Investigating Evolutionary Dynamics of RHA1 Operons

Yong Chen
Rowan University

Dandan Geng

Kristina Ehrhardt

Shaoqiang Zhang

Follow this and additional works at: https://rdw.rowan.edu/csm_facpub

Part of the Genetics and Genomics Commons

Let us know how access to this document benefits you - share your thoughts on our feedback form.

Recommended Citation
Chen, Yong; Geng, Dandan; Ehrhardt, Kristina; and Zhang, Shaoqiang, "Investigating Evolutionary Dynamics of RHA1 Operons" (2016). Faculty Scholarship for the College of Science & Mathematics. 131. https://rdw.rowan.edu/csm_facpub/131

This Article is brought to you for free and open access by the College of Science & Mathematics at Rowan Digital Works. It has been accepted for inclusion in Faculty Scholarship for the College of Science & Mathematics by an authorized administrator of Rowan Digital Works. For more information, please contact jiras@rowan.edu, rdw@rowan.edu.
Investigating Evolutionary Dynamics of RHA1 Operons

Yong Chen1,2,*, Dandan Geng1,3, Kristina Ehhardt2,4 and Shaoqiang Zhang3,*

1National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. 2Department of Biological Sciences, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX, USA. 3College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China. 4Bioengineering Department, The University of Texas at Dallas, Richardson, TX, USA. *Corresponding Author.

ABSTRACT: Grouping genes as operons is an important genomic feature of prokaryotic organisms. The comprehensive understanding of the operon organizations would be helpful to decipher transcriptional mechanisms, cellular pathways, and the evolutionary landscape of prokaryotic genomes. Although thousands of prokaryotes have been sequenced, genome-wide investigation of the evolutionary dynamics (division and recombination) of operons among these genomes remains unexplored. Here, we systematically analyzed the operon dynamics of Rhodococcus jostii RHA1 (RHA1), an oleaginous bacterium with high potential applications in biofuel, by comparing 340 prokaryotic genomes that were carefully selected from different genera. Interestingly, 99% of RHA1 operons were observed to exhibit evolutionary events of division and recombination among the 340 compared genomes. An operon that encodes all enzymes related to histidine biosynthesis in RHA1 (His-operon) was found to be segmented into smaller gene groups (sub-operons) in diverse genomes. These sub-operons were further reorganized with different functional genes as novel operons that are related to different biochemical processes. Comparatively, the operons involved in the functional categories of lipid transport and metabolism are relatively conserved among the 340 compared genomes. At the pathway level, RHA1 operons found to be significantly conserved were involved in ribosome synthesis, oxidative phosphorylation, and fatty acid synthesis. These analyses provide evolutionary insights of operon organization and the dynamic associations of various biochemical pathways in different prokaryotes.

KEYWORDS: operon, Rhodococcus jostii RHA1, evolutionary dynamics, pathway

Background

In prokaryotic genomes, an operon is a functional unit of multiple neighboring genes under the control of a single promoter and terminator.1,2 Typically, about half of the protein-coding genes are organized into operons, representing one of the main strategies of gene organization, regulation, and transcription in prokaryotes.3–5 The functions and transcription of many operons have been studied extensively because of which extensive biological insight has been achieved. For example, the studies of two operons that are related to tryptophan6 and histidine7 syntheses have revealed new and sophisticated mechanisms of transcription control. Additionally, genes grouped in operons are widely found to have similar biological functions, indicating that clustering of genes involved in a biosynthetic route is a common feature of prokaryotic genomes.2–8,9 Furthermore, the information of operon organization, regulation, interactions, and dynamics have been used for identifying functionally linked genes10–12 annotating gene functions13,14 explaining the genome expansion/reduction,15,16 and facilitating the synthetic modification of biochemical processes.17–20

Operon organizations are considered to be well maintained even across phylogenetically distant genomes, as the proximity of functionally related genes offers more efficient regulation.5,6 Dynamic events such as division or recombination are also widely observed,6–8,21,22 suggesting that some operons might be a recent invention of evolution and others might result from convergent evolution. For example, a detailed examination of the repABC operon revealed that each member of this operon has its own evolutionary dynamics.23 Evolutionary models of operons such as tryptophan24 have been proposed to study their abundance, distribution of sizes, and evolutionary dynamics over time.25 Thus, a better understanding of the genome-wide operon organization and their dynamics among a large number of genomes will provide essential information not only for understanding experimental designs but also for understanding the evolutionary organization of prokaryotic genomes.

Experimental determination of operons is time consuming, and therefore, several computational methods have been presented to predict genome-wide operons by using a number of genomic/genetic features,26–29 including intergenic
distance, 30,31 conservation of gene order, 32,33 functional relationships, 34 and transcriptional data. 35 These methods have achieved high accuracies based on the validations of experimentally defined operons, 26–29 eg, 90.2% and 93.7% in Bacillus subtilis (B. subtilis) and Escherichia coli genomes, respectively. 30 With more genomes sequenced, the applications of these methods have allowed high-quality predicted operons and broad coverage of prokaryotic genomes. So far, the experimentally validated as well as computationally predicted operons of thousands of sequenced prokaryotes have been collected in several operon databases, 36–40 providing the opportunity to comprehensively understand the operons of prokaryotic genomes.

Although abundant information on operons is available, there is lack of genome-wide comparison of operons to understand their evolutionary dynamics based on the landscape of prokaryotic genomes. As a part of our demonstration, we analyzed the operon dynamics of Rhodococcus jostii RHA1 (RHA1) among 340 prokaryotic genomes. RHA1 is a soil actinomycete with exceptional abilities to synthesize, store, and degrade large types of lipids. 41–44 It has become a model bacterium to understand the pathway of lipid metabolism for biofuel development. 45–47 In this work, the aims of analyzing RHA1 operons are twofold: one is to provide insights of the evolutionary dynamics of RHA1 operons based on a diverse set of prokaryotic genomes and the second is to discover whether operon evolution contributed to the exceptional ability of lipid metabolism in RHA1 cells. We compared all RHA1 operons and their organization with 340 genomes to understand their dynamic evolution. Subsequently, we categorized the functional conservation of RHA1 operons and found that the operons related to lipid transport and metabolism are significantly conserved.

Materials and Methods
Selection of 341 bacterial genomes. To properly compare operon structures among genomes, we first need to carefully select the genomes. Currently, more than 5,000 prokaryotic genomes have been completely sequenced and are available in the NCBI database, 48 but they are largely imbalanced and biased to pathogenic bacteria such as E. coli and Mycobacterium smegmatis. Here, we selected only the genome with the largest DNA sequence from each genus of domain bacteria and archaea, where the genus was a well-used evolutionary distance for comparative genomic analysis. 49,50 Selecting only the largest genome of a genus will be beneficial not only to procure abundant information of genes but also to avoid the redundancy and imbalance of sequenced genomes among genera. In total, 341 genomes (Supplementary Table 1) of different genera were selected and downloaded from the NCBI database (NCBI release of August 2015). We used the RHA1 genome as the reference genome and the other 340 genomes as comparing genomes. All the genes/proteins mentioned in the paper are labeled by NCBI Geninfo Identifier numbers or if available by their official symbol names. Usually, one gene of a bacterial genome codes one protein, so we did not differentiate between gene and protein throughout the paper.

Comparative and phylogenetic analysis of RHA1 operons. We first aligned each of the 9,145 genes of RHA1 with all the genes of the 340 comparing genomes by using the BLASTP program. 51 For an RHA1 gene, we defined its homology gene in another comparing genome as the best-matched gene, which has the smallest e-value. If the e-values of genes of the comparing genome are all greater than 1e-05, then it is considered that no homology has been detected. 29 For all of the 341 selected genomes, their possible operons were predicted using the operon prediction program with the default parameters 30 and these operons are available at the DOOR database. 36,52 The 9,145 RHA1 genes were predicted to belong to 5,556 operons. A single gene is also considered an operon (termed a single-gene operon). Among the 5,556 operons, 55 operons have no homologies found in the 340 comparing genomes, while each of the other 5,501 operons includes at least one gene with homologies found in at least one of these comparing genomes. When comparing an operon A of RHA1 with operon B of another genome, we compared the genes of A and their homologies in B. Three dynamic types of an operon pair A and B were considered: deletion, elongation, and unchanged (Fig. 1). Operon A was defined as unchanged from operon B if all gene homologies of A were all found in operon B and vice versa. If only a subset of gene homologies of A were found in operon B, operon A was called deleted. If the gene homologies of A were all found in operon B and the gene number of A was less than B, B was defined as an elongation of A. As an extreme type of deletion, if no homologies of operon A were found in a comparing genome, A was called absent in this genome. For each of the RHA1 operons, we

---

**Figure 1.** Schematic view of dynamic changes of operons. Notes: Illustration of three possible types when comparing operon A and operon B. (i) Deletion: gene-3 was deleted in operon B. (ii) Unchanged: operon A and operon B had similar genes. (iii) Elongation: operon A was elongated to operon B where gene-4’ denotes the newly added gene in operon B.
compared it with all the operons of a comparing genome to detect its dynamic types, and recorded the number of genomes in which this type was observed. We then defined the ratio of deletion (or elongation or unchanged) for an operon as the proportion of the genomes with its deletion (or elongation or unchanged) observed among total genomes with any one of the three dynamic types. To describe the dynamic landscape of 5,556 operons within the 340 genomes, we constructed a $5,556 \times 340$ matrix (termed as the operon comparative matrix) by setting the state “Elongated” as 2, “Unchanged” as 1, “Deleted” as $-1$, and “Absent” as 0. A two-way hierarchical clustering method was performed on this matrix to analyze the evolutionary similarity of RHA1 operons.

Analyzing the functional conservation of operons. We used the Clusters of Orthologous Groups (COG) to classify the genes/operons of RHA1 into 17 functional categories. We defined an operon belonging to a functional category if most of the genes of this operon belong to this category. If each of the genes has a different COG, the operon is classified into the category “S: Function Unknown”. We then classified the 5,556 operons into 17 COG categories (operon groups). For each of the 5,556 operons, we calculated the number of genomes that the operon was kept Unchanged. Clearly, the greater the Unchanged number is for an operon, the more conserved it is among the 340 genomes. We then tested for each operon group (termed as $X$) if it is significantly conserved with all the other 16 operon groups (termed as $Y$). Mathematically, suppose there are $m$ and $n$ operons in $X$ and $Y$, we can achieve two vectors $(x_1, x_2, ..., x_m)$ and $(y_1, y_2, ..., y_n)$, where $x_i, i = 1, 2, ..., m$ and $y_i, i = 1, 2, ..., n$ are the number of genomes that operons were kept unchanged. Since the operon numbers the operon numbers of $X$ and $Y$ are usually different, we performed the two samples Kolmogorov–Smirnov test (K–S test) between them to test if they are significantly different or not. For each category, this statistical procedure can be considered a test based on sampling with replacement. We also classified the 17 categories into four super functional groups: information storage and processing, cellular processes and signaling, metabolism, and the poorly characterized group all according to the COG database. Similar to the statistical procedure used above, we tested whether the operons of a super functional group are more significantly conserved than the collection of operons from the other three groups. Gene phylogenetic analysis was performed using MEGA4.0. Functional enrichment analyses of gene sets were performed by utilizing the DAVID database.

Analyzing the operon conservation of pathways. All of the pathways from the 341 genomes were downloaded from the KEGG database (released in August 2015). There are a total of 109 pathways from RHA1, which include 4,148 operons. Similar to the operon analysis of COG functional categories mentioned above, we tested the conservation of the 109 pathways by using K–S test individually.

Results
Comparative analysis of RHA1 operons with 340 prokaryotic genomes. To investigate the evolutionary dynamics of operon structures, a comparative and phylogenetic analysis of RHA1 operons was performed on all operons of the 340 comparing prokaryotic genomes. In the RHA1 genome, the 9,145 genes were organized as 5,556 operons, and the gene number distribution of these operons is similar to the distribution of the total 701,360 operons from the 340 comparing genomes (Fig. 2A). The ratio of operons with at least two genes among the total 341 genomes is $\sim$35%, which is consistent with previous operon analysis. For an RHA1 operon, three possible dynamic types are considered if it is observed to be partially deleted, elongated, or unchanged in another comparison genome (“Materials and methods” section). In total, 99% (5,501) of the 5,556 operons were observed with deletion or elongation types among at least one of the comparing 340 genomes. The 3,603 single-gene operons were frequently observed to be elongated by combining with different genes, achieving the highest ratio of elongation as 52.19% (Fig. 2B). For larger operons, the ratio of elongation decreased, whereas the ratio of deletion increased. Surprisingly, we found that 19 larger operons (Supplementary Table 2), each with at least 10 genes, remained unchanged in a relatively high ratio of genomes, indicating that these operons were highly conserved in more genomes. We then performed functional enrichment analysis on the 209 genes of these 19 operons and found that they were mainly involved in 10 categories including NADH activities, rRNA/RNA binding, and structural constituents of ribosomes (lower figure, Fig. 2B). Five of these 10 categories are significantly related to NADH activity ($P < 1e-06$, Fisher’s exact test), suggesting that the molecular functions of NADH activities are evolutionarily conserved among prokaryotic organisms.

To further understand the bias of the three dynamic types, we constructed a $5,556 \times 340$ matrix to record the dynamic types of each operon within the 340 genomes. We analyzed the matrix using a two-way clustering method and manually annotated the operon clusters with their dominant dynamic events (elongation, deletion, or absence) among the 340 genomes. We found that all 5,556 operons were clustered into four groups. Approximately 20% of operons tend to elongate, 25% of operons tend to be deleted, and 5% of operons are mixed, either having deletions or elongations. This sums up more than 50% of the RHA1 operons, whose dynamic events may contribute to obtaining novel biological functions or regulatory modules in the different genomes. We also checked the genomes within different clusters, and found that they are relatively congruent with taxonomic classification from the NCBI database (as an example, see the 12 genera in Fig. 2C).

Evolutionarily dynamics of the His-operon: a case study. We selected an RHA1 operon for detailed analysis of its dynamic events among compared genomes. This operon is composed of 11 genes, including eight enzymes related to...
Chen et al

Evolutionary Bioinformatics

2016:12

histidine biosynthesis, one subB protein, and two hypothetical proteins (termed as “His-operon”). After investigating and clustering the His-operon with their homologies, we found that the two hypothetical genes (with Geninfo Identifier number of 111018037 and 111018041) were almost absent in 340 comparing genomes, suggesting that they could be newly obtained from the RHA1 genome (Fig. 3A). The His-operon turned out to be divided into several sub-operons in other genomes, even in two strains with close evolutionary distance to RHA1, Mycobacterium smegmatis MC2 155 (Mycobacterium) and Nocardia farcinica IFM 10152 (Nocardia). In the Mycobacterium genome, the His-operon was divided into two sub-operons: one keeps the main body of the His-operon and the other includes one separated gene 111018040 (hisC1). In the Nocardia genome, the His-operon was divided into three operons, where the main body of the His-operon in Mycobacterium was further separated into two smaller operons (Fig. 3A). Interestingly, different recombination events of hisC1 with other genes are observed in Mycobacterium and Nocardia (Fig. 3B). In Mycobacterium, hisC1 has a homologous gene 118470683 that was recombined with four genes annotated as MaoC, CoA transferase, amidohydrolase 3, and DNA-binding protein in an operon. Meanwhile, the homologies of these four genes in RHA1 are all separated from each other along the chromosome (indicated in red color, Figure 3B). In Nocardia, hisC1 has a homologous gene 54026425 that was recombined with three genes annotated as two ccrB proteins and one phosphoglucomutase in an operon (indicated in green color, Fig. 3B). The homologies of these three genes in RHA1 are all separated from each other along the chromosome (indicated in red color, Figure 3B). In Nocardia, hisC1 has a homologous gene 54026425 that was recombined with three genes annotated as two ccrB proteins and one phosphoglucomutase in an operon (indicated in green color, Fig. 3B). The homologies of these three genes in RHA1 are all separated from each other along the chromosome (indicated in red color, Figure 3B). In Nocardia, hisC1 has a homologous gene 54026425 that was recombined with three genes annotated as two ccrB proteins and one phosphoglucomutase in an operon (indicated in green color, Fig. 3B). The homologies of these three genes in RHA1 are all separated from each other along the chromosome (indicated in red color, Figure 3B).

Figure 2. Comparing RHA1 operons with those of 340 prokaryotic genomes.

Notes: (A) The gene number distribution of RHA1 operons and the operon union of 340 comparing genomes. (B) The statistical analysis of three dynamic types of the 5,501 RHA1 operons. The ratio of deletion, elongation, and unchanged was calculated as an average of the corresponding ratios for all operons with the same gene numbers. Functional enrichment analysis of 209 genes of the 19 larger operons (Supplementary Table 2) was performed using the DAVID database. (C) Clustered results of RHA1 operons and 340 comparing genomes. For an operon, its dynamic events in genomes are presented as different colored dots. The dominant dynamic events of operons were manually marked as Elongation, Deletion, and Absence. The operons were mainly clustered into three groups as elongation (red), deletion (green), and absence (blue), and their proportions of operons are noted in elliptical circles, respectively. The genomes of 12 genera with close evolutionary distances are clustered into one group. The bottom lists 10 genomes where most of the RHA1 operons are absent.
Investigating operon dynamics

340 genomes, we also noticed that a number of operons tend to remain unchanged (see the columns dominated with white spots in Fig. 2C). We investigated on the functions of these conserved operons, which are of great interest, since these well-maintained operons could contribute to important cellular processes and thus be essential for the evolution of prokaryotic organisms. We categorized all the genes in the 5,556 operons into COG categories, and then associated each operon to the COG category that the majority of its genes belonged to. We then checked the operons of four super COG functional categories (“Materials and methods” section) and found that the metabolism group is the most conserved. In detail, the operons of the metabolism group remained unchanged in an average of 99.5 of 430 genomes, which is significantly larger than the average number (47.5) for the operons of the other three groups \( (P = 8.12 \times 10^{-189}, \text{K-S test}) \). The metabolism super group includes eight basic categories (Table 1). All of them were tested to be significantly conserved \( (P < 1 \times 10^{-03}, \text{K-S test}) \), where the category of lipid transport and metabolism is the most conserved \( (P = 1.87 \times 10^{-39}) \).

Functional investigation of the conserved operons of RHA1 pathways. We also tested the functional conservation of operons based on biological pathways. From the KEGG pathway database,\(^{38}\) we downloaded 109 pathways for the RHA1 genome, which included 4,148 operons. All these pathways can be divided into three super groups: metabolism (92 pathways, 3,792 operons), genetic information processing (13 pathways, 218 operons), and environmental information processing (4 pathways, 138 operons). By analyzing the operons of these 109 pathways, we found that five pathways are significantly conserved, including ribosome (1.85e-06), oxidative

Table 1. Statistical analysis of eight metabolism categories.

<table>
<thead>
<tr>
<th>COG</th>
<th>FUNCTIONAL DESCRIPTION</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Lipid transport and metabolism</td>
<td>1.87E-39</td>
</tr>
<tr>
<td>C</td>
<td>Energy production and conversion</td>
<td>1.87E-33</td>
</tr>
<tr>
<td>E</td>
<td>Amino acid transport and metabolism</td>
<td>2.87E-25</td>
</tr>
<tr>
<td>G</td>
<td>Carbohydrate transport and metabolism</td>
<td>5.14E-19</td>
</tr>
<tr>
<td>Q</td>
<td>Secondary metabolites biosynthesis, transport and catabolism</td>
<td>1.15E-17</td>
</tr>
<tr>
<td>H</td>
<td>Coenzyme transport and metabolism</td>
<td>2.04E-13</td>
</tr>
<tr>
<td>P</td>
<td>Inorganic ion transport and metabolism</td>
<td>1.42E-10</td>
</tr>
<tr>
<td>F</td>
<td>Nucleotide transport and metabolism</td>
<td>4.42E-04</td>
</tr>
</tbody>
</table>

Figure 3. Evolutionary dynamics of His-operon. Notes: (A) Heatmap of the homologies of the 11 His-operon genes among the 340 selected genomes. The genomes clustered by using the e-values of the 11 His-operon genes that were obtained from the BLASTP program.\(^{31}\) These 11 genes were enriched in 25 genera (red cluster) and almost absent in 47 genera (blue cluster). Operon structures of the His-operon are shown for 25 genomes. Homologous genes are depicted in identical colors. (B) Divided operon structures of His-operon in M. smegmatis MC2 155 and N. farcinica IFM 10152. The e-values between homologous genes are noted adjacent to the gene boxes.
Table 2. Statistical analysis of five RHA1 pathways.

<table>
<thead>
<tr>
<th>KEGG</th>
<th>FUNCTION</th>
<th>P VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>00310</td>
<td>Ribosome</td>
<td>1.85E-06</td>
</tr>
<tr>
<td>00190</td>
<td>Oxidative phosphorylation</td>
<td>3.08E-06</td>
</tr>
<tr>
<td>00061</td>
<td>Fatty acid biosynthesis</td>
<td>5.51E-05</td>
</tr>
<tr>
<td>00523</td>
<td>Polyketide sugar unit biosynthesis</td>
<td>2.15E-04</td>
</tr>
<tr>
<td>00550</td>
<td>Peptidoglycan biosynthesis</td>
<td>8.39E-04</td>
</tr>
</tbody>
</table>

Discussion

The evolution of operons has been well studied in microorganisms such as E. coli; however, there is lack of genome-wide comparison of operon organization among a large number of prokaryotic genomes. Here, we have systematically categorized the conservation of RHA1 operons based on their dynamic types among the 340 compared genomes. The deletion and elongation of RHA1 operons are widely observed among diverse genomes, indicating that the organization of genes belonging to the same biological pathway followed different routes in different prokaryotes. Furthermore, the clustering analysis of the total 341 genomes based on the dynamic types of RHA1 operons largely matches with the taxonomic results from the NCBI database, suggesting that the majority of operons are inherited vertically.

Although a large amount of research and data are available regarding the structure, distribution, and functions of operons, the formation and dynamics of operons are still unclear. Our results confirmed that recombination events (such as deletion and elongation) are widely observed for most operons, supporting a highly dynamic view of operon formation and evolution. Divergent evolutionary events, including horizontal gene transfer, point mutations, and homologous recombination, have been hypothesized to be major force to drive operon formation and dynamics. Thus, it is interesting to further investigate the different rates of how these evolutionary events are involved in operon recombination among prokaryotic genomes. Our results also suggest that there could be a high false-positive ratio of identifying functionally linked genes or annotating gene functions using the information of operon organization, since genes performing different functions can form an operon and the operon structures are dynamically changing. As this is the case, we may need to integrate more different/independent information (such as co-evolution of genes, transcriptome) and employ better mathematical models to improve the precision of predictions.

RHA1 is known as a “lipid factory” for its high ability of synthesis and storage of diverse lipids. Our results provide potential evidence to explain its exceptional ability of lipid processes. First, we found that genes involved in the highly conserved operons mainly participate in eight COG functional categories of metabolism. Specially, several larger and conserved operons are functionally enriched in NADH dehydrogenase activity and the ribosome complex. Second, the His operon is well maintained as a whole-pathway operon, while its members are separated and recombined with different genes as new operons in diverse organisms. In general, we found that most of the operons related to metabolism tend to keep more gene members since they are often observed to be deleted in the 340 compared genomes. Based on the hypothesis that the genes in an operon are usually regulated as a unit, operons that embraced more functionally related members could provide high efficiency in biochemical processes. Therefore, the completeness of the RHA1 operons could be contributing to its high ability of lipid processes. RHA1 has been considered to have a high potential in biofuel development. To define its main pathways of lipid metabolism, such as triacylglycerol synthesis, a large number of transcriptomic analysis and biochemical experiments have been performed. Our comparative evaluations of the dynamic organizations of RHA1 operons could help to understand the pathways of lipid synthesis by mining combined operons among different genomes, and thus to improve the development of biofuel.

Acknowledgment

The authors gratefully acknowledge Prof. Pingsheng Liu for his insightful suggestions on the functional analysis.

Author Contributions

Conceived and designed the experiments: YC, SZ. Analyzed the data: YC, DG. Wrote the first draft of the manuscript: YC, DG. Made critical revisions and approved the final version: YC, DG, KE, SZ. All the authors reviewed and approved the final manuscript.

Supplementary Material

Supplementary Table 1. Detailed information of 341 genomes (xls file).

Supplementary Table 2. Gene list and annotation of 19 operons (xls file).

REFERENCES

61. Hernandez MA, Mohn WW, Martinez E, Rost E, Alvarez AF, Alvarez HM. Biosynthesis of storage compounds by Rhodococcus jostii RHA1 and global iden- tification of genes involved in their metabolism. BMC Genomics. 2008;9:600.