Materials and Methods (Unabridged)

Isolation of bacterial strains from a national onion disease survey and preliminary identification using 16S rRNA gene sequencing.

P. agglomerans strains conforming to the naming convention YYST####, indicating the two-digit year, two letter U.S.A. state abbreviation, and sequential isolate number (e.g., 20WA0189), were isolated as part of a United States national survey of bacteria associated with diseased onion plants and bulbs conducted during 2020 and 2021 in 11 onion-growing states (CA, CO, GA, ID, NM, NY, OR, PA, TX, UT, and WA). Isolates were recovered from symptomatic onion foliage and bulbs using standardized protocols. Briefly, small sections of leaf or bulb tissue (2 mm x 2 mm) were excised along the margin of the necrotic tissues using a sterile scalpel. The excised tissue was macerated in 100 µl of sterile water, and a sterile inoculation loop was used to streak macerate suspension onto nutrient agar (beef extract 3 g/L, peptone 5 g/L, agar 15 g/L, pH 7.0) or onion extract agar medium [1]. After 48 hours of incubation at 28°C, pure cultures were generated through re-streaking and sub-culturing, as necessary. Pure cultures were preserved from nutrient broth (NB) cultures as 15% glycerol cryostocks stored at -80°C. Initial genus identification of isolates was conducted by 16S rRNA gene PCR assay and sequencing. DNA template for the PCR assay was prepared by combining 10 µl of overnight NB culture with 100 µl of sterile dH_2O, followed by 95°C incubation for 10 minutes, and centrifugation of cell debris. See Table S1 for 16S rRNA primer sequences and annealing temperatures. 16S rRNA gene amplicons sequences were determined by Sanger sequencing using either the forward or reverse PCR primer (Table S1). Following sequencing and quality trimming, the bacterial strains were classified to genus using the SILVA Alignment, Classification and Tree service (https://www.arb-silva.de/aligner/). Strains were phenotyped further using the red onion scale necrosis (RSN) assay [2]. The study included 382 survey isolates representing a diversity of geographic locations in the U.S.A., sampling dates, onion tissue types, and RSN phenotypes, as well as 47 strains isolated from onion plants and bulbs and water sources near onion fields donated from the collection of Steven Beer at Cornell University, and 73 *Pantoea* strains previously isolated from onion or other sources by coauthors

of this study. *Pantoea* strains (Table S4) were screened using a species-specific PCR assay [3] (Table S1) to identify *P. agglomerans* strains. PCR amplification and sequencing of the *infB* gene was used for secondary species confirmation for 32 *Panteoa* strains [3, 4].

Bacterial culturing

P. agglomerans and *Escherichia coli* strains were cultured in Luria-Bertani (LB) broth (containing 10 g/L tryptone, 5 g/L yeast extract, and 5 (University of Georgia, UGA) or 10 (U.S. Department of Agriculture, USDA) g/L NaCl) or agar (with 15 (UGA) or 16 (USDA) g/L agar) at temperatures of 28℃ and 37℃, respectively. Antibiotics were added to the medium at the following final concentrations: gentamicin (Gm) at 10 µg/mL, kanamycin (Kan) at 50 µg/mL, streptomycin (Sm) at 100 µg/mL and rifampicin (Rp) at 50 µg/mL.

Whole genome sequencing and assembly

Bacterial genomic DNA was extracted from overnight LB liquid cultures using a modification of the protocol from [5] or either the Puregene Yeast/Bact. Kit B (Qiagen) or the Wizard HMW DNA Extraction Kit (Promega), following the manufacturer's recommended Gram-negative bacterial DNA extraction protocol. For strain AR1a, DNA was extracted using a phenol: chloroform: isoamyl alcohol extraction and ethanol precipitation as in [6]. Genomic DNA was submitted to MicrobesNG, SeqCenter, or the Biotechnology Resource Center (BRC) Genomics Facility (RRID:SCR_021727) at the Cornell Institute of Biotechnology for Illumina short read sequencing. Oxford Nanopore long read sequencing was conducted either by Plasmidsaurus or in house at the USDA Agricultural Research Service (ARS) in Ithaca, New York. Genome sequencing at different facilities was conducted as described below.

1. MicrobesNG

The samples were sequenced paired-end on an Illumina sequencer to generate 250 bp reads that were put through a standard analysis pipeline established at MicrobesNG (Birmingham, U.K). Briefly, the reads were mapped to the closest available reference genome identified by Kraken (https://github.com/DerrickWood/kraken), and assembled *de novo* using SPAdes

(https://github.com/ablab/spades) to ensure quality assembly and 30X coverage. Resulting assemblies were annotated using Prokka (https://github.com/tseemann/prokka).

2. SeqCenter (Previously MiGS)

The Illumina DNA Prep kit and IDT 10 bp UDI indices were used to prepare sample libraries that were sequenced on an Illumina NextSeq 2000. The 151-bp-long, paired-end reads were demultiplexed and subjected to quality control and adapter trimming with BCL Convert (v. 3.9.3) (Illumina).

3. Long-read sequencing performed by Plasmidsaurus

The Oxford Nanopore Technologies Ligation Sequencing Kit version SQK-LSK114 was used to prepare samples for sequencing on GridION 10.4.1 flowcells (FLO-MIN114) with the "super accuracy" basecaller in MinKNOW.

4. Short read sequencing performed by Cornell BRC Genomics Facility

For strains BH6c, SUH 1, and CB1, Illumina libraries were prepared using NEBNext Multiplex Oligos for Illumina (New England Biolabs), NEBNext Ultra FS DNA Library Prep Kit for Illumina (New England Biolabs), and AMPure XP beads (Beckman Coulter) according to manufacturer's directions. Quality control of libraries was performed using BioAnalyzer (Cornell Biotechnology Resource Center Genomics Core) to calculate average library size and relative concentration of primers. Libraries were quantified using Qubit dsDNA HS assay (Thermo Fisher Scientific) and were pooled and submitted to Cornell BRC. For other strains sequenced at the BRC, purified genomic DNA was quantified using the Qubit dsDNA BR Assay and submitted. The BRC used the Nextera Flex kit (Illumina) to generate sequencing libraries that were sequenced using the MiSeq or NextSeq 500 instruments (Illumina) to generate ~150-bp paired end reads.

5. Long-read sequencing performed by Mt. Sinai Medical Center

AR1a was sequenced on a PacBio PSII (Pacific Bioscience) at the Icahn School of Medicine, Mount Sinai, using methods described in [7] and [6].

6. Long-read sequencing performed by USDA-ARS

Genomic DNA quality was checked for potential degradation by agarose gel electrophoresis. DNA was quantified using a Qubit dsDNA BR Assay (ThermoFisher Scientific). Oxford Nanopore long-read DNA sequencing was conducted using the Spot-ON Flow Cell R9 version in a MinION Mk1B sequencing device. The library was prepared using the Rapid Barcoding Kit SQK-RBK and the flow cell was primed using the Flow Cell Priming Kit EXP-FLP002, according to the manufacturer's instructions (Oxford Nanopore Technologies).

Basecalling was performed on the raw Nanopore data using Guppy (https://community.nanoporetech.com/downloads/guppy/). Indexed libraries were demultiplexed with “qcat” (https://github.com/nanoporetech/qcat). The specific version of each program used for each strain can be found with the NCBI SRA record for each sequence deposited.

Genome Assembly and Annotation

Performed by Plasmidsaurus

Resulting raw reads were filtered with Filtlong (https://github.com/rrwick/Filtlong) to remove the lowest quality 5% of reads as well as reads with a length <2000 nucleotides. The filtered reads were subsampled to 400 Mb using rasusa (https://github.com/mbhall88/rasusa) and the bacterial genome assembly was carried out using the flye assembler (https://github.com/fenderglass/Flye) with -nano-hq and -read-error .02. Subsequent to assembly, polishing of the reconciled consensus was performed using medaka (https://github.com/nanoporetech/medaka). Finally, the genome annotations were made with bakta (https://github.com/oschwengers/bakta).

Performed at the University of Georgia (UGA)

FastQC (v. 0.11.8) [8] was used to evaluate the quality of the raw reads, and Trimmomatic (v. 0.39) [9] was employed to filter out low-quality reads. The filtered paired-end reads were then assembled into contigs using SPAdes (v3.15.3) [10], and the quality and completeness of the assemblies were assessed using QUAST (v. 5.0.2) [11] and BUSCO (v. 5.2.2) [12].

Performed by the USDA-ARS in Ithaca, NY

The AR1a genome was assembled using Flye 2.6 [13] as described by [6].

The details of the assembly of each genome varied among strains. However, in general, each method followed the one suggested in the Trycycler documentation (https://github.com/rrwick/Trycycler): First, FiltLong (https://github.com/rrwick/Filtlong) was used to trim and filter the Nanopore reads. Next, filtered reads were subsampled randomly to 50x coverage (assuming a 5 Mb genome) to make 15 read subsets, and 5 draft assemblies each were generated using Flye [13], MiniPolish [14], and Raven [15] for a total of 15 draft assemblies. Trycycler [16] was used to cluster and reconcile the 15 assemblies into a single consensus assembly. Medaka (https://github.com/nanoporetech/medaka) was used to polish the consensus assembly with the trimmed and filtered long reads. FASTP [17] was used to trim and filter the raw Illumina reads, and NextPolish [18] was used to polish the assembly with trimmed and filtered Illumina reads. Custom scripts were used to “normalize” the final assembly: (a) instances of the phi-X phage genome were removed; (b) the contigs were renamed, “chromosome”, “plasmidA”, “plasmidB”, etc., based on sequence length; (c) sequences were rotated to put, e.g., DnaA near position 1 on the positive strand.

The quality of each genome was assessed by annotating it with PGAP [19] and evaluating its proteome with BUSCO [12]. Finally, an independent, “bottom-up” assembly was generated using Unicycler [20]. The Trycycler and Unicycler assemblies were compared using “dnadiff” from Mummer to identify potential assembly anomalies.

The specific version of each program used for each strain and can be found with the NCBI Assembly record for that genome.

Data availability of genomes

The sequencing and assembly details, metadata, genome assemblies and corresponding annotations as well as the raw read data were deposited in the NCBI GenBank database under BioProject numbers PRJNA1069770 (UGA) and PRJNA642846 (USDA). The accessions of the deposited genomes can be found in Table S16.

Roary

In addition to 81 assembled genomes resulting from this study, 100 publicly available *P. agglomerans* genome assemblies were obtained from the NCBI GenBank database. The selection of these additional assemblies was based on their BUSCO scores and the availability of strain information, including the source, date, and location of isolation. The strain information for all 81 *P. agglomerans* genomes sequenced for this analysis can be found in Table S16, and information for the 100 publicly available genomes used for this analysis can be found in Table S3. Following the assembly, all 181 genomes were annotated for Gram negative bacteria using Prokka (v. 1.14.5) [21].

The core genes of the 181 annotated *P. agglomerans* genomes were identified with Roary (v. 3.13.0) [22]. The concatenated nucleotide sequences of core genes were aligned with MAFFT [23] which was used to construct a phylogenetic tree using FastTree (v. 2.1.11) [24]. The general time reversible (GTR) nucleotide substitution model was used to construct an approximately maximum likelihood phylogenetic tree rooted at the midpoint, and the reliability of the internal branches was assessed by approximate likelihood-ratio test (aLRT) as the bootstrap method. The resulting tree was visualized and edited in iTOL (v. 6.8) [25] available at https://itol.embl.de/.

Average nucleotide identity

FastANI (v. 1.3.3) [26] was employed to compute the pairwise average nucleotide identity (ANI) across the 187 *P. agglomerans* genomes. In addition to the Roary-analyzed 181 strains, six additional genomes of *P. agglomerans* from Steven Beer's collection at Cornell University were

included in the analysis. Following calculation, the ANI values were clustered hierarchically using the average linkage method and the one-minus Pearson correlation metric. Subsequently, a correlation plot was generated for all genomes using the ANI values by utilizing the Pearson correlation matrix option within Morpheus, a tool accessible at: https://software.broadinstitute.org/morpheus/.

Phylogenetic analysis of a housekeeping gene, *gyrB*

Phylogenetic analysis of *gyrB* (DNA gyrase β-subunit) was conducted for 181 *P. agglomerans* strains by constructing a maximum-likelihood tree. The full-length nucleotide sequences of the *gyrB* gene (2,409 bp) were extracted from the 181 whole genome sequenced (WGS) *P. agglomerans* strains using the BLAST function in Geneious Prime (v. 2013.1.2). The *gyrB* gene sequences from other *Pantoea* species were obtained from NCBI GenBank, with corresponding GenBank accession numbers provided in Table S16. The *gyrB* gene sequences were aligned using the MAFFT alignment tool (v. 7.48), followed by trimming the alignment in Geneious Prime. The trimming process specifically targeted the regions corresponding to the partial *gyrB* sequence used for analysis by [27] (745 bp) or [4] (688 bp). The nucleotide substitution model for each alignment was chosen based on the Akaike Information Criterion (AIC) using the Smart Model Selection function in the online PhyML [28] 3.0 tool. Branch support was evaluated through 1,000 replicate bootstrap analysis, and the tree was rooted with an outgroup, *Tatumella ptyseos* NCTC 11468^{T}. The tree was visualized and edited in iTOL (v. 6.8). The partial *gyrB* sequence of *P. agglomerans* strains without genome sequences (non-WGS) was determined by PCR amplification and sequencing, as previously described [27]. See Table S1 for primer sequences.

Identification of characterized secondary metabolite and disease-associated gene clusters in sequenced genomes.

Using BLAST+ [29] (v. 2.11.0), the nucleotide sequences of the virulence and secondary metabolite gene clusters such as HiVir [30], allicin tolerance (*alt*) gene cluster [2], copper tolerance gene cluster (this study) were searched against the genomes of 181 *P. agglomerans*

strains to elucidate the genomic signatures of onion-pathogenicity. We also examined well-characterized secondary metabolite gene clusters known to be present in *Pantoea* species [31], including pantocin A (U81376.2) [32], *Pantoea* Natural Product 3 (MN807451.1) [33] and herbicolin (CP068441.1) [34]. A cluster was considered present if 80% of the total gene cluster length was covered in the queried genome with a minimum of 70% nucleotide similarity. Following these criteria, we compiled a metadata table (Table S14) indicating the presence and absence of these clusters. This table was then used as an input annotation file for iTOL (version 6.8, available at https://itol.embl.de/) to integrate it with the phylogenetic tree constructed using core genes.

Selection of phylotype strains

To compute a set of potential representative strains for each phylogroup, we first computed the MASH [35] distance for each pair of genomes in the phylograph using `mashtree` v. 1.4.6 [36]. Then, for each genome, we computed the average MASH distance to all other genomes within the phylogroup, generating a list of 10 strains. In a separate analysis, the gene group presence/absence matrix generated by the Roary analysis described above was used to compute the average Manhattan distance between each pair of closed genomes within each phylogroup. Within each phylogroup, strains with small MASH to other phylogroup strains, with large sets of shared orthologs, and that were also available in public culture collections were considered phylotype strains.

Design and use of HiVir detection 'PanHiVir' primers.

The full-length HiVir gene clusters (from *hvrA* to *hvrK)* from 31 HiVir-positive *P. agglomerans* strains with whole genome sequencing data, as well as from the genomes of other *Pantoea* species, such as *P. ananatis* 97-1R (GCA_002952035.2), LMG 2665^{T} (GCF_000710035.2), *P. allii* 20TX0020 (GCA_022585405.1), and *P. stewartii* subspecies *indologenes* pv. *cepacicola* PNA 03-3 (GCA_003201175.1), were downloaded from NCBI GenBank. The extraction and alignment were conducted using BLAST and MAFFT alignment functions in Geneious Prime (version 2013.1.2). Regions of perfect conservation within the HiVir gene cluster across these four *Pantoea* species

were identified manually. Subsequently, we designed PanHiVir primers (Table S1) targeting the *hvrD* gene to detect the HiVir gene cluster in these *Pantoea* species. These primers were used in colony PCR assays for a total of 502 *Pantoea* strains (Table S4). The colony PCR reaction setup and protocol were similar to the *P. agglomerans*-specific PCR assay, with the exception of the annealing temperature (Table S1).

copC detection PCR assay and phylogenetic analysis

The nucleotide sequences of *copABCD* genes from a total of 65 *P. agglomerans* strains that contained copper tolerance gene clusters were aligned using MAFFT in Geneious Prime (v. 2013.1.2). For the design of *copC* detection primers, we selected conserved regions flanking the *copC* gene. Subsequently, we performed colony PCR assays using these *copC* primers to screen an additional 43 *P. agglomerans* strains that lacked whole genome sequencing. The details of the *copC* primer sequences and the annealing temperatures can be found in Table S4. A total of 67 *copC* gene sequences (381 bp) from 65 *P. agglomerans* strains (comprising 53 with whole genome sequencing data and 12 without) were extracted and subjected to phylogenetic analysis following the procedure previously described for the *gyrB* gene. However, unlike the *gyrB* maximum-likelihood tree, the *copC* tree was rooted at its midpoint without an outgroup.

Clinker diagrams

The annotated HiVir and copper tolerance gene clusters from the closed genome strains of *P. agglomerans* were extracted and exported in GenBank format using Geneious Prime (v. 2013.1.2). The exported files were uploaded to the online comparative gene cluster analysis toolbox (available at: https://cagecat.bioinformatics.nl/tools/clinker) to generate a Clinker [37] diagram. The minimum alignment sequence identity was set to a default value of 0.3 and the coloring scheme of diagram was changed manually.

Analysis of plasmid sequences

The NCBI dataset program v. 15.23.0 (https://www.ncbi.nlm.nih.gov/datasets) was used to download all “complete” *P. agglomerans* (taxon = 549) genomes. In addition, the complete

genomes assembled as part of this study were included to produce a single pool of complete genomes for analysis. To ensure that the genome annotations were consistent, each genome in the pool was reannotated with Prokka v. 1.14.6 [21]. In order to improve the quality of the output, each genome's RefSeq or locally-generated PGAP [19] annotations, as appropriate, were provided to Prokka (--proteins GENOME.gbk) to use as a set of primary annotations. Next, BUSCO v. 5.5.0 [12] was used to evaluate the quality of each genome and its Prokka annotation using the enterobacterales lineage database. Any genome with a "completeness" (C) score of < 95% or a "duplicated single-copy" (D) score of > 5% was removed from the pool. In order to ensure the genomes in the pool were identified correctly as *P. agglomerans*, a pair-wise ANI analysis of all genomes in the pool was performed using FastANI v. 1.33 [26]. Any genome that did not have an ANI score $\geq$ 95%, with respect to the *P. agglomerans* type strain (GCF_019048385.1), was removed from the pool.

Having constructed a pool of 35 high-quality, complete, consistently-reannotated *P. agglomerans* genomes (Table S17), each genome's sequence (.fna), proteome (.faa), and annotation (.gff) file was split using custom scripts into separate files for the chromosome and each individual plasmid. The MASH [35] distance between each pair of replicons was computed using Mashtree v. 1.4.5 [36]. In addition to producing a table of the MASH distance between each part, Mashtree also produces a tree in Newick format. The mashtree_bootstrap.pl script with 100 reps was used to assign bootstrap values to the branches of the tree.

Using the MASH distances, replicons were assigned to plasmid clusters using the following method: a graph was constructed in which each replicon was represented by a node, and edges were created between pairs of nodes when the MASH distance was less than or equal to an arbitrary cutoff. For this analysis, we used a MASH distance of 0.05, which very loosely corresponds to a 95% identity. A custom script was used to compute connected components within this graph; each connected component represents a cluster of plasmids with similar k-mer statistics. Hereafter, we refer to these plasmid clusters as "plasters" and this analysis as the "plaster analysis".

Some circular plasmid figures were generated using BRIG (BLAST Ring Image Generator) [38].

Combined hierarchical clustering and heatmap diagrams of MASH distances (Figures S5, S6, S7, S8, S9, and S10) were produced using R v. 4.3.1 [39] with the ComplexHeatmap v. 2.14.0 library [40].

Predictions of phage of plasmids

Phage predictions were generated with PHASTER [41] and PHAST [42].

Prediction of conjugative and mobilizable plasmids

For each of the 35 closed, high-quality *P. agglomerans* genomes, we used a custom script (gbk2gembase) to convert the annotation generated by Prokka into a proteome data file specifically formatted for MacSyFinder. Next, for each genome, we used this proteome data file as input to MacSyFinder v. 2.1.2 [43, 44]. We analyzed each proteome with the CONJScan/Plasmids v. 2.0.1 models [45]. We specified that all submodels should be considered. We also specified that all replicons were circular.

Panreplicon analysis

In order to determine whether or not the replicons in each plaster share a set of common, or core, genes, we ran Roary v. 3.13.0 [22] with the set of individual replicons (not genomes) as input. In order to produce results for every gene group across all replicons, regardless of cardinality, Roary was run with the parameter, -cd 0.0. A custom script was used to post-process the Roary output in order to compute the percentage inclusion of each gene group within each plaster.

In the full-genome context, genes in the pangenome typically are classified based on the percentage of strains found to have a copy. For instance, the *hard core* of the pangenome may be defined as the set of genes found in 99 - 100% of the analyzed genomes, the *soft_core* as genes found in 95% or more of the genomes, the *shell* as the genes found in 15% or more of the genomes, and the rest of the genes as the *cloud*. To understand the gene complement of each plaster group's replicons, we extended this definition as follows: the *hard core* of a plaster group is a set of genes found in 99 - 100% of the replicons that are members of this plaster

group; similarly, the *soft core* of a plaster group are the genes found in 95% of the members of the plaster group; the definitions of *shell* and *cloud* for plaster groups are similar.

Genetic manipulation

Spontaneous rifampicin resistant isolates were recovered by plating approximately 10^8 CFU (200-300 µl of overnight shaken cultures in LB broth or a suspension in sterile water of cells recovered from LB plate cultures) onto LB rifampicin (LB Rp) amended agar medium. Antibiotic resistance was confirmed by dilution spotting suspensions onto both LB and LB Rp media and counting colonies to confirm similar growth rates. The rifampicin resistant strains also were assessed for RSN phenotypes compared to the wildtype (WT) strains prior to use in further experiments.

The in-frame deletion of the *hvrA* (phosphoenopyruvate mutase) gene from *P. agglomerans* AR1a was conducted following the established allelic exchange protocols previously utilized for *P. ananatis* [3, 46] to exchange the *hvrA* gene with a synthesized DNA fragment (Table S1) corresponding to the *hvrA* deletion cloned into pR6KT2G using BP clonase II. The Δ*hvrA* mutant was recovered in a two-step process by selection of GmR pR6KT2G single-crossover exconjugants followed by sucrose counter-selection to recover double-crossover mutants. The *P. agglomerans hvrA* deletion mutant strains were genotyped using PCR assays and verified by PCR amplicon sequencing (Table S1).

Two KanR marked derivatives of the pCB1C plasmid in *P. agglomerans* CB1 were generated by allelic exchange using the pKNG101 vector [47] to insert the kanamycin resistance gene from pBS44 [48] either between LD072_23365 and LD072_23370 (pCB1C-Akan) or replacing pCB1C LD072_23445 and LD072_23450 (pCB1C-BKan). DNA fragments were amplified by PCR assays and then joined and cloned into pBC SK^- using NEBuilder HiFi DNA Assembly (Table S1). After sequencing confirmation, the correctly assembled fragments were sub-cloned into pKNG101 by SalI-HF and SpeI-HF restriction digest and T4 ligation. The KanR double cross-in mutants of CB1 pCB1C were recovered in a two-step process using selection of KanR SmR pKNG101 *P. agglomerans* CB1 single-crossover exconjugants, followed by sub-culturing without streptomycin selection, and screening for KanR SmS putative double-crossover mutants and

PCR genotyping (Table S1). Successful creation of *P. agglomerans* CB1 pCB1C-AKan and CB1 pCB1C-BKan was confirmed by Illumina sequencing at the Cornell Biotechnology Resource Center, as described above.

Native mobilization of the Kan marked pCB1C virulence plasmid.

P. agglomerans CB1 pCB1C-AKan or CB1 pCB1C-BKan donor strains and Rp-resistant AR8b, MMD61212-C, FC61912-B, and J22c recipient strains were suspended from fresh LB plate cultures to an OD_{600} of 0.4 in sterile water. Approximately 10^8 CFU of each donor and recipient strain were combined, and the bacterial suspensions were then pelleted by centrifugation and re-suspended in 1 mL of LB. After 19 hours of incubation with shaking at 28℃, 0.2 mL of culture was plated onto LB Kan Rp agar medium to recover potential pCB1C-AKan or pCB1C-BKan exconjugants. Oxford Nanopore sequencing was conducted as described above to confirm the presence of pC-AKan in *P. agglomerans* AR8b, MMD61212-C, and FC61912-B exconjugants.

To estimate pC-AKan conjugation efficiency, water suspensions of the *P. agglomerans* CB1 pCB1C-Akan donor and the Rp-resistant MMD61212-C were prepared as described above and combined. The population of each parent in the combined suspension was determined by dilution series spot plating onto LB Rp and LB Kan agar medium. A 0.2 mL aliquot of the combined suspension was spread onto water agar medium and incubated overnight at 28℃. Bacteria were recovered from the water agar in 0.4 mL of sterile water. The population of each parent after incubation on water agar was estimated by dilution series spot plating onto LB Rp and LB Kan agar media and 0.3 mL was plated onto LB Kan Rp agar medium to recover exconjugants. Putative exconjugants were screened by PCR assay to confirm the presence of pCB1C-AKan (Table S1).

Bacterial inoculum preparation

P. agglomerans strains selected for phenotyping assays were streaked onto LB or LM agar media (consisting of 10 g/L tryptone, 6 g/L yeast extract, 1.193 g/L KH_2PO_4, 0.6 g/L NaCl, 0.4 g/L $MgSO_4 \cdot 7H_20$, and agar 15 g/L) containing appropriate antibiotics from -80℃ cultures and incubated overnight at 28℃. To create bacterial inoculum, a single colony of each strain was

inoculated into 3 mL of LB or LM broth and incubated overnight with agitation at 200 rpm. Following incubation, the bacterial cells were harvested by centrifugation and the cell pellets were resuspended and adjusted to an OD_{600} of 0.2-0.3 using sterile 1X PBS buffer. Alternatively, bacteria were pulled from LB agar plates grown overnight at 28°C using sterile cotton-tipped applicators, and inoculated into autoclaved, high-purity water and adjusted to an OD_{600} of 0.2 - 0.3.

Disease assays

Red onions were procured through the UGA dining services or purchased from local grocery stores in Athens, GA or Ithaca, NY. RSN assays were conducted for all *Pantoea* strains using a previously described protocol [49] with slight modifications. The onion scales were stab inoculated using either a sterile pipette tip or a toothpick that had been dipped in bacterial inoculum. Onion plants (cv. Century) were grown in the greenhouse in Athens, GA as described by [49] and inoculated according to the cut tip inoculation method for foliar necrosis assays [49].

Red onion bulb necrosis assays were conducted using a modification of the onion bulb necrosis assay procedures described in [30]. Red onions without any visible wounds or external symptoms of infection were selected for the assay. After removing the dried outer scale layers, the surface of the onion bulb was disinfected by wiping with 70% ethanol. A sterilized, full-length, dissecting needle was used to puncture the shoulder on one side of the onion bulb, approximately 45 mm towards the center of the bulb. Following this, 10 µl of bacterial inoculum was pipetted into the puncture wound. Sterile 1X PBS buffer was used as a negative control treatment. Three separate onion bulbs were used per treatment. Bulbs were incubated upright at 28°C for one week, bisected vertically through the site of inoculation, and the cut surfaces was photographed.

For gain-of-function foliar disease assays with ex-conjugant strains, onion plants were grown as described by [50]. Plants were moved to a growth chamber set at 28°C with 14 hours of light/day and ambient humidity, one day prior to inoculation. Bacterial inoculum was prepared in water to an OD_{600} = 0.2, as previously described. Plants were brought into the lab, and four

plants were inoculated per treatment (strains or sterile water), and five leaves were inoculated per plant, starting with the youngest leaf > 10 cm above the soil line that was wide enough to be inoculated with a toothpick without detaching the tissue above the inoculation site. Leaves were inoculated approximately 10 cm above the soil line by stabbing the leaves with toothpicks dipped in inoculum. After inoculation, plants were returned to the growth chamber. Lesions on all leaves were measured 2, 3, 4, and 5 days after inoculation. The experiment was completed twice. Lesions from the fifth oldest inoculated leaves were not included in the analysis, as these leaves began to senesce during the experiment, and the lesions could not be distinguished from the naturally senescent tissue.

A linear mixed effects model was used to compare lesion size on day 5. The model included a fixed effect of treatment and a random effect of plant due to the fact that measurements were taken on 4 leaves per plant. Post-hoc comparisons among treatments were performed using Tukey's honestly significant difference (HSD) test. A p-value < 0.05 was considered statistically significant.

Statistical Analysis of Leaf Assay Data

For leaf assays, endpoint (5 days post inoculation) leaf lesion length was pooled from eight inoculated plants, four leaves per plant from two experiments. Lesion lengths were averaged, and standard error was plotted for error bars. In general, leaves inoculated with CB1 harboring the kanamycin-resistance cassette containing plasmid, and all of the transconjugants developed lesions over five days. None of the original strains used as recipients caused lesions in any inoculated leaves. Two of 32 water-inoculated leaves developed lesions over the course of the experiment.

Only treatments CB1 pCB1C-AKan; the transconjugants of AR8b Rp^r, MMD61212-C Rp^r, FC61912-B Rp^r (also containing pCB1C-AKan); and water were included in the model since there were no lesions (all 0 values, no variability) in other treatments. The overall effect of treatment was statistically significant.

($F_{4,155}$ = 57.5, p < 0.0001) Tukey post-hoc tests indicated that the mean lesion length for treatments CB1 pCB1C-AKan and the transconjugants were statistically significantly different than water, but there were no differences in CB1 pCB1C-AKan and the transconjugants at 5 days post inoculation

The 95% confidence intervals for CB1 pCB1C-AKan and the transconjugants at day 5 did not include 0, but the confidence interval for water did, indicating that water was not significantly different than 0 and not significantly different than leaves inoculated with the strains recipient strains lacking pCB1C-AKan (Figure 3c).

Copper tolerance assay

A copper tolerance assay was performed using 123 strains of *P. agglomerans* (81 whole genome sequenced (WGS) and 42 non-WGS strains) (Table S22), following established protocols [51, 52] with minor adjustments. Casitone Yeast Extract (CYE) agar medium, (composed of 1.7 g/L casitone, 0.35 g/L yeast extract, and 15 g/L agar) was supplemented with filter-sterilized copper sulfate ($CuSO_4 \cdot 5H_2O$) at final concentrations of 50, 100, 150, or 200 µg/ml. To prevent formation of highly mucoid colonies that can interfere with interpretation of confluent growth, glycerol was omitted from the medium. As a control treatment, CYE medium without copper sulfate was used. The bacterial inoculum was prepared as previously outlined and arrayed in 96-well microtiter plates, and 1 µl of the inoculum was spotted onto CYE agar medium in gridded square petri plates using a multichannel pipettor. Afterward, the plates were dried in a biosafety cabinet and incubated at 28°C for 2 days. Bacterial growth at each copper concentration was observed and scored for confluent growth, non-confluent growth, or no growth. The assay was run with two technical replicates and was carried out three times.

References

1. Zaid AM, Bonasera JM, Beer SV. OEM—a new medium for rapid isolation of onion-pathogenic and onion-associated bacteria. *Journal of Microbiological Methods* 2012; **91**: 520–526.

2. Stice SP, Stumpf SD, Gitaitis RD, Kvitko BH, Dutta B. *Pantoea ananatis* genetic diversity analysis reveals limited genomic diversity as well as accessory genes correlated with onion pathogenicity. *Frontiers in Microbiology* 2018; **9**: 184.

3. Shin GY, Smith A, Coutinho TA, Dutta B, Kvitko BH. Validation of species-specific PCR assays for the detection of *Pantoea ananatis*, *P. agglomerans*, *P. allii*, and *P. stewartii*. *Plant Disease* 2022; **106**: 2563–2570.

4. Brady C, Cleenwerck I, Venter S, Vancanneyt M, Swings J, Coutinho T. Phylogeny and identification of *Pantoea* species associated with plants, humans and the natural environment based on multilocus sequence analysis (MLSA). *Systematic and Applied Microbiology* 2008; **31**: 447–460.

5. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 1988; **16**: 1215–1215.

6. Asselin JE, Bonasera JM, Helmann TC, Beer SV, Stodghill PV. Complete genome sequence resources for the onion pathogen, *Pantoea ananatis* OC5a. *Phytopathology* 2021; **111**: 1885–1888.

7. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, et al. Klebsiella pneumoniae carbapenemase (KPC)-producing k. Pneumoniae at a single institution: Insights into endemicity from whole-genome sequencing. *Antimicrobial Agents and Chemotherapy* 2015; **59**: 1656–1663.

8. Andrews S. FastQC. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

10. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics* 2020; **70**.

449 11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome
450 assemblies. *Bioinformatics* 2013; **29**: 1072–1075.

451 12. Seppey M, Manni Mé, Zdobnov EM. BUSCO: Assessing genome assembly and annotation
452 completeness. *Gene Prediction* 2019; 227–245.

453 13. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
454 repeat graphs. *Nature Biotechnology* 2019; **37**: 540–546.

455 14. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome
456 sequencing [version 4; peer review: 4 approved]. *F1000Research* 2021; **8**: 2138.

457 15. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nature*
458 *Computational Science* 2021; **1**: 332–336.

459 16. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Trycycler: Consensus
460 long-read assemblies for bacterial genomes. *Genome Biology* 2021; **22**: 266.

461 17. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor.
462 *Bioinformatics* 2018; **34**: i884–i890.

463 18. Hu J, Fan J, Sun Z, Liu S. NextPolish: A fast and efficient genome polishing tool for long-
464 read assembly. *Bioinformatics* 2019; **36**: 2253–2255.

465 19. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI
466 prokaryotic genome annotation pipeline. *Nucleic Acids Research* 2016; gkw569.

467 20. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies
468 from short and long sequencing reads. *PLOS Computational Biology* 2017; **13**: e1005595.

469 21. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**:
470 2068–2069.

471 22. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid
472 large-scale prokaryote pan genome analysis. *Bioinformatics* 2015; **31**: 3691–3693.

23. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 2013; **30**: 772–780.

24. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* 2010; **5**: e9490.

25. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 2021; **49**: W293–W296.

26. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 2018; **9**: 5114.

27. Bonasera JM, Asselin JE, Beer SV. Identification of bacteria pathogenic to or associated with onion (*Allium cepa*) based on sequence differences in a portion of the conserved gyrase b gene. *Journal of Microbiological Methods* 2014; **103**: 138–143.

28. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 2010; **59**: 307–321.

29. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinf* 2009; **10**: 421.

30. Asselin JE, Bonasera JM, Beer SV. Center rot of onion (*Allium cepa*) caused by *Pantoea ananatis* requires *pepM*, a predicted phosphonate-related gene. *Mol Plant-Microbe Interact* 2018; **31**: 1291–1300.

31. Kirk A, Stavrinides J. Distribution and comparative genomic analysis of antimicrobial gene clusters found in *Pantoea*. *Frontiers in Microbiology* 2024; **15**.

32. Smits THM, Duffy B, Blom J, Ishimaru CA, Stockwell VO. Pantocin A, a peptide-derived antibiotic involved in biological control by plant-associated *Pantoea* species. *Archives of Microbiology* 2019; **201**: 713–722.

33. Williams AN, Stavrinides J. *Pantoea* Natural Product 3 is encoded by an eight-gene biosynthetic gene cluster and exhibits antimicrobial activity against multi-drug resistant *Acinetobacter baumannii* and *Pseudomonas aeruginosa*. *Microbiological Research* 2020; **234**: 126412.

34. Kamber T, Lansdell TA, Stockwell VO, Ishimaru CA, Smits THM, Duffy B. Characterization of the biosynthetic operon for the antibacterial peptide herbicolin in *Pantoea vagans* biocontrol strain C9-1 and incidence in *Pantoea* species. *Applied and Environmental Microbiology* 2012; **78**: 4412–4419.

35. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology* 2016; **17**.

36. Katz L, Griswold T, Morrison S, Caravas J, Zhang S, Bakker H, et al. Mashtree: A rapid comparison of whole genome sequence files. *Journal of Open Source Software* 2019; **4**: 1762.

37. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* 2021; **37**: 2473–2475.

38. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* 2011; **12**: 402.

39. R Core Team. R: A language and environment for statistical computing. 2023. R Foundation for Statistical Computing, Vienna, Austria.

40. Gu Z. Complex heatmap visualization. *iMeta* 2022; **1**: e43.

41. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research* 2016; **44**: W16–W21.

42. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A fast phage search tool. *Nucleic Acids Research* 2011; **39**: W347–W352.

43. Abby SS, Néron B, Ménager H, Touchon M, Rocha EPC. MacSyFinder: A program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS ONE* 2014; **9**: e110726.

44. Néron B, Denise R, Coluzzi C, Touchon M, Rocha EPC, Abby SS. MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. *Peer Community Journal* 2023; **3**: e28.

45. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. *Scientific Reports* 2016; **6**: 23080.

46. Stice SP, Thao KK, Khang CH, Baltrus DA, Dutta B, Kvitko BH. Thiosulfinate tolerance is a virulence strategy of an atypical bacterial pathogen of onion. *Current Biology* 2020; **30**: 3130–3140.e6.

47. Kaniga K, Delor I, Cornelis GR. A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria: Inactivation of the blaA gene of *Yersinia enterocolitica*. *Gene* 1991; **109**: 137–141.

48. Swingle B, Thete D, Moll M, Myers CR, Schneider DJ, Cartinhour S. Characterization of the PvdS-regulated promoter motif in *Pseudomonas syringae* pv. *tomato* DC3000 reveals regulon members and insights regarding PvdS function in other pseudomonads. *Molecular Microbiology* 2008; **68**: 871–889.

49. Shin GY, Dutta B, Kvitko B. The genetic requirements for HiVir-mediated onion necrosis by *Pantoea ananatis*, a necrotrophic plant pathogen. *Molecular Plant-Microbe Interactions* 2023; **36**: 381–391.

50. Bonasera JM, Asselin JE, Beer SV. Lactic acid bacteria cause a leaf blight and bulb decay of onion (*Allium cepa*). *Plant Disease* 2017; **101**: 29–33.

51. Sundin GW. Copper resistance in *Pseudomonas syringae* pv. *syringae* from cherry orchards and its associated transfer in vitro and in planta with a plasmid. *Phytopathology* 1989; **79**: 861.

52. Tho KE, Brisco-McCann E, Wiriyajitsomboon P, Sundin GW, Hausbeck MK. Bacteria associated with onion foliage in Michigan and their copper sensitivity. *Plant Health Progress* 2019; **20**: 170–177.