

# Northumbria Research Link

Citation: Sangal, Vartul, Jones, Amanda, Goodfellow, Michael, Hoskisson, Paul, Kämpfer, Peter and Sutcliffe, Iain (2015) Genomic analyses confirm close relatedness between *Rhodococcus defluvii* and *Rhodococcus equi* (*Rhodococcus hoagii*). *Archives of Microbiology*, 197 (1). pp. 113-116. ISSN 0302-8933

Published by: Springer

URL: <http://dx.doi.org/10.1007/s00203-014-1060-5> <<http://dx.doi.org/10.1007/s00203-014-1060-5>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/18365/>

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.)

[www.northumbria.ac.uk/nrl](http://www.northumbria.ac.uk/nrl)



Short Communication

**Genomic analyses confirm close relatedness between *Rhodococcus defluvii* and  
*Rhodococcus equi* (*Rhodococcus hoagii*)**

Vartul Sangal<sup>1\*</sup>, Amanda L. Jones<sup>1</sup>, Michael Goodfellow<sup>2</sup>, Paul A. Hoskisson<sup>3</sup>, Peter  
Kämpfer<sup>4</sup>, Iain C. Sutcliffe<sup>1</sup>

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

<sup>2</sup>School of Biology, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

<sup>3</sup>Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161  
Cathedral Street, Glasgow G4 0RE, UK

<sup>4</sup>Institut für Angewandte Mikrobiologie, Justus-Liebig-Universität, Giessen, D-35392,  
Germany

\*Correspondence: Vartul Sangal, Faculty of Health and Life Sciences, Northumbria  
University, Northumberland Building, Newcastle upon Tyne – NE1 8ST, UK.

Tel: +44 191 243 7173; e-mail: [vartul.sangal@northumbria.ac.uk](mailto:vartul.sangal@northumbria.ac.uk)

Keywords: *Rhodococcus equi*, *Rhodococcus defluvi*, genome, average nucleotide identity,  
average amino-acid identity

1 **Abstract**

2 *Rhodococcus defluvi* strain Ca11<sup>T</sup> was isolated from a bioreactor involved in extensive  
3 phosphorus removal. We have sequenced the whole genome of this strain and our  
4 comparative genomic and phylogenetic analyses confirm its close relatedness with  
5 *Rhodococcus equi* (*Rhodococcus hoagii*) strains, which share >80% of the gene content. The  
6 *R. equi* virulence plasmid is absent though most of the chromosomal *R. equi* virulence-  
7 associated genes are present in *R. defluvi* Ca11<sup>T</sup>. These data suggest that although *R. defluvi*  
8 is an environmental organism, it has the potential to colonise animal hosts.

9 *Rhodococcus defluvii* is a Gram-positive, mycolic acid-containing, rod shaped  
10 actinobacterium that has been described as a new member of the heterogeneous genus  
11 *Rhodococcus* (Jones and Goodfellow 2012; Kämpfer et al. 2014). The type strain of this  
12 species, Ca11<sup>T</sup> (=DSM 45893<sup>T</sup> =LMG27563<sup>T</sup>), was isolated from a wastewater treatment  
13 bioreactor involved in phosphorus removal. Strain Ca11<sup>T</sup> showed the highest 16S rRNA  
14 sequence similarity (98.9%) and corresponding DNA-DNA relatedness value (51.3%;  
15 reciprocal 38.1%) to the type strain of *Rhodococcus equi* (*Rhodococcus hoagii*; Kämpfer et  
16 al., 2014). The nomenclature of these taxa is currently a matter of debate as the priority of the  
17 name *R. hoagii* over *R. equi* (or *vice versa*) is under review by the Judicial Commission of the  
18 International Committee on Systematics of Prokaryotes (Garrity 2014) while the bacterial  
19 genus name *Rhodococcus* is considered to be illegitimate (Tindall 2014). For clarity, we here  
20 refer to the *R. equi*/*R. hoagii* taxon as *R. equi*.

21 In this study, we have sequenced the genome of *R. defluvii* strain Ca11<sup>T</sup> and  
22 performed comparative analyses with the genome sequences of *R. equi* strains C7<sup>T</sup> (Sangal et  
23 al. 2014), 103S (Letek et al. 2010) and ATCC 33707 (Qin et al. 2010) [GenBank accession  
24 numbers APJC00000000, NC\_014659 and NZ\_CM001149, respectively]. Genomic DNA  
25 extracted from 1.5ml of culture grown for 48 h at 30°C in Brain-Heart Infusion broth (Oxoid)  
26 was sequenced on an Illumina MiSeq instrument, according to the manufacturer's  
27 instructions. A total of 2,156,061 reads with an average read length of 238 bp were assembled  
28 into 267 contigs (>200 bp) using CLC Genomic Workbench (Qiagen). The size of assembly  
29 was 5,134,337 bp with an average 75-fold coverage.

30 The size of the draft genome and G+C content of *R. defluvii* strain Ca11<sup>T</sup> (5.13 Mb,  
31 68.71%) are similar to those of *R. equi* strains C7<sup>T</sup> (5.20 Mb, 68.79%), 103S (5.04 Mb,  
32 68.82%) and ATCC 33707 (5.26 Mb, 68.77%). However, the genome sequence has only  
33 been completed for strain 103S and so these values may slightly vary for other strains if their

34 genomes are finished. Using the RAST pipeline (Aziz et al. 2008), the Ca11<sup>T</sup> genome was  
35 annotated to have 4,796 features including 4,740 protein coding sequences. The genomes of  
36 *R. equi* strains were also re-annotated using the RAST pipeline to allow an equivalence of  
37 annotation. The Ca11<sup>T</sup> genome was found to share 4,166 genes with the three *R. equi* strains  
38 (3,720 with bi-directional and 446 with uni-directional protein BLAST hits; Aziz et al. 2012).  
39 It also shared an additional 128 genes with at least one *R. equi* strain but not with all three.  
40 446 genes were specific to *R. defluvii* Ca11<sup>T</sup> that were absent in the *R. equi* genomes; 361 of  
41 these encode hypothetical proteins and six belong to mobile genetic elements (transposase,  
42 phage associated or mobile element proteins). A BLAST search of 75 randomly selected  
43 hypothetical proteins of *R. defluvii* against the NCBI protein database using default settings  
44 revealed homologies for most of them with hypothetical proteins in other rhodococci or other  
45 bacterial species (data not shown), indicating that not all are unique to *R. defluvii* Ca11<sup>T</sup>. The  
46 remaining 79 genes specific to *R. defluvii* Ca11<sup>T</sup> (compared to the *R. equi* strains) can  
47 typically be related to known metabolic activities (Table S1), including a gene encoding  
48 alkylphosphonate utilization protein PhnA. The *phn* operon gene products are involved in the  
49 cleavage of carbon-phosphorus bonds in alkylphosphonates (Chen et al. 1990). However, the  
50 presence of the *phnA* gene in strain Ca11<sup>T</sup> is unlikely to be associated with phosphorus  
51 removal in the bioreactor from which it was isolated because the other genes of this operon  
52 are missing. Three homologs of *phnB* and two homologs of *phnE* genes were present  
53 elsewhere in the Ca11<sup>T</sup> genome but they are shared with the *R. equi* strains. A number of  
54 other genes involved in phosphorus metabolism are also common between *R. defluvii* and the  
55 three *R. equi* strains.

56 An operon in the genome of strain Ca11<sup>T</sup> that encodes Ter family proteins (TerA,  
57 TerB, TerC-like and two TerD) and associated biosynthetic enzymes is absent from the  
58 genomes of the three *R. equi* strains (Table S1). Comparable loci have previously been

59 suggested to be involved in biosynthesis of nucleoside-like metabolites (Anantharaman et al.  
60 2012). The protein BLAST search revealed the presence of homologs of these genes in other  
61 rhodococci and actinomycetes, suggesting a potential horizontal acquisition of this operon by  
62 *R. defluvii*. Alternatively, this operon may have been lost by *R. equi* as it has adapted to a  
63 pathogenic lifestyle. Two of the genes specific to *R. defluvii* Ca11<sup>T</sup> (compared to the *R. equi*  
64 strains) encode phospholipase C enzymes. Phospholipases C are the virulence factors that  
65 induce alveolar macrophage necrosis, resulting in cell death (Assis et al. 2014). As noted  
66 above, most of the genes specific to strain Ca11<sup>T</sup> encode hypothetical proteins and it is  
67 possible that some of these uncharacterized proteins contribute to functional variations  
68 between *R. defluvii* and *R. equi*.

69 Rhodococci are generally involved in environmental processes such as the  
70 degradation of organic and xenobiotic substances, except for the pathogens *R. equi* and  
71 *Rhodococcus fascians* (Bell et al. 1998; Alvarez 2010). The pathogenicity of these two  
72 species has been associated with the presence of large plasmids encoding virulence proteins  
73 (Takai et al. 2000; Letek et al. 2008; Francis et al. 2012; Stes et al. 2013). The virulence  
74 plasmid in *R. equi* is 80-90 Kb in size and carries a pathogenicity island encoding virulence  
75 associated proteins (Vap) while plasmid free strains were found to be avirulent (Takai et al.  
76 2000). A sequence BLAST-based functional comparison using the SEED server (Aziz et al.  
77 2012) revealed the absence of Vap proteins (VapA, C-I proteins from plasmid pVAPA1037  
78 and VapB, J-M from pVAPB1593; Letek et al. 2008) in the draft genome sequence of *R.*  
79 *defluvii*, suggesting the absence of the virulence plasmid in strain Ca11<sup>T</sup>. However, 228 of the  
80 243 *R. equi* chromosomal virulence-related genes defined by Letek *et al.* (2010) are present  
81 in strain Ca11<sup>T</sup> (Table S2), including the *esx* cluster. The *paa* operon that was identified in *R.*  
82 *equi* strain ATCC 33707 and which may be involved in pathogenesis in humans (Sangal et al.  
83 2014) is absent from *R. defluvii* strain Ca11<sup>T</sup>. The presence of a high proportion of virulence-

84 related genes in the genome of strain Ca11<sup>T</sup> suggests that this organism may also have the  
85 potential to colonise animal hosts. Indeed, it is noted that three additional bacterial strains  
86 with 16S rRNA gene sequences identical to that of strain Ca11<sup>T</sup> have been isolated from  
87 salmon intestines (Skrodenyte-Arbaciauskiene, V. & Virbickas T. Genbank accession  
88 numbers HM244990, HM244992 and HM244993).

89 A phylogenetic analysis was performed using PhyloPhlAn (Segata et al. 2013)  
90 including *Rhodococcus erythropolis* PR4 (Sekine et al. 2006), *Rhodococcus jostii* RHA1  
91 (McLeod et al. 2006), *Nocardia brasiliensis* ATCC 700358 (Vera-Cabrera et al. 2012) and  
92 *Corynebacterium diphtheriae* NCTC 05011 (Sangal et al. 2012) were used as outgroups.  
93 PhyloPhlAn automatically extracts the sequences of the 400 most conserved universal  
94 proteins that were identified by off-line pre-processing of all available microbial genomes by  
95 Segata et al. (2013). It generates highly robust phylogenetic trees from a concatenated  
96 alignment of computationally selected subset of amino-acid sequences with highest entropy  
97 and an appropriate relative contribution of the most conserved residues from each protein  
98 following a maximum likelihood maximization approach (gamma model of rate  
99 heterogeneity) with 20 bootstrap replicates using RAxML (Stamatakis 2006). Our  
100 PhyloPhlAn analysis showed that *R. defluvii* Ca11<sup>T</sup> shared a phyletic line with *R. equi* that  
101 was relatively distant from the other rhodococci and from *N. brasiliensis* (Fig. 1). BLAST-  
102 based average nucleotide identities (ANIb) between the genomes of *R. defluvii* Ca11<sup>T</sup> and the  
103 *R. equi* strains were 82.96-83.25% (Richter and Rosselló-Móra 2009) and average amino acid  
104 identities (AAI) varied between 85.31-85.45%. The ANIb and AAI values between *R.*  
105 *defluvii* and the other rhodococci (*R. jostii* RHA1 and *R. erythropolis* PR4) were < 76% and  
106 < 72%, respectively. The digital DNA-DNA hybridization (dDDH) distances were calculated  
107 using the genome-to-genome distance calculator at the GGDC 2.0 web server (Auch et al.  
108 2010; Meier-Kolthoff et al. 2013). GGDC values mimic conventional DNA-DNA

109 hybridization values and have been shown to have very high correlation with 16S rRNA  
110 sequence distances (Auch et al. 2010; Meier-Kolthoff et al. 2013). GGDC 2.0 uses three  
111 different formulae to calculate the distances and the results of formula-2, which has been  
112 recommended for analysing draft genomes (Auch et al. 2010), were considered in this study.  
113 The dDDH values between *R. defluvii* and *R. equi* strains C7<sup>T</sup>, 103S and ATCC 33707 were  
114  $26.9 \pm 3.02$ ,  $27 \pm 3.02$  and  $27.1 \pm 3.01$ , respectively. The *R. defluvii* genome showed lower  
115 dDDH similarities with the *R. erythropolis* PR4 ( $20.2 \pm 2.73$ ) and *R. jostii* RHA1 ( $20.7 \pm$   
116  $2.81$ ) genomes, values that are comparable to the dDDH distances from *N. brasiliensis* ATCC  
117 00358 ( $20.4 \pm 2.63$ ) and *C. diphtheriae* NCTC 05011 ( $21 \pm 2.53$ ). Cumulatively, these results  
118 suggest that *R. defluvii* is more closely related to *R. equi* than to other rhodococci, as  
119 previously concluded from 16S rRNA gene sequence analysis (Kämpfer et al. 2014).

120 In addition to the nomenclatural issues highlighted above, it has been proposed that *R.*  
121 *equi* should be reclassified as '*Prescottella equi*' (Jones et al. 2013b; Jones et al. 2013a).  
122 However, the genus name '*Prescottella*' cannot be validated until the Judicial Commission  
123 reports on whether the species epithet *equi* should be conserved over *hoagii* (Garrity 2014).  
124 Based on the phylogenetic and genomic distances between *R. defluvii* and the other  
125 rhodococci (Fig. 1), *R. defluvii* could eventually be reclassified as a second species within  
126 '*Prescottella*'. However, this conclusion needs further support from analyses of a larger  
127 collection of genomes of *Rhodococcus* species.

128 In summary, we report the genome sequence of the type strain of the recently  
129 identified species, *R. defluvii* strain Ca11<sup>T</sup>. The strain is phylogenetically closely related to *R.*  
130 *equi* strains with high similarities both at the nucleotide and functional levels. The whole  
131 genome shotgun sequence has been deposited at DDBJ/EMBL/GenBank under the Accession  
132 number JPOC00000000. The version described in this study is the first version,  
133 JPOC01000000.



134

135

### **Acknowledgements**

136 VS is supported by an Anniversary Research Fellowship from Northumbria University,

137 Newcastle upon Tyne. The authors would like to thank anonymous reviewers for their

138 constructive comments and suggestions. We also thank NU-OMICS facility for assistance in

139 genome sequencing and J. Gibson for IT assistance.

140

141

## References

142 Alvarez HM (2010) Biology of *Rhodococcus*. Springer, Heidelberg doi: 10.1007/978-3-642-  
143 12937-7

144 Anantharaman V, Iyer LM, Aravind L (2012) Ter-dependent stress response systems: novel  
145 pathways related to metal sensing, production of a nucleoside-like metabolite, and  
146 DNA-processing. Mol Biosyst 8:3142-3165 doi: 10.1039/c2mb25239b

147 Assis PA et al. (2014) *Mycobacterium tuberculosis* expressing phospholipase C subverts  
148 PGE2 synthesis and induces necrosis in alveolar macrophages. BMC Microbiol  
149 14:128 doi: 10.1186/1471-2180-14-128

150 Auch AF, von Jan M, Klenk HP, Göker M (2010) Digital DNA-DNA hybridization for  
151 microbial species delineation by means of genome-to-genome sequence comparison.  
152 Stand Genomic Sci 2:117-134 doi: 10.4056/signs.531120

153 Aziz RK et al. (2008) The RAST Server: rapid annotations using subsystems technology.  
154 BMC Genomics 9:75 doi: 10.1186/1471-2164-9-75

155 Aziz RK et al. (2012) SEED servers: high-performance access to the SEED genomes,  
156 annotations, and metabolic models. PLoS One 7:e48053 doi:  
157 10.1371/journal.pone.0048053

158 Bell KS, Philp JC, Aw DW, Christofi N (1998) The genus *Rhodococcus*. J Appl Microbiol  
159 85:195-210

160 Chen CM, Ye QZ, Zhu ZM, Wanner BL, Walsh CT (1990) Molecular biology of carbon-  
161 phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved  
162 in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. J Biol Chem  
163 265:4461-4471

164 Francis I et al. (2012) pFiD188, the linear virulence plasmid of *Rhodococcus fascians* D188.  
165 Mol Plant Microbe Interact 25:637-647 doi: 10.1094/MPMI-08-11-0215

166 Garrity GM (2014) Conservation of *Rhodococcus equi* (Magnusson 1923) Goodfellow and  
167 Alderson 1977 and rejection of *Corynebacterium hoagii* (Morse 1912) Ebersson 1918.  
168 Int J Syst Evol Microbiol 64:311-312 doi: 10.1099/ijs.0.059741-0

169 Jones AL, Goodfellow M (2012) Genus IV *Rhodococcus* (Zopf 1891) emended. Goodfellow,  
170 Alderson and Chun 1998a. In: Goodfellow M et al. (eds) Bergey's Manual of  
171 Systematic Bacteriology, 2 edn. Springer, New York, pp 437-464

172 Jones AL, Sutcliffe IC, Goodfellow M (2013a) *Prescottia equi* gen. nov., comb. nov.: a new  
173 home for an old pathogen. Antonie Van Leeuwenhoek 103:655-671 doi:  
174 10.1007/s10482-012-9850-8

175 Jones AL, Sutcliffe IC, Goodfellow M (2013b) Proposal to replace the illegitimate genus  
176 name *Prescottia* Jones et al. 2013 with the genus name *Prescottella* gen. nov. and to  
177 replace the illegitimate combination *Prescottia equi* Jones et al. 2013 with

- 178 *Prescottella equi* comb. nov. Antonie Van Leeuwenhoek 103:1405-1407 doi:  
179 10.1007/s10482-013-9924-2
- 180 Kämpfer P, Dott W, Martin K, Glaeser SP (2014) *Rhodococcus defluvii* sp. nov., isolated  
181 from wastewater of a bioreactor and formal proposal to reclassify [*Corynebacterium*  
182 *hoagii*] and *Rhodococcus equi* as *Rhodococcus hoagii* comb. nov. Int J Syst Evol  
183 Microbiol 64:755-761 doi: 10.1099/ijs.0.053322-0
- 184 Letek M et al. (2010) The genome of a pathogenic *Rhodococcus*: cooptive virulence  
185 underpinned by key gene acquisitions. PLoS Genet 6 doi:  
186 10.1371/journal.pgen.1001145
- 187 Letek M et al. (2008) Evolution of the *Rhodococcus equi vap* pathogenicity island seen  
188 through comparison of host-associated *vapA* and *vapB* virulence plasmids. J Bacteriol  
189 190:5797-5805 doi: 10.1128/JB.00468-08
- 190 McLeod MP et al. (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights  
191 into a catabolic powerhouse. Proc Natl Acad Sci U S A 103:15582-15587 doi:  
192 10.1073/pnas.0607048103
- 193 Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M (2013) Genome sequence-based species  
194 delimitation with confidence intervals and improved distance functions. BMC  
195 Bioinformatics 14:60 doi: 10.1186/1471-2105-14-60
- 196 Qin X et al. (2010) *Rhodococcus equi* ATCC 33707, whole genome shotgun sequencing. In,  
197 2010 edn, <http://www.ncbi.nlm.nih.gov/nuccore/325556670>
- 198 Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic  
199 species definition. Proc Natl Acad Sci U S A 106:19126-19131 doi:  
200 10.1073/pnas.0906412106
- 201 Sangal V, Jones AL, Goodfellow M, Sutcliffe IC, Hoskisson PA (2014) Comparative  
202 genomic analyses reveal a lack of a substantial signature of host adaptation in  
203 *Rhodococcus equi* ("*Prescottella equi*"). Pathog Dis 71:352-356 doi: 10.1111/2049-  
204 632X.12126
- 205 Sangal V, Tucker NP, Burkovski A, Hoskisson PA (2012) Draft genome sequence of  
206 *Corynebacterium diphtheriae* biovar intermedius NCTC 5011. J Bacteriol 194:4738  
207 doi: 10.1128/JB.00939-12
- 208 Segata N, Bornigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for  
209 improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304  
210 doi: 10.1038/ncomms3304
- 211 Sekine M et al. (2006) Sequence analysis of three plasmids harboured in *Rhodococcus*  
212 *erythropolis* strain PR4. Environ Microbiol 8:334-346 doi: 10.1111/j.1462-  
213 2920.2005.00899.x

- 214 Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses  
215 with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690 doi:  
216 10.1093/bioinformatics/btl446
- 217 Stes E, Francis I, Pertry I, Dolzblasz A, Depuydt S, Vereecke D (2013) The leafy gall  
218 syndrome induced by *Rhodococcus fascians*. *FEMS Microbiol Lett* 342:187-194 doi:  
219 10.1111/1574-6968.12119
- 220 Takai S et al. (2000) DNA sequence and comparison of virulence plasmids from  
221 *Rhodococcus equi* ATCC 33701 and 103. *Infect Immun* 68:6840-6847
- 222 Tindall BJ (2014) A note on the genus name *Rhodococcus* Zopf 1891 and its homonyms. *Int*  
223 *J Syst Evol Microbiol* 64:1062-1064 doi: 10.1099/ijs.0.060624-0
- 224 Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Perez-Maya AA, Ocampo-Candiani J  
225 (2012) Complete genome sequence of *Nocardia brasiliensis* HUJEG-1. *J Bacteriol*  
226 194:2761-2762 doi: 10.1128/JB.00210-12

227 **Figure Legend**

228 **Figure 1.** Phylogenetic tree (radial, un-rooted) derived from 400 universal proteins using the  
229 program PhyloPhlAn showing the relatedness of *R. defluvii* Ca11<sup>T</sup> with *R. equi* and  
230 representatives of other closely related taxa. Scale bar shows normalized fraction of total  
231 branch lengths as described by Segata et al. (2013).

