

# Northumbria Research Link

Citation: Sangal, Vartul, Jones, Amanda, Goodfellow, Michael, Sutcliffe, Iain and Hoskisson, Paul (2014) Comparative genomic analyses reveal a lack of a substantial signature of host adaptation in *Rhodococcus equi* ('Prescottella equi'). *Pathogens and Disease*, 71 (3). pp. 352-356. ISSN 2049-632X

Published by: Wiley-Blackwell

URL: <http://dx.doi.org/10.1111/2049-632X.12126> <<http://dx.doi.org/10.1111/2049-632X.12126>>

This version was downloaded from Northumbria Research Link:  
<http://nrl.northumbria.ac.uk/15405/>

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

**Comparative genomic analyses reveal a lack of a substantial signature of host adaptation in *Rhodococcus equi* (“*Prescottella equi*”)**

Vartul Sangal<sup>1,2</sup>, Amanda L. Jones<sup>2</sup>, Michael Goodfellow<sup>3</sup>, Iain C. Sutcliffe<sup>2</sup> and Paul A. Hoskisson<sup>1</sup>

<sup>1</sup>Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow G4 0RE, UK

<sup>2</sup>Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

<sup>3</sup>School of Biology, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

Correspondence: Paul A. Hoskisson, Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow G4 0RE, UK

Tel.: +44 (0)141 548 2819; Fax: +44 (0)141 548 4124; e-mail: paul.hoskisson@strath.ac.uk

Keywords: *Rhodococcus equi*, “*Prescottella equi*”, opportunistic pathogen, genome, metabolic differences

## Abstract

1  
2  
3  
4  
5  
6  
7  
8  
9  
10

*Rhodococcus equi* (“*Prescottella equi*”) is a pathogenic actinomycete primarily infecting horses but has emerged as an opportunistic human pathogen. We have sequenced the genome of the type strain of this species, *R. equi* strain C7<sup>T</sup>, and compared the genome with that of another foal isolate 103S and of a human isolate ATCC 33707. The *R. equi* strains are closely related to each other and yet distantly related to other rhodococci and *Nocardia brasiliensis*. The comparison of gene contents among *R. equi* strains revealed minor differences that could be associated with host adaptation from foals to humans, including the presence of a *paa* operon in the human isolate which is potentially involved in pathogenesis.

11 *Rhodococcus equi*, a Gram-positive coccobacillus that was formerly classified as  
12 *Corynebacterium equi*, is primarily a foal pathogen that causes fatal lymphadenitis and  
13 ulcerative enteritis in young foals (Prescott, 1991; Giguere, *et al.*, 2011; Vazquez-Boland, *et*  
14 *al.*, 2013). It has recently emerged as an important opportunistic human pathogen causing  
15 significant mortality among immunocompromised patients (Kedlaya, *et al.*, 2001;  
16 Yamshchikov, *et al.*, 2010) and is also associated with a wide range of other vertebrate hosts  
17 and with environmental reservoirs (Prescott, 1991).

18 In this study, we have sequenced the genome of the *R. equi* Type strain C7<sup>T</sup> that was isolated  
19 from the lung abscess of a foal in 1922 (Goodfellow & Alderson, 1977). Comparative analyses  
20 were performed against the available genomes of *R. equi*, namely against those of strains 103S  
21 (Letek, *et al.*, 2010) and ATCC 33707 (Qin, *et al.*, 2010; GenBank Accession numbers  
22 NC\_014659 and NZ\_CM001149, respectively). The *R. equi* C7<sup>T</sup> genome was sequenced on a  
23 Roche 454 GS-Junior instrument as described previously (Sangal, *et al.*, 2012a; Sangal, *et al.*,  
24 2012b). A total of 172,478 reads with an average read length of 464 bp were assembled into  
25 48 contigs (>200 bp) by GS *de novo* assembler v 2.7. The size of assembly was 5.197 MB  
26 with an average of 19 fold coverage. The genome was annotated using the NCBI PGAAP  
27 pipeline ([http://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](http://www.ncbi.nlm.nih.gov/genome/annotation_prok/)) and RAST (Aziz, *et al.*,  
28 2008). The draft genome of *R. equi* strain C7<sup>T</sup> is estimated to have 5,033 features including  
29 4,978 coding sequences and 55 tRNA genes. This genome has been submitted to GenBank  
30 and has the accession number APJC00000000.

31 The *R. equi* 103S and ATCC 33707 genomes were also annotated using RAST to ensure  
32 equivalence of annotation for comparative analyses. The CGView comparative tool analysis  
33 (Grant, *et al.*, 2012) revealed a very high conservation of both the nucleotide (Fig. 1A) and the  
34 protein sequences (Fig. 1B) among these genomes. Indeed, the genome-wide protein Blast  
35 searches, using the SEED Viewer (Aziz, *et al.*, 2012), revealed that only 69, 152 and 187

36 genes, mostly encoding hypothetical proteins, are specific to strains 103S, C7<sup>T</sup> and ATCC  
37 33707, respectively that returned no hits in the other *R. equi* genomes (Supplementary Table 1).  
38 Some of these strain specific genes belonged to mobile genetic elements (integrase, transposase  
39 and repeats) and only 10-13 genes are predicted to be involved in cellular activities  
40 (Supplementary Table 1). Interestingly, strain ATCC 33707 possessed a set of phenylacetate  
41 CoA ligase/oxygenase genes (PaaD, F-I, Z) that are absent in the genomes of strains 103S and  
42 C7<sup>T</sup>. These genes have been implicated in virulence in various pathogens (Law, et al., 2008;  
43 Teufel, et al., 2010). Strain ATCC 33707 was isolated from a human abscess and strains 103S  
44 and C7<sup>T</sup> from equine hosts. Putative PaaE is common to all three isolates and PaaJ and PaaK  
45 subunits of strain ATCC 33707 shared ~30% identities with two proteins (PaaD-like protein  
46 DUF59 involved in Fe-S cluster assembly and probable phenylacetic acid degradation NADH  
47 oxidoreductase PaaE, respectively; data not shown) in the genomes of strains 103S and C7<sup>T</sup>.  
48 The genes encoding PaaD and PaaZ have been identified on an horizontally acquired region by  
49 the Alien Hunter (Vernikos & Parkhill, 2006) and a lack of these genes in equine isolates  
50 suggest that *paa* operon was potentially acquired by strain ATCC 33707 which may be  
51 associated with pathogenesis in human hosts. Alternatively, the *paa* operon may not be  
52 essential for pathogenesis in foals and may have been lost in equine isolates through genome  
53 reduction. Clearly, a wider analysis of a larger collection of *R. equi* isolates from humans and  
54 foals should be performed to confirm this finding.

55 Strain 103S shares 42 genes with strain C7<sup>T</sup> that are absent from the genome of strain ATCC  
56 33707 and 48 genes with strain ATCC 33707 that are absent from the genome of strain C7<sup>T</sup>.  
57 Most of these genes encode hypothetical proteins and only 6-12 of these are potentially  
58 involved in cellular activities, including nucleotide/nucleoside (especially RNA) metabolism,  
59 carbohydrates and sulphur metabolism, and respiration (Supplementary Table 1). A gene  
60 encoding NADH-ubiquinone oxidoreductase chain H (fig|685727.5.peg.1900) and one

61 encoding a glycosyltransferase (fig|685727.5.peg.4060) are common between the genomes of  
62 strains 103S and ATCC 33707 but absent in that of strain C7<sup>T</sup>. Similarly, two genes involved in  
63 sulphur metabolism (fig|685727.5.peg.436, fig|685727.5.peg.1276) are common between the  
64 genomes of strains C7<sup>T</sup> and 103S but missing from the genome of strain ATCC 33707. Strains  
65 C7<sup>T</sup> and ATCC 33707 share 84 genes that are absent from the genome of strain 103S, of which  
66 23 are likely involved in a variety of metabolic activities (Supplementary Table 1). In addition,  
67 some genes showed poor identity (<50% protein sequence identity) between the different  
68 genomes that may also partly contribute to the functional divergence of these strains (data not  
69 shown). These variations potentially reflect the minor differences in lifestyle of individual  
70 strains. Overall, the two equine isolates and the human isolate share most of their gene content  
71 with 109-258 variable genes between strain pairs (Supplementary Table 1). These results are  
72 consistent with a recent report suggesting that *R. equi* isolates are very closely related to each  
73 other (Vazquez-Boland, *et al.*, 2013).

74 Single nucleotide polymorphisms (SNPs) were extracted from the genomic alignment of all  
75 three strains using Mauve (Darling, *et al.*, 2004) and were annotated using 103S as the  
76 reference genome by TRAMS (Reumerman, *et al.*, 2013), after excluding the SNPs with  
77 missing alleles. There were 49,903 SNPs between the genomes of 103S and C7<sup>T</sup>, of which  
78 4,976 were intergenic. These SNPs introduced 28,458 synonymous and 12,804  
79 nonsynonymous changes in the genome of strain C7<sup>T</sup> including 21 stop codons (Supplementary  
80 Table 2). Nine nonsense SNPs were present in genes encoding hypothetical proteins and two  
81 nonsense SNPs were present in a gene encoding a MCE family protein, Mce1F (Supplementary  
82 Table 2). A third nonsense SNP terminated another MCE family membrane protein. MCE  
83 family proteins were originally implicated as having an important role in the entry and survival  
84 of *Mycobacterium tuberculosis* inside host cells (Arruda, *et al.*, 1993) although complex  
85 phenotypes are associated with *mce* loci (Clark *et al.* 2013) and it is likely that the direct

86 function of the MCE proteins is in transport of hydrophobic solutes such as sterols (Mohn *et al.*  
87 2008). Other nonsense SNPs resulted in the inactivation of genes involved in putative ABC  
88 transporter ATP binding protein, putative esterase, YndJ, pyruvate oxidase, ribosyl  
89 nicotinamide transporter (PnuC-like), polyhydroxyalkanoic acid synthase, epoxide hydrolase,  
90 acetate kinase and long-chain fatty acid CoA ligase (Supplementary Table 2). It is not clear  
91 whether these changes reflect a difference in ability of this strain to utilize different carbon  
92 sources or if they have accumulated over long term culture to adapt to an *in vitro* lifestyle.

93 There are 43,350 SNPs between the genomes of strains 103S and ATCC 33707 that introduced  
94 26,312 synonymous and 10,379 nonsynonymous changes including 19 stop codons  
95 (Supplementary Table 2). Seven nonsynonymous changes in genes for an ABC transporter  
96 ATP binding protein, a putative esterase, an acetate kinase and four hypothetical proteins are  
97 shared with the genome of strain C7<sup>T</sup> (Supplementary Table 2). Stop codons also inactivated  
98 eight other hypothetical proteins and genes encoding 3-oxoacyl-[acyl-carrier protein]  
99 reductase, secreted peptidase and an oxetanocin A resistance protein (Supplementary Table 2).  
100 3-oxoacyl-[acyl-carrier protein] reductase is involved in fatty acid biosynthesis by reducing  $\beta$ -  
101 ketoacyl-ACP substrates to beta-hydroxyacyl-ACP products (Lai and Cronan, 2004) but it is  
102 unlikely to affect the strain fitness because there are multiple copies of the gene that encode  
103 this enzyme in the genome.

104 The draft genome of strain C7<sup>T</sup> is ~153 Kb larger than the chromosome of strain 103S but  
105 ~59 Kb smaller than the genome of the ATCC 33707 strain. Strains C7<sup>T</sup> and 103S were  
106 isolated from foals (Goodfellow & Alderson, 1977; de la Pena-Moctezuma & Prescott, 1995)  
107 and strain ATCC 33707 from a human skin abscess (Prescott, 1981). The C7<sup>T</sup> strain was  
108 isolated in 1922 but isolation dates are unknown for the remaining two strains. The smaller  
109 chromosome in strain 103S may represent degradation of the accessory genome that may not  
110 be essential for host infection. To identify horizontal gene acquisition by these strains, the

111 genomes were analysed by Alien Hunter (Vernikos & Parkhill, 2006) which identified 72  
112 acquired regions in 103S (~662 Kb), 144 in C7<sup>T</sup> (~1.2 Mb) and 115 in ATCC 33707 (~972 Kb;  
113 Supplementary Fig. 1). Most of the imports identified in the genome of strain 103S were also  
114 identified in those of strains C7<sup>T</sup> and ATCC 33707, including two large regions that were  
115 previously described (Letek, *et al.*, 2010). The draft status of the C7<sup>T</sup> (48 contigs) and ATCC  
116 33707 (35 contigs) genomes might be largely responsible for identification of higher numbers  
117 of imports as some gaps may have divided them into more fragments. However, the total length  
118 of imports was also variable though most of the genes were conserved among these strains. The  
119 predicted horizontally acquired regions contributed to 376 of the total 582 variable genes  
120 identified by genome-wide protein Blast searches (Supplementary Table 1). The host jumps of  
121 *Staphylococcus aureus* have been characterized by gain or loss of accessory gene pool via  
122 mobile genetic elements (Lowder, *et al.*, 2009; Resch, *et al.*, 2013) and horizontal gene transfer  
123 also appears to have little impact on functional variation among *R. equi* strains (Supplementary  
124 Table 1). Some of the imported genes may be involved in virulence in particular hosts, but their  
125 functions remain to be characterized as most of them encode hypothetical proteins.

126 The virulence plasmid, which is essential for virulence in *R. equi* (Takai, *et al.*, 2000; Letek,  
127 *et al.*, 2008; Vazquez-Boland, *et al.*, 2013), is missing from strains C7<sup>T</sup> and ATCC 33707,  
128 both of which are avirulent (Takai, *et al.*, 1985; Prescott, 1991). The virulence plasmid has  
129 presumably been lost from these strains at some point during long term culture *in vitro*.

130 The phylogenetic relatedness of *R. equi* strains was investigated using PhyloPhlAn, a  
131 program designed to accurately calculate phylogenies from >400 proteins that can potentially  
132 resolve taxonomic groupings (Segata *et al.*, 2013). We also included the genome sequences  
133 of *Rhodococcus erythropolis* PR4 (Sekine, *et al.*, 2006; Letek, *et al.*, 2010), *Rhodococcus*  
134 strain RHA1 (McLeod, *et al.*, 2006), recently formally classified as *Rhodococcus jostii* (Jones  
135 *et al.*, 2013a) and *Nocardia brasiliensis* ATCC 700358 (Vera-Cabrera, *et al.*, 2012) and used



136 *Corynebacterium diphtheriae* NCTC 03529 (Sangal, *et al.*, 2012b) as the outgroup  
137 (Accession numbers NC\_012490, CP000431, NC\_018681 and AJGI00000000, respectively).  
138 All of the *R. equi* strains grouped closely together but distantly to the other rhodococci and *N.*  
139 *brasiliensis* (Fig. 2), which supports our recent proposal of reclassifying *R. equi* into a  
140 separate genus as “*Prescottella equi*” (Jones, *et al.*, 2013b; Jones, *et al.*, 2013c). However, the  
141 name *Prescottella* has still to be validated, although the taxonomic status of *R. equi* has been  
142 questioned (Kämpfer *et al.* 2013). Further analyses incorporating a wider range of genomes of  
143 rhodococci and closely related bacteria are needed to remove any ambiguity over the proposal  
144 for the prospective new genus.

145 In summary, we report the draft genome of the type strain of an important pathogen, *R. equi*  
146 strain C7<sup>T</sup>. *R. equi* strains are phylogenetically closely related and the nucleotide sequence  
147 and functional characteristics are highly conserved within this taxon. *R. equi* has recently  
148 emerged as an important opportunistic human pathogen but the comparison of equine and  
149 human isolates revealed only small differences in the gene content that may be associated  
150 with the host adaptation. The whole Genome Shotgun project of *R. equi* strain C7<sup>T</sup> has been  
151 deposited at DDBJ/EMBL/GenBank under the accession no. APJC000000000. The version  
152 described in this paper is the first version, APJC01000000.

153

#### 154 **Acknowledgements**

155 VS is supported by an Anniversary Research Fellowship at Northumbria University,  
156 Newcastle upon Tyne.

## References

- 157  
158  
159 Arruda S, Bomfim G, Knights R, Huima-Byron T & Riley LW (1993) Cloning of an *M.*  
160 *tuberculosis* DNA fragment associated with entry and survival inside cells. *Science*  
161 261: 1454-1457.
- 162 Aziz RK, Bartels D, Best AA, *et al.* (2008) The RAST Server: rapid annotations using  
163 subsystems technology. *BMC Genomics* **9**: 75.
- 164 Aziz RK, Devoid S, Disz T, *et al.* (2012) SEED servers: high-performance access to the  
165 SEED genomes, annotations, and metabolic models. *PLoS One* **7**: e48053.
- 166 Clark LC, Seipke RF, Prieto P, Willemsse J, van Wezel GP, Hutchings MI & Hoskisson PA  
167 (2013) Mammalian cell entry genes in *Streptomyces* may provide clues to the  
168 evolution of bacterial virulence. *Sci Rep* **3**: 1109.
- 169 Darling AC, Mau B, Blattner FR & Perna NT (2004) Mauve: multiple alignment of  
170 conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394-1403.
- 171 de la Pena-Moctezuma A & Prescott JF (1995) A physical map of the 85 kb virulence  
172 plasmid of *Rhodococcus equi* 103. *Can J Vet Res* **59**: 229-231.
- 173 Giguere S, Cohen ND, Chaffin MK, Hines SA, Hondalus MK, Prescott JF & Slovis NM  
174 (2011) *Rhodococcus equi*: clinical manifestations, virulence, and immunity. *J Vet*  
175 *Intern Med* **25**: 1221-1230.
- 176 Goodfellow M & Alderson G (1977) The actinomycete-genus *Rhodococcus*: a home for the  
177 "rhodochrous" complex. *J Gen Microbiol* **100**: 99-122.
- 178 Grant JR, Arantes AS & Stothard P (2012) Comparing thousands of circular genomes using  
179 the CGView Comparison Tool. *BMC Genomics* **13**: 202.
- 180 Jones AL, Davies J, Fakuda M, Brown R, Lim J & Goodfellow M (2013a) *Rhodococcus*  
181 *jostii*: A home for *Rhodococcus* strain RHA1. *Antonie van Leeuwenhoek* **104**:435-440.

182 Jones AL, Sutcliffe IC & Goodfellow M (2013b) *Prescottia equi* gen. nov., comb. nov.: a  
183 new home for an old pathogen. *Antonie Van Leeuwenhoek* **103**: 655-671.

184 Jones AL, Sutcliffe IC & Goodfellow M (2013c) Proposal to replace the illegitimate genus  
185 name *Prescottia* Jones et al. 2013 with the genus name *Prescottella* gen. nov. and to  
186 replace the illegitimate combination *Prescottia equi* Jones et al. 2013 with  
187 *Prescottella equi* comb. nov. *Antonie Van Leeuwenhoek* **103**: 1405-1407.

188 Kämpfer P, Dott W, Martin K & Glaeser SP (2013) *Rhodococcus defluvii* sp. nov. isolated  
189 from wastewater of a bioreactor and formal proposal to reclassify '*Corynebacterium*'  
190 *hoagii* and *Rhodococcus equi* as *Rhodococcus hoagii*. *Int J Syst Evol Microbiol*. In  
191 press.

192 Kedlaya I, Ing MB & Wong SS (2001) *Rhodococcus equi* infections in immunocompetent  
193 hosts: case report and review. *Clin Infect Dis* **32**: E39-46.

194 Lai CY & Cronan JE (2004) Isolation and characterization of beta-ketoacyl-acyl carrier  
195 protein reductase (fabG) mutants of *Escherichia coli* and *Salmonella enterica* serovar  
196 Typhimurium. *J Bacteriol* **186**: 1869-1878.

197 Law RJ, Hamlin JN, Sivro A, McCorrister SJ, Cardama GA & Cardona ST (2008) A  
198 functional phenylacetic acid catabolic pathway is required for full pathogenicity of  
199 *Burkholderia cenocepacia* in the *Caenorhabditis elegans* host model. *J Bacteriol* **190**:  
200 7209-7218.

201 Letek M, Ocampo-Sosa AA, Sanders M, *et al.* (2008) Evolution of the *Rhodococcus equi*  
202 vap pathogenicity island seen through comparison of host-associated vapA and vapB  
203 virulence plasmids. *J Bacteriol* **190**: 5797-5805.

204 Letek M, Gonzalez P, Macarthur I, *et al.* (2010) The genome of a pathogenic *Rhodococcus*:  
205 cooptive virulence underpinned by key gene acquisitions. *PLoS Genet* **6**: e1001145.

206 Lowder BV, Guinane CM, Ben Zakour NL, *et al.* (2009) Recent human-to-poultry host  
207 jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc Natl Acad Sci*  
208 *U S A* **106**: 19545-19550.

209 McLeod MP, Warren RL, Hsiao WW, *et al.* (2006) The complete genome of *Rhodococcus*  
210 sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A*  
211 **103**: 15582-15587.

212 Mohn WW, van der Geize R, Stewart GR, Okamoto S, Liu J, Dijkhuizen L & Eltis LD  
213 (2008) The actinobacterial *mce4* locus encodes a steroid transporter. *J Biol Chem* 283:  
214 35368-35374.

215 Prescott JF (1981) Capsular serotypes of *Corynebacterium equi*. *Can J Comp Med* 45: 130-  
216 134.

217 Prescott JF (1991) *Rhodococcus equi*: an animal and human pathogen. *Clin Microbiol Rev*  
218 **4**: 20-34.

219 Qin X, Bachman B, Battles P, *et al.* (2010) *Rhodococcus equi* ATCC 33707, whole genome  
220 shotgun sequencing. <http://www.ncbi.nlm.nih.gov/nucore/325556670>.

221 Resch G, Francois P, Morisset D, *et al.* (2013) Human-to-bovine jump of *Staphylococcus*  
222 *aureus* CC8 is associated with the loss of a beta-hemolysin converting prophage and  
223 the acquisition of a new staphylococcal cassette chromosome. *PLoS One* **8**: e58187.

224 Reumerman RA, Tucker NP, Herron PR, Hoskisson PA & Sangal V (2013) Tool for Rapid  
225 Annotation of Microbial SNPs (TRAMS): a simple program for rapid annotation of  
226 genomic variation in prokaryotes. *Antonie Van Leeuwenhoek* **104**: 431-434.

227 Sangal V, Tucker NP, Burkovski A & Hoskisson PA (2012a) Draft genome sequence of  
228 *Corynebacterium diphtheriae* biovar intermedius NCTC 5011. *J Bacteriol* **194**: 4738.

229 Sangal V, Tucker NP, Burkovski A & Hoskisson PA (2012b) The draft genome sequence of  
230 *Corynebacterium diphtheriae* bv. mitis NCTC 3529 reveals significant diversity  
231 between the primary disease-causing biovars. *J Bacteriol* **194**: 3269.

232 Segata N, Bornigen D, Morgan XC & Huttenhower C (2013) PhyloPhlAn is a new method  
233 for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**:  
234 2304.

235 Sekine M, Tanikawa S, Omata S, *et al.* (2006) Sequence analysis of three plasmids  
236 harboured in *Rhodococcus erythropolis* strain PR4. *Environ Microbiol* **8**: 334-346.

237 Takai S, Michizoe T, Matsumura K, Nagai M, Sato H & Tsubaki S (1985) Correlation of in  
238 vitro properties of *Rhodococcus (Corynebacterium) equi* with virulence for mice.  
239 *Microbiol Immunol* **29**: 1175-1184.

240 Takai S, Hines SA, Sekizaki T, *et al.* (2000) DNA sequence and comparison of virulence  
241 plasmids from *Rhodococcus equi* ATCC 33701 and 103. *Infect Immun* **68**: 6840-6847.

242 Teufel R, Mascaraque V, Ismail W, *et al.* (2010) Bacterial phenylalanine and phenylacetate  
243 catabolic pathway revealed. *Proc Natl Acad Sci U S A* **107**: 14390-14395.

244 Vazquez-Boland JA, Giguere S, Hapeshi A, Macarthur I, Anastasi E & Valero-Rello A  
245 (2013) *Rhodococcus equi*: The many facets of a pathogenic actinomycete. *Vet*  
246 *Microbiol.* In press.

247 Vera-Cabrera L, Ortiz-Lopez R, Elizondo-Gonzalez R, Perez-Maya AA & Ocampo-  
248 Candiani J (2012) Complete genome sequence of *Nocardia brasiliensis* HUJEG-1. *J*  
249 *Bacteriol* **194**: 2761-2762.

250 Vernikos GS & Parkhill J (2006) Interpolated variable order motifs for identification of  
251 horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands.  
252 *Bioinformatics* **22**: 2196-2203.

253 Yamshchikov AV, Schuetz A & Lyon GM (2010) *Rhodococcus equi* infection. *Lancet*  
254 *Infect Dis* **10**: 350-359.  
255

256

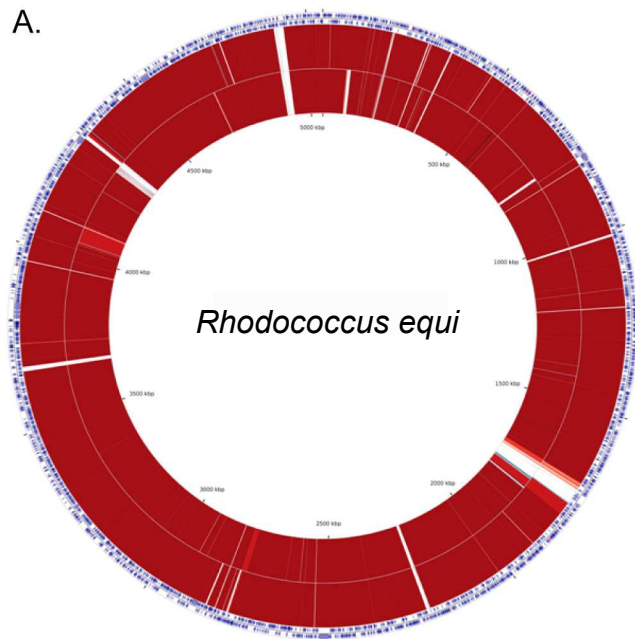
## Figure Legends

257

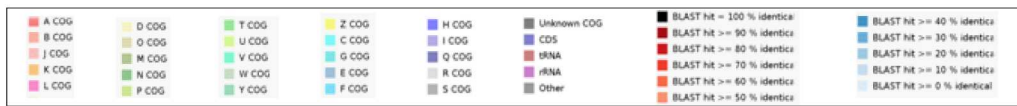
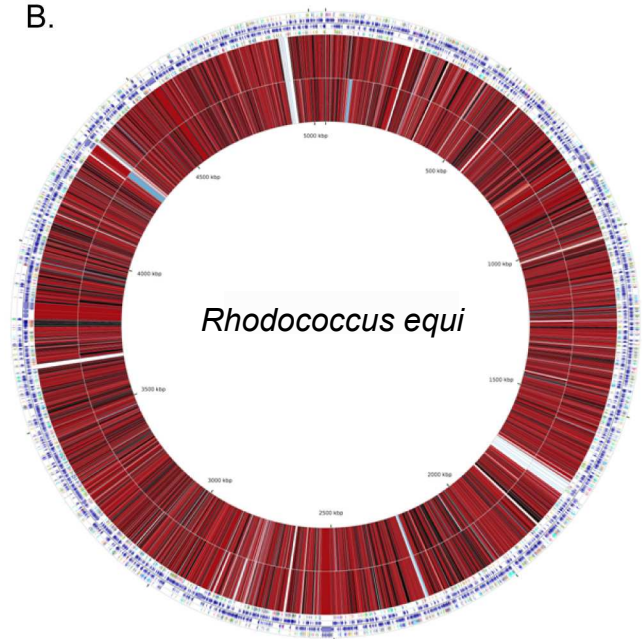
258 **Figure 1.** Genomic blast maps of *R. equi* strain 103S against ATCC 33707 and C7<sup>T</sup>. (A)  
259 DNA blast map, (B) CDS blast map. The colour scale (inset) shows the level of sequence  
260 identity with the respective sequences from strain C7<sup>T</sup> (middle circle) and strain ATCC 33707  
261 (inner circle).

262 **Figure 2.** Phylogenetic tree from >400 universal proteins showing the genomic relationship  
263 of *R. equi* to representatives of closely related taxa. Scale bar shows mean nucleotide  
264 substitutions per site.

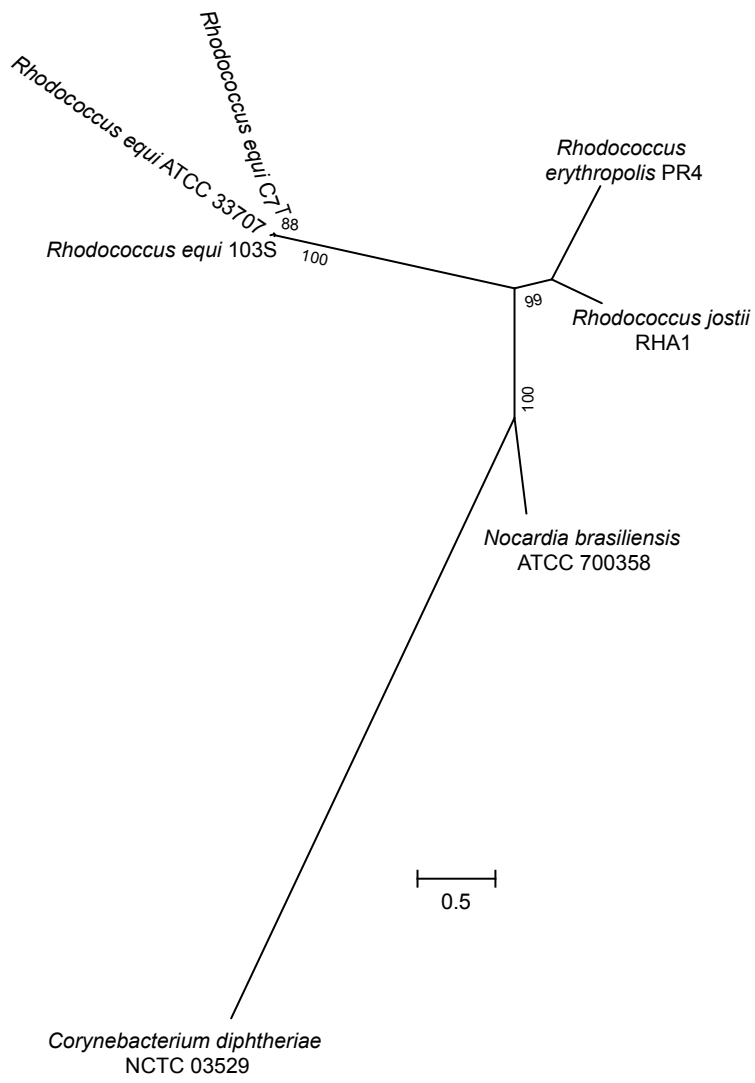
A.

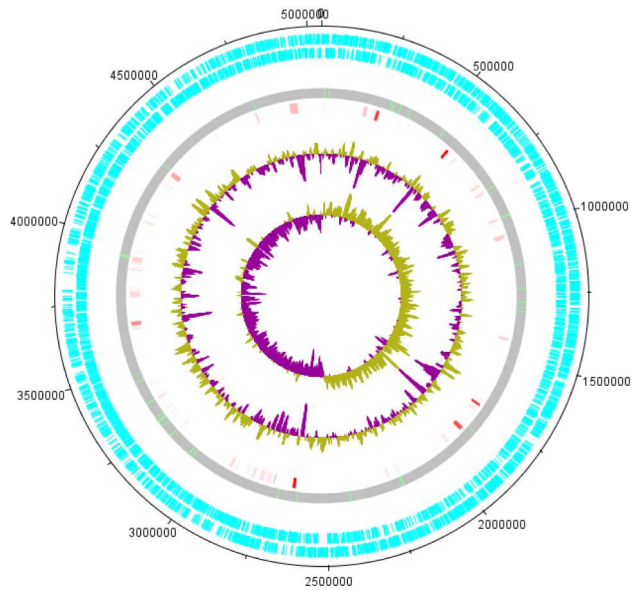
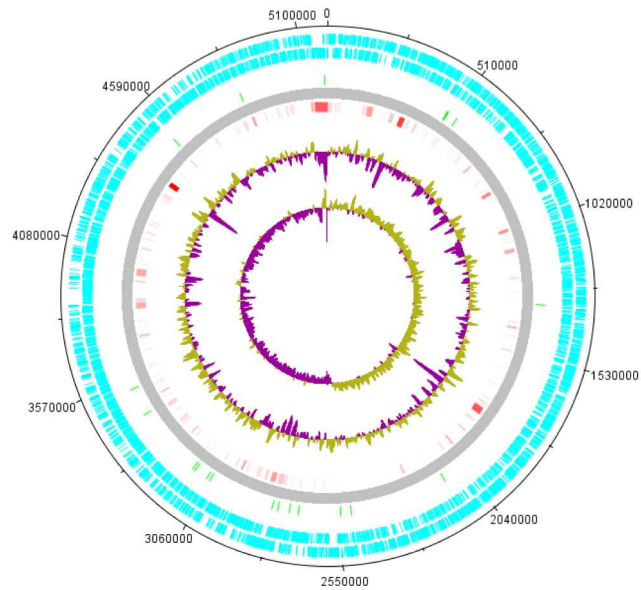
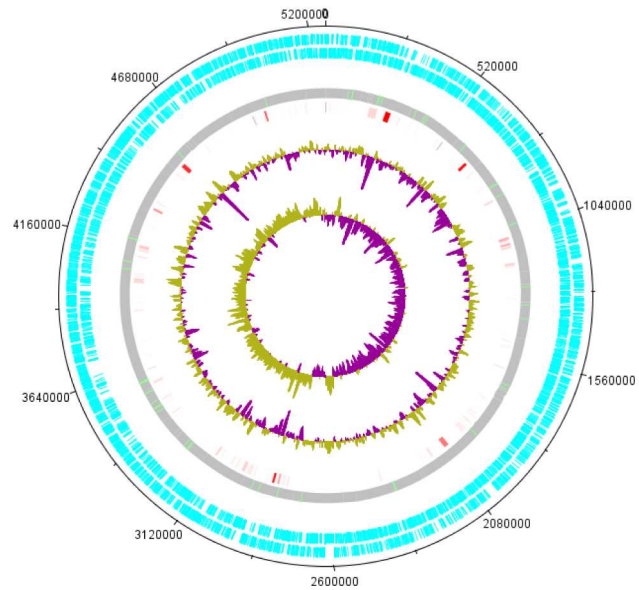


B.







*A. R. equi* 103S*B. R. equi* C7T*C. R. equi* ATCC 33707

Supplementary Figure 1. Circular diagrams of *R. equi* chromosomes. Outer two rings are the coding sequences on forward and reverse strands. The genomic islands predicted by AlienHunter are shown in red colour followed by GC plot and GC skew, respectively.

Supplementary Table 1. Differences in the gene content between 103S, ATCC33707 and 103S based on the bidirectional protein blast searches

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	41	fig 685727.5.peg.65 (Q)	fig 525370.5.peg.4858 (97.5)		ATCC33707
hypothetical protein	216	fig 685727.5.peg.69 (Q)	fig 525370.5.peg.4854 (100)		ATCC33707
Hydantoin racemase (EC 5.1.99.-)	252	fig 685727.5.peg.75 (Q)	fig 525370.5.peg.4848 (99.6)		
hypothetical protein	64	fig 685727.5.peg.98 (Q)		fig 1219013.3.peg.93 (98.41)	
hypothetical protein	44	fig 685727.5.peg.101 (Q)	fig 525370.5.peg.4821 (100)		
FIG022958: hypothetical protein	210	fig 685727.5.peg.122 (Q)	fig 525370.5.peg.4977 (100)		
Permease of the drug/metabolite transporter (DMT) superfamily	51	fig 685727.5.peg.242 (Q)		fig 1219013.3.peg.240 (97.73)	
hypothetical protein	69	fig 685727.5.peg.264 (Q)		fig 1219013.3.peg.268 (100)	C7
hypothetical protein	110	fig 685727.5.peg.265 (Q)		fig 1219013.3.peg.269 (98.11)	C7
hypothetical protein	151	fig 685727.5.peg.266 (Q)		fig 1219013.3.peg.270 (98.67)	C7
hypothetical protein	39	fig 685727.5.peg.300 (Q)	fig 525370.5.peg.3529 (100)		
hypothetical protein	56	fig 685727.5.peg.306 (Q)			
hypothetical protein	178	fig 685727.5.peg.313 (Q)			
Nitric-oxide reductase (EC 1.7.99.7), quinol-dependent	778	fig 685727.5.peg.328 (Q)			
hypothetical protein	221	fig 685727.5.peg.329 (Q)			
hypothetical protein	56	fig 685727.5.peg.404 (Q)			
Arylsulfatase (EC 3.1.6.1)	48	fig 685727.5.peg.436 (Q)		fig 1219013.3.peg.462 (100)	
hypothetical protein	42	fig 685727.5.peg.461 (Q)		fig 1219013.3.peg.487 (100)	C7
hypothetical protein	85	fig 685727.5.peg.477 (Q)	fig 525370.5.peg.713 (100)		
hypothetical protein	98	fig 685727.5.peg.484 (Q)		fig 1219013.3.peg.514 (98.97)	
hypothetical protein	42	fig 685727.5.peg.509 (Q)	fig 525370.5.peg.2715 (100)		
hypothetical protein	66	fig 685727.5.peg.558 (Q)			
hypothetical protein	50	fig 685727.5.peg.563 (Q)			
hypothetical protein	71	fig 685727.5.peg.650 (Q)			
hypothetical protein	149	fig 685727.5.peg.759 (Q)		fig 1219013.3.peg.838 (98.65)	
putative ABC transporter ATP binding protein	612	fig 685727.5.peg.791 (Q)		fig 1219013.3.peg.4071 (28.26)	
hypothetical protein	62	fig 685727.5.peg.892 (Q)		fig 1219013.3.peg.987 (90.16)	
hypothetical protein	71	fig 685727.5.peg.921 (Q)	fig 525370.5.peg.2059 (100)		
hypothetical protein	39	fig 685727.5.peg.947 (Q)			
hypothetical protein	40	fig 685727.5.peg.960 (Q)			103S
hypothetical protein	61	fig 685727.5.peg.961 (Q)			103S
Putative membrane protein	182	fig 685727.5.peg.981 (Q)		fig 1219013.3.peg.1085 (98.34)	
hypothetical protein	51	fig 685727.5.peg.1008 (Q)		fig 1219013.3.peg.1113 (100)	
hypothetical protein	43	fig 685727.5.peg.1059 (Q)			
hypothetical protein	202	fig 685727.5.peg.1103 (Q)	fig 525370.5.peg.3802 (97.5)		
hypothetical protein	45	fig 685727.5.peg.1129 (Q)		fig 1219013.3.peg.1237 (100)	
hypothetical protein	53	fig 685727.5.peg.1285 (Q)			
hypothetical protein	77	fig 685727.5.peg.1334 (Q)	fig 525370.5.peg.3954 (98.67)		
hypothetical protein	43	fig 685727.5.peg.1461 (Q)			
hypothetical protein	60	fig 685727.5.peg.1552 (Q)		fig 1219013.3.peg.1665 (98.31)	C7
hypothetical protein	53	fig 685727.5.peg.1639 (Q)		fig 1219013.3.peg.1757 (100)	C7
hypothetical protein	200	fig 685727.5.peg.1640 (Q)		fig 1219013.3.peg.1757 (98.99)	
hypothetical protein	139	fig 685727.5.peg.1641 (Q)	fig 525370.5.peg.1120 (42.97)		
hypothetical protein	152	fig 685727.5.peg.1645 (Q)	fig 525370.5.peg.1116 (84.21)		ATCC33707-103S
hypothetical protein	321	fig 685727.5.peg.1646 (Q)	fig 525370.5.peg.1084 (31.85)		ATCC33707-103S
hypothetical protein	487	fig 685727.5.peg.1647 (Q)			103S
hypothetical protein	255	fig 685727.5.peg.1649 (Q)	fig 525370.5.peg.1109 (98.82)		ATCC33707-103S
DNA helicase related protein	1828	fig 685727.5.peg.1655 (Q)	fig 525370.5.peg.1095 (23.18)		ATCC33707-103S
hypothetical protein	92	fig 685727.5.peg.1656 (Q)			103S
hypothetical protein	924	fig 685727.5.peg.1658 (Q)			103S
Helicase, C-terminal:Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	2130	fig 685727.5.peg.1659 (Q)			103S
Type II restriction enzyme, methylase subunits	1530	fig 685727.5.peg.1660 (Q)			103S
putative ATP-dependent helicase	960	fig 685727.5.peg.1661 (Q)	fig 525370.5.peg.2722 (22.42)		
hypothetical protein	1135	fig 685727.5.peg.1662 (Q)			
hypothetical protein	114	fig 685727.5.peg.1664 (Q)			103S
hypothetical protein	444	fig 685727.5.peg.1665 (Q)			103S
hypothetical protein	63	fig 685727.5.peg.1666 (Q)			103S
hypothetical protein	118	fig 685727.5.peg.1667 (Q)			103S
hypothetical protein	94	fig 685727.5.peg.1669 (Q)	fig 525370.5.peg.1077 (86.02)		ATCC33707-103S
hypothetical protein	173	fig 685727.5.peg.1670 (Q)	fig 525370.5.peg.1076 (88.95)		
hypothetical protein	55	fig 685727.5.peg.1681 (Q)	fig 525370.5.peg.1064 (100)		
cyclase/dehydrase	173	fig 685727.5.peg.1707 (Q)	fig 525370.5.peg.1038 (100)		
hypothetical protein	72	fig 685727.5.peg.1708 (Q)	fig 525370.5.peg.1037 (98.59)		
putative isomerase	122	fig 685727.5.peg.1709 (Q)	fig 525370.5.peg.1036 (100)		
NADH-ubiquinone oxidoreductase chain H (EC 1.6.5.3)	435	fig 685727.5.peg.1900 (Q)	fig 525370.5.peg.4090 (99.08)		
hypothetical protein	38	fig 685727.5.peg.1908 (Q)			

Table S1 continued...

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	65	fig 685727.5.peg.1998 (Q)			
COG1359: Uncharacterized conserved protein	79	fig 685727.5.peg.2112 (Q)			103S
hypothetical protein	61	fig 685727.5.peg.2114 (Q)			103S
hypothetical protein	123	fig 685727.5.peg.2116 (Q)			103S
Non-heme chloroperoxidase (EC 1.11.1.10)	52	fig 685727.5.peg.2118 (Q)			
alpha/beta hydrolase fold( EC:1.11.1.10)	64	fig 685727.5.peg.2119 (Q)			
hypothetical protein	38	fig 685727.5.peg.2121 (Q)			
hypothetical protein	82	fig 685727.5.peg.2133 (Q)			
hypothetical protein	58	fig 685727.5.peg.2218 (Q)			
hypothetical protein	38	fig 685727.5.peg.2265 (Q)		fig 1219013.3.peg.2377 (100)	
hypothetical protein	116	fig 685727.5.peg.2347 (Q)	fig 525370.5.peg.4220 (88.89)		
hypothetical protein	57	fig 685727.5.peg.2390 (Q)	fig 525370.5.peg.4177 (100)		
hypothetical protein	444	fig 685727.5.peg.2393 (Q)			
hypothetical protein	43	fig 685727.5.peg.2401 (Q)			
hypothetical protein	82	fig 685727.5.peg.2435 (Q)	fig 525370.5.peg.1805 (98.77)		
hypothetical protein	84	fig 685727.5.peg.2475 (Q)	fig 525370.5.peg.1704 (98.78)		ATCC33707-103S
hypothetical protein	67	fig 685727.5.peg.2533 (Q)		fig 1219013.3.peg.2646 (98.48)	
hypothetical protein	42	fig 685727.5.peg.2566 (Q)			103S
hypothetical protein	56	fig 685727.5.peg.2569 (Q)	fig 525370.5.peg.1612 (100)		ATCC33707-103S
hypothetical protein	86	fig 685727.5.peg.2580 (Q)		fig 1219013.3.peg.2691 (100)	C7-103S
FIG01136303: hypothetical protein	115	fig 685727.5.peg.2585 (Q)		fig 1219013.3.peg.2699 (95.12)	103S
hypothetical protein	86	fig 685727.5.peg.2586 (Q)	fig 525370.5.peg.1595 (95.56)		ATCC33707-103S
hypothetical protein	84	fig 685727.5.peg.2615 (Q)	fig 525370.5.peg.1965 (89.16)		ATCC33707-103S
hypothetical protein	260	fig 685727.5.peg.2617 (Q)		fig 1219013.3.peg.2734 (96.91)	C7-103S
hypothetical protein	100	fig 685727.5.peg.2618 (Q)		fig 1219013.3.peg.2735 (94.95)	C7-103S
putative acetyltransferase	143	fig 685727.5.peg.2658 (Q)			103S
hypothetical protein	195	fig 685727.5.peg.2660 (Q)			103S
hypothetical protein	288	fig 685727.5.peg.2684 (Q)		fig 1219013.3.peg.2803 (100)	C7-103S
hypothetical protein	333	fig 685727.5.peg.2685 (Q)		fig 1219013.3.peg.2804 (99.7)	C7-103S
hypothetical protein	356	fig 685727.5.peg.2687 (Q)		fig 1219013.3.peg.2806 (98.87)	C7-103S
hypothetical protein	425	fig 685727.5.peg.2688 (Q)		fig 1219013.3.peg.4267 (29.87)	103S
hypothetical protein	171	fig 685727.5.peg.2701 (Q)		fig 1219013.3.peg.2818 (97.06)	C7
hypothetical protein	38	fig 685727.5.peg.3012 (Q)			
hypothetical protein	48	fig 685727.5.peg.3094 (Q)			
hypothetical protein	43	fig 685727.5.peg.3095 (Q)		fig 1219013.3.peg.3228 (92.86)	C7
hypothetical protein	64	fig 685727.5.peg.3272 (Q)		fig 1219013.3.peg.3411 (96.83)	
hypothetical protein	48	fig 685727.5.peg.3310 (Q)	fig 525370.5.peg.3144 (100)		ATCC33707-103S
hypothetical protein	38	fig 685727.5.peg.3318 (Q)			
hypothetical protein	63	fig 685727.5.peg.3447 (Q)			103S
hypothetical protein	65	fig 685727.5.peg.3448 (Q)			103S
hypothetical protein	124	fig 685727.5.peg.3449 (Q)			103S
hypothetical protein	256	fig 685727.5.peg.3450 (Q)			103S
hypothetical protein	281	fig 685727.5.peg.3451 (Q)			103S
hypothetical protein	613	fig 685727.5.peg.3452 (Q)		fig 1219013.3.peg.1777 (32.39)	C7-103S
hypothetical protein	38	fig 685727.5.peg.3480 (Q)		fig 1219013.3.peg.3618 (100)	
hypothetical protein	97	fig 685727.5.peg.3522 (Q)		fig 1219013.3.peg.3660 (97.92)	
YgjD/Kae1/Qri7 family, required for threonylcarbamoyladenine (t(6)A) formation in tRNA	347	fig 685727.5.peg.3523 (Q)		fig 1219013.3.peg.3661 (99.71)	
hypothetical protein	167	fig 685727.5.peg.3799 (Q)	fig 525370.5.peg.1498 (96.99)		
hypothetical protein	161	fig 685727.5.peg.3853 (Q)			103S
hypothetical protein	45	fig 685727.5.peg.4055 (Q)			103S
hypothetical protein	48	fig 685727.5.peg.4057 (Q)			103S
GumJ protein	504	fig 685727.5.peg.4058 (Q)			103S
hypothetical protein	363	fig 685727.5.peg.4059 (Q)			103S
hypothetical protein	44	fig 685727.5.peg.4063 (Q)	fig 525370.5.peg.673 (95.35)		ATCC33707-103S
hypothetical protein	106	fig 685727.5.peg.4065 (Q)			103S
hypothetical protein	47	fig 685727.5.peg.4066 (Q)	fig 525370.5.peg.671 (89.13)		ATCC33707-103S
hypothetical protein	90	fig 685727.5.peg.4069 (Q)	fig 525370.5.peg.668 (100)		ATCC33707-103S
hypothetical protein	54	fig 685727.5.peg.4140 (Q)	fig 525370.5.peg.4557 (91.07)		
hypothetical protein	53	fig 685727.5.peg.4187 (Q)		fig 1219013.3.peg.4353 (98.08)	
hypothetical protein	38	fig 685727.5.peg.4392 (Q)	fig 525370.5.peg.513 (100)		
hypothetical protein	44	fig 685727.5.peg.4462 (Q)		fig 1219013.3.peg.4643 (97.67)	C7-103S
hypothetical protein	39	fig 685727.5.peg.4465 (Q)		fig 1219013.3.peg.4646 (100)	C7-103S
hypothetical protein	42	fig 685727.5.peg.4466 (Q)		fig 1219013.3.peg.4647 (97.56)	C7-103S
hypothetical protein	39	fig 685727.5.peg.4517 (Q)		fig 1219013.3.peg.4700 (100)	
hypothetical protein	88	fig 685727.5.peg.4520 (Q)		fig 1219013.3.peg.4703 (93.1)	
hypothetical protein	45	fig 685727.5.peg.4524 (Q)	fig 525370.5.peg.382 (97.73)		ATCC33707

Table S1 continued...

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	48	fig 685727.5.peg.4559 (Q)	fig 525370.5.peg.347 (95.74)		
hypothetical protein	56	fig 685727.5.peg.4616 (Q)			103S
hypothetical protein	48	fig 685727.5.peg.4617 (Q)			103S
hypothetical protein	41	fig 685727.5.peg.4618 (Q)			103S
similar to Ketosteroid isomerase-related protein	136	fig 685727.5.peg.4620 (Q)			103S
hypothetical protein	45	fig 685727.5.peg.4621 (Q)			103S
Transposase	80	fig 685727.5.peg.4622 (Q)			103S
transposase, IS4 family protein	99	fig 685727.5.peg.4623 (Q)			103S
Mobile element protein	71	fig 685727.5.peg.4624 (Q)			103S
hypothetical protein	85	fig 685727.5.peg.4627 (Q)			103S
hypothetical protein	466	fig 685727.5.peg.4628 (Q)			103S
hypothetical protein	245	fig 685727.5.peg.4629 (Q)			103S
Putative DNA-binding protein	277	fig 685727.5.peg.4631 (Q)	fig 525370.5.peg.1035 (31.78)		103S
hypothetical protein	41	fig 685727.5.peg.4632 (Q)			103S
hypothetical protein	849	fig 685727.5.peg.4634 (Q)			103S
hypothetical protein	255	fig 685727.5.peg.4635 (Q)			103S
Integrase	279	fig 685727.5.peg.4636 (Q)			103S
hypothetical protein	932	fig 685727.5.peg.4637 (Q)			103S
hypothetical protein	138	fig 685727.5.peg.4639 (Q)	fig 525370.5.peg.287 (96.35)		ATCC33707-103S
hypothetical protein	42	fig 685727.5.peg.4734 (Q)	fig 525370.5.peg.4931 (100)		ATCC33707-103S
rRNA small subunit 7-methylguanosine (m7G) methyltransferase GidB	224	fig 685727.5.peg.4735 (Q)	fig 525370.5.peg.4930 (96.41)		ATCC33707-103S
hypothetical protein	40		fig 525370.5.peg.120 (Q)		
hypothetical protein	41		fig 525370.5.peg.134 (Q)		
hypothetical protein	40		fig 525370.5.peg.344 (Q)		
hypothetical protein	38		fig 525370.5.peg.440 (Q)		
hypothetical protein	611		fig 525370.5.peg.464 (Q)		ATCC33707
ATP-dependent DNA helicase UvrD/PcrA	567		fig 525370.5.peg.465 (Q)		ATCC33707
hypothetical protein	39		fig 525370.5.peg.479 (Q)		ATCC33707
hypothetical protein	306		fig 525370.5.peg.621 (Q)		
hypothetical protein	583		fig 525370.5.peg.622 (Q)		
hypothetical protein	45		fig 525370.5.peg.645 (Q)		ATCC33707
Glycosyltransferase	130	fig 685727.5.peg.4060 (32.74)	fig 525370.5.peg.675 (Q)		ATCC33707
FIG00511468: hypothetical protein	535		fig 525370.5.peg.676 (Q)		ATCC33707
hypothetical protein	112		fig 525370.5.peg.681 (Q)		ATCC33707
3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	259		fig 525370.5.peg.751 (Q)		ATCC33707
hypothetical protein	211		fig 525370.5.peg.752 (Q)	fig 1219013.3.peg.4138 (24.56)	ATCC33707
Phenylacetic acid degradation protein PaaD, thioesterase	140		fig 525370.5.peg.754 (Q)		ATCC33707
hypothetical protein	105		fig 525370.5.peg.755 (Q)		ATCC33707
Aldehyde dehydrogenase (EC 1.2.1.3), PaaZ	682		fig 525370.5.peg.756 (Q)		ATCC33707
Transcriptional regulator, TetR family	179		fig 525370.5.peg.757 (Q)		ATCC33707
Phenylacetate-CoA oxygenase, PaaG subunit	317		fig 525370.5.peg.761 (Q)		
Phenylacetate-CoA oxygenase, PaaH subunit	108		fig 525370.5.peg.762 (Q)		
Phenylacetate-CoA oxygenase, PaaI subunit	306		fig 525370.5.peg.763 (Q)		
Phenylacetate-coenzyme A ligase (EC 6.2.1.30) PaaF	433		fig 525370.5.peg.766 (Q)		
hypothetical protein	798	fig 685727.5.peg.1668 (82.28)	fig 525370.5.peg.1078 (Q)		ATCC33707
hypothetical protein	253		fig 525370.5.peg.1079 (Q)		ATCC33707
hypothetical protein	97		fig 525370.5.peg.1080 (Q)		ATCC33707
hypothetical protein	160		fig 525370.5.peg.1081 (Q)		ATCC33707
hypothetical protein	104		fig 525370.5.peg.1082 (Q)		ATCC33707
hypothetical protein	142		fig 525370.5.peg.1083 (Q)		ATCC33707
hypothetical protein	318		fig 525370.5.peg.1085 (Q)		ATCC33707
hypothetical protein	49		fig 525370.5.peg.1086 (Q)		ATCC33707
hypothetical protein	51		fig 525370.5.peg.1087 (Q)		ATCC33707
ATP/GTP binding protein	627		fig 525370.5.peg.1088 (Q)		ATCC33707
hypothetical protein	283		fig 525370.5.peg.1089 (Q)		ATCC33707
hypothetical protein	296		fig 525370.5.peg.1090 (Q)		ATCC33707
hypothetical protein	164		fig 525370.5.peg.1092 (Q)		ATCC33707
hypothetical protein	749		fig 525370.5.peg.1096 (Q)		ATCC33707
hypothetical protein	290		fig 525370.5.peg.1097 (Q)		ATCC33707
hypothetical protein	388		fig 525370.5.peg.1098 (Q)		ATCC33707
hypothetical protein	340		fig 525370.5.peg.1099 (Q)		
hypothetical protein	843		fig 525370.5.peg.1100 (Q)		
FIG00995584: hypothetical protein	656		fig 525370.5.peg.1101 (Q)		
hypothetical protein	917		fig 525370.5.peg.1102 (Q)		
FIG00995267: hypothetical protein	1191		fig 525370.5.peg.1103 (Q)		ATCC33707
FIG00998105: hypothetical protein	922		fig 525370.5.peg.1104 (Q)		ATCC33707

Table S1 continued...

Function	Protein Length	1035 Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	97		fig 525370.5.peg.1105 (Q)		ATCC33707
hypothetical protein	78		fig 525370.5.peg.1111 (Q)		ATCC33707
hypothetical protein	120		fig 525370.5.peg.1112 (Q)		ATCC33707
hypothetical protein	201		fig 525370.5.peg.1113 (Q)		ATCC33707
hypothetical protein	98		fig 525370.5.peg.1114 (Q)		ATCC33707
hypothetical protein	45		fig 525370.5.peg.1115 (Q)		ATCC33707
Partial REP13E12 repeat protein	195		fig 525370.5.peg.1224 (Q)		ATCC33707
hypothetical protein	73		fig 525370.5.peg.1425 (Q)		
hypothetical protein	39		fig 525370.5.peg.1441 (Q)		ATCC33707
hypothetical protein	44		fig 525370.5.peg.1444 (Q)		ATCC33707
hypothetical protein	99		fig 525370.5.peg.1599 (Q)		ATCC33707
hypothetical protein	128		fig 525370.5.peg.1600 (Q)		ATCC33707
hypothetical protein	242		fig 525370.5.peg.1721 (Q)		ATCC33707
hypothetical protein	55		fig 525370.5.peg.1722 (Q)		ATCC33707
hypothetical protein	85		fig 525370.5.peg.1723 (Q)		ATCC33707
hypothetical protein	91		fig 525370.5.peg.1724 (Q)		ATCC33707
Phage Gp37Gp68 protein	39		fig 525370.5.peg.1725 (Q)		ATCC33707
hypothetical protein	240		fig 525370.5.peg.1726 (Q)		ATCC33707
hypothetical protein	515		fig 525370.5.peg.1727 (Q)		ATCC33707
hypothetical protein	206		fig 525370.5.peg.1728 (Q)		ATCC33707
putative phage tail	357		fig 525370.5.peg.1729 (Q)		ATCC33707
hypothetical protein	134		fig 525370.5.peg.1730 (Q)		
hypothetical protein	199		fig 525370.5.peg.1732 (Q)		
hypothetical protein	375		fig 525370.5.peg.1733 (Q)		
hypothetical protein	480		fig 525370.5.peg.1734 (Q)	fig 1219013.3.peg.4933 (21.38)	
hypothetical protein	290		fig 525370.5.peg.1735 (Q)		
tail length tape measure protein	1610		fig 525370.5.peg.1736 (Q)	fig 1219013.3.peg.4935 (25.45)	ATCC33707
hypothetical protein	121		fig 525370.5.peg.1737 (Q)		ATCC33707
hypothetical protein	148		fig 525370.5.peg.1738 (Q)		ATCC33707
phage major tail protein	186		fig 525370.5.peg.1739 (Q)		ATCC33707
hypothetical protein	126		fig 525370.5.peg.1740 (Q)		ATCC33707
hypothetical protein	68		fig 525370.5.peg.1741 (Q)		
hypothetical protein	78		fig 525370.5.peg.1742 (Q)		
hypothetical protein	120		fig 525370.5.peg.1743 (Q)		
hypothetical protein	139		fig 525370.5.peg.1744 (Q)		
hypothetical protein	152		fig 525370.5.peg.1745 (Q)		
hypothetical protein	308		fig 525370.5.peg.1746 (Q)		ATCC33707
hypothetical protein	137		fig 525370.5.peg.1747 (Q)		ATCC33707
hypothetical protein	422		fig 525370.5.peg.1749 (Q)		ATCC33707
phage terminase, large subunit, putative	490		fig 525370.5.peg.1750 (Q)		ATCC33707
hypothetical protein	116		fig 525370.5.peg.1751 (Q)		ATCC33707
hypothetical protein	46		fig 525370.5.peg.1752 (Q)		ATCC33707
hypothetical protein	107		fig 525370.5.peg.1753 (Q)		ATCC33707
hypothetical protein	86		fig 525370.5.peg.1754 (Q)		ATCC33707
hypothetical protein	66		fig 525370.5.peg.1755 (Q)		ATCC33707
hypothetical protein	126		fig 525370.5.peg.1756 (Q)		ATCC33707
hypothetical protein	222		fig 525370.5.peg.1757 (Q)	fig 1219013.3.peg.4952 (31.48)	ATCC33707
hypothetical protein	65		fig 525370.5.peg.1758 (Q)		ATCC33707
hypothetical protein	149		fig 525370.5.peg.1759 (Q)		ATCC33707
hypothetical protein	143		fig 525370.5.peg.1760 (Q)		ATCC33707
hypothetical protein	92		fig 525370.5.peg.1761 (Q)		ATCC33707
hypothetical protein	59		fig 525370.5.peg.1762 (Q)		ATCC33707
hypothetical protein	129		fig 525370.5.peg.1763 (Q)		ATCC33707
hypothetical protein	112		fig 525370.5.peg.1764 (Q)		
hypothetical protein	104		fig 525370.5.peg.1766 (Q)		
hypothetical protein	88		fig 525370.5.peg.1767 (Q)		
hypothetical protein	174		fig 525370.5.peg.1768 (Q)		
Cell division protein FtsK	378		fig 525370.5.peg.1770 (Q)	fig 1219013.3.peg.2065 (54.84)	
hypothetical protein	332		fig 525370.5.peg.1771 (Q)		ATCC33707
hypothetical protein	53		fig 525370.5.peg.1772 (Q)		ATCC33707
hypothetical protein	83		fig 525370.5.peg.1773 (Q)		ATCC33707
hypothetical protein	92		fig 525370.5.peg.1774 (Q)		ATCC33707
hypothetical protein	102		fig 525370.5.peg.1775 (Q)		ATCC33707
hypothetical protein	124		fig 525370.5.peg.1776 (Q)		ATCC33707
hypothetical protein	177		fig 525370.5.peg.1777 (Q)		ATCC33707
hypothetical protein	68		fig 525370.5.peg.1778 (Q)		ATCC33707

Table S1 continued...

Function	Protein Length	1035 Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	79		fig 525370.5.peg.1779 (Q)		ATCC33707
hypothetical protein	328		fig 525370.5.peg.1887 (Q)	fig 1219013.3.peg.2816 (96.15)	
hypothetical protein	46		fig 525370.5.peg.1888 (Q)		
hypothetical protein	50		fig 525370.5.peg.1898 (Q)		ATCC33707
hypothetical protein	242		fig 525370.5.peg.2073 (Q)		
hypothetical protein	59		fig 525370.5.peg.2089 (Q)		
hypothetical protein	49		fig 525370.5.peg.2105 (Q)		ATCC33707
hypothetical protein	76		fig 525370.5.peg.2106 (Q)		ATCC33707
hypothetical protein	71		fig 525370.5.peg.2107 (Q)		ATCC33707
hypothetical protein	258		fig 525370.5.peg.2108 (Q)		ATCC33707
hypothetical protein	43		fig 525370.5.peg.2109 (Q)		ATCC33707
hypothetical protein	42		fig 525370.5.peg.2110 (Q)		ATCC33707
hypothetical protein	66		fig 525370.5.peg.2111 (Q)		ATCC33707
hypothetical protein	45		fig 525370.5.peg.2112 (Q)		ATCC33707
hypothetical protein	60		fig 525370.5.peg.2113 (Q)		ATCC33707
hypothetical protein	65		fig 525370.5.peg.2114 (Q)		ATCC33707
hypothetical protein	249		fig 525370.5.peg.2117 (Q)		ATCC33707
hypothetical protein	43		fig 525370.5.peg.2118 (Q)		ATCC33707
hypothetical protein	59		fig 525370.5.peg.2119 (Q)		ATCC33707
hypothetical protein	72		fig 525370.5.peg.2120 (Q)		ATCC33707
hypothetical protein	143		fig 525370.5.peg.2121 (Q)		ATCC33707
hypothetical protein	120		fig 525370.5.peg.2122 (Q)		ATCC33707
hypothetical protein	152		fig 525370.5.peg.2123 (Q)		ATCC33707
hypothetical protein	51		fig 525370.5.peg.2124 (Q)		ATCC33707
hypothetical protein	40		fig 525370.5.peg.2125 (Q)		ATCC33707
hypothetical protein	130		fig 525370.5.peg.2127 (Q)		ATCC33707
hypothetical protein	58		fig 525370.5.peg.2128 (Q)		ATCC33707
hypothetical protein	56		fig 525370.5.peg.2129 (Q)		ATCC33707
hypothetical protein	113		fig 525370.5.peg.2130 (Q)		ATCC33707
Phage terminase	515		fig 525370.5.peg.2131 (Q)		ATCC33707
hypothetical protein	424		fig 525370.5.peg.2132 (Q)		ATCC33707
hypothetical protein	476		fig 525370.5.peg.2133 (Q)		ATCC33707
dehydrogenase / Cytidine deaminase (EC 3.5.4.5)	147		fig 525370.5.peg.2134 (Q)		
hypothetical protein	314		fig 525370.5.peg.2135 (Q)		
hypothetical protein	129		fig 525370.5.peg.2136 (Q)		
hypothetical protein	124		fig 525370.5.peg.2137 (Q)		
hypothetical protein	110		fig 525370.5.peg.2138 (Q)		ATCC33707
Sigma 54-dependent transcriptional activator	94		fig 525370.5.peg.2139 (Q)		ATCC33707
hypothetical protein	133		fig 525370.5.peg.2140 (Q)		ATCC33707
hypothetical protein	268		fig 525370.5.peg.2141 (Q)		ATCC33707
hypothetical protein	88		fig 525370.5.peg.2142 (Q)		ATCC33707
hypothetical protein	38		fig 525370.5.peg.2143 (Q)		ATCC33707
hypothetical protein	98		fig 525370.5.peg.2144 (Q)		ATCC33707
hypothetical protein	149		fig 525370.5.peg.2145 (Q)		ATCC33707
hypothetical protein	151		fig 525370.5.peg.2146 (Q)		ATCC33707
hypothetical protein	1178		fig 525370.5.peg.2147 (Q)	fig 1219013.3.peg.4935 (33.5)	ATCC33707
hypothetical protein	112		fig 525370.5.peg.2150 (Q)		ATCC33707
hypothetical protein	123		fig 525370.5.peg.2152 (Q)		ATCC33707
hypothetical protein	92		fig 525370.5.peg.2153 (Q)		ATCC33707
hypothetical protein	62		fig 525370.5.peg.2154 (Q)		ATCC33707
hypothetical protein	118		fig 525370.5.peg.2155 (Q)		ATCC33707
hypothetical protein	288		fig 525370.5.peg.2156 (Q)		ATCC33707
hypothetical protein	112		fig 525370.5.peg.2157 (Q)		ATCC33707
hypothetical protein	111		fig 525370.5.peg.2158 (Q)		ATCC33707
hypothetical protein	115		fig 525370.5.peg.2159 (Q)		ATCC33707
hypothetical protein	329		fig 525370.5.peg.2160 (Q)		ATCC33707
hypothetical protein	73		fig 525370.5.peg.2161 (Q)		ATCC33707
hypothetical protein	72		fig 525370.5.peg.2162 (Q)		ATCC33707
hypothetical protein	53		fig 525370.5.peg.2163 (Q)		ATCC33707
hypothetical protein	40		fig 525370.5.peg.2164 (Q)		ATCC33707
Integrase-like protein	275		fig 525370.5.peg.2165 (Q)		ATCC33707
hypothetical protein	38		fig 525370.5.peg.2166 (Q)		ATCC33707
hypothetical protein	249		fig 525370.5.peg.2167 (Q)		ATCC33707
hypothetical protein	54		fig 525370.5.peg.2168 (Q)		ATCC33707
hypothetical protein	119		fig 525370.5.peg.2169 (Q)		ATCC33707
hypothetical protein	67		fig 525370.5.peg.2298 (Q)		

Table S1 continued...

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	330		fig 525370.5.peg.2559 (Q)		ATCC33707
hypothetical protein	52		fig 525370.5.peg.2598 (Q)		ATCC33707
hypothetical protein	58		fig 525370.5.peg.2640 (Q)		ATCC33707
Arsenical pump-driving ATPase (EC 3.6.3.16)	380		fig 525370.5.peg.2730 (Q)	fig 1219013.3.peg.532 (100)	
hypothetical protein	82		fig 525370.5.peg.2971 (Q)		
hypothetical protein	79		fig 525370.5.peg.3033 (Q)		
hypothetical protein	61		fig 525370.5.peg.3517 (Q)		
Acyl dehydratase	128		fig 525370.5.peg.3541 (Q)		
FIG00827952: hypothetical protein	420		fig 525370.5.peg.3544 (Q)		
hypothetical protein	52		fig 525370.5.peg.3659 (Q)		
hypothetical protein	38		fig 525370.5.peg.3771 (Q)		ATCC33707
hypothetical protein	38		fig 525370.5.peg.3875 (Q)		
hypothetical protein	164		fig 525370.5.peg.3918 (Q)		
hypothetical protein	80		fig 525370.5.peg.4018 (Q)		ATCC33707
hypothetical protein	619		fig 525370.5.peg.4388 (Q)		
hypothetical protein	44		fig 525370.5.peg.4466 (Q)		
Mobile element protein	285		fig 525370.5.peg.4685 (Q)		
transposase for IS3517	109		fig 525370.5.peg.4686 (Q)		
hypothetical protein	49		fig 525370.5.peg.4845 (Q)		
2,3-dihydroxybiphenyl 1,2-dioxygenase	293	fig 685727.5.peg.72 (100)	fig 525370.5.peg.4851 (Q)		
hypothetical protein	62		fig 525370.5.peg.4873 (Q)		
rRNA small subunit 7-methylguanosine (m7G) methyltransferase GidB	226	fig 685727.5.peg.4735 (95.52)	fig 525370.5.peg.4928 (Q)		ATCC33707-103S
hypothetical protein	54		fig 525370.5.peg.4839 (97.3)	fig 1219013.3.peg.76 (Q)	
hypothetical protein	109			fig 1219013.3.peg.177 (Q)	C7
Glycerol-3-phosphate regulon repressor, DeoR family	277		fig 525370.5.peg.2517 (99.64)	fig 1219013.3.peg.258 (Q)	ATCC33707-C7
hypothetical protein	71		fig 525370.5.peg.2514 (98.57)	fig 1219013.3.peg.261 (Q)	ATCC33707-C7
Glycerol-3-phosphate ABC transporter, periplasmic glycerol-3-phosphate-binding protein (TC 3.A.1.1.3)	394		fig 525370.5.peg.2513 (100)	fig 1219013.3.peg.262 (Q)	ATCC33707-C7
Integrase	287		fig 525370.5.peg.1780 (28.82)	fig 1219013.3.peg.304 (Q)	C7
hypothetical protein	290			fig 1219013.3.peg.305 (Q)	C7
hypothetical protein	346			fig 1219013.3.peg.306 (Q)	C7
hypothetical protein	52			fig 1219013.3.peg.307 (Q)	C7
hypothetical protein	54			fig 1219013.3.peg.308 (Q)	C7
hypothetical protein	114			fig 1219013.3.peg.309 (Q)	C7
hypothetical protein	156			fig 1219013.3.peg.310 (Q)	C7
hypothetical protein	205			fig 1219013.3.peg.311 (Q)	C7
hypothetical protein	51			fig 1219013.3.peg.312 (Q)	C7
hypothetical protein	257			fig 1219013.3.peg.314 (Q)	C7
Transcriptional regulator	351			fig 1219013.3.peg.315 (Q)	C7
hypothetical protein	136			fig 1219013.3.peg.316 (Q)	C7
hypothetical protein	332			fig 1219013.3.peg.317 (Q)	C7
hypothetical protein	305			fig 1219013.3.peg.318 (Q)	C7
hypothetical protein	133			fig 1219013.3.peg.319 (Q)	C7
hypothetical protein	80			fig 1219013.3.peg.320 (Q)	C7
hypothetical protein	66			fig 1219013.3.peg.321 (Q)	C7
hypothetical protein	66			fig 1219013.3.peg.322 (Q)	C7
hypothetical protein	156			fig 1219013.3.peg.334 (Q)	
hypothetical protein	101		fig 525370.5.peg.3516 (97)	fig 1219013.3.peg.335 (Q)	
hypothetical protein	194		fig 525370.5.peg.794 (96.02)	fig 1219013.3.peg.435 (Q)	ATCC33707-C7
Putative polysaccharide deacetylase	294		fig 525370.5.peg.790 (99.66)	fig 1219013.3.peg.439 (Q)	ATCC33707-C7
L-asparaginase (EC 3.5.1.1)	337		fig 525370.5.peg.789 (100)	fig 1219013.3.peg.440 (Q)	
hypothetical protein	163			fig 1219013.3.peg.494 (Q)	
hypothetical protein	65			fig 1219013.3.peg.506 (Q)	
hypothetical protein	91			fig 1219013.3.peg.530 (Q)	
WhiB-type transcription regulator	53			fig 1219013.3.peg.531 (Q)	
Arsenical pump-driving ATPase (EC 3.6.3.16)	347		fig 525370.5.peg.2729 (100)	fig 1219013.3.peg.533 (Q)	
FIG011121: hypothetical protein	44		fig 525370.5.peg.2728 (100)	fig 1219013.3.peg.534 (Q)	
Enoyl-CoA hydratase (EC 4.2.1.17)	269		fig 525370.5.peg.2695 (100)	fig 1219013.3.peg.566 (Q)	
hypothetical protein	369		fig 525370.5.peg.2693 (100)	fig 1219013.3.peg.568 (Q)	
hypothetical protein	350		fig 525370.5.peg.2687 (100)	fig 1219013.3.peg.574 (Q)	ATCC33707-C7
hypothetical protein	131		fig 525370.5.peg.2681 (97.69)	fig 1219013.3.peg.580 (Q)	C7
Transcriptional regulator, IclR family	273		fig 525370.5.peg.2674 (99.26)	fig 1219013.3.peg.587 (Q)	
hypothetical protein	170		fig 525370.5.peg.2667 (98.82)	fig 1219013.3.peg.594 (Q)	ATCC33707-C7
Beta-mannanase-like protein	329		fig 525370.5.peg.2627 (98.78)	fig 1219013.3.peg.633 (Q)	
Cellulose synthase catalytic subunit	658		fig 525370.5.peg.2625 (99.24)	fig 1219013.3.peg.635 (Q)	
hypothetical protein	146		fig 525370.5.peg.2623 (99.31)	fig 1219013.3.peg.637 (Q)	



Table S1 continued...

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
FIG00994841: hypothetical protein	90		fig 525370.5.peg.2460 (100)	fig 1219013.3.peg.675 (Q)	
hypothetical protein	51		fig 525370.5.peg.2410 (94)	fig 1219013.3.peg.725 (Q)	C7
hypothetical protein	121			fig 1219013.3.peg.766 (Q)	
hypothetical protein	73		fig 525370.5.peg.2349 (94.44)	fig 1219013.3.peg.787 (Q)	
hypothetical protein	84		fig 525370.5.peg.2301 (94.74)	fig 1219013.3.peg.835 (Q)	
hypothetical protein	147			fig 1219013.3.peg.837 (Q)	
hypothetical protein	64		fig 525370.5.peg.2262 (98.41)	fig 1219013.3.peg.874 (Q)	
hypothetical protein	109		fig 525370.5.peg.2259 (100)	fig 1219013.3.peg.877 (Q)	
hypothetical protein	45		fig 525370.5.peg.2220 (100)	fig 1219013.3.peg.916 (Q)	
hypothetical protein	40		fig 525370.5.peg.2192 (100)	fig 1219013.3.peg.944 (Q)	
hypothetical protein	53			fig 1219013.3.peg.967 (Q)	C7
hypothetical protein	41			fig 1219013.3.peg.968 (Q)	C7
hypothetical protein	40			fig 1219013.3.peg.969 (Q)	C7
hypothetical protein	43			fig 1219013.3.peg.970 (Q)	C7
hypothetical protein	51			fig 1219013.3.peg.999 (Q)	C7
Methyltransferase type 12	252			fig 1219013.3.peg.1001 (Q)	C7
hypothetical protein	38			fig 1219013.3.peg.1097 (Q)	C7
Prophage Lp2 protein 6	374			fig 1219013.3.peg.1189 (Q)	C7
hypothetical protein	73		fig 525370.5.peg.3791 (98.61)	fig 1219013.3.peg.1219 (Q)	
hypothetical protein	38			fig 1219013.3.peg.1227 (Q)	
hypothetical protein	393		fig 525370.5.peg.3707 (98.08)	fig 1219013.3.peg.1306 (Q)	
hypothetical protein	72			fig 1219013.3.peg.1476 (Q)	
hypothetical protein	43			fig 1219013.3.peg.1480 (Q)	
OsmC-like protein	136			fig 1219013.3.peg.1729 (Q)	
hypothetical protein	230			fig 1219013.3.peg.1758 (Q)	C7
hypothetical protein	55			fig 1219013.3.peg.1762 (Q)	C7
hypothetical protein	38			fig 1219013.3.peg.1764 (Q)	C7
hypothetical protein	38			fig 1219013.3.peg.1768 (Q)	C7
hypothetical protein	62			fig 1219013.3.peg.1769 (Q)	C7
hypothetical protein	77			fig 1219013.3.peg.1770 (Q)	C7
YeeC-like protein	402			fig 1219013.3.peg.1771 (Q)	C7
FIG006126: DNA helicase, restriction/modification system component YeeB	657			fig 1219013.3.peg.1772 (Q)	C7
FIG045374: Type II restriction enzyme, methylase subunit YeeA	901			fig 1219013.3.peg.1773 (Q)	C7
hypothetical protein	230			fig 1219013.3.peg.1774 (Q)	C7
hypothetical protein	275			fig 1219013.3.peg.1775 (Q)	C7
hypothetical protein	187			fig 1219013.3.peg.1776 (Q)	C7
Fe-S protein, homolog of lactate dehydrogenase SO1521	107		fig 525370.5.peg.1073 (95.74)	fig 1219013.3.peg.1778 (Q)	C7
hypothetical protein	45			fig 1219013.3.peg.1779 (Q)	C7
hypothetical protein	184			fig 1219013.3.peg.1815 (Q)	C7
hypothetical protein	513			fig 1219013.3.peg.1822 (Q)	C7
hypothetical protein	190			fig 1219013.3.peg.1823 (Q)	C7
hypothetical protein	242			fig 1219013.3.peg.1824 (Q)	C7
hypothetical protein	233			fig 1219013.3.peg.1831 (Q)	
KH domain RNA binding protein YlqC	81		fig 525370.5.peg.946 (100)	fig 1219013.3.peg.1916 (Q)	
hypothetical protein	43			fig 1219013.3.peg.2163 (Q)	
hypothetical protein	40			fig 1219013.3.peg.2285 (Q)	
hypothetical protein	60			fig 1219013.3.peg.2297 (Q)	
hypothetical protein	59			fig 1219013.3.peg.2322 (Q)	
hypothetical protein	58			fig 1219013.3.peg.2417 (Q)	
hypothetical protein	69			fig 1219013.3.peg.2511 (Q)	C7
hypothetical protein	377		fig 525370.5.peg.4061 (37.2)	fig 1219013.3.peg.2531 (Q)	
hypothetical protein	157			fig 1219013.3.peg.2544 (Q)	C7
Protein of unknown function DUF262	606			fig 1219013.3.peg.2545 (Q)	C7
Uncharacterized protein SCO1141	138			fig 1219013.3.peg.2547 (Q)	C7
hypothetical protein	78		fig 525370.5.peg.1796 (98.57)	fig 1219013.3.peg.2556 (Q)	
hypothetical protein	40			fig 1219013.3.peg.2693 (Q)	C7
hypothetical protein	46			fig 1219013.3.peg.2694 (Q)	C7
Glycosyltransferase	334		fig 525370.5.peg.1985 (98.21)	fig 1219013.3.peg.2712 (Q)	
hypothetical protein	71			fig 1219013.3.peg.2722 (Q)	C7
hypothetical protein	74			fig 1219013.3.peg.2732 (Q)	C7
hypothetical protein	120			fig 1219013.3.peg.2733 (Q)	C7
hypothetical protein	133		fig 525370.5.peg.1925 (100)	fig 1219013.3.peg.2776 (Q)	C7
hypothetical protein	38		fig 525370.5.peg.1924 (100)	fig 1219013.3.peg.2777 (Q)	ATCC33707-C7
hypothetical protein	60		fig 525370.5.peg.1907 (98.31)	fig 1219013.3.peg.2794 (Q)	
hypothetical protein	67			fig 1219013.3.peg.2846 (Q)	C7
hypothetical protein	57			fig 1219013.3.peg.2868 (Q)	

Table S1 continued...

Function	Protein Length	1035 Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	141	<a href="#">fig 685727.5.peg.266 (32)</a>		<a href="#">fig 1219013.3.peg.2873 (Q)</a>	
hypothetical protein	203		<a href="#">fig 525370.5.peg.4357 (99.5)</a>	<a href="#">fig 1219013.3.peg.2924 (Q)</a>	
Oxidoreductase, short-chain dehydrogenase/reductase family (EC 1.1.1.-)	211			<a href="#">fig 1219013.3.peg.2961 (Q)</a>	
hypothetical protein	150			<a href="#">fig 1219013.3.peg.3004 (Q)</a>	
hypothetical protein	85		<a href="#">fig 525370.5.peg.4395 (100)</a>	<a href="#">fig 1219013.3.peg.3026 (Q)</a>	
hypothetical protein	148			<a href="#">fig 1219013.3.peg.3074 (Q)</a>	
hypothetical protein	39			<a href="#">fig 1219013.3.peg.3084 (Q)</a>	C7
hypothetical protein	106		<a href="#">fig 525370.5.peg.3360 (97.12)</a>	<a href="#">fig 1219013.3.peg.3227 (Q)</a>	C7
hypothetical protein	43		<a href="#">fig 525370.5.peg.3348 (100)</a>	<a href="#">fig 1219013.3.peg.3240 (Q)</a>	
hypothetical protein	38			<a href="#">fig 1219013.3.peg.3280 (Q)</a>	
Aspartyl-tRNA(Asn) amidotransferase subunit A (EC 6.3.5.6) / Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.7)	67		<a href="#">fig 525370.5.peg.3297 (98.48)</a>	<a href="#">fig 1219013.3.peg.3292 (Q)</a>	
hypothetical protein	47		<a href="#">fig 525370.5.peg.3287 (95.65)</a>	<a href="#">fig 1219013.3.peg.3302 (Q)</a>	C7
hypothetical protein	70			<a href="#">fig 1219013.3.peg.3342 (Q)</a>	
hypothetical protein	56			<a href="#">fig 1219013.3.peg.3346 (Q)</a>	
hypothetical protein	63			<a href="#">fig 1219013.3.peg.3393 (Q)</a>	C7
possible sugar phosphate isomerase/ epimerase	274		<a href="#">fig 525370.5.peg.3096 (99.63)</a>	<a href="#">fig 1219013.3.peg.3494 (Q)</a>	
FIG00995642: hypothetical protein	365		<a href="#">fig 525370.5.peg.3094 (99.18)</a>	<a href="#">fig 1219013.3.peg.3496 (Q)</a>	
hypothetical protein	38			<a href="#">fig 1219013.3.peg.3626 (Q)</a>	
putative transposase	89		<a href="#">fig 525370.5.peg.2750 (80.23)</a>	<a href="#">fig 1219013.3.peg.3841 (Q)</a>	
putative acyltransferase domain protein	66			<a href="#">fig 1219013.3.peg.3842 (Q)</a>	
hypothetical protein	39			<a href="#">fig 1219013.3.peg.3914 (Q)</a>	
Ferrichrome ABC transporter	322		<a href="#">fig 525370.5.peg.1527 (99.38)</a>	<a href="#">fig 1219013.3.peg.3919 (Q)</a>	
hypothetical protein	48		<a href="#">fig 525370.5.peg.1526 (95.12)</a>	<a href="#">fig 1219013.3.peg.3920 (Q)</a>	
hypothetical protein	50		<a href="#">fig 525370.5.peg.1518 (100)</a>	<a href="#">fig 1219013.3.peg.3928 (Q)</a>	ATCC33707-C7
Adenylate kinase (EC 2.7.4.3)	173			<a href="#">fig 1219013.3.peg.3948 (Q)</a>	
hypothetical protein	248		<a href="#">fig 525370.5.peg.1440 (77.42)</a>	<a href="#">fig 1219013.3.peg.4001 (Q)</a>	ATCC33707-C7
hypothetical protein	38			<a href="#">fig 1219013.3.peg.4007 (Q)</a>	
hypothetical protein	40			<a href="#">fig 1219013.3.peg.4017 (Q)</a>	C7
hypothetical protein	40			<a href="#">fig 1219013.3.peg.4090 (Q)</a>	
hypothetical protein	679		<a href="#">fig 525370.5.peg.1248 (99.56)</a>	<a href="#">fig 1219013.3.peg.4197 (Q)</a>	ATCC33707-C7
hypothetical protein	723			<a href="#">fig 1219013.3.peg.4209 (Q)</a>	
hypothetical protein	76			<a href="#">fig 1219013.3.peg.4212 (Q)</a>	C7
hypothetical protein	457			<a href="#">fig 1219013.3.peg.4213 (Q)</a>	C7
hypothetical protein	75			<a href="#">fig 1219013.3.peg.4214 (Q)</a>	C7
hypothetical protein	380			<a href="#">fig 1219013.3.peg.4215 (Q)</a>	C7
putative glycosyl transferase	336			<a href="#">fig 1219013.3.peg.4216 (Q)</a>	C7
hypothetical protein	73			<a href="#">fig 1219013.3.peg.4219 (Q)</a>	C7
hypothetical protein	222			<a href="#">fig 1219013.3.peg.4221 (Q)</a>	C7
hypothetical protein	39			<a href="#">fig 1219013.3.peg.4246 (Q)</a>	
Believed to be an alternative form of N-formylglutamate deformylase (EC 3.5.1.68), but experimentally invalidated	435		<a href="#">fig 525370.5.peg.619 (97.15)</a>	<a href="#">fig 1219013.3.peg.4270 (Q)</a>	
hypothetical protein	860		<a href="#">fig 525370.5.peg.616 (99.28)</a>	<a href="#">fig 1219013.3.peg.4273 (Q)</a>	
hypothetical protein	229		<a href="#">fig 525370.5.peg.614 (98.68)</a>	<a href="#">fig 1219013.3.peg.4275 (Q)</a>	
ABC transporter, periplasmic spermidine putrescine-binding protein PotD (TC 3.A.1.11.1)	336		<a href="#">fig 525370.5.peg.610 (100)</a>	<a href="#">fig 1219013.3.peg.4279 (Q)</a>	ATCC33707-C7
Transcriptional regulator, IclR family	248		<a href="#">fig 525370.5.peg.609 (99.19)</a>	<a href="#">fig 1219013.3.peg.4280 (Q)</a>	ATCC33707-C7
Gentisate 1,2-dioxygenase (EC 1.13.11.4)	357			<a href="#">fig 1219013.3.peg.4369 (Q)</a>	
FMN reductase (EC 1.5.1.29)	199	<a href="#">fig 685727.5.peg.1276 (27.27)</a>		<a href="#">fig 1219013.3.peg.4371 (Q)</a>	
Transcriptional regulator, IclR family	278		<a href="#">fig 525370.5.peg.2082 (27.54)</a>	<a href="#">fig 1219013.3.peg.4374 (Q)</a>	
hypothetical protein	60			<a href="#">fig 1219013.3.peg.4377 (Q)</a>	
hypothetical protein	48		<a href="#">fig 525370.5.peg.4657 (97.87)</a>	<a href="#">fig 1219013.3.peg.4390 (Q)</a>	
hypothetical protein	102		<a href="#">fig 525370.5.peg.540 (92.08)</a>	<a href="#">fig 1219013.3.peg.4546 (Q)</a>	
hypothetical protein	233		<a href="#">fig 525370.5.peg.510 (97.5)</a>	<a href="#">fig 1219013.3.peg.4573 (Q)</a>	ATCC33707-C7
hypothetical protein	48			<a href="#">fig 1219013.3.peg.4606 (Q)</a>	C7
hypothetical protein	318		<a href="#">fig 525370.5.peg.401 (99.37)</a>	<a href="#">fig 1219013.3.peg.4686 (Q)</a>	
regulatory protein, TetR	208		<a href="#">fig 525370.5.peg.400 (98.55)</a>	<a href="#">fig 1219013.3.peg.4687 (Q)</a>	
hypothetical protein	39			<a href="#">fig 1219013.3.peg.4711 (Q)</a>	
hypothetical protein	401		<a href="#">fig 525370.5.peg.345 (98.75)</a>	<a href="#">fig 1219013.3.peg.4743 (Q)</a>	
hypothetical protein	222		<a href="#">fig 525370.5.peg.328 (98.19)</a>	<a href="#">fig 1219013.3.peg.4759 (Q)</a>	ATCC33707-C7
hypothetical protein	93			<a href="#">fig 1219013.3.peg.4810 (Q)</a>	
Enoyl-CoA hydratase (EC 4.2.1.17)	213		<a href="#">fig 525370.5.peg.3654 (31.58)</a>	<a href="#">fig 1219013.3.peg.4824 (Q)</a>	C7
phage integrase family protein	398		<a href="#">fig 525370.5.peg.1780 (28.89)</a>	<a href="#">fig 1219013.3.peg.4900 (Q)</a>	C7
hypothetical protein	56			<a href="#">fig 1219013.3.peg.4901 (Q)</a>	C7
hypothetical protein	72			<a href="#">fig 1219013.3.peg.4902 (Q)</a>	C7
hypothetical protein	201			<a href="#">fig 1219013.3.peg.4903 (Q)</a>	C7
hypothetical protein	129			<a href="#">fig 1219013.3.peg.4905 (Q)</a>	C7
hypothetical protein	117			<a href="#">fig 1219013.3.peg.4906 (Q)</a>	C7

Table S1 continued...

Function	Protein Length	103S Feature (blast hit)	ATCC33707 Feature (blast hit)	C7 Feature (blast hit)	HGT
hypothetical protein	110			fig 1219013.3.peg.4907 (Q)	C7
hypothetical protein	134			fig 1219013.3.peg.4908 (Q)	C7
hypothetical protein	332			fig 1219013.3.peg.4909 (Q)	C7
hypothetical protein	123			fig 1219013.3.peg.4910 (Q)	C7
hypothetical protein	97			fig 1219013.3.peg.4911 (Q)	C7
hypothetical protein	47			fig 1219013.3.peg.4924 (Q)	C7
hypothetical protein	364		fig 525370.5.peg.2599 (24.66)	fig 1219013.3.peg.4925 (Q)	ATCC33707
hypothetical protein	146			fig 1219013.3.peg.4926 (Q)	C7
hypothetical protein	138			fig 1219013.3.peg.4927 (Q)	C7
hypothetical protein	220			fig 1219013.3.peg.4928 (Q)	C7
hypothetical protein	107		fig 525370.5.peg.1731 (30.09)	fig 1219013.3.peg.4929 (Q)	C7
hypothetical protein	143			fig 1219013.3.peg.4930 (Q)	C7
hypothetical protein	251		fig 525370.5.peg.2151 (69.96)	fig 1219013.3.peg.4931 (Q)	C7
hypothetical protein	123			fig 1219013.3.peg.4932 (Q)	C7
Phage minor tail protein	548		fig 525370.5.peg.2149 (34.43)	fig 1219013.3.peg.4933 (Q)	ATCC33707-C7
Phage minor tail protein # Gp27	304		fig 525370.5.peg.2148 (29.34)	fig 1219013.3.peg.4934 (Q)	ATCC33707-C7
hypothetical protein	188			fig 1219013.3.peg.4936 (Q)	C7
hypothetical protein	246			fig 1219013.3.peg.4937 (Q)	C7
hypothetical protein	112			fig 1219013.3.peg.4938 (Q)	C7
FIG00545832: hypothetical protein	109			fig 1219013.3.peg.4939 (Q)	C7
hypothetical protein	99			fig 1219013.3.peg.4940 (Q)	C7
hypothetical protein	137			fig 1219013.3.peg.4941 (Q)	C7
hypothetical protein	71			fig 1219013.3.peg.4942 (Q)	C7
hypothetical protein	128			fig 1219013.3.peg.4943 (Q)	C7
hypothetical protein	335			fig 1219013.3.peg.4944 (Q)	C7
hypothetical protein	200			fig 1219013.3.peg.4945 (Q)	C7
hypothetical protein	593			fig 1219013.3.peg.4946 (Q)	C7
hypothetical protein	534			fig 1219013.3.peg.4947 (Q)	C7
hypothetical protein	433			fig 1219013.3.peg.4948 (Q)	C7
hypothetical protein	166			fig 1219013.3.peg.4949 (Q)	C7
hypothetical protein	238			fig 1219013.3.peg.4950 (Q)	C7
hypothetical protein	100			fig 1219013.3.peg.4951 (Q)	C7
hypothetical protein	225		fig 525370.5.peg.2126 (38)	fig 1219013.3.peg.4952 (Q)	ATCC33707-C7
hypothetical protein	111			fig 1219013.3.peg.4953 (Q)	C7
hypothetical protein	71			fig 1219013.3.peg.4954 (Q)	C7
hypothetical protein	124			fig 1219013.3.peg.4955 (Q)	C7
hypothetical protein	205			fig 1219013.3.peg.4957 (Q)	C7
hypothetical protein	43			fig 1219013.3.peg.4958 (Q)	C7
hypothetical protein	116			fig 1219013.3.peg.4959 (Q)	C7
hypothetical protein	107			fig 1219013.3.peg.4960 (Q)	C7
hypothetical protein	63			fig 1219013.3.peg.4961 (Q)	C7
hypothetical protein	121			fig 1219013.3.peg.4962 (Q)	C7
hypothetical protein	41			fig 1219013.3.peg.4963 (Q)	C7
hypothetical protein	117			fig 1219013.3.peg.4964 (Q)	C7
hypothetical protein	90			fig 1219013.3.peg.4965 (Q)	C7
hypothetical protein	165			fig 1219013.3.peg.4966 (Q)	C7
hypothetical protein	122			fig 1219013.3.peg.4967 (Q)	C7
hypothetical protein	92			fig 1219013.3.peg.4968 (Q)	C7
hypothetical protein	99			fig 1219013.3.peg.4969 (Q)	C7
hypothetical protein	281		fig 525370.5.peg.2116 (29.37)	fig 1219013.3.peg.4970 (Q)	ATCC33707-C7
Phage-related protein, predicted endonuclease	312		fig 525370.5.peg.2115 (29.96)	fig 1219013.3.peg.4971 (Q)	ATCC33707-C7
hypothetical protein	91			fig 1219013.3.peg.4972 (Q)	C7
hypothetical protein	68			fig 1219013.3.peg.4973 (Q)	C7
hypothetical protein	98			fig 1219013.3.peg.4974 (Q)	C7
hypothetical protein	102			fig 1219013.3.peg.4975 (Q)	C7
Integrase	352		fig 525370.5.peg.1714 (24.2)	fig 1219013.3.peg.4976 (Q)	ATCC33707

Note: 1. Genes used as the query for blast searches are marked with (Q) whereas the blast score is mentioned with the target features.

2. Two genes of strain 103S, one of ATCC33707 and four C7 genes showed similarities with more than one features in another genome. These genes and their target genes are labelled in green, blue and red, respectively.

3. The last column HGT represents "Horizontal Gene Transfer" and we mentioned the strain IDs where the genes are borne by the genomic islands that are predicted by the program Alien Hunter.

Supplementary Table 2. A list of nonsense SNPs in *R. equi* strains C7 and ATCC33707 using 103S as the reference

103S				C7		ATCC 33707	
Position	Ref. allele	Locus/gene	Product	SNP allele	SNP type	SNP allele	SNP type
717240	C	fig 685727.5.peg.690	CONSERVED 13E12 REPEAT FAMILY PROTEIN			a	nonsense
777219	C	fig 685727.5.peg.749	MCE-family protein Mce1F	a	nonsense		
777362	C	fig 685727.5.peg.749	MCE-family protein Mce1F	t	nonsense		
778256	G	fig 685727.5.peg.751	MCE associated membrane protein	a	nonsense		
832375	C	fig 685727.5.peg.791	putative ABC transporter ATP binding protein	t	nonsense	t	nonsense
953696	C	fig 685727.5.peg.895	3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)			t	nonsense
1018281	A	fig 685727.5.peg.965	hypothetical protein	t	nonsense		
1046258	C	fig 685727.5.peg.994	putative esterase	t	nonsense	t	nonsense
1060941	C	fig 685727.5.peg.1008	hypothetical protein			a	nonsense
1086868	G	fig 685727.5.peg.1037	hypothetical protein			a	nonsense
1420888	G	fig 685727.5.peg.1369	YndJ	a	nonsense		
1441851	G	fig 685727.5.peg.1390	hypothetical protein			a	nonsense
1474013	G	fig 685727.5.peg.1422	Pyruvate oxidase (EC 1.2.3.3)	a	nonsense		
1719095	C	fig 685727.5.peg.1649	hypothetical protein	t	nonsense		
1766963	A	fig 685727.5.peg.1671	Ribosyl nicotinamide transporter, PnuC-like	g	nonsense	c	nonsynonymous
1806315	C	fig 685727.5.peg.1711	secreted peptidase			a	nonsense
1939596	G	fig 685727.5.peg.1836	hypothetical protein			t	nonsense
2159015	G	fig 685727.5.peg.2048	oxetanocin A resistance protein			a	nonsense
2593202	C	fig 685727.5.peg.2435	hypothetical protein	t	nonsense		
2646430	C	fig 685727.5.peg.2481	Polyhydroxyalkanoic acid synthase	t	nonsense		
2754494	C	fig 685727.5.peg.2584	Epoxide hydrolase (EC 3.3.2.9)	t	nonsense		
2755296	A	fig 685727.5.peg.2585	FIG01136303: hypothetical protein	g	nonsense	g	nonsense
2777171	C	fig 685727.5.peg.2615	hypothetical protein	t	nonsense		
3092582	C	fig 685727.5.peg.2906	hypothetical protein			a	nonsense
3296283	C	fig 685727.5.peg.3095	hypothetical protein			t	nonsense
3395747	G	fig 685727.5.peg.3187	hypothetical protein			a	nonsense
3735874	G	fig 685727.5.peg.3515	hypothetical protein	a	nonsense	a	nonsense
4133706	C	fig 685727.5.peg.3890	hypothetical protein	t	nonsense	t	nonsense
4160134	C	fig 685727.5.peg.3917	Acetate kinase (EC 2.7.2.1)	t	nonsense	t	nonsense
4300318	G	fig 685727.5.peg.4054	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	a	nonsense		
4350321	C	fig 685727.5.peg.4088	FIG00863843: hypothetical protein	t	nonsense	t	nonsense
4379112	C	fig 685727.5.peg.4112	hypothetical protein			a	nonsense
4959414	C	fig 685727.5.peg.4661	hypothetical protein	g	nonsense		