Short Communication

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

Vartul Sangal¹*, Amanda L. Jones¹, Michael Goodfellow², Paul A. Hoskisson³, Peter Kämpfer⁴, Iain C. Sutcliffe¹

¹Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
²School of Biology, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK
³Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow G4 0RE, UK
⁴Institut 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

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

Rhodococcus defluvii strain Ca11T was isolated from a bioreactor involved in extensive phosphorus removal. We have sequenced the whole genome of this strain and our comparative genomic and phylogenetic analyses confirm its close relatedness with Rhodococcus equi (Rhodococcus hoagii) strains, which share >80% of the gene content. The R. equi virulence plasmid is absent though most of the chromosomal R. equi virulence-associated genes are present in R. defluvii Ca11T. These data suggest that although R. defluvii is an environmental organism, it has the potential to colonise animal hosts.
*Rhodococcus defluvii* is a Gram-positive, mycolic acid-containing, rod shaped actinobacterium that has been described as a new member of the heterogeneous genus *Rhodococcus* (Jones and Goodfellow 2012; Kämpfer et al. 2014). The type strain of this species, Ca11\(^\mathrm{T}\) (=DSM 45893\(^\mathrm{T}\) =LMG27563\(^\mathrm{T}\)), was isolated from a wastewater treatment bioreactor involved in phosphorus removal. Strain Ca11\(^\mathrm{T}\) showed the highest 16S rRNA sequence similarity (98.9%) and corresponding DNA-DNA relatedness value (51.3%; reciprocal 38.1%) to the type strain of *Rhodococcus equi* (*Rhodococcus hoagii*; Kämpfer et al., 2014). The nomenclature of these taxa is currently a matter of debate as the priority of the name *R. hoagii* over *R. equi* (or vice versa) is under review by the Judicial Commission of the International Committee on Systematics of Prokaryotes (Garrity 2014) while the bacterial genus name *Rhodococcus* is considered to be illegitimate (Tindall 2014). For clarity, we here refer to the *R. equi*/*R. hoagii* taxon as *R. equi*.

In this study, we have sequenced the genome of *R. defluvii* strain Ca11\(^\mathrm{T}\) and performed comparative analyses with the genome sequences of *R. equi* strains C7\(^\mathrm{T}\) (Sangal et al. 2014), 103S (Letek et al. 2010) and ATCC 33707 (Qin et al. 2010) [GenBank accession numbers APJC00000000, NC_014659 and NZ_CM001149, respectively]. Genomic DNA extracted from 1.5ml of culture grown for 48 h at 30\(^\circ\)C in Brain-Heart Infusion broth (Oxoid) was sequenced on an Illumina MiSeq instrument, according to the manufacturer’s instructions. A total of 2,156,061 reads with an average read length of 238 bp were assembled into 267 contigs (>200 bp) using CLC Genomic Workbench (Qiagen). The size of assembly was 5,134,337 bp with an average 75-fold coverage.

The size of the draft genome and G+C content of *R. defluvii* strain Ca11\(^\mathrm{T}\) (5.13 Mb, 68.71%) are similar to those of *R. equi* strains C7\(^\mathrm{T}\) (5.20 Mb, 68.79%), 103S (5.04 Mb, 68.82%) and ATCC 33707 (5.26 Mb, 68.77%). However, the genome sequence has only been completed for strain 103S and so these values may slightly vary for other strains if their
genomes are finished. Using the RAST pipeline (Aziz et al. 2008), the Ca11T genome was
annotated to have 4,796 features including 4,740 protein coding sequences. The genomes of
R. equi strains were also re-annotated using the RAST pipeline to allow an equivalence of
annotation. The Ca11T genome was found to share 4,166 genes with the three R. equi strains
(3,720 with bi-directional and 446 with uni-directional protein BLAST hits; Aziz et al. 2012).
It also shared an additional 128 genes with at least one R. equi strain but not with all three.
446 genes were specific to R. defluvii Ca11T that were absent in the R. equi genomes; 361 of
these encode hypothetical proteins and six belong to mobile genetic elements (transposase,
phage associated or mobile element proteins). A BLAST search of 75 randomly selected
hypothetical proteins of R. defluvii against the NCBI protein database using default settings
revealed homologies for most of them with hypothetical proteins in other rhodococci or other
bacterial species (data not shown), indicating that not all are unique to R. defluvii Ca11T. The
remaining 79 genes specific to R. defluvii Ca11T (compared to the R. equi strains) can
typically be related to known metabolic activities (Table S1), including a gene encoding
alkylphosphonate utilization protein PhnA. The phn operon gene products are involved in the
cleavage of carbon-phosphorus bonds in alkylphosphonates (Chen et al. 1990). However, the
presence of the phnA gene in strain Ca11T is unlikely to be associated with phosphorus
removal in the bioreactor from which it was isolated because the other genes of this operon
are missing. Three homologs of phnB and two homologs of phnE genes were present
elsewhere in the Ca11T genome but they are shared with the R. equi strains. A number of
other genes involved in phosphorus metabolism are also common between R. defluvii and the
three R. equi strains.

An operon in the genome of strain Ca11T that encodes Ter family proteins (TerA,
TerB, TerC-like and two TerD) and associated biosynthetic enzymes is absent from the
genomes of the three R. equi strains (Table S1). Comparable loci have previously been
suggested to be involved in biosynthesis of nucleoside-like metabolites (Anantharaman et al. 2012). The protein BLAST search revealed the presence of homologs of these genes in other rhodococci and actinomycetes, suggesting a potential horizontal acquisition of this operon by \textit{R. defluvii}. Alternatively, this operon may have been lost by \textit{R. equi} as it has adapted to a pathogenic lifestyle. Two of the genes specific to \textit{R. defluvii} Ca11$^T$ (compared to the \textit{R. equi} strains) encode phospholipase C enzymes. Phospholipases C are the virulence factors that induce alveolar macrophage necrosis, resulting in cell death (Assis et al. 2014). As noted above, most of the genes specific to strain Ca11$^T$ encode hypothetical proteins and it is possible that some of these uncharacterized proteins contribute to functional variations between \textit{R. defluvii} and \textit{R. equi}.

\textbf{Rhodococci} are generally involved in environmental processes such as the degradation of organic and xenobiotic substances, except for the pathogens \textit{R. equi} and \textit{Rhodococcus fascians} (Bell et al. 1998; Alvarez 2010). The pathogenicity of these two species has been associated with the presence of large plasmids encoding virulence proteins (Takai et al. 2000; Letek et al. 2008; Francis et al. 2012; Stes et al. 2013). The virulence plasmid in \textit{R. equi} is 80-90 Kb in size and carries a pathogenicity island encoding virulence associated proteins (Vap) while plasmid free strains were found to be avirulent (Takai et al. 2000). A sequence BLAST-based functional comparison using the SEED server (Aziz et al. 2012) revealed the absence of Vap proteins (VapA, C-I proteins from plasmid pVAPA1037 and VapB, J-M from pVAPB1593; Letek et al. 2008) in the draft genome sequence of \textit{R. defluvii}, suggesting the absence of the virulence plasmid in strain Ca11$^T$. However, 228 of the 243 \textit{R. equi} chromosomal virulence-related genes defined by Letek \textit{et al.} (2010) are present in strain Ca11$^T$ (Table S2), including the \textit{esx} cluster. The \textit{paa} operon that was identified in \textit{R. equi} strain ATCC 33707 and which may be involved in pathogenesis in humans (Sangal et al. 2014) is absent from \textit{R. defluvii} strain Ca11$^T$. The presence of a high proportion of virulence-
related genes in the genome of strain Ca11\textsuperscript{T} suggests that this organism may also have the potential to colonise animal hosts. Indeed, it is noted that three additional bacterial strains with 16S rRNA gene sequences identical to that of strain Ca11\textsuperscript{T} have been isolated from salmon intestines (Skrodenyte-Arbaciauskiene, V. & Virbickas T. Genbank accession numbers HM244990, HM244992 and HM244993).

A phylogenetic analysis was performed using PhyloPhlAn (Segata et al. 2013) including \textit{Rhodococcus erythropolis} PR4 (Sekine et al. 2006), \textit{Rhodococcus jostii} RHA1 (McLeod et al. 2006). \textit{Nocardia brasiliensis} ATCC 700358 (Vera-Cabrera et al. 2012) and \textit{Corynebacterium diphtheriae} NCTC 05011 (Sangal et al. 2012) were used as outgroups. PhyloPhlAn automatically extracts the sequences of the 400 most conserved universal proteins that were identified by off-line pre-processing of all available microbial genomes by Segata et al. (2013). It generates highly robust phylogenetic trees from a concatenated alignment of computationally selected subset of amino-acid sequences with highest entropy and an appropriate relative contribution of the most conserved residues from each protein following a maximum likelihood maximization approach (gamma model of rate heterogeneity) with 20 bootstrap replicates using RAxML (Stamatakis 2006). Our PhyloPhlAn analysis showed that \textit{R. defluvii} Ca11\textsuperscript{T} shared a phyletic line with \textit{R. equi} that was relatively distant from the other rhodococci and from \textit{N. brasiliensis} (Fig. 1). BLAST-based average nucleotide identities (ANI\textsubscript{b}) between the genomes of \textit{R. defluvii} Ca11\textsuperscript{T} and the \textit{R. equi} strains were 82.96-83.25\% (Richter and Rosselló-Móra 2009) and average amino acid identities (AAI) varied between 85.31-85.45\%. The ANI\textsubscript{b} and AAI values between \textit{R. defluvii} and the other rhodococci (\textit{R. jostii} RHA1 and \textit{R. erythropolis} PR4) were < 76\% and <72\%, respectively. The digital DNA-DNA hybridization (dDDH) distances were calculated using the genome-to-genome distance calculator at the GGDC 2.0 web server (Auch et al. 2010; Meier-Kolthoff et al. 2013). GGDC values mimic conventional DNA-DNA
hybridization values and have been shown to have very high correlation with 16S rRNA sequence distances (Auch et al. 2010; Meier-Kolthoff et al. 2013). GGDC 2.0 uses three different formulae to calculate the distances and the results of formula-2, which has been recommended for analysing draft genomes (Auch et al. 2010), were considered in this study. The dDDH values between *R. defluvii* and *R. equi* strains C7T, 103S and ATCC 33707 were 26.9 ± 3.02, 27 ± 3.02 and 27.1 ± 3.01, respectively. The *R. defluvii* genome showed lower dDDH similarities with the *R. erythropolis* PR4 (20.2 ± 2.73) and *R. jostii* RHA1 (20.7 ± 2.81) genomes, values that are comparable to the dDDH distances from *N. brasiliensis* ATCC 00358 (20.4 ± 2.63) and *C. diphtheriae* NCTC 05011 (21 ± 2.53). Cumulatively, these results suggest that *R. defluvii* is more closely related to *R. equi* than to other rhodococci, as previously concluded from 16S rRNA gene sequence analysis (Kämpfer et al. 2014).

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

In summary, we report the genome sequence of the type strain of the recently identified species, *R. defluvii* strain Ca11T. The strain is phylogenetically closely related to *R. equi* strains with high similarities both at the nucleotide and functional levels. The whole genome shotgun sequence has been deposited at DDBJ/EMBL/GenBank under the Accession number JPOC00000000. The version described in this study is the first version, JPOC01000000.
Acknowledgements

VS is supported by an Anniversary Research Fellowship from Northumbria University, Newcastle upon Tyne. The authors would like to thank anonymous reviewers for their constructive comments and suggestions. We also thank NU-OMICS facility for assistance in genome sequencing and J. Gibson for IT assistance.
References


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


Takai S et al. (2000) DNA sequence and comparison of virulence plasmids from Rhodococcus equi ATCC 33701 and 103. Infect Immun 68:6840-6847


Figure Legend

Figure 1. Phylogenetic tree (radial, un-rooted) derived from 400 universal proteins using the program PhyloPhlAn showing the relatedness of *R. defluvii* Ca11^T^ with *R. equi* and representatives of other closely related taxa. Scale bar shows normalized fraction of total branch lengths as described by Segata et al. (2013).
Rhodococcus defluvii Ca11T

Rhodococcus equi 103S

Rhodococcus jostii RHA1

Rhodococcus erythropolis PR4

Nocardia brasiliensis ATCC 700358

Corynebacterium diphtheriae NCTC 05011