

# Northumbria Research Link

Citation: Paterson, Megan, Ranasinghe, Diyanath, Blom, Jochen, Dover, Lynn, Sutcliffe, Iain, Lopes, Bruno and Sangal, Vartul (2019) Genomic analysis of a novel *Rhodococcus* (Prescottella) equi isolate from a bovine host. *Archives of Microbiology*, 201 (9). pp. 1317-1321. ISSN 0302-8933

Published by: Springer

URL: <https://doi.org/10.1007/s00203-019-01695-z> <<https://doi.org/10.1007/s00203-019-01695-z>>

This version was downloaded from Northumbria Research Link:  
<https://nrl.northumbria.ac.uk/id/eprint/40054/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



# Genomic analysis of a novel *Rhodococcus (Prescottella) equi* isolate from a bovine host

Megan L. Paterson<sup>1</sup> · Diyanath Ranasinghe<sup>1</sup> · Jochen Blom<sup>2</sup> · Lynn G. Dover<sup>1</sup> · Iain C. Sutcliffe<sup>1</sup> · Bruno Lopes<sup>3</sup> · Vartul Sangal<sup>1</sup>

Received: 7 March 2019 / Revised: 6 June 2019 / Accepted: 20 June 2019 / Published online: 13 July 2019  
© The Author(s) 2019

## Abstract

*Rhodococcus (Prescottella) equi* causes pneumonia-like infections in foals with high mortality rates and can also infect a number of other animals. *R. equi* is also emerging as an opportunistic human pathogen. In this study, we have sequenced the genome of a novel *R. equi* isolate, B0269, isolated from the faeces of a bovine host. Comparative genomic analyses with seven other published *R. equi* genomes, including those from equine or human sources, revealed a pangenome comprising of 6876 genes with 4141 genes in the core genome. Two hundred and 75 genes were specific to the bovine isolate, mostly encoding hypothetical proteins of unknown function. However, these genes include four copies of *terA* and five copies of *terD* genes that may be involved in responding to chemical stress. Virulence characteristics in *R. equi* are associated with the presence of large plasmids carrying a pathogenicity island, including genes from the *vap* multigene family. A BLAST search of the protein sequences from known virulence-associated plasmids (pVAPA, pVAPB and pVAPN) revealed a similar plasmid backbone on two contigs in bovine isolate B0269; however, no homologues of the main virulence-associated genes, *vapA*, *vapB* or *vapN*, were identified. In summary, this study confirms that *R. equi* genomes are highly conserved and reports the presence of an apparently novel plasmid in the bovine isolate B0269 that needs further characterisation to understand its potential involvement in virulence properties.

**Keywords** *Rhodococcus equi* · Virulence · Pathogenicity island · Bovine · Pneumonia · Plasmid

## Introduction

*Rhodococcus equi* (“*Rhodococcus hoagii*”/“*Prescottella equi*”), is a Gram-positive, obligate aerobic mycolic-acid containing actinomycete. *R. equi* strains are

phylogenomically distinct from other rhodococci and have been proposed to be classified into a novel genus, *Prescottella*, along with *Rhodococcus defluvii* (Jones et al. 2013; Sangal et al. 2015, 2016). The formal nomenclature of this taxon is still waiting clarification (Garrity 2014; Goodfellow et al. 2015). For simplicity, here we refer to the *R. equi*/*R. hoagii*/*P. equi* taxon as *R. equi*.

*Rhodococcus equi* primarily causes pyogranulomas and ulcerative enteritis in young foals (Prescott 1991; Vazquez-Boland et al. 2013) but can also cause sub-maxillary lymphadenitis and respiratory lymph node abscesses in a range of other animals, most notably porcine and bovine species (Vazquez-Boland et al. 2013; Valero-Rello et al. 2015; Ribeiro et al. 2017). It is also notable as an opportunistic human pathogen which is responsible for considerable mortality among immunocompromised patients (Yamshchikov et al. 2010; Giguere et al. 2011). Due to its significant economic impact on the equine breeding industry, recent research has focused on understanding the host–pathogen interaction and the mechanisms of pathogenesis of *R.*

Communicated by Erko Stackebrandt.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00203-019-01695-z>) contains supplementary material, which is available to authorized users.

✉ Vartul Sangal  
vartul.sangal@northumbria.ac.uk

<sup>1</sup> Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

<sup>2</sup> Bioinformatics and Systems Biology, Justus-Liebig-Universität, Giessen, Germany

<sup>3</sup> School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK

*equi* strains in different hosts (von Bargen and Haas 2009; Vazquez-Boland et al. 2013; Sangal et al. 2014). Notably, the nature of the pathogenicity island carried by the virulence plasmid significantly influences the host association of *R. equi* strains (Valero-Rello et al. 2015; MacArthur et al. 2017; Ribeiro et al. 2017).

In this study, we have sequenced the genome of a novel *R. equi* strain, B0269 that was isolated in 2014 from the faeces of a bovine host in Scotland. Bovine faecal sample (1:10 diluted with sterile saline) was homogenised in BHI enrichment broth and was incubated at ambient temperature for 1 h with occasional agitation. 100 µL of this sample was plated onto blood agar plate (E&O Laboratories, UK) and incubated at 37 °C for 48 h. A single colony was sub-cultured on another blood agar plate and a loopful of the culture was used for genomic DNA extraction using Wizard® Genomic DNA Purification Kit (Promega, USA). The DNA was quantified using the Quant-iT™ PicoGreen™ dsDNA assay kit (ThermoFisher Scientific, UK). The final concentration of genomic DNA was ~ 30 ng/µl. The genome sequencing was performed on an Illumina Hi-Seq 2000 (Illumina Inc., USA) at the Wellcome Trust Sanger Institute, UK. A total of 4,266,424 paired-end reads with an average read length of 100 bp were assembled into 30 contigs using Velvet (Zerbino and Birney 2008) and were annotated using the RAST pipeline (Aziz et al. 2008; Overbeek et al. 2014; Brettin et al. 2015). The draft genome is 5.7 Mb in size with a 68.4 mol% GC content and 5487 features (5430 coding sequences and 57 tRNA genes) that are comparable to previously sequenced *R. equi* strains (Anastasi et al. 2016; Sangal et al. 2016). The genome sequence of strain B0269 has been deposited at the ENA database under the accession number ERR646794.

For comparative genomic analyses, the publicly available genome sequences of seven *R. equi* strains were obtained from GenBank, i.e., strain 103S, ATCC 33707, C7<sup>T</sup>, N1288, N1295, N1301 and DSM 20295 (Accession numbers: NC\_014659, NZ\_CM001149, APJC00000000; LRQY00000000; NZ\_LRQZ00000000; NZ\_LRRA00000000; NZ\_LRRF00000000, respectively; Letek et al. 2010; Sangal et al. 2016). These strains were isolated from equine hosts except for strain ATCC 33707, which was isolated from a human, N1288 from a swine host, and N1301 from environment (Qin et al. 2010; Sangal et al. 2014, 2016). Strain DSM 20295 was first described as *Corynebacterium hoagii* in the year 1912 but the source of this strain is unknown (Morse 1912; Kämpfer et al. 2014). To have an equivalence of annotation, these genomes were re-annotated using the RAST pipeline (Aziz et al. 2008; Overbeek et al. 2014; Brettin et al. 2015) and were compared using EDGAR (Blom et al. 2016). Pairwise average amino-acid and nucleotide identities (AAI and ANI) of strain B0269 against 103S, ATCC 33707, C7<sup>T</sup>, N1288, N1295, N1301,

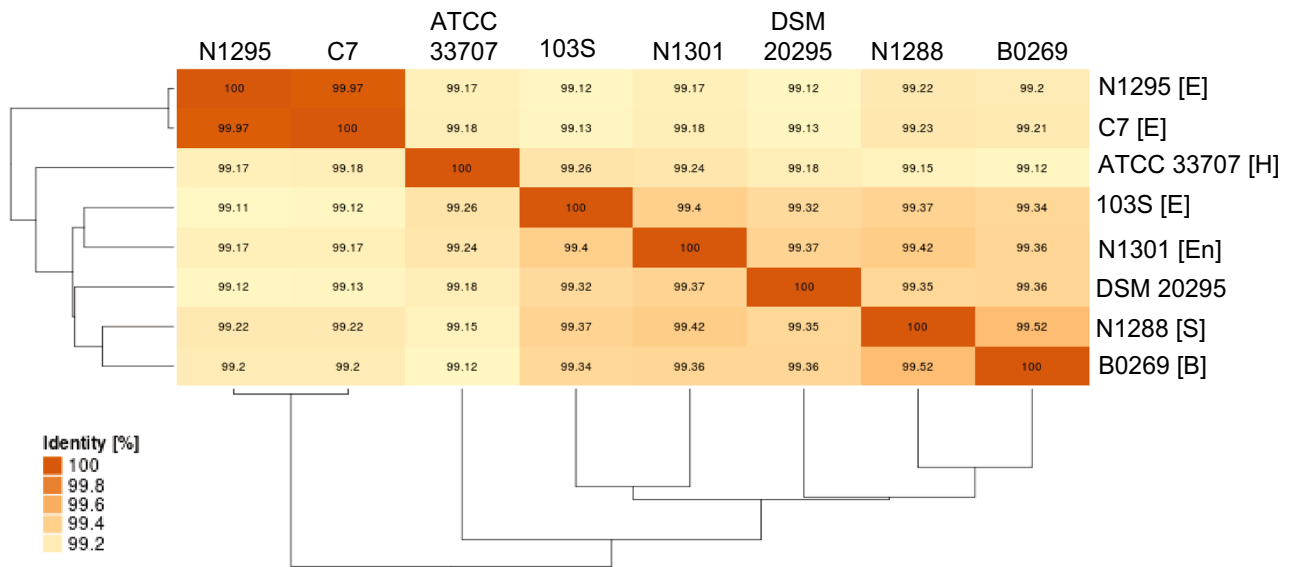
and DSM 20295 were calculated using BLAST-based algorithms implemented in EDGAR (Blom et al. 2016).

The pan genome is comprised of 6876 genes, of which 4141 genes belong to the core genome. The core genome is slightly larger than the one calculated by Anastasi et al. (2016), who identified 8174 genes (homologous gene-clusters) in the pan genome including 3858 core genes. Anastasi et al. (2016) used Get\_Homologues V2.0 (Contreras-Moreira and Vinuesa 2013) and OrthoMCL algorithm with a 70% sequence identity and 75% coverage in protein homology to define orthologs. In this study, we used EDGAR that applies a more robust approach to determine orthologous genes by calculating Blast Score Ratio Values (Lerat et al. 2003) through an intensively iterative process. The pan genomes are more stringently calculated by pairwise comparison of gene contents of a selected reference using Reciprocal Blast Hits that are filtered according to the orthology criterion based on the Blast Score Ratio Values (Blom et al. 2009, 2016). Therefore, the minor variation in the size of core and pan genomes is likely contributed by the difference in the approach used for calculating the pan genome between these studies.

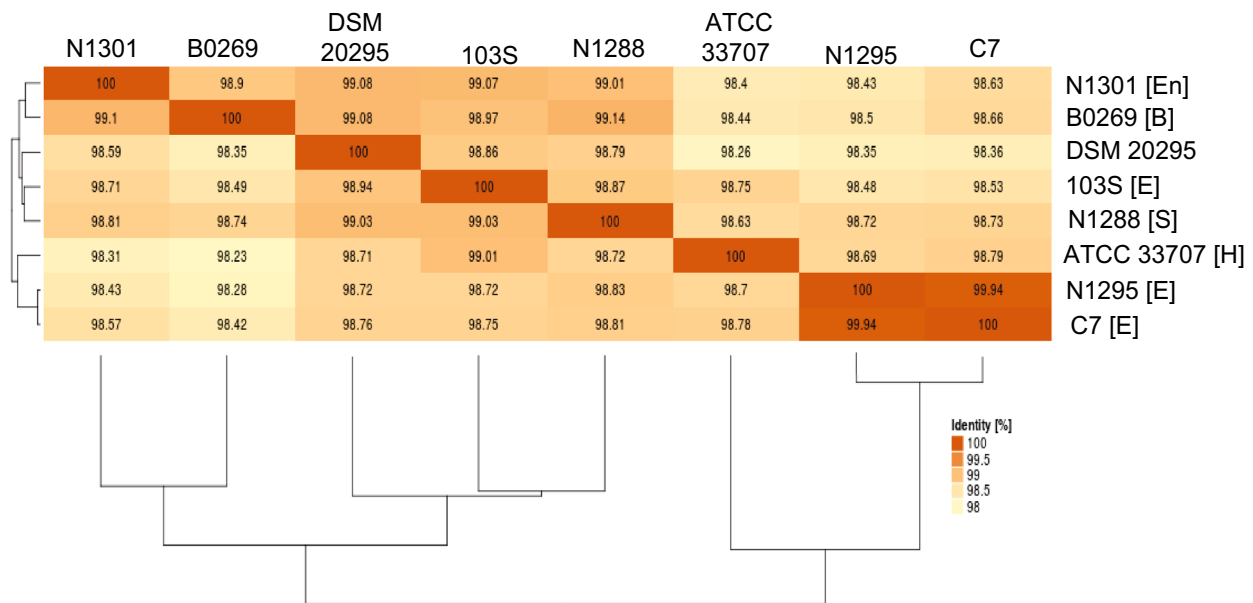
A maximum-likelihood tree was constructed from the concatenated protein sequence alignment of the core genome using IQ-Tree with 100,000 iterations of ultra-fast bootstrap and 100,000 SH-like approximate likelihood ratio test (Minh et al. 2013; Nguyen et al. 2015). The phylogenetic tree was visualised using the Interactive Tree Of Life (Letunic and Bork 2016), showing a close relatedness of strain B0269 with other *R. equi* isolates (Supplementary Fig. 1). AAI and ANI values of > 99% and > 98%, respectively are consistent with the identification of strain B0269 as *R. equi* and confirm a very high degree of genomic conservation within the species (Fig. 1), as observed previously (Sangal et al. 2015; Anastasi et al. 2016).

Only two hundred and seventy-five genes are found to be specific to *R. equi* strain B0269 that were absent from the other *R. equi* isolates included in this study (Supplementary Table 1). One hundred and thirty-one of these genes encode hypothetical proteins, 9 genes belong to mobile genetic elements (two mobile element proteins and 7 phage-associated genes) while the remaining 135 genes have predictable functions including four copies of *terA* and five copies of *terD* genes. The roles of *ter* gene-clusters remain elusive but they have been implicated to be involved in multiple activities including resistance to tellurite and other xenobiotic compounds, responding to chemical stress and anti-viral defence mechanisms (Anantharaman et al. 2012). *Ter* family proteins have been found in the closely related species *R. defluvi*, but not in other *R. equi* strains (Sangal et al. 2015). The *ter* operon has also been found to help *Yersinia pestis* survive within macrophages (Ponnusamy and Clinkenbeard 2015)

### A Pairwise average amino-acid identities among *R. equi* genomes isolated from different sources



### B Pairwise average nucleotide identities among *R. equi* genomes isolated from different sources



**Fig. 1** Heat maps showing **a** pairwise average amino-acid identities (AAI) and **b** pairwise average nucleotide identities (ANI) among *R. equi* genomes. The source of isolation where known, are mentioned

in parentheses next to the strain designations: [B] bovine, [E] equine, [En] environment, [H] human and [S] swine hosts

and, therefore, could contribute to the virulence of strain B0269, although we note that strains without *ter* genes survive and multiply within macrophages (Rahman et al. 2005; von Bargen and Haas 2009; Vazquez-Boland et al. 2013). The *ter* region in strain B0269 is larger (~ 7 Kb; reB0269\_Peg1159–reB0269\_Peg1166) than the one in *R.*

*defluvii* strain Ca11<sup>T</sup> (~ 4 Kb region; figl6666666.64062.peg.1365–figl6666666.64062.peg.1370; Sangal et al. 2015), with average GC content of 64.77 and 65.8 mol%, respectively (Supplementary Fig. 2). Furthermore, additional *ter* genes are present on the same contig in strain B0269 (reB0269\_Peg1150–reB0269\_Peg1152) and on

different contigs in strain Ca11<sup>T</sup> (figl6666666.64062.pcg.103 and figl6666666.64062.pcg.2208). The discontinuous distribution of the *ter* operon in *R. equi/R. defluvi* strains (proposed genus *Prescottella*) suggests this operon may have been acquired by horizontal gene transfer independently by strain B0269.

Three types of virulence plasmids have been identified among *R. equi* isolates (Valero-Rello et al. 2015; MacArthur et al. 2017). Equine and porcine isolates generally harbour circular plasmids, pVAPA and pVAPB, respectively, while a linear pVAPN plasmid has been identified among bovine isolates (Valero-Rello et al. 2015; Ribeiro et al. 2017). *R. equi* strains with any of these plasmids are capable of human infection. Environmental *R. equi* isolates commonly lack the virulence plasmids (Ribeiro et al. 2017). A BLAST search of the protein sequences from pVAPA, pVAPB and pVAPN revealed a plasmid backbone similar to that of pVAPN to be present in strain B0269 with 45 out of 140 genes showing > 50% query coverage (alignment length\*100/query length) and > 70% sequence similarities to genes on contigs 2, 9, 11, 12 and 18 (Supplementary Table 2). In contrast, only five pVAPA genes and four pVAPB genes showed > 70% sequence similarities to genes on these contigs. One hundred and seventy-three of the 275 B0269 specific genes are present on these contigs. Sixty-eight of them encoded hypothetical proteins, four genes encoded ABC transporter components for either iron or peptides substrates, and the remaining genes were related to various cellular functions without any obvious association with virulence. This suggests that these proteins may confer novel functionalities to this plasmid type.

Contigs 9 and 18 did not map onto the chromosomal sequence of *R. equi* reference strain 103S when the draft assembly of strain B0269 was aligned using MAUVE (Darling et al. 2010), again consistent with these being plasmid-derived sequences. Only 12% of contig 12 sequence shared similarities with the chromosomal sequence of strain 103S and this contig also likely belongs to the novel plasmid. Ninety-two percent of the contig 2 and 41% of the contig 11 sequences mapped on the chromosome of strain 103S. Only three genes from each of these contigs showed similarity to the plasmid sequences and therefore, the unaligned regions on these contigs may represent genomic islands. In addition, the smaller contigs 23–30 also did not map on the chromosome of 103S and may potentially belong to the plasmid. As noted above, only 45 of the 140 pVAPN genes showed similarities with the proteins in B0269, suggesting that this strain likely possesses a novel plasmid similar to pVAPN but bearing with a distinctive overall gene complement that should be further characterised to understand its potential role in pathogenesis.

## Conclusions

Genomic analyses of eight *R. equi* isolates from diverse sources (environment, equine, bovine and human) confirms that the *R. equi* genome is highly conserved. Bovine strain B0269 possesses multiple copies of *terA* and *terD* genes that are absent from other *R. equi* strains and the functions of these remain to be determined. This strain also apparently carries a novel large plasmid that has a genetic backbone similar to the virulence-associated plasmid pVAPN recovered from other bovine strains; however, further characterisation is needed to understand its potential involvement in virulence properties.

**Acknowledgements** MLP is supported by Research Development Fund from Northumbria University to VS. The authors would like to thank J. Gibson for IT assistance.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Anantharaman V, Iyer LM, Aravind L (2012) Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. *Mol BioSyst* 8:3142–3165. <https://doi.org/10.1039/c2mb25239b>
- Anastasi E, MacArthur I, Scortti M, Alvarez S, Giguere S, Vazquez-Boland JA (2016) Pangenome and phylogenomic analysis of the pathogenic actinobacterium *Rhodococcus equi*. *Genome Biol Evol* 8:3140–3148. <https://doi.org/10.1093/gbe/evw222>
- Aziz RK et al (2008) The RAST server: rapid annotations using subsystems technology. *BMC Genom* 9:75. <https://doi.org/10.1186/1471-2164-9-75>
- Blom J et al (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinform* 10:154. <https://doi.org/10.1186/1471-2105-10-154>
- Blom J et al (2016) EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 44:W22–W28. <https://doi.org/10.1093/nar/gkw255>
- Brettin T et al (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>
- Contreras-Moreira B, Vinuesa P (2013) GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. <https://doi.org/10.1128/AEM.02411-13>
- Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.



- PLoS ONE 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>
- Garrity GM (2014) Conservation of *Rhodococcus equi* (Magnusson 1923) Goodfellow and Alderson 1977 and rejection of *Corynebacterium hoagii* (Morse 1912) Ebersson 1918. *Int J Syst Evol Microbiol* 64:311–312. <https://doi.org/10.1099/ijs.0.059741-0>
- Giguere S et al (2011) *Rhodococcus equi*: clinical manifestations, virulence, and immunity. *J Vet Intern Med* 25:1221–1230. <https://doi.org/10.1111/j.1939-1676.2011.00804.x>
- Goodfellow M, Sangal V, Jones AL, Sutcliffe IC (2015) Charting stormy waters: a commentary on the nomenclature of the equine pathogen variously named *Prescottella equi*, *Rhodococcus equi* and *Rhodococcus hoagii*. *Equine Vet J*. <https://doi.org/10.1111/evj.12399>
- Jones AL, Sutcliffe IC, Goodfellow M (2013) *Prescottia equi* gen. nov., comb. nov.: a new home for an old pathogen. *Antonie Van Leeuwenhoek* 103:655–671. <https://doi.org/10.1007/s10482-012-9850-8>
- Kämpfer P, Dott W, Martin K, Glaeser SP (2014) *Rhodococcus defluvi* sp. nov., isolated from wastewater of a bioreactor and formal proposal to reclassify [*Corynebacterium hoagii*] and *Rhodococcus equi* as *Rhodococcus hoagii* comb. nov. *Int J Syst Evol Microbiol* 64:755–761. <https://doi.org/10.1099/ijs.0.053322-0>
- Lerat E, Daubin V, Moran NA (2003) From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol* 1:E19. <https://doi.org/10.1371/journal.pbio.0000019>
- Letek M et al (2010) The genome of a pathogenic rhodococcus: cooperative virulence underpinned by key gene acquisitions. *PLoS Genet* 6:e1001145. <https://doi.org/10.1371/journal.pgen.1001145>
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>
- MacArthur I, Anastasi E, Alvarez S, Scotti M, Vazquez-Boland JA (2017) Comparative genomics of *Rhodococcus equi* virulence plasmids indicates host-driven evolution of the vap pathogenicity Island. *Genome Biol Evol* 9:1241–1247. <https://doi.org/10.1093/gbe/evx057>
- Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Morse ME (1912) A study of the diphtheria group of organisms by the biometrical method. *J Infect Dis* 11:253–285
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Overbeek R et al (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>
- Ponnusamy D, Clinkenbeard KD (2015) Role of tellurite resistance operon in filamentous growth of *Yersinia pestis* in macrophages. *PLoS ONE* 10:e0141984. <https://doi.org/10.1371/journal.pone.0141984>
- Prescott JF (1991) *Rhodococcus equi*: an animal and human pathogen. *Clin Microbiol Rev* 4:20–34
- Qin X et al (2010) *Rhodococcus equi* ATCC 33707, whole genome shotgun sequencing. In: 2010 edn. <http://www.ncbi.nlm.nih.gov/nucleotide/325556670>. Accessed Nov 2017
- Rahman MT, Parreira V, Prescott JF (2005) In vitro and intra-macrophage gene expression by *Rhodococcus equi* strain 103. *Vet Microbiol* 110:131–140. <https://doi.org/10.1016/j.vetmic.2005.08.005>
- Ribeiro MG et al (2017) Novel bovine-associated pVAPN plasmid type in *Rhodococcus equi* identified from lymph nodes of slaughtered cattle and lungs of people living with HIV/AIDS. *Transbound Emerg Dis* 1:21. <https://doi.org/10.1111/tbed.12785>
- Sangal V, Jones AL, Goodfellow M, Sutcliffe IC, Hoskisson PA (2014) Comparative genomic analyses reveal a lack of a substantial signature of host adaptation in *Rhodococcus equi* (“*Prescottella equi*”). *Pathog Dis* 71:352–356. <https://doi.org/10.1111/2049-632X.12126>
- Sangal V, Jones AL, Goodfellow M, Hoskisson PA, Kämpfer P, Sutcliffe IC (2015) Genomic analyses confirm close relatedness between *Rhodococcus defluvi* and *Rhodococcus equi* (*Rhodococcus hoagii*). *Arch Microbiol* 197:113–116. <https://doi.org/10.1007/s00203-014-1060-5>
- Sangal V et al (2016) Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. *Sci Rep* 6:38392. <https://doi.org/10.1038/srep38392>
- Valero-Rello A et al (2015) An Invertron-like linear plasmid mediates intracellular survival and virulence in bovine isolates of *Rhodococcus equi*. *Infect Immun* 83:2725–2737. <https://doi.org/10.1128/IAI.00376-15>
- Vazquez-Boland JA, Giguere S, Hapeshi A, MacArthur I, Anastasi E, Valero-Rello A (2013) *Rhodococcus equi*: the many facets of a pathogenic actinomycete. *Vet Microbiol* 167:9–33. <https://doi.org/10.1016/j.vetmic.2013.06.016>
- von Bargen K, Haas A (2009) Molecular and infection biology of the horse pathogen *Rhodococcus equi*. *FEMS Microbiol Rev* 33:870–891. <https://doi.org/10.1111/j.1574-6976.2009.00181.x>
- Yamshchikov AV, Schuetz A, Lyon GM (2010) *Rhodococcus equi* infection. *Lancet Infect Dis* 10:350–359. [https://doi.org/10.1016/S1473-3099\(10\)70068-2](https://doi.org/10.1016/S1473-3099(10)70068-2)
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.