Original citation:

Permanent WRAP url:
http://wrap.warwick.ac.uk/55954

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license and may be reused according to the conditions of the license. For more details see: http://creativecommons.org/licenses/by/3.0/

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk
Ion Torrent-based transcriptional assessment of a Corynebacterium pseudotuberculosis equi strain reveals denaturing high-performance liquid chromatography a promising rRNA depletion method

Thiago L. P. Castro,1 Nubia Seyffert,1 Rommel T. J. Ramos,2 Silvanira Barbosa,2 Rodrigo D. O. Carvalho,1 Anne Cybelle Pinto,1 Adriana Ribeiro Carneiro,2 Wanderson Marques Silva,1 Luis G. C. Pacheco,3 Christopher Downson,4 Maria P. C. Schneider,2 Anderson Miyoshi,1 Vasco Azevedo1 and Artur Silva2*

1Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.
2Genome and Proteome Network of the State of Pará, Universidade Federal do Pará, Brazil.
3Institute of Health Sciences, Universidade Federal da Bahia, Salvador, Brazil.
4School of Life Sciences, University of Warwick, Coventry, UK.

Summary

Corynebacterium pseudotuberculosis equi is a Gram-positive pathogenic bacterium which affects a variety of hosts. Besides the great economic losses it causes to horse-breeders, this organism is also known to be an important infectious agent to cattle and buffaloes. As an outcome of the efforts in characterizing the molecular basis of its virulence, several complete genome sequences were made available in recent years, enabling the large-scale assessment of genes throughout distinct isolates. Meanwhile, the RNA-seq stood out as the technology of choice for comprehensive transcriptome studies, which may bring valuable information regarding active genomic regions, despite of the still impeditive associated costs. In an attempt to increase the use of generated reads per instrument run, by effectively eliminating unwanted rRNAs from total RNA samples without relying on any commercially available kits, we applied denaturing high-performance liquid chromatography (DHPLC) as an alternative method to assess the transcriptional profile of C. pseudotuberculosis. We have found that the DHPLC depletion method, allied to Ion Torrent sequencing, allows mapping of transcripts in a comprehensive way and identifying novel transcripts when a de novo approach is used. These data encourage us to use DHPLC in future transcriptional evaluations in C. pseudotuberculosis.

Introduction

Corynebacterium pseudotuberculosis is a Gram-positive and facultative intracellular bacterium that does not sporulate, has fimbriae, and displays pleomorphic cells which vary from coccoid to filamentary rods, measuring up to 0.6 µm per 3 µm in size (Jones and Collins, 1986; Baird and Fontaine, 2007). Two biovars have been described for C. pseudotuberculosis, ovis and equi; they are mainly distinguished by their ability to reduce nitrate and tropism for infecting particular kinds of animals (Biberstein et al., 1971).

Caprine, ovine, equine and bovine are among the most prominent hosts, to whom the infection by C. pseudotuberculosis can cause distinct clinical signals. The nitrate-negative biovar ovis, the most studied of C. pseudotuberculosis, causes caseous lymphadenitis in sheep and goats, a chronic disease characterized by the abscessation of external and internal lymphnodes. On the other hand, the biovar equi commonly tests positive for nitrate reduction, and is associated to distinct diseases in a variety of hosts, including ulcerative lymphangitis in horses and contagious acne in cattle (Paton et al., 1994; Braverman et al., 1999; Spier et al., 2004). Corynebacterium pseudotuberculosis biovar equi is also the causative agent of oedematous skin disease in buffaloes, characterized by superficial swellings which culminate in the formation of abscesses and release of serous bloody exudates (Barakat et al., 1984).

Recent efforts in elucidating the molecular features associated to host tropism and virulence have led to the acquisition of whole-genome sequences of six different strains of C. pseudotuberculosis equi, all of them made public in the National Center for Biotechnology Information databases (NCBI). However, additional studies...
The first RNA-seq work involving non-coding RNAs (ncRNAs), what allows a more precise annotation of the transcribed, as well as estimating their transcriptional levels. By performing RNA-seq of an organism, it is also possible to identify operon structures and set new coding sequences (CDS), including untranslated regulatory elements (ncRNAs), what allows a more precise annotation of features present in the genome (Sorek and Cossart, 2010). The first RNA-seq work involving *C. pseudotuberculosis* was published recently by Pinto and colleagues (2012). The authors describe a set of genes from the 1002 strain which are co-regulated in response to adverse environmental conditions, such as high temperature and elevated osmolarity in the growth medium.

Despite being of great value for structural genomics, RNA-seq experiments are limited by the large quantity of reads which correspond to ribosomal RNA sequences. This is explained by the fact that more than 80% of the overall synthesized RNA in bacteria and eukaryotes is ribosomal (Westermann et al., 2012). Therefore, it is usual to submit total RNA to rRNA subtraction prior to the cDNA library construction, using commercially available kits based on rRNA exonuclease digestion or hybridization with specific probes. In the specific case of eukaryotic transcriptomes, it is also possible to enrich mRNA molecules with poly(A) selection (Martin and Wang, 2011).

An alternative methodology for rRNA depletion

DHPLC depletion of rRNA and qRT-PCR validation

All extracted total RNA samples from *C. pseudotuberculosis* 31 showed similar chromatogram profiles after DHPLC analysis, such as the one illustrated in Fig. 1A. This is an important result for us, once in order to standardize the DHPLC rRNA depletion methodology we had to set a reliable interval representing our fraction of interest. Taking into account the WAVE® DHPLC System manufacturer’s recommendations (Transgenomic) and the separation profiles generated for *C. pseudotuberculosis*, the retention period from 10 up to 14 min constituted the fraction to be collected for further evaluations of depletion in this work. The collected material was reanalysed using DHPLC, and a prominent peak likely to be formed by mRNA molecules came out in the chromatographic profile (Fig. 1B).

Quantitative reverse transcription PCR assays (qRT-PCR) showed partial removal of 16S rRNA after DHPLC fractioning, as occurred with samples treated in a parallel way using the commercial depletion kit RiboMinus (Chen and Duan, 2011) (Fig. 2). The rates of depletion with DHPLC ranged from 78.2% up to 92.2% in relation to the initial rRNA levels observed in non-treated total RNA samples. For the RiboMinus treatment, the rates ranged from 58.8% up to 88%, with a greater experimental variation in depletion efficiency when compared with what was obtained with DHPLC but similar to what was obtained with the combination of DHPLC and DHPLC depletion.
seen for DHPLC. This leads us to suggest that the DHPLC fractionation method is probably less sensitive to minor inter-sample variations than the hybridization subtraction-based RiboMinus kit. However, it was not possible to state DHPLC as the most efficient depletion method, since it significantly demonstrated better performance in only two out of four independent experiments. In this context, it is prudent to consider that for our model organism, DHPLC provides at least the same depletion levels of a well-established and broadly trusted commercially available kit.

Fig. 1. Chromatograms of a RNA sample before and after DHPLC fractionation for rRNA removal. A. Total RNA sample chromatographic profile. The two highest spikes correspond to ribosomal RNAs, and the red dashed lines delimitate the interval of run corresponding to the collected sample (mRNA) for downstream applications. B. Chromatographic profile of sample submitted to depletion with DHPLC. The green dashed lines virtually represent the previously selected area for sample collection. The proportion between the selected peak and the rRNA-related spikes appears to be more balanced in this case.

Fig. 2. Percentages of 16S rRNA depletion for each of four independently prepared biological replicates, in relation to the sigA gene mRNA levels. DHPLC and RiboMinus treatments were performed in a parallel way for each replicate. The asterisks indicate the cases where the levels of depletion observed for DHPLC were significantly different from those seen for RiboMinus, according to the software REST 2009.
Reference-based transcriptional assessment

The transcriptome of C. pseudotuberculosis 31 was assessed based on the RiboMinus- and DHPLC-depleted RNA samples, both derived from the same extraction but independently converted into cDNA for Ion Torrent sequencing. For RiboMinus, we estimated 58.75% of all sequences with Phred quality score greater or equal to 15 as being related to rRNA-coding genes (198 366 out of 337 645 reads), while for DHPLC the proportion of rRNA-related sequences decreased to 45.08% of total with Phred 15 score (183 963 out of 408 082 reads). The reads which have mapped against the CDS from the reference genome but are not ribosomal totalize 53 679 and 74 029 for RiboMinus and DHPLC respectively. The average length of total mapped reads was 143.35 nucleotide bases for RiboMinus and 144.59 bases for DHPLC.

For both transcriptomes, we investigated the representativeness of biological processes and molecular functions attributed by Gene Ontology (GO) (Fig. 3). By comparing the samples, we noticed some similarity in the distribution of transcribed genes across the different classes considered, and this was expected since the same biological condition was addressed. However, the majority of classes presented a higher number of different transcripts in the case of DHPLC, what can be partially attributed to the larger number of non-ribosomal mapped reads acquired for the sample treated by this method. The process represented by the greatest number of transcripts was oxidation-reduction, whereas only a few transcribed elements were related to pathogenesis or cell adhesion, what is supported by the fact that the bacteria had been cultured without interacting with any hosts.

Taking into account the functional classification, while transcripts for the class of structural constituents of ribosomes were the most abundant ones in the RiboMinus-treated sample, there was a greater representativeness for the class of nucleic acid binding in the case of DHPLC depletion, including genes directly involved in transcriptional regulation.

Many of the identified transcripts were shared by both transcriptomes; among them we highlight the antigen 84, which is reported to be essential for Mycobacterium smegmatis polarization during cellular division, besides acting in the modulation of cell shape in pleiomorphic actinobacteria (Nguyen et al., 2007). Other interesting detected transcript codifies for a protein similar to the diphteria toxin (DT). The DT is described as the main virulence factor produced by Corynebacterium diphtheriae, and other proteins which are similar to DT in molecular weight and immunological structure have been detected in other pathogenic corynebacteria (Wong and Groman, 1984; Efstratiou et al., 1998). We can also mention the transcript for superoxide dismutase, known to be important for survival of M. tuberculosis inside activated macrophage cells (Piddington et al., 2001). Finally, we cite the transcripts for heat shock proteins (HSPs), which are involved in thermal stress adaptation and are also capable of inducing both humoral and cellular immune responses in the host (Pinho et al., 2009).

Among the transcripts uniquely identified following DHPLC fractionation we highlight those coding for a thioredoxin-related protein, engaged in oxidation-reduction processes; an urease accessory protein, which has been correlated to pathogenicity in corynebacteria; and the cell adhesion involved manganese ABC transporter substrate-binding protein and periplasmic zinc-binding protein TroA (Nolden et al., 2000; Hantke, 2005; Papp-Wallace and Maguire, 2006; Colbert et al., 2006). As uniquely identified in the RiboMinus-treated sample, we emphasize the transcripts linked to the transcriptional regulator sigma factor K and the detoxification-related mycothione glutathione reductase (Patel and Blanchard, 1999).

Analysis of the genome of C. pseudotuberculosis 31 using the Rfam software allowed us to predict eleven non-coding functional RNAs (ncRNAs), other than rRNAs and IRNAs, which belong to three distinct known families: riboswitches, ribozymes and small RNAs (Table 1). Many of these ncRNAs are codified by bacteria of the genus Streptomyces and Clostridium (Weinberg et al., 2010; Chen et al., 2011), but their functions are not yet well understood. Among the predicted features, eight ncRNAs were confirmed to be transcribed, of which the tmRNA appeared as the most represented one in both RiboMinus- and DHPLC-related transcriptomes. The Bacillus subtilis tmRNA is required for growth at high temperature condition, beyond under cadmium and ethanol in vitro exposure (Muto et al., 2000). Other ncRNA relatively known and whose transcript was detected only in the RiboMinus-treated sample is the 6C.1, which plays a role in dormancy in Streptomyces coelicolor (Swiercz et al., 2008). In conclusion, although it may seem that the RiboMinus performed slightly better than DHPLC in maintaining ncRNA molecules in the pool of RNAs subjected to sequencing, further evaluations are necessary to better compare the efficiency of the two depletion methods on this issue. We also speculate that alternative fractions following the DHPLC treatment can be tested on their content of ncRNAs or any other specific group of RNAs.

De novo strategy for new transcripts screening

One key possibility arising from transcriptome assessments is the identification of non-predicted CDS in the genome. Aiming to evaluate how our RiboMinus- and DHPLC-depleted RNA-sequencing data perform to allow
the detection of new genes in *C. pseudotuberculosis* 31, we assembled all reads that had not mapped against the CDS from the reference genome into contigs (Table 2). Despite observing a higher number of non-mapped reads for the DHPLC sample, and a higher total of assembled contigs in this case as well, the removal of redundant sequences evidenced a larger amount of different contigs for RiboMinus (3836 against 1247 for DHPLC) (Table 3). One explanation for such inversion resides in the fact that larger contigs could be generated for DHPLC, what is in accordance with its greater N50 value (455 against 399 for RiboMinus).

Following redundancy removal, we found that 487 DHPLC assembled contigs and 3058 RiboMinus assembled ones presented no matches against CDS of *C. pseudotuberculosis* 31 (Table 3). These considerable amounts of dissimilar sequences could be explained by intrinsic sequencing errors, which have caused penalties to matches under the criteria adopted by us (alignments should be larger than or equal to 30 bp and present

**Fig. 3.** Classification of reads mapped against the CDS of *C. pseudotuberculosis* 31, according to terms of Gene Ontology. A. The identified transcripts were grouped by the biological processes to which their products are associated with. The horizontal bars indicate the amounts of grouped transcripts (genes) for each category, considering transcriptome results for both RiboMinus- and DHPLC-treated samples. B. In a similar way as before, the identified transcripts were sorted by their association with different classes of molecular functions.
Furthermore, our reference genome underwent automatic annotation process based on previously annotated genomes from other *C. pseudotuberculosis* strains, so it is plausible to consider that a number of CDS may have not been annotated in strain 31.

To analyse the similarity of all unknown contigs to the non-redundant nucleotide collection from the NCBI, we decided to use more strict criteria: only matches of at least 100 bp and with e-value lower than or equal to 1e-05 were considered. In spite of sequencing errors, which may have hampered our analysis, we could identify 15 different genes that were not previously annotated in the *C. pseudotuberculosis* 31 genome, six out of them related to hypothetical proteins in other organisms and one of unknown function (Table 4). Among the newly identified features are various genes coding for surface-anchored proteins, which participate in the bacterial pilus assembly and belong in their majority to the Spa family. Predicted to be present in the *C. pseudotuberculosis* ovis FRC41 strain, these proteins were proved to mediate, in *C. diphteriae* and other pathogens, bacterial attachment and colonization of host tissues (Gaspar and Ton-That, 2006; Trost et al., 2010). The remaining new transcripts, exclusively identified for the DHPLC-depleted sample, codify for anaerobic dehydrogenase and subunits alpha and beta of nitrate reductase, enzyme which plays an important role in the nitrogen assimilation (Richardson et al., 2001).

With respect to these findings, we highlight that the DHPLC depletion method associated with Ion Torrent sequencing allowed us to successfully discover non-annotated genes in the genome of *C. pseudotuberculosis* 31, as occurred when RiboMinus was used. The next step to be taken is to thoroughly analyse and correct all annotation errors, collaborating with upcoming studies regarding this organism.

In case the differences observed between the RiboMinus- and DHPLC-treated samples are caused by peculiarities of each depletion procedure, additional transcriptome evaluations are necessary for indicating to what extent these could alter the set of transcripts to be sequenced. Recent improvements on Ion Torrent technology will certainly benefit such work, by providing larger

<table>
<thead>
<tr>
<th>ncRNAs predicted in the genome</th>
<th>Corresponding strand</th>
<th>Rfam bits score</th>
<th>Genomic stance (bp)</th>
<th>RiboMinus-treated sample</th>
<th>DHPLC-treated sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>tmRNA.1</td>
<td>+</td>
<td>136.29</td>
<td>605 717–606 098</td>
<td>1497</td>
<td>444</td>
</tr>
<tr>
<td>RNaseP_bact_a.1</td>
<td>–</td>
<td>250.04</td>
<td>1 601 151–1 601 563</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>cspA.1</td>
<td>+</td>
<td>55.15</td>
<td>231 046–231 384</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>msK.1</td>
<td>–</td>
<td>54.15</td>
<td>398 399–398 458</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bacteria_small SRP.1</td>
<td>+</td>
<td>55.35</td>
<td>149 269–149 361</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>mraW.1</td>
<td>–</td>
<td>48.99</td>
<td>1 528 518–1 528 622</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6C.1</td>
<td>+</td>
<td>32.18</td>
<td>223 511–223 577</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cobalamin.2</td>
<td>–</td>
<td>72.55</td>
<td>724 215–724 451</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Cobalamin.1</td>
<td>+</td>
<td>74.77</td>
<td>924 216–924 412</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TPP.1</td>
<td>+</td>
<td>57.38</td>
<td>793 275–793 405</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>yybP-ykoY.1</td>
<td>–</td>
<td>40.90</td>
<td>1 362 468–1 362 614</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

a. In reference to the whole-genome sequence of *C. pseudotuberculosis* 31 (NC_017730.1).

e-value lower than or equal to 1e-05). Furthermore, our reference genome underwent automatic annotation process based on previously annotated genomes from other *C. pseudotuberculosis* strains, so it is plausible to consider that a number of CDS may have not been annotated in strain 31.

To analyse the similarity of all unknown contigs to the non-redundant nucleotide collection from the NCBI, we decided to use more strict criteria: only matches of at least 100 bp and with e-value lower than or equal to 1e-05 were considered. In spite of sequencing errors, which may have hampered our analysis, we could identify 15 different genes that were not previously annotated in the *C. pseudotuberculosis* 31 genome, six out of them related to hypothetical proteins in other organisms and one of unknown function (Table 4). Among the newly identified features are various genes coding for surface-anchored proteins, which participate in the bacterial pilus assembly and belong in their majority to the Spa family. Predicted to be present in the *C. pseudotuberculosis* ovis FRC41 strain, these proteins were proved to mediate, in *C. diphteriae* and other pathogens, bacterial attachment and colonization of host tissues (Gaspar and Ton-That, 2006; Trost et al., 2010). The remaining new transcripts, exclusively identified for the DHPLC-depleted sample, codify for anaerobic dehydrogenase and subunits alpha and beta of nitrate reductase, enzyme which plays an important role in the nitrogen assimilation (Richardson et al., 2001).

With respect to these findings, we highlight that the DHPLC depletion method associated with Ion Torrent sequencing allowed us to successfully discover non-annotated genes in the genome of *C. pseudotuberculosis* 31, as occurred when RiboMinus was used. The next step to be taken is to thoroughly analyse and correct all annotation errors, collaborating with upcoming studies regarding this organism.

In case the differences observed between the RiboMinus- and DHPLC-treated samples are caused by peculiarities of each depletion procedure, additional transcriptome evaluations are necessary for indicating to what extent these could alter the set of transcripts to be sequenced. Recent improvements on Ion Torrent technology will certainly benefit such work, by providing larger

<table>
<thead>
<tr>
<th>rRNA depletion method</th>
<th>Non-mapped reads useda</th>
<th>Assembly round</th>
<th>Total of contigs assembled</th>
<th>N50 of the contigs</th>
<th>Total base pairs assembled</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiboMinus</td>
<td>85 600</td>
<td>MergeK25</td>
<td>674</td>
<td>282</td>
<td>172 401</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK27</td>
<td>1 225</td>
<td>283</td>
<td>318 679</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK29</td>
<td>2 358</td>
<td>288</td>
<td>623 789</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK31</td>
<td>4 595</td>
<td>291</td>
<td>1 229 563</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK33</td>
<td>9 032</td>
<td>296</td>
<td>2 454 627</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK35</td>
<td>17 822</td>
<td>310</td>
<td>5 195 511</td>
</tr>
<tr>
<td>DHPLC</td>
<td>150 090</td>
<td>MergeK25</td>
<td>1 681</td>
<td>429</td>
<td>564 199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK27</td>
<td>2 740</td>
<td>441</td>
<td>949 402</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK29</td>
<td>4 799</td>
<td>445</td>
<td>1 706 112</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK31</td>
<td>8 834</td>
<td>447</td>
<td>3 200 016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK33</td>
<td>16 948</td>
<td>450</td>
<td>6 210 773</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MergeK35</td>
<td>32 659</td>
<td>456</td>
<td>12 134 747</td>
</tr>
</tbody>
</table>

a. Reads which did not match against CDS from *C. pseudotuberculosis* 31 (query coverage < 60% and similarity < 70%).

© 2013 The Authors. *Microbial Biotechnology* published by Blackwell Publishing Ltd and Society for Applied Microbiology, *Microbial Biotechnology, 6*, 168–177
reads in a more efficient high-throughput way, and therefore minimizing the adverse effects caused by errors in sequences. Still, we speculate that the DHPLC rRNA depletion method could also be applied to the RNA-seq analysis intended for differential expression purposes, representing a cost-effective alternative for those groups which have access to the required equipment, since the expenses with depletion kits surely overcome those associated with the reagents for chromatography. On the other hand, in cases where only restricted amounts of total RNA are available, the DHPLC method is not likely to provide sufficient material for downstream applications. In conclusion, this work outlines a new rRNA depletion strategy which will be strongly considered to be applied in our future works about the transcriptional activity of C. pseudotuberculosis.

Experimental procedures

Bacterial strain, growth condition and preparation of bacterial pellet

A Corynebacterium pseudotuberculosis equi strain, identified as 31, was previously isolated from an infected buffalo in Egypt and confirmed by biochemical and molecular methods. This strain was regularly cultivated in Brain Heart Infusion broth (BHI, HiMedia Laboratories) and stored at −80°C with 50% glycerol. For the transcriptional assessments, Tween 80 (Sigma-Aldrich) was added to the BHI medium at 0.05% to avoid clumping of cells in suspension. The inoculated bacteria were grown at 37°C, 140 r.p.m., until the exponential phase was reached (OD600nm = 0.4). For each culture, 20 ml were centrifuged at 5000 r.p.m. for 5 min, and then 5 ml of RNAlater® stabilizing reagent (Ambion) were added prior to storing at −80°C.

RNA extraction from C. pseudotuberculosis C31

Stabilized bacteria were thawed on ice and then pelleted at 5000 r.p.m. for 5 min. Then they were suspended in RLT buffer (Qiagen) and submitted to mechanical disruption, using glass beads with 0.1 mm of diameter (VK01, Berthold Technologies). Two homogenization cycles were performed at 6500 r.p.m., 30 s each, with interval of 30 s between them, in a Precellys®24 equipment (Bertin Technologies). The total RNA was purified using RNase-free Water, yielding at least 140 ng according to the Qubit® Fluorometric Quantitation kit (Invitrogen). Additionally, the RNA integrity was checked by electrophoresis in 1% agarose gels, in an RNase-free environment.

Table 3. Statistics for the de novo approach after removal of redundant contigs using the Simplifier software.

<table>
<thead>
<tr>
<th>Non-redundant contigs (transcripts)</th>
<th>rRNA depletion method</th>
<th>DHPLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>3 836</td>
<td>1 247</td>
</tr>
<tr>
<td>Total base pairs</td>
<td>1 392 183</td>
<td>449 947</td>
</tr>
<tr>
<td>N50</td>
<td>399</td>
<td>455</td>
</tr>
<tr>
<td>Smaller</td>
<td>1712</td>
<td>2 237</td>
</tr>
<tr>
<td>With hits*</td>
<td>778 (407 620 bp)</td>
<td>760 (295 852 bp)</td>
</tr>
<tr>
<td>Without hits*</td>
<td>3 058 (984 563 bp)</td>
<td>487 (154 095 bp)</td>
</tr>
</tbody>
</table>

* Alignments performed against CDS of C. pseudotuberculosis 31. Were considered negligible alignments < 30 bp with e-value > 1e-0.5.

Table 4. New gene products found among the C. pseudotuberculosis 31 assembled transcripts.

<table>
<thead>
<tr>
<th>Gene product</th>
<th>Number of matching contigs*</th>
<th>Average size of contigs (bp)</th>
<th>Matching strain/product</th>
<th>Depletion method used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothetical protein</td>
<td>20</td>
<td>1510</td>
<td>Propionibacterium acnes HL053PA2 (HMPREF9565_01483 protein)</td>
<td>DHPLC/RiboMinus</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>31</td>
<td>1235</td>
<td>Propionibacterium acnes HL074PA1 (HMPREF9574_00607 protein)</td>
<td>RiboMinus</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>2</td>
<td>782</td>
<td>Corynebacterium glutamicum R (cgr 1268 protein) (BACCAP_003932 protein)</td>
<td>DHPLC/RiboMinus</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>1</td>
<td>366</td>
<td>Propionibacterium acnes HL074PA1 (HMPREF9574_00607 protein)</td>
<td>DHPLC/RiboMinus</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>2</td>
<td>1040</td>
<td>Collinsella aerofaciens ATCC 25986 (COLAER_01671 protein)</td>
<td>RiboMinus</td>
</tr>
<tr>
<td>Conserved hypothetical protein</td>
<td>1</td>
<td>572</td>
<td>Lactococcus jensenii JV-V16</td>
<td>RiboMinus</td>
</tr>
<tr>
<td>Unknown protein</td>
<td>1</td>
<td>344</td>
<td>Streptococcus suis 98AHAS3</td>
<td>RiboMinus</td>
</tr>
<tr>
<td>spaF gene product</td>
<td>9</td>
<td>1732</td>
<td>Propionibacterium pseudotuberculosis FRC41</td>
<td>DHPLC</td>
</tr>
<tr>
<td>spaE gene product</td>
<td>6</td>
<td>880</td>
<td>Propionibacterium pseudotuberculosis FRC41</td>
<td>DHPLC</td>
</tr>
<tr>
<td>spaY gene product</td>
<td>3</td>
<td>642</td>
<td>Propionibacterium pseudotuberculosis FRC41</td>
<td>DHPLC</td>
</tr>
<tr>
<td>spaC gene product</td>
<td>1</td>
<td>958</td>
<td>Corynebacterium ulcerans 809</td>
<td>DHPLC</td>
</tr>
<tr>
<td>Surface-anchored fimbrial protein</td>
<td>8</td>
<td>958</td>
<td>Corynebacterium ulcerans BR-AD22</td>
<td>DHPLC</td>
</tr>
<tr>
<td>Anaerobic dehydrogenase</td>
<td>1</td>
<td>332</td>
<td>Corynebacterium glutamicum ATCC 13032</td>
<td>DHPLC</td>
</tr>
<tr>
<td>Nitrate reductase beta subunit</td>
<td>2</td>
<td>554</td>
<td>Propionibacterium acnes HL074PA1</td>
<td>DHPLC</td>
</tr>
<tr>
<td>Nitrate reductase alpha subunit</td>
<td>4</td>
<td>651</td>
<td>Corynebacterium diphtheriae HC04</td>
<td>DHPLC</td>
</tr>
</tbody>
</table>

* Only matches ≥ 100 bp with e-value ≤ 1e-05 were considered.
mental replicates, the generated cycle-threshold values were 60°C for 15 s, and 95°C for 15 s. For each of the four experi-
dissociation curves were performed as follows: 95°C for 15 s, PCR System supplied with the software SDS, version 2.4.
primer (2 pmol), 2 (50 reagent (Agilent Technologies), 0.375
the Qubit® Fluorometric Quantitation kit (Invitrogen). All
rRNA depletion
For the rRNA depletion through DHPLC, four subsamples of 20 µg of total RNA each, provenient from the same extrac-
tion, were separately injected into the WAVE® System 4500
the buffers A [0.1 M triethylammonium acetate (TEAA) at
acetonitrile] were used to elute samples at the flow rate of
pH 7.0] and B [0.1 M TEAA (at pH 7.0) containing 25%
DNA pellets were dried and resuspended with RNase-free water.
The chromatograms were analysed using the Wave®
Small-animal metabolism, rRNA-coding region in the genome of
oligonucleotide primers. The first pair anneals to the 16S
were removed using the Dynabeads® MyOne™ Streptavidin
Kit, Life Technologies), and the spheres without template
was broken with successive washes (Ion Template Solutions
preparation was carried out at fully denaturing condition (75°C).
The affinity script OPCR cDNA Synthesis Kit (Agilent Tech-
negom®). This procedure was necessary to achieve, at the
end of the process, depleted RNA in sufficient amount for
the tested conditions. The reactions for relative quantifica-
tion, the mean and standard error values calculated for
the qRT-PCR assays were performed in accordance with
Bustin and colleagues (2009). Following the relative quanti-
fication, the 16S rRNA were converted into percentages, which were
considered for all comparisons between the depletion
GraphPad Prism, version 5.0 (GraphPad Software, San
RNA-seqencing
For RNA-seqencing, we started from a single total RNA
sample, which was subdivided and passed through two inde-
pendent rRNA depletion procedures: one with the RiboMinus
kit and the other based on the DHPLC method. The cDNA
libraries were prepared using Ambion® RNA-Seq Library Con-
struction Kit. Briefly, 200 ng of fresh depleted RNA were frag-
mented with RNase III. After clean up, specific adapters (Ion
adapter mix, Life Technologies) were hybridized with 100 ng
each sample. The reverse transcription was carried out
following the manufacturer’s recommendations, and the
resulting cDNA samples were purified. Only molecules with
more than 100 bp were selected using the Agencourt
AMPure® XP reagent (Beckman). After this stage, the cDNA
was amplified and then purified with PureLink® PCR Micro Kit
(Invitrogen). Quantification of amplified DNA samples was
performed with the Qubit® Fluorometric Quantitation kit (Invitrogen).
Subsequently, the IKA® Ultra-Turrax® Tube Drive system
was applied for an emulsion PCR, performed with reagents
from the Ion Template Reagents Kit and the Ion Template
Preparation Kit (Life Technologies). In brief, the DNA frag-
ments were bonded onto spheres (Ion Sphere™ particles)
and then amplified. Following the amplification, the emulsion
was broken with successive washes (Ion Template Solutions
Kit, Life Technologies), and the spheres without template
were removed using the Dynabeads® MyOne™ Streptavidin
C1 Magnetic Beads (Invitrogen). The spheres-bonded ampli-
fied fragments were mixed with the Ion Torrent sequencing
primers and the sequencing polymerase enzyme, using the
Ion Sequencing Reagents Kit (Life Technologies). The result-
ing mixtures were loaded on the Ion 314™ Chip for sequenc-
ing with the Ion Torrent Personal Genome Machine™ (PGM)
(Life Technologies). All procedures adopted are in agreement
with what is recommended by the handbooks of Ion
sequencing Reagent Kit and Ion Xpress™ Template Kit (Life
Technologies).
Bioinformatic analysis
The Ion Torrent-generated reads were trimmed and filtered
according to in-house bioinformatic scripts. Sequences pre-
senting an average Phred quality score inferior to 15 were
eliminated by the Quality Assessment Software (Ramos
et al., 2011). The reference-based assessment relied on the
gene C. pseudotuberculosis, a 92 µl amplicon, and the second is complement-
tary to the sigma factor A locus (sigA gene). Both pairs provided amplification efficiencies greater than 0.95 under
the tested conditions. The reactions for relative quantification
were prepared as follows: 7.5 µl Brilliant SYBR® Green qPCR
reagent (Agilent Technologies), 0.375 µl ROX reference dye
(50× diluted), 1 µl forward primer (2 pmol), 1 µl reverse
primer (2 pmol), 2 µl template cDNA (10× diluted), and ultra-
pure water to final 15 µl volume. All amplifications were
performed using the Applied Biosystems 7900HT Real-Time
PCR System supplied with the software SDS, version 2.4.
Dissociation curves were performed as follows: 95°C for 15 s,
60°C for 15 s, and 95°C for 15 s. For each of the four experi-
mental replicates, the generated cycle-threshold values were
loaded into the REST 2009 software (Qiagen), and then the
16S rRNA levels were assessed in relation to the correspond-
ing housekeeping sigA gene transcription levels, as
described by Pfaffl and colleagues (2002). In general lines,
the qRT-PCR assays were performed in accordance with
Bustin and colleagues (2009). Following the relative quanti-
fication, the mean and standard error values calculated for
the 16S rRNA were converted into percentages, which were
considered for all comparisons between the depletion
methods. The outcome data were plotted with the software
© 2013 The Authors. Microbial Biotechnology published by Blackwell Publishing Ltd and Society for Applied Microbiology, Microbial
Biotechnology, 6, 168–177
similarity with annotated CDS were considered for our analysis. To categorize the gene products into biological processes and molecular functions, the transcripts were processed with the Blast2GO software (Conesa et al., 2005). The outcome data were plotted using GraphPad Prism.

With respect to the ncRNAs identification, genomes RefSeq of the CMNR group members were obtained from the NCBI, and then a local database was built for predictions with the GeneMark software (Lukashin and Borodovsky, 1998). Subsequently, the RFam software (http://rfam.sanger.ac.uk/) was used for prediction of ncRNAs. The identification of transcribed ncRNAs was performed as described for the reference-based assessment.

For the de novo strategy, we used all non-mapped reads from the reference-based analysis on the Velvet assembler (Zerbino and Birney, 2008) combined with the Oases software (Schulz et al., 2012). In brief, initial assemblies with different k-mers (25, 27, 29, 31, 33 and 35) were generated with Velvet. Subsequently, the transcripts obtained in each assembly were added into a single file, which was then submitted to new assemblies with Oases (k-mers 25–35). All assembled contigs for each rRNA depletion method were combined into a single multifasta file, and then undesired redundancies were removed by the Simplifier software (Ramos et al., 2012). The resulting contigs were blasted against the CDS of C. pseudotuberculosis 31, and were considered negligible alignments of less than 30 bp and which presented e-value greater than 1e-05. The dissimilar transcripts were then selected and blasted against the NCBI non-redundant nucleotide collection, and were considered new genes the hits covering 100 bp or more of queries and presenting e-value lower than or equal to 1e-05.

Acknowledgements
We thank Salah Abdel Karim Selim and Mohamed Salah Moawad (University of Cairo) for providing the C. pseudotuberculosis 31 strain.

Conflict of interest
None declared.

References


