Pangenome and phylogenomic analysis of the pathogenic actinobacterium Rhodococcus equi

Citation for published version:

Digital Object Identifier (DOI):
10.1093/gbe/evw222

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genome Biology and Evolution

Publisher Rights Statement:
© The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Pangenome and Phylogenomic Analysis of the Pathogenic Actinobacterium *Rhodococcus equi*

Elisa Anastasi¹,†, Iain MacArthur¹,‡, Mariela Scortti¹,², Sonsiray Alvarez¹, Steeve Giguère³, and José A. Vázquez-Boland¹,²,⁴,*

¹Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom
²Edinburgh Medical School (Biomedical Sciences), University of Edinburgh, Edinburgh, United Kingdom
³Department of Large Animal Medicine, University of Georgia, Georgia, USA
⁴Grupo de Patogenómica Bacteriana, Universidad de León, León, Spain

†These authors contributed equally to the study.
*Corresponding author: E-mail: v.boland@ed.ac.uk.

Accepted: September 2, 2016

Data deposition: This project has been deposited at GenBank under accessions LWTV00000000, LWIC00000000, LWBN00000000, LWHS00000000, LWHT00000000, LWJH00000000, LWJH00000000, LWKH00000000, LWKH00000000, LWHY00000000, LWJH00000000, LWIA00000000, LWIB00000000, LWTO00000000, LXFH00000000, LXFG00000000, LWHR00000000, LWTP00000000, LWV00000000, LWTV00000000, LWTR00000000, LWTS00000000, LWTT00000000, LWTV00000000, and LWTV00000000.

Abstract

We report a comparative study of 29 representative genomes of the animal pathogen *Rhodococcus equi*. The analyses showed that *R.* *equi* is genetically homogeneous and clonal, with a large core genome accounting for ≈80% of an isolates’ gene content. An open pangenome, even distribution of accessory genes among the isolates, and absence of significant core–genome recombination, indicated that gene gain/loss is a major driver of *R.* *equi* genome evolution. Traits previously predicted to be important in *R.* *equi* physiology, virulence and niche adaptation were part of the core genome. This included the lack of a phospho-enol-pyruvate:carbohydrate transport system (PTS), unique among the rhodococci except for the closely related *Rhodococcus defluvii*, reflecting selective PTS gene loss in the *R.* *equi*–*R.* *defluvii* sublineage. Thought to be asaccharolytic, *rbsCB* and *glcP* non-PTS sugar permease homologues were identified in the core genome and, albeit inefficiently, *R.* *equi* utilized their putative substrates, ribose and (irregularly) glucose.

There was no correlation between *R.* *equi* whole-genome phylogeny and host or geographical source, with evidence of global spread of genomvars. The distribution of host-associated virulence plasmid types was consistent with the exchange of the plasmids (and corresponding host shifts) across the *R.* *equi* population, and human infection being zoonotically acquired. Phylogenomic analyses demonstrated that *R.* *equi* occupies a central position in the *Rhodococcus* phylogeny, not supporting the recently proposed transfer of the species to a new genus.

Key words: *Rhodococcus equi*, pangenome analysis, comparative genomics, genome diversity and evolution, phylogenomics, *Corynebacteriales*, *Actinobacteria*.

Introduction

The soil-dwelling actinobacterium *Rhodococcus equi* is the causative agent of a purulent bronchopneumonic disease that affects foals in equine farms worldwide. In addition to horses, *R.* *equi* can also infect other animal species and is associated with severe opportunistic infections in immunocompromised people (Prescott 1991; von Bargen and Haas 2009; Vazquez-Boland et al. 2013). We previously reported the complete genome sequence of an equine isolate of *R.* *equi* (strain 103S). This work provided key information about the genome structure of the pathogen and the mechanisms of rhodococcal niche-adaptive genome plasticity and virulence evolution (Letek et al. 2010). Here we present the first comprehensive comparative genomic analysis of *R.* *equi*, involving...
multiple isolates from different sources. Our new study provides insight into the core features, diversity, population structure and genome evolution of R. equi. It also clarifies the phylogenetic position of the species, repeatedly questioned based on equivocal 16S rDNA and numerical phenetic studies (Goodfellow et al. 1998; Gurtler et al. 2004; Jones and Goodfellow 2012; Jones et al. 2013b), unambiguously confirming R. equi is a bona fide member of the genus Rhodococcus.

Materials and Methods

Bacteria

The isolates sequenced in this study (supplementary table S1, Supplementary Material online) were selected to include at least two representatives from each of the seven major R. equi genogroups defined by Asel PFGE genotyping (Vazquez-Boland et al. 2008 and our unpublished data) plus the type strain of the species, DSM 20307T (= ATCC 6939T = ATCC 25729T = NBRC 101255T). Isolates of different animal sources (equine, bovine, porcine, ovine, human), geographical origin (13 countries) and host-associated virulence plasmid type carriage (pVAPA, pVAPB, pVAPN) (Takai et al. 2000; Letek et al. 2008; Valero-Rello et al. 2015) were analyzed.

Genome Sequencing and Analysis

Rhodococcus equi DNA was isolated from exponential cultures in BHI (OD600 ≈ 1.0) using the GenElute™ kit (Sigma–Aldrich). Shotgun 101-bp pair-end DNA sequencing was performed at Beijing Genomics Institute (BGI, China) using TruSeq DNA PCR-Free Sample library preparation kit on Illumina HiSeq 2000 instruments. Strains 2274 to 2288 (supplementary table S1, Supplementary Material online) were sequenced at the genomics facility of the University of Georgia (USA) as previously described (Anastasi et al. 2015). Adaptors and low quality reads were trimmed using Scythe (https://github.com/vsbuffalo/scythe) and Sickle (https://github.com/najoshi/sickle), respectively, and assembled using SPAdes (Bankevich et al. 2012). Annotation was performed using Prokka V1.11 (Seemann 2014) and the complete 103S genome (Letek et al. 2010) was analyzed. Twenty-seven de novo determined R. equi whole-genome shotgun assemblies, the available draft genome of ATCC 33707, and the complete 103S genome (Letek et al. 2010) were analyzed (supplementary table S1, Supplementary Material online). The average CDS number was 4,933 (range 4,525–5,325), similar to the gene content of the manually annotated 5.04-Mbp 103S genome (Gorin et al. 2007; Richter and Rosselló-Móra 2009; Kim et al. 2014). This corresponded to 100% 16S rDNA sequence identity (1,519 nt) across all the isolates.

In comparison, the ANI values with members of the two other main Rhodococcus lines of descent as defined based on 16 rDNA phylogenies (McMinn et al. 2000; Jones and Goodfellow 2012), that is, the “erythropolis” clade (R. erythropolis, R. jostii, R. opacus and R. fascians included in the analysis) and the “Rhodochrous” clade (Rhodococcus rhodochrous, Rhodococcus rhodniii, Rhodococcus ruber, and Rhodococcus pyridinivorans included in the analysis), were 72.27–74.58% and 68.55–75.15%, respectively. The ANI with the recently described R. equi close relative, Rhodococcus defluvi (strain Ca11') (Kampfer et al. 2014), was 82.96%. This corresponded to 16S rDNA identity values of 96–98% and 95–97% for representatives of the “erythropolis” and “Rhodochrous” clades, respectively, and 99% for R. defluvi.

The above data correlate with a strong degree of genome similarity and synteny conservation in BLASTn alignments (fig. 1), indicating that R. equi is a genetically homogeneous species.

Results and Discussion

Rhodococcus equi Is Genetically Homogeneous

The core genome shared by all 29 R. equi strains comprises 3,858 homologous gene clusters (HGC) (fig. 1A), equivalent to 81.5% of the 103S genome or 78.2% of the average gene content of the analyzed isolates, reflecting a low degree of intraspecies genomic variability. A core genome size estimation plot starts plateauing at about 25–27 genomes (fig. 2B).
indicating that the number of core genes is close to its maximum. The core genome contributes to 47.21% of the pangenome of the analyzed isolates ($n = 8,174$ HGCs). About 35% of the pangenome is constituted by “cloud” HGCs, with a predominance of genes present in only one genome (fig. 2A), accounting for the species’ genome variability. This is consistent with the pangenome size plot, which increases almost linearly as new genomes are added (fig. 2B). The 4,316 HGC of the accessory pangenome are evenly distributed among the 29 isolates (fig. 2C), indicating a homogeneous pattern of genome evolution with similar rates of gene gain/loss processes across the $R. equi$ population.

A KEGG functional classification showed similar overall distribution of categories between the core and the accessory genome, except for a proportional enrichment of genes involved in genetic information processing and nucleotide and cofactor/vitamins metabolism in the core genome, and in xenobiotic degradation, lipid metabolism and environmental information processing in the accessory genome (fig. 2D).

**Specific Core Genome Features**

We investigated whether specific traits identified in the 103S genome as potentially important for $R. equi$ (Letek et al. 2010) belonged to the species’ core genome (supplementary table S2, Supplementary Material online). The absence of PTS sugar transport components (EI, HPr, EII complex/permeases) (Letek et al. 2010) was confirmed as a general feature of $R. equi$. This is likely due to gene loss because PTS components were present in all tested genomes from the other main lines of descent of the genus *Rhodococcus*. The PTS was also absent from the closely related $R. defluvii$ Ca11T (Kampfer et al. 2014), within the same terminal clade as $R. equi$ in the *Rhodococcus* phylogeny (see below fig. 4), indicating that the gene loss event likely took place in the common ancestor of both species.

Two putative non-PTS sugar transporter genes were identified in the $R. equi$ core genome: REQ19940-60 (103S annotation) encoding an RbsCB-like monosaccharide/ribose (xylose/arabinose) ATP-binding Cassette (ABC) transporter and cognate putative sugar kinase, and REQ20500 encoding...
a Major Facilitator Superfamily (MFS) permease similar to the Streptomyces coelicolor glucose transporter GlcP (van Wezel et al. 2005) (supplementary table S2, Supplementary Material online). Phenotype MicroArray (PMA) carbon source utilization tests (Bochner 2009) showed positive reactions for D-ribose, 2-deoxy-D-ribose, D-xylose (and its C₀-2 carbon epimer L-lyxose), and D/L-arabinose (supplementary fig. S1 A, Supplementary Material online). To exclude false positives due to abiotic dye reduction, growth curves were also performed in a chemically defined medium (mREMM, see supplementary fig. S1, Supplementary Material online, for details) using as a control L-lactate, a main carbon source for R. equi (Letek et al. 2010). Here only D-ribose consistently promoted R. equi growth, although after a protracted lag phase and to a lesser extent than L-lactate (supplementary fig. S1 B, Supplementary Material online). In some experiments, delayed, weak growth was also observed with α-D-glucose (supplementary fig. S1 B, Supplementary Material online). Thus, while thought to be unable to metabolize carbohydrates (Letek et al. 2010), R. equi might utilize some sugars, albeit less efficiently than L-lactate and other preferred carbon sources (i.e., acetate and in general short- and long-chain monocarboxylates and fatty acids [Letek et al. 2010 and our unpublished observations]).

Virtually, all 103S loci potentially involved in tolerance to desiccation and oxidative stress, and thus important for R. equi survival in dry soil and transmission by aerosolized dust (Muscatello et al. 2007; Vazquez-Boland et al. 2013), were also found to be part of the core genome (supplementary table S2, Supplementary Material online). The same applies to the intrinsic resistome identified in 103S (9/10 β-lactamases, 5/5 aminoglycoside phosphotransferases and 4/4 multidrug efflux systems were conserved in all strains) (supplementary table S2, Supplementary Material online). Indeed, in vitro resistance to a number of antimicrobials, particularly
β-lactams and quinolones, has been observed in 103S (Letek et al. 2010) and reported in the literature for R. equi (Nordmann and Ronco 1992; Mascellino et al. 1994; Soriano et al. 1998; Makrai et al. 2000; Jacks et al. 2003; Jones and Goodfellow 2012).

All putative virulence-associated loci found in 103S, including those identified as HGT islands, that is, mce2, srt1, srt2 and the pilus and capsule biosynthesis determinants (Letek et al. 2010), also belonged to the R. equi core genome (Supplementary table S2 and fig. S2, Supplementary Material online). Two large HGT regions previously identified in 103S, likely generated by multiple horizontal gene acquisitions (Letek, et al. 2010), were also at least partially conserved in all isolates (fig. 1 and supplementary fig. S3, Supplementary Material online). Since these genomic islands are all at the same chromosomal location in the genomes analyzed, the corresponding HGT events clearly occurred before R. equi diversification into sublineages (see below). The maintenance of a foreign DNA signature indicates a relatively recent acquisition, consistent with an evolutionarily young species.

**Rhodococcus equi Core Genome Diversity and Population Structure**

The species’ phylogeny was reconstructed by analysis of single nucleotide polymorphisms in alignments of the draft genomes to the 103S reference genome. All R. equi isolates branched radially at a short distance (≈0.001–0.002 substitutions per
site between nodes of the major species’ sublineages, denoting strong intraspecies genetic relatedness (fig. 3). The high degree of relatedness is most evident in a genomic ML tree including R. defluvii Ca11T (fig. 3B and C), a species most closely related to R. equi according to 16S rDNA phylogenies (Kampfer et al. 2014) and whole genome comparisons (see above and supplementary fig. S7, Supplementary Material online). A recombination analysis showed no evidence of significant core–genome exchanges between strains (supplementary fig. S4, Supplementary Material online).

Comparison of a parsimony tree based on a gene presence/absence matrix (supplementary fig. S5, Supplementary Material online) and the ML core–genome tree (fig. 3C) showed similar relationships between strains, indicating that the different R. equi sublineages tend to be associated with a similar accessory proteome composition. Overall, the above data is consistent with a clonal diversification pattern and a recent evolutionary origin for R. equi.

There was no obvious association between core–genome phylogenotypes and host source, whereas the latter was clearly linked with the host-associated plasmid type (fig. 3C). No correlation between genomic types and the geographical origin of the isolates was observed. This is illustrated by the equine strains DSM20307T and PAM1271 or the bovine strains PAM1354 and PAM1557, which essentially share the same core and accessory genome while originating from Sweden and Canada, or Ireland and Japan, respectively (fig. 3C and supplementary fig. S5, Supplementary Material online).

Rhodococcus Phylogenetics

In a whole-genome phylogeny, the genus Rhodococcus appears as a distinct, well-defined monophyletic grouping of the Corynebacterales (fig. 4 and supplementary fig. S6, Supplementary Material online). Rhodococcus equi isolates are clustered together in a Rhodococcus subclade (no. 3 or “equi” subclade) that contains two sister sublineages, one comprising R. equi and R. defluvii Ca11T, confirming their close relatedness (Kampfer et al. 2014), and the other, Rhodococcus triatomae BKS15-14 and an unclassified isolate (fig. 4 and supplementary fig. S6, Supplementary Material online). Two other Rhodococcus subclades correspond to the 16S rDNA monophyletic groupings “rhodochrous” (subclade 1, with two sublineages: one encompassing R. ruber, another the type species of the genus, R. rhodochrous, and Rhodococcus pyridinivorans) and “erythropolis” (subclade 2, also with two sublineages: one with R. opacus, R. jostii, Rhodococcus imtechensis and Rhodococcus wratislavensis, the other comprising R. erythropolis and Rhodococcus qingshengii). Of note, subclades 2 (“erythropolis/jostii-opacus”) and 3 (“equi”) are sister lineages of a main Rhodococcus sub-division at the top of the genus tree (fig. 4 and supplementary fig. S6, Supplementary Material online). Supplementary figure S7, Supplementary Material online, illustrates the genetic relatedness between R. equi and representative members of Rhodococcus subclades 1, 2 and 3 in pairwise DNA sequence alignments.

Rhodococcus rhodnii LMG 5362 and R. fascians isolates define respectively two novel, more distantly related Rhodococcus subclades (nos. 4 and 5), the latter (“fascians”) branching off at an early bifurcation in the genus phylogeny (fig. 4).

Rhodococcus and Nocardia form two clearly differentiated clades under a common node in the intermediate branchings of the Corynebacterales (fig. 4 and supplementary fig. S6, Supplementary Material online). Both genera belong to a well-supported phyletic line that also comprises Smaragdicoccus niigatensis DSM44881T, classified in the Nocardiaeae (as is Rhodococcus), as well as Mycobacterium spp. and Amycolicicoccus subflavus (Hoyosella subflava) DQ5394AT, classified in the Mycobacteriaeae (Ludwig et al. 2012). Another major Corynebacterales phylogenomic subdivision is formed by members of the genera Tsukamurella, of the monogeneric Tsukamurellaceae, and Gordonia and Williamsia, in some taxonomies classified within the Nocardiaeae (Ludwig et al. 2012). The phylogenomic data therefore indicate that the Nocardiaeae taxon is polyphyletic and call for a reclassification of the genera Rhodococcus, Nocardia and Smaragdicoccus into a same (Mycobacteriaeae) family together with Amycolicicoccus (Hoyosella) and Mycobacterium.

Conclusions

Our whole-genome comparative analyses show that R. equi is largely monomorphic, not supporting the commonly held view that R. equi is heterogeneous (McMinn et al. 2000; Jones and Goodfellow 2012; Jones et al. 2013b) and its isolates phylogenetically very diverse (Gurtler et al. 2004). The tendency of the core–genome sublineages to associate with a specific composition of the accessory genome and the lack of significant core–genome recombination indicate that R. equi evolution is primarily clonal. Although the accessory genome represents a relatively small fraction of an isolates’ gene content (~20%), R. equi possesses an open panenome that constitutes the basis of its genomic variability. The coincidence of the gaps in the genomic alignments with HGT islands in the complete 103S genome sequence indicates that lateral genetic exchanges have played a key role in the shaping of the R. equi accessory genome.

Our analyses show no evidence of phylogeographic correlation but instead of ample global circulation of genomotypes, probably linked to international livestock trade. The distribution of the host-associated virulence plasmid types in the R. equi phylogeny is consistent with the dynamic conjugal exchange of the plasmids across the R. equi population (Tripathi et al. 2012; Valero-Rello et al. 2015) and their key role in animal host tropism (Vazquez-Boland et al. 2013;...
Valero-Rello et al. 2015). Strains sharing the same core and accessory genotypetype and virulence plasmid type were associated with both the corresponding adapted animal host and people (e.g., pVAPB-carrying 1413 and 1533 isolates, pVAPN-carrying 1354 and 1557 isolates) (fig. 3 and supplementary fig. S5, Supplementary Material online), strongly supporting that *R. equi* infection is zoonotically transmitted to humans (Ocampo-Sosa et al. 2007; Vazquez-Boland et al. 2013). Further illustrating the remarkable uniformity of *R. equi*, virtually all major determinants predicted in 103S to be important for the species’ biology, virulence and niche adaptation (Letek et al. 2010) were part of the core genome. This includes the absence of a PTS and other specific metabolic traits such as the ΔthiC thiamin auxotrophic mutation or lactate utilization via a lutABC operon (Letek et al. 2010). These features may represent an adaptation to, and competitive advantage within the main saprophytic habitats of *R. equi*, manure-rich soil and the intestine (Muscatello, et al. 2007; Vazquez-Boland, et al. 2013), where microbially derived thiamine, and lactate and short-chain fatty acids produced by carbohydrate-fermenting microbiota, are presumably abundant.

Finally, our phylogenomic analyses resolve the lingering problem of *R. equi* taxonomy (Goodfellow et al. 1998; McMinn et al. 2000; Gurtler et al. 2004; Jones and...
Goodfellow 2012; Ludwig et al. 2012). It is evident from our data that R. equi is not at the periphery or outwith the genus Rhodococcus, closer to the Nocardia, as previously claimed (Goodfellow et al. 1998; McMinn et al. 2000; Jones et al. 2013b), but deeply embedded in the rhodococcal phylogeny. Indeed, the “equi-defluvi-triatomae” subclade (no. 3) forms with its sister “erythropolis/jostii-opacus” subclade (no. 2) a major monophyletic subdivision central to the genus Rhodococcus (fig. 4). In complete genome comparisons, R. equi 103S shows the same degree of pairwise homology to R. erythropolis PR4 and R. jostii RHA1 as these two subclade 2 members between themselves (Letek et al. 2010; Vazquez-Boland et al. 2013). This means that the recent proposal of transferring R. equi to a new genus “Prescotella”, with “Prescotella equi” as its sole species (Jones et al. 2013a, 2013b), would only be justified if new genera were also created for each R. erythropolis and R. jostii. Such an atomization of the genus Rhodococcus is unwarranted, because the rhodococci form, in the Corynebacteriales phylogenetic tree (see supplementary fig. S6, Supplementary Material online), a distinct monophyletic grouping equivalent in rank and diversity to other well-established genera, such as Corynebacterium, Gordonia or Mycobacterium.

Supplementary Material
Supplementary figures S1–S7 and tables S1–S3 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
We are greatly indebted to N. Fujita, National Institute of Technology and Evaluation (NITE), Japan, for making available the draft genome sequence of R. rhodochrous NBRC16069T. We also thank D. Lewis for her contribution in establishing our laboratory’s global R. equi isolate collection, and B. Contreras-Moreira for help with Get_Homologues software. This work was supported by the Horserace Betting Levy Board (grant nos. vet/pri/712 and vet/pri/753; to J.V.-B.) and core BBSRC funding from the Roslin Institute (BB/J004227/1). E.A. was supported by a BBSRC-funded Zoetis-sponsored CASE PhD studentship from the Centre for Infectious Diseases of the University of Edinburgh.

Literature Cited


Associate editor: Howard Ochman