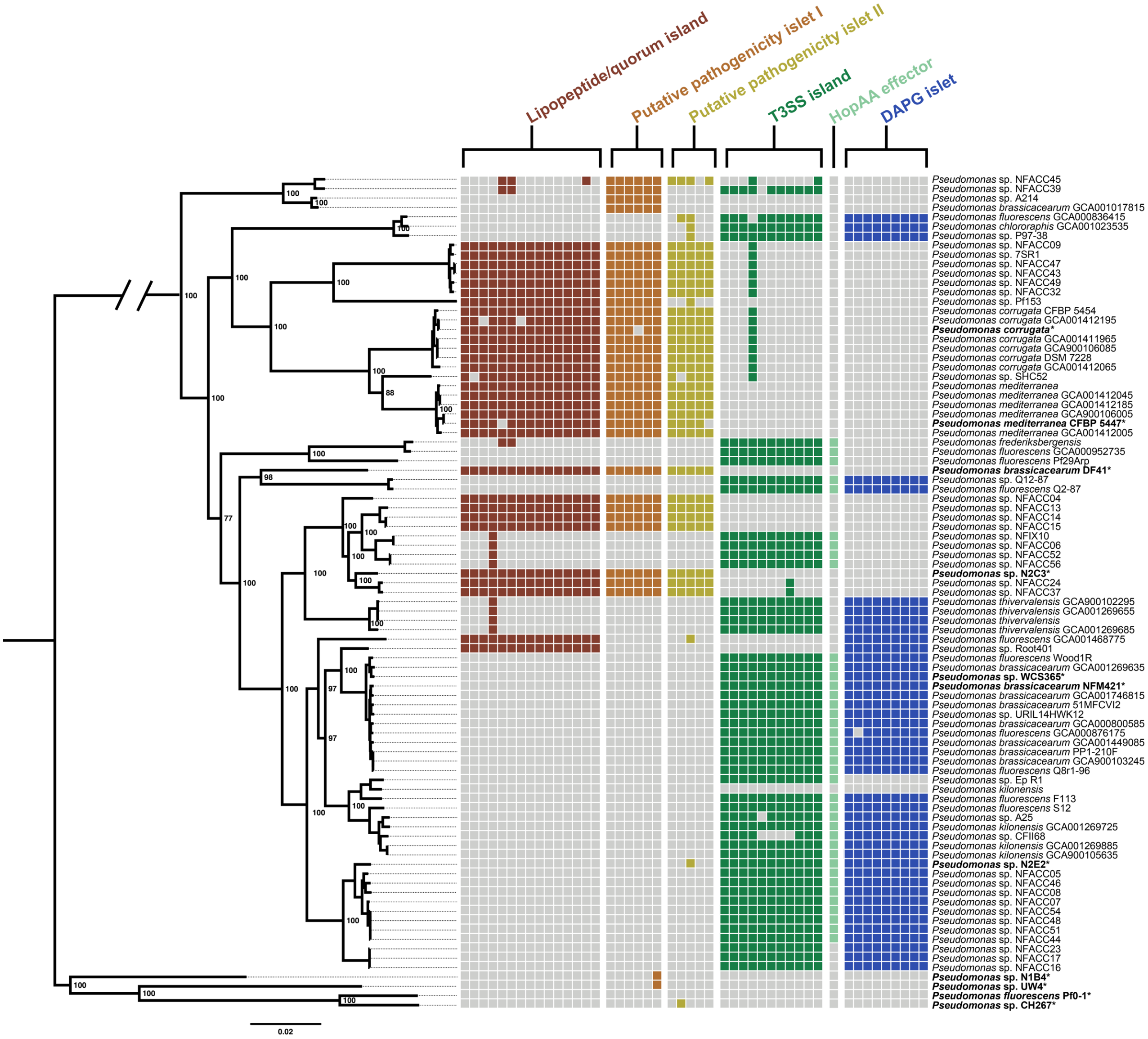

Figure 3. **Polyphyletic distribution of pathogenic and beneficial genomic islands within the *bcm* clade.** A species tree was built for the *bcm* clade using a concatenated alignment of 2,030 conserved genes. Color-coded squares represent presence and absence of individual genes associated with each locus based on PyParanoid presence-absence data. In the case of both the LPQ and T3SS, only a subset of genes to the island are shown. This is because some genes in both islands are part of larger homolog groups not specific to each island. In the case of the LPQ, these are the non-ribosomal peptide synthetase genes and in the case of the T3SS it is the structural components of the secretion apparatus. Bold taxon names highlight strains experimentally tested for beneficial/commensal or pathogenic behavior. All six loci exhibit a highly polyphyletic distribution in extant strains suggesting that the pathogenic and commensal/beneficial lifestyles have a complex evolutionary history in this clade.

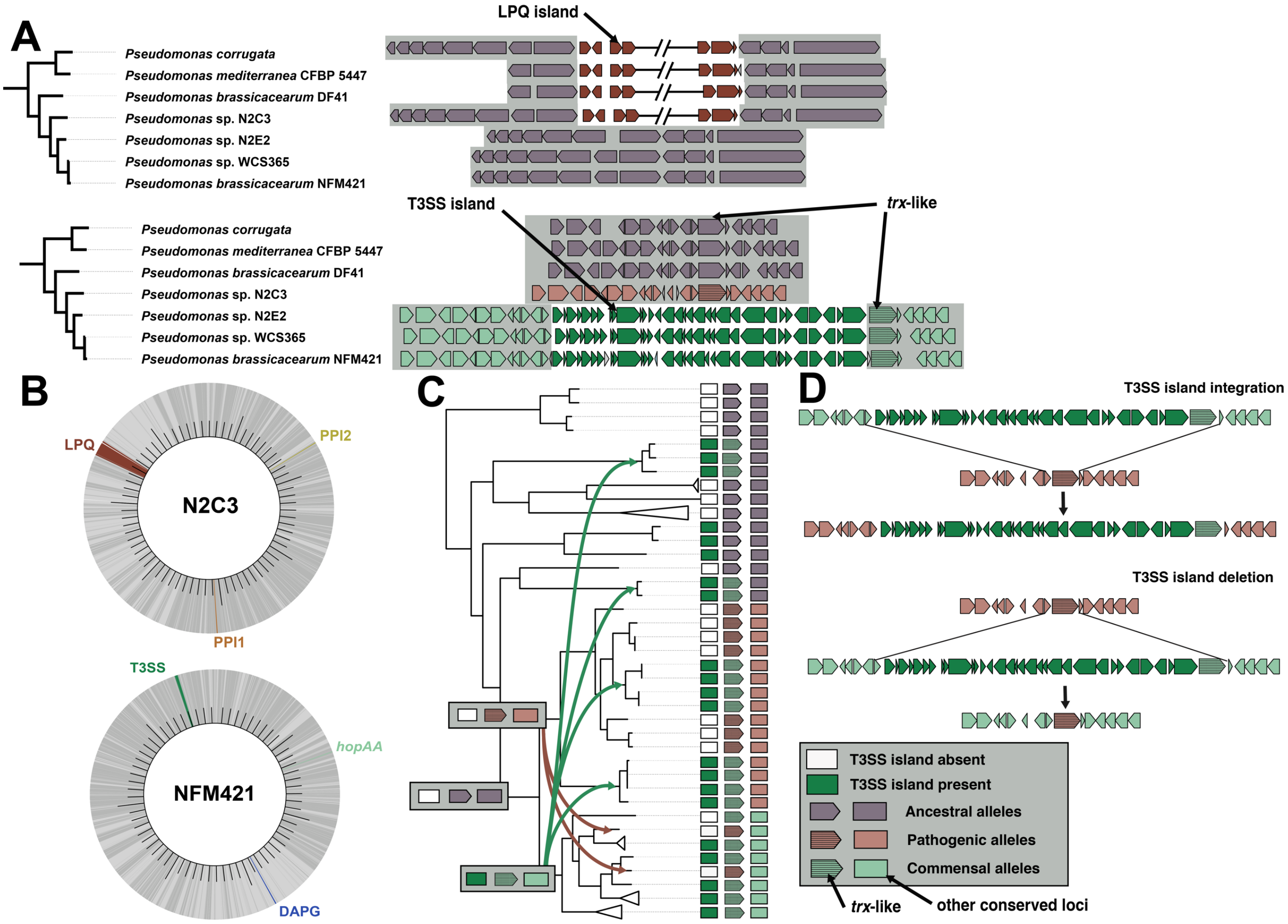

Figure 4. **Phylogenomic evidence for independent island gain and loss through homologous recombination of flanking regions.** (**A**) Insertion sites for each locus are conserved throughout the *bcm* species complex. Here only the LPQ and T3SS island insertion sites are shown. The tree shows the evolutionary relationship of the *bcm* strains. The grey background shows the conserved regions that flank both islands are adjacent in strains lacking the island. Additionally, the location of the *trx*-like gene is shown adjacent to the T3SS which is used as a marker for recombination. (**B**) Circular genome plots of two representative complete genomes from the pathogenic strain N2C3 and the beneficial strain NFM421. Darker grey regions of the genome represent the core genome of the *bcm* clade, while light grey denotes the variable genome showing that these six loci make up only a small amount of the variable loci in any given strain. (**C**) A species phylogeny for the *bcm* clade adapted from Figure 3 to show evidence of homologous recombination of the T3SS island inferred through phylogenetic incongruencies between the species tree and the phylogeny of the *trx*-like gene (see Figure S11). Dark green filled-in rectangles denote taxa that have the T3SS island, while white empty squares are missing the island. Based on the phylogenetic evidence, we propose a parsimonious model for evolution of the T3SS island where gain of the T3SS island led to divergence of an ancestral lineage (purple arrow/rectangle) into a beneficial/commensal lineage (light green arrow/rectangle) and a pathogenic lineage (light red

arrow/rectangle). However, subsequent gain (green curved arrows) or loss (red curved arrows) of the T3SS island has led to phylogenetic incongruence between the adjacent *trx*-like gene and the conserved loci used to construct the species tree (i.e. the *bcm* clade clonal phylogeny). (**D**) A diagram of the proposed homologous recombination mechanism leading to island gain (top) and loss (bottom). Both events lead to insertion of a *trx*-like allele with a distinct history from the surrounding conserved chromosomal loci.

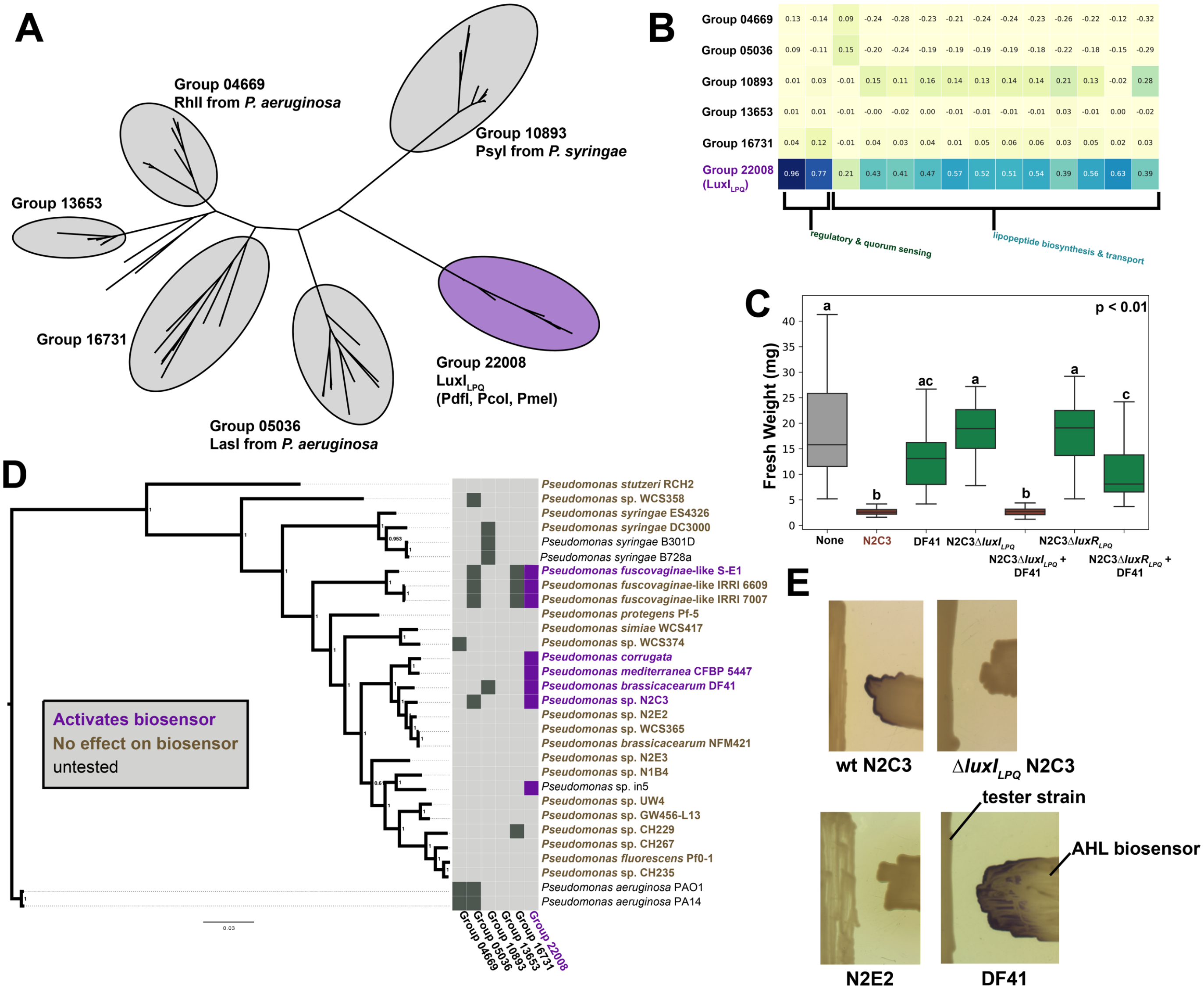

***Figure 5.* A lipopeptide-associated quorum sensing in divergent *Pseudomonas* spp.** (**A**) Gene tree showing the phylogenetic relationship of the 6 AHL synthase homology groups identified by PyParanoid. Light gray squares indicate the absence of a gene and colored squares (dark gray or purple) show the presence of a gene. Group 22008 contained the LuxI$_{\mathrm{LPQ}}$ protein sequence from N2C3 as well as protein sequences from other LPQ+ strains (PdfI, PcoI, PmeI) (*32*, *33*, *53*). (**B**) Correlation coefficients between each AHL synthase homology group and the 14 other LPQ island homology groups across the entire *Pseudomonas* clade, showing that the presence of LuxI$_{\mathrm{LPQ}}$ is correlated with the capacity for lipopeptide biosynthesis across the entire *Pseudomonas* genus. LuxI$_{\mathrm{LPQ}}$ may serve as a common signal associated with lipopeptide production across diverse strains. (**C**) Fresh weight of seedlings inoculated with wild-type strains (N2C3 and DF41), N2C3 quorum-sensing mutants (N2C3Δ*luxI*$_{LPQ}$ – AHL synthase deletion, N2C3Δ*luxR*$_{LPQ}$ – AHL-binding transcriptional regulator deletion), or N2C3 quorum-sensing mutants in conjunction with wild-type DF41. (**D**) Species tree based on 217 housekeeping genes showing the distribution of 6 AHL

synthase homology groups identified using PyParanoid. Bold taxon names in brown had no effect on the CV026 biosensor while bold taxon names in purple activated the biosensor, showing that the presence of PyParanoid Group 22008 corresponds with CV026 activation. (**E**) Streak test examples showing differential activation of pigment production in *Chromobacterium violaceum* CV026, which is an AHL biosensor that responds to short-chain AHL molecules ranging from C4-AHL to C8-AHL, with strongest activation by C6-AHL.

References

1. A. Hirsch, Plant-microbe symbioses: a continuum from commensalism to parasitism. *Symbiosis*. **37**, 345–363 (2004).
2. X.-F. Xin, B. Kvitko, S. Y. He, Pseudomonas syringae: what it takes to be a pathogen. *Nat. Rev. Microbiol.* **16**, 316 (2018).
3. K. M. Jones, H. Kobayashi, B. W. Davies, M. E. Taga, G. C. Walker, How rhizobial symbionts invade plants: The Sinorhizobium - Medicago model. *Nat. Rev. Microbiol.* **5**, 619–633 (2007).
4. E. A. Savory *et al.*, Evolutionary transitions between beneficial and phytopathogenic Rhodococcus challenge disease management. *Elife*. **6**, 1–28 (2017).
5. O. Tenaillon, D. Skurnik, B. Picard, E. Denamur, The population genetics of commensal Escherichia coli. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
6. D. Lin, B. Koskella, Friend and foe: Factors influencing the movement of the bacterium Helicobacter pylori along the parasitism-mutualism continuum. *Evol. Appl.* **8**, 9–22 (2015).
7. I. L. Quibod *et al.*, Rice-infecting pseudomonas genomes are highly accessorized and harbor multiple putative virulence mechanisms to Cause Sheath Brown Rot. *PLoS One*. **10**, 1–25 (2015).
8. E. A. Trantas *et al.*, Comparative genomic analysis of multiple strains of two unusual plant pathogens: Pseudomonas corrugata and Pseudomonas mediterranea. *Front. Microbiol.* **6**, 1–19 (2015).
9. C. H. Haney, B. S. Samuel, J. Bush, F. M. Ausubel, Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nat. Plants*. **1**, 15051 (2015).
10. N. Fromin, W. Achouak, J. M. Thiéry, T. Heulin, The genotypic diversity of Pseudomonas brassicacearum populations isolated from roots of Arabidopsis thaliana: Influence of plant genotype. *FEMS Microbiol. Ecol.* **37**, 21–29 (2001).
11. R. L. Berendsen *et al.*, Unearthing the genomes of plant-beneficial Pseudomonas model strains WCS358, WCS374 and WCS417. *BMC Genomics*. **16**, 1–23 (2015).
12. J. Sikorski, H. Jahr, W. Wackernagel, The structure of a local population of phytopathogenic Pseudomonas brassicacearum from agricultural soil indicates development under purifying selection pressure. *Environ. Microbiol.* **3**, 176–186 (2001).
13. A. Belimov, I. Dodd, V. Safronova, N. Hontzeas, W. Davies, Pseudomonas brassicacearum strain Am3 containing 1-aminocyclopropane-1-carboxylate deaminase can show both pathogenic and growth-promoting properties in its interaction with tomato. *J. Exp. Bot.* **58**, 1485–1495 (2007).
14. D. Bulgarelli *et al.*, Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature*. **488**, 91–95 (2012).
15. D. S. Lundberg *et al.*, Defining the core Arabidopsis thaliana root microbiome. *Nature*. **488**, 86–90 (2012).
16. C. H. Haney *et al.*, Rhizosphere-associated Pseudomonas induce systemic resistance to herbivores at the cost of susceptibility to bacterial pathogens. *Mol. Ecol.* **27**, 1833–1847 (2018).
17. C. M. Scarlett, J. T. Fletcher, P. Roberts, R. A. Lelliott, Tomato pith necrosis caused by Pseudomonas corrugata n. sp. *Ann. Appl. Biol.* **88**, 105–114 (1978).
18. J. Kim, O. Choi, W.-I. Kim, First Report of Sheath Brown Rot of Rice Caused by *Pseudomonas fuscovaginae* in Korea. *Plant Dis.* **99**, 1033–1033 (2015).
19. F. P. Geels, B. Schippers, Selection of Antagonistic Fluorescent Pseudomonas spp. and their Root Colonization and Persistence following Treatment of Seed Potatoes. *J. Phytopathol.* **108**, 193–206 (1983).
20. S. De Weert, L. C. Dekkers, I. Kuiper, G. V. Bloemberg, B. J. J. Lugtenberg, Generation of Enhanced Competitive Root-Tip-Colonizing Pseudomonas Bacteria through Accelerated Evolution. *J. Bacteriol.* **186**, 3153–3159 (2004).
21. F. Kamilova, G. Lamers, B. Lugtenberg, Biocontrol strain Pseudomonas fluorescens WCS365 inhibits germination of Fusarium oxysporum spores in tomato root exudate as well as subsequent formation of new spores. *Environ. Microbiol.* **10**, 2455–2461 (2008).
22. W. Achouak *et al.*, Pseudomonas brassicacearum sp. nov. and Pseudomonas thivervalensis sp. nov., two root-associated bacteria isolated from Brassica napus and Arabidopsis thaliana. *Int. J. Syst. Evol. Microbiol.* **50**, 9–18 (2000).

23. M. P. Thorgersen *et al.*, Molybdenum Availability Is Key to Nitrate Removal in Contaminated Groundwater Environments. *Appl. Environ. Microbiol.* **81**, 4976–83 (2015).
24. M. N. Price *et al.*, Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*. **557**, 503–509 (2018).
25. C. Hesse *et al.*, Genome-based evolutionary history of *Pseudomonas* spp. *Environ. Microbiol.* (2018), doi:10.1111/1462-2920.14130.
26. K. Papenfort, B. L. Bassler, Quorum sensing signal-response systems in Gram-negative bacteria. *Nat. Rev. Microbiol.* **14**, 576–588 (2016).
27. B. K. Scholz-Schroeder, M. L. Hutchison, I. Grgurina, D. C. Gross, The contribution of syringopeptin and syringomycin to virulence of Pseudomonas syringae pv. syringae strain B301D on the basis of sypA and syrB1 biosynthesis mutant analysis. *Mol. Plant-Microbe Interact.* **14**, 336–348 (2001).
28. M. L. Hutchison, D. C. Gross, Lipopeptide phytotoxins produced by Pseudomonas syringae pv. syringae: comparison of the biosurfactant and ion channel-forming activities of syringopeptin and syringomycin. *Mol. Plant. Microbe. Interact.* **10**, 347–354 (1997).
29. D. A. Baltrus *et al.*, Dynamic Evolution of Pathogenicity Revealed by Sequencing and Comparative Genomics of 19 Pseudomonas syringae Isolates. *PLoS Pathog.* **7**, e1002132 (2011).
30. C. P. Strano *et al.*, Pseudomonas corrugata crpCDE is part of the cyclic lipopeptide corpeptin biosynthetic gene cluster and is involved in bacterial virulence in tomato and in hypersensitive response in Nicotiana benthamiana. *Mol. Plant Pathol.* **16**, 495–506 (2015).
31. H. K. Patel *et al.*, Identification of virulence associated loci in the emerging broad host range plant pathogen Pseudomonas fuscovaginae. *BMC Microbiol.* **14**, 1–13 (2014).
32. G. Licciardello *et al.*, N-acyl-homoserine-lactone quorum sensing in tomato phytopathogenic Pseudomonas spp. is involved in the regulation of lipodepsipeptide production. *J. Biotechnol.* **159**, 274–282 (2012).
33. G. Licciardello *et al.*, Pseudomonas corrugata contains a conserved N-acyl homoserine lactone quorum sensing system; its role in tomato pathogenicity and tobacco hypersensitivity response. *FEMS Microbiol. Ecol.* **61**, 222–234 (2007).
34. C. Berry, W. G. D. Fernando, P. C. Loewen, T. R. de Kievit, Lipopeptides are essential for Pseudomonas sp. DF41 biocontrol of Sclerotinia sclerotiorum. *Biol. Control.* **55**, 211–218 (2010).
35. C. F. Michelsen *et al.*, Nonribosomal Peptides, Key Biocontrol Components for Pseudomonas fluorescens In5, Isolated from a Greenlandic Suppressive Soil. **6**, 1–9 (2015).
36. D. Garrido-Sanz *et al.*, Genomic and genetic diversity within the Pseudomonas fluoresces complex. *PLoS One*. **11** (2016), doi:10.1371/journal.pone.0150183.
37. C. Collins, X. Didelot, A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Comput. Biol.* **14**, 1–21 (2018).
38. D. V. Mavrodi *et al.*, Structural and functional analysis of the type III secretion system from Pseudomonas fluorescens Q8r1-96. *J. Bacteriol.* **193**, 177–189 (2011).
39. M. Marchi *et al.*, Genomic analysis of the biocontrol strain *Pseudomonas fluorescens* Pf29Arp with evidence of T3SS and T6SS gene expression on plant roots. *Environ. Microbiol. Rep.* **5**, 393–403 (2013).
40. K. R. Munkvold, A. B. Russell, B. H. Kvitko, A. Collmer, Pseudomonas syringae pv. tomato DC3000 type III effector HopAA1-1 functions redundantly with chlorosis-promoting factor PSPTO4723 to produce bacterial speck lesions in host tomato. *Mol. Plant. Microbe. Interact.* **22**, 1341–1355 (2009).
41. D. M. Weller *et al.*, Role of 2,4-diacetylphloroglucinol-producing fluorescent Pseudomonas spp. in the defense of plant roots. *Plant Biol.* **9**, 4–20 (2007).
42. I. C. Clark, R. A. Melnyk, A. Engelbrektson, J. D. Coates, Structure and evolution of chlorate reduction composite transposons. *MBio*. **4** (2013), doi:10.1128/mBio.00379-13.
43. R. A. Melnyk, J. D. Coates, The Perchlorate Reduction Genomic Island: Mechanisms and Pathways of Evolution by Horizontal Gene Transfer. *BMC Genomics*. **16** (2015), doi:10.1186/s12864-015-2011-5.
44. Y. Cui *et al.*, Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen vibrio parahaemolyticus. *Mol. Biol. Evol.* **32**, 1396–1410 (2015).

45. M. Juhas *et al.*, Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**, 376–393 (2009).
46. W. D. Hamilton, The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16 (1964).
47. W. D. Hamilton, The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**, 17–52 (1964).
48. A. Gardner, S. A. West, Greenbeards. *Evolution (N. Y).* **64**, 25–38 (2010).
49. J. E. Strassmann, O. M. Gilbert, D. C. Queller, Kin Discrimination and Cooperation in Microbes. *Annu. Rev. Microbiol.* **65**, 349–367 (2011).
50. D. Wall, Kin Recognition in Bacteria. *Annu. Rev. Microbiol.* **70**, 143–160 (2016).
51. K. P. Rumbaugh *et al.*, Kin selection, quorum sensing and virulence in pathogenic bacteria. *Proc. R. Soc. B Biol. Sci.* **279**, 3584–3588 (2012).
52. A. Latifi *et al.*, Multiple homologues of LuxR and LuxI control expression of virulence determinants and secondary metabolites through quorum sensing in Pseudomonas aeruginosa PAO1. *Mol. Microbiol.* **17**, 333–343 (1995).
53. C. L. Berry *et al.*, Characterization of the Pseudomonas sp. DF41 quorum sensing locus and its role in fungal antagonism. *Biol. Control.* **69**, 82–89 (2014).
54. S. Schubert *et al.*, Role of intraspecies recombination in the spread of pathogenicity islands within the Escherichia coli species. *PLoS Pathog.* **5**, 1–10 (2009).
55. O. X. Cordero, L.-A. Ventouras, E. F. DeLong, M. F. Polz, Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc. Natl. Acad. Sci.* **109**, 20059–20064 (2012).
56. M. Marchetti *et al.*, Experimental Evolution of a Plant Pathogen into a Legume Symbiont. *PLoS Biol.* **8**, e1000280 (2010).
57. F. M. Freimoser, C. Pelludat, M. N. P. Remus-Emsermann, Tritagonist as a new term for uncharacterised microorganisms in environmental systems. *ISME J.* **10**, 1–3 (2016).
58. J. Wiedenbeck, F. M. Cohan, Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35**, 957–976 (2011).
59. J. Li, D. H. Ovakim, T. C. Charles, B. R. Glick, An ACC Deaminase Minus Mutant of Enterobacter cloacae UW4No Longer Promotes Root Elongation. *Curr. Microbiol.* **41**, 101–105 (2000).