1 Full Title

2 Repeated gain and loss of a single gene modulates the evolution of vascular pathogen lifestyles

3

4 Short Title

5 Gene gain and loss drive transitions in pathogen lifestyle

6

7 Authors

8 Emile Gluck-Thaler^{1,†}, Aude Cerutti^{2,†}, Alvaro Perez-Quintero^{3,†}, Jules Butchacas^{1,4}, Verónica
9 Roman-Reyna^{1,4}, Vishnu Narayanan Madhavan⁵, Deepak Shantharaj⁶, Marcus V. Merfa⁶, Céline
10 Pesce^{7,8,9}, Alain Jauneau¹⁰, Taca Vancheva^{7,8}, Jillian M. Lang³, Caitilyn Allen¹¹, Valerie Verdier⁷,
11 Lionel Gagnevin⁷, Boris Szurek⁷, Sébastien Cunnac⁷, Gregg Beckham¹², Leonardo de la Fuente⁶,
12 Hitendra Kumar Patel⁵, Ramesh V Sonti⁵, Claude Bragard⁸, Jan E. Leach³, Laurent D. Noël²,
13 Jason C. Slot^{1,4}, Ralf Koebnik^{7,*}, Jonathan M. Jacobs^{1,4*}

14

15 Affiliations

16 ¹Department of Plant Pathology, The Ohio State University, Columbus, OH 43210, USA.

17 ²LIPM, Université de Toulouse, INRA, CNRS, Université Paul Sabatier, Castanet-Tolosan,
18 France.

19 ³Agricultural Biology, Colorado State University, Fort Collins, CO, USA.

20 ⁴Infectious Disease Institute, The Ohio State University, Columbus, OH 43210, USA.

21 ⁵CSIR-Centre for Cellular and Molecular Biology, Hyderabad, 500007, India

22 ⁶Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849, USA.

23 ⁷Institut de Recherche pour le Développement, UMR Interactions Plantes Microorganismes
24 Environnement, Montpellier, France.

25 ⁸Earth & Life Institute, Université Catholique Louvain-la-Neuve, Louvain-la-Neuve, Belgium.

26 ⁹Department of Microbiology, University of New Hampshire, Durham, NH, USA.

27 ¹⁰Institut Fédératif de Recherche 3450, Plateforme Imagerie, Pôle de Biotechnologie Végétale,

28 Castanet-Tolosan, France.

29 ¹¹Department of Plant Pathology, University of Wisconsin—Madison, Madison, WI 53706, USA.

30 ¹²National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401, USA.

31

32 *Correspondence to: Jonathan M. Jacobs, Ralf Koebnik

33

34 †Authors equally contributed to this work.

35

36

37 Abstract

38 Vascular pathogens travel long distances through host veins leading to life-threatening, systemic
 39 infections. In contrast, non-vascular pathogens remain restricted to infection sites, triggering
 40 localized symptom development. The contrasting features of vascular and non-vascular diseases
 41 suggest distinct etiologies, but the basis for each remains unclear. Here, we show that the
 42 hydrolase CbsA acts as a phenotypic switch between vascular and non-vascular plant
 43 pathogenesis. *cbsA* was enriched in genomes of vascular phytopathogenic bacteria in the
 44 Xanthomonadaceae family and absent in most non-vascular species. CbsA expression allowed
 45 non-vascular *Xanthomonas* to cause vascular blight while *cbsA* mutagenesis resulted in reduction
 46 of vascular or enhanced non-vascular symptom development. Phylogenetic hypothesis testing
 47 further revealed that *cbsA* was lost in multiple non-vascular lineages and more recently gained by
 48 some vascular subgroups, suggesting that vascular pathogenesis is ancestral. Our results overall
 49 demonstrate how the gain and loss of single loci can facilitate the evolution of complex ecological
 50 traits.

51

52

53 MAIN TEXT

54

55 Introduction

56

57 Pathogenic microorganisms cause diseases of animals and plants. Some pathogenic species
58 colonize the host vasculature, which leads to systemic infection, while others remain localized to
59 non-vascular tissues. Complex structural and biochemical differences between vascular and non-
60 vascular tissues suggest that pathogens have multiple distinct adaptations to either environment,
61 yet the genetic and evolutionary bases of such adaptations are largely unknown.

62

63 Adaptations often occur through wholesale gain and loss of specific genes, resulting in more rapid
64 evolution compared with incremental changes at the DNA sequence level alone (1). In bacteria,
65 gene gain occurs primarily through horizontal gene transfer while gene loss or pseudogenization
66 occurs through multiple mechanisms, including transposon-mediated insertions and sequence
67 deletions in open reading frames (2–4). Especially for loci encoding ecologically relevant traits,
68 gene gain and loss effectively act as phenotypic switches, enabling rapid shifts between what
69 otherwise seem like complex lifestyles (3). For example, transitions between plant pathogenic and
70 commensal *Pseudomonas* (5), transitions between mutualist and parasitic phenotypes in nitrogen-
71 fixing bacteria (6, 7) and transitions between mutualistic and plant pathogenic *Rhodococcus* (8)
72 have all been shown to reproducibly occur through the gain and loss of genomic islands
73 containing multiple genes all contributing to the same phenotype. Such rapid evolutionary
74 dynamics have profound implications for our understanding of disease ecology and disease
75 management strategies.

76

77 In plants, vascular xylem and non-vascular parenchyma tissues represent distinct niches. Xylem is
 78 comprised of dead cells with highly reinforced walls organized into cylinders that provide plants
 79 with structural integrity and a means of long-distance fluid transport. In contrast, parenchyma
 80 tissues are composed of living cells and gas filled intercellular spaces. Xylem fluid consists
 81 primarily of water and mineral nutrients, and is thought to be nutrient limiting, although many
 82 vascular pathogens can use it to reach high densities (9). Xylem tissue runs throughout the plant,
 83 enabling the distribution of water from roots to leaves, but also serving as a potential pathway for
 84 rapid, systemic transport of pathogens.

85
 86 *Xanthomonas* (Gammaproteobacteria) is diverse genus of plant-associated Gram-negative
 87 bacteria that cause vascular and non-vascular diseases of over 200 monocot and dicot plant hosts
 88 (10). *Xanthomonas* species are separated into subgroups called pathovars (pv.) based on their
 89 phenotypic behavior such as symptom development (e.g. vascular or non-vascular) or host range
 90 (10). Vascular xanthomonads invade the water transporting xylem; non-vascular *Xanthomonas*
 91 species cause localized symptoms by colonizing the mesophyll. Although often closely related,
 92 the genetic determinants distinguishing vascular from non-vascular *Xanthomonas* lineages at the
 93 intraspecific level are not clear.

94
 95 Here, we used *Xanthomonas* as a model to study the etiology of plant vascular pathogenesis
 96 because this genus contains multiple independent pairs of strains from the same species that cause
 97 either vascular or non-vascular diseases. This enabled us to disentangle genetic features that are
 98 shared due to ancestry and those that may be shared due to common tissue-specific lifestyles.
 99 Given the tendency of bacteria to evolve through the gain and loss of genes organized into
 00 clusters or genomic islands, we hypothesized that vascular and non-vascular pathogenesis emerge
 01 through the gain and loss of small numbers of linked loci. Surprisingly, we found evidence

02 supporting the most extreme version of this hypothesis, where transitions between vascular and
03 non-vascular lifestyles are mediated by the repeated gain and loss of a single gene that acts as a
04 phenotypic switch.

05

06 **Results**

07

08 ***cbsA* is significantly associated with vascular pathogenesis**

09

10 We first identified high priority candidate genes associated with transitions to vascular and non-
11 vascular lifestyles. We classified predicted proteins from 59 publicly available whole genome
12 sequences of *Xanthomonas* and *Xylella* species into ortholog groups (OGs). We then conducted
13 an analysis of trait evolution across a SNP-based phylogeny where for each OG we tested the
14 hypothesis that transitions to vascular or non-vascular lifestyles were dependent on that OG's
15 presence or absence (Figure 1). The phylogenetic relationships between vascular and non-
16 vascular pathovars indicated that xylem pathogenesis is paraphyletic, i.e., not limited to a single
17 clade, an individual *Xanthomonas* sp., or host plant genus (Figure 1; Supplemental Figure 1&2).
18 Instead, vascular diseases of many host plant families are caused by different pathovars across the
19 *Xanthomonas* genus. We identified two OGs whose presence was strongly associated (Log Bayes
20 Factor >10) with the distribution of tissue-specific lifestyles (Figure 1; Supplemental Figure 1,
21 Supplemental Table 1&2). One OG (OG0003492) was highly associated with vascular
22 pathogenesis, while the other (OG0002818) was associated with non-vascular pathogenesis. For
23 this study, we focused on vascular pathogen-enriched OG0003492, which encodes a cell wall
24 degrading cellobiohydrolase (EC 3.2.1.4, glycosylhydrolase family GH6) called CbsA (11, 12).

25

Next, phylogenetic analysis of CbsA sequences revealed that distinct monophyletic lineages within this gene family are alternatively found in either vascular or non-vascular pathogen genomes (Figure 1B; Supplemental Figure 3). Within *Xanthomonas*, CbsA sequences form two distinct clades: the first contains sequences found in both vascular and non-vascular pathogen genomes, and the second contains sequences found exclusively in vascular pathogen genomes. All vascular pathogens with a CbsA homolog found in the first clade also possess a CbsA homolog found in the second clade, effectively possessing two copies of the CbsA gene (Figure 1B, Supplemental Figure 3). The association of specific CbsA clades with specific pathogen lifestyles, in addition to its occasional presence in multiple copies in vascular pathogen genomes, suggests that CbsA sequences found in either of these two clades have distinct biological functions, with sequences that are exclusive to vascular pathogens likely contributing to vascular pathogenesis.

Heterologous expression of *cbsA* bestows vascular pathogenesis to a non-vascular pathogen

Because *cbsA* was present in vascular and largely absent from non-vascular *Xanthomonas* species, we hypothesized that *cbsA* was either: A) gained by vascular *Xanthomonas* species or B) lost by non-vascular *Xanthomonas* species. To experimentally test the alternate models, we examined the effects of manipulating *cbsA* on the contrasting tissue-specific behavior of two closely related barley pathogens from the same species: vascular *Xanthomonas translucens* pvs. *translucens* (Xtt) and non-vascular *undulosa* (Xtu).

Xtt and Xtu both cause non-vascular bacterial leaf streak (BLS) disease of barley (13). However, only Xtt can colonize the xylem which leads to long distance bacterial blight (BB) symptom development (Figure 2A-C) (13, 14). Upon leaf clipping, only Xtt produces distant vascular BB;

meanwhile Xtu symptoms remain near the site of inoculation (Figure 2A). Moreover, Xtt strains contain an intact copy of *cbsA*, while *X. translucens* pv. *undulosa* contains a copy of *cbsA* that is disrupted in the 5' region by a transposase (Supplemental Figure 4).

As Xtt possesses *cbsA* while Xtu lacks an intact copy, we tested if the expression of CbsA promotes vascular symptom development in Xtu. Xtu miniTn7::*cbsA*_{Xtt}, a single insertion variant with an intact copy of *cbsA* from Xtt, caused distant leaf lesions of approximately 4.5 cm (Figure 2A-B). Moreover, expression of the characterized CbsA ortholog from the vascular rice pathogen Xoo (Xtu miniTn7::*cbsA*_{Xoo}) also permitted Xtu to cause distant symptom development consistent with a vascular pathogenic lifestyle. Using GFP-expressing strains, we reproducibly observed Xtu miniTn7::*cbsA*_{Xoo} inside the xylem similar to Xtt (Figure 2C). Wild-type Xtu did not produce vascular symptoms and was not detected in distant xylem vessels (Figure 2C). Therefore, the gain of *cbsA* from either of two different vascular pathogens is sufficient to promote xylem-mediated colonization and distant infection of leaves by non-vascular Xtu.

Impact of *cbsA* mutagenesis on vascular pathogenesis is dependent on genetic background

We found that the Xtt Δ *cbsA* mutant was still capable of causing vascular leaf blight, suggesting other unknown factors support vascular pathogenesis beyond CbsA alone (Figure 2D&F). However, while Xtt Δ *cbsA* could still cause systemic symptom development, the mutation of this cellulase altered this strain's pathogenic behavior by promoting the development of non-vascular, water-soaked lesions downstream of the xylem blight on 90% of infected leaves compared with only 10% of leaves on plants infected with wild-type vascular Xtt (Figure 2E&F). These water-soaked symptoms are typical of non-vascular disease development in Xtt and Xtu (13, 14).

Therefore, while vascular disease development is not completely abolished by *cbsA* mutagenesis, the absence of *cbsA* increased the development of non-vascular disease symptoms.

These results did not match previous reports that *cbsA* deletion mutants in *X. oryzae* pv. *oryzae* and *R. solanacearum* have reduced systemic virulence and vascular pathogenesis (15, 16). We therefore replicated and expanded upon these previous findings by mutating *cbsA* in *Xanthomonas oryzae* pv. *oryzae* and *Xylella fastidiosa* (Xanthomonadaceae). *X. oryzae* pv. *oryzae* causes bacterial blight of rice with systemic symptoms similar to Xtt on barley. *Xylella fastidiosa*, an insect-vectored, xylem pathogen, is the causal agent of Pierce's disease of grape and the emerging olive quick decline disease. *X. oryzae* pv. *oryzae* and *X. fastidiosa* deletion mutants were severely reduced in vascular symptom development, confirming and building upon previous reports (Supplemental Figure 5)(15). The variable effects of mutagenizing *cbsA* in Xtt versus *X. oryzae* pv. *oryzae* and *X. fastidiosa* indicate that the robustness of vascular phenotypes is lineage dependent within *Xanthomonas*, with certain species likely possessing multiple determinants in addition to *cbsA* that contribute to vascular pathogenesis.

The genomic location of *cbsA* alternates between four distinct neighborhoods

Across all examined genomes, *cbsA* is found embedded in one of four genomic neighborhood types with conserved gene synteny (Figure 1B&3). The localization of *Xylella fastidiosa*'s and *X. vasicola*'s *cbsA* in type 1 neighborhoods, combined with a lack of evidence suggesting horizontal gene transfer between these two species (Figure 1B), provides support that *cbsA* was present and organized in a type 1 context in the last common ancestor of *Xanthomonas* and *Xylella*. Based on this inference, it is likely that *cbsA* was then re-located into type 2, type 3 and type 4 neighborhoods through separate cis-transposition events as *Xanthomonas* spp. diversified. The

00 timing of transposition events 3 and 4 are uncertain due to lack of resolution in species-level
 01 relationships, but likely occur near to where indicated on the species tree (Figure 3). Within the
 02 gamma-proteobacteria, all known vascular pathogens in our dataset have a copy of *cbsA* localized
 03 in the context of type 1, 2, or 4 neighborhoods. Within *Xanthomonas*, sequences from the clade of
 04 CbsA homologs found in both vascular and non-vascular pathogens are located in type 3
 05 neighborhoods, while sequences from the clade of CbsA homologs found exclusively in vascular
 06 pathogen genomes are located in type 4 neighborhoods, further supporting the hypothesis that
 07 sequences belonging to either of these two clades have separate functions (Figure 2).

09 ***cbsA* has been independently gained by lineages now displaying vascular lifestyles**

11 *cbsA* and varying lengths of adjacent sequence experienced three horizontal transfers in the
 12 *Xanthomonas* genus mediated by homologous recombination events in flanking gene
 13 neighborhoods (events 7,8,9 in Figure 3, Supplemental Figure 6-8). Two transfers from what was
 14 likely the ancestor of the vascular pathogen *X. phaseoli* are coincident with the emergence of
 15 vascular lifestyles in xylem-adapted *X. campestris* pv. *campestris* and *X. citri* pv. *phaseoli*, and
 16 occurred within the context of type 4 neighborhoods (events 8 and 9 Figure 3; Supplemental
 17 Figure 6-8). The third transfer occurred in the context of a type 3 neighborhood, where neither the
 18 donor lineage of *X. vesicatoria* nor the recipient lineage of *X. citri* have been reported to be
 19 capable of vascular pathogenesis.

21 ***cbsA* was horizontally transferred from vascular gamma- to beta-proteobacteria**

23 We found additional evidence that *cbsA* was horizontally transferred from gamma-proteobacterial
 24 Xanthomonadaceae to the beta-proteobacterial xylem plant pathogens *R. solanacearum* and

25 *Xylophilus ampelinus* (Figure 3). *cbsA* sequences in both *X. translucens* pv. *translucens* and *R.*
 26 *solanacearum* are flanked on one or both sides by transposable elements (Figure 1B), providing a
 27 plausible mechanism for mediating horizontal transfer through transposition between these distant
 28 lineages. However, we could not test this specific hypothesis with confidence because the
 29 phylogenies of the transposable elements in question are complex and contain signatures of
 30 extensive horizontal transfer between strains.

31

32 ***cbsA* has been repeatedly lost from lineages now displaying non-vascular lifestyles**

33

34 At least 10 losses of *cbsA* are required to parsimoniously explain its distribution across the beta-
 35 and gamma-proteobacteria when taking into account all HGT events supported by phylogenetic
 36 hypothesis testing (Figure 3; Supplemental Tables 4-6). While the majority of losses are inferred
 37 using parsimony criteria (e.g. losses in non-vascular strains of *X. hortorum* and *X. fragariae*;
 38 Methods), several *cbsA* pseudogenes present in extant species directly support the hypothesis of
 39 repeated, independent losses through distinct inactivation mechanisms. For example, *cbsA* was
 40 independently pseudogenized in the non-vascular *X. translucens* pv. *undulosa* and *X. sacchari*
 41 through sequence deletions in its 5' coding region (Supplemental Figure 4&6). In contrast,
 42 transposable elements have disrupted the 5' region of *cbsA* in non-vascular *X. oryzae* pv.
 43 *oryzicola*, and are present in the type 4 neighborhoods of certain non-vascular *X. citri* subsp. *citri*
 44 and *X. fuscans* subsp. *aurantifolii* isolates that lack a copy of *cbsA* (Supplemental Figure 6). These
 45 examples of multiple, independent disruptions to *cbsA* in lineages displaying non-vascular
 46 lifestyles suggest that non-vascular pathogenesis convergently evolved through repeated gene
 47 loss.

48

49 **Discussion**

50

51 Systemic pathogens traverse host veins to move long distances, leading to life-threatening
 52 systemic infections. In contrast, non-vascular pathogens remain restricted to the site of infection,
 53 triggering localized symptom development with far fewer implications for host health. Although
 54 complex differences between these modes of infection suggest they have radically different
 55 origins, the results we present here suggest that vascular and non-vascular pathogenesis are two
 56 points on an evolutionary continuum, a finding with important implications for understanding and
 57 predicting pathogen evolution (Figure 4). By integrating comparative genomic, phylogenetic, and
 58 functional genetic analyses, we found evidence that vascular and non-vascular plant pathogenic
 59 lifestyles emerge from the repeated gain and loss of a single gene that can act as a phenotypic
 60 switch.

61

62 Our functional and phylogenetic results suggest that *cbsA* contributes to the evolution of
 63 *Xanthomonas* vascular pathogenicity, but to varying extent depending on the species considered.
 64 Xylem-specific pathogens, including *X. fastidiosa*, *X. oryzae* pv. *oryzae* and *R. solanacearum*,
 65 require CbsA for vascular pathogenesis, whereas Xtt, which induces both vascular and non-
 66 vascular disease symptoms, appears to use other factors beyond CbsA to colonize xylem
 67 vasculature. That the phenotypic outcomes of CbsA acquisition are dependent on genetic
 68 background suggests that there exist multiple evolutionary routes to vascular pathogenesis, and
 69 highlights the particularities of specific host-pathogen interactions. Nevertheless, the
 70 preponderance of phenotypic and phylogenetic evidence supports the hypothesis that *cbsA* was
 71 present in the last common ancestor of *Xanthomonas* and *Xylella*, has since played not only a
 72 historical but possibly a contemporary role in driving the emergence and re-emergence of tissue-
 73 specific behavior in the Xanthomonadaceae.

74

While we document repeated gains and losses of *cbsA*, the conditions that favor phenotypes resulting from either its presence or absence remain to be determined. Although *cbsA* homologs are among the highest expressed genes during xylem pathogenesis (9, 17), and are required for vascular pathogenesis in several species (Supplemental Figure 5), the contributions of CbsA to pathogen fitness remain unclear. Current theory suggests that there may be a fitness cost to retaining this gene and the vascular lifestyle it enables, given that CbsA induces immune responses and can prime the plant against *Xanthomonas* infection (15). Furthermore, cell wall degradation products, such as the CbsA enzymatic biproduct cellobiose, could act as a danger-associated molecular pattern in the plant mesophyll and may induce plant defenses through WRKY transcription factors (18). We therefore speculate that *cbsA*'s absence may be selected for to dampen recognition by the host and/or the elicitation of host immunity; however, these hypotheses remain to be tested.

Gene loss is a fundamental mechanism of adaptation (19). Especially for loci with large effects such as *cbsA*, only a minimal number of loss events are required to incur drastic changes to phenotype. Adaptive phenotypes arising through loss of function may emerge over shorter timescales compared with adaptive phenotypes arising through gains in function, as genes typically have more mutational opportunities for losing functions than for gaining functions (20). Even within our own limited dataset, we observed multiple mutational routes in the form of sequence deletions and transposable element insertions that led to the convergent loss of *cbsA* in different non-vascular pathogen lineages, which suggests that non-vascular phenotypes readily emerge in the Xanthomonadaceae.

Although there may be fewer mutational routes for gaining gene functions compared with losing them, our phylogenetic analyses revealed that rates of gain and loss may be balanced by latent

patterns in genome architecture, such as the conservation of synteny. Homologous recombination in bacteria is typically studied within species, and is considered to be important for maintaining genetic diversity in what would otherwise be clonal lineages (21). Less considered are the impacts of homologous recombination across species. Our results add to a growing body of literature suggesting that, while perhaps less common than intraspecific homologous recombination (22, 23), interspecific gene exchange facilitated by homologous recombination at syntenic loci is an important mechanism of adaptation (24). All three *cbsA* HGT events within *Xanthomonas* occurred through homologous recombination in syntenic neighborhoods flanking *cbsA* presence/absence polymorphisms, and two of these resulted in the reversal of an ancestral loss event (Figure 2), suggesting that synteny conservation potentiates not only gene gain but the reversal of lineage-specific gene loss. By effectively increasing an individual strain's ability to access cross-species pan-genomic material, the conservation of synteny is likely to be an important accelerator of ecological adaptation.

Overall, our study provides an integrated evolutionary and functional framework for studying the genetic bases of transitions between vascular and non-vascular pathogen lifestyles (Figure 4). Our experiments demonstrate that the acquisition of *cbsA* is sufficient for long-distance systemic pathogenesis in specific *Xanthomonas* pathogens. Conversely, the loss of *cbsA*, while not necessary to abolish vascular disease development, is sufficient for the development of non-vascular disease symptoms. We add to a growing body of literature that suggests that transitions between distinct bacterial ecotypes may be mediated by the recurrent gain and loss of few loci (5, 8). Although it remains to be determined how the processes of rapid gene gain and loss impact vascular and non-vascular evolution in other pathogenic microbes, our work suggests that these evolutionary events play an important role in shaping bacterial adaptation to specific host tissues.

25 **Materials and Methods**

27 **Comparative genomics for identification of vascular pathogen-specific genes**

29 Using Orthofinder v2.2.3 (25), we first created ortholog groups (OGs) from all predicted amino
 30 acid sequences derived from 171 complete and 8 partially complete publicly available assemblies
 31 from the Xanthomonadaceae and representative lineages across the beta- and gamma-
 32 proteobacteria in order to obtain a comprehensive comparative genomic dataset (Table S1).
 33 Consensus functional annotations for each OG were obtained by determining the most frequent
 34 protein family domain present among the members of the OG using InterProScan version 5.25-
 35 64.0 (26). Predicted proteins across all genomes were classified into 36,905 OGs using
 36 Orthofinder (Supplemental Table 2) (25)

38 Genomes were classified as vascular, non-vascular or unknown based on available information in
 39 the literature (Supplemental Table 1). The *Xanthomonas* species included xylem and parenchyma
 40 pathogens that infect diverse dicot and monocot crops such as rice, wheat, barley, cabbage,
 41 tomato, citrus and common bean. A distant vascular grape and citrus Xanthomonadaceae
 42 bacterium, *Xylella fastidiosa*, was also analyzed.

44 For analyses limited to the Xanthomonadaceae, we built a more resolved SNP-based parsimony
 45 tree using kSNP3 (27) from a set of publicly available complete and annotated genomes from
 46 different species in the Xanthomonadaceae family (optimum kmer size = 21; Supplemental Table
 47 1). Using the kSNP3 as a reference, associations were identified between the presence/absence of
 48 each ortholog group in the analyzed genomes and the vascular/non-vascular trait using
 49 BayesTraitsV3 (28). The likelihood that both traits (vascularity vs. gene presence) evolved

independently was compared to the likelihood they evolved dependently. Evidence of dependent evolution was assessed as Log Bayes Factors = 2(log marginal likelihood dependent model – log marginal likelihood independent model), and genes were considered to have strong evidence of dependent evolution with a Log Bayes Factor >10.

Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table S7. *Escherichia coli* strains were grown at 37°C in Luria-Bertani medium. *X. translucens* or *X. oryzae* cells were grown at 28°C on solid or liquid nutrient broth or peptone-sucrose rich media (14). When necessary, media were supplemented with gentamicin (15 µg/ml), kanamycin (25 µg/ml) or spectinomycin (50 µg/ml). See Table S7 for specific strains used in this study.

Recombinant DNA techniques

Total genomic and plasmid DNA were isolated by standard methods. *E. coli* and *Xanthomons* species were transformed as previously described (14). To construct complementation vectors of *cbsA_{Xtt}* and *cbsA_{Xoo}*, the gene regions including the native promoters were PCR-amplified from *X. translucens* pv. *translucens* str. UPB886. Each were cloned into pUC18miniTn7T to create pUC18miniTn7T::*cbsA_{Xtt}* and pUC18miniTn7T::*cbsA_{Xoo}* (29). For gene expression, *X. translucens* pv. *undulosa* strains were transformed with miniTn7 plasmids and pTNS1 to promote transposition and single gene insertion, and each was confirmed as described (29). We were unable to insert *cbsA* via miniTn7 *X. translucens* pv. *translucens* strain UPB886. We therefore sequenced *X. translucens* pv. *translucens* Δ *cbsA* with long read PacBio SMRT sequencing (Supplemental Figure 9). There were no notable differences in sequence between wild-type UPB886 and the Δ *cbsA*

mutant. For visualization of bacteria by fluorescence microscopy, *Xanthomonas* bacteria (Table S7) were transformed with vectors for GFP expression (pNEO-GFP) (30). See Supplemental Tables 7 & 8 for specific strains and primers, respectively, used in this study.

Plant growth conditions, inoculation methods and live imaging with confocal microscopy

Barley (*Hordeum vulgare* L. cv. Morex or Betzes) were grown in growth chambers with cycles of 16 hours of light per day at 22-24°C. Rice (*Oryza sativa* cv. Nipponbare) were grown in growth chambers with 16 hours of light per day at 28°C 70% relative humidity or in the greenhouse. Plant seeds were directly germinated in potting mix. For either barley or rice, one leaf per plant was inoculated by leaf-clipping 7-10 days after seeds were planted with a water-based inoculum (OD₆₀₀=0.1) or water as a control as previously described (14). Disease symptoms were assessed using at least n = 5 replications per condition. Statistical differences were evaluated using the one-way ANOVA with Tukey's multiple comparison test or Student's *t*-test when appropriate. Symptom development was evaluated 21 days post-inoculation.

For bacterial localization, barley plant leaves were inoculated as above. Whole leaf tissue was imaged 5-14 days post inoculation with a Leica SP2 AOBS (Wetzlar, Germany) laser scanning confocal microscope with 40X oil objective. Barley leaves were cut directly adjacent to the inoculation zone for asymptomatic plants and immediately downstream of symptoms for symptomatic plants. Plant tissue was mounted onto a glass slide with water and covered with a glass coverslip. A 488 nm laser was used for GFP excitation and emitted fluorescence was collected between 505 and 540 nm. A 405 nm and 633 nm lasers were used for autofluorescence and emitted fluorescence was collected between 410 and 460 nm to define plant cell structures and between 650 and 700 nm for chlorophyll. Three to six plants were examined per biological

replicate per treatment over three total biological replicates. Representative confocal images represent maximal projections calculated from 15 to 25 confocal planes acquired in the z dimension (increments of at least 0.5 mm).

.

Phylogenetic analyses

.

To decrease redundancy among strain- or species-specific genomes in our dataset while maintaining sample power, we built a preliminary 50% majority-rule consensus tree based on the maximum likelihood (ML) phylogenies of 139 amino acid alignments of single copy orthologs. We used this tree to guide our selection of at most 3 representative genomes from each *Xanthomonas* pathovar, ultimately arriving at a final dataset of 86 genomes (Table S1). Using this de-replicated genomic dataset, we then built a final 50% majority-rule consensus tree based on 81 amino acid-based ML phylogenies of single copy orthologs that had greater than 60% average bootstrap support, our rationale being that consolidating multiple gene trees with high support increases the robustness of species-level phylogenetic inference (31). We rooted the final consensus tree at the bifurcation between the beta- and gamma-proteobacteria.

.

All nucleotide and amino acid alignments were generated using MAFFT v7.047 with options '--auto' for automatic selection of best alignment strategy (32) and trimmed using trimAL v1.4 with options '-automated1' for heuristic method selection and '-gt 0.25' for removing all sites with gaps in $\geq 75\%$ of sequences (33). Sequences with gaps in $\geq 30\%$ of sites were removed. All ML trees were built using IQTREE v1.6.9 with option '-m MFP' to find the best-fitting model of sequence evolution (34). Majority-rule consensus trees were built using RAxML v8.2.11 (35).

.

Analysis of *cbsA* homologs and neighboring genomic regions

.25

.26 In order to determine the precise mechanisms and relative order through which *cbsA* was gained
 .27 and lost from the genomes in our dataset, we analyzed the evolutionary history and structural
 .28 features of all gene neighborhoods that flank *cbsA*. Using custom scripts, we first explored the
 .29 gene neighborhoods surrounding *cbsA* homologs (+/- 15kb) in the 86-genome dataset for
 .30 conserved synteny, as defined by orthogroup content conservation. For each of the four conserved
 .31 neighborhood types that we identified, we then re-searched all genomes for regions composed of
 .32 these genes, thus identifying all instances of each neighborhood in each genome, regardless of
 .33 whether *cbsA* was present or not (Table S3). In doing so, we could then leverage phylogenetic
 .34 evidence from flanking genes to support or reject competing hypotheses of gene duplications,
 .35 horizontal gene transfers and losses that may have resulted in *cbsA*'s extant distribution.

.36

.37 We built nucleotide-based ML phylogenies of *cbsA* and the genes from each neighborhood type,
 .38 and manually reconciled their evolutionary histories with the consensus species tree using a
 .39 combination of parsimony-based gene tree-species tree reconciliation and likelihood-based
 .40 phylogenetic testing (Supplemental Figures 6-8; Supplemental Tables 4-6). In order to robustly
 .41 root the *cbsA* tree for reconciliation analysis, we first retrieved the top 1000 hits in the NCBI nr
 .42 protein database (last accessed: 03/09/18) to the *cbsA* sequence in *X. campestris* (accession:
 .43 WP_076057318) and used them to build a midpoint rooted ML tree (available on the Figshare
 .44 repository). This tree was then used as a reference to root subsequent ML trees that focused only
 .45 on this study's clade of *cbsA* sequences of interest. We additionally built a ML tree with *cbsA*
 .46 sequences from the full 179 genome dataset in order to verify the final topology of the *cbsA* tree
 .47 built with the 86 genome de-replicated dataset (available on the Figshare repository). All other
 .48 gene trees were midpoint rooted.

.49

.50 All genomic regions were further annotated for transposable elements with BLAST using the
.51 ISFinder database in order to ensure a comprehensive structural annotation of mobile elements
.52 (36). Nucleotide sequences of the genomic regions that were missing *cbsA* were searched using
.53 BLASTn with a *cbsA* query to ensure any missing or incomplete *cbsA* coding regions were
.54 identified. The Mixture Model and Hidden Markov Model from the PhyML package were used to
.55 detect homologous recombination breakpoints in the untrimmed nucleotide alignments that were
.56 then manually inspected and refined if necessary (37).

.57

.58 **Phylogenetic hypothesis testing**

.59

.60 In each tree with a topology that suggested HGT, we compared the likelihood of the most likely
.61 tree obtained through a standard ML search (representing the hypothesis of HGT) with the
.62 likelihood of a constrained tree where sequences were forced to adhere to a topology that would
.63 be expected under a scenario of vertical inheritance (representing the hypothesis of no HGT). In
.64 this way, we could probabilistically assess whether a scenario of vertical inheritance or HGT best
.65 explained the observed sequence data. We used the approximately unbiased (AU) test with
.66 100,000 re-samplings using the RELL method (38) as implemented in IQTREE v1.6.9 (34) to
.67 identify the most likely tree among a set of constrained and optimal trees. The null hypothesis that
.68 the constrained tree had the largest observed likelihood was rejected at $\alpha \leq 0.05$. Practically, this
.69 meant that we inferred HGT by showing that the constrained ML tree was significantly worse
.70 (smaller log likelihood) than the optimal ML tree. Constrained ML tree searches were conducted
.71 using IQTREE v1.6.9 (34) by supplying a trimmed nucleotide alignment and a non-
.72 comprehensive, multifurcating constraint tree specifying the monophyly of particular sequences
.73 of interest to which the resulting ML tree was forced to adhere to (Figures S4-6; see Tables S4-6
.74 for all constraint criteria).

.75

.76 **Data visualization**

.77

.78 All phylogenetic trees were visualized using ETE3 v3.0.0b32 (39). All genomic regions were
.79 visualized using Easyfig (40).

.80

.81 **Supplemental Materials**

.82

.83 **References and Notes**

- .84 1. J. Iranzo, Y. I. Wolf, E. V. Koonin, I. Sela, Gene gain and loss push prokaryotes beyond
.85 the homologous recombination barrier and accelerate genome sequence divergence. *Nat.*
.86 *Commun.* (2019), doi:10.1038/s41467-019-13429-2.
- .87 2. E. V. Koonin, Y. I. Wolf, Genomics of bacteria and archaea: The emerging dynamic view
.88 of the prokaryotic world. *Nucleic Acids Res.* (2008), doi:10.1093/nar/gkn668.
- .89 3. A. T. Maurelli, R. E. Fernández, C. A. Bloch, C. K. Rode, A. Fasano, “Black holes” and
.90 bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella*
.91 spp. and enteroinvasive *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3943–8 (1998).
- .92 4. C.-H. Kuo, H. Ochman, Deletional Bias across the Three Domains of Life. *Genome Biol.*
.93 *Evol.* **1**, 145–62 (2009).
- .94 5. R. A. Melnyk, S. S. Hossain, C. H. Haney, Convergent gain and loss of genomic islands
.95 drive lifestyle changes in plant-associated *Pseudomonas*. *ISME J.* (2019),
.96 doi:10.1038/s41396-019-0372-5.
- .97 6. S. S. Porter, J. Faber-Hammond, A. P. Montoya, M. L. Friesen, C. Sackos, Dynamic
.98 genomic architecture of mutualistic cooperation in a wild population of *Mesorhizobium*.
.99 *ISME J.* (2019), doi:10.1038/s41396-018-0266-y.

7. K. G. Nandasena, G. W. O'Hara, R. P. Tiwari, J. G. Howieson, Rapid in situ evolution of nodulating strains for *Biserrula pelecinus* L. through lateral transfer of a symbiosis island from the original mesorhizobial inoculant. *Appl. Environ. Microbiol.* (2006), doi:10.1128/AEM.00889-06.
8. E. A. Savory, S. L. Fuller, A. J. Weisberg, W. J. Thomas, M. I. Gordon, D. M. Stevens, A. L. Creason, M. S. Belcher, M. Serdani, M. S. Wiseman, N. J. Grünwald, M. L. Putnam, J. H. Chang, Evolutionary transitions between beneficial and phytopathogenic rhodococcus challenge disease management. *Elife* (2017), doi:10.7554/eLife.30925.
9. J. M. Jacobs, L. Babujee, F. Meng, A. Milling, C. Allen, The in planta transcriptome of *Ralstonia solanacearum*: Conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio.* **3** (2012), doi:10.1128/mBio.00114-12.
10. M.-A. Jacques, M. Arlat, A. Boulanger, T. Boureau, S. Carrère, S. Cesbron, N. W. G. Chen, S. Cociancich, A. Darrasse, N. Denancé, M. Fischer-Le Saux, L. Gagnevin, R. Koebnik, E. Lauber, L. D. Noël, I. Pieretti, P. Portier, O. Pruvost, A. Rieux, I. Robène, M. Royer, B. Szurek, V. Verdier, C. Vernière, Using Ecology, Physiology, and Genomics to Understand Host Specificity in *Xanthomonas*: French Network on *Xanthomonads* (FNX). *Annu. Rev. Phytopathol.* **54** (2016), doi:10.1146/annurev-phyto-080615-100147.
11. L. Tayi, S. Kumar, R. Nathawat, A. S. Haque, R. V. Maku, H. K. Patel, R. Sankaranarayanan, R. V. Sonti, A mutation in an exoglucanase of *Xanthomonas oryzae* pv. *oryzae*, which confers an endo mode of activity, affects bacterial virulence, but not the induction of immune responses, in rice. *Mol. Plant Pathol.* (2018), doi:10.1111/mpp.12620.
12. G. T. Beckham, J. Ståhlberg, B. C. Knott, M. E. Himmel, M. F. Crowley, M. Sandgren, M. Sørli, C. M. Payne, Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Curr. Opin. Biotechnol.* (2014), , doi:10.1016/j.copbio.2013.12.002.

13. C. Bragard, E. Singer, A. Alizadeh, L. Vauterin, H. Maraite, J. Swings, *Xanthomonas translucens* from Small Grains: Diversity and Phytopathological Relevance. *Phytopathology*. **87**, 1111–1117 (1997).
14. C. Pesce, J. M. Jacobs, E. Berthelot, M. Perret, T. Vancheva, C. Bragard, R. Koebnik, Comparative Genomics Identifies a Novel Conserved Protein, HpaT, in Proteobacterial Type III Secretion Systems that Do Not Possess the Putative Translocon Protein HrpF. *Front. Microbiol.* **8**, 1177 (2017).
15. G. Jha, R. Rajeshwari, R. V Sonti, Functional interplay between two *Xanthomonas oryzae* pv., *oryzae* secretion systems in modulating virulence on rice. *Mol. Plant. Microbe Interact.* **20**, 31–40 (2007).
16. H. Liu, S. Zhang, M. a Schell, T. P. Denny, Pyramiding unmarked deletions in *Ralstonia solanacearum* shows that secreted proteins in addition to plant cell-wall-degrading enzymes contribute to virulence. *Mol. Plant. Microbe Interact.* **18**, 1296–1305 (2005).
17. J. F. González, G. Degrassi, G. Devescovi, D. De Vleeschauwer, M. Höfte, M. P. Myers, V. Venturi, A proteomic study of *Xanthomonas oryzae* pv. *oryzae* in rice xylem sap. *J. Proteomics*. **75**, 5911–5919 (2012).
18. C. de A. Souza, S. Li, A. Z. Lin, F. Boutrot, G. Grossmann, C. Zipfel, S. C. Somerville, Cellulose-Derived Oligomers Act as Damage-Associated Molecular Patterns and Trigger Defense-Like Responses. *Plant Physiol.* (2017), doi:10.1104/pp.16.01680.
19. R. Albalat, C. Cañestro, Evolution by gene loss. *Nat. Rev. Genet.* (2016), , doi:10.1038/nrg.2016.39.
20. A. K. Hottes, P. L. Freddolino, A. Khare, Z. N. Donnell, J. C. Liu, S. Tavazoie, Bacterial Adaptation through loss of Function. *PLoS Genet.* (2013), doi:10.1371/journal.pgen.1003617.
21. B. J. Shapiro, J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabó, M.

F. Polz, E. J. Alm, Population genomics of early events in the ecological differentiation of bacteria. *Science* (80-.). (2012), doi:10.1126/science.1218198.

22. C. L. Huang, P. H. Pu, H. J. Huang, H. M. Sung, H. J. Liaw, Y. M. Chen, C. M. Chen, M. B. Huang, N. Osada, T. Gojobori, T. W. Pai, Y. T. Chen, C. C. Hwang, T. Y. Chiang, Ecological genomics in *Xanthomonas*: The nature of genetic adaptation with homologous recombination and host shifts. *BMC Genomics* (2015), doi:10.1186/s12864-015-1369-8.

23. N. Potnis, P. P. Kandel, M. V. Merfa, A. C. Retchless, J. K. Parker, D. C. Stenger, R. P. P. Almeida, M. Bergsma-Vlami, M. Westenberg, P. A. Cobine, L. De La Fuente, Patterns of inter- and intrasubspecific homologous recombination inform eco-evolutionary dynamics of *Xylella fastidiosa*. *ISME J.* (2019), doi:10.1038/s41396-019-0423-y.

24. E. A. Newberry, R. Bhandari, G. V. Minsavage, S. Timilsina, M. O. Jibrin, J. Kemble, E. J. Sikora, J. B. Jones, N. Potnis, Independent evolution with the gene flux originating from multiple *Xanthomonas* species explains genomic heterogeneity in *Xanthomonas perforans*. *Appl. Environ. Microbiol.* (2019), doi:10.1128/AEM.00885-19.

25. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* (2015), doi:10.1186/s13059-015-0721-2.

26. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, S. Hunter, InterProScan 5: Genome-scale protein function classification. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu031.

27. S. N. Gardner, T. Slezak, B. G. Hall, kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* (2015), doi:10.1093/bioinformatics/btv271.

28. M. Pagel, A. Meade, BayesTraits. 2005 *IEEE Comput. Syst. Bioinforma. Conf. Work.*

Poster Abstr. (2005), doi:10.1109/CSBW.2005.110.

29. K. H. Choi, H. P. Schweizer, mini-Tn7 insertion in bacteria with single attTn7 sites:

Example *Pseudomonas aeruginosa*. *Nat. Protoc.* (2006), doi:10.1038/nprot.2006.24.

30. S.-W. Han, C.-J. Park, S.-W. Lee, P. C. Ronald, An efficient method for visualization and growth of fluorescent *Xanthomonas oryzae* pv. *oryzae* in planta. *BMC Microbiol.* **8**, 164 (2008).

31. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* (2013), doi:10.1038/nature12130.

32. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* (2013), doi:10.1093/molbev/mst010.

33. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* (2009), doi:10.1093/bioinformatics/btp348.

34. L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* (2015), doi:10.1093/molbev/msu300.

35. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (2014), doi:10.1093/bioinformatics/btu033.

36. P. Siguier, ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* (2006), doi:10.1093/nar/gkj014.

37. B. Boussau, L. Guéguen, M. Gouy, A mixture model and a Hidden Markov Model to simultaneously detect recombination breakpoints and reconstruct phylogenies. *Evol. Bioinforma.* (2009).

38. H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.*

(2002), doi:10.1080/10635150290069913.

39. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* (2016), doi:10.1093/molbev/msw046.

40. M. J. Sullivan, N. K. Petty, S. A. Beatson, Easyfig: A genome comparison visualizer. *Bioinformatics* (2011), doi:10.1093/bioinformatics/btr039.

Acknowledgments

General: The authors are grateful to the French Xanthomonads Network, Jeffery Chang (Oregon State), Stephen Cohen (Ohio State) and Tiffany Lowe-Power (UC—Davis) for fruitful intellectual discussions.

Funding: An NSF Postdoctoral Fellowship in Biology (1306196) to MJM, a US Fulbright Scholar Award to Belgium to MJM, a USDA-NIFA Postdoctoral Fellowship (2017-67012-26116) to MJM, a COST SUSTAIN travel grant to MJM and a NSF-NIFA joint PBI grant (2018-05040) to MJM, JML and JEL. NSF (DEB-1638999) to JCS. A Fonds de Recherche du Quebec-Nature et Technologies Doctoral Research Scholarship to EGT. AJ and LDN are supported by the NEPHRON project (ANR-18-CE20-0020-01). This work was supported by a Ph.D. grant from the French Ministry of National Education and Research to AC. LIPM is part of the TULIP LabEx (ANR-10-LABX-41; ANR-11-IDEX-0002-02).

Author contributions: MJM conceptualized and MJM and RK supervised the conducted research. EGT, AC and APQ equally conducted research and provided formal analysis. CP, TV and JML conducted additional research. MJM, EGT, APQ, JS, LDN wrote the original draft, while RK, CA, LG, BS, SC, VV, JEL, CB and GB participated in reviewing and editing the manuscript.

Competing interests: Authors declare no competing interests

Data and materials availability: All data trimmed alignments, optimal and constrained maximum likelihood tree files, orthogroup assignments, and custom scripts are available on the Figshare data repository (DOI: 10.6084/m9.figshare.8218703).

Figures

Figure 1. The cellobiohydrolase CbsA is associated with transitions to vascular pathogenic lifestyles in Gram-negative pathogens. A) Highest-ranking associations between ortholog group presence/absences and evolutionary transitions between vascular and non-vascular lifestyles in the *Xanthomonadaceae*. A genome-based SNP phylogeny is shown to the left, with strains from the same species condensed into clades. Classifications of each strain as vascular (blue), non-vascular (yellow) or unknown (gray) are depicted to the right of each tip, followed by a heatmap summarizing, for each strain, the presence (black) or absence (white) of the two gene ortholog groups whose distributions are most strongly supported to be dependent on vascular lifestyle status (determined by model testing through the ranking of log Bayes Factors; Methods). Additional figure details can be found in Supplemental Figures 1&5. B) A phylogenetic tree based on *CbsA* amino acid sequences from strains with whole genome sequences found in (A), where branches on the tree are color coded according to pathogenic lifestyle. To the right of each tip is a schematic depicting the neighborhood type in which that particular *cbsA* sequence is found, where the four possible neighborhood types are defined based on conserved synteny (indicated by color-coded gene models corresponding to specific ortholog groups). Vascular

51 bacteria possess *cbsA* homologs located in type 1, 2, and 4 neighborhoods, while non-vascular
 52 bacteria possess *cbsA* homologs found primarily in type 3 neighborhoods. Note that strains of the
 53 vascular pathogen *X. campestris* pv. *campestris* have two copies of *cbsA* located in either type 3
 54 or type 4 neighborhoods.

55

56

Figure 2. Experimental gain and loss of CbsA facilitates transitions between vascular and non-vascular pathogenic lifestyles. A) Addition of either *cbsA* from vascular *X. translucens* pv. *translucens* (Xtt) or *cbsA* from vascular *Xanthomonas oryzae* pv. *oryzae* (Xoo) to non-vascular *X. translucens* pv. *undulosa* (Xtu) permits development of chlorotic lesions indicative of vascular disease on barley 21 days post-inoculation (dpi) B) Corresponding vascular lesion lengths, with significant differences among treatments indicated by a-d (n = 6, P < 0.02) C) Representative confocal images of vascular bundles downstream of leaf lesions on barley 12 dpi with green fluorescent protein (GFP) transformed strains demonstrate gain of vascular colonization by Xtu *cbsA*_{Xoo}. Green indicates bacterial cells expressing GFP; magenta indicates chlorophyll autofluorescence outlining non-vascular mesophyll cells; cyan indicates autofluorescence outlining xylem cell walls or phenylpropanoid accumulation in mesophyll cells. D-E) Lesion lengths or incidence of non-vascular water soaked lesions were quantified after barley leaf clipping 14 dpi with Xtt Δ *cbsA*. Bars in E) represent percent leaves showing symptoms with dots included to display individual leaf lesions incidence. F) Images of symptomatic barley leaves infected with Xtt and Xtt Δ *cbsA*, where water soaked lesions are indicated with black arrows indicating non-vascular symptom development.

A

B

175
 176 **Figure 3. Repeated horizontal transfer, transposition, and gene loss events drive the**
 177 **distribution of *cbsA* in gram-negative bacteria.** A) A 50% majority rule consensus tree
 178 summarizing 81 conserved single copy ortholog trees is shown to the left, with the names of the
 179 75 individual isolates consolidated into relevant taxonomic groupings. Inferred horizontal gene
 180 transfer (HGT), transposition, and loss events are drawn and numbered on the tree, and further
 181 described in B). The matrix to the right of tree indicates the presence/absence of one of four
 182 distinct genomic neighborhood types (shaded/unshaded cells) in which *cbsA* homologs are found
 183 within a given genome (presence of *cbsA* indicated by an overlaid red arrow). Note that in many
 184 cases, all of the constituent genes making up a specific neighborhood are present in a given
 185 genome save for *cbsA* (indicated by the absence of an overlaid red arrow). This tree has been
 186 lightly edited for viewing purposes by removing several taxa from outside the Xanthamonadales,
 187 and can be viewed in its entirety in Supplemental Figure 3. B) The sequence of inferred
 188 evolutionary events drawn onto the tree in A). Genomic neighborhood types are represented by
 189 schematics where gene models are color-coded according to ortholog group. The color-coding of
 190 neighborhood types is consistent across both panels.

191

Figure 4. The evolution of vascular and non-vascular pathogenesis in plant-associated *Xanthomonas* bacteria is driven by the gain and loss of *cbsA*. Our combined phenotypic and phylogenetic analyses support a model where vascular and non-vascular pathogenesis exist as two points on the same evolutionary continuum that is traversed by either the acquisition or loss of a single cellobiohydrolase, *cbsA*.

A

bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.24.058529>; this version posted April 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

B

CbsA

Neighborhood type

A **B** **C** **D** **E** **F**

Systemic

Non-systemic

Disease spectrum

Vascular

Niche adaptation

Non-vascular

Gene loss/inactivation

cbsA

~~*cbsA*~~

Horizontal
transfer

cbsA