3 Biotech (2020) 10:376
https://doi.org/10.1007/s13205-020-02371-6

ORIGINAL ARTICLE

In silico functional and evolutionary analyses of rubber oxygenases (RoxA and RoxB)

Vikas Sharma[1] · **Fauzul Mobeen**[1] · **Tulika Prakash**[1]

Received: 26 July 2019 / Accepted: 28 July 2020 / Published online: 5 August 2020

Abstract

The study presents an in silico identification of poly (*cis*-1,4-isoprene) cleaving enzymes, viz., RoxA and RoxB in bacteria, followed by their functional and evolutionary exploration using comparative genomics. The orthologs of these proteins were found to be restricted to Gram-negative beta-, gamma-, and delta-proteobacteria. Toward the evolutionary propagation, the *RoxA* and *RoxB* genes were predicted to have evolved via a common interclass route of horizontal gene transfer in the phylum Proteobacteria (delta → gamma → beta). Besides, recombination, mutation, and gene conversion were also detected in both the genes leading to their diversification. Further, the differential selective pressure is predicted to be operating on entire *RoxA* and *RoxB* genes such that the former is diversifying further, whereas the latter is evolving to reduce its genetic diversity. However, the structurally and functionally important sites/residues of these genes were found to be preventing changes implying their evolutionary conservation. Further, the phylogenetic analysis demonstrated a sharp split between the RoxA and RoxB orthologs and indicated the emergence of their variant as another type of putative rubber oxygenase (RoxC) in the class Gammaproteobacteria. A detailed in silico analysis of the signature motifs and residues of Rox sequences exhibited important differences as well as similarities among the RoxA, RoxB, and putative RoxC sequences. Although RoxC appears to be a hybrid of RoxA and RoxB, the signature motifs and residues of RoxC are more similar to RoxB.

Keywords Evolution · Horizontal gene transfer · Natural rubber degradation · Phylogeny · Rubber oxygenases · Selective pressure

Introduction

Natural rubber (NR) is a biopolymer of *cis*-1,4-isoprene and has been commercially exploited because of its wide range of domestic and industrial applications. As NR does not accumulate in the environment, NR-degradation must occur. Bioremediation of NR is a potential eco-friendly alternative strategy to cope with the environmental and health hazards associated with the traditional methods of NR-degradation. Several reports on bacterial biodegradation of NR have been published which have revealed two distinct groups of NR-degrading bacteria, viz., Gram-positive Actinobacteria and Gram-negative Proteobacteria (Jendrossek and Birke 2019). The key enzyme involved in the NR-degradation process is different in both the Gram-positive and Gram-negative bacteria with the former harboring latex clearing protein (Lcp), a *b*-type mono-heme cytochrome and the latter harboring rubber oxygenase (Rox) enzymes (*c*-type di-heme cytochrome) (Jendrossek and Birke 2019). The distribution of Lcp has been explored in several bacteria, whereas the studies on Rox distribution are limited to only a few members, including *Steroidobacter cummioxidans* 35Y, *Rhizobacter gummiphilus* NS21, *Haliangium ochraceum* SMP-2, *Corallococcus coralloides* B035, and *Myxococcus fulvus* HW1. Among all NR-degrading bacteria, Gram-negative *S. cummioxidans* 35Y (previously *Xanthomonas* sp. 35Y) is one of the fastest-growing strains, when cultivated

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-020-02371-6) contains supplementary material, which is available to authorized users.

✉ Tulika Prakash
tulika@iitmandi.ac.in

Vikas Sharma
vikas.sharma.biotech@gmail.com

Fauzul Mobeen
faizgwalior@gmail.com

[1] School of Basic Sciences, Indian Institute of Technology Mandi, Kamand, Mandi 175005, Himachal Pradesh, India

on rubber latex (Jendrossek and Birke 2019; Sharma et al. 2018; Tsuchii and Takeda 1990). Despite harboring a faster NR-degrading bacterium, the Gram-negative group of NR-degraders is much less studied. Thus, there is a need to explore this bacterial group to search for potential candidates which might have better NR-degradation efficiency in comparison to the known NR-degrading microbes.

In the Gram-negative NR-degrading bacteria, two paralogous copies of Rox enzymes viz., RoxA and RoxB, have been demonstrated to function synergistically (Jendrossek and Birke 2019). The RoxA and RoxB proteins resemble each other with respect to their molecular masses for mature proteins (71.5 kDa for RoxA and 70.3–70.8 kDa for RoxB) and cofactor content (2 *c*-type hemes). However, they share only 38% amino acid sequence identity and differ in their functional properties. For example, RoxB cleaves poly (*cis*-1,4-isoprene) randomly into a distinctive product spectrum of C_{20}, C_{25}, C_{30}, and higher oligo-isoprenoids (Birke and Jendrossek 2014; Birke et al. 2018) in an *endo*-type manner (Fig. 1). In contrast, RoxA cleaves poly (*cis*-1,4-isoprene) from one end in an *exo*-type fashion into 12-oxo-4,8-dimethyltrideca-4,8-diene-1-al (ODTD), a C_{15} oligo-isoprenoid, as the major end product. Besides, RoxA can also degrade Lcp- and RoxB-derived higher oligo-isoprenoid units into smaller ODTD units, which upon oxidoreduction can be easily taken up by the bacterial cells for further metabolism. Thus, the primary functions of RoxA and RoxB remain the same, which is NR-cleavage; however, the mechanisms for this process are quite different for each enzyme. These observations indicate substantial evolutionary events of divergence, propagation, and co-evolution of these proteins.

The emergence of RoxA and RoxB paralogs is the result of the continuous adaptive evolution of Rox proteins in different bacterial hosts. A complete understanding of the functional divergence of Rox proteins requires comprehensive evolutionary exploration of these proteins as demonstrated in the case of other genes/proteins (Grove 2010; Gruen et al. 2019; Juárez-Vázquez et al. 2017; Noor et al. 2014). The

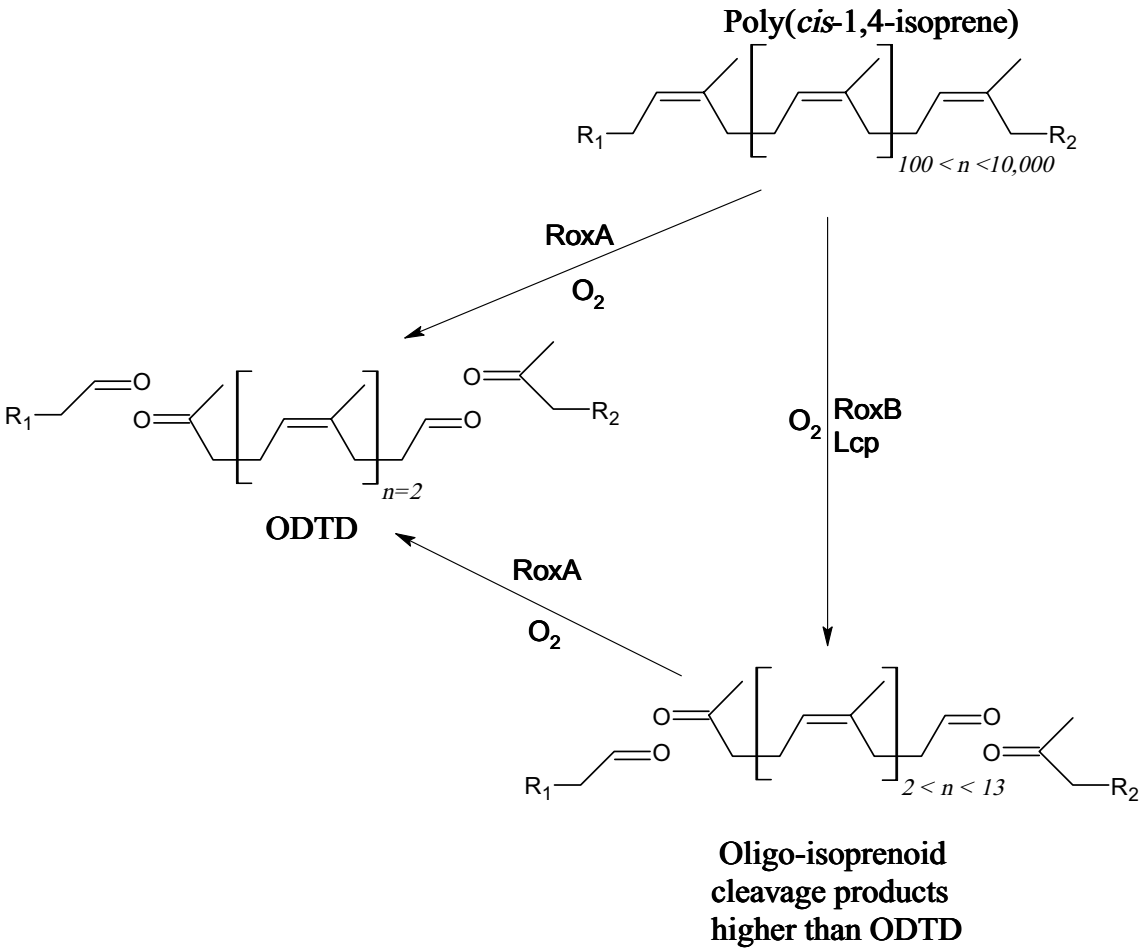


Fig. 1 Oxidative cleavage of Poly(*cis*-1,4-isoprene) by RoxA, RoxB, and Lcp enzymes. Poly(*cis*-1,4-isoprene) ($100 < n < 10{,}000$) is oxidatively cleaved by RoxA, RoxB, and Lcp to oligo-isoprenoids with terminal keto- and aldehyde groups, where cleaved products differ in the number of intact isoprene units (n) with $2 \leq n < 13$. ODTD: 12-oxo-4,8-dimethyltrideca-4,8-diene-1-al

evolution of prokaryotic genomes can be described within the framework of population genetics by exploring their genetic relatedness (phylogeny) and the processes of horizontal gene transfer (HGT), recombination, selective pressure, gene conversion, and mutation. Of these, phylogeny is a major tool to study the evolution of a genetically related group of organisms by the comparison of orthologous sequences. HGT is an important process of trait accumulation via the exchange of genetic material for adaptive purposes (Arber 2014). The traits, which undergo HGT, must be naturally selected to prevent their loss by simple genetic drift (Perry and Wright 2014). For adaptive evolution, the traits once accumulated in the genome undergo the process of diversification, which is attributed to point mutations (Bryant et al. 2012), events of gene conversion (Darmon and Leach 2014), and recombination (Arber 2014; Boc and Makarenkov 2011). Importantly, all these evolutionary attributes remain completely unexplored for Rox proteins, except for the phylogenetic analysis, which has been carried out in a few studies only on a limited number of bacteria (Birke et al. 2017; Jendrossek and Reinhardt 2003; Kasai et al. 2017).

The observations mentioned above prompted us to perform comprehensive analyses of the RoxA and RoxB proteins of Gram-negative NR-degrading bacteria to explore their distribution and functional and evolutionary traits in the prokaryotic domain using in silico approaches of comparative genomics. To this end, we predicted the orthologs of key rubber-cleaving enzymes, viz., RoxA and RoxB, in the bacterial domain to identify the putative Gram-negative bacterial strains with the potential to degrade NR. We further evaluated the functional homology of the predicted orthologs with well-characterized RoxA and RoxB enzymes. We also studied the evolution of *RoxA* and *RoxB* genes, which might have led to the emergence of these functionally distinct paralogs. Toward this, we carried out HGT, intragenic recombination, recombination rate, mutation rate, gene conversion, and selective pressure analyses.

Materials and methods

Retrieval of genome sequences

In this study, 38 publicly available bacterial genomes of complete or draft quality (Table S1) were used for the analyses. The nucleotide, coding genes, and protein sequences of *S. cummioxidans* 35Y were downloaded from the Figshare repository (https://doi.org/10.6084/m9.figshare.6226466). For the remaining bacteria, the data were downloaded from the NCBI databases (GenBank (https://ftp.ncbi.nlm.nih.gov/genbank) or RefSeq (https://ftp.ncbi.nlm.nih.gov/refseq), version as of February 2018). A flowchart of methodology and bioinformatics tools used for the present analysis is given in Fig. S1.

Assignment of orthologs

For comparative analysis, assignment of orthologs, and similarity searches, an E-value cutoff of 10^{-4}, a threshold of 30% amino acid sequence identity, and 60% coverage of both query and hit in the alignment (BLASTP) were used throughout the analyses unless otherwise specified.

Mining of RoxA and RoxB orthologs

Online BLASTP v2.8.0+ (Camacho et al. 2009) was used with default parameters to search for orthologs of the RoxA (STC_00358) and RoxB (STC_02518) protein sequences of *S. cummioxidans* 35Y against the NCBI database of non-redundant protein sequences (NR) (https://blast.ncbi.nlm.nih.gov/Blast.cgi, version as of February 2018). An ortholog was assigned to RoxA or RoxB designation based on its highest amino acid sequence identity, either with RoxA or RoxB protein sequence of *S. cummioxidans* 35Y.

Sequence alignment

Nucleotide and protein sequences were aligned using the T-Coffee server (Di Tommaso et al. 2011) with default parameters. The codon alignment of nucleotide sequences was constructed by PAL2NAL (Suyama et al. 2006) based on protein alignments. Unipro UGENE v1.24.1 (Okonechnikov et al. 2012) was used for the visualization of alignments. The consensus positions of amino acid residues in a given alignment were used throughout the analyses.

Heatmap construction

GENE-E matrix visualization and analysis program (https://www.broadinstitute.org/cancer/software/GENE-E) was used to generate the heatmap and hierarchical clustering based on amino acid sequence identities (%) among RoxA and RoxB orthologs which were calculated using BLASTP v2.8.0+ (Camacho et al. 2009).

Phylogenetic tree reconstruction

For all the phylogenetic analyses, *Thermosulfurimonas dismutans* was used as the outgroup. FigTree v1.4.2 (Rambaut 2014) was used for tree visualization.

Gene tree

For all the phylogenetic analyses, the neighbor-joining method (Saitou and Nei 1987) implemented in MEGA6

(Tamura et al. 2013) was used to infer evolutionary history. Robustness of the phylogenetic trees was evaluated using the bootstrap resampling method of Felsenstein (Felsenstein 1985). The bootstrap consensus trees inferred from 1000 replicates were taken to represent the evolutionary history of the taxa analyzed. The trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer phylogenetic trees. All the positions containing gaps and missing data were eliminated. The evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

For phylogenetic tree reconstruction of RoxA and RoxB protein orthologs, their aligned protein sequences were used. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 90 amino acid sequences with a total of 220 positions in the final dataset.

For HGT analyses, all the gene trees were constructed using nucleotide alignments. The evolutionary distances were computed using the maximum composite likelihood method (Tamura et al. 2004) and are in units of the number of base substitutions per site. Codon positions included were 1st + 2nd + 3rd + Noncoding. The phylogenetic analyses of *RoxA* and *RoxB* orthologous genes involved 24 nucleotide sequences with a total of 885 positions and 29 nucleotide sequences with a total of 759 positions in the final dataset, respectively.

Species tree

For HGT analysis, all the species trees were obtained using the standalone version of Up-to-date Bacterial Core Gene set (UBCG) software (Na et al. 2018) with default parameters. The software employs a set of 92 single-copy core genes commonly present in all bacterial genomes for the construction of the genome-based phylogenetic tree. To this end, UBCG software finds the genes in bacterial genomes using PRODIGAL (Hyatt et al. 2010) and extracts the core genes using HMMER3 (Eddy 2011). Further, it generates multiple-sequence alignment of each of the orthologous core gene sets using MAFFT (Katoh and Standley 2013). These MSAs are concatenated followed by excluding the alignment positions having gap characters more than 50%. Finally, it calculates the maximum likelihood phylogenetic tree using FastTree (Price et al. 2010) as a default tool. The Gene Support Index (GSI) (Na et al. 2018) implemented in UBCG was used to evaluate the robustness of the phylogenetic trees. GSI indicates how many genes support the branch in the concatenated phylogenetic tree. The GSI values of 50 or greater were considered statistically significant and indicated support for a clade in our analysis.

Functional analyses

For the functional analysis of the orthologs of RoxA and RoxB proteins, their protein sequences were scanned with the standalone version of InterProScan v5.10–50.0 (Jones et al. 2014) to classify them into families and predicting their domains and important sites. Inmembrane program (Perry and Ho 2013) was used for the prediction of the subcellular localization in Gram-negative bacteria with default parameters. This program includes a transmembrane helix predictor (TMMOD) (Kahsay et al. 2005), a secretion signal predictor (SignalP) (Petersen et al. 2011), a lipoprotein signal predictor (LipoP) (Juncker et al. 2003), and a sequence alignment for protein profiles (HMMER) (Durbin et al. 1998). SecretomeP v2.0 (Bendtsen et al. 2005) was used for the prediction of non-classically secreted proteins of Gram-negative bacteria. Here, the SecP score ≥ 0.5 was used as the determining value to indicate possible secretion.

Horizontal gene transfer analysis

Complete HGT events were analyzed with Tree and Reticulogram reconstruction (T-REX) server (Boc et al. 2012) using the bipartition dissimilarity (BD) as the optimization criterion (i.e., minimum cost) (Boc et al. 2010). For a given pair of species and gene trees, the BD test calculates their proximity by several distance-based measures and predicts minimum cost scenario HGTs by the progressive reconciliation of those trees using a refinement of the Robinson and Foulds distance (Boc et al. 2010; Robinson and Foulds 1981). T-REX was optimized by taking into account the evolutionary events, including gene duplication and deletion. The HGT detection was carried out with the bootstrap resampling option available in T-REX to assess the support of the tree branches. Importantly, bootstrap score values of 100% were considered statistically significant and showed strong support for all the predicted gene transfers.

Intragenic recombination, recombination rate, mutation rate, and gene conversion analyses

The codon alignments of both *RoxA* and *RoxB* genes were screened for recombination breakpoints using the GARD algorithm (Pond et al. 2006) implemented in the HyPhy package (Pond et al. 2005) on the Datamonkey server (Pond and Frost 2005). The presence of recombination was statistically verified using the pairwise homoplasy test (PHI) (Bruen et al. 2006) implemented in SplitsTree v.4.14.8 (Huson and Bryant 2005). We further statistically tested the assumption that the variation in *RoxA* and *RoxB* genes is neutral using the Tajima's D test (Tajima 1989). More specifically, we tested the neutrality of mutations in the entire coding sequence of *RoxA* and *RoxB* genes, as

well as independently for its major sub-regions comprising important structural and functional sites/residues. The codon alignments of *RoxA* and *RoxB* genes and their major sub-regions were used for the Tajima's D test of neutrality (Tajima 1989) using the algorithm implemented in MEGA6 (Tamura et al. 2013). In addition, seven methods implemented in the Recombination Detection Program (RDP) suite v4.97 (Martin et al. 2015), including 3Seq (Boni et al. 2007), Bootscan (Martin et al. 2005), Chimaera (Posada 2002), GENECONV (Padidam et al. 1999), MaxChi (Smith 1992), RDP (Martin and Rybicki 2000), and SisScan (Gibbs et al. 2000), were also applied to detect potential recombinant sequences, parental sequences, and recombination breaking points in both *RoxA* and *RoxB* genes. Analyses were performed with default settings for the different detection methods and each possible event was considered statistically significant when a supporting p value was less than 0.05 in more than one detection method.

The mutation rate (theta, Watterson's mutation parameter) and recombination parameter (rho) of *RoxA* and *RoxB* gene sets were estimated by the LDHat package (McVean et al. 2002) implemented in RDP suite v4.97 (Martin et al. 2015). To detect the events of gene conversion, codon alignment of *Rox* genes (*RoxA* + *RoxB*) was analyzed for signatures of interparalog gene conversion by a previously described method (Betrán et al. 1997) implemented in DNA Sequence Polymorphism (DnaSP) software v6.12.01 (Librado and Rozas 2009).

Selective pressure analyses

Selective pressure (ω) was calculated as the ratio of non-synonymous (K_a) to synonymous (K_s) substitution rates (Hu and Banzhaf 2008; Hurst 2002; Pond et al. 2005). The non-synonymous substitutions in a nucleotide sequence cause a change in the translated amino acid sequence, whereas the synonymous substitutions do not alter it. The value of ω greater than one implies positive or Darwinian or diversifying selection of a gene which drives the changes to increase its genetic diversity (Shapiro and Alm 2008). The value of ω less than one implies negative or purifying or stabilizing selection of a gene which eliminates its deleterious mutations that arise in a population to reduce its genetic diversity. The value of ω equal to one of a gene indicates its neutral selection.

The codon alignment was used to calculate K_a and K_s using the Nei–Gojobori (NG) method (Nei and Gojobori 1986) implemented in KaKs_Calculator v1.2 (Zhang et al. 2006) with default parameters. To assess the selection at codon level for both *RoxA* and *RoxB* genes, their codon alignments were subjected to selective pressure analyses using the HyPhy algorithm (Pond et al. 2005) implemented in MEGA6 (Tamura et al. 2013).

Results

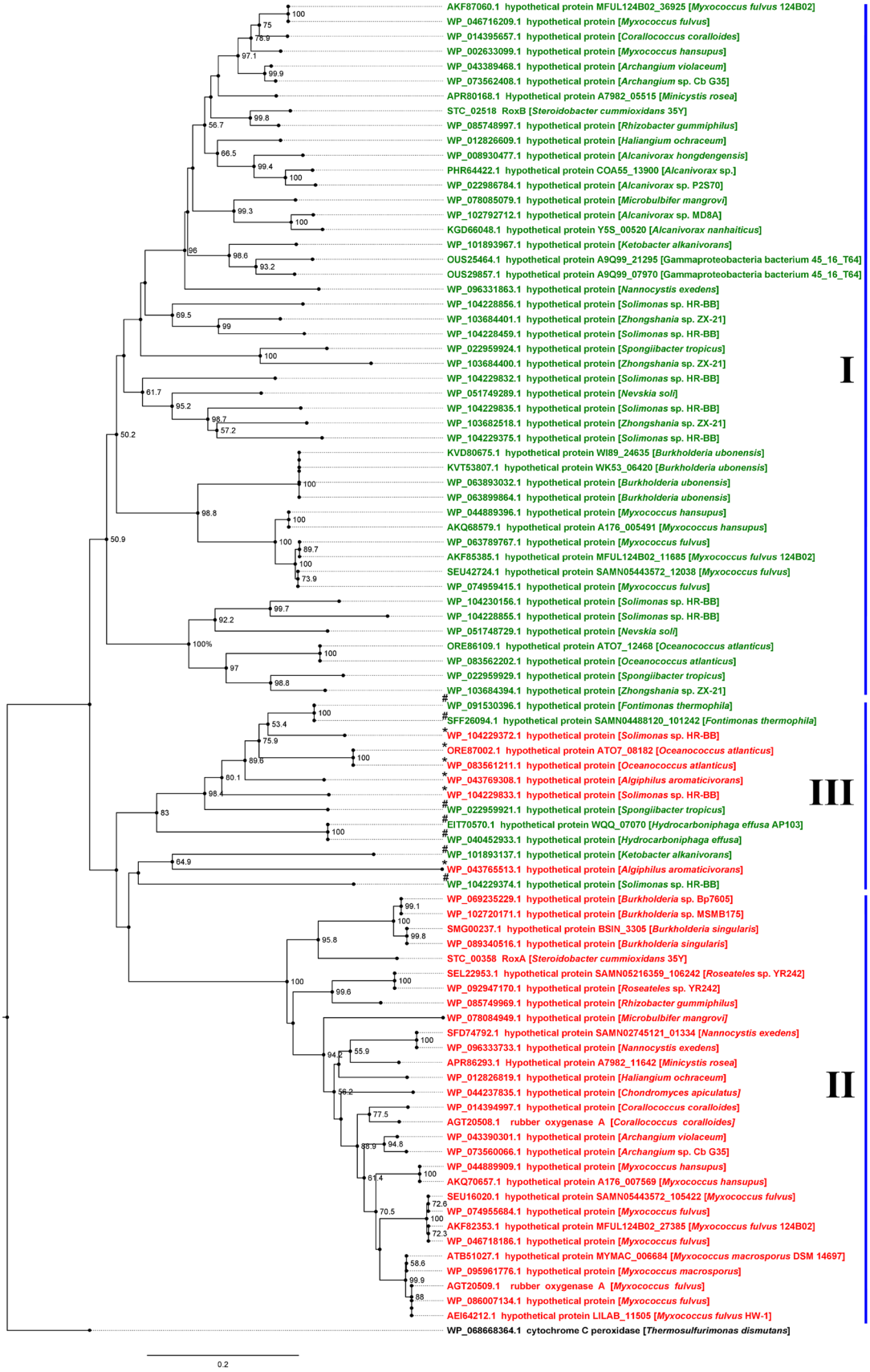
Distribution of RoxA and RoxB proteins among prokaryotes

We identified the orthologs of rubber oxygenases (RoxA and RoxB) and explored their distribution and evolutionary relatedness among prokaryotes. The orthologous protein sequences of RoxA and RoxB were subjected to comprehensive functional analyses to examine their resemblances with the well-characterized RoxA and RoxB protein sequences of *S. cummioxidans* 35Y.

In silico identification and exploration of phylogenetic relatedness of RoxA and RoxB orthologs in Gram-negative bacteria

A sequence similarity-based approach was used to identify the orthologous protein sequences of rubber oxygenases (RoxA and RoxB) of *S. cummioxidans* 35Y in the NR database. An ortholog was assigned to RoxA or RoxB designation based on its highest amino acid sequence identity, either with the previously well-characterized RoxA or RoxB protein sequence of *S. cummioxidans* 35Y. From this analysis, 87 orthologs (34 similar to RoxA and 53 similar to RoxB) were identified, in addition to the RoxA and RoxB of *S. cummioxidans* 35Y. These orthologs ($n = 89$) belonged to the Gram-negative members of the beta, gamma, and delta classes of the phylum Proteobacteria (Table S2). Deltaproteobacteria (57%) was the most abundant class, followed by Gammaproteobacteria to which the RoxA orthologs ($n = 35$) were assigned. An opposite trend was observed for the RoxB orthologs ($n = 54$), wherein Gammaproteobacteria (63%) was the most abundant class followed by Deltaproteobacteria. However, Betaproteobacteria was the least abundant class assigned to both the RoxA and RoxB orthologs.

To explore evolutionary relatedness, a comprehensive phylogenetic analysis of 89 RoxA and RoxB orthologous proteins was performed. Since the cytochrome *c* peroxidase A (CcpA) is distantly related to RoxA protein (Seidel et al. 2013), we used this enzyme/gene of *Thermosulfurimonas dismutans* as the outgroup for phylogenetic analysis. We observed three separate clusters in the phylogenetic tree (Fig. 2). The first cluster belonged to RoxB orthologs ($RoxB_I$), the second cluster belonged to RoxA orthologs ($RoxA_{II}$), and the third cluster had both RoxA and RoxB sequences ($RoxA_{III}$ and $RoxB_{III}$). The presence of the third cluster in the phylogenetic tree indicated the emergence of another type of putative rubber oxygenases in Gram-negative bacteria. It should be noted that the


AKF87060.1 hypothetical protein MFUL124B02_36925 [Myxococcus fulvus 124B02]
WP_046716209.1 hypothetical protein [Myxococcus fulvus]
WP_014395657.1 hypothetical protein [Corallococcus coralloides]
WP_002633099.1 hypothetical protein [Myxococcus hansupus]
WP_043389468.1 hypothetical protein [Archangium violaceum]
WP_073562408.1 hypothetical protein [Archangium sp. Cb G35]
APR80168.1 Hypothetical protein A7982_05515 [Minicystis rosea]
STC_02518 RoxB [Steroidobacter cummioxidans 35Y]
WP_085748997.1 hypothetical protein [Rhizobacter gummiphilus]
WP_012826609.1 hypothetical protein [Haliangium ochraceum]
WP_008930477.1 hypothetical protein [Alcanivorax hongdengensis]
PHR64422.1 hypothetical protein COA55_13900 [Alcanivorax sp.]
WP_022986784.1 hypothetical protein [Alcanivorax sp. P2S70]
WP_078085079.1 hypothetical protein [Microbulbifer mangrovi]
WP_102792712.1 hypothetical protein [Alcanivorax sp. MD8A]
KGD66048.1 hypothetical protein Y5S_00520 [Alcanivorax nanhaiticus]
WP_101893967.1 hypothetical protein [Ketobacter alkanivorans]
OUS25464.1 hypothetical protein A9Q99_21295 [Gammaproteobacteria bacterium 45_16_T64]
OUS29857.1 hypothetical protein A9Q99_07970 [Gammaproteobacteria bacterium 45_16_T64]
WP_096331863.1 hypothetical protein [Nannocystis exedens]
WP_104228856.1 hypothetical protein [Solimonas sp. HR-BB]
WP_103684401.1 hypothetical protein [Zhongshania sp. ZX-21]
WP_104228459.1 hypothetical protein [Solimonas sp. HR-BB]
WP_022959924.1 hypothetical protein [Spongiibacter tropicus]
WP_103684400.1 hypothetical protein [Zhongshania sp. ZX-21]
WP_104229832.1 hypothetical protein [Solimonas sp. HR-BB]
WP_051749289.1 hypothetical protein [Nevskia soli]
WP_104229835.1 hypothetical protein [Solimonas sp. HR-BB]
WP_103682518.1 hypothetical protein [Zhongshania sp. ZX-21]
WP_104229375.1 hypothetical protein [Solimonas sp. HR-BB]
KVD80675.1 hypothetical protein WI89_24635 [Burkholderia ubonensis]
KVT53807.1 hypothetical protein WK53_06420 [Burkholderia ubonensis]
WP_063893032.1 hypothetical protein [Burkholderia ubonensis]
WP_063899864.1 hypothetical protein [Burkholderia ubonensis]
WP_044889396.1 hypothetical protein [Myxococcus hansupus]
AKQ68579.1 hypothetical protein A176_005491 [Myxococcus hansupus]
WP_063789767.1 hypothetical protein [Myxococcus fulvus]
AKF85385.1 hypothetical protein MFUL124B02_11685 [Myxococcus fulvus 124B02]
SEU42724.1 hypothetical protein SAMN05443572_12038 [Myxococcus fulvus]
WP_074959415.1 hypothetical protein [Myxococcus fulvus]
WP_104230156.1 hypothetical protein [Solimonas sp. HR-BB]
WP_104228855.1 hypothetical protein [Solimonas sp. HR-BB]
WP_051748729.1 hypothetical protein [Nevskia soli]
ORE86109.1 hypothetical protein ATO7_12468 [Oceanococcus atlanticus]
WP_083562202.1 hypothetical protein [Oceanococcus atlanticus]
WP_022959929.1 hypothetical protein [Spongiibacter tropicus]
WP_103684394.1 hypothetical protein [Zhongshania sp. ZX-21]
WP_091530396.1 hypothetical protein [Fontimonas thermophila]
SFF26094.1 hypothetical protein SAMN04488120_101242 [Fontimonas thermophila]
WP_104229372.1 hypothetical protein [Solimonas sp. HR-BB]
ORE87002.1 hypothetical protein ATO7_08182 [Oceanococcus atlanticus]
WP_083561211.1 hypothetical protein [Oceanococcus atlanticus]
WP_043769308.1 hypothetical protein [Algiphilus aromaticivorans]
WP_104229833.1 hypothetical protein [Solimonas sp. HR-BB]
WP_022959921.1 hypothetical protein [Spongiibacter tropicus]
EIT70570.1 hypothetical protein WQQ_07070 [Hydrocarboniphaga effusa AP103]
WP_040452933.1 hypothetical protein [Hydrocarboniphaga effusa]
WP_101893137.1 hypothetical protein [Ketobacter alkanivorans]
WP_043765513.1 hypothetical protein [Algiphilus aromaticivorans]
WP_104229374.1 hypothetical protein [Solimonas sp. HR-BB]
WP_069235229.1 hypothetical protein [Burkholderia sp. Bp7605]
WP_102720171.1 hypothetical protein [Burkholderia sp. MSMB175]
SMG00237.1 hypothetical protein BSIN_3305 [Burkholderia singularis]
WP_089340516.1 hypothetical protein [Burkholderia singularis]
STC_00358 RoxA [Steroidobacter cummioxidans 35Y]
SEL22953.1 hypothetical protein SAMN05216359_106242 [Roseateles sp. YR242]
WP_092947170.1 hypothetical protein [Roseateles sp. YR242]
WP_085749969.1 hypothetical protein [Rhizobacter gummiphilus]
WP_078084949.1 hypothetical protein [Microbulbifer mangrovi]
SFD74792.1 hypothetical protein SAMN02745121_01334 [Nannocystis exedens]
WP_096333733.1 hypothetical protein [Nannocystis exedens]
APR86293.1 Hypothetical protein A7982_11642 [Minicystis rosea]
WP_012826819.1 hypothetical protein [Haliangium ochraceum]
WP_044237835.1 hypothetical protein [Chondromyces apiculatus]
WP_014394997.1 hypothetical protein [Corallococcus coralloides]
AGT20508.1 rubber oxygenase A [Corallococcus coralloides]
WP_043390301.1 hypothetical protein [Archangium violaceum]
WP_073560066.1 hypothetical protein [Archangium sp. Cb G35]
WP_044889909.1 hypothetical protein [Myxococcus hansupus]
AKQ70657.1 hypothetical protein A176_007569 [Myxococcus hansupus]
SEU16020.1 hypothetical protein SAMN05443572_105422 [Myxococcus fulvus]
WP_074955684.1 hypothetical protein [Myxococcus fulvus]
AKF82353.1 hypothetical protein MFUL124B02_27385 [Myxococcus fulvus 124B02]
WP_046718186.1 hypothetical protein [Myxococcus fulvus]
ATB51027.1 hypothetical protein MYMAC_006684 [Myxococcus macrosporus DSM 14697]
WP_095961776.1 hypothetical protein [Myxococcus macrosporus]
AGT20509.1 rubber oxygenase A [Myxococcus fulvus]
WP_086007134.1 hypothetical protein [Myxococcus fulvus]
AEI64212.1 hypothetical protein LILAB_11505 [Myxococcus fulvus HW-1]
WP_068668364.1 cytochrome C peroxidase [Thermosulfurimonas dismutans]
I
III
II
0.2

◂**Fig. 2** Phylogenetic relationships among the 89 orthologs of rubber oxygenases (RoxA and RoxB) in the phylum Proteobacteria. The phylogenetic tree was constructed based on amino acid sequences of the RoxA and RoxB orthologs by using the neighbor-joining method. The cytochrome *c* peroxidase amino acid sequence of *Thermosulfurimonas dismutans* was used as the outgroup. Numbers at nodes indicate the levels of bootstrap support (%) based on a maximum likelihood analysis of 1000 resampled datasets; values of less than 50% are not shown. Bar, 0.2 substitutions per site. Solid lines represent the lengths of branches; dotted lines are used to align the tip labels for better visualization; the circle represents the node; red fonts represent RoxA orthologs, green fonts represent RoxB orthologs; * represents the RoxA orthologs in cluster III, # represents the RoxB orthologs in cluster III; I: cluster I orthologs of RoxB ($RoxB_I$); II: cluster II orthologs of RoxA ($RoxA_{II}$); III: cluster III orthologs of RoxA ($RoxA_{III}$) and RoxB ($RoxB_{III}$) and are termed as putative RoxC

catalytic activities of RoxA and RoxB enzymes, which form a part of the cluster I ($RoxB_I$) and cluster II ($RoxA_I$), respectively, are experimentally confirmed in *S. cummioxidans* 35Y (Jendrossek and Birke 2019) and *R. gummiphilus* NS21 (Birke et al. 2018). However, the catalytic activities of RoxA and RoxB enzymes of all members of cluster III ($RoxA_{III}$ and $RoxB_{III}$) remain unknown. These three clusters were further confirmed by an amino acid identity-based heatmap of RoxA and RoxB orthologs (Fig. S2). Of these three clusters, the first and second clusters possessed rubber oxygenases of the beta, gamma, and delta classes of the phylum Proteobacteria, whereas the third cluster belonged to the class Gammaproteobacteria only. It is interesting to note that a few Proteobacteria were found to harbor orthologs of both RoxA and RoxB proteins, whereas few others had either a RoxA or a RoxB protein encoded in their genomes (Table S2).

In silico functional analyses of Rox orthologs for signature motifs and subcellular localization

The functional annotations of most of the RoxA and RoxB protein orthologs (~ 96%) remain unknown as these proteins are annotated as "hypothetical proteins" (Table S2). To predict the functions of these orthologs, we performed comprehensive analyses of their functional motifs, domains, and structurally conserved sites/residues. The RoxA and RoxB orthologs were found to exhibit conservation at most of their respective signature motifs and residues (Table 1). Importantly, all RoxA and RoxB orthologs were found to harbor two conserved CXXCH motifs (Figs. S3 and S4). The occurrence of these cytochrome motifs was further confirmed by the InterProScan analysis of all RoxA and RoxB orthologs. This analysis predicted the presence of a cytochrome *c*-like domain (IPR009056) as the conserved motif that possesses electron transfer activity (GO:0009055) and heme-binding (GO:0020037) functions in all RoxA and RoxB orthologs (data not showed). All RoxA and RoxB orthologs were also found to harbor a conserved MauG motif (PXXHNXXXP) (Fig. S5). Besides, the axial histidine ligands of heme 1 (Fig. S3) and heme 2 (Figs. S4 and S6) were also found to be conserved at the corresponding positions among all the RoxA and RoxB orthologs. Other important residues, viz., A251, I252, L254, I255, F301, A316, and F317 (Seidel et al. 2013) were not found to be conserved at their corresponding positions in all the RoxA and RoxB orthologs (Table 1 and Fig. S7). However, A251, F301, and F317 residues were substituted only by other non-polar and hydrophobic amino acids in all the orthologs which indicated functional conservation of these amino acids toward a significant role in the enzyme function.

We further identified the signature motifs and residues, which were specific to the Rox sequences belonging to either $RoxB_I$ or $RoxA_{II}$ or $RoxA_{III}$ and $RoxB_{III}$ clusters (Table 1). Notably, the $RoxA_{II}$ cluster sequences harbored the CSXCH motif of heme 1, whereas it was CXXCH in the sequences of the $RoxB_I$ cluster. Similarly, the sequences of the $RoxA_{II}$ cluster harbored the CASCH motif of heme 2, whereas the sequences of the $RoxB_I$ cluster possessed the CXXCH motif at this position. The MauG motif, viz., PXFHNXSVP, was conserved among the sequences of $RoxA_{II}$ cluster, whereas it was changed to PYXHNXXXP in the sequences of $RoxB_I$ cluster. Interestingly, the signature motifs and residues of $RoxA_{II}$ sequences were found to harbor more conserved sites in comparison to those of $RoxB_I$ sequences. In addition, the non-conserved sites of these motifs and residues of the sequences of $RoxB_I$ and $RoxA_{II}$ clusters were also found to have a prevalence of non-synonymous substitutions.

The sequences of the third cluster ($RoxA_{III}$ and $RoxB_{III}$) harbored the CXXCH motif of heme 1, which was different from that of the sequences of $RoxA_{II}$ cluster (CSXCH), but was more similar to that of $RoxB_I$ cluster (CXXCH) with a difference at the third position only (Table 1). In the case of the $RoxB_I$ cluster, the amino acid at this position was substituted only by those amino acids, which were similar in their physicochemical properties viz., polarity and hydropathy, whereas in case of $RoxA_{III}$ and $RoxB_{III}$ sequences, this amino acid was substituted by amino acids different in their physicochemical properties. Further, the heme 2 motif of $RoxA_{III}$ and $RoxB_{III}$ sequences (CAXCH) was different from the corresponding motifs of both the $RoxB_I$ and $RoxA_{II}$ clusters. Similarly, the MauG motif of $RoxA_{III}$ and $RoxB_{III}$ sequences (PYXHNXXVP) was different from those present in $RoxB_I$ and $RoxA_{II}$ cluster sequences. Another important difference lies in the residual positions of F301 and F317 (of RoxA) among the sequences of the three clusters. In the present study, we have found that F301 remains conserved in the sequences of both $RoxA_{II}$ and $RoxB_I$ clusters (Table 1). However, in the sequences of $RoxA_{III}$ and $RoxB_{III}$ cluster, F301 is substituted by other hydrophobic and non-polar amino acids. Further, F317, which remains conserved

Table 1 A list of the functional signatures (motifs/residues) of the RoxA and RoxB orthologs

Signature motif/residue	Function of motif/residue	Equivalent position in RoxA of *S. cummioxidans* 35Y (Birke et al. 2017; Seidel et al. 2013)	Consensus			
			Position in alignment	Amino acid residue ($RoxA_{II}$)	Amino acid residue ($RoxA_{III}$ and $RoxB_{III}$)	Amino acid residue ($RoxB_{I}$)
CXXCH motif 1 or first heme-binding motif	This motif enables a covalent attachment of the apo-peptides of RoxA and RoxB to the heme groups with the help of cytochrome *c* maturation (ccm) proteins, resulting in the mature holo-RoxA and holo-RoxB enzymes (*c*-cytochromes) (Jendrossek and Birke 2019; Jendrossek and Reinhardt 2003; Birke et al. 2013). Besides, the N-terminal heme (heme 1), attached to the first heme-binding motif CXXCH is coordinated by the axial histidine ligand of heme 1	C191	612	C	C	C
		S192	613	S	Xc	Xc
		A193	614	Xc	Xc	Xs
		C194	615	C	C	C
		H195	616	H	H	H
Alanine (A) residue	These residues are involved in the formation of hydrophobic cavity surrounding the putative active site on the distal side of heme (Braaz et al. 2005; Kasai et al. 2017; Seidel et al. 2013)	A251	722	A	A	Xs
Isoleucine (I) residue		I252	723	I	Xc	Xc
Leucine (L) residue		L254	725	Xs	Xc	–
Isoleucine (I) residue		I255	726	Xs	Xc	Xc
Phenylalanine (F) residue		F301	792	F	Xs	F
Alanine (A) residue		A316	807	A	Xc	Xc
Phenylalanine (F) residue	This residue is involved in the binding of oxygen molecule (Birke et al. 2012; Seidel et al. 2013; Kasai et al. 2017)	F317	808	F	Xs	Xs
CXXCH motif 2 or Second heme-binding motif	This motif enables a covalent attachment of the apo-peptides of RoxA and RoxB to the heme groups with the help of cytochrome c maturation (ccm) proteins, resulting in the mature holo-RoxA and holo-RoxB enzymes (c-cytochromes) (Jendrossek and Birke 2019; Jendrossek and Reinhardt 2003; Birke et al. 2013). Besides, the C-terminal heme (heme 2), attached to the second heme-binding motif CXXCH is coordinated by the axial histidine ligand of heme 2	C390	921	C	C	C
		A391	922	A	A	Xc
		S392	923	S	Xc	Xc
		C393	924	C	C	C
		H394	925	H	H	H

Table 1 (continued)

Signature motif/residue	Function of motif/residue	Equivalent position in RoxA of *S. cummioxidans* 35Y (Birke et al. 2017; Seidel et al. 2013)	Consensus			
			Position in alignment	Amino acid residue ($RoxA_{II}$)	Amino acid residue ($RoxA_{III}$ and $RoxB_{III}$)	Amino acid residue ($RoxB_{I}$)
MauG motif	This motif with a conserved histidine residue is the characteristic of a diheme cytochrome and is involved in a specific cytochrome c incorporation/maturation process (Jendrossek and Birke 2019)	P514	1116	P	P	P
		Y515	1117	Xc	Y	Y
		F516	1118	F	Xs	Xs
		H517	1119	H	H	H
		N518	1120	N	N	N
		G519	1121	Xc	Xs	Xs
		S520	1122	S	Xs	Xc
		V521	1123	V	V	Xs
		P522	1124	P	P	P
Histidine (H) residue	The C-terminal heme (heme 2), attached to the second heme-binding motif CXXCH is also coordinated by this axial histidine ligand of heme 2 (Jendrossek and Birke 2019)	H641	1273	H	H	H

It demonstrates the conservation and variation at the important functional sites of amino acid sequences of these orthologs. Xc denotes that an amino acid residue is substituted by another amino acid with different physiochemical properties, viz., polarity and hydropathy; Xs denotes that an amino acid residue is substituted by another amino acid with similar physiochemical properties, viz., polarity and hydropathy. $RoxB_{I}$ represents RoxB orthologs in cluster I, $RoxA_{II}$ represents RoxA orthologs in cluster II, $RoxA_{III}$ represents RoxA orthologs in cluster III, and $RoxB_{III}$ represents RoxB orthologs in cluster III

in the sequences of $RoxA_{II}$ cluster, is substituted by other hydrophobic and non-polar amino acids in the sequences of both $RoxB_{I}$ and $RoxA_{III}$ and $RoxB_{III}$ clusters. Besides, few other signature residues were also found to be different in the $RoxA_{III}$ and $RoxB_{III}$ sequences from those present in $RoxB_{I}$ and $RoxA_{II}$ clusters (Table 1). Similar to the $RoxB_{I}$ sequences, signature motifs and residues of $RoxA_{III}$ and $RoxB_{III}$ sequences were found to harbor less conserved sites in comparison to those of $RoxA_{II}$ sequences. Based on the substitutions at the signature motifs and residues, $RoxA_{III}$ and $RoxB_{III}$ sequences were more similar to $RoxB_{I}$ sequences as compared to $RoxA_{II}$ sequences. Besides, the non-conserved sites of these motifs and residues of the $RoxA_{III}$ and $RoxB_{III}$ sequences were found to have a prevalence of non-synonymous substitutions similar to that present in the $RoxB_{I}$ and $RoxA_{II}$ cluster sequences.

In addition to the conserved structural motifs and cytochrome-specific functions, the RoxA and RoxB enzymes must be extracellular to cleave natural poly (*cis*-1,4-isoprene) into oligo-isoprenoids. In this direction, a comprehensive subcellular localization analysis predicted all the 89 RoxA and RoxB orthologs to be extracellular. These extracellular proteins were predicted to be secreted out of bacterial cells either via classical (~ 30%) or non-classical (~ 70%) pathways of secretion (Table S2). Interestingly, all the $RoxB_{III}$ and $RoxA_{III}$ proteins of the third cluster were found to be secreted only by the non-classical pathway, whereas the classical and non-classical pathways were two optional routes predicted for the proteins of the $RoxB_{I}$ and $RoxA_{II}$ clusters. We further predicted the secretory mechanisms of the rubber oxygenases of *S. cummioxidans* 35Y. The RoxB of *S. cummioxidans* 35Y was found to harbor a signal peptide and a signal peptidase cleavage site and thus it was predicted to be secreted via the Sec-dependent manner of the classical pathways. In contrast, the RoxA was found to lack a signal peptide; however, it possessed a signal peptide cleavage site and a transmembrane helix. This indicated that RoxA might be localized in the inner membrane or periplasm of the bacterial cell. Our further analysis predicted RoxA to be secreted via the non-classical pathway of secretion. However, an earlier study has reported a putative secretion of RoxA via a Sec-dependent manner (Birke et al. 2017). Thus, the secretory mechanisms of RoxA need further investigation.

Evolutionary analyses of the *RoxA* and *RoxB* genes

Among the RoxA and/or RoxB harboring bacteria as shown in clusters I, II, and III of the phylogenetic tree (Fig. 2), we selected only those species and strains whose complete or draft quality genomes were publicly available at the time of analysis. In all genomes, an ortholog was assigned to RoxA or RoxB designation only based on its highest amino acid sequence identity, either with the previously experimentally validated RoxA or RoxB protein sequence of *S. cummioxidans* 35Y. Thus, 37 genomes were selected, including the genome of *S. cummioxidans* 35Y. Among the selected 37 Proteobacteria, 9 and 14 genomes were found to harbor RoxA and RoxB protein-encoding genes, respectively, whereas only 14 genomes were found to possess both these genes (Table S1). In these 14 genomes, the *RoxA* and *RoxB* genes were inparalogs; however, they did not occur in a cluster. Thus, the *RoxA* and *RoxB* genes were individually subjected to HGT, intragenic recombination, recombination rate, mutation rate, gene conversion, and selective pressure analyses to evaluate their propagation and evolution in the phylum Proteobacteria.

Horizontal gene transfer analysis of the *RoxA* and *RoxB* genes

To investigate the occurrence of HGT events, we compared the species trees to the individual gene trees of *RoxA* and *RoxB*. For a given pair of species and gene trees for *RoxA* or *RoxB*, T-REX predicted minimum cost scenario HGTs by a progressive reconciliation of trees using bipartition dissimilarity (BD). The GSI values of 50 or greater were considered statistically significant and indicated support for a clade in the analysis of phylogenetic trees.

For the *RoxA* gene, five HGTs were predicted, including two interclass events in the phylum Proteobacteria (Fig. 3). Bootstrapping value of 100% showed the statistical validation of all the obtained gene transfers. One of the interclass events was between the classes Deltaproteobacteria (donor) and Gammaproteobacteria (acceptor) and the other event was between the classes Gammaproteobacteria (donor) and Betaproteobacteria (acceptor). Interestingly, *S. cummioxidans* 35Y was among the members of the class Gammaproteobacteria, which accepted the *RoxA* gene from the class Deltaproteobacteria through a clade consisting of *Chondromyces apiculatus* DSM 436, *Minicystis rosea* DSM 24000, *Haliangium ochraceum* DSM 14365, and *Nannocystis exedens* ATCC 25963. Further, *S. cummioxidans* 35Y was the only member of the class Deltaproteobacteria, which donated the *RoxA* gene to the class Betaproteobacteria through a clade comprising the members of the genus *Burkholderia*.

For the *RoxB* gene, 12 HGTs were predicted, including five interclass HGTs (Fig. 4). Once again, bootstrap score values of 100% showed strong support for all the predicted gene transfers. Of the five interclass events, three were between the classes Deltaproteobacteria (donor) and Gammaproteobacteria (acceptor) and two were between the classes Gammaproteobacteria (donor) and Betaproteobacteria (acceptor). Similar to the HGT events of *RoxA*, *S. cummioxidans* 35Y was among the members of the class

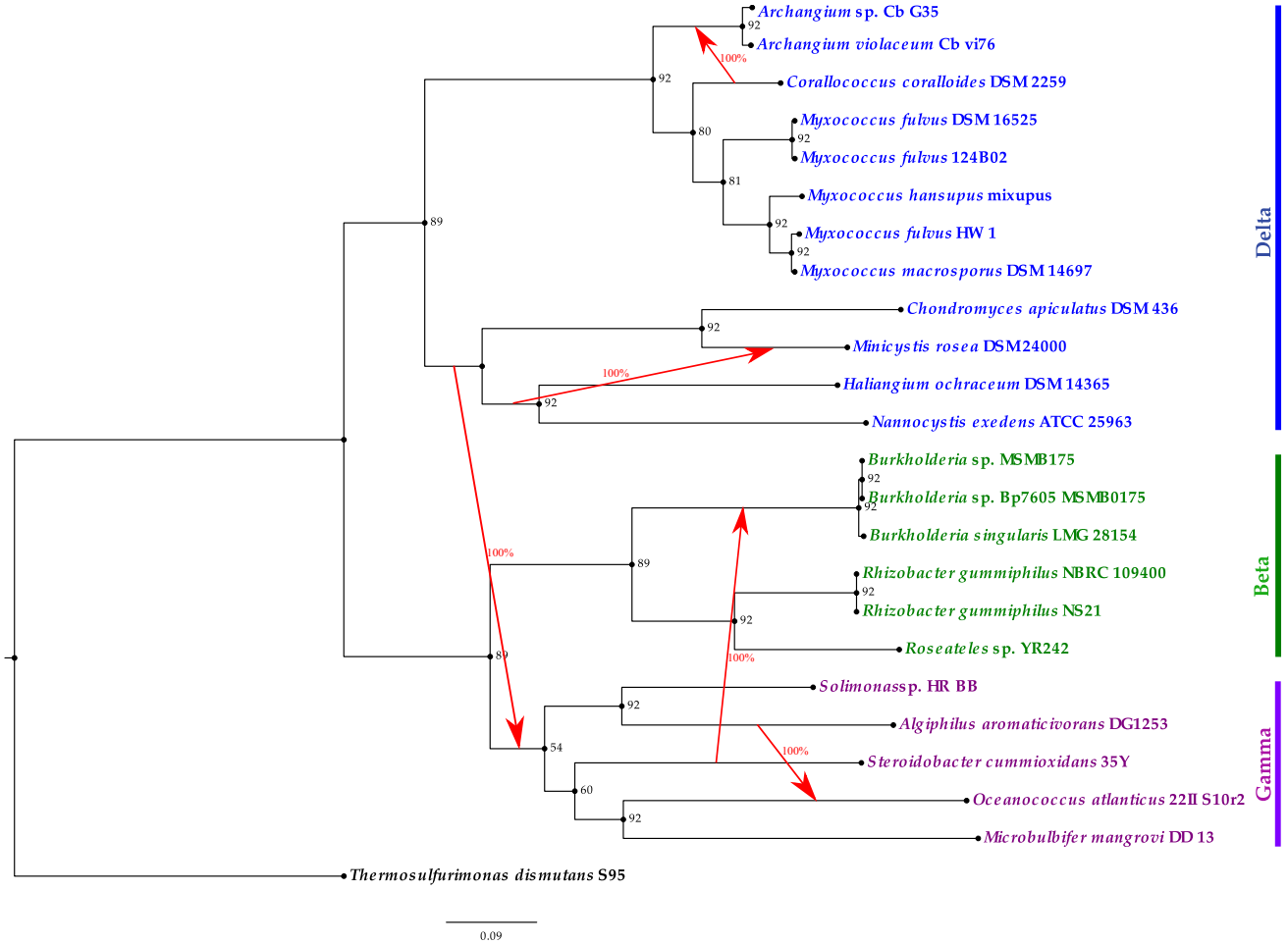


Fig. 3 A species phylogenetic tree demonstrating five horizontal gene transfer events of the *RoxA* gene in the phylum Proteobacteria**.** An Up-to-Date Bacterial Core Gene (UBCG) phylogeneomic tree reconstructed with 92 bacterial core genes was used to infer the species phylogenetic tree. *Thermosulfurimonas dismutans* S95 was used as the outgroup. The tree is formatted using FigTree software. The numbers at nodes indicate the levels of gene support index (GSI); values of less than 50 are not shown. Bar, 0.09 substitutions per site. Arrows indicate the direction of gene transfer, which is inferred using the "Tree and Reticulogram reconstruction server (T-REX)" adopting the bipartition dissimilarity (BD) optimization criterion. HGT bootstrap scores (in percentage) are indicated near the arrow lines of the corresponding HGTs. Delta (blue): class Deltaproteobacteria; beta (green): class Betaproteobacteria; gamma (violet): class Gammaproteobacteria

Gammaproteobacteria, which accepted the *RoxB* gene from the class Deltaproteobacteria. Further, *S. cummioxidans* 35Y and *Nevskia soli* DSM 19509 were the only Gammaproteobacteria which donated the *RoxB* gene to *R. gummiphilus* NS21 and the members of the genus *Burkholderia* of the class Betaproteobacteria, respectively.

Intragenic recombination, recombination rate, mutation rate, and gene conversion analyses of the *RoxA* and *RoxB* genes

For adaptive evolution, the genes accumulated through HGT generally undergo diversification primarily by the process of intragenic recombination. To this end, we found evidence for recombination breakpoints at positions 315, 556, 852, and 1478 in the *RoxA* gene (Fig. S8). The PHI test strongly supported the occurrence of recombination in the *RoxA* gene sequence ($p < 0.05$). The tests for phylogenetic incongruence to the left and right of the respective recombination breakpoints were significant ($p < 0.1$ each). Similarly, we found evidence for the recombination breakpoints at positions 220, 546, 830, 1108, and 1320 in the *RoxB* gene (Fig. S9). Besides, the PHI test also strongly supported the occurrence of recombination in the *RoxB* gene sequence ($p < 0.05$). The tests for phylogenetic incongruence to the left and right of the respective recombination breakpoints were significant for all the breakpoints ($p < 0.1$ each), except for the breakpoint at position 830. These data suggested a notable impact of recombination events on neutrality tests for both *RoxA* and *RoxB* genes. Subsequently, we employed the Tajima's D statistics test to assess the neutrality at the molecular level on the entire *RoxA* and *RoxB* genes (Table S3) as well as

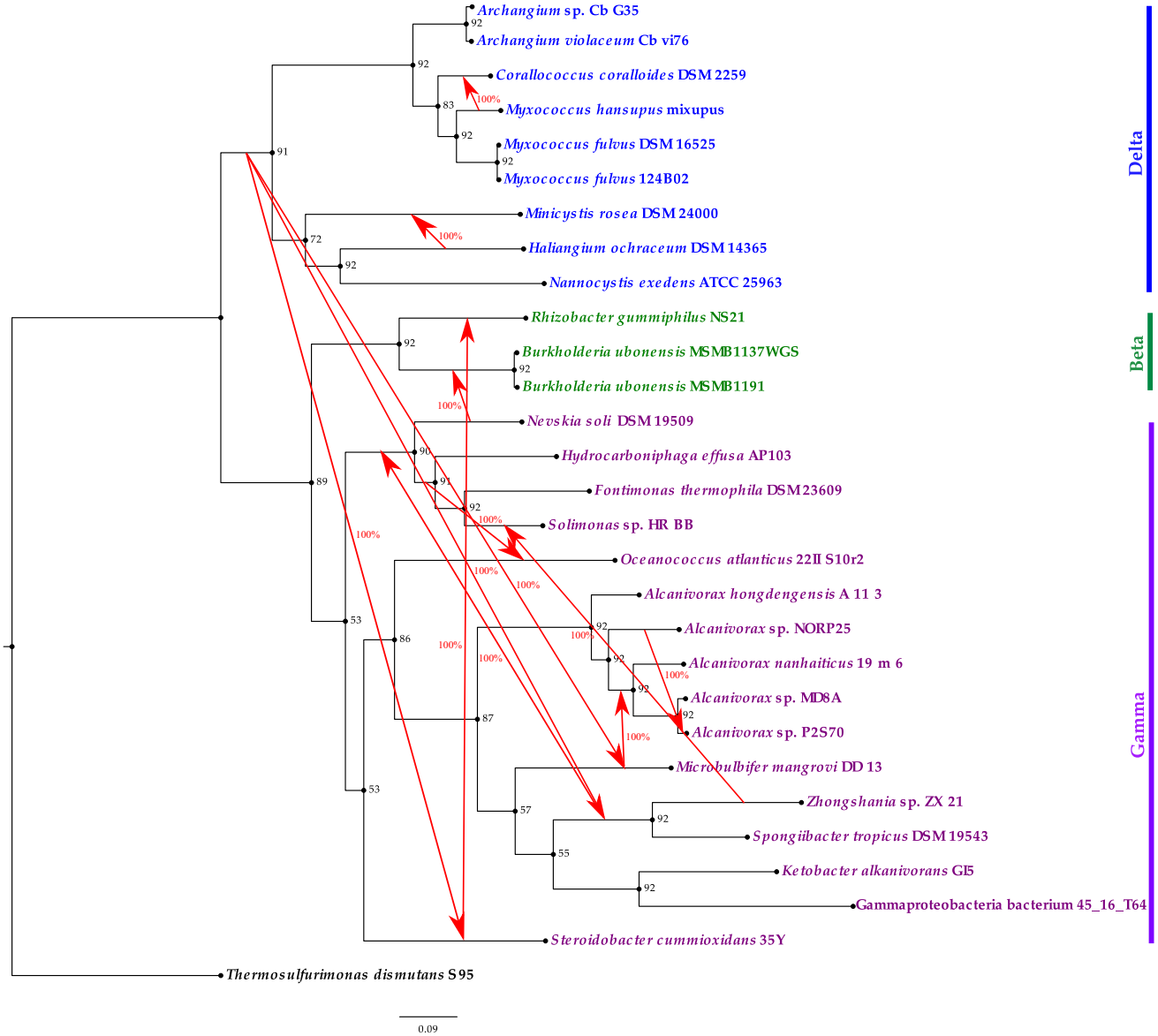


Fig. 4 A species phylogenetic tree demonstrating 12 horizontal gene transfer events of the *RoxB* gene. An Up-to-Date Bacterial Core gene (UBCG) phylogeneomic tree reconstructed with 92 bacterial core genes has been used to infer species phylogeny. *Thermosulfurimonas dismutans* S95 was used as the outgroup. Tree was formatted using FigTree software. Numbers at nodes indicate the levels of gene support index (GSI); values of less than 50 are not shown. Bar, 0.09 substitutions per site. Arrows indicate the direction of gene transfer, which is inferred using the "Tree and Reticulogram reconstruction server (T-REX)" adopting the bipartition dissimilarity (BD) optimization criterion. HGT bootstrap scores (in percentage) are indicated near the arrow lines of the corresponding HGTs. Delta (blue): class Deltaproteobacteria; beta (green): class Betaproteobacteria; gamma (violet): class Gammaproteobacteria

on all the sub-regions of *RoxA* and *RoxB* genes containing structurally and functionally important sites/residues (Tables S3 and S4). The values from the Tajima's D test were significantly positive in all cases which rejects the neutral theory of molecular evolution. This suggests that the *RoxA* and *RoxB* genes and their structurally and functionally important sub-regions evolve under the operation of balancing selection to maintain variation for fitness across heterogeneous environments.

Further, *RoxA* and *RoxB* genes were individually analyzed for the detection of potential recombinant sequences, identification of potential parental sequences, and localization of possible recombination breakpoints. Significant recombination events were detected in ~ 70% (16/23 sequences) and ~ 57% (16/28 sequences) of *RoxA* and *RoxB* gene sequences, respectively, with a high degree of confidence ($p < 0.05$ for more than one recombination detection programs) (Tables S5 and S6, Figs. S8 and S9). We

further predicted five generations of *RoxA* gene with 7, 10, 3, 4, and 1 sequence(s) in the first, second, third, fourth, and fifth generations, respectively (Table S7). Similarly, we predicted four generations of *RoxB* gene with 12, 10, 3, and 1 sequence(s) in the first, second, third, and fourth generations, respectively, whereas the generations of two sequences of *RoxB* gene remained undetermined (Table S8). These results suggest that the sequence diversification of both *RoxA* and *RoxB* genes might be an ongoing process governed by continuous recombination events. The numbers of analyzed sequences of both *RoxA* and *RoxB* genes were, however, too low to trace both the parents of a recombinant sequence, as one parent remained unknown most of the time. In addition, there is also a possibility of misidentification of the potential recombinant sequences, since one of the identified parents itself might be of a recombinant origin (Tables S5 and S6). Thus, a thorough investigation should involve more sequences of both *RoxA* and *RoxB* genes to infer the true recombinants and their parents.

The genetic diversification of a bacterial gene is also attributed to point mutations in addition to the events of recombination. Therefore, we analyzed the relative contribution of recombination and point mutation to the diversification of *RoxA* and *RoxB* genes. The values of the average recombination rate per base pair of *RoxA* and *RoxB* genes were 0.192 and 2.11, respectively, and the values of the average mutation rate per base pair of *RoxA* and *RoxB* genes were 0.09 and 0.05, respectively. This indicated that recombination occurred more frequently than point mutation in both the *RoxA* and *RoxB* genes. This also suggested that the *RoxB* gene underwent more events of recombination and fewer events of point mutations than the *RoxA* gene. Besides recombination and point mutation, gene conversion is another prominent method of diversification. To this end, a total of 247 interparalog gene conversion tracts of various lengths were detected (Table S9) by a previously described method (Betrán et al. 1997). Of these, 86 tracts were present in the *RoxA* gene and 161 in the *RoxB* gene. This indicated the occurrence of gene conversion events between these two types of rubber oxygenases. The inferred gene conversion tracts of the *RoxA* gene were found to have variable lengths (nucleotide distance) ranging from 2 to 771. Similarly, the lengths of gene conversion tracts of the *RoxB* gene were between 2 and 397.

Selective pressure analyses of the *RoxA* and *RoxB* genes

To further evaluate the evolution of *RoxA* and *RoxB* genes, we determined the selective pressures operating on these genes. The selective pressure ($\omega = K_a/K_s$) is measured as the ratio of the number of nonsynonymous substitutions per nonsynonymous site (K_a), to the number of synonymous substitutions per synonymous site (K_s) (Hu and Banzhaf 2008; Hurst 2002; Pond et al. 2005). The ratio of K_a/K_s is expected to exceed unity only if the natural selection promotes changes in the protein sequence (diversifying selection), whereas a ratio of less than unity is expected only if the natural selection suppresses protein changes (purifying selection). An intense positive selective pressure ($\omega = 3.51$) was found to be operating on the *RoxA* gene. This indicated a divergence of the *RoxA* gene by propagating beneficial mutations, which in turn might lead to its adaptive evolution. In contrast, the *RoxB* gene was found to be under an intense negative selective pressure ($\omega = 0.24$), indicating its purifying selection. Further, we determined the selective pressure operating on the structurally and functionally important sites/residues of *RoxA* and *RoxB* genes in terms of d_N/d_S, which is analogous to K_a/K_s. This analysis predicted that the important functional sites/residues of both *RoxA* and *RoxB* genes were under purifying selective pressure (Table S10). The summary of an overall evolution of *RoxA* and *RoxB* genes is given in Fig. 5.

Discussion

In our analysis, we observed that the *RoxA* and *RoxB* genes are restricted only to the Gram-negative members of the beta, gamma, and delta classes of the phylum Proteobacteria. This suggests a taxon-specific evolution of these genes in a given group of bacteria. We found evidence of HGT events of the *RoxA* and *RoxB* genes among the Gram-negative members of the phylum Proteobacteria. These observations indicate HGT as the possible mode of propagation of the *RoxA* and *RoxB* genes in different taxonomic lineages. *RoxA* and *RoxB* genes have originated in the class Deltaproteobacteria and have later propagated to the class Gammaproteobacteria and then to the class Betaproteobacteria via a common interclass route of HGT (Fig. 5). Interestingly, Deltaproteobacteria is also one of the classes to be evolved first under the phylum Proteobacteria, whereas the classes Betaproteobacteria and Gammaproteobacteria have evolved later (Gupta 1998). This suggests that *RoxA* and *RoxB* genes might have originated at the early time points in the evolution of the phylum Proteobacteria.

HGTs are an important source of genome evolution in the prokaryotes (Arber 2014). Once transferred, the genes and the acceptor bacteria continue to evolve, often resulting in the type strain with better traits. Bacterial HGT can involve either complete or partial host gene replacement followed by intragenic recombination (Arber 2014; Boc and Makarenkov 2011). The process of intragenic recombination is well known to play a significant role in the evolution of bacterial genes (Costa et al. 2014; Hagblom et al. 1985). To this end, the *RoxA* and *RoxB* genes in different bacterial hosts were

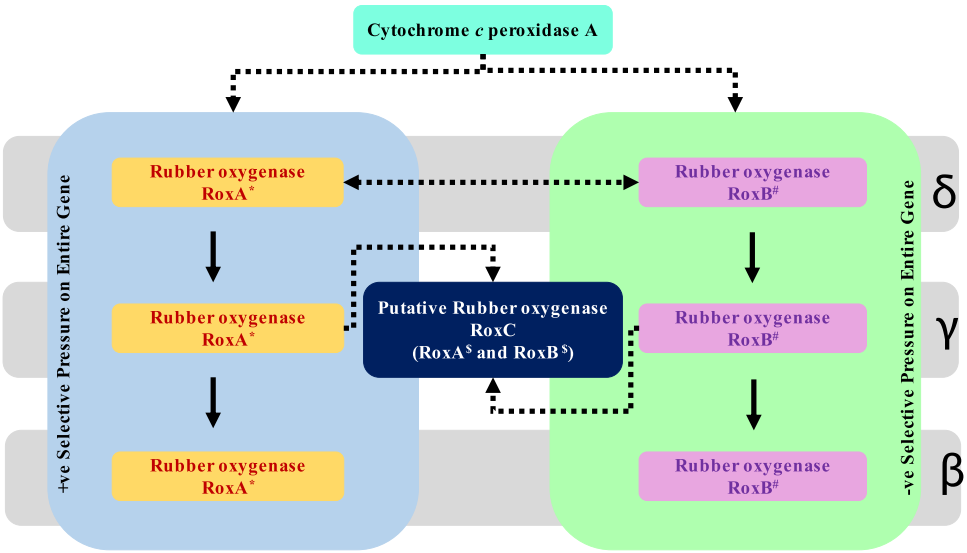


Fig. 5 A cartoon demonstrating the evolution of the *RoxA* and *RoxB* genes in the phylum Proteobacteria. Solid arrows represent the direction of evolution as predicted using in silico methods, whereas dotted arrows represent the predictions hypothesized from our observations made during the analyses. RoxC represents the novel putative rubber oxygenase. β: Betaproteobacteria, δ: Deltaproteobacteria, γ: Gammaproteobacteria. * represents *RoxA* genes belonging to cluster II ($RoxA_{II}$), # represents *RoxB* genes belonging to cluster I ($RoxB_{I}$), and $ represents *RoxA* and *RoxB* genes belonging to cluster III ($RoxA_{III}$ and $RoxB_{III}$) and are termed as putative *RoxC* gene. The important structural and functional sites/residues of both *RoxA* and *RoxB* genes are under the operation of purifying selective pressure. Besides, these genes have also evolved under the effect of intragenic recombination, gene conversion, and point mutation

found to undergo several events of intragenic recombination with *RoxB* consisting of more recombination breakpoints than *RoxA*. The different events of recombination allow paralogs to diverge in their functions, which in turn facilitates their separation into independently evolving gene lineages (Francino 2012). The evolution of a bacterial gene is not only carried out by recombination events, but point mutations also play a major role in this process. Interestingly, the recombinations were found to occur more frequently than point mutations in both the *RoxA* and *RoxB* genes. However, the rates of recombinations and point mutations are different for both these genes. The *RoxB* gene demonstrated more events of recombination and fewer events of point mutation than the *RoxA* gene. Interestingly, this observation is similar to those in the case of HGT and intragenic recombination analyses, which predicted more HGT events and breakpoints for the *RoxB* than for the *RoxA* gene. These results suggest that the genetic diversification of the *RoxA* and *RoxB* genes might have been primarily driven by recombination with the *RoxB* gene being highly sensitive to HGT and recombination events. In addition, gene conversion tracts between the *RoxA* and *RoxB* genes suggest the occurrence of intergenic conversion events. This is consistent with the fact that intergenic conversion occurs between paralogs, which possess sufficient sequence homology (Mansai et al. 2011). Such events of gene conversion promote co-evolution in gene families and maintain similarities between repeated genes (Paulsson et al. 2017).

HGT is one of the driving forces behind the events of speciation in a given environment (Zhang et al. 2016). The evolutionary propagation of the *RoxA* and *RoxB* genes in different bacterial hosts is in corroboration with the fact that new species or strains with beneficial or adaptive combinations are likely to compete better than their neighbors and so proliferate (Nielsen et al. 2014; Van Elsas et al. 2003). The HGT of the *RoxA* and *RoxB* genes in new strains, followed by natural selection of these phenotypes might have played a central role in the evolution of the NR-degradation potential of bacteria. In the absence of selective pressure in a given environment, an HGT event can be lost by simple genetic drift (Perry and Wright 2014). This supports our observation that the evolutionarily emerged populations of the *RoxA* and *RoxB* genes have been under the operation of selective pressure, which prevents their loss during evolution. Interestingly, the *RoxA* and *RoxB* occur as paralogous genes, which tend to evolve different functions under differential selective pressures (Ho-Huu et al. 2012). The differential selective pressures determine whether the underlying mutations are beneficial or detrimental in a given niche at a given time which leads to the evolution of different variants of a given gene (Van Elsas et al. 2003). Being consistent with this fact, the *RoxA* and *RoxB* genes have indeed evolved different functions which indicate the operation of differential selective pressures on these duplicated gene copies. This is in corroboration with earlier studies which have demonstrated contrasted patterns of selective pressure on

the paralogous gene pairs (Ho-Huu et al. 2012; Van Zee et al. 2016). Toward this, the *RoxA* gene is found to be under an intense positive selective pressure and the *RoxB* gene is under an intense negative selective pressure. Thus, the *RoxA* gene is diversifying further, which might lead to the evolution of its new variants, whereas the *RoxB* gene undergoes purifying selection by eliminating its deleterious mutations which might lead to the retention of its best-fit variants. The operations of differential selective pressures on the *RoxA* and *RoxB* genes also suggest their independent evolution. Further, in corroboration with a previous study, which showed a distant relationship between the *RoxA* and *CcpA* genes (Seidel et al. 2013), we propose that the *CcpA* gene has evolved into two paralogous genes, viz., *RoxA* and *RoxB*. These genes have further propagated in various bacterial lineages via the process of HGT and continue to evolve under differential selective pressures (*RoxA* is diversifying, whereas *RoxB* is purifying) in their respective bacterial hosts.

The evolution of the RoxA and RoxB proteins under differential selective pressures is also reflected in terms of the types of substitutions occurring in their primary amino acid sequences. A diversifying selective pressure on the entire *RoxA* gene reflects the occurrence of more non-synonymous substitutions throughout the gene. However, an operation of purifying selective pressure on the structurally and functionally important sites/residues indicates their high conservation. This suggests that these sites either might not be changing or might be accumulating more synonymous substitutions. This is in corroboration with earlier studies where positively selected surface-exposed variable proteins of *Neisseria* were found to harbor certain conserved sites under the operation of purifying selection (Wachter and Hill 2016). Our analysis also indicates that although the *RoxA* gene undergoes a diversifying selection, still it tends to conserve its signature sites/residues to retain its functionality. In contrast, a purifying selection of the entire *RoxB* gene indicates the occurrence of more synonymous substitutions throughout the gene. Further, both synonymous and non-synonymous substitutions are found to occur at its functional signature sites/residues at similar frequencies. However, an operation of purifying selective pressure on the structurally and functionally important amino acid sites/residues of *RoxB* genes indicates their conservation. It is important to note that the *RoxA* and *RoxB* are paralogous genes which have evolved different functions. These observations suggest that despite a purifying selection of the *RoxB* gene, it accumulates few non-synonymous changes at its signature sites/residues, which might cause its structural and functional diversification from the *RoxA* gene. Thus, the operation of differential selective pressures on *RoxA* and *RoxB* genes might be the major driving force behind their diversification.

An independent evolution of the RoxB and RoxA proteins is suggested by their occurrence in two separate clusters (clusters I and II) in the putative rubber oxygenases-derived phylogenetic tree. This further supports an evolutionary divergence of these proteins. The phylogenetic separation of rubber oxygenases in a third distinct cluster (cluster III) indicates the evolution of a novel rubber oxygenase in Gram-negative bacteria. Based on the significant amino acid sequence identity with the RoxA or RoxB of *S. cummioxidans* 35Y, clusters I and II belong to RoxB and RoxA orthologs, respectively, whereas cluster III possesses both RoxA and RoxB orthologs. RoxB sequences of cluster I and RoxA sequences of cluster II are distributed in delta, beta, and gamma classes of the phylum Proteobacteria, whereas RoxA and RoxB of cluster III are restricted only to the class Gammaproteobacteria. This suggests that the continuous evolution of the RoxA and RoxB proteins might have led to the emergence of their variant in the class Gammaproteobacteria. Despite the significant amino acid sequence identity with the RoxA or RoxB of *S. cummioxidans* 35Y, rubber oxygenases of cluster III possess variations in some of their signature motifs and residues. For example, rubber oxygenases of cluster III possess significant differences at the residual positions of F301 and F317 with respect to the sequences of the clusters I and II. It has been shown that F301 and F317 are important residues which are located in close vicinity in the 3D space to the heme site at the N terminal (heme 1 motif) and are critical for its stable binding to dioxygen molecule and correct positioning to cleave poly(*cis*-1,4-isoprene) (Schmitt et al. 2019). Besides, it has also been demonstrated in a previous study on RoxA that variations at F301 and F317 lead to destabilization of dioxygen coordination, resulting in poor or missing catalytic activity. To this end, F301, which remains conserved in the sequences of both clusters I and II, has undergone synonymous substitution in the sequences of cluster III. The F317, which is conserved in the sequences of cluster II, has also undergone synonymous substitution in the sequences of clusters I and III. These variations in the sequences of the third cluster can lead to significant structural alterations in the encoded proteins and can alter their catalytic activities. Based on the substitutions at the signature motifs and residues, rubber oxygenases of cluster III are more similar to RoxB sequences of cluster I than to those RoxA sequences of cluster II. Further, all the rubber oxygenases of cluster III have been found to be secreted only by the non-classical pathway, whereas the classical and non-classical pathways are two optional routes predicted for the rubber oxygenases of clusters I and II. This makes rubber oxygenases of cluster III different from those of clusters I and II in their mode of secretion. Thus, rubber oxygenases of cluster III possess few similarities and dissimilarities to both the RoxB (cluster I) and RoxA (cluster II). The distinct features of rubber oxygenases of cluster III suggest them as a novel protein; hence, we have termed it as a putative RoxC. The catalytic activities

of RoxA and RoxB have been previously demonstrated (Jendrossek and Birke 2019); however that of putative RoxC protein needs to be experimentally validated.

Conclusions

The predicted Rox proteins are restricted to only the Gram-negative members of the beta, gamma, and delta classes of the phylum Proteobacteria. Toward the evolutionary propagation, both *RoxA* and *RoxB* genes might have originated in the class Deltaproteobacteria and have horizontally transferred to the class Gammaproteobacteria and then to the class Betaproteobacteria. Further, both *RoxA* and *RoxB* genes are predicted to have evolved under differential selective pressures, leading to their structural and functional diversification from each other. Besides, the diversification of *RoxA* and *RoxB* genes is also attributed to the evolutionary processes of recombination, point mutation, and gene conversion. The continuous evolution of Rox proteins, viz., RoxA and RoxB, might have led to the emergence of their variant as another type of putative rubber oxygenase (RoxC) in the class Gammaproteobacteria. Interestingly, putative RoxC sequences have significant sequence homology with either of the two rubber oxygenases (RoxA and RoxB) of *S. cummioxidans* 35Y. Thus, it needs further experimental investigation to verify the putative function of RoxC as a novel rubber oxygenase. Further, all the rubber oxygenases are predicted to be secreted out of the bacterial cells, either via a classical or non-classical pathway suggesting the conservation of the extracellular nature of these enzymes. Overall, the present study identifies the rubber oxygenases in the Gram-negative bacteria, followed by the comprehensive exploration of their evolutionary attributes. We anticipate that in silico results of the current study will be useful for researchers to experimentally verify the predictions in the potential Gram-negative rubber-degrading species.

Acknowledgements TP acknowledges IIT Mandi for financial and technical support. VS acknowledges the Ministry of Human Resource Development (MHRD), India, for providing the research fellowship.

Author contributions TP and VS conceived or designed the study, TP and VS performed research, TP, VS, and FM analyzed data, TP and VS wrote the paper.

Funding This research received no external funding.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Arber W (2014) Horizontal gene transfer among bacteria and its role in biological evolution. Life (Basel) 4:217–224

Bendtsen JD, Kiemer L, Fausbøll A, Brunak S (2005) Non-classical protein secretion in bacteria. BMC Microbiol 5:58

Betrán E, Rozas J, Navarro A, Barbadilla A (1997) The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics 146:89–99

Birke J, Jendrossek D (2014) Rubber oxygenase (RoxA) and latex clearing protein (Lcp) cleave rubber to different products and use different cleavage mechanisms. Appl Environ Microbiol 80:5012–5020

Birke J, Hambsch N, Schmitt G, Altenbuchner J, Jendrossek D (2012) Phe317 is essential for rubber oxygenase RoxA activity. Appl Environ Microbiol 78:7876–7883

Birke J, Röther W, Schmitt G, Jendrossek D (2013) Functional identification of rubber oxygenase (RoxA) in soil and marine myxobacteria. Appl Environ Microbiol 79:6391–6399

Birke J, Röther W, Jendrossek D (2017) RoxB is a novel type of rubber oxygenase that combines properties of rubber oxygenase RoxA and latex clearing protein (Lcp). Appl Environ Microbiol 83:e00721–e1717

Birke J, Röther W, Jendrossek D (2018) *Rhizobacter gummiphilus* NS21 has two rubber oxygenases (RoxA and RoxB) acting synergistically in rubber utilisation. Appl Microbiol Biotechnol 102:10245–10257

Boc A, Makarenkov VJ (2011) Towards an accurate identification of mosaic genes and partial horizontal gene transfers. Nucleic Acids Res 39:e144–e144

Boc A, Philippe H, Makarenkov V (2010) Inferring and validating horizontal gene transfer events using bipartition dissimilarity. Syst Biol 59:195–211

Boc A, Diallo AB, Makarenkov V (2012) T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res 40:W573–W579

Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176:1035–1047

Braaz R, Armbruster W, Jendrossek D (2005) Heme-dependent rubber oxygenase RoxA of *Xanthomonas* sp. cleaves the carbon backbone of poly (cis-1, 4-isoprene) by a dioxygenase mechanism. Apwbiol 71:2473–2478

Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665–2681

Bryant J, Chewapreecha C, Bentley SD (2012) Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. Future Microbiol 7:1283–1296

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421

Costa J, Teixeira PG, d'Avó AF, Júnior CS, Veríssimo A (2014) Intragenic recombination has a critical role on the evolution of *Legionella pneumophila* virulence-related effector *sidJ*. PLoS One 9:e109840

Darmon E, Leach DR (2014) Bacterial genome instability. Microbiol Mol Biol Rev 78:1–39

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM et al (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res 39:W13–W17

Durbin R, Eddy SR, Krogh A, Mitchison G (1998) Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge University Press, Cambridge

Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Francino MP (2012) The ecology of bacterial genes and the survival of the new. Int J Evol Biol 2012:394026

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16:573–582

Grove A (2010) Functional evolution of bacterial histone-like HU proteins. Curr Issues Mol Biol 13:1–12

Gruen DS, Wolfe JM, Fournier GP (2019) Paleozoic diversification of terrestrial chitin-degrading bacterial lineages. BMC Evol Biol 19:34

Gupta RS (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62:1435–1491

Hagblom P, Segal E, Billyard E, So M (1985) Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. Nature 315:156–158

Ho-Huu J, Ronfort J, De Mita S, Bataillon T, Hochu I, Weber A et al (2012) Contrasted patterns of selective pressure in three recent paralogous gene pairs in the *Medicago* genus (L.). BMC Evol Biol 12:195

Hu T, Banzhaf W (2008) Nonsynonymous to Synonymous Substitution Ratio Ka/Ks: measurement for rate of evolution in evolutionary computation. In: Proceedings of 10th International Conference on Parallel Problem Solving from Nature (PPSN X). Springer-Verlag GmbH, Berlin, Germany, pp 448–457.

Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18:486–487

Huson DH, Bryant D (2005) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119

Jendrossek D, Birke J (2019) Rubber oxygenases. Appl Microbiol Biotechnol 103:125–142

Jendrossek D, Reinhardt S (2003) Sequence analysis of a gene product synthesized by *Xanthomonas* sp. during growth on natural rubber latex. FEMS Microbiol Lett 224:61–65

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240

Juárez-Vázquez AL, Edirisinghe JN, Verduzco-Castro EA, Michalska K, Wu C, Noda-García L, Babnigg G, Endres M, Medina-Ruíz S, Santoyo-Flores J, Carrillo-Tripp M, Ton-That H, Joachimiak A, Henry CS, Barona-Gómez F (2017) Evolution of substrate specificity in a retained enzyme driven by gene loss. Elife 6:e22679

Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662

Kahsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. Bioinformatics 21:1853–1858

Kasai D, Imai S, Asano S, Tabata M, Iijima S, Kamimura N et al (2017) Identification of natural rubber degradation gene in *Rhizobacter gummiphilus* NS21. Biosci Biotechnol Biochem 81:614–620

Katoh K, Standley DM (2013) MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Mansai SP, Kado T, Innan H (2011) The rate and tract length of gene conversion between duplicated genes. Genes 2:313–331

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563

Martin D, Posada D, Crandall K, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retrov 21:98–102

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003

McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160:1231–1241

Na S-I, Kim YO, Yoon S-H, Ha S-m, Baek I, Chun J (2018) UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J Microbiol 56:281–285

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nielsen KM, Bøhn T, Townsend JP (2014) Detecting rare gene transfer events in bacterial populations. Front Microbiol 4:415

Noor S, Changey F, Oakeshott JG, Scott C, Martin-Laurent F (2014) Ongoing functional evolution of the bacterial atrazine chlorohydrolase AtzA. Biodegradation 25:21–30

Okonechnikov K, Golosova O, Fursov M, Team U (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225

Paulsson J, El Karoui M, Lindell M, Hughes D (2017) The processive kinetics of gene conversion in bacteria. Mol Microbiol 104:752–760

Perry AJ, Ho BK (2013) Inmembrane, a bioinformatic workflow for annotation of bacterial cell-surface proteomes. Source Code Biol Med 8:9

Perry JA, Wright GD (2014) Forces shaping the antibiotic resistome. BioEssays 36:1179–1184

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786

Pond SLK, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533

Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol Biol Evol 19:708–717

Price MN, Dehal PS, Arkin AP (2010) FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490

Rambaut A (2014) FigTree v1.4.2, A graphical viewer of phylogenetic trees. https://tree.bio.ed.ac.uk/software/figtree Accessed 8 March 2018.

Robinson DR, Foulds LR (1981) Comparison of phylogenetic trees. Math Biosci 53:131–147

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schmitt G, Birke J, Jendrossek D (2019) Towards the understanding of the enzymatic cleavage of polyisoprene by the dihaem-dioxygenase RoxA. AMB Express 9:166

Seidel J, Schmitt G, Hoffmann M, Jendrossek D, Einsle O (2013) Structure of the processive rubber oxygenase RoxA from *Xanthomonas* sp. Proc Natl Acad Sci USA 110:13833–13838

Shapiro BJ, Alm EJ (2008) Comparing patterns of natural selection across species using selective signatures. PLoS Genet 4:e23

Sharma V, Siedenburg G, Birke J, Mobeen F, Jendrossek D, Prakash T (2018) Metabolic and taxonomic insights into the Gram-negative natural rubber degrading bacterium *Steroidobacter cummioxidans* sp. nov., strain 35Y. PLoS One 13:0197448

Smith JM (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101:11030–11035

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tsuchii A, Takeda K (1990) Rubber-degrading enzyme from a bacterial culture. Appl Environ Microbiol 56:269–274

Van Elsas JD, Turner S, Bailey MJ (2003) Horizontal gene transfer in the phytosphere. New Phytol 157:525–537

Van Zee JP, Schlueter JA, Schlueter S, Dixon P, Sierra CA, Hill CA (2016) Paralog analyses reveal gene duplication events and genes under positive selection in *Ixodes scapularis* and other ixodid ticks. BMC Genomics 17:241

Wachter J, Hill S (2016) Positive selection pressure drives variation on the surface-exposed variable proteins of the pathogenic *Neisseria*. PLoS One 11:e0161348

Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom Proteom Bioinf 4:259–263

Zhang YZ, Li Y, Xie BB, Chen XL, Yao QQ, Zhang XY et al (2016) Nascent genomic evolution and allopatric speciation of *Myroides profundi* D25 in its transition from Land to Ocean. mBio 7:e01946–e11915

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166