Evolution of the Rhodococcus equi vap pathogenicity island seen through comparison of host-associated vapA and vapB virulence plasmids

JOURNAL OF BACTERIOLOGY, Sept. 2008, p. 5797–5805 Vol. 190, No. 17
0021-9193/08/$08.00+0 doi:10.1128/JB.00468-08

Evolution of the *Rhodococcus equi vap* Pathogenicity Island Seen through Comparison of Host-Associated *vapA* and *vapB* Virulence Plasmids▽†

Michal Letek,[1,2,3]‡ Alain A. Ocampo-Sosa,[2,3]‡ Mandy Sanders,[4] Ursula Fogarty,[3] Tom Buckley,[3] Desmond P. Leadon,[3] Patricia González,[1,3] Mariela Scortti,[1,2,5] Wim G. Meijer,[6] Julian Parkhill,[4] Stephen Bentley,[4] and José A. Vázquez-Boland[1,2,5]*

Division of Microbial Pathogenesis, Centre for Infectious Diseases, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom[1]; Bacterial Molecular Pathogenesis Group, Faculty of Medical and Veterinary Sciences, University of Bristol, Bristol, United Kingdom[2]; Irish Equine Centre, Johnstown, Naas, Ireland[3]; Pathogen Genomics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, United Kingdom[4]; Grupo de Patogenómica Bacteriana, Instituto de Biología Molecular y Genómica, Universidad de León, León, Spain[5]; and School of Biomolecular and Biomedical Science, University College Dublin, Dublin, Ireland[6]

Received 6 April 2008/Accepted 18 June 2008

The pathogenic actinomycete *Rhodococcus equi* harbors different types of virulence plasmids associated with specific nonhuman hosts. We determined the complete DNA sequence of a *vapB*$^+$ plasmid, typically associated with pig isolates, and compared it with that of the horse-specific *vapA*$^+$ plasmid type. pVAPB1593, a circular 79,251-bp element, had the same housekeeping backbone as the *vapA*$^+$ plasmid but differed over an ≈22-kb region. This variable region encompassed the *vap* pathogenicity island (PAI), was clearly subject to selective pressures different from those affecting the backbone, and showed major genetic rearrangements involving the *vap* genes. The pVAPB1593 PAI harbored five different *vap* genes (*vapB* and *vapJ* to *-M*, with *vapK* present in two copies), which encoded products differing by 24 to 84% in amino acid sequence from the six full-length *vapA*$^+$ plasmid-encoded Vap proteins, consistent with a role for the specific *vap* gene complement in *R. equi* host tropism. Sequence analyses, including interpolated variable-order motifs for detection of alien DNA and reconstruction of Vap family phylogenetic relationships, suggested that the *vap* PAI was acquired by an ancestor plasmid via lateral gene transfer, subsequently evolving by *vap* gene duplication and sequence diversification to give different (host-adapted) plasmids. The *R. equi* virulence plasmids belong to a new family of actinobacterial circular replicons characterized by an ancient conjugative backbone and a horizontally acquired niche-adaptive plasticity region.

Rhodococcus equi is a member of the mycolic acid-containing group of actinobacteria, or mycolata, which also includes the *Corynebacterium*, *Gordonia*, *Mycobacterium*, and *Nocardia* genera (18). Like many other actinomycetes, *R. equi* is ubiquitous in nature and lives as a saprophyte in soil (25, 35, 41). The genus *Rhodococcus* is a genetically diverse taxon that may be empirically classified into two groups (4, 23): the nonpathogenic, or environmental, rhodococci, exemplified by *Rhodococcus erythropolis*, which includes metabolically versatile bacteria of industrial interest (32), and the pathogenic rhodococci, with two species, the plant pathogen *Rhodococcus fascians* (22) and the animal pathogen *R. equi* (25, 35, 41). All rhodococci typically harbor large conjugative plasmids encoding niche-adaptive functions, such as various primary and secondary metabolic processes in the environmental species and host colonization (virulence) factors in the pathogenic ones. Some of these extrachromosomal replicons are linear megaplasmids of up to 1 Mb in size, whereas others are circular plasmids of ≈100 kb (34, 55).

R. equi can be isolated from pulmonary and extrapulmonary pyogranulomatous infections in various mammalian hosts. It causes equine purulent bronchopneumonia, or rattles, a severe respiratory disease of foals characterized by extensive abscessation of the lung parenchyma, lymphadenitis, and a high mortality rate. *R. equi* is also an opportunistic human pathogen associated with AIDS and immunosuppression. Human rhodococcal infection is clinically and pathologically similar to pulmonary tuberculosis (36, 41).

R. equi is a facultative intracellular parasite that replicates within macrophages, causing their destruction by necrosis (25, 33). The pathogenicity of *R. equi* is associated with the presence of a large, circular plasmid of 80 to 90 kb (52). The loss of this plasmid results in an inability to cause disease in foals and to replicate in macrophages in vitro and in mouse tissues in vivo (21, 26). Plasmids from equine isolates encode virulence-associated protein A (VapA) (48), an in vivo-expressed, 17.4-kDa surface lipoprotein antigen (6, 10, 51). Deletion of the *vapA* gene attenuates *R. equi* virulence to an extent similar

* Corresponding author. Mailing address: Division of Microbial Pathogenesis, Centre for Infectious Diseases, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom. Phone: 44 (0)131 651 3619. Fax: 44 (0)131 650 6564. E-mail: v.boland@ed.ac.uk.

† Supplemental material for this article may be found at http://jb.asm.org/.

‡ These authors are alphabetically arranged and contributed equally to the work.

▽ Published ahead of print on 7 July 2008.

to that for plasmid curing, suggesting a major role for VapA in pathogenesis (21, 28). An analysis of plasmid DNA sequences from two virulent equine isolates, strains 103 and ATCC 33701, showed *vapA* to be present in a pathogenicity island (PAI) encoding other Vap proteins, VapC to VapH (50). These Vaps display substantial sequence similarity, particularly at their carboxy termini, and no homologs have been found in other organisms. The specific roles of the different Vap proteins in *R. equi* virulence remain unknown.

Nonequine *R. equi* isolates often produce a larger (20-kDa) variant surface lipoprotein antigen, VapB (10, 49), which is very similar in primary structure to VapA (78% amino acid sequence identity) (39). VapA and VapB (and corresponding genes) are mutually exclusive, i.e., they do not occur in the same isolate, suggesting that they are allelic variants of one locus that has divergently evolved in two different plasmid subpopulations (38). In a recent study based on detection of *vapA* and *vapB* gene markers by PCR, we showed that *R. equi* virulence plasmids can be classified into three general types: *vapA*$^+$, *vapB*$^+$, and *vapAB* negative (38). Each of these plasmid types was almost exclusively associated with a specific nonhuman animal host, i.e., horse, pig, and cattle, respectively (38). Moreover, no *vapB*$^+$ (pig-type) plasmid was found in equine isolates, no *vapA*$^+$ (horse-type) plasmid was identified among bovine isolates, and no *vapAB*-negative (bovine-type) plasmid was found among porcine isolates. By contrast, the three plasmid types were common among *R. equi* strains from humans (11%, 55%, and 34% of plasmid-positive isolates, respectively), a host in which the infection is opportunistic and associated with immunosuppression. These data suggested the existence of host immunity-driven selection of specific plasmid types in *R. equi* (38) and supported a model in which *R. equi* host tropism is dictated by the plasmid in naturally susceptible animal species.

Comparative analysis of virulence plasmid types associated with specific animal hosts may therefore provide valuable insight into the molecular mechanisms underlying *R. equi* pathogenesis and host specificity. We report here the nucleotide sequence of a *vapB*-type plasmid from *R. equi*. We show that this plasmid, pVAPB1593, and the previously sequenced *vapA* (equine)-type plasmid from strains 103 and ATCC 33701 (50) share the same backbone but differ in terms of their *vap* PAIs, suggesting that differences in this locus and the Vap proteins that it encodes may play an important role in *R. equi* host tropism.

MATERIALS AND METHODS

Bacteria, growth conditions, and chemicals. The *R. equi vapB*$^+$ PAM 1593 isolate used in this study has been described previously (38). The bacteria were grown at 30°C for 24 to 48 h in brain heart infusion medium (Difco-BD, Detroit, MI). All chemicals were obtained from Sigma (Sigma-Aldrich, Poole, England, United Kingdom) unless otherwise stated.

Plasmid DNA isolation. Plasmid pVAPB1593 DNA was isolated from early-stationary-phase broth cultures of PAM 1593 by using a Qiagen midi prep kit according to a low-copy-number plasmid protocol (42). Lysozyme (5 μg/ml) was added to the bacterial cell suspension in buffer P1, and the mixture was incubated at 37°C for 2 h before processing. DNA quality and concentration were assessed visually by agarose electrophoresis and spectrophotometrically by calculating the ratio of absorbances at 260 and 280 nm.

DNA sequencing. The pVAPB1593 DNA was sheared by sonication and size fractionated. A random library was constructed in pUC19 and the complete plasmid sequence determined by a shotgun method, as previously described (7). DNA sequences were collected from paired end reads of pUC19 library clones by using BigDye Terminator chemistry and ABI 3730 sequencing machines (Applied Biosystems, Foster City, CA). Sequence gaps were bridged by read pairs or end sequenced PCR. Manual error checking and finishing of the sequence were performed to ensure that every base was covered by at least two clones with high-quality sequence in each direction, and that every consensus base had a Phred quality score (19) of >30 (i.e., <1/1,000 chance of error).

Sequence analysis and annotation. Plasmid DNA sequences were assembled, finished, and annotated using Artemis software (http://www.sanger.ac.uk/Software/Artemis/) (46). Glimmer2 software (17) was used for the initial prediction of coding sequences (CDSs). These initial predictions were curated manually in Artemis, then translated into protein sequences, and used for searches of the NCBI database with BLASTp (2) and searches of the Pfam (3) and Prosite (27) databases of protein motifs. Putative secreted and membrane proteins were identified with SignalP 3.0 (5) and TMHMM (31), respectively (http://www.cbs.dtu.dk/services/). The sequence of pVAPB1593 was compared with those of other plasmids by using tBLASTx (1), and the comparisons were visualized using the Artemis comparison tool (ACT) (14). Multiple DNA and protein sequence alignments were generated with ClustalW (16) (European Bioinformatics Institute, EMBL; http://www.ebi.ac.uk/) and T-Coffee (37) (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi). Informational noise was limited using a conservative annotation approach, allocating putative functions only when homology and protein domain searches provided consistent, unambiguous returns.

Nucleotide sequence accession number. The nucleotide sequence of the pVAPB1593 plasmid and the revised annotation of pVAPA1037 have been deposited in EMBL/GenBank under accession numbers AM947676 and AM947677, respectively.

RESULTS AND DISCUSSION

General description and annotation of pVAPB1593 and reannotation of pVAPA1037. The plasmid harbored by *R. equi* PAM 1593 is circular, is 79,251 bp in size, and contains 72 CDSs, including four pseudogenes, equivalent to a coding density of 78.7%. The overall structure of pVAPB1593 is very similar to that of the *vapA*-type plasmids from the equine isolates ATCC 33701 and 103 (50). As the p103 and p33701 plasmids are virtually identical in size (80,609 and 80,610 bp) and nucleotide sequence, they clearly are the same element and will be referred to hereafter simply as pVAPA1037. Comparative gene maps of pVAPB1593 and pVAPA1037 are shown in Fig. S1 in the supplemental material. Detailed gene annotation for pVAPB1593 is available in Table S1 in the supplemental material, together with a revised annotation for pVAPA1037 based on the same criteria. The reannotated pVAPA1037 sequence contains 73 CDSs, including eight pseudogenes. A function could be predicted for 39% of the proteins encoded by the plasmids (see Table S1 in the supplemental material), and 71% of the CDSs of pVAPB1593 (65% for pVAPA1037) had homologs in other bacteria. The greatest similarity was with products from other actinobacteria, particularly *Rhodococcus* spp. (55% of genes), followed by *Mycobacterium* spp. (6%) and *Nocardia* spp. (4%).

pVAPB1593 and pVAPA1037 differ in a variable region (VR). A segment corresponding to ≈75% of the plasmid sequence was highly conserved in pVAPB1593 and pVAPA1037 (95% DNA sequence identity). This section corresponded to the previously described conjugation and replication-and-partition regions (50) and displayed only minor differences in genetic structure, including two CDSs specific to pVAPB1593 (pVAPB_0101 and -0362) and two genes, one from each plasmid (pVAPB_0340 and pVAPA_0270), that had become corrupted (see Table S1 in the supplemental material).

In contrast, significant differences in DNA sequence (43%

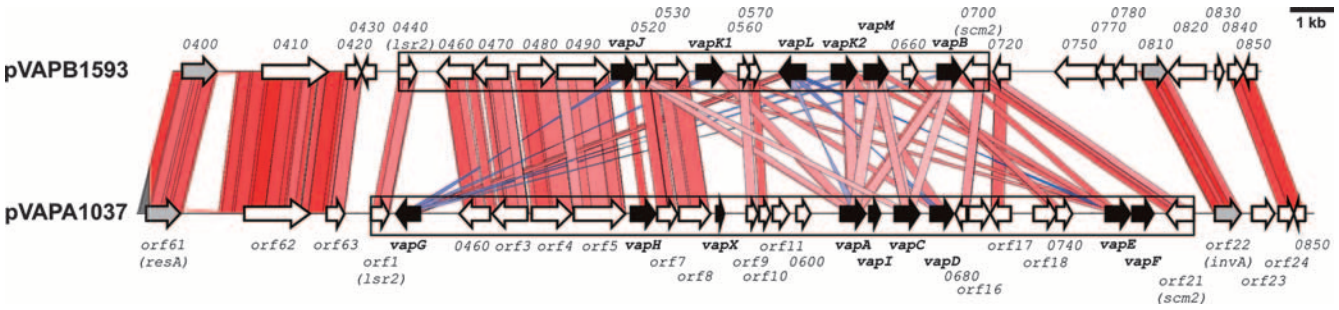

FIG. 1. Genetic structure and ACT analysis of the VR of the *R. equi* virulence plasmids pVAPB1593 and pVAPA1037. Regions with significant similarity (tBLASTx) are connected by colored lines (red, sequences in direct orientation; blue, sequences in reverse orientation). The intensity of the color indicates the strength of the sequence homology (pink/light blue, lowest; red/deep blue, highest). The virulence-associated *vap* genes are shown in black, and the mobility-related *resA*- and *invA*-like genes are shown in gray. The *vap* PAI, as defined based on Alien Hunter output (see Fig. S1 in the supplemental material), is boxed. Gene designations for pVAPB1593 are according to standardized annotation nomenclature adopted in this study for *R. equi* virulence plasmids (pVAP) (see Table S1 in the supplemental material), and those for pVAPA1037 are according to the nomenclature used by Takai et al. (50) (except for the newly identified genes, in which standardized nomenclature has been used) (see Table S1 in the supplemental material). The reference bar at the top right indicates 1 kb.

identity) and genetic structure were observed in a discrete region encompassing the PAI identified by Takai et al. in pVAPA1037 (50) (Fig. 1). Although clearly much more divergent than the conserved housekeeping backbone, a number of features of this VR were common to pVAPB1593 and pVAPA1037, such as the presence of *vap* genes (see below) and several conserved non-*vap* genes (see Table S1 in the supplemental material; also Fig. 1), indicating a common origin. Interestingly, most of the VR was clearly picked up, in both pVAPB1593 and pVAPA1037, by the Alien Hunter algorithm (http://www.sanger.ac.uk/Software/analysis/alien_hunter/) (see Fig. S1 in the supplemental material; also see below), which identifies horizontally acquired DNA by reliably capturing local compositional biases based on a variable-order motif distributions method (54).

Comparative analyses of DNA sequences and genetic structure based on ACT (Fig. 1) indicated that the VRs from pVAPB1593 and pVAPA1037 diversified through gene duplications, translocations, and inversions, together with insertion/deletion events. The major gene rearrangements involved the *vap* genes (Fig. 1). These genes have very similar nucleotide sequences (mean identity in the conserved region, 56%; range, 41 to 99%) (see Fig. S2 in the supplemental material) and therefore form direct repeats which may promote DNA recombination. An example in pVAPB1593 is provided by the *vapK1* gene and its almost exact duplicate *vapK2*, which, together with their flanking sequences, form an 873-bp-long perfect direct repeat (Fig. 1). There is also evidence of gene decay processes: pVAPA_0460, pVAPB_0560, and pVAPB_0720 are frameshifted and encode shorter products; pVAPB_0830 is truncated at its 3′ end; and pVAPA_vapF, -vapI, and -vapX show various degrees of degeneration (see below). Gene corruptions were in some cases associated with duplication events. Thus, pVAPA_0600 and -0740 are 5′-truncated and degenerate homologs, respectively, of pVAPB_0660, and pVAPA_0680 encodes an 87-residue degenerate portion of the pVAPA_0800 (*orf21*)/pVAPB_0700 product (see Table S1 in the supplemental material). Most of the pseudogenes identified in pVAPB1593 and pVAPA1037 were present in the VR.

DNA sequence conservation between pVAPB1593 and pVAPA1037 extended beyond the genes to the intergenic sequences in the plasmid backbone region. This was not the case in the VR, in which substantial intergenic sequence divergence was observed. This, together with the significantly lower level of CDS similarity between pVAPB1593/pVAPA1037 gene homologs (excluding the *vap* genes from the analysis) in the VR than in the conserved backbone (85.00 ± 8.75 versus 96.34 ± 8.78 mean percent amino-acid sequence identity; $P < 0.01$, Student's *t* test), indicates that these two DNA regions are subject to different evolutionary pressures in both plasmids.

Boundaries of the *vap* PAI. One of the primary criteria defining PAI elements is the atypical composition of their DNA with respect to the surrounding genetic environment, owing to their lateral acquisition (20, 24). Computational predictions (see above) localized the horizontal gene transfer event boundaries to the 5′ regions upstream from the *lsr2* gene (pVAPA/B_0440, or *orf1*) on the left and the *scm2* gene (pVAPB_0800/pVAPA_0700, or *orf21*) on the right (Fig. 1 and 2).

Thus, the *vap* PAI of pVAPB1593 would be 15.5 kb in size and encompass 17 CDSs, including six *vap* genes and one pseudogene, and that of pVAPB1037 would be 21.3 kb in size (not 27.5 kb as previously suggested by Takai et al. [50]) and encompass 26 CDSs, including six full-length *vap* genes, three *vap* pseudogenes (see below), and four other pseudogenes.

Despite the significant differences in genetic organization, a 6.2-kb region encompassing pVAPA/B_0460 to -0530 is conserved in the two *vap* PAIs (Fig. 1). Five of these CDSs (pVAPA/B_0480 to -0530) form the *virR* operon, with a *vap* gene in the middle (*vapHJ*) and a regulatory gene at each end: pVAPA/B_0480, or *virR*, encoding a LysR-type transcription factor, and pVAPA/B_0530, or *orf8*, encoding a response regulator (12). Both transcriptional regulators have been shown to be involved in controlling *vap* gene expression and virulence in pVAPA1037 (11, 12, 43, 45).

The conserved non-*vap* gene complement of the PAI also includes pVAPA/B_0440 (*lsr2*), encoding Lsr2, a DNA-bridging histone-like protein also present in *Mycobacterium* spp. and other actinobacteria (15); pVAPA/B_0470 (*orf23*), encoding a putative *S*-adenosylmethionine-dependent methyltransferase;

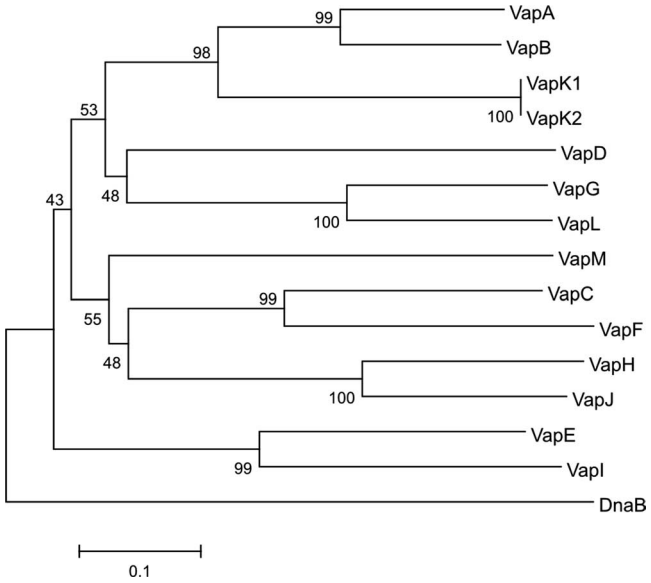

FIG. 2. Neighbor-joining unrooted phylogenetic tree of the *vap* multigene family constructed from a ClustalW alignment of the mature Vap proteins encoded by pVAPB1593 and pVAPA1037. For the analyses, a full-length VapF protein was reconstructed by correcting the two frameshift mutations in the 3′ region of the gene. Bootstrap values (10,000 replicates) are indicated at the nodes. Outgroup, putative DnaB protein from pVAPB1593 (pVAPB_0390). The bar indicates genetic distance.

pVAPA/B_490 (*orf5*), encoding a putative major facilitator superfamily transporter; and pVAPB_0700/pVAPA_0800 (*orf21*), encoding a protein with similarity to AroQ class chorismate mutase (CM) enzymes (see Table S1 in the supplemental material). The conservation of these genes in the PAIs of both the *vapA*- and the *vapB*-type virulence plasmids is intriguing and points to a potential role in pathogenesis. Particularly enigmatic is pVAPB_0700/pVAPA_0800 (*orf21*), here designated *scm2* (see Table S1 in the supplemental material). The *scm2* product belongs to an emerging family of proteins, designated *AroQ, which differ from the typical cytosolic AroQ CM enzymes involved in the shikimate pathway for aromatic amino acid biosynthesis (EC 5.4.99.5) in that they have a predicted signal sequence. *AroQ family members are encoded in the genomes of other bacterial pathogens, such as *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, and *Salmonella enterica* serovar Typhimurium and have been experimentally shown to be secreted and to have CM mutase activity (13, 30), and in *S. enterica* serovar Typhimurium, the coding gene was found to be selectively induced in vivo in infected mice, suggesting a role in virulence (9). Although the existence of a *AroQ-mediated periplasmic pathway for biosynthesis of phenylalanine from chorismate has been documented (13), and this may theoretically facilitate bacterial growth in the aromatic amino acid-deprived vacuolar compartment of the macrophage, mammals do not possess the shikimate pathway, and CM substrates (chorismate and prephenate) are unlikely to be present in animal tissues (44). Smc2 may share a common enzymatic activity with AroQ enzymes (ciclohexadienyl mutase) (13) but differ in substrate range and be involved in some unknown aspect of in vivo bacterial survival. Homologs of *lsr2* (two genes) and *scm2* (one gene, *scm1*) are present in the *R. equi* chromosome (our unpublished observations), and it is unclear whether there is functional redundancy between the chromosome- and plasmid-encoded products. The plasmidic *lsr2* and *scm2* genes may just be housekeeping, fitness-enhancing determinants which relocated to the PAI after the introgression of the horizontally acquired virulence-associated DNA in the pVAP ancestor.

The rest of the non-*vap* genes present in the PAI encode products of unknown functions (five in pVAPB1593, nine in pVAPA1037) (Fig. 1; also see Table S1 in the supplemental material).

New members of the *vap* multigene family. Six *vap* family genes were identified in the pVAPB1593 PAI: the previously described *vapB* (39, 49), plus five new *vap* multigene family members, which we named *vapJ*, *vapK1*, *vapK2*, *vapM*, and *vapL* (Fig. 1; also see Tables S1 and S2 in the supplemental material). All of the pVAPB1593 *vap* genes encode full-length Vap proteins with a potential signal sequence followed by an ≈50- to 70-residue nonconserved, structurally disordered region predicted to contain the lipidation site (10, 49, 50) and a conserved C-terminal region likely to form the functionally active domain of the protein (see Fig. S2 in the supplemental material).

The reannotated pVAPA1037 plasmid has a revised *vap* gene complement comprising nine CDSs (Fig. 1). Six of these encode full-length Vap proteins (*vapA*, *-C*, *-D*, *-E*, *-G*, and *-H*), and the other three correspond to *vap* genes displaying various degrees of degeneration: *vapF*, which has a 5′ truncation and two frameshift mutations at the 3′ end, resulting in an unsecreted Vap protein with a 39-residue nonconserved C terminus; *vapI* (11, 40), which is a 5′- and 3′-truncated *vap* gene encoding an 80-residue Vap polypeptide having no predicted signal sequence and lacking most of the conserved C-terminal domain; and a highly corrupted *vap* pseudogene (45), here designated *vapX*, encoding a 28-residue peptide from the central region of the conserved Vap C-terminal domain (see Table S2 in the supplemental material).

Overall, the pVAPB1593-encoded Vap proteins display a mean of 37.71 to 41.42% sequence identity (range, 16 to 76%) to the pVAPA1037-encoded Vaps. The interplasmid Vap sequence identities were similar to the mean internal Vap sequence identities for each plasmid, suggesting that Vap family protein diversification is subject to similar evolutionary forces in the two pVAP plasmids. However, Vap sequence variability was somewhat greater within pVAPA1037 (mean identity of 21.66 to 38.5%; range, 8 to 48%) than within pVAPB1593 (mean identity 37.75 to 43.5%; range, 33 to 99%) (see Table S3 in the supplemental material).

Evolution of the *vap* genes. The highest sequence identity scores in pairwise comparisons between all of the Vaps from pVAPB1593 and pVAPA1037 corresponded to proteins from different plasmids in some cases and to proteins from the same plasmid in others (see Table S3 in the supplemental material). We investigated the evolution and phylogenetic relationships of the *vap* multigene family in pVAPB1593 and pVAPA1037 by constructing a multiple alignment and a neighbor-joining tree for the mature Vap proteins, using the amino acid p-distance model (Fig. 2). This procedure was repeated with the aligned C-terminal conserved domains and with the *vapI* and

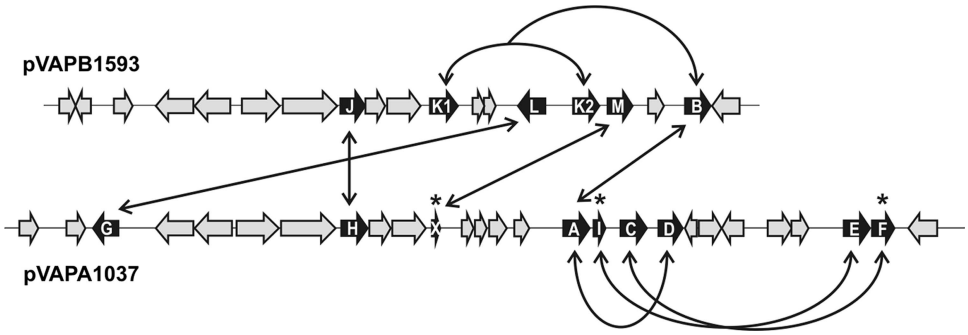

FIG. 3. Model of the evolutionary dynamics of the *vap* multigene family. On a schematic representation of the *vap* PAIs from pVAPB1593 and pVAPA1037, *vap* genes (in black) presumably derived from vertical evolution of a preexisting common ancestral determinant are connected by straight arrows, and those probably originated by gene duplication in the specific plasmid are connected by curved arrows (see the text for details). The connection between *vapM* and the *vapX* pseudogene was deduced from a phylogenetic tree constructed with the 28-residue VapX product aligned with the corresponding protein fragments from other members of the Vap family (not shown). Asterisks indicate degenerate *vap* genes; note in pVAPA1037 that one gene in each of the duplicated *vapIE* and *vapCF* tandems is undergoing decay.

vapX pseudogene products and the corresponding fragments of all the Vap proteins (not shown). Comparing the results obtained with the genetic structure and synteny of the PAIs from pVAPB1593 and pVAPA1037 made it possible to reconstruct the probable evolutionary dynamics of the *vap* multigene family in these plasmids (Fig. 3).

Thus, *vapA* and *vapB* seem to be divergent allelic variants of the same ancestral *vap* gene, consistent with the observed mutual exclusivity of their corresponding PCR gene markers among *R. equi* isolates (38, 39). Similarly, *vapG* and *vapL*, both of which are transcribed in the opposite direction relative to the other *vap* genes, appear to be derived from a common ancestral gene despite mapping to very different sites in their corresponding PAIs (Fig. 3). The same would apply to *vapH* and *vapJ*, in this case clearly supported by the conserved synteny of the surrounding *virR* operon genes in pVAPA1037 and pVAPB1593. VapK is closely related to VapAB and therefore probably arose by an initial duplication of the ancestral *vapAB* gene early in the evolution of pVAPB1593, followed by a second, relatively recent duplication of the *vapK* locus to generate two almost identical gene copies. It appears that *vapD* is the counterpart of *vapK* in pVAPA1037 and that *vapE* and the *vapF* pseudogene arose by en block duplication of *vapI* (currently also a pseudogene) and *vapC* in pVAPA1037. Finally, the product of the *vapX* pseudogene is most closely related to VapM (Fig. 3). In summary, the PAI of the last common ancestor of pVAPB1593 and pVAPA1037 probably contained four precursor *vap* genes (from left to right, *vapHJ*, *vapXM*, *vapGL*, and *vapAB*) (Fig. 3); these underwent further duplication and sequence diversification (and in some cases decay) events in different plasmid subpopulations, possibly in response to specific host-driven evolutionary forces. Considering synteny and phylogenetic data, *vap* gene evolution was clearly accompanied by substantial genetic rearrangements in the PAI (Fig. 1).

Conserved architecture for the large, circular rhodococcal plasmids: common housekeeping backbone and horizontally acquired VR. About 40% of the pVAPB1593 and pVAPA1037 genes had homologs in the only other example of a large, circular plasmid sequenced to date in the genus *Rhodococcus*, the 104-kb pREC1 plasmid from strain PR4 of the environmental species *Rhodococcus erythropolis* (47). The pVAP backbone can be subdivided into three modules: (i) a conjugation module, with genes encoding a putative methylase/helicase, TraA and TraG/TraD conjugal transfer protein homologs, and a type I DNA topoisomerase; (ii) an unknown-function model, encoding hypothetical proteins, many of which may be membrane associated; and (iii) a replication/partition module, with genes encoding ParA, DnaB, and ParB gene homologs (see Fig. S1 and Table S1 in the supplemental material). A multiple comparison by ACT showed that the pREC1 and pVAP plasmids are colinear, with two regions of similarity: one highly conserved and syntenic, encompassing the conjugation and unknown-function modules, and the other much less conserved, corresponding to the replication/partition module (Fig. 4; also see Fig. S1 and Table S1 in the supplemental material). There is in addition a region of dissimilarity, or a VR. This pREC1-specific VR, which encodes products involved in the β-oxidation of fatty acids, together with many hypothetical proteins, was also identified as laterally acquired DNA by Alien Hunter analysis (see Fig. S1 in the supplemental material). Unlike the pVAP VR, it is not immediately adjacent to the conserved backbone but is inserted in the middle of the replication module, splitting it into two sections; thus, in pREC1 the *rep* and *parA* genes lie ≈37 kb apart from the *parB* and *dnaB* genes. Interestingly, the left section of the pREC1 replication module was also identified as laterally acquired DNA, together with the contiguous VR (see Fig. S1 in the supplemental material). In addition, the Rep and ParA proteins encoded by this left section of the pREC1 replication module are unrelated to their pVAP counterparts (no significant protein similarity detected with BLASTp), whereas the ParB and DnaB proteins from pREC1 and the pVAP plasmids are homologous (43% and 32% identities, respectively, similar to the mean level of identity for pVAP/pREC1 homologs) (see Table S1 in the supplemental material). Therefore, although gene synteny is conserved, the left section of the pREC1 replication/partition module seems to be derived from a replicon different from that of the pVAP plasmids. The replication region of the pVAP/pREC1 backbone seems thus to be a hot spot for the integration of foreign DNA. Consistent with this, a putative phage excisionase/recombinase gene is present, at

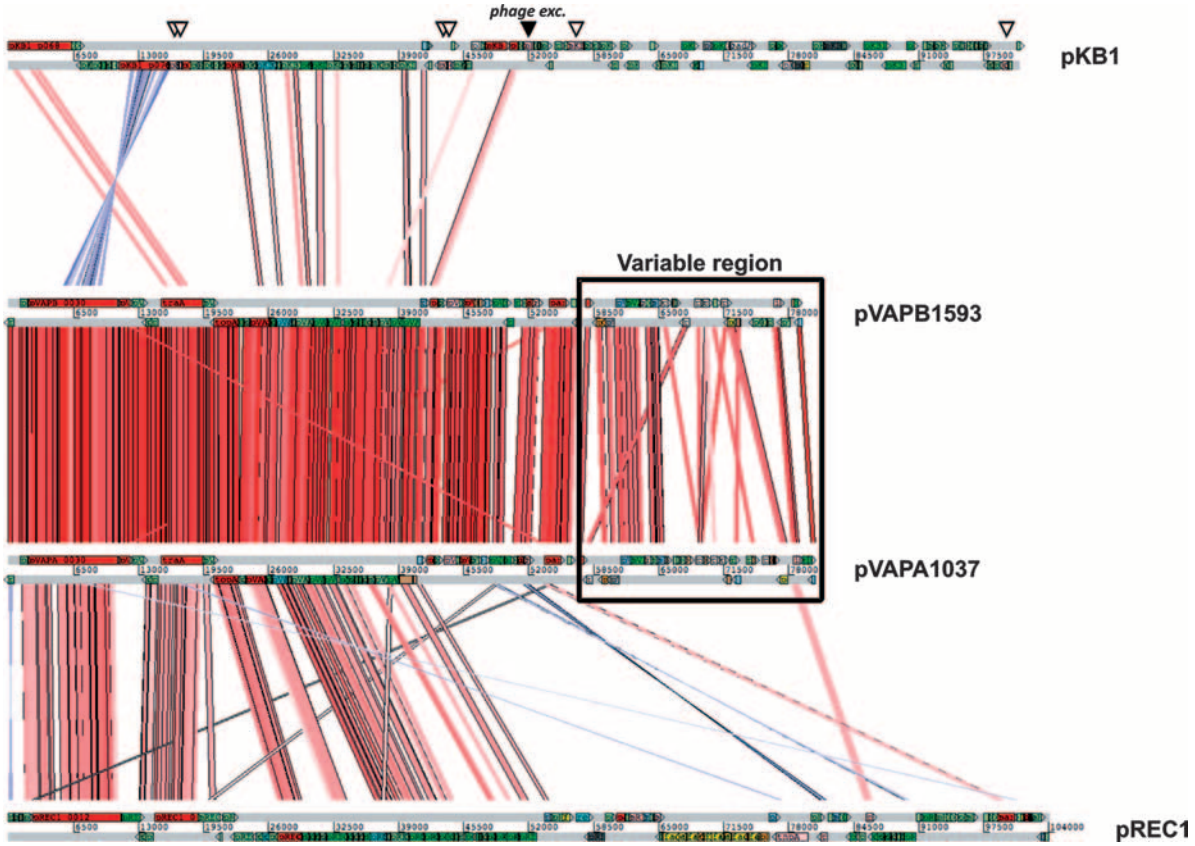

FIG. 4. ACT analysis of the rhodococcal pVAP and pREC1 plasmids and *Gordonia westfalica* pKB1 plasmid (8). See the legends to Fig. 1 and to Fig. S1 in the supplemental material for an explanation of gene color codes and ACT output. To facilitate the visualization of the colinearity and gene synteny, the plasmids were aligned on the first conserved gene of the backbone. Arrowheads indicate the mobility genes present in pKB1 (the black arrowhead shows the conserved phage excisionase [phage exc.] gene also present in the rhodococcal plasmids) (see Fig. S1 and Table S4 in the supplemental material).

an identical position downstream from the *parA* gene, in the replication modules of both the pVAP (CDS 0310) and the pREC1 (CDS 0059) plasmids (see Fig. S1 and Table S1 in the supplemental material).

Interestingly, putative DNA resolvase (*resA*-like) and invertase/resolvase (*invA*-like) genes, normally found in transposons and other mobile genetic elements, are present at or near the end of the pVAP VR (see Fig. S1 and Table S1 in the supplemental material). Homologous genes are also present in pERC1, not flanking—but rather located within—the VR, close to each other, together with additional DNA mobilization/recombinase determinants (see Fig. S1 and Table S1 in the supplemental material). This suggests an evolutionary model in which the initial insertion of a mobile element or elements carrying *resA* and *invA* homologs in an ancestor of the rhodococcal plasmids provided a flexible landing platform for the horizontal acquisition of DNA. Unlike the conjugal transfer/plasmid replication backbone, which is constrained by obvious housekeeping selective pressures for conservation, the newly acquired alien DNA would have been free to evolve under various niche-specific pressures, leading to functional diversification, as found in the contemporary niche-adapted VRs of the pVAP and pREC1 plasmids.

The *R. equi* pVAP plasmids belong to a distinct family of actinobacterial replicons. We finally investigated the possible presence of plasmids similar to the *R. equi* pVAP virulence plasmids and *R. erythropolis* pREC1 in other bacteria. We used the products of the conserved unknown-function module as a search signature and the conservation of backbone synteny as an identity criterion. A *c*onjugation–*u*nknown-function–*r*eplication/partition–*V*R (CURV) modular arrangement was identified in the pKB1 plasmid from another mycolata species, *Gordonia westfalica* (Fig. 4) (8). pKB1 is also circular, similar in size to the rhodococcal plasmids (101 kb), and its conserved backbone encodes proteins 36% identical on average (range, 25 to 65%) to the corresponding pVAP homologs. The conserved genes include a pVAPA/B_0310/pREC1_0059 (putative phage excisionase) homolog located in a similar position in the replication module, providing further evidence of a common origin for pKB1 and the rhodococcal plasmids. The pKB1-specific VR encodes products putatively involved in heavy metal ion detoxification and redox processes, together with many hypothetical or membrane proteins (8). As in the rhodococcal plasmids, the VR is also flanked by mobility genes, but their putative products (a transposase and a bacteriophage-related protein) are different from the *resA*- and *invA*-like gene

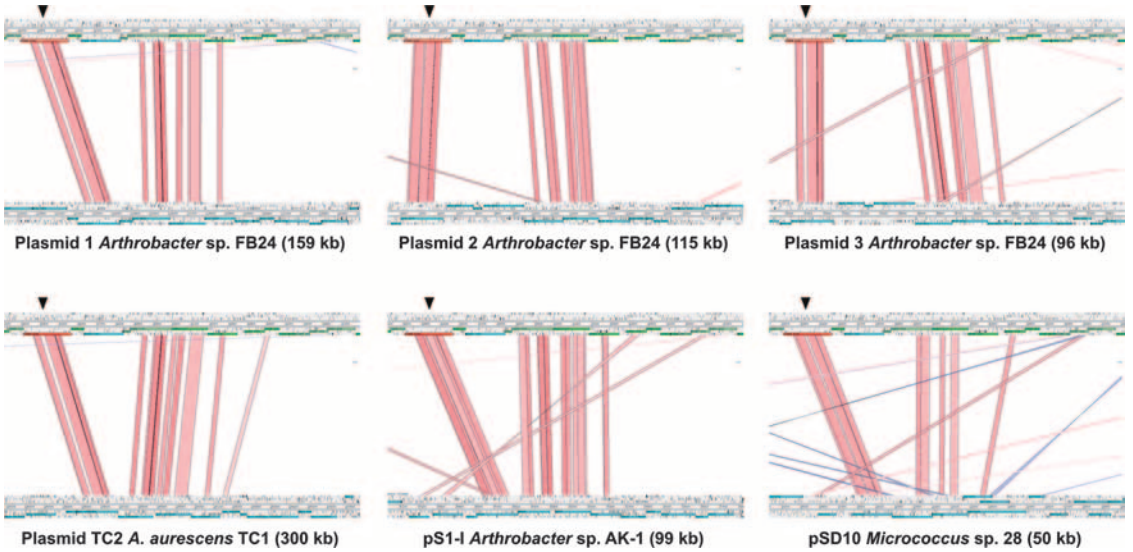

FIG. 5. ACT analysis of the conserved signature region of the CURV plasmids, a new family of large, circular actinobacterial plasmids. This highly syntenic region includes the conserved gene encoding a putative conjugal transfer protein, TraG/TraD (indicated by an arrow), and spans the second half of the conjugation and the first half of the unknown-function plasmid modules (see the text for details). An ≈15-kb fragment of each of the indicated plasmids from *Arthrobacter* and *Micrococcus* spp. (total size in brackets) was compared with the corresponding fragment from pVAPB1593 (above). See also Table S4 in the supplemental material.

products (Fig. 4). Thus, functional plasticity in this group of plasmids involves a VR which is acquired horizontally through diverse mechanisms.

The similar modular architectures of the pVAP, pREC1, and pKB1 plasmids, with a conserved syntenic backbone containing a cluster of unique genes of unknown functions near the conjugation module, suggests that they belong to one family of large, circular replicons. The CURV signature was also identified in a number of large, circular plasmids from *Arthrobacter* spp. (e.g., *Arthrobacter aurescens* plasmid TC2, *Arthrobacter* sp. strain FB24 plasmids 1, 2, and 3, or *Arthrobacter* sp. strain AK-1 plasmid pSI-1) (29) and in the pSD10 circular plasmid from *Micrococcus* sp. strain 28 (NCBI accession no. NC_004954), but surrounded by a much more degenerate conserved backbone (see Table S4 in the supplemental material), consistent with the greater phylogenetic distance separating these actinobacterial genera from *Rhodococcus* (53). All these plasmids seem to belong to an ancient family of replicons that was widespread among the actinobacterial ancestors and diversified during the speciation process. A distinctive feature of this CURV plasmid family is the presence of a highly conserved syntenic gene cluster which always includes a conserved conjugal transfer TraG/TraD protein-coding gene and spans the second and first halves of the conjugation and unknown-function modules (Fig. 5).

Conclusions. Our previous findings suggested a model in which *R. equi* host tropism is determined by the type of pVAP plasmid harbored by the bacteria (38). Consistent with this model, we recently found that a specific *R. equi* chromosomal pulsed-field gel electrophoresis macrorestriction type can be isolated from different animal species in association with the corresponding host-adapted plasmid type (A. Ocampo-Sosa, D. Lewis, and J. A. Vazquez-Boland, unpublished results). The complete *vapB*-type plasmid sequence reported here and its comparison with the *vapA*-type plasmid sequence provide valuable insight into the evolution of the virulence-associated *vap* PAI and the possible mechanisms underlying host tropism in *R. equi*. We show that the *vapA*- and *vapB*-type virulence plasmids, each associated with a specific nonhuman animal host (38), share a conserved backbone but differ in their *vap* PAIs. The *vapB* PAI encodes a set of new different Vap proteins, suggesting that the specific *vap* multigene family complement carried by the virulence plasmid may play a critical role in the specificity of the interaction of *R. equi* with its animal hosts. Sequence analyses indicated that the *vap* genes had evolved by gene duplication and sequence diversification and are responsible for the genetic plasticity of the PAI through their recombination-prone, duplicated sequences. Comparison of the sequences of the two *R. equi* virulence plasmids with those of a third rhodococcal plasmid from the nonpathogenic species *R. erythropolis*, and with other circular plasmids from more-distantly related mycolata (*Gordonia*) and nonmycolata (*Arthrobacter* and *Micrococcus*), led to the identification of a new group of large extrachromosomal elements widespread among actinobacteria. This family of circular plasmids, the CURV family, is characterized by an ancient housekeeping backbone—presumably transferable by conjugation—and a horizontally acquired plasticity region encoding niche-adaptive functions.

ACKNOWLEDGMENTS

This work was supported by the Horserace Betting Levy Board (grant vet/prj/712), the NDP structural fund 2002-2006, the Research Stimulus Fund (grant RSF 06-379), and the Grayson-Jockey Club Research Foundation.

REFERENCES

1. **Abbott, J. C., D. M. Aanensen, and S. D. Bentley.** 2007. WebACT: an online genome comparison suite. Methods Mol. Biol. **395:**57–74.

2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy.** 2004. The Pfam protein families database. Nucleic Acids Res. **32:**D138–D141.
4. **Bell, K. S., J. C. Philp, D. W. Aw, and N. Christofi.** 1998. The genus *Rhodococcus*. J. Appl. Microbiol. **85:**195–210.
5. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
6. **Benoit, S., A. Benachour, S. Taouji, Y. Auffray, and A. Hartke.** 2001. Induction of *vap* genes encoded by the virulence plasmid of *Rhodococcus equi* during acid tolerance response. Res. Microbiol. **152:**439–449.
7. **Bentley, S. D., C. Corton, S. E. Brown, A. Barron, L. Clark, J. Doggett, B. Harris, D. Ormond, M. A. Quail, G. May, D. Francis, D. Knudson, J. Parkhill, and C. A. Ishimaru.** 2008. Genome of the actinomycete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. J. Bacteriol. **190:**2150–2160.
8. **Bröker, D., M. Arenskötter, A. Legatzki, D. H. Nies, and A. Steinbüchel.** 2004. Characterization of the 101-kilobase-pair megaplasmid pKB1, isolated from the rubber-degrading bacterium *Gordonia westfalica* Kb1. J. Bacteriol. **186:**212–225.
9. **Bumann, D.** 2002. Examination of *Salmonella* gene expression in an infected mammalian host using the green fluorescent protein and two-colour flow cytometry. Mol. Microbiol. **43:**1269–1283.
10. **Byrne, B. A., J. F. Prescott, G. H. Palmer, S. Takai, V. M. Nicholson, D. C. Alperin, and S. A. Hines.** 2001. Virulence plasmid of *Rhodococcus equi* contains inducible gene family encoding secreted proteins. Infect. Immun. **69:**650–656.
11. **Byrne, G. A., C. A. Boland, E. P. O'Connell, and W. G. Meijer.** 2008. Differential mRNA stability of the *vapAICD* operon of the facultative intracellular pathogen *Rhodococcus equi*. FEMS Microbiol. Lett. **280:**89–94.
12. **Byrne, G. A., D. A. Russell, X. Chen, and W. G. Meijer.** 2007. Transcriptional regulation of the *virR* operon of the intracellular pathogen *Rhodococcus equi*. J. Bacteriol. **189:**5082–5089.
13. **Calhoun, D. H., C. A. Bonner, W. Gu, G. Xie, and R. A. Jensen.** 2001. The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella typhimurium* and *Pseudomonas aeruginosa*. Genome Biol. **2:**research0030.1–0030.16.
14. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21:**3422–3423.
15. **Chen, J. M., H. Ren, J. E. Shaw, Y. J. Wang, M. Li, A. S. Leung, V. Tran, N. M. Berbenetz, D. Kocincova, C. M. Yip, J. M. Reyrat, and J. Liu.** 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. Nucleic Acids Res. **36:**2123–2135.
16. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson.** 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. **31:**3497–3500.
17. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
18. **Embley, T. M., and E. Stackebrandt.** 1994. The molecular phylogeny and systematics of the actinomycetes. Annu. Rev. Microbiol. **48:**257–289.
19. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
20. **Gal-Mor, O., and B. B. Finlay.** 2006. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell. Microbiol. **8:**1707–1719.
21. **Giguère, S., M. K. Hondalus, J. A. Yager, P. Darrah, D. M. Mosser, and J. F. Prescott.** 1999. Role of the 85-kilobase plasmid and plasmid-encoded virulence-associated protein A in intracellular survival and virulence of *Rhodococcus equi*. Infect. Immun. **67:**3548–3557.
22. **Goethals, K., D. Vereecke, M. Jaziri, M. Van Montagu, and M. Holsters.** 2001. Leafy gall formation by *Rhodococcus fascians*. Annu. Rev. Phytopathol. **39:**27–52.
23. **Gurtler, V., B. C. Mayall, and R. Seviour.** 2004. Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? FEMS Microbiol. Rev. **28:**377–403.
24. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54:**641–679.
25. **Hondalus, M. K.** 1997. Pathogenesis and virulence of *Rhodococcus equi*. Vet. Microbiol. **56:**257–268.
26. **Hondalus, M. K., and D. M. Mosser.** 1994. Survival and replication of *Rhodococcus equi* in macrophages. Infect. Immun. **62:**4167–4175.
27. **Hulo, N., C. J. Sigrist, V. Le Saux, P. S. Langendijk-Genevaux, L. Bordoli, A. Gattiker, E. De Castro, P. Bucher, and A. Bairoch.** 2004. Recent improvements to the PROSITE database. Nucleic Acids Res. **32:**D134–D137.
28. **Jain, S., B. R. Bloom, and M. K. Hondalus.** 2003. Deletion of *vapA* encoding Virulence Associated Protein A attenuates the intracellular actinomycete *Rhodococcus equi*. Mol. Microbiol. **50:**115–128.
29. **Jerke, K., C. H. Nakatsu, F. Beasley, and A. Konopka.** 2008. Comparative analysis of eight *Arthrobacter* plasmids. Plasmid **59:**73–85.
30. **Kim, S. K., S. K. Reddy, B. C. Nelson, G. B. Vasquez, A. Davis, A. J. Howard, S. Patterson, G. L. Gilliland, J. E. Ladner, and P. T. Reddy.** 2006. Biochemical and structural characterization of the secreted chorismate mutase (Rv1885c) from *Mycobacterium tuberculosis* $H_{37}R_v$: an *AroQ enzyme not regulated by the aromatic amino acids. J. Bacteriol. **188:**8638–8648.
31. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
32. **Larkin, M. J., R. De Mot, L. A. Kulakov, and I. Nagy.** 1998. Applied aspects of *Rhodococcus* genetics. Antonie van Leeuwenhoek **74:**133–153.
33. **Lührmann, A., N. Mauder, T. Sydor, E. Fernandez-Mora, J. Schulze-Luehrmann, S. Takai, and A. Haas.** 2004. Necrotic death of *Rhodococcus equi*-infected macrophages is regulated by virulence-associated plasmids. Infect. Immun. **72:**853–862.
34. **McLeod, M. P., R. L. Warren, W. W. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. Jones, R. Holt, F. S. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn, and L. D. Eltis.** 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc. Natl. Acad. Sci. USA **103:**15582–15587.
35. **Meijer, W. G., and J. F. Prescott.** 2004. *Rhodococcus equi*. Vet. Res. **35:**383–396.
36. **Muscatello, G., D. P. Leadon, M. Klayt, A. Ocampo-Sosa, D. A. Lewis, U. Fogarty, T. Buckley, J. R. Gilkerson, W. G. Meijer, and J. A. Vazquez-Boland.** 2007. *Rhodococcus equi* infection in foals: the science of 'rattles'. Equine Vet. J. **39:**470–478.
37. **Notredame, C., D. G. Higgins, and J. Heringa.** 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205–217.
38. **Ocampo-Sosa, A. A., D. A. Lewis, J. Navas, F. Quigley, R. Callejo, M. Scortti, D. P. Leadon, U. Fogarty, and J. A. Vazquez-Boland.** 2007. Molecular epidemiology of *Rhodococcus equi* based on *traA*, *vapA*, and *vapB* virulence plasmid markers. J. Infect. Dis. **196:**763–769.
39. **Oldfield, C., H. Bonella, L. Renwick, H. I. Dodson, G. Alderson, and M. Goodfellow.** 2004. Rapid determination of *vapA*/*vapB* genotype in *Rhodococcus equi* using a differential polymerase chain reaction method. Antonie van Leeuwenhoek **85:**317–326.
40. **Polidori, M., and A. Haas.** 2006. VapI, a new member of the Rhodococcus equi Vap family. Antonie van Leeuwenhoek **90:**299–304.
41. **Prescott, J. F.** 1991. *Rhodococcus equi*: an animal and human pathogen. Clin. Microbiol. Rev. **4:**20–34.
42. **Qiagen.** 2005. Very low-copy plasmid/cosmid purification protocol. Qiagen plasmid purification handbook, 3rd ed. Qiagen, Hilden, Germany.
43. **Ren, J., and J. F. Prescott.** 2004. The effect of mutation on Rhodococcus equi virulence plasmid gene expression and mouse virulence. Vet. Microbiol. **103:**219–230.
44. **Richards, T. A., J. B. Dacks, S. A. Campbell, J. L. Blanchard, P. G. Foster, R. McLeod, and C. W. Roberts.** 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot. Cell **5:**1517–1531.
45. **Russell, D. A., G. A. Byrne, E. P. O'Connell, C. A. Boland, and W. G. Meijer.** 2004. The LysR-type transcriptional regulator VirR is required for expression of the virulence gene *vapA* of *Rhodococcus equi* ATCC 33701. J. Bacteriol. **186:**5576–5584.
46. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945.
47. **Sekine, M., S. Tanikawa, S. Omata, M. Saito, T. Fujisawa, N. Tsukatani, T. Tajima, T. Sekigawa, H. Kosugi, Y. Matsuo, R. Nishiko, K. Imamura, M. Ito, H. Narita, S. Tago, N. Fujita, and S. Harayama.** 2006. Sequence analysis of three plasmids harboured in *Rhodococcus erythropolis* strain PR4. Environ. Microbiol. **8:**334–346.
48. **Sekizaki, T., S. Takai, Y. Egawa, T. Ikeda, H. Ito, and S. Tsubaki.** 1995. Sequence of the *Rhodococcus equi* gene encoding the virulence-associated 15-17-kDa antigens. Gene **155:**135–136.
49. **Takai, S., T. Anzai, Y. Fujita, O. Akita, M. Shoda, S. Tsubaki, and R. Wada.** 2000. Pathogenicity of *Rhodococcus equi* expressing a virulence-associated 20 kDa protein (VapB) in foals. Vet. Microbiol. **76:**71–80.
50. **Takai, S., S. A. Hines, T. Sekizaki, V. M. Nicholson, D. A. Alperin, M. Osaki, D. Takamatsu, M. Nakamura, K. Suzuki, N. Ogino, T. Kakuda, H. Dan, and J. F. Prescott.** 2000. DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. Infect. Immun. **68:**6840–6847.
51. **Takai, S., M. Iie, Y. Watanabe, S. Tsubaki, and T. Sekizaki.** 1992. Virulence-associated 15- to 17-kilodalton antigens in *Rhodococcus equi*: temperature-dependent expression and location of the antigens. Infect. Immun. **60:**2995–2997.
52. **Takai, S., T. Sekizaki, T. Ozawa, T. Sugawara, Y. Watanabe, and S. Tsubaki.** 1991. Association between a large plasmid and 15- to 17-kilodalton antigens in virulent *Rhodococcus equi*. Infect. Immun. **59:**4056–4060.

53. **Ventura, M., C. Canchaya, A. Tauch, G. Chandra, G. F. Fitzgerald, K. F. Chater, and D. van Sinderen.** 2007. Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. Microbiol. Mol. Biol. Rev. **71:** 495–548.
54. **Vernikos, G. S., and J. Parkhill.** 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics **22:**2196–2203.
55. **Warren, R., W. W. Hsiao, H. Kudo, M. Myhre, M. Dosanjh, A. Petrescu, H. Kobayashi, S. Shimizu, K. Miyauchi, E. Masai, G. Yang, J. M. Stott, J. E. Schein, H. Shin, J. Khattra, D. Smailus, Y. S. Butterfield, A. Siddiqui, R. Holt, M. A. Marra, S. J. Jones, W. W. Mohn, F. S. Brinkman, M. Fukuda, J. Davies, and L. D. Eltis.** 2004. Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus* sp. strain RHA1. J. Bacteriol. **186:**7783–7795.