MPMI Vol. 21, No. 12, 2008, pp. 1549–1560. doi:10.1094/MPMI-21-12-1549. *e*-Xtra*

Niche-Specificity and the Variable Fraction of the *Pectobacterium* Pan-Genome

J. D. Glasner,[1] M. Marquez-Villavicencio,[2] H.-S. Kim,[2] C. E. Jahn,[2] B. Ma,[1] B. S. Biehl,[1] A. I. Rissman,[1] B. Mole,[3] X. Yi,[4] C.-H. Yang,[4] J. L. Dangl,[3,5,6,7] S. R. Grant,[5,6] N. T. Perna,[1] and A. O. Charkowski[2]

[1]Genome Center of Wisconsin and [2]Department of Plant Pathology, University of Wisconsin–Madison, Madison, WI 53706, U.S.A.; [3]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, U.S.A.; [4]Department of Biological Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53211, U.S.A.; [5]Department of Biology, [6]Curriculum in Genetics, and [7]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill

Submitted 11 April 2008. Accepted 11 August 2008.

We compare genome sequences of three closely related soft-rot pathogens that vary in host range and geographical distribution to identify genetic differences that could account for lifestyle differences. The isolates compared, *Pectobacterium atrosepticum* SCRI1043, *P. carotovorum* WPP14, and *P. brasiliensis* 1692, represent diverse lineages of the genus. *P. carotovorum* and *P. brasiliensis* genome contigs, generated by 454 pyrosequencing ordered by reference to the previously published complete circular chromosome of *P. atrosepticum* genome and each other, account for 96% of the predicted genome size. Orthologous proteins encoded by *P. carotovorum* and *P. brasiliensis* are approximately 95% identical to each other and 92% identical to *P. atrosepticum*. Multiple alignment using Mauve identified a core genome of 3.9 Mb conserved among these *Pectobacterium* spp. Each core genome is interrupted at many points by species-specific insertions or deletions (indels) that account for approximately 0.9 to 1.1 Mb. We demonstrate that the presence of a *hrpK*-like type III secretion system-dependent effector protein in *P. carotovorum* and *P. brasiliensis* and its absence from *P. atrosepticum* is insufficient to explain variability in their response to infection in a plant. Additional genes that vary among these species include those encoding peptide toxin production, enzyme production, secretion proteins, and antibiotic production, as well as differences in more general aspects of gene regulation and metabolism that may be relevant to pathogenicity.

Additional keywords: Erwinia, host range, soft rot.

Comparative genomics can reveal physiological and functional variation among bacteria that provides insight into their ability to exploit distinct ecological niches. For plant-pathogenic bacteria, key factors related to interaction with a plant host have emerged from comparisons with closely related animal-associated bacteria. For example, identification of gene clusters, or islands, that distinguish a single plant-pathogenic *Pectobacterium* genome from those of other enterobacteria that infect animals, such as *Escherichia* and *Salmonella* spp., revealed known and novel virulence factors, many of which are also found in more distantly related plant pathogens (Bell et al. 2004; Toth et al. 2006). Comparison of more closely related genomes can be particularly useful for illuminating niche adaptation, in part because the level of observed genetic variation is lower, simplifying reconstruction of both the evolutionary history and the phenotypic consequences of individual polymorphisms. Intraspecific comparisons of genomes from strains of both *Pseudomonas syringae* and *Xanthomonas campestris* with diverse host ranges revealed differences, including the type and number of type III secreted proteins (Feil et al. 2005; Joardar et al. 2005; Thieme et al. 2005). These can affect host range either by suppressing host defenses and enabling the growth of a pathogen or, conversely, they can trigger a strong resistance response blocking pathogen growth if an individual type III secreted protein is recognized by a cognate resistance protein in the host (Jones and Dangl 2006).

There are no previous intraspecific genome sequence comparisons for plant-associated enterobacteria; however, several groups of animal-associated enterobacteria have been sampled fairly extensively. Early comparisons revealed that only 40% of the total distinct protein-coding sequences are shared among the model *Escherichia coli* K12 strain, an enterohemorrhagic strain, and a uropathogenic strain (Welch et al. 2002), despite the fact that over 3.5 Mb of the 4.5- to 5.5-Mb core genome is conserved among all three. This pattern has been observed in other groups of bacteria (Kettler et al. 2007; Tettelin et al. 2005), leading to the concept of a pan-genome, or the complete collection of genes in the species, only a fraction of which are found in any given strain. Importantly, as additional genomes (lineages) of a species are added to comparisons, the fraction of genes conserved, or core genome, declines relatively slowly for most bacteria, reflecting the relative stability of core metabolic and information-processing-related gene content. In contrast, some "cosmopolitan" species, such as *E. coli* and *Streptococcus agalactiae*, have a much larger and dynamic pan-genome, with the variable fraction composed of the sum total of many distinct lineage-specific gene islands, acquired through horizontal gene transfer. Mathematical models predict that even relatively well-sampled lineages such as *E. coli* will continue to yield novel genes as more strains are sequenced (Willenbrock et al. 2007). For plant-pathogenic enterobacteria with one published genome sequence, characterization of the pan-genome is in its infancy.

The soft rot pectobacteria (previously known as *Erwinia carotovora*) are economically important plant-pathogenic entero-

Corresponding author: J. D. Glasner; Telephone: (608) 890-0171; Fax: (608) 265-9485; E-mail: jglasner@wisc.edu

Nucleotide sequence data for the assembled contigs are available in the GenBank database under accession numbers ABVX00000000 and ABVY00000000.

*The *e*-Xtra logo stands for "electronic extra" and indicates that a supplementary figure is published online.

bacteria. *Pectobacterium* spp. cause a spectrum of disease symptoms (termed wilt, soft rot, and blackleg) on a wide range of monocot and dicot host plants. These diseases are responsible for large economic losses during potato and ornamental production. *Pectobacterium* spp. have been isolated from numerous plant hosts, soil, and both surface and ground water (McCarter-Zorner et al. 1984, 1985). *Pectobacterium* spp. have also been found in association with a variety of invertebrates, ranging from fruit flies to snails (Harrison et al. 1977; Molina et al. 1974; Phillips and Kelman 1982). These invertebrates play an important role in the spread of soft rot bacteria between plant hosts and may also respond to the plant pathogen. For example, Basset and associates (2000) showed that fruit flies have an immune response when inoculated with *Pectobacterium* spp.

The genus *Pectobacterium* was recently resurrected and four *E. carotovora* subspecies were renamed as different *Pectobacterium* spp.: *Pectobacterium atrosepticum*, *P. betavasculorum*, *P. carotovorum*, and *P. wasabiae* (Gardan et al. 2003). Concurrently, a potential fifth species, originally named *E. carotovora* subsp. *brasiliensis*, was described as causing blackleg on potato in Brazil (Duarte et al. 2004). This species is phylogenetically distinct from the other four *Pectobacterium* spp. (Ma et al. 2007) and will be referred to herein as *P. brasiliensis.* The pathogens *P. carotovorum, P. atrosepticum*, *P. brasiliensis*, and *P. wasabiae* all cause disease on potato. Multiple species may cause disease in the same field and even on the same individual plant. *P. carotovorum* is found in many climates worldwide, *P. atrosepticum* is found in cool climates worldwide, and *P. brasiliensis* has only been reported in Brazil, Israel, and the United States but is likely to have a wider distribution (Ma et al. 2007). Of the potato-infecting species, *P. carotovorum*, *P. brasiliensis*, and *P. wasabiae* have all been reported to cause disease on other plant species and, thus, are broad-host-range pathogens, whereas *P. atrosepticum* appears to be limited to potato and closely related solanaceous crops. The genetic differences that limit the host range of *P. atrosepticum* compared with other *Pectobacterium* spp. are unknown.

In order to identify variation between subspecies that could provide clues to the nature of differences in host range, nutritional requirements, and species niches in the pectobacteria, we compared the genome sequences of three isolates that were phlyogentically well distributed. A phylogenetic analysis based on sequencing fragments of seven housekeeping genes from bacterial soft rot isolates collected from diverse locations around the world revealed that the *Pectobacterium* spp. could be classified into a monophyletic group distinct from the broad-host-range soft rot pathogen *Dickeya* sp. (*Erwinia chrysanthemi*) and the tree pathogen *Brenneria* sp. The pectobacteria could be divided into five clades related by monophyletic descent (Ma et al. 2007). We chose the previously sequenced strain *P. atrosepticum* SCRI1043 representing clade V, *P. carotovorum* WPP14 representing clade II, and *P. brasiliensis* 1692 representing clade I. Conserved genes from these species were, on average, 95% identical. This degree of homology allowed us to assemble multiple sequence contigs derived by 454-platform pyrosequencing (454 Life Sciences, Branford, CT, U.S.A.) and Newbler analysis of genomic DNA from *P. brasiliensis* and *P. carotovorum* into ordered genome sequences by alignment with the genome of *P. atrosepticum* SCRI1043. Although we used the *P. atrosepticum* SCRI1043 genome to establish the structure of the other two genomes, the depth of sequence economically available with pyrosequencing allowed us to identify and position genes unique to each species as well as conserved sequences.

Comparison of the three genomes revealed a common core genome (representing approximately 80% of the nucleotides in each species) dotted with islands carrying diverse sequences. The genes in unique islands were enriched for proteins of DNA replication, mostly of phage origin, and in regulatory genes. Unlike the xanthomonads and *Pseudomonas syringae*, and even the blight-causing *E. amylovora* (Triplett et al. 2006), the predicted type III secreted proteins were highly conserved within the group. Variability that could be associated with differences in host range was seen in genes for type IV secretion systems (T4SS), putative phytotoxins, taxis and motility genes, and cell surface proteins.

RESULTS AND DISCUSSION

We used 454-platform massively parallel pyrosequencing (Margulies et al. 2005) to generate draft genome sequences for *Pectobacterium carotovorum* and *P. brasiliensis*. Comparisons with a published complete genome for *P. atrosepticum* and with each other allowed us to order and orient the large number of contigs onto a scaffold based on the *P. atrosepticum* sequence, thereby constructing virtual genomes for *P. carotovorum* and *P. brasiliensis* and facilitating examination of higher order features, such as the presence, absence, or rearrangement of gene islands (Fig. 1).

Assembly and ordering contigs in the draft genomes.

Summary statistics are shown in Table 1. Three runs of the 454 GS20 instrument on a DNA preparation for *P. carotovorum* resulted in 116,320,270 bp of usable sequence, or 23.3-fold coverage of the genome assuming a final size of 5 Mb, similar to *P. atrosepticum* and most other enterobacteria. Two runs using the same instrumentation yielded 86,099,090 bp and a level of 17.2-fold coverage for *P. brasiliensis*. De novo assembly for each genome was performed using the standard Newbler assembler, resulting in 4,746,006 and 4,775,163 bp in contigs greater than 500 bp in length for *P. carotovorum* and *P. brasiliensis*, respectively. We employed two approaches to predict the order and orientation of the large number of contigs from each draft genome, Projector 2.0 (van Hijum et al. 2005) and Mauve 2.0 (Darling 2004). Projector 2.0 uses BLAST searches of strategic segments of each contig against a complete (or draft) reference genome to predict their relative locations and also outputs a list of polymerase chain reaction (PCR) primer pairs designed to link adjacent contigs. We tested 188 pairs of primers with *P. carotovorum* and confirmed 84 predicted contig linkages (data not shown). It is likely that more could be confirmed by optimizing PCR conditions.

Upon aligning the Projector 2.0 ordered contig sets for both draft genomes with the complete *P. atrosepticum* genome, we observed that Mauve detected additional contigs that could be ordered, including several cases where the *P. carotovorum* and *P. brasiliensis* contigs could be ordered relative to each other even if they were not homologous to regions of *P. atrosepticum*. Starting with the Projector 2.0 ordered set, we iteratively aligned and reordered contigs until no additional members of existing locally collinear blocks were found. After running both Projector 2.0 and Mauve, order and orientation can be predicted for a total of 273/731 *P. carotovorum* contigs and 169/1370 *P. brasiliensis* contigs. These ordered contigs account for 96.6 and 96.8% of their respective genomes. Mauve alignments for the unordered and ordered contig sets are compared in Figure 1. The final alignment of ordered contigs is available through ASAP and was used for a variety of the analyses described below.

Comparison of *P. carotovorum* draft genome with physical chromosome map.

A preliminary physical map of the *P. carotovorum* chromosome was constructed using restriction enzyme mapping,

mutagenesis, and DNA hybridization (Yap et al. 2004), and this map was compared with the assembled draft sequence. Yap and associates (2004) digested *P. carotovorum* genomic DNA with I-*Ceu*I, which recognizes a conserved sequence in bacterial 23S rRNA genes, and observed seven I-*Ceu*I fragments by pulsed field gel electrophoresis (PFGE), suggesting that *P. carotovorum*, like most *Pectobacterium* strains, encodes seven rRNA operons. Sizes of the smallest six fragments were estimated and compared with the draft genome alignment. All six fragments could be identified in the draft genome and all were roughly the same size as observed by PFGE; thus, the draft genome sequence appears to represent the majority of the genome sequence. The physical map obtained by Yap and associates (2004) also correctly represented the order of the seven I-*Ceu*I fragments in the WPP14 genome.

Gene prediction, annotation, and comparative genomics.

In order to minimize redundant effort and maximize consistency across genomes, we used a comparative approach to genome annotation. TBLASTN searches using the predicted proteins from the complete *P. atrosepticum* genome as queries against the draft genomes were used to identify and annotate the boundaries of all intact open reading frames (ORF). This analysis also revealed many gene fragments that would be annotated as pseudogenes if they occurred in complete genomes; however, we expect that, in these draft genomes, most of these disrupted reading frames are caused by sequencing errors arising from both known (Margulies et al. 2005) and potentially uncharacterized sources of error associated with pyrosequencing or assembly. PCR and resequencing using traditional Sanger sequencing chemistry confirmed that 10 of 10 examples were sequencing errors (data not shown). We filtered Glimmer 2 results to remove predicted ORF that overlapped with the set inherited from *P. atrosepticum* in order to identify lineage-specific ORF. Orthologs were predicted using pairwise reciprocal BLASTP searches retaining only the best hits that include greater than 60% of both predicted proteins and show greater than 55% identity. Proteins with a second match that had an *E* value within one order of magnitude were labeled homologs rather than orthologs. Additionally, shorter (30 to 60% aligned) matches with comparable levels of sequence identity to orthologs were labeled homologs. This allowed us to accommodate gene fragments. Annotations for *P. carotovorum* and *P. brasiliensis* genes with *P. carotovorum* orthologs were propagated directly without further manual review. Annotations for lineage-specific genes were generated using the

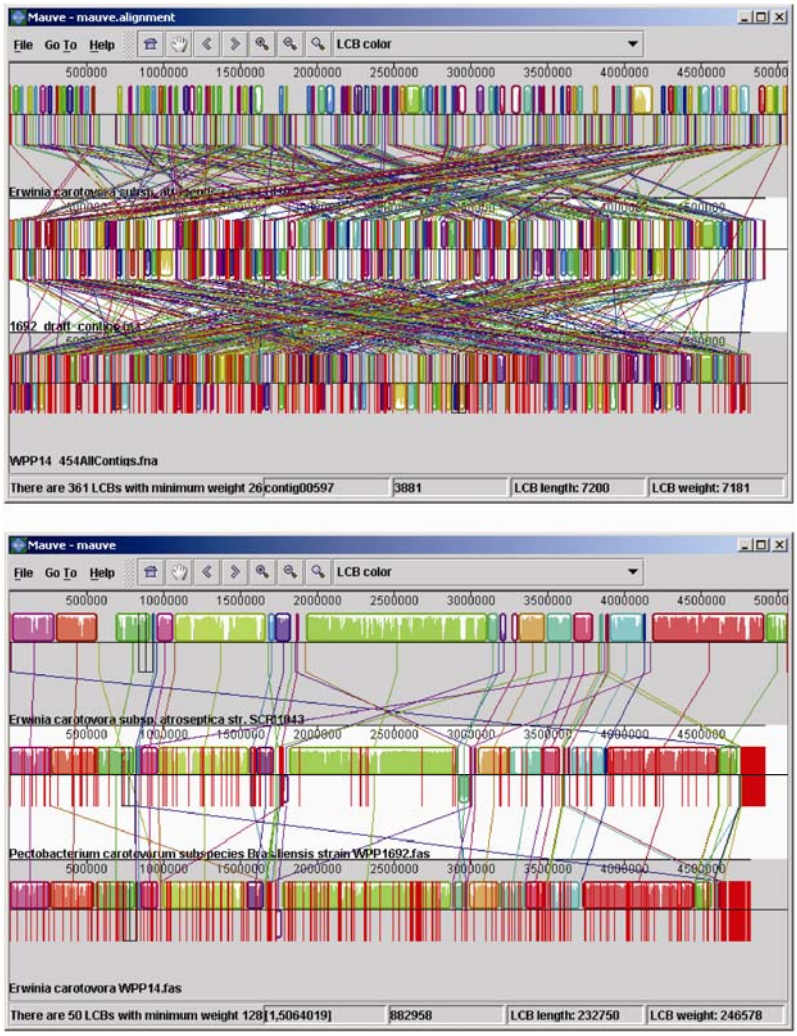


Fig. 1. Mauve alignments of the *Pectobacterium atrosepticum* genome and the draft genomes of *P. brasiliensis* and *P. carotovorum* before (top image) and after (bottom image) contig reordering. Each alignment has three panels, one for each genome (*P. atrosepticum*, *P. brasiliensis*, and *P. carotovorum*), composed of colored segments corresponding to the boundaries of locally collinear blocks with lines connecting the center of homologous blocks in each genome. Vertical red lines in the *P. brasiliensis* and *P. carotovorum* panels indicate contig boundaries. A significant number of small, unordered lineage-specific contigs appear as a dense red region at the end of the *P. brasiliensis* and *P. carotovorum* genome panels. The reduction in the number of locally collinear blocks from 361 in the top alignment to 50 in the bottom alignment and increase in average block length is clearly visible.

RAST server. Genes found exclusively in *P. carotovorum* or *P. brasiliensis* and genes found in both but absent from *P. atrosepticum* that also had a significant match (*E* value > 0.00001) to a potentially informative (did not contain "unknown," "hypothetical," "phage," "orf," or "transposase") GenPept entry were flagged for manual review by human experts. All annotations, automated and manual, and orthologs relationships were managed and distributed using ASAP (Glasner et al. 2006). The assembled contigs were also deposited in GenBank.

Alignment of the three *Pectobacterium* genomes using Mauve (Figs. 1 and 2) revealed that approximately 77% of the complete *P. atrosepticum* chromosome is present in *P. brasiliensis* and *P. carotovorum*. This is comparable with the fraction of core genome relative to the total pan-genome in the previous three *E. coli* strain comparisons (Welch et al. 2002); however, the variable fraction of the pan-genome is more evenly distributed among the three *Pectobacterium* strains and subset genome pairs. Approximately 5.4% of the *P. brasiliensis* and *P. carotovorum* sequences match each other but are not conserved with *P. atrosepticum*. This is close to double the amount of sequence shared by *P. atrosepticum* and either *P. brasiliensis* or *P. carotovorum*, but not both, supporting the closer relationship of *P. brasiliensis* and *P. carotovorum* observed by Ma and associates (2007). This estimate of the size of the core genome based on genome alignment is more reliable than what can be achieved by counting BLASTP-predicted homologs, in part because it includes intergenic regions and in part because it is more robust with regard to pseudogenes and sequencing errors. An extreme example is a region encoding two large (>7,000 amino acids each) predicted components of a nonribosomal synthase in *P. atrosepticum*. No homologous proteins are detected in either *P. brasiliensis* or *P. carotovorum*, even though DNA sequences homologous to almost the entire 42-kb region are present in both draft genomes. The large number of relatively small contigs found in this region of the draft genomes suggests that this type of highly repetitive gene poses assembly problems with short pyrosequencing reads. Keeping this caveat in mind, at least 3,251 of 4,492 predicted *P. atrosepticum* genes have orthologs (or close homologs) in all three species, and an additional 177 and 192 genes have matches in *P. brasiliensis* or *P. carotovorum*, respectively. Likely sequencing errors are expected to have an even greater impact on inference of orthologs between the two draft genomes, where we detect 190 to 195 conserved genes. Average amino acid identity between *P. carotovorum* and *P. brasiliensis* is higher (95%) than observed between orthologs of either *P. carotovorum* or *P. brasiliensis* and *P. atrosepticum* (92%), underscoring the closer relationship between these two lineages. We strongly suggest that users consult the multiple alignment if they are interested in conservation of a particular gene or region.

Pectobacterium islands and lineage-specific islands.

Bell and associates (2004) described 17 islands in the *P. atrosepticum* genome absent from other enterobacteria. We examined the extent to which these islands are conserved among the three *Pectobacterium* spp. (Fig. 3). We also defined islands that differentiate these three genomes using the Mauve alignment. For the two draft genomes, we allowed islands to span multiple ordered contigs. Frequently, small islands can be gathered into larger "variable regions" that correspond to a likely single evolutionary event, such as acquisition of a prophage, or a cluster of genes related to a common biological process, such as O-antigen synthesis. Analysis of the functional classes (Gene Ontology [GO] terms) of genes represented in gene islands specific to one or two genomes (Fig. 4) revealed that the predicted function for genes in indels were in similar proportions to those in the core genome with, the exceptions of genes for DNA replication and genes for transcriptional regulation. The abundance of genes for phage in the indels accounts for the overrepresentation of genes for DNA replication. The abundance of unique transcriptional regulatory genes could reflect species adaptation to specific environments.

Manual review of the computational predictions and visual inspection using the Mauve alignment viewer produced Supplementary Figure 1, which describes the 42 variable regions (VR001 to VR042). This list includes both regions with a computational predicted island longer than 10 kb and clusters of shorter islands that collectively span a region greater than 10 kb in at least one of the *Pectobacterium* spp. Genes at the end of each island are listed, as well as putative functions for the genes encoded in the island. The *P. atrosepticum* horizontally acquired islands (HAI) identified by Bell and associates (2004)

Table 1. Summary statistics for three sequenced *Pectobacterium* genomes[a]

Genome	*P. atrosepticum* SCRI1043	*P. brasiliensis* 1692	*P. carotovorum* WPP14
Sequencing method	Dye terminator chemistry on ABI3700 automated sequencers	Two plates of 454 GS20 sequencing	Three plates of 454 GS20 sequencing
Assembly method	assembled using PHRAP and GAP4 for finishing	Newbler (version 1.0.53.12)	Newbler (version 01.0.51.02)
Total number of reactions	54,600 paired-end reads (two small insert libraries); 25,800 paired-end reads (medium insert library); 500 paired-end reads (large insert library)	804,834	1,080,725
Number of reads assembled	80,900	782,719	1,057,457
Approximate read length	500	110	110
Total number of bases sequenced	40,450,000	86,099,090	116,320,270
Finished sequence	Final closed sequence composed of 106,500 reads; average 10.2-fold coverage	Unclosed sequence with approximately 17.2-fold coverage	Unclosed sequence with approximately 23.3-fold coverage
Total number of contigs	1	1,370	731
Total base pairs	5,064,019	4,918,574	4,823,187
Number of large contigs (≈>500 bp)	1	143	272
Total bp in large contigs	5,064,019	4,775,163	4,746,006
Average contig size	na	34,353	17,775
N50 contig size	na	75,304	38,023
Largest contig size	na	350,808	102,016
Number of predicted proteins	4,492	4,387	4,245
Number of pseudogenes identified	21	nd	150
Reported	Bell et al. 2004	This study	This study

[a] Abbreviations: na = not available and nd = not determined.

are indicated in parentheses. The variable region (VR_000) designations provide an identifier for a polymorphic region flanked by conserved chromosomes regardless of whether or not there are islands (homologous or nonhomologous) at that location. More details about the content and biological significance of these regions are described below. *P. atrosepticum* ORF are referred to using the designation given in Bell and associates (2004) and ORF in the draft genome sequences are indicated by their ASAP feature ID number. When orthologous ORF are present in all three genomes, only the *P. atrosepticum* ORF number is used.

Many previously defined *P. atrosepticum* HAI are not conserved in *P. carotovorum* or *P. brasiliensis*.

The results presented in the previous section demonstrate that, although the HAI differentiate *P. atrosepticum* from other closely related animal-associated enterobacterial genomes (Bell et al. 2004), a substantial number of genes in those regions are not essential determinants of plant pathogenicity or even the specific soft-rot phenotypes associated with *Pectobacterium* spp. as a group. Some of the HAI identified previously in *P. atrosepticum* are entirely absent from both *P. brasiliensis* and *P. carotovorum*, including HAI2, HAI3, HAI4, and HAI9. Although many of these are phage-related, HAI2 includes genes for a type IV pilus as well as the *cfa* cluster (PA0511-0614). For other HAI, parts are conserved in one or the other new genomes but not both. Our multiple alignment suggests that several have undergone rearrangements in one or more genomes. Large parts of HAI14, including the nitrogen fixation genes, are present only in *P. atrosepticum*. Both *P. atrosepticum* and *P. brasiliensis* encode metabolic genes and transporters at the same relative chromosomal position but also a phage, suggesting that this location may also be a frequent site of insertion of horizontally acquired genes. The HAI7 type IV system of *P. atrosepticum* (PA1598-1679) is replaced by two different insertions in *P. carotovorum* and *P. brasiliensis*, as expected if this is a hot spot for variation; however, *P. brasiliensis* has T4SS (AED 4454-4501) genes, as part of an apparent integrated plasmid at a different site. In contrast HAI5, HAI6, HAI8, HAI15, and HAI 17 are completely or largely conserved among all three genomes. These HAI encode exoplysaccharide and the O antigen (HAI5); a nonribosomal peptide phytotoxin (HAI6); the type 3 secretion system (T3SS) and HecAB agluttinin (HAI8), the *aggA* agglutination adhesion island (HAI15); and a phage insertion (HAI17).

Host range determinants in *Pectobacterium* spp.

Soft-rot pathogens are notorious as broad-host-range pathogens capable of decaying a broad range of monocot and dicot plants. *P. atrosepticum* is notable because it is a well-characterized narrow-host-range pathogen, found widely on potato and rarely on other solanaceous crops. In other bacterial plant pathogens, such as *Xanthomonas* and *Pseudomonas* spp., where individual strains often can infect only a few plant hosts, pathogenesis is often restricted by recognition of T3SS effectors by host plants. Thus, in these systems, in at least some cases, host range limitation is due to host recognition of

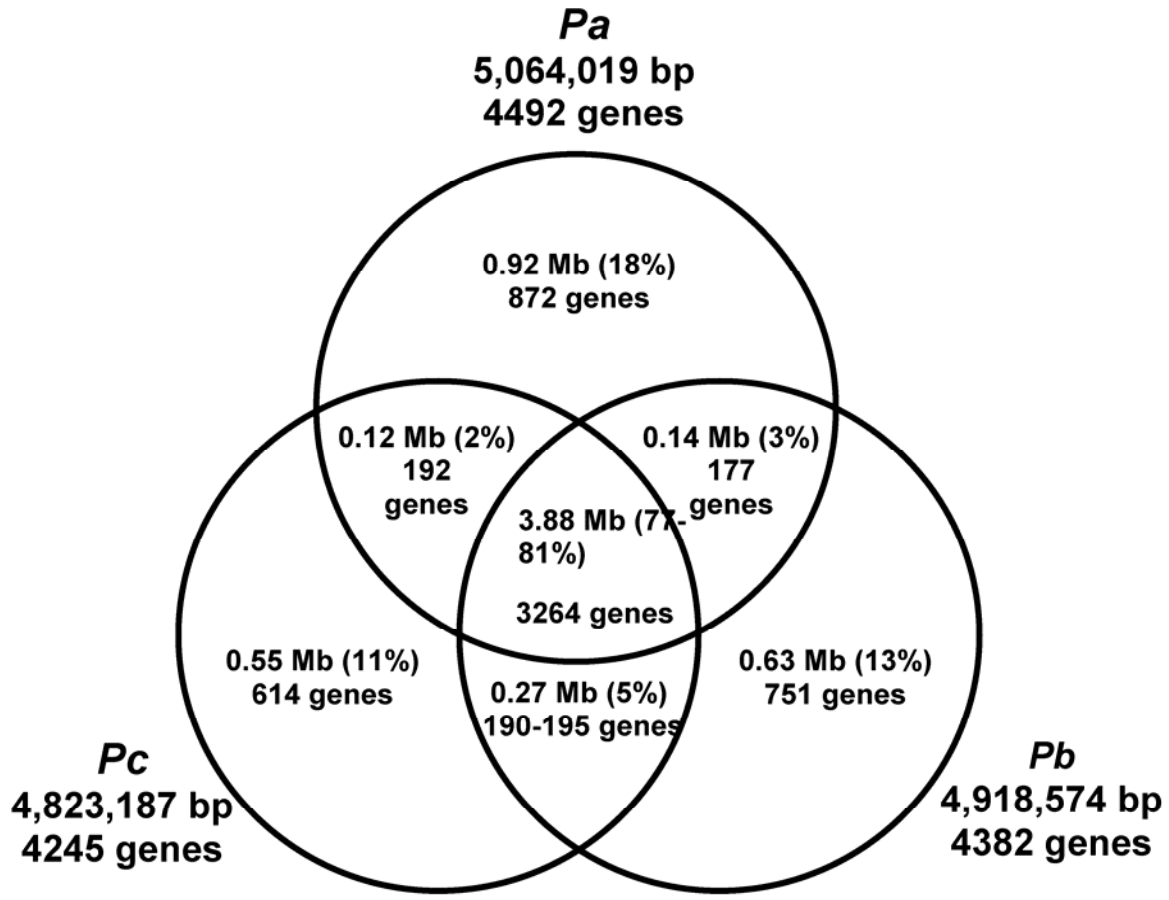


Fig. 2. Venn diagram illustrating the total amounts of shared and lineage-specific nucleotide sequence observed in the Mauve multiple alignment and predicted orthologous genes. Gene fragments resulting from likely sequence errors in the two draft genomes leads to gene count differences between species even within orthologous regions. For simplicity, the gene count for *Pectobacterium atrosepticum* (*Pa*) is shown for all shared regions except the comparison between *P. carotovorum* (*Pc*)and *P. brasiliensis* (*Pb*), where counts are shown for both species.

genes encoded by the pathogen and not because the pathogen is lacking genes required to infect a particular host species.

The draft genome sequences of *P. brasiliensis* and *P. carotovorum* provide some insight into the host range limitations of *P. atrosepticum*. Unlike *Pseudomonas* and *Xanthomonas* pathogens, the *Pectobacterium* spp. appear to encode few T3SS effector proteins; therefore, it is unlikely that these proteins are limiting the host range of *P. atrosepticum* to potato. In fact, the

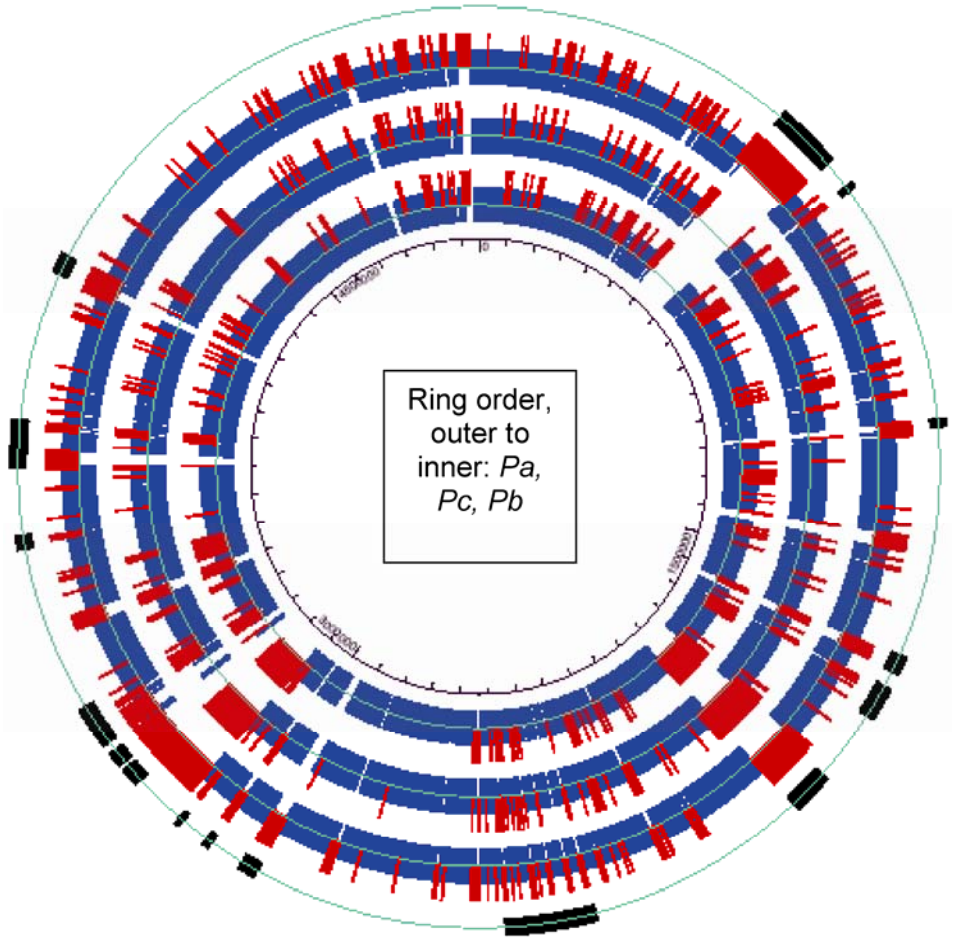


Fig. 3. Circular diagram comparing three *Pectobacterium* genomes. This plot is shown using the coordinate system of the complete *Pectobacterium atrosepticum* genome. The outer ring illustrates the position of HAI (black boxes) previously reported by Bell and associates (2004). The next ring represents the *Pectobacterium atrosepticum* (*Pa*) genome with blue corresponding to regions conserved across all three genomes and red corresponding to regions present in only one or two of the three genomes. The next ring represents the *P. carotovorum* (*Pc*) genome and the innermost ring is the *P. brasiliensis* (*Pb*) genome.

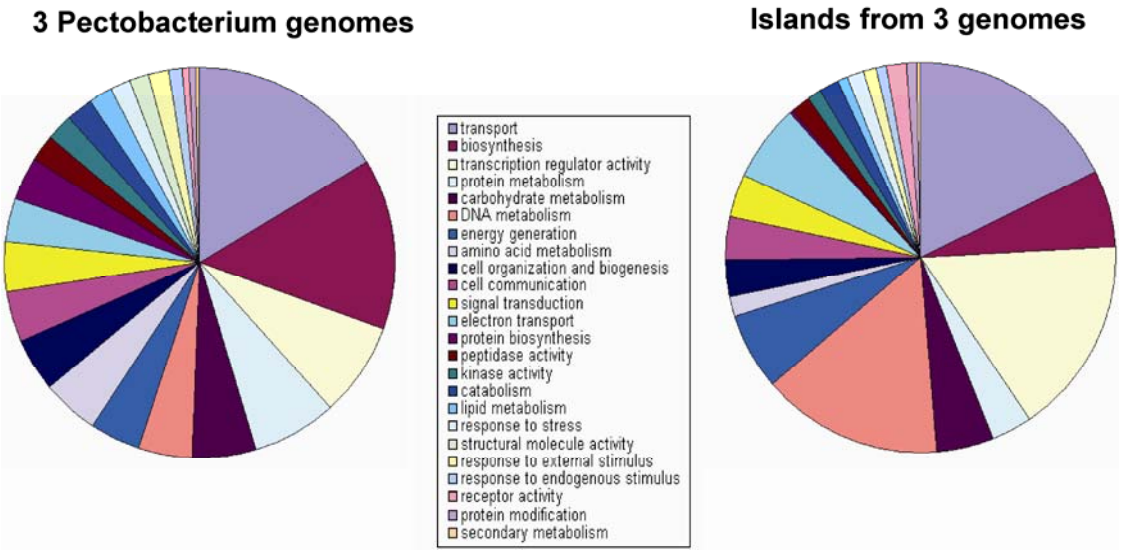


Fig. 4. Pie charts comparing Gene Ontology (GO) categories between genes found in all *Pectobacterium* genomes and genes unique to one of the three genomes. These pie charts show the distribution of genes in GO term categories. The chart on the left shows the distribution for all genes in all three *Pectobacterium* genomes and the chart on the right shows the distribution for genes unique in one of the three *Pectobacterium* spp. The key in the middle shows the color code for the GO term categories shown in the pie charts.

P. atrosepticum T3SS is even more restricted than *P. brasiliensis* and *P. carotovorum* (see below). The most striking difference among the three species is that *P. atrosepticum* encodes genes for production of a putative phytotoxin that are lacking from the other two species. The action or recognition of this toxin by other plant species could restrict the host range of *P. atrosepticum* in a manner similar to that seen with many toxin-producing fungal pathogens. The genomes of the broad-host-range *P. brasiliensis* and *P. carotovorum* pathogens appear to encode more plant cell-wall-degrading enzymes, an additional polyketide and peptide synthetase, and several large genes of unknown function, all of which could contribute to pathogenesis on a wide range of plant species. These draft genome sequences do not clearly show whether the limitation of *P. atrosepticum* to potato is due to this pathogen lacking mechanisms required for pathogenesis on other species or due to *P. atrosepticum* inducing defenses of other species; however, the draft genomes do provide a number of high-priority targets for answering this fundamental question about soft-rot pathogenesis.

Plant cell wall degradation.

Three *Pectobacterium* genomes encode nearly identical type II secretion systems (T2SS) and similar sets of plant cell-wall-degrading enzymes. *Pectobacterium* spp. wilt and rot symptoms are caused by enzymes, including pectate lyases, polygalacturonases, cellulases, and a rhamnogalacturonase, which degrade the structural components of the plant cell wall (Toth and Birch 2005). Multiple layers of regulation control synthesis of plant cell-wall-degrading enzymes in soft-rot pathogens, and the known regulators, including KdgR, ExpRI, RexZ, Crp, and H-NS, are conserved among the three *Pectobacterium* spp. However, there are also important differences in regulators among the three strains, with *P. atrosepticum* encoding two tandem copies of *pecT* and lacking an AraC-family transcriptional regulator likely to regulate the pectin methyl esterase gene *pmeB*. In addition, PecSM and Pir, key regulators found in *Dickeya* spp. (Nomura et al. 1998; Praillet et al. 1997), are not present in *Pectobacterium* spp.

Most of the plant cell-wall-degrading enzymes, as well as Svx, a protein of unknown function homologous to *X. campestris* AvrXca, are secreted through the T2SS (Corbett et al. 2005). All three *Pectobacterium* spp. encode homologous T2SS gene clusters in the same locus and all three also encode a pectate lyase and a polygalacturonase adjacent to the T2SS gene cluster. All three species also encode orthologous plant cell-wall-degrading enzymes, including 10 pectate lyases, 1 pectin lyase, 4 polygalacturonases, 2 cellulases, and 1 rhamnogalacturonase (Table 2). The soft-rot and stem-rot symptoms caused by these three species are similar and the conservation of the plant cell wall enzymes is likely to account for much of the similarity in symptoms.

Both *P. carotovorum* and *P. brasiliensis* encode putative cell-wall-degrading enzymes that are not present in *P. atrosepticum*. For example, both *P. carotovorum* and *P. brasiliensis* have an indel that consists of a GntR regulatory protein homolog (ADT-0003398; AED-0003444), a putative permease that could import the digested polymer (ADT-0002526; AED-0001909), a gene encoding a putative polysaccharide deacetylase (ADT-0003396; AED-0003442), and an Asp/Glu racemase (ADT-0003397; AED-0003443) which could degrade a host polymer. In addition, *P. carotovorum* encodes a cluster of enzyme homologs most closely related to *Clostridium* genes that may also play a role in plant cell wall degradation, including a glycoside hydrolase and a xylan β-1,4-xylosidase (ADT-0002737-8). Pectolytic *Clostridium* spp. are often present in decaying root and tuber crops along with *Pectobacterium* spp.; thus, it is not surprising that there would be evidence of horizontal gene transfer between these genera. In addition to the two genes mentioned above, 11 additional *P. carotovorum* genes and two *P. brasiliensis* genes are most similar to *Clostridium* genes.

Metalloproteases contribute to *Pectobacterium* spp. virulence and multiple *Pectobacterium* gene islands encode novel proteases.

Pectobacterium proteases also contribute to plant cell wall degradation (Marits et al. 1999) (Table 2), and numerous proteases not found in related animal pathogens are present in *Pectobacterium* spp. Those proteases that have been experimentally examined are secreted via a type 1 secretion system (T1SS) (Delepelaire and Wandersman 1991; Wandersman et al. 1990). Many of the *Pectobacterium* proteases are also found in *Dickeya* spp. and a few are present in *Pseudomonas syringae*, suggesting that their importance in virulence may be underestimated. Plants defend against proteases by production of enzymes inhibitors (Quilis et al. 2007); the presence of numerous proteases in gene islands and the likely co-evolution of microbial peptidases and plant inhibitors are reminiscent of the co-evolution of type III effector proteins and plant disease resistance genes.

Suppression of plant defenses—the T3SS and phytotoxins.

All three species encode homologous T3SS, but *Pectobacterium atrosepticum* lacks hrpK. In contrast to hemibiotrophic plant pathogens such as *Pseudomonas syringae* and *X. campestris*, the T3SS of pectobacteria does not appear to be essential for growth on potato, because *Pectobacterium wasabiae* does not encode a T3SS but can cause disease on potato (Kim, H. S., Perna, N. T., Ma, B., and Charkowski, A. O., *unpublished data*). However, the T3SS does contribute to virulence of other *Pectobacterium* spp. (Holeva et al. 2004; Rantakari et al. 2001). Mutation of genes required for the secretion apparatus or conserved effector proteins from *P. atrosepticum* 1039 leads to a reduction in virulence on potato (Holeva et al. 2004). The T3SS of the three sequenced *Pectobacterium* spp. are homologous and in the same locus. All genes known to be required for functional T3SS are present in all three species; however, only two of the three, *P. carotovorum* and *P. brasiliensis*, elicit a type III-dependent hypersensitive response (HR) on tobacco plants, which indicates that resistance genes in tobacco recognize and respond to type III effector proteins delivered from *Pectobacterium* spp.

We examined the three genomes for differences in putative T3SS effectors to account for this difference. All three *Pectobacterium* spp. encode homologs to T3SS-secreted proteins identified in *P. atrosepticum*. Both *P. carotovorum* and *P. brasiliensis*, but not *P. atrosepticum*, encode HrpK, a T3SS-secreted protein that aids in translocation of effectors across the plant cell wall in *Pseudomonas syringae* (Petnicki-Ocwieja et al. 2005). Mutation of *hrpK* in *Pectobacterium carotovorum* WPP14 did not eliminate the ability of this strain to elicit the HR, and expression of the *P. carotovorum* WPP14 *hrpK* gene from a plasmid in *P. atrosepticum* did not confer HR elicitation onto *P. atrosepticum* (not shown). Thus, the lack of HR elicitation by *P. atrosepticum* SCRI1043 is not due to lack of *hrpK*.

Other secretion systems.

Plant-pathogenic bacteria encode numerous secretion systems that contribute to virulence in addition to the T1SS, T2SS, and T3SS. The T4SS, which is required for plasmid conjugation and which, like the T3SS, is capable of secreting proteins to the extracellular milieu and of translocating proteins into host cells, is present in different locations in *P. atrosepticum* (HAI17) and *P. brasiliensis* (VR028 AED4454-4501). A remnant of a T4SS is present in *P. carotovorum* (ADT-0003337).

Currently, it is not possible to predict which proteins travel the T4SS; thus, although it appears that the *P. atrosepticum* T4SS may contribute to virulence (Bell et al. 2004), whether or not proteins secreted via the T4SS are conserved among *P. atrosepticum* and *P. brasiliensis* remains unknown.

The type V secretion system (T5SS), which includes autotransporter and two-partner secretion, is the simplest of the secretion systems (Henderson et al. 2004). T5SS play important roles in pathogenicity of many bacterial pathogens, including *Dickeya* spp. where the type V-secreted HecA hemagglutinin promotes attachment to leaf surfaces (Rojas et al. 2002). Several large proteins encoded in *Pectobacterium* indels are likely to be secreted via T5SS, including serine protease, hemolysin, and hemagglutinin homologs.

Polyketide and peptide synthetases.

Secondary metabolites produced by polyketide or peptide synthetases are important fitness and virulence factors in *Pseudomonas syringae*, where these exceptionally large proteins produce toxins active against plants and microbes as well as siderophores critical for obtaining iron. The *cfa* locus (HAI2) of *Pectobacterium atrosepticum* is absent from *P. brasiliensis* and *P. carotovorum*. A *P. atrosepticum* region in HAI6 encodes genes (PA1487-1488) similar to the pore-forming phytotoxin syringomycin synthetase from *Pseudomonas syringae* (Bender et al. 1999) and is conserved among all three genomes but fragmented in the two draft genomes. In addition, *Pectobacterium carotovorum* and *P. brasiliensis* encode a polyketide or peptide synthetase system not found in *P. atrosepticum* (VR006; ADT-0002503-2509), which is most similar to ones from distantly related gram-positive bacteria and cyanobacteria, including several that produce toxins. This variability in secondary metabolite production capabilities suggests differences in how the *Pectobacterium* spp. interact with either their plant hosts or competing microbes, and calls for further characterization of the synthesis products, analysis of the distribution of these systems among *Pectobacterium* isolates, and cross-species growth inhibition assays.

Table 2. *Pectobacterium* plant cell-wall-degrading enzymes[a]

Enzyme	*Pectobacterium atrosepticum*	*P. brasiliensis*	*P. carotovorum*
Cellulases			
	ABL-0062482	ADT-0000571	AED-0001316
bcsZ	ABL-0064711	ADT-0001191	AED-0002581
Pectate and pectin lyases			
pelA	ECA 4067	AED-0002421	ADT-0004257
pelB	ECA 4068	AED-0003514	ADT-0004256
pelC	ECA 4069	AED-0002430	ADT-0004255
pelI	ECA 1094	AED-0002063	ADT-0001495
pelW	ECA 2402	AED-0001427	ADT-0002894/5
pelX	ECA 4510	AED-0002687	ADT-0001203
pelZ	ECA 4070	AED-0002419	ADT-0002215
	ECA 3112	AED-0001793	ADT-0000798
	ECA 2135	AED-0001225	ADT-0001743
	ECA 2553	AED-0003665	ADT-0003928
	Not present	AED-0002413	Not present
pnl	ECA 1499	AED-0000858	ADT-0001600
Polygalacturonases			
pehA	ECA 1095	AED-0002061	ADT-0000261
pehK	ECA 3552	AED-0003871	ADT-0003117
pehN	ECA 1190	AED-0000675	ADT-0000300
pehX	ECA 3111	AED-0001794	ADT-0000797
peh fragment	ECA 0663	Not present	Not present
Proteases			
	ECA 0879	AED-0000493	ADT-0001419
	ECA 1450	AED-0003719	ADT-0003664
	ECA 2007	AED-0001126	ADT-0001719
	ECA 2785	AED-0004128	ADT-0004008
	ECA 3211	AED-0001839	ADT-0001994
	ECA 2160	AED-0001247	ADT-0001749
	ECA 2074	AED-0003289	ADT-0002858
	ECA 1988	AED-0001062	ADT-0000521
	ECA 1290	AED-0000730	ADT-0001454
	ECA 4192	AED-0002472	ADT-0001132
	ECA 4397	AED-0001980	ADT-0003103
	ECA 2771	AED-0003468	ADT-0003994
	ECA 0388	AED-0000276	ADT-0001318
	ECA 0134	AED-0004146	ADT-0003438
	ECA 0133	AED-0000078	ADT-0001251
	ECA 2802	AED-0001639	ADT-0000726
	ECA 2964	AED-0001716	Not present
	Not present	AED-0003397	Not present
	Not present	AED-0004092	Not present
	ECA 1004	Not present	ADT-0003580
	Not present	AED-0003242	ADT-0002516
	Not present	Not present	ADT-0004360
	Not present	AED-2871	ADT-3651
	ECA 0980	Not present	Not present
	Not present	AED-0004079	Not present
	Not present	Not present	ADT-2640
	Not present	Not present	ADT-3526 and ADT-2470

[a] *Pectobacterium* protease homologs not also present in *Escherichia coli* are included as potential plant cell-wall-degrading enzymes.

Motility and taxis.

The *Pectobacterium* spp. are all capable of swimming and swarming and the entire flagellar gene cluster is encoded in one locus, similar to that of *Dickeya dadantii* and *Yersinia* spp. but differing from *Escherichia coli* and *Salmonella enterica*. *Pectobacterium* spp. encode two flagellin homologs. In all three *Pectobacterium* spp., one of the flagellin genes is located in the midst of the flagella gene cluster while the second flagellin is encoded elsewhere in a locus that contains numerous indels and rearrangements. All three flagellin genes encode a flg22 peptide sequence, suggesting that *Pectobacterium* spp. could activate Fls2-mediated defenses in *Arabidopsis* and other host plants able to recognize this particular flagellin peptide.

One hallmark that sets plant pathogens apart from animal pathogens is the large number of taxis proteins encoded in plant pathogens. Related enteric animal pathogens encode 5 to 12 methyl-accepting chemotaxis (MCP) receptors whereas, in contrast, all of the soft-rot pathogens encode over 30 MCP receptors, with *P. atrosepticum*, *P. carotovorum*, and *P. brasiliensis* encoding 36, 39, and 34 taxis receptors, respectively. Taxis receptors function as heterotrimers of homodimers (Parkinson et al. 2005). Thus, the five receptors commonly found in *E. coli* could form up to 35 different combinations of heterotrimers, although some combinations are more likely than others due to an approximately 10-fold higher concentration of two of the receptors compared with the other three. The high number of receptors found in plant pathogens means that thousands of combinations of heterotrimers are possible.

Invertebrate interactions.

Many enterobacteria have intimate interactions with insects and, since the early part of the 1900s, insects have been suspected of spreading soft-rot disease. For example, *Pectobacterium* spp. have been cultured from corn maggot eggs and mouth parts (Gnanamanickam 2006), and transgenic plants with resistance to corn maggots also suffer less from soft-rot pathogens. The related pathogen *D. dadantii* 3937 is notable in that it encodes four *Bacillus thuringiensis* toxin homologs, at least one of which is active against aphids (Grenier et al. 2006). The genetics of interactions between *Pectobacterium* spp. and invertebrates have only recently been examined. A single locus, which includes Evf and its regulator, Evr, improves survival of *Pectobacterium* spp. in fruit fly guts (Basset et al. 2000; Muniz et al. 2007). None of the sequenced *Pectobacterium* strains encode either a *Bacillus* toxin homolog or an Evf. *P. atrosepticum*, *P. brasiliensis*, and *P. carotovorum* all encode the regulator Evr although, in *P. atrosepticum*, *evr* is a pseudogene. *P. atrosepticum* may have an additional mechanism for invertebrate interactions because it encodes a cluster of genes homologous to the hemin storage proteins required by *Yersinia* spp. for transmission by fleas (Hinnebusch et al. 1996).

Competition with other microbes—antibiotics and bacteriocins.

Pectobacterium spp. differ in the mechanisms they use to compete with other microbes in the environment. Some strains of *Pectobacterium* produce the carbapenem antibiotic 1-carbapen-2-em-3-carboxylic acid (Car), an indication that the *Pectobacterium* sp. is in competition with potato endophytes and secondary invaders for resources (Parker et al. 1982). Of the three sequenced strains, the *car* genes are only found in *P. brasiliensis*. *Pectobacterium* strains also produce multiple forms of carotovoricin, a phage-tail-like bacteriocin which kills closely related strains and species (Nguyen et al. 2001, 2002). *P. atrosepticum* encodes two bacteriocin resistance proteins which are not homologous to each other, and only one of these two is also found in *P. carotovorum* and *P. brasiliensis*. Both *P. carotovorum* and *P. brasiliensis* encode one gene homologous to the *Yersinia* bacteriocin pesticin (ASAP ID ADT-0002424; AED-0003952) and a second, unrelated bacteriocin homolog is adjacent to the pesticin homolog in *P. brasiliensis* (AED-0003953). No bacteriocins are apparent in the *P. atrosepticum* genome.

Variability in nutrient acquisition and metabolism.

Pectobacterium strains vary in the carbon sources they can use (Yap et al. 2004) and numerous *Pectobacterium* indels appear to be involved in metabolite uptake or degradation, providing a genetic explanation for some of these differences. For example, *P. atrosepticum* and *P. brasiliensis* but not *P. carotovorum* encode galactonate and gluconate metabolism islands as well as a sucrose isomerase. Thus, these *P. atrosepticum* and *P. brasiliensis* strains are likely to be able to metabolize more different, common plant sugars than *P. carotovorum*. This variation suggests that competition among *Pectobacterium* strains goes beyond how well they attack plants and includes which plant nutrients they are able to metabolize.

A protein family unique to the soft-rot erwinia.

A family of large proteins ranging between 515 and 732 residues was found in all of the soft-rot enterobacterial genomes but is lacking from all other genomes in GenBank. Two members of this protein family are found in *P. atrosepticum* (ECA1185 and ECA1186) and *P. brasiliensis* (AED-0003594 and 2989), while *P. carotovorum* encodes three members (AADT-002625, 3321, and 4109) and *D. dadantii* 3937 encodes one homolog (ABF-0020188). These proteins each have a conserved F5/8 type C domain that may aid in binding galactose (Newstead et al. 2005). Other proteins with this domain bind to cell membranes; thus, these proteins could be involved in interactions with the plant cell wall or cell membrane.

Conclusions.

We characterized the pan-genome of *Pectobacterium* by comparing three genomes from strains that are well separated phylogenetically within the genus. Using 454-platform pyrosequencing to a depth of approximately 20-fold genome coverage allowed assembly into roughly 200 contigs representing approximately 95% of each draft genome. The availability of a high-quality assembled genome from a related type strain made it possible for us to align the 200 contigs of each genome into a useful chromosome assembly using the Mauve program. The genes unique to each isolate were identified and aligned, making the information most useful for genome comparison available at an affordable cost. However, without completing the genome sequences by closing gaps and resequencing regions of low coverage or ambiguous sequence, it is not possible to rule out the existence of missing or erroneous data, determine the extent of rearrangements among chromosomes, and identify extrachromosomal elements. As the use of 454 and other "next-generation" sequencing technologies increases, it is important to weigh the costs and benefits of collecting additional data such as dual-ended sequences of fragments, optical map information, or directed sequences of gaps by standard PCR or cloning methods. Currently, the costs of collecting this additional data are roughly equivalent to the costs of obtaining the initial sequence data and warrant consideration when completion of the genome is critical for comparison or interpretation of sequence data. In the case of the *Pectobacterium* genomes presented here, little was known about the pan-genomic content of members of this genus before this report, and analysis of partially completed genomes was sufficient to identify major

differences between isolates that will form the basis for further characterization to see if they explain phenotypic variation.

We found that each strain differed in approximately 11 to 18% of the genome. Regulatory genes were more abundant in the class of uniquely represented genes. This diversity in gene regulatory mechanisms may reflect adaptation to specific ecological niches where each organism must respond to different environmental stimuli such as varying temperatures and alternative stressful conditions. Not surprisingly, these regulators are often clustered with other strain-specific genes, suggesting potential targets of regulation, and many of these variable genes encode products involved in metabolic processes.

A number of genomic differences that we identified might have consequences for virulence of the organisms. We found variation in the content of plant cell-wall-degrading enzymes, the most obvious players in causing the symptoms of soft-rot disease, as well as diversity in the content of genes encoding metalloproteases, which are also important for degradation of host cell walls. Each strain has unique genes for production of phytoxins, another likely determinant of host range. Differences in the secretion abilities of these organisms are suggested by variation in the presence and location of type IV (and other) secretion systems among the genomes. Each genome had a large but not fully overlapping collection of genes for motility and chemotaxis. Genes for surface proteins such as O-antigens, pili, and adhesins showed significant variability among genomes for factors that could either suppress or trigger host defenses. A large fraction of genes that differ between these organisms are of unknown function, including a family of proteins that appears unique to the soft rot pathogens. Figuring out which of the factors identified in this study are key determinants of the phenotypic differences between these pathogens will require sequencing of additional genomes and experimental characterization of gene functions, particularly sensitive assays for even subtle roles in virulence. The *P. atrosepticum*, *P. brasiliensis*, and *P. carotovorum* isolates examined in this report represent relative extremes with respect to the phylogenetic diversity within this genus, and their genomes vary considerably in genomic content. Resolving which changes are responsible for niche specialization is challenging given the magnitude of genomic differences. Sequences of a few additional members of the genus, particularly isolates more closely related to each of the groups already sampled or isolates with similar phenotypes, would likely help pinpoint a smaller number of important factors.

MATERIALS AND METHODS

Genomic analysis of *P. carotovorum* and *P. brasiliensis*.

P. carotovorum strain WPP14 (Yap et al. 2004) and *P. brasiliensis* strain 1692 (Duarte et al. 2004) were streaked from frozen cultures onto Luria-Bertani (LB) plates. Genomic DNA was isolated (Sambrook and Russell 2001) from cultures of *P. carotovorum* and *P. brasiliensis* grown in LB broth in shaking incubators at 25 and 37°C, respectively. *P. carotovorum* was sequenced in three standard runs on a 454 GS20 by 454 Life Sciences (Branford, CT, U.S.A.) and assembled with Newbler software version 1.0.51.02. *P. brasiliensis* was sequenced in two standard runs on a GS20 by the Washington University Genome Center and assembled with Newbler software version 1.0.53.12. Contigs were reordered using Projector 2.0 (van Hijum et al. 2005) and Mauve 2.0 (Darling 2004) which was also employed for multiple genome alignment. Primer pairs for 188 gaps in the *P. carotovorum* genome sequence were designed using Projector 2.0 and used in PCR with *P. carotovorum* genomic DNA. Selected PCR products were sequenced by standard dye-deoxy sequencing on an ABI capillary sequencer.

We used a combination of TBLASTN (Altschul et al. 1997) searches with predicted proteins from *P. atrosepticum* SCRI1043 (Bell et al. 2004) and Glimmer2 (Delcher et al. 1999) to predict ORF in the *P. carotovorum* and *P. brasiliensis* genomes. Manual inspection of predicted ORF was used to remove some ORF that overlapped other ORF and small ORF (<250 bp) that lacked significant BLASTP (*E* value <0.00001) matches against the GenPept database. We used InterProScan (Quevillon 2005) to identify protein domains in the *Pectobacterium* genomes and obtained GO (Ashburner et al. 2000) terms associated with each domain from the InterPro to GO term mapping available from the GO website.

Sequence data and analyses are available for download from the ASAP database (Glasner et al. 2006).

Expression of *P. carotovorum hrpK* in *P. atrosepticum*.

We PCR amplified the *hrpK* gene from *P. carotovorum*, including a 135-bp upstream region relative to the translation start site of *hrpK* that contains a putative HrpL-binding motif, using primers P0492 (5′-taa gag tca gga gct agt gtg gcg gag ctc agg gtt-3′) and P0493 (5′-taa gct ggc gca tta gcg cga att cgg aat att g-3′). The PCR fragment was cloned into pCPP50 (Bogdanove et al. 1998), resulting in plasmid p50*hrpK*. This plasmid was electroporated into *P. atrosepticum* and a single ampicillin-resistant colony was isolated for hypersensitive response assays.

Deletion of *hrpK* from *P. carotovorum*.

Regions flanking the *P. carotovorum hrpK* gene were PCR amplified with the left primer set P0504 (5′-gtg ctg gat ccg cta ata tca tca tac-3′) and P0505 (5′-cgt act ctg cga agc ttc ccg tcc cca ttc tgc tgt tgt ca-3′) and right primer set P0506 (5′-gga agc ttc gca gag tac gat tcc caa acc gcg cta atg c-3′) and P0507 (5′-gtc tgc cgg atc cac gtt taa cga t-3′). These two PCR fragments were used as templates for crossover PCR with primers P0504 and P0507. The 2.9-kb product was cloned into pGEMT-easy (Promega Corp., Madison, WI, U.S.A.), resulting in a plasmid pTAΔ*hrpK*_ABCD. A chloramphenicol resistance and GFP cassette obtained from pTAgfp::cm was cloned into the *Hin*dIII site of pTAΔ*hrpK*_ABCD, producing pTAΔ*hrpK*::gfp::cm This plasmid was electroporated into *P. carotovorum* for allelic-exchange mutagenesis following the methods described by Ried and Collmer (1987). The *hrpK* deletion mutation was confirmed by PCR and Southern blot analysis. Methods for electroporation, restriction endonuclease digestion, PCR, cloning, and Southern blotting were performed as described by Sambrook and Russell (2001).

HR assay in *Nicotiana tabacum*.

An assay for the HR was performed essentially as described in Bauer et al. (Bauer et al. 1994). Six- to seven-week-old *Nicotiana tabacum* cv. Xanthi leaves were infiltrated with either *P. atrosepticum*, *P. atrosepticum* with p50*hrpK*, *P. carotovorum, Pc*Δ*hrpK*, or water as a negative control. Plants were visually assayed for HR elicitation after 24 h.

ACKNOWLEDGMENTS

We would like to thank the National Science Foundation Environmental Genomics Program for funding this work through award number 0412599.

LITERATURE CITED

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389-3402.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J.

M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25(1):25-29.

Basset, A., Khush, R. S., Braun, A., Gardan, L., Boccard, F., Hoffmann, J. A., and Lemaitre, B. 2000. The phytopathogenic bacteria *Erwinia carotovora* infects *Drosophila* and activates an immune response. Proc. Natl. Acad. Sci. U.S.A. 97(7):3376-3381.

Bauer, D. W., Bogdanove, A. J., Beer, S. V., and Collmer, A. 1994. *Erwinia chrysanthemi hrp* genes and their involvement in soft rot pathogenesis and elicitation of the hypersensitive response. Mol. Plant-Microbe Interact. 7(5):573-81.

Bell, K. S., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J., Holeva, M. C., Thomson, N. R., Bentley, S. D., Churcher, L. J. C., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M. A., Sanders, M., Walker, D., Whitehead, S., Salmond, G. P. C., Birch, P. R. J., Parkhill, J., and Toth, I. K. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci. U.S.A. 101(30):11105-11110.

Bender, C. L., Alarcon-Chaidez, F., and Gross, D. C. 1999. *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol. Mol. Biol. Rev. 63(2):266-292.

Bogdanove, A. J., Bauer, D. W., and Beer, S. V. 1998. *Erwinia amylovora* secretes DspE, a pathogenicity factor and functional AvrE homolog, through the Hrp (type III secretion) pathway. J. Bacteriol. 180(8):2244-2247.

Corbett, M., Virtue, S., Bell, K., Birch, P., Burr, T., Hyman, L., Lilley, K., Poock, S., Toth, I., and Salmond, G. 2005. Identification of a new quorum-sensing-controlled virulence factor in *Erwinia carotovora* subsp. *atroseptica* secreted via the type II targeting pathway. Mol. Plant-Microbe Interact. 18(4):334-342.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394-1403.

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27(23):4636-4641.

Delepelaire, P., and Wandersman, C. 1991. Characterization, localization and transmembrane organization of the three proteins PrtD, PrtE and PrtF necessary for protease secretion by the gram-negative bacterium *Erwinia chrysanthemi.* Mol. Microbiol. 5(10):2427-2434.

Duarte, V., de Boer, S. H., Ward, L. J., and de Oliveira, A. M. 2004. Characterization of atypical *Erwinia carotovora* strains causing blackleg of potato in Brazil. J. Appl. Microbiol. 96(3):535-545.

Feil, H., Feil, W. S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J. E., Lapidus, A., Detter, J. C., Land, M., Richardson, P. M., Kyrpides, N. C., Ivanova, N., and Lindow, S. E. 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. tomato DC3000. Proc. Natl. Acad. Sci. U.S.A. 102(31):11064-11069.

Gardan, L., Gouy, C., Christen, R., and Samson, R. 2003. Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. Nov. and *Pectobacterium wasabiae* sp. nov. Int. J. Syst. Evol. Microbiol. 53:381-391.

Glasner, J. D., Rusch, M., Liss, P., Plunkett, G., Cabot, E. L., 3rd, Darling, A. Anderson, B. D., Infield-Harm, P., Gilson, M. C., and Perna, N. T. 2006. ASAP: A resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Res. 34:D41-45.

Gnanamanickam, S. S. 2006. Plant-Associated Bacteria. Springer, Dordrecht, The Netherlands.

Grenier, A. M., Duport, G., Pages, S., Condemine, G., and Rahbe, Y. 2006. The phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) is a pathogen of the pea aphid. Appl. Environ. Microbiol. 72:1956-1965.

Harrison, M., Quinn, C., Ann Sells, I., and Graham, D. 1977. Waste potato dumps as sources of insects contaminated with soft rot coliform bacteria in relation to re-contamination of pathogen-free potato stocks. Potato Res. 20(1):37-52.

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. 2004. Type V protein secretion pathway: The autotransporter story. Microbiol. Mol. Biol. Rev. 68(4):692-744.

Hinnebusch, B. J., Perry, R. D., and Schwan, T. G. 1996. Role of the *Yersinia pestis* hemin storage (*hms*) locus in the transmission of plague by fleas. Science 273:367-370.

Holeva, M. C., Bell, K. S., Hyman, L. J., Avrova, A. O., Whisson, S. C., Birch, P. R. J., and Toth, I. K. 2004. Use of a pooled transposon mutation grid to demonstrate roles in disease development for *Erwinia carotovora* subsp *atroseptica* putative type III secreted effector (DspE/A) and helper (HrpN) proteins. Mol. Plant-Microbe Interact. 17(9):943-950.

Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., Daugherty, S. C., DeBoy, R., Durkin, A. S., Giglio, M. G., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S., Crabtree, J., Creasy, T., Davidsen, T., Haft, D. H., Zafar, N., Zhou, L. W., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C. M., Chatterjee, A. K., Cartinhour, S., Schneider, D. J., Mansfield, J., Collmer, A., and Buell, C. R. 2005. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. J. Bacteriol. 187(18):6488-6498.

Jones, J. D. G., and Dangl, J. L. 2006. The plant immune system. Nature 444:323-329.

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P., and Chisholm, S. W. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet. 3(12):e231.

Ma, B., Hibbing, M. E., Kim, H.-S., Reedy, R. M., Yedidia, I., Breuer, J., Breuer, J., Glasner, J. D., Perna, N. T., Kelman, A., and Charkowski, A. O. 2007. The host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Phytopathology 97:1150-1163.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J. Lohman, K. L. Lu, H. Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W. Nickerson, E., Nobile, J. R. Plant, R. Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Marits, R., Koiv, V., Laasik, E., and Mae, A. 1999. Isolation of an extracellular protease gene of *Erwinia carotovora* subsp. *carotovora* strain SCC3193 by transposon mutagenesis and the role of protease in phytopathogenicity. Microbiology 145:1959-1966.

McCarter-Zorner, N. J., Franc, G. D., Harrison, M. D., Michaud, J. E., Quinn, C. E., Sells, I. A., and Graham, D. C. 1984. Soft rot *Erwinia* in surface and underground waters in Southern Scotland and in Colorado, United States. J. Appl. Bacteriol. 57:95-105.

McCarter-Zorner, N. J., Harrison, M. D., Franc, G. D., Quinn, C. E., Sells, I. A., and Graham, D. C. 1985. Soft rot *Erwinia* bacteria in the rhizosphere of weeds and crop plants in Colorado, United States and Scotland. J. Appl. Bacteriol. 59:357-368.

Molina, J. M., Harrison, D., and Brewer, J. W. 1974. Transmission of *Erwinia carotovora* var. *atroseptica* by *Drosophila melanogaster* Meig. I. Acquisition and transmission of the bacterium. Am. Potato J. 51:245-250.

Muniz, C. A., Jaillard, D., Lemaitre, B., and Boccard, F. 2007. *Erwinia carotovora* Evf antagonizes the elimination of bacteria in the gut of *Drosophila* larvae. Cell. Microbiol. 9(1):106-119.

Newstead, S. L., Watson, J. N.. Bennet, A. J., and Taylor, G. 2005. Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. Acta Crystallogr. D. Biol. Crystallogr. 61:1483-1491.

Nguyen, H. A., Kaneko, J., and Kamio, Y. 2002. Temperature-dependent production of carotovoricin Er and pectin lyase in phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. Biosci. Biotechnol. Biochem. 66(2):444-447.

Nguyen, H. A., Tomita, T., Hirota, M., Kaneko, J., Hayashi, T., and Kamio, Y. 2001. DNA inversion in the tail fiber gene alters the host range specificity of carotovoricin Er, a phage-tail-like bacteriocin of phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. J. Bacteriol. 183(21):6274-6281.

Nomura, K., Nasser, W., Kawagishi, H., and Tsuyumu, S. 1998. The *pir* gene of *Erwinia chrysanthemi* EC16 regulates hyperinduction of pectate lyase virulence genes in response to plant signals. Proc. Natl. Acad. Sci. U.S.A. 95(24):14034-14039.

Parker, W. L., Rathnum, M. L., Wells, J. S., Jr., Trejo, W. H., Principe, P. A., and Sykes, R. B. 1982. SQ 27,860, a simple carbapenem produced by species of *Serratia* and *Erwinia*. J. Antibiot. (Tokyo) 35(6):653-660.

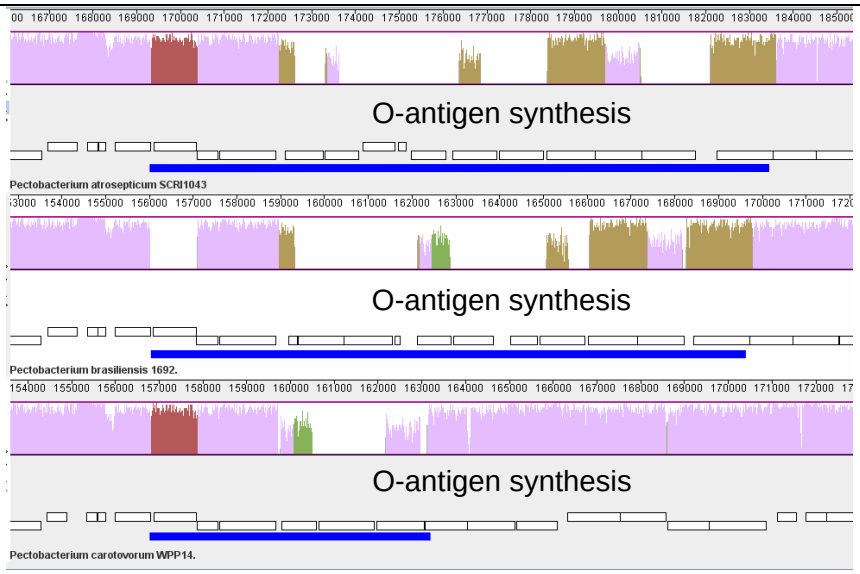
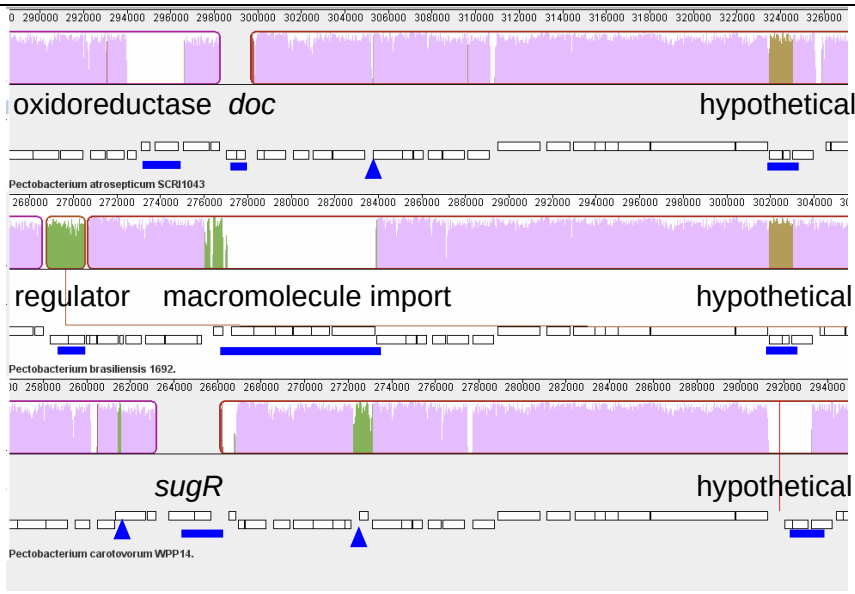
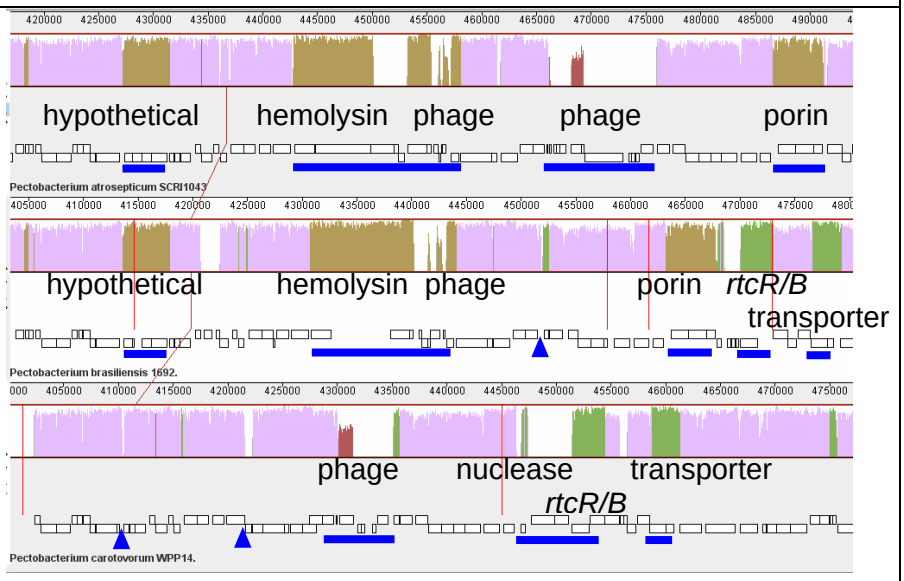
Parkinson, J. H., Ames, P., and Studdert, C. A. 2005. Collaborative signaling by bacterial chemoreceptors. Curr. Opin. Microbiol. 8:116-121.

Petnicki-Ocwieja, T., van Dijk, K., and Alfano, J. R. 2005. The *hrpK* op-

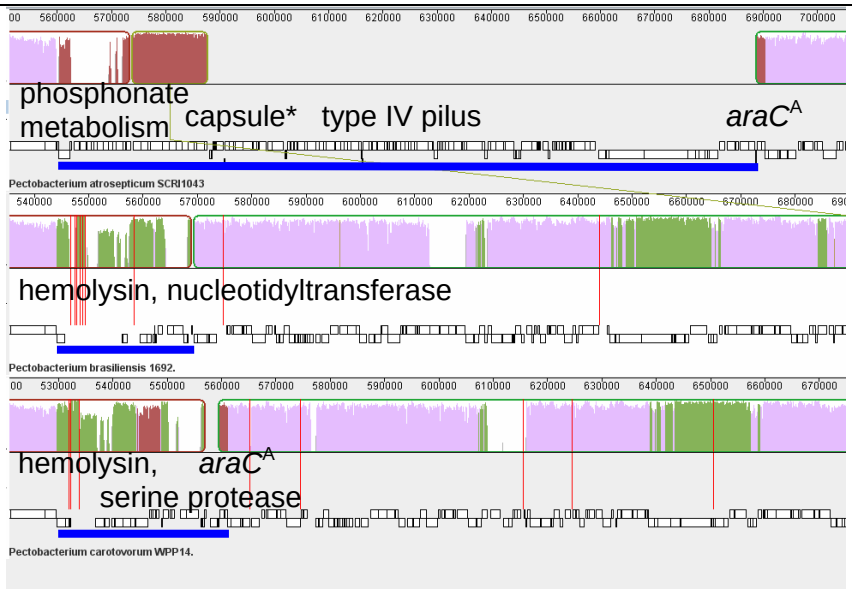
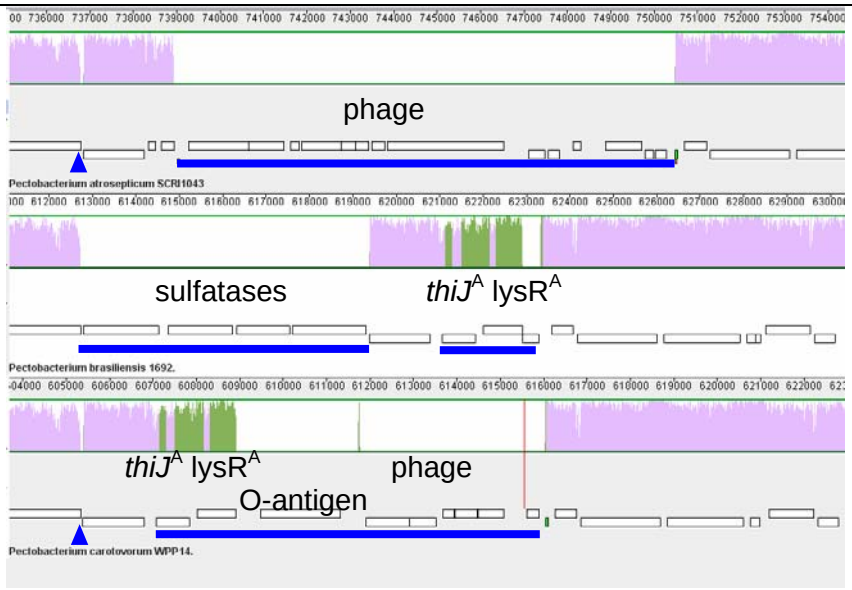
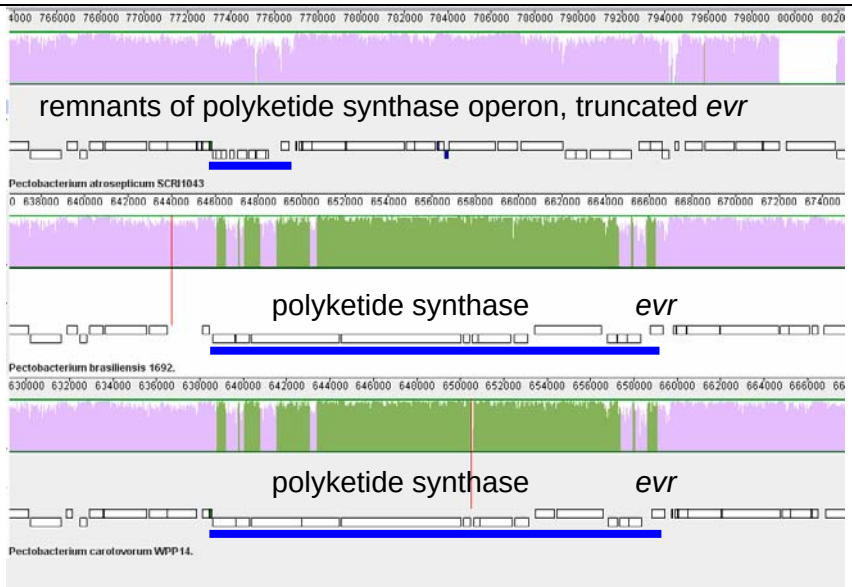
eron of *Pseudomonas syringae* pv. *tomato* DC3000 encodes two proteins secreted by the type III (Hrp) protein secretion system: HopB1 and HrpK, a putative type III translocator. J. Bacteriol. 187(2):649-663.
Phillips, J. A., and Kelman, A. 1982. Direct fluorescent antibody stain procedure applied to insect transmission of *Erwinia carotovora*. Phytopathology 72(7):898-901.
Praillet, T., Reverchon, S., and Nasser, W. 1997. Mutual control of the PecS/PecM couple, two proteins regulating virulence-factor synthesis in *Erwinia chrysanthemi*. Mol. Microbiol. 24(4):803-814.
Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. 2005. InterProScan: Protein domains identifier. Nucleic Acids Res. 33:W116-120.
Quilis, J., Meynard, D., Vila, L., Aviles, F. X., Guiderdoni, E., and San Segundo, B. 2007. A potato carboxypeptidase inhibitor gene provides pathogen resistance in transgenic rice. Plant Biotechnol. J. 5(4):537-553.
Rantakari, A., Virtaharju, O., Vahamiko, S., Taira, S., Palva, E. T., Saarilahti, H. T., and Romantschuk, M. 2001. Type III secretion contributes to the pathogenesis of the soft- rot pathogen *Erwinia carotovora*: Partial characterization of the *hrp* gene cluster. Mol. Plant-Microbe Interact. 14(8):962-968.
Ried, J. L., and Collmer, A. 1987. An *nptI-sacB-sacR* cartridge for constructing directed, unmarked mutations in gram-negative bacteria by marker-exchange-eviction mutagenesis. Gene 57:239-246.
Rojas, C. M., Ham, J.-H., Deng, W.-L., Doyle, J. J., and Collmer, A. 2002. HecA is a member of a class of adhesins produced by diverse pathogenic bacteria and contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. Proc. Natl. Acad. Sci. U.S.A. 99:13142-13147.
Sambrook, J., and Russell, D. W. 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, U.S.A.
Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., Madupu, R., Brinkac, L. M., Dodson, R. J., Rosovitz, M. J., Sullivan, S. A., Daugherty, S. C., Haft, D. H., Selengut, J., Gwinn, M. L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K. J. Smith, S., Utterback, T. R., White, O., Rubens, C. E., Grandi, G., Madoff, L. C., Kasper, D. L., Telford, J. L., Wessels, M. R., Rappuoli, R., and Fraser. C. M. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. U.S.A. 102 (39):13950-13955.
Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A. C., Meyer, F., Mittenhuber, G., Nies, D. H., Niesbach-Klosgen, U., Patschkowski, T., Ruckert, C., Rupp, O., Schneiker, S., Schuster, S. C., Vorholter, F. J., Weber, E., Puhler, A., Bonas, U., Bartels, D., and Kaiser, O. 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. J. Bacteriol. 187(21):7254-7266.
Toth, I. K., and Birch, P. R. J. 2005. Rotting softly and stealthily. Curr. Opin. Plant Biol. 8:424-429.
Toth, I. K., Pritchard, L., and Birch, P. R. J. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. Annu. Rev. Phytopathol. 44:305-336.
Triplett, L. R., Zhao, Y., and Sundin, G. W. 2006. Genetic differences between blight-causing Erwinia species with differing host specificities, identified by suppression subtractive hybridization. Appl. Environ. Microbiol. 72(11):7359-7364.
van Hijum, S. A. F. T., Zomer, A. L., Kuipers, O. P., and Kok, J. 2005. Projector 2: Contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. Nucleic Acids Res. 33:W560-W566.
Wandersman, C., Delepelaire, P., and Letoffe, S. 1990. Secretion processing and activation of *Erwinia chrysanthemi* proteases. Biochimie 72(2-3):143-146.
Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., Perna, N. T., Mobley, H. L., Donnenberg, M. S., and Blattner, F. R. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 99(26):17020-17024.
Willenbrock, H., Hallin, P. F., Wassenaar, T. M., and Ussery, D. W. 2007. Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. Genome Biol. 8(12):R267.
Yap, M.-N., Barak, J. D., and Charkowski, A. O. 2004. Genomic diversity of *Erwinia carotovora* subsp. *carotovora* and its correlation with virulence. Appl. Environ. Microbiol. 70:3013-3023.

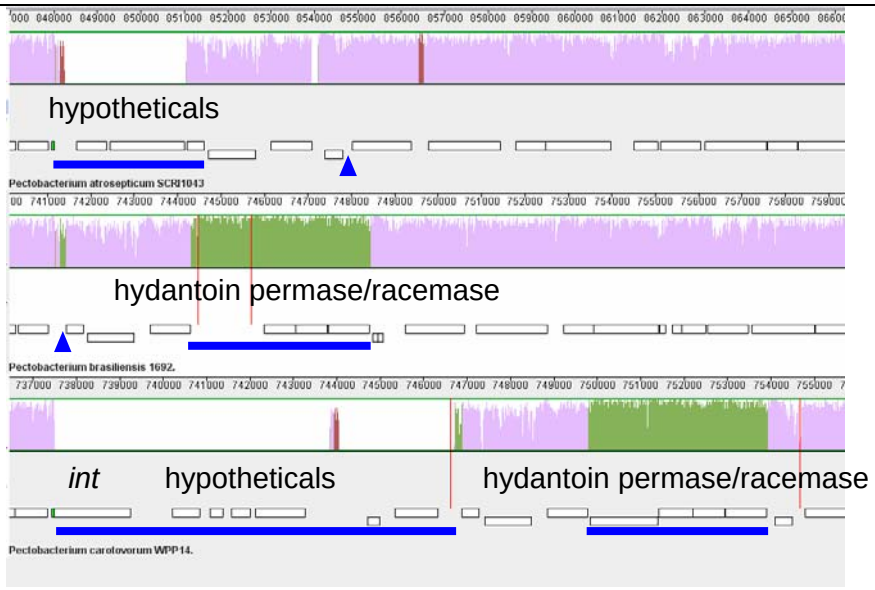
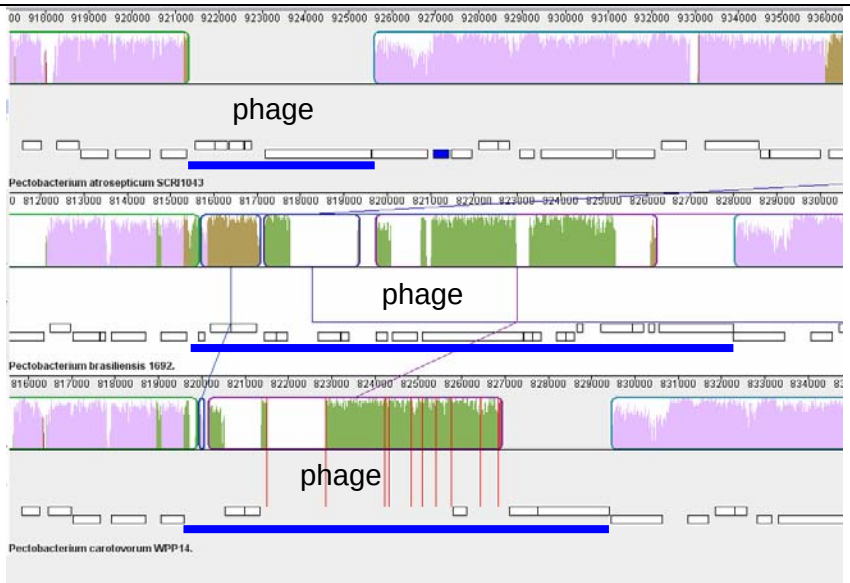
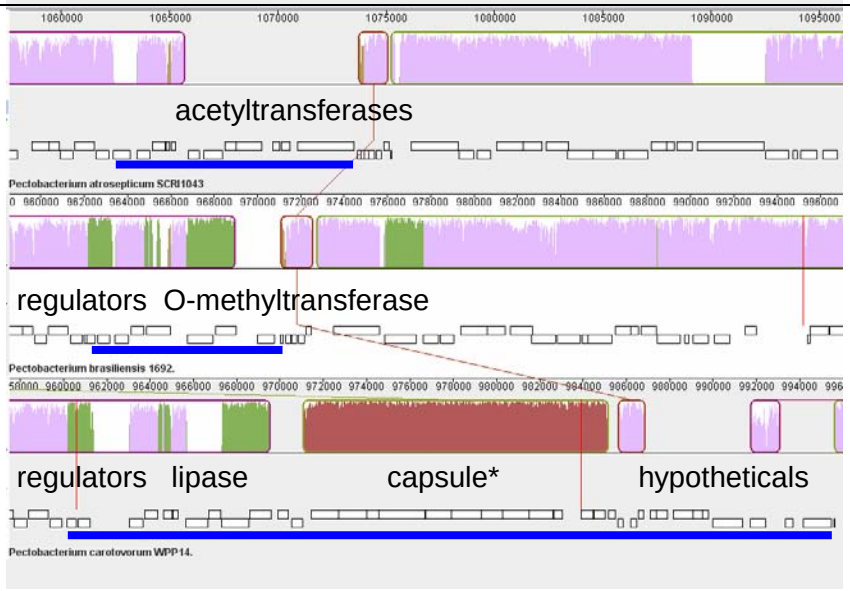
AUTHOR-RECOMMENDED INTERNET RESOURCES

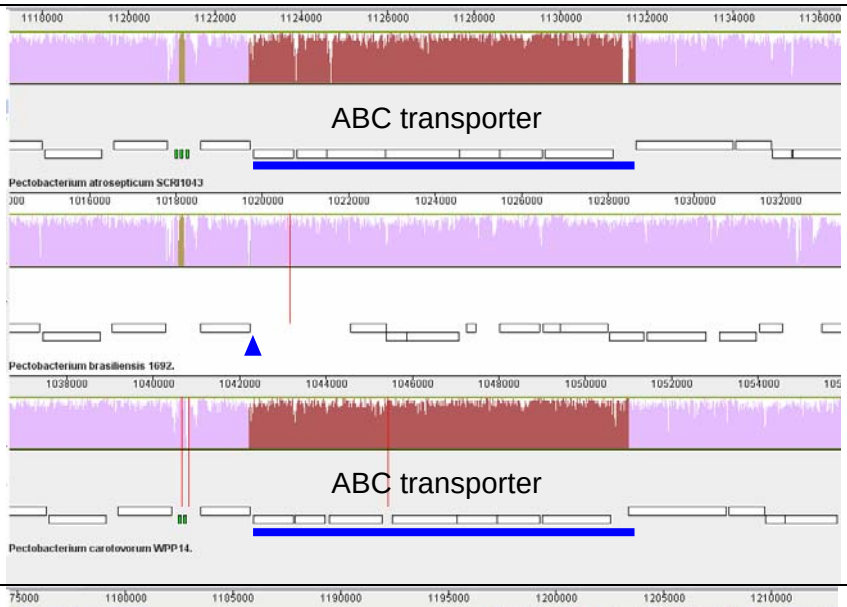
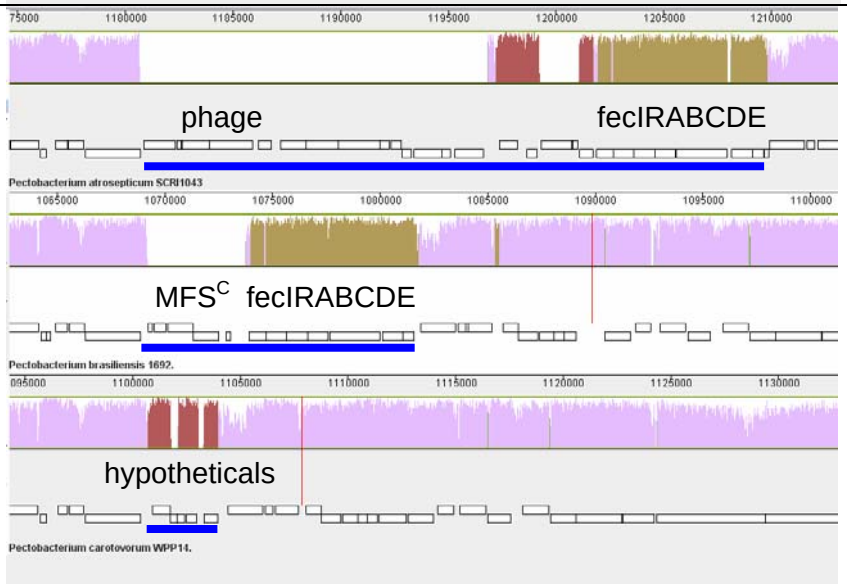
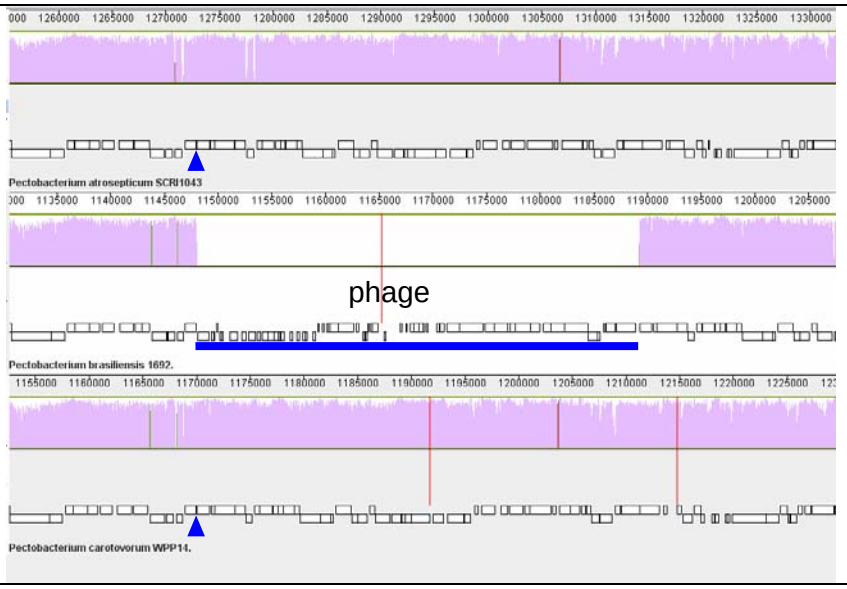
The University of Wisconsin's ASAP database:
asap.ahabs.wisc.edu/asap/home.php
Gene Ontology map of InterPro families:
www.geneontology.org/external2go/interpro2go
RAST (Rapid Annotation using Subsystem Technology) server:
rast.nmpdr.org

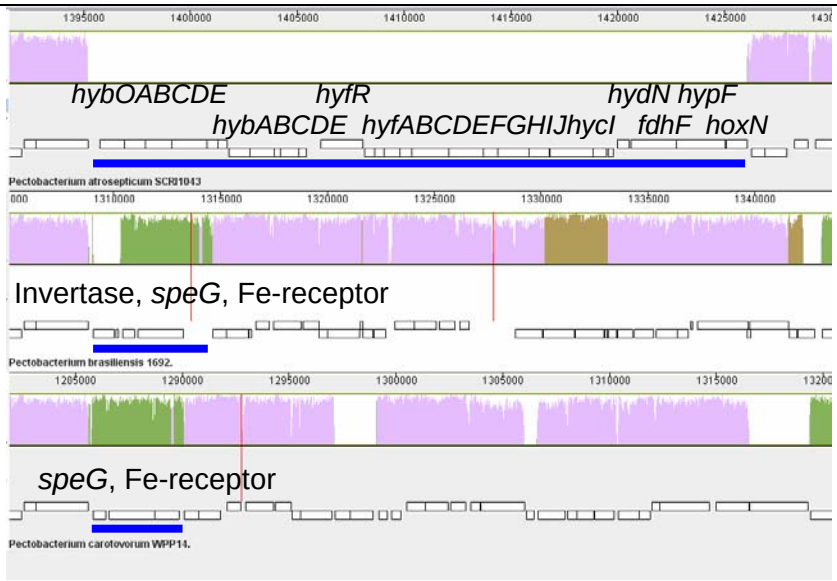
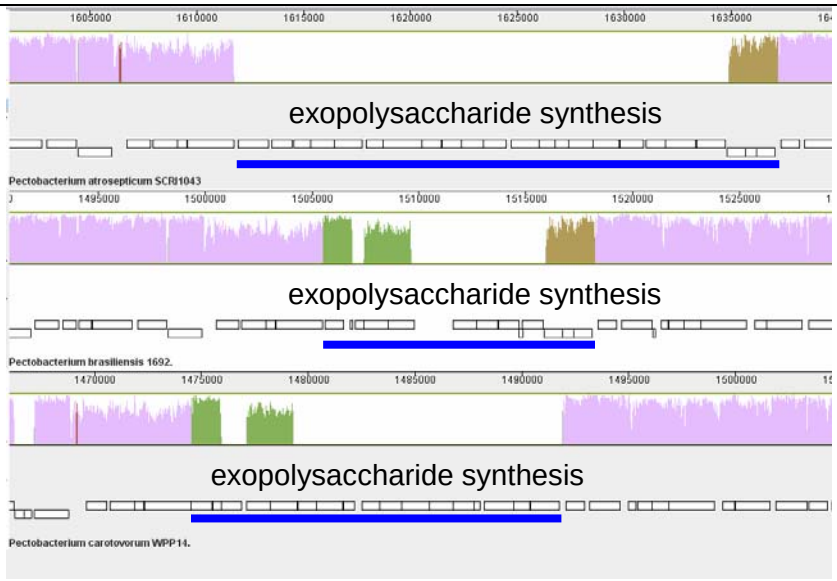
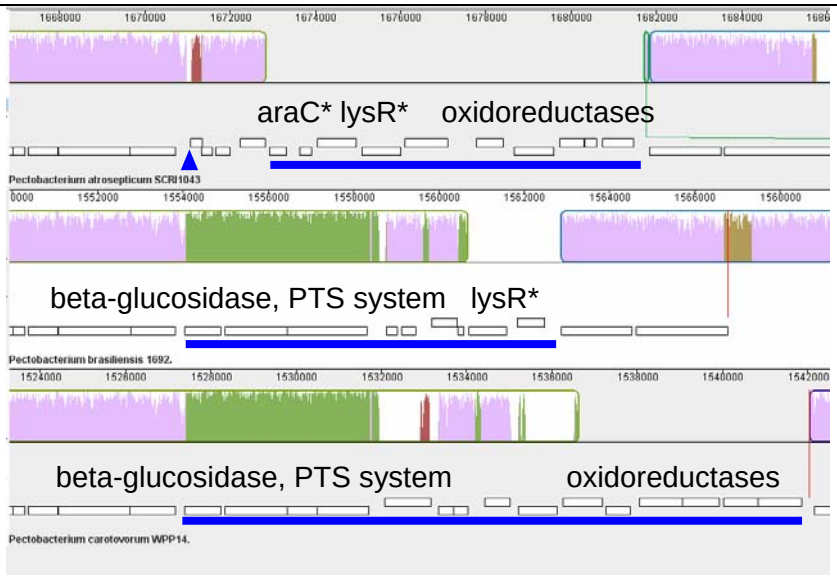
Variable Region	
VR-001 *Pa* 0148-0163 *Pb* AED-0003652-0086 *Pc* ADT-0002352-3443	
VR-002 *Pa* 0247-0276 *Pb* AED-0000195-0190 *Pc* ADT-0000083-2389	
VR-003 *Pa* 0375-0434 *Pb* AED-0000259-0304 *Pc* ADT-0002415-2447	

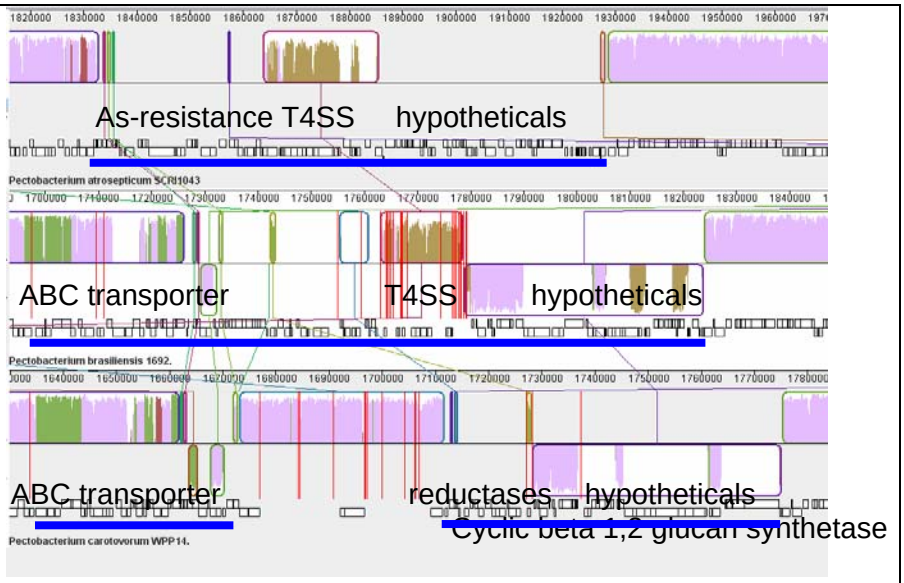
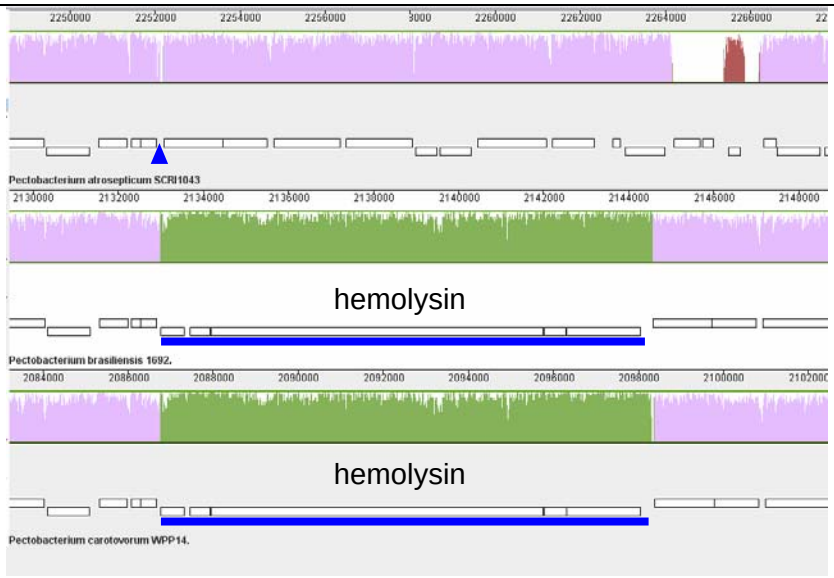
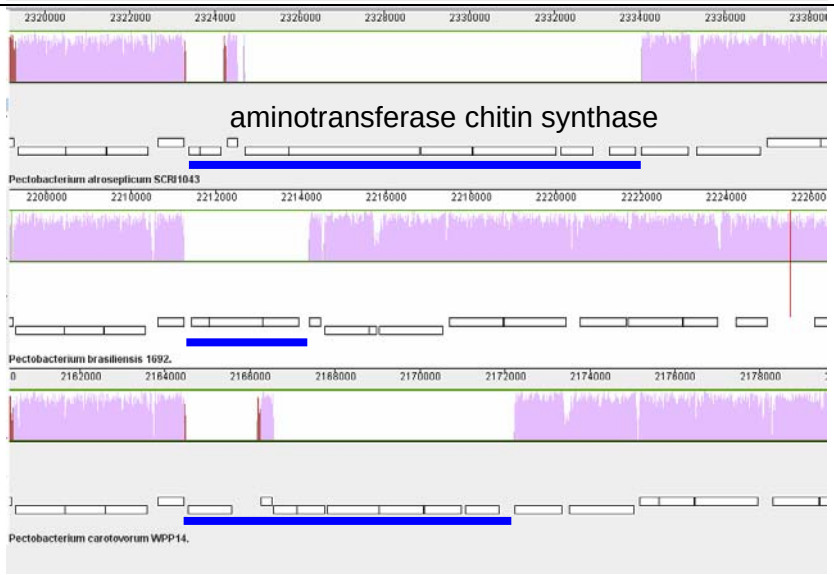
Supplemental Figure 1. *Pectobacterium* variable regions (VR). Islands that are larger than 5 kb, or clusters of shorter islands that collectively span a region greater than 10 kb in at least one of the *Pectobacterium* species are described. Genes flanking each VR are listed as well as putative functions for the genes encoded in the island (note that gene identifiers may not be consecutive numbers in the *Pb* and *Pc* draft sequences and that the gene identifiers from *Pa* correspond to locus tag numbers with the prefix ECA from Bell et al. (Bell et al. 2004). The variable region (VR_000) designations provide an identifier for a polymorphic region flanked by conserved chromosome irrespective of whether or not there are VRs (homologous or nonhomologous) at that location. The figures on the right are based on the Mauve alignment and visualization. For each genome boxes indicate genes (above the line transcribed from the plus strand, below the line on the minus strand). Homologous chromosomal regions are aligned in all three genomes. The height of the colored bars above the genes show the level of nucleotide identity in conserved regions and the color indicates which genomes share a region, with a pink color indicating a region conserved in all three genomes, brown conserved in *Pa* and *Pb*, green conserved in *Pb* and *Pc*, and red conserved in *Pa* and *Pc*. Blue bars beneath the genes indicate VRs and blue arrows indicate the site of insertion/deletion of a VR in a genome lacking the region. Labels above the genes indicate predicted functions encoded in the VR.

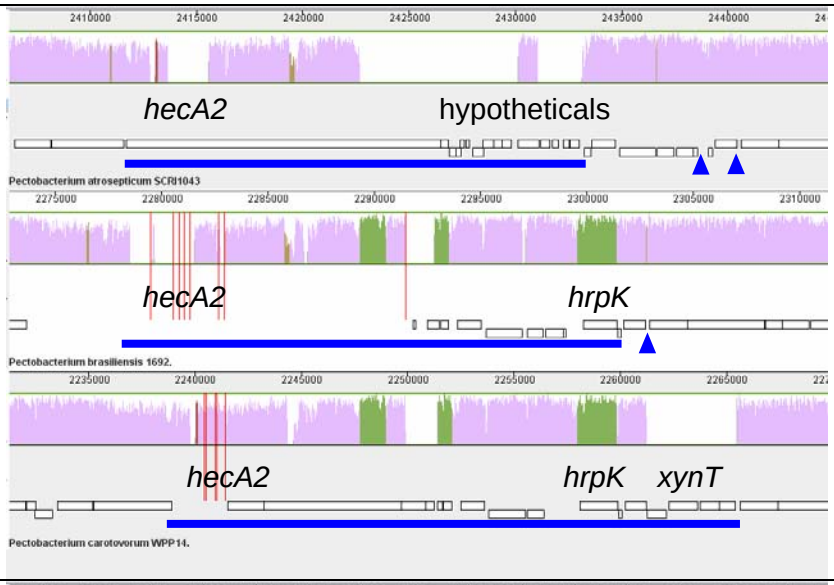
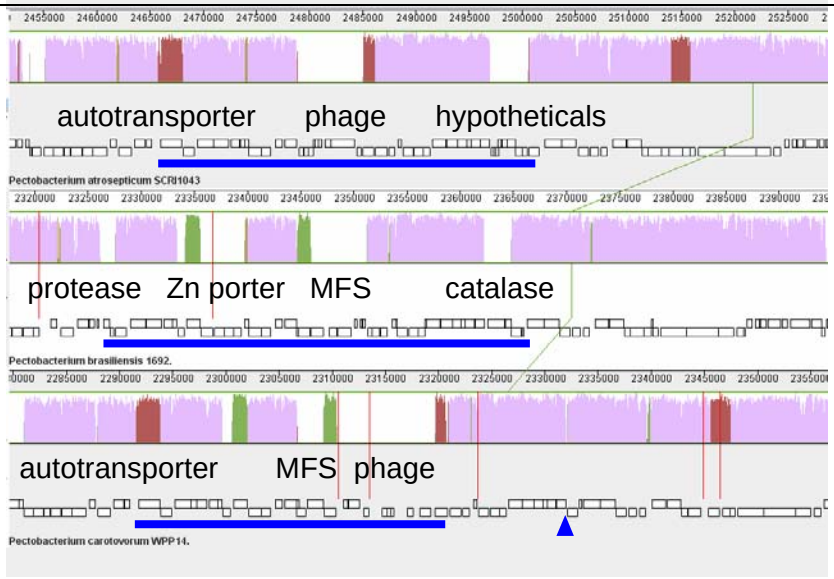
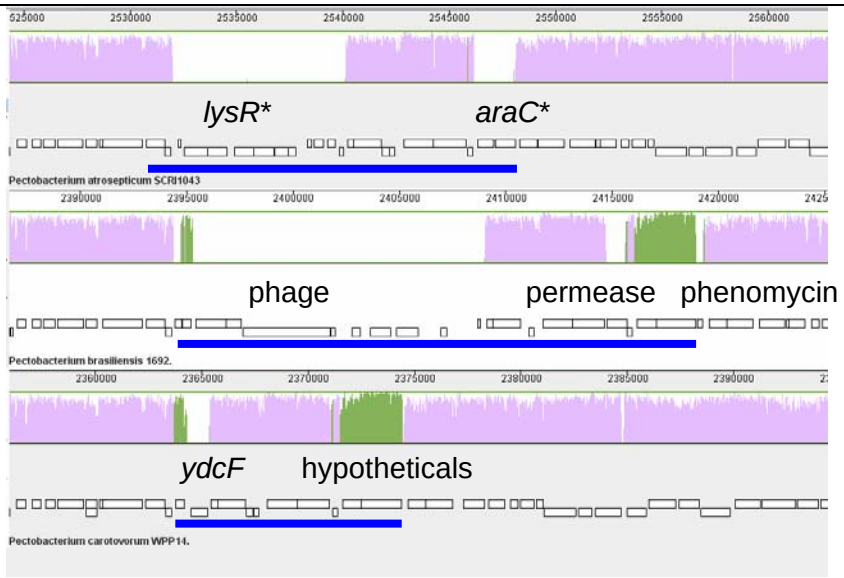
VR-004 *Pa* 0483-0617 *Pb* AED-0002887-2736 *Pc* ADT-0002461-1340 *capsule orthologs are in PC VR0009	
VR-005 *Pa* 0661-0679 *Pb* AED-0004265-4273 *Pc* ADT-0001354-3358	
VR-006 *Pa* 0702-0709 *Pb* AED-0000384-0385 *Pc* ADT-0001363-0178	

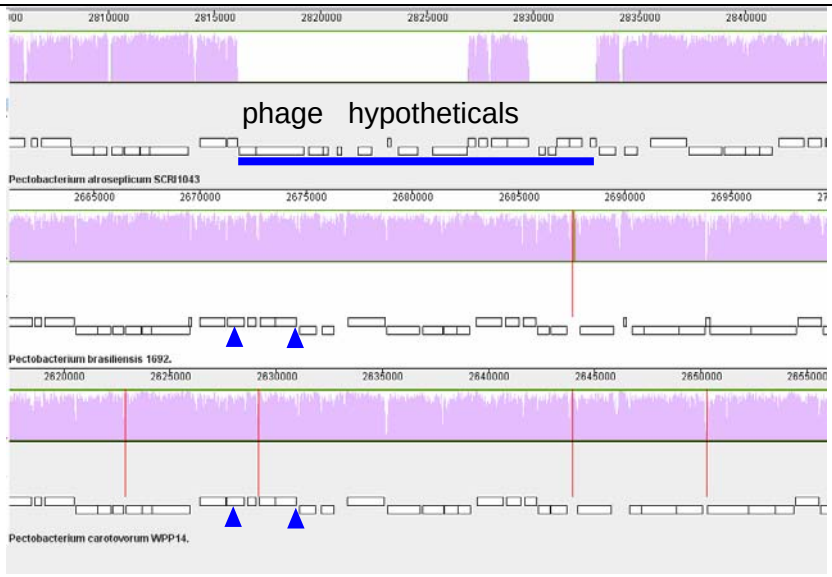
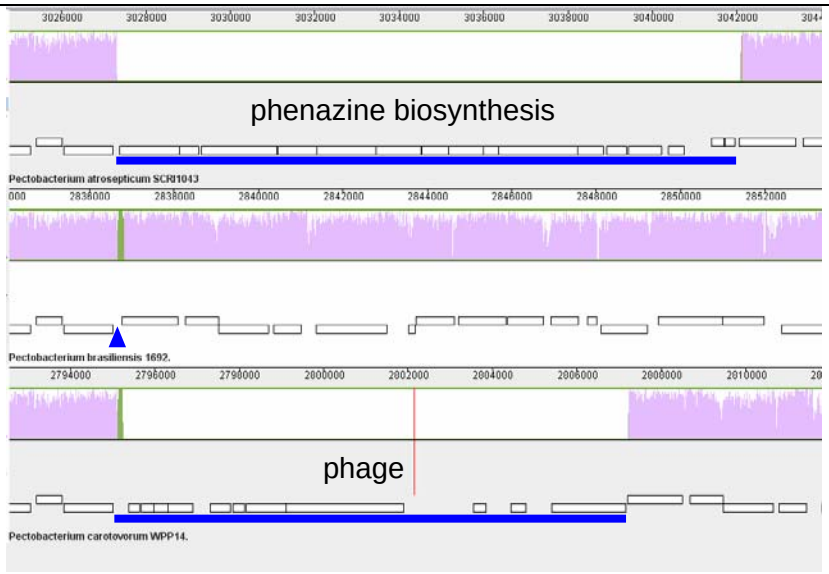
VR-007 *Pa* 0776-0782 *Pb* AED-0000418-3340 *Pc* ADT-0003391-2525	
VR-008 *Pa* 0829-0782 *Pb* AED-0000455-3340 *Pc* ADT-0002528-2525	
VR-009 *Pa* 0948-0968 *capsule homologs in *Pa* VR0004 *Pb* AED-0003156-3169 *Pc* ADT-0003572-3342	

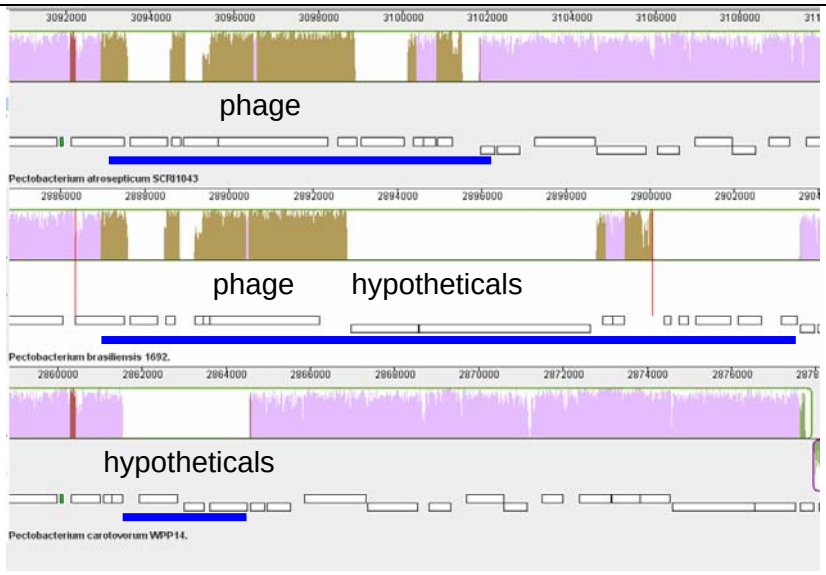
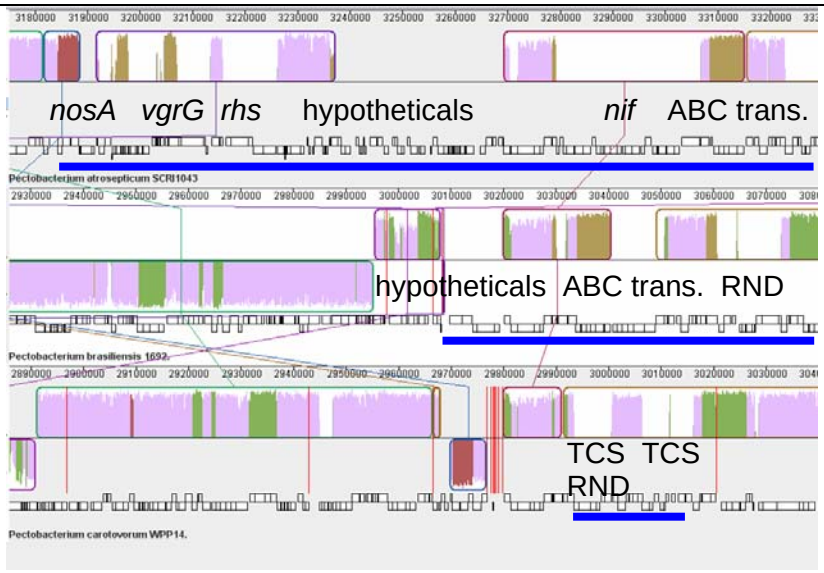
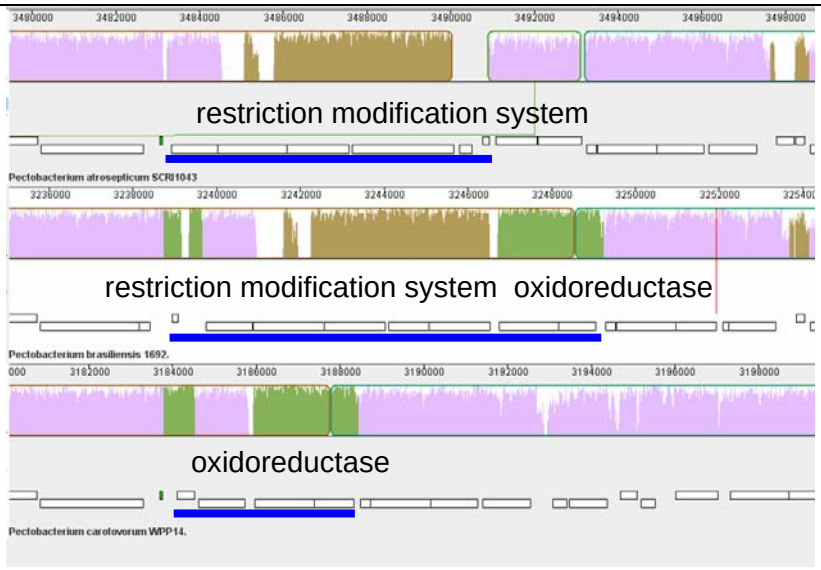
VR-010 *Pa* 1001-1009 *Pb* AED-0002825-0000580 *Pc* ADT-0000245-1464	
VR-011 *Pa* 1053-1081 *Pb* AED-0000593-4175 *Pc* ADT-0000256-3581	
VR-012 *Pa* 1119-1120 *Pb* AED-0002071-0639 *Pc* ADT-0001504-1505	

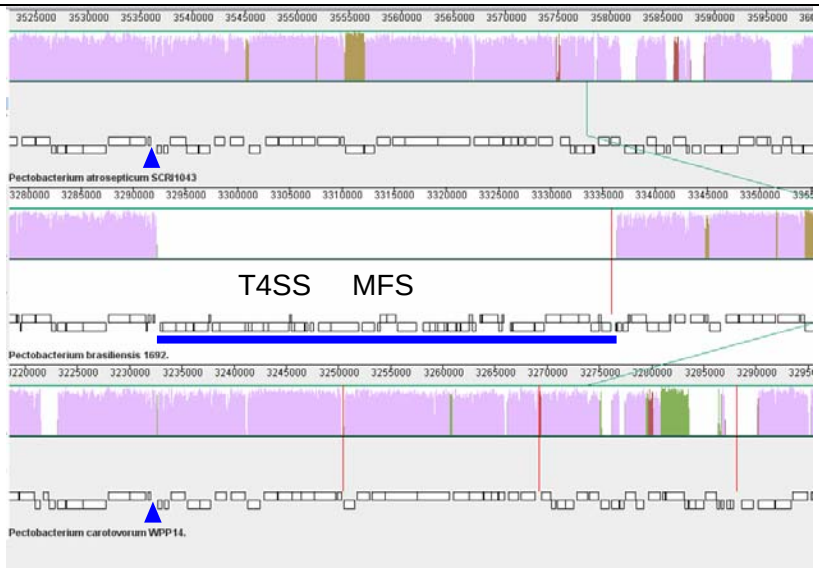
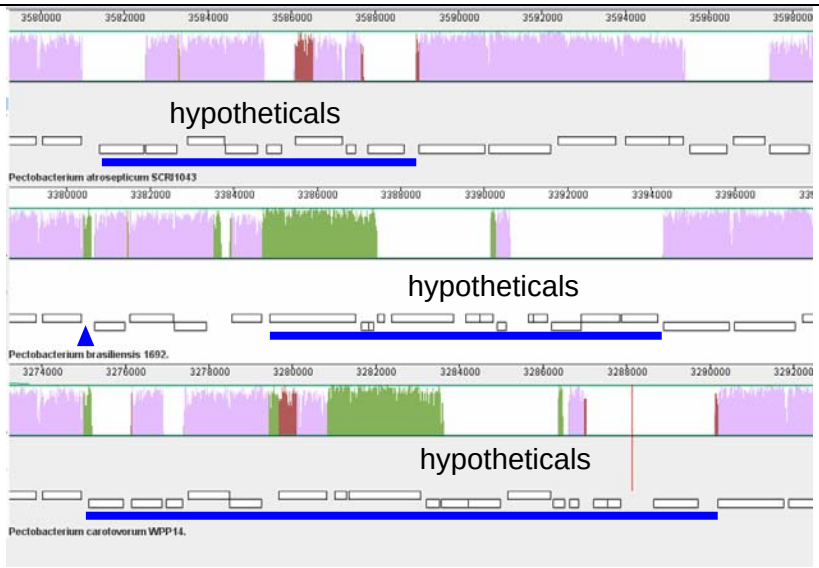
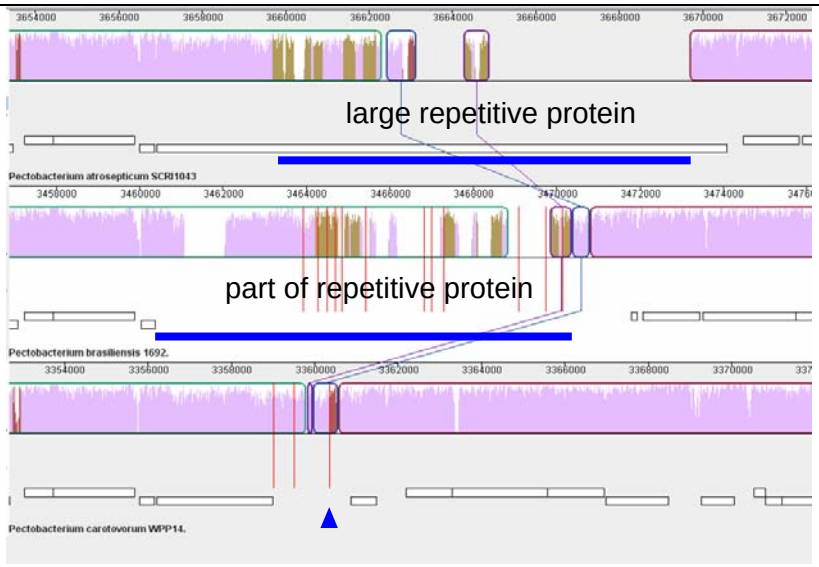
VR-013 *Pa* 1224-1253 *Pb* AED-0000669-0692 *Pc* ADT-0001538-0311	
VR-014 *Pa* 1419-1443 *Pb* AED-0002833-0842 *Pc* ADT-0001578-3667	
VR-015 *Pa* 1470-1486 A homolog of a non-ribosomal peptide synthetase located at the edge of *Pa* and *Pb* VR16 is in *Pc* VR17 *Pb* AED-0000822-0837 *Pc* ADT-0000388-3733 (rearrangement in this region in *Pc*)	

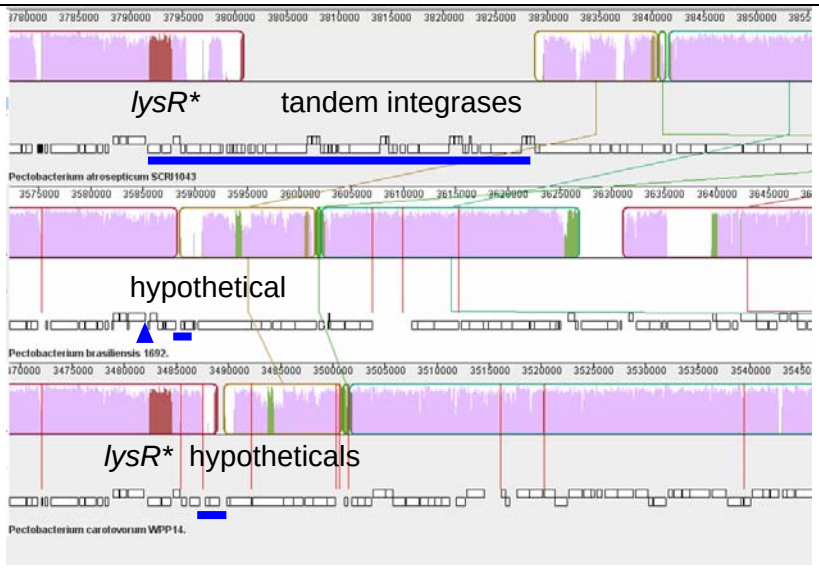
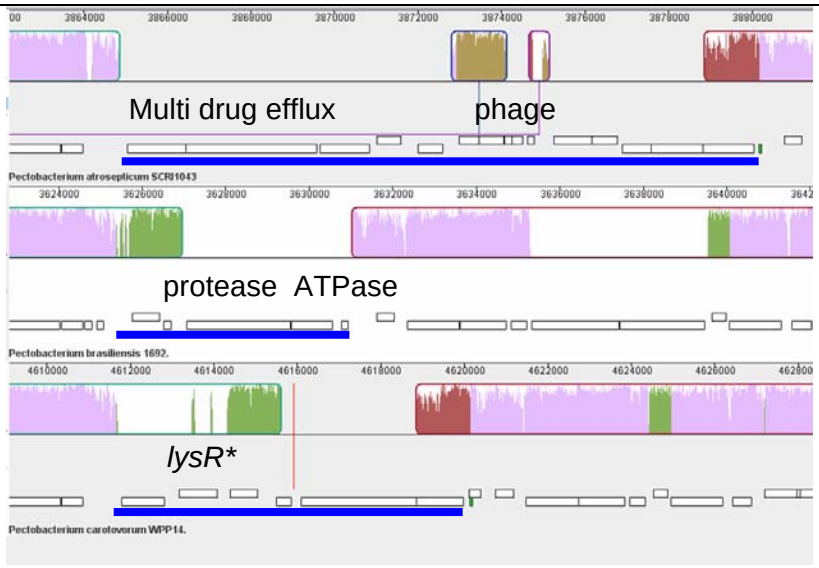
VR-016 *Pa* 1560-1681 *Pb* AED-0003634-2828 VR contains alternate flagellin found in *Pa* VR 29 *Pc* ADT-0000405-1628 VR contains alternate flagellin found in *Pa* VR 29	
VR-017 *Pa* 1985-1986 *Pb* AED-0001015-1064 *Pc* ADT-0000518-0519	
VR-018 *Pa* 2041-2051 *Pb* AED-0001040-0985 *Pc* ADT-0000529-1730	

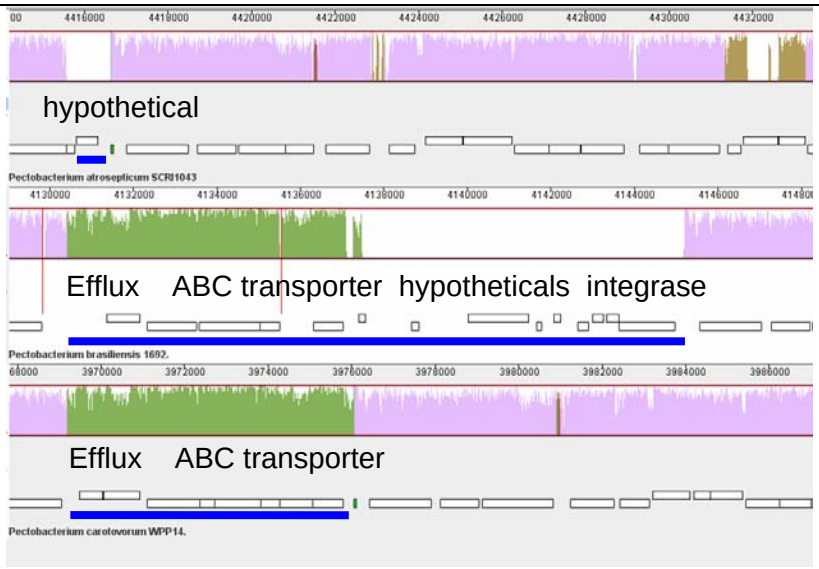
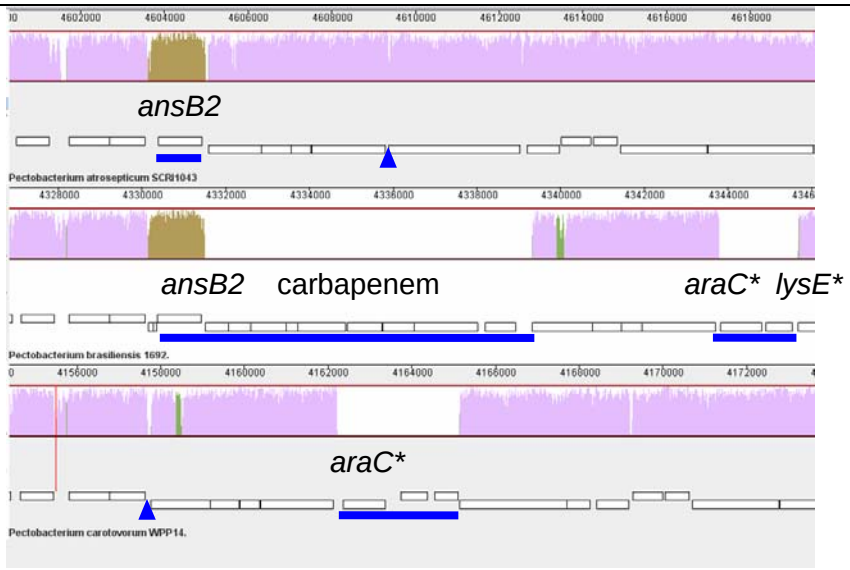
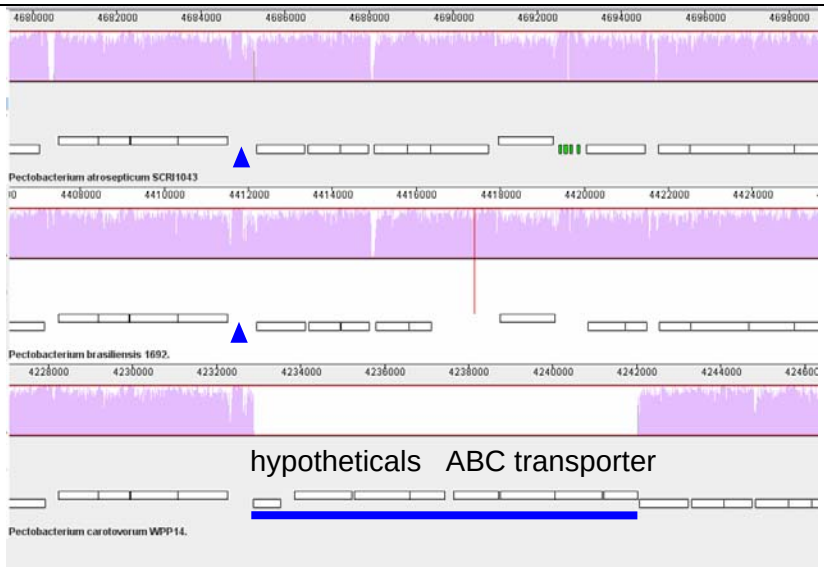
VR-019 *Pa* 2041-2051 *Pb* AED-0001040-0985 *Pc* ADT-0000529-1730	
VR-020 *Pa* 2162-2203 *Pb* AED-0001245-1324 *Pc* ADT-0000552-1759	
VR-021 *Pa* 2227-2248 *Pb* AED-0004306-2807 *Pc* ADT-0003878-3872	

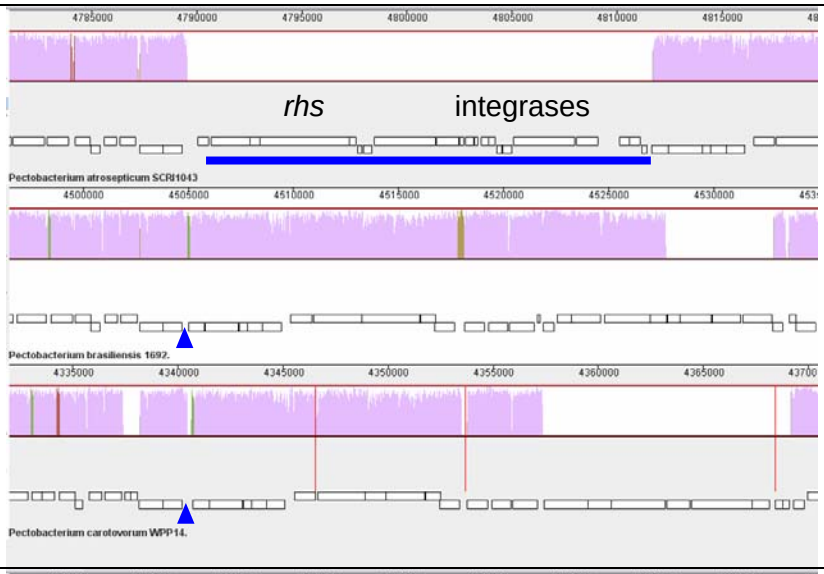
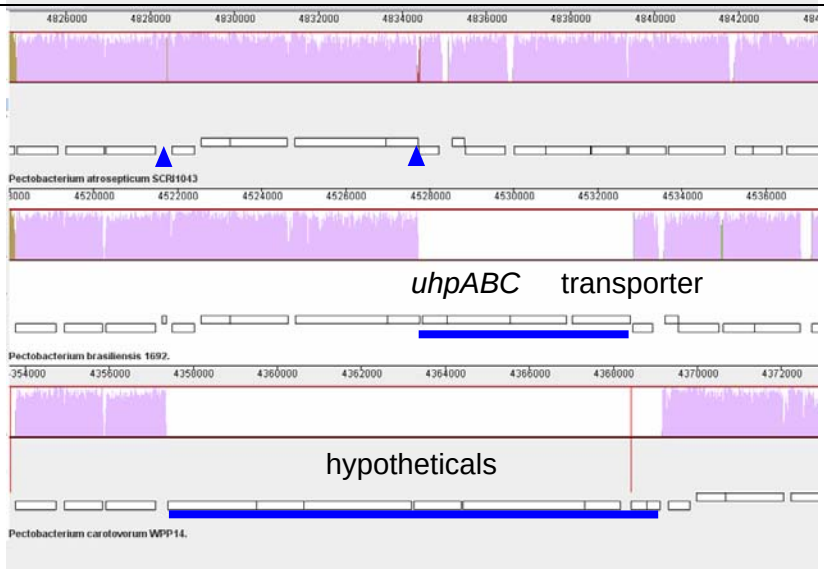
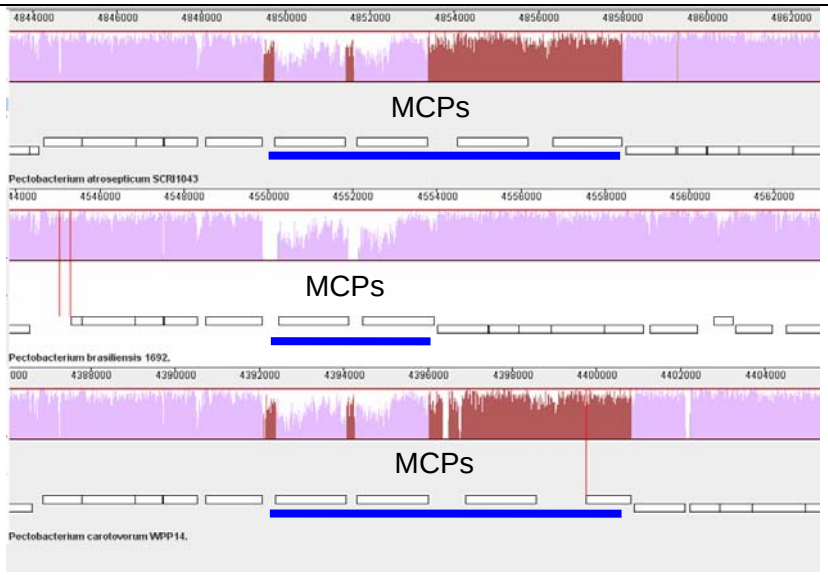
VR-022 *Pa* 2498-2518 *Pb* AED-0004392-1266 *Pc* ADT-0002923-0656	
VR-023 *Pa* 2597-2597A *Pb* AED-0001494-3676 *Pc* ADT-0001866-3943	
VR-024 *Pa* 2691-2708 *Pb* AED-0003691-3692 *Pc* ADT-0001885-2969	

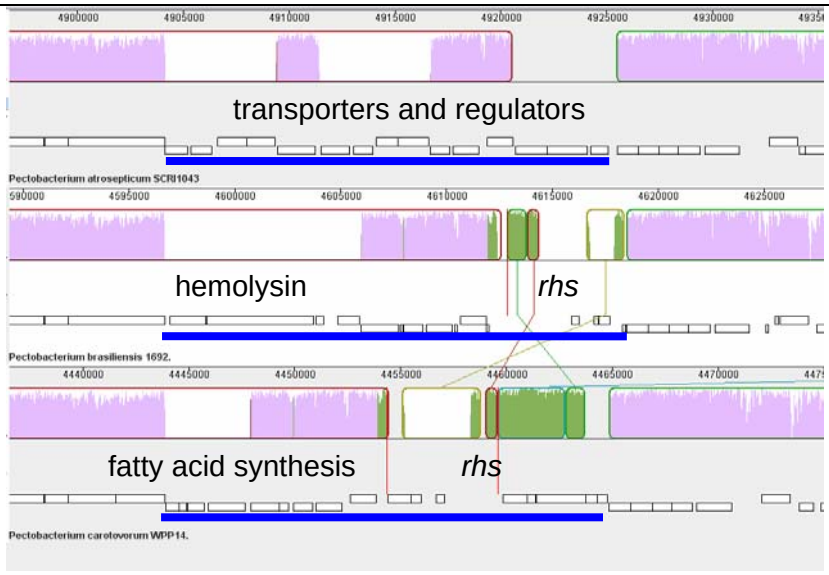
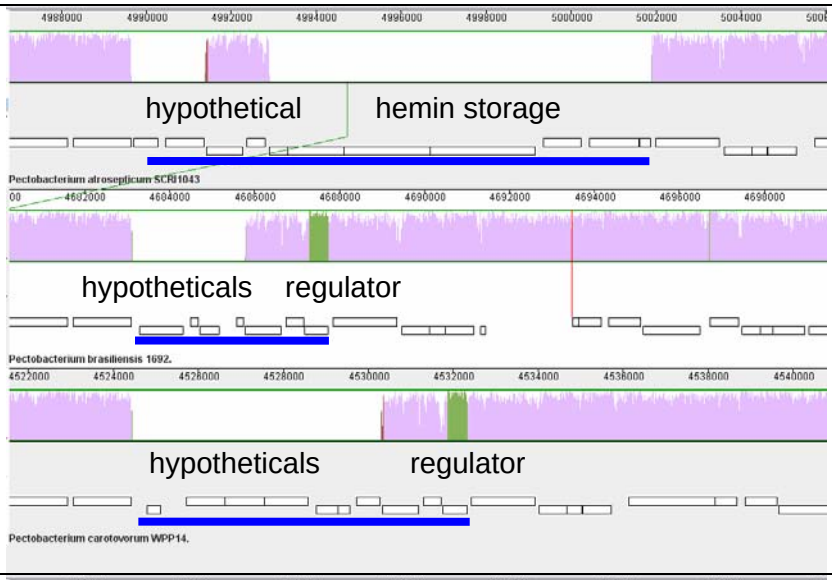
VR-025 *Pa* 2750-2760 *Pb* AED-0002838-1619 *Pc* ADT-0002982-1899	
VR-026 *Pa* 2840-2986 VR 29 contains alternate flagellin found in *Pb* and *Pc* VR 17 *Pb* AED-0004100-1680 (rearrangement in this region in *Pb*) *Pc* ADT-0003770-1936	
VR-027 *Pa* 3118-3128 (rearrangement in this region in *Pa*) *Pb* AED-0003490-1789 *Pc* ADT-0002572-0804	

VR-028 *Pa* 3166-3167 *Pb* AED-0002706-1803 *Pc* ADT-0001980-0815	
VR-029 *Pa* 3199-3208 *Pb* AED-0000184-1855 *Pc* ADT-0004113-0826	
VR-030 This region is potentially misassembled *Pa* 3265-3267 *Pb* AED-0001827-2844 *Pc* ADT-0004123-41234	

VR-031 *Pa* 3374-3421 *Pb* AED-0003349-3356 *Pc* ADT-0003089-4151	
VR-032 *Pa* 3445-3461 *Pb* AED-0002943-1987 *Pc* ADT-0000887-4442	
VR-033 *Pa* 3690-3745 *Pb* AED-0003103-3096 *Pc* ADT-0003607-2586	

VR-034 *Pa* 3943-3746 *Pb* AED-0003095-2279 *Pc* ADT-0003166-1022	
VR-035 *Pa* 4125-4131 *Pb* AED-0004540-3574 *Pc* ADT-0001808-3196	
VR-036 *Pa* 4191-4192 *Pb* AED-0002464-2472 *Pc* ADT-0002249-1132	

VR-037 *Pa* 4274-4296 *Pb* AED-0002544-3608 *Pc* ADT-0002262-4314	
VR-038 *Pa* 4309-4296 *Pb* AED-0002548-3608 *Pc* ADT-0002268-4314	
VR-039 *Pa* 4331-4336 *Pb* AED-0002590-2577 *Pc* ADT-0001182-2277	

VR-040 *Pa* 4374-4296 *Pb* AED-0002560-3608 *Pc* ADT-0003253-4314	
VR-041 *Pa* 4446-4457 *Pb* AED-0002609-4203 *Pc* ADT-0001199-3262	
VR-042 *Pa* 4477-4508 *Pb* AED-0002681-2649 *Pc* ADT-0002303-3403