Identification of potential genes associated with magnetosome biogenesis in the CtrA regulon in *Magnetospirillum magneticum* AMB-1

Yizi Yang

Nanjing Medical University

Ning Gu (✉ guning@seu.edu.cn)

Southeast University <https://orcid.org/0000-0003-0047-337X>

Research Article

Keywords: magnetotactic bacteria, magnetosome, magnetosome-related gene, network, microarray bioinformatics, CtrA regulon

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1231040/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Magnetotactic bacteria (MTB) are worth studying because of magnetosome biomineralization. Magnetosome biogenesis in MTB is controlled by multiple genes known as magnetosome-related genes. Recent advances in microarray bioinformatics provide a unique opportunity for studying functions of magnetosome-related genes and networks that they are involved in. Furthermore, microarray analysis can also help identify genes potentially associated with magnetosome biogenesis. To find more magnetosome-related genes in the CtrA regulon, we analyzed expression data of 8 samples of *Magnetospirillum magneticum* AMB-1 in the GSE35625 dataset in NCBI GEO. We identified 10 potential magnetosome-related genes using a nearest neighbor search strategy on both protein-protein interaction and gene co-expression networks. We also discovered and compared two co-expression modules which most known magnetosome-related genes belong to. These analyses provided several clues for the importance of energy on regulating co-expression module structures. In addition, we hypothesized that many magnetosome-related genes may develop new abilities to interact with other biomolecules including their functions in magnetosome biogenesis during evolution, since they are not hub genes in networks. Overall, our study provides useful information for researches on magnetotactic bacteria and magnetosome.

Introduction

Magnetosomes are specialized organelles containing magnetic crystals of magnetite or greigite, and help host bacteria orient in magnetic fields. Many bacteria are able to produce magnetosomes, and consequently they are called magnetotactic bacteria (MTB). These bacteria are polyphyletic, and belong to multiple phyla, such as *Proteobacteria* and *Nitrospirae* (Lin et al. 2018). MTB display a stunning diversity in cell morphology and physiology. They can be found in aquatic environments across all over the world. Different MTB species can have different magnetosome crystals. These magnetosomes display a variety of shapes, sizes and compositions across different bacteria.

Magnetosome formation is controlled by many different genes in MTB. More than forty genes have been found to be responsible for magnetosome biogenesis (Uebe and Schuler 2016). The number of genes responsible for magnetosome biomineralization can be changed when magnetosomes are synthesized by different bacteria. A group of *mam* genes, given its name based on encoding protein that are associated with the magnetosome membrane, are found to be important in magnetosome biogenesis (Frankel and Bazylinski 2006). According to previous studies, nine genes (*mamA*, *-B*, *-E*, *-K*, *-M*, *-O*, *-P*, *-Q*, *-I*) are present in all genomes of magnetotactic bacteria that have been studied to date (Uebe and Schuler 2016). Moreover, the *mamL* gene is additionally present in all magnetotactic bacteria that produce only magnetite crystals (Lefevre et al. 2013). Some of these genes are crucial for magnetosome biogenesis. A deletion of *mamB* caused the most severe aberrant magnetosome membrane phenotype (Murat et al. 2010; Raschdorf et al. 2016; Uebe et al. 2011). Other genes such as *mamE* and *mamM* are also essential for magnetosome formation. These genes form a group of essential genes (*mamB*, *-E*, *-M*, *-O*, *-Q*, *-L*) for magnetite (Lohsse et al. 2014; Taoka and Fukumori 2018; Uebe and Schuler 2016). By contrast, several

genes are not essential in magnetosome formation, such as *mamJ* and *mamK*. These genes are not absolutely necessary for magnetosome formation, and therefore they are called non-essential genes.

Proteins encoded by essential genes and non-essential genes work together to help produce magnetosome in bacteria. However, a complete list of all essential and non-essential genes has not been obtained for several reasons. Firstly, the number of the genes related to magnetosome biogenesis varies among bacteria. Some genes, such as *limE*, *limJ* and *limO* are found uniquely in *Magnetospirillum magneticum* AMB-1, whereas several novel genes, for instance *man* genes, are detected in the phylum *nitrospirae* (Lin et al. 2014). Secondly, functions of the same genes are changed across multiple bacteria, which were confirmed by gene deletion experiments. The phenotypes of the AMB *mamK*-deletion were much weaker than that of the MSR *mamK*-deletion mutant (Uebe and Schuler 2016). Thirdly, although several hypothesized mechanisms of magnetosome formation were proposed by various researchers in recent years, detailed molecular mechanisms of magnetosome biogenesis are still not known. There are many knowledge gaps required to be filled before getting a list of all magnetosome-related genes.

Advances in microarray-based bioinformatic analysis may help understand the complex mechanisms of magnetosome biogenesis. These analyses can screen differential expressed genes (DEGs) and elucidate the mechanism of a specific type of cancer, such as esophageal squamous cell carcinoma (Deng and Wu 2018). Similarly, these analyses can also help reveal the mechanisms of magnetosome biogenesis in MTB. Multiple DEGs can be identified by comparing the gene expression profiles of normal and mutated bacteria. This analysis was used to identify the genes related to the CtrA pathway in *Magnetospirillum*. and consequently, revealed that genes in the magnetosome island are among those in the CtrA regulon in AMB-1 (Greene et al. 2012). Therefore, it is possible to identify DEGs associated with magnetosome formation in the CtrA regulon when comparing gene expression profiles of wildtype bacteria and *ctrA* deletion strains.

Magnetosome biogenesis entails multiple processes and molecular mechanisms. It requires network and pathway analyses of identified DEGs to gain a better understanding of underlying mechanisms and molecular events in MTB. Some tools are designed for these analyses, including STRING v11.0 for protein-protein interaction (PPI) networks construction (Szklarczyk et al. 2019), Cytoscape for network import and illustration (Shannon et al. 2003), WGCNA for description of correlation patterns among genes cross microarray samples and modular construction of highly correlated DEGs (Langfelder and Horvath 2008), and etc.. With the help of these tools, a more comprehensive understanding of magnetosome biogenesis might be obtained.

Materials And Methods

Microarray data, sequence resources, and gene lists

We downloaded the gene expression profiles data GSE35625 of *Magnetospirillum magneticum* AMB-1 (Greene et al. 2012) from the GEO database using two Bioconductor packages, i.e. Biobase (Gentleman et al. 2005) and GEOquery (Davis and Meltzer 2007). The dataset of GSE35625 contained 12 mutated

samples and 4 wild-type samples, while 4 of 12 mutated samples belonged to the *ctrA* deletion strain or the deletion strain complemented with empty vector. The expression matrix, sample treatments, platform annotations were also downloaded from the GEO database. Furthermore, all genes and encoded proteins in the genome of *Magnetospirillum magneticum* AMB-1 were downloaded from the NCBI genome database. Additionally, 2 lists of known magnetosome-related genes were obtained from the article entitled 'Magnetosome biogenesis in magnetotactic bacteria' (Uebe and Schuler 2016) and the book chapter entitled 'Molecular Mechanism of Magnetic Crystal Formation in Magnetotactic Bacteria' (Arakaki et al. 2018).

Identification of DEGs in mutated bacteria

The gene expression profiles data were analyzed using the R packages Limma (Smyth 2005), ggplot2 (Wickham 2011), reshape2 (Wickham 2012) and pheatmap (Kolde and Kolde 2015) to identify DEGs related to magnetosome in the CtrA regulon. The identification process of DEGs was similar to GEO2R method (Barrett et al. 2013), but the difference lies in that the average of expression values of the same GENE_NAME in the expression matrix was calculated to detect DEGs in our study. Unannotated ORFs detected in DEG calling such as 'LD8 1061118-1061810' were removed from the list of DEGs, since there is no information stored in the database such as STRING for these ORFs. A group.xls file was constructed for the wild-type group (WT) and the mutated group (the *ctrA* deletion). A FDR<0.05 and a |logFC| >1 were set as threshold of screening for DEGs and identifying common DEGs both in wild-type MTB and in a mutated sample without involving the *ctrA* pathway. Identified DEGs were compared with lists of known magnetosome-related genes.

GO and KEGG pathway enrichment analysis

Gene ontology and KEGG pathway enrichment analysis of identified DEGs was performed using the Gene Set Enrichment Analysis for prokaryotes (GSEA-Pro v3) web tool (Carvalho et al. 2017). GSEA-Pro auto detected threshold values of ratio data on the basis of GO classification. Other results, such as COG, EggNOG and KEGG were also obtained using GSEA-Pro with default setting. By comparing these results, a list of functional similar gene was obtained for known magnetosome-related genes.

PPI network and cluster analysis

DEGs' protein sequences were obtained from the downloaded files from NCBI. PPI networks of DEGs in *Magnetospirillum magneticum* AMB-1 were investigated using STRING with a combined confidence score of > 0.4. Clusters analysis of PPI networks was performed using the MCODE Cytoscape plugin (Bader and Hogue 2003). The interaction relationship table were imported into the Cytoscape software to visualize these PPI networks. Essential genes were identified using the CytoHubba Cytoscape plugin (Chin et al. 2014). Six supported methods including MCC, DMNC, Closeness, Radiality, Betweenness, and Stress were applied to this study. We believed that potential magnetosome-related genes rank closely with known magnetosome-related genes in ranking lists. After searching 20 nearest neighbors of magnetosome-related genes in ranking for essential nodes, we extracted 6 lists of nearest neighbors, and more than 20

neighbors of magnetosome-related genes were identified if multiple neighbors have same rank. A list of common neighbors was obtained when different analysis results were compared to each other. We compared two lists of neighbors and functional similar genes to find potential members of magnetosome-related genes.

WGCNA and functional enrichment analysis in detected modules

Co-expression networks of identified DEGs were constructed using the WGCNA package. A Gene_expressionMatrix.xls file was created when calling DEGs. We extracted DEGs' expression data across all samples in GSE35625, and used these data to create WGCNA modules. No trait-related analysis was performed, since no sample in GSE35625 had experimental trait data. The soft-thresholding power value is one of the most important parameters, for the reason that it is able to affect the scale-free fit index and the average connectivity degree of the modules constructed. We used the pickSoftThreshold function to help pick an appropriate power for network construction. We set the minModuleSize parameter to 30. Key modules were identified based on correlations among modules. The genes in these modules were submitted to GSEA-Pro for GO and KEGG pathway enrichment analysis. Key modules were also imported into Cytoscape for further visualization and analysis. Same six methods (MCC, DMNC, Closeness, Radiality, Betweenness, and Stress) were used to find hub nodes in WGCNA modules. Nearest neighbors of known magnetosome-related genes in the CtrA regulon were searched using the aforementioned strategy in gene neighbor search in the PPI network.

Results

Identification of DEGs in mutated bacteria

Using several R packages and setting $FDR < 0.05$ and a $|\log FC| > 1$, 806 DEGs were screened between the wild-type and *ctrA* deleted bacteria in GSE35625, including 428 up-regulated genes and 378 down-regulated genes. A density plot and a volcano plot were shown in Fig1. Additionally, 6 Unannotated ORFs were also screened. A deg.txt file of identified DEGs was created using R and can be found in Supplementary data FileS1.

(Fig1 approximately here)

GO and KEGG pathway enrichment analysis

759 DEGs were successfully converted to standard locus tags, and the rest were removed due to sequence redundancy in 2015 in the NCBI databases. All 759 DEGs excluding 6 unannotated ORFs were submitted to GSEA-Pro v3 to perform the GO and KEGG pathway enrichment analysis. There were three parts of the GO enrichment analysis: biological process (BP), cellular component (CC) and molecular function (MF). In general, the DEGs were strongly enriched in diverse BPs, such as translation (GO:0006412), protein repair (GO:0030091), ribosome assembly (GO:0042255), cellular respiration (GO:0045333), chemotaxis (GO:0006935), pathogenesis (GO:0009405), and bacterial-type flagellum-

dependent cell motility (GO:0071913). The DEGs were strongly enriched in multiple CCs, such as cytoplasm (GO:0005737), large ribosomal subunit (GO:0015934), plasma membrane respiratory chain complex 1 (GO:0045272), and bacterial-type flagellum basal body (GO:0009425). The DEGs were also strongly enriched in many MFs, for instance rRNA binding (GO:0019843), endoribonuclease activity (GO:0004521), translation initiation factor activity (GO:0003743), structural constituent of ribosome (GO:0003735), ATP binding (GO:0005524), metal ion binding (GO:0046872), motor activity (GO:0003774), and etc..

We also submitted up-regulated and down-regulated genes respectively to GSEA-Pro v3. The up-regulated DEGs were strongly enriched in most of these aforementioned terms, such as translation (GO:0006412), ribosome (GO:0005840), and rRNA binding (GO:0019843). Moreover, the up-regulated DEGs were also particularly enriched in both energy conversion and molecule biosynthesis plus various enzyme activities, for instance UDP-glucose metabolic process (GO: 0006011), ribosome biogenesis (GO:0042254), undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase activity (GO:0099621), and UTP:glucose-1-phosphate uridylyltransferase activity (GO:0003983).

By contrast, the down-regulated DEGs were particularly enriched in several biological processes and transfer or transportation functions, such as DNA replication (GO:0006260), sodium ion transport (GO: 0006814), acetate transmembrane transporter activity (GO:0015123), glycolate transmembrane transporter activity (GO:0043879), 3-deoxy-manno-octulosonate cytidylyltransferase activity (GO:0008690). The down-regulated DEGs were also enriched in plasma membrane (GO:0005886) and bacterial-type flagellum basal body (GO:0009425).

KEGG pathway analysis showed that the identified 759 DEGs were strongly enriched in the multiple pathways of energy conversion, molecular biosynthesis and environmental sensory activity in general, such as glycerophospholipid metabolism (mag00564), RNA polymerase (mag03020), and two-component system (mag02020). Particularly, two enriched pathways were specifically identified for the up-regulated DEGs, including glyoxylate and dicarboxylate metabolism (mag00630), ubiquinone and other terpenoid-quinone biosynthesis (mag00130). These pathways are related to a diversity in biosynthesis of amino acids and molecular machines and energy conversions. Among the enriched pathways, bacterial flagellar assembly (mag02040) and chemotaxis (mag02030) are directly related to bacterial movement.

13 common magnetosome-related genes in the CtrA regulon were identified when we compared 759 DEGs with known lists of magnetosome-related genes, including *mamF-like* (*amb0412*), *mamD-like* (*amb0400*), *mamJ* (*amb0964*), *mamS* (*amb0975*), *mamU* (*amb0977*, *amb2502*), *mamW* (*amb0981*), *mamQ2* (*amb1005*), *ftsZm* (*amb1015*), *mamX* (*amb1017*), *mamY* (*amb1018*), *mms5* (*amb1027*), *magA* (*amb3990*). Except *mamJ*, *mamW* and *mms5*, other 10 genes were recognized and assigned to 5 GO terms (integral component of membrane, cytoplasm, phospholipid biosynthetic process, ATP binding, metal ion binding) using GSEA-Pro v3. Furthermore, 11 types of genes clusters (COG, GO, KEGG, Operon, PFAM, InterPro, Superfamily, Keywords, eggNOG_COG, ENOG, Regulon) were obtained via GSEA-Pro v3.

We compared gene clusters which known magnetosome-related genes belong to, and identified multiple functional similar genes when these genes cooccurred in two or more clusters. These genes are shown in Table 1.

Table 1 Functional similar genes of known magnetosome-related genes.

Known Genes	Functional Similar Genes
<i>mamF-like</i> (AMB_RS02120, <i>amb0412</i>)	<i>amb0188, amb0273, amb0537, recA, amb0630, dnaA, amb0674, carA, kup1, amb0894, pstB2, amb1385, amb1428, amb1466, amb1551, amb1865, amb2097, amb2222, nrdR3, plsX, gyrA, coaD, nhaA, pyrH, lolD, amb2686, amb2770, gltX2, amb2976, amb3001, mnmA, hscA, anmK, ruvB, ruvA, hisG, amb3387, amb3618, amb3655, amb3712, amb3802, murF, murD, amb3924, amb4015, amb4061, amb4090, htpG, amb4368, amb4397, amb1062, amb2736</i>
<i>mamD-like</i> (AMB_RS02045, <i>amb0400</i>)	
<i>mamS</i> (AMB_RS04995, <i>amb0975</i>)	
<i>mamU</i> (AMB_RS05005, <i>amb0977</i>)	
<i>mamU</i> (AMB_RS12580, <i>amb2502</i>)	
<i>mamQ2</i> (AMB_RS05155, <i>amb1005</i>)	
<i>ftsZm</i> (AMB_RS05200, <i>amb1015</i>)	
<i>mamX</i> (AMB_RS05210, <i>amb1017</i>)	
<i>mamY</i> (AMB_RS05215, <i>amb1018</i>)	
<i>magA</i> (AMB_RS20195, <i>amb3990</i>)	

PPI network and cluster analysis

759 DEGs' protein sequences were submitted to the STRING database, and a PPI network in AMB-1 was constructed to identify interesting genes and biological clusters that may play roles in magnetosome biogenesis in the CtrA regulon. Moreover, some nodes did not connect to other nodes in the networks. These nodes were removed or hided from the networks. This PPI network was visualized using Cytoscape

as shown in Fig2. A total of 477 nodes and 4100 edges were screened from the PPI network. 9 of 13 known magnetosome-related genes were removed due to no connection with other nodes. Among 4 genes, *mamJ* has the highest node degree, while *ftsZm* is second highest, followed by 2 *mamU* nodes. The cluster analysis of this PPI network using MCODE with default parameters indicated that most magnetosome-related genes do not belong to any cluster, except *mamJ*. A small cluster of *mamJ*, *amb1158*, *amb1183*, and *amb0395* was found, and it is not known for magnetosome biogenesis. This result may suggest many magnetosome-related genes are not hub genes in a whole old network, and they can develop new abilities to interact with other genes during evolution, and consequently form new parts of the whole network (Fig3).

This network was analyzed using 6 methods (DMNC, MCC, Closeness, Radiality, Stress, Betweenness) supported by Cytoscape. The most important nodes are usually ribosome and translation-associated genes. Important genes having other types of functions can also be identified. Considering non-essential network status of magnetosome-related genes, we decided to search neighborhoods of known magnetosome genes in ranking lists. After searching and comparing 20 nearest neighbors of magnetosome-related gene, we identified 44 nodes, which are common nearest neighbors of magnetosome-related genes at least by 3 methods (Table 2). After being compared with the list of functional similar genes, 8 genes were commonly found in the two lists. These genes include *dnaA*, *lolD*, *amb1551*, *amb0674*, *amb3712*, *amb4015*, *amb0273*, and *amb4397*. Almost all genes have functions of ATP binding, whereas some of them have functions of regulating membrane properties. Besides, some genes also have functions of becoming cytoplasm components, metal ion binding, lipoprotein transport, and etc.

(Fig2 approximately here)

(Fig3 approximately here)

Table 2 Nearest neighbor nodes of known magnetosome-related genes in AMB-1. They were identified by at least 3 methods provided by CytoHubba.

Known Genes	Nearest Neighbors of Known Genes
<i>mamJ</i> (<i>amb0964</i>)	<i>amb0049, amb0429, dnaA, amb0452, amb1551, amb3710, amb2771, amb2827, amb0475, purH, amb1609, amb3286, amb1381, amb3825, amb2831, lolD, amb1158, amb0674, amb0235, amb3235, msrQ, amb0086, amb0104, amb3712, amb0018, amb1483, amb3185, amb2878, amb3242, amb4015, amb0562, amb0140, amb0273, amb2859, amb2998, amb1127, amb1273, amb0365, amb1962, amb0495, amb3786, amb4397, amb1623, amb3910</i>
<i>ftsZm</i> (<i>amb1015</i>)	
<i>mamU</i> (<i>amb0977</i>)	
<i>mamU</i> (<i>amb2502</i>)	

Six methods include MCC, DMNC, Closeness, Radiality, Betweenness, and Stress.

WGCNA analysis and functional enrichment analysis in detected modules

To further search potential magnetosome-related genes, co-expression modules of 4508 genes in AMB-1 were also constructed using the WGCNA R software. When the soft-thresholding power value equals 15, the scale-free fit index is greater than 0.9 and the average connectivity degree wasn't too low. Consequently, the WGCNA identified 21 co-expression modules using this power value. Some modules contain very large numbers of genes such as the blue and ivory modules. Known magnetosome-related genes were found not only in these two modules but also in other modules such as the steelblue, bisque4, darkred, and royalblue modules, highlighting a continuous time event of magnetosome formation. However, all 13 aforementioned magnetosome-related genes in the CtrA regulon were found in either the blue or ivory module.

The relationships among the co-expression modules are shown in a hierarchical clustering dendrogram of the eigengenes and a heatmap of the eigengene adjacency in Fig4 Panel (a) and Panel (b). The blue module has 15 known magnetosome-related genes in and outside the CtrA regulon, and this module has positive correlations with 8 modules such as the darkorange2, thistle1 and royalblue modules. However, this module are negatively related to the steelblue and ivory modules. By contrast, the ivory module is positively correlated with 8 modules such as the steelblue, salmon4, lightcyan1 and red modules. Most magnetosome-related genes can be grouped into the blue-module genes (for example *mamU*, *mamW*, *ftsZm*, *mamX*, *mamY*, *mamQ2*, *mamF-like*) and ivory-module genes (for example *mamJ*, *mamS*, *magA*, *mms5*, *mamD-like*). Both the essential and non-essential genes can be found in these two groups.

(Fig4 approximately here)

GO and KEGG pathway enrichment analysis was also conducted using gene lists of the blue and ivory modules. However, no significant results were found for both modules, indicating extremely diverse functions of genes in these two modules. After comparing functional annotations of the genes in two modules, we found that they belong to completely same 19 GO classes including one or more classes for cellular process and signaling, information storage and processing, metabolism, and poorly characterized. However, multiple genes in the blue module have functions of binding ions of more diamagnetic metal species than those in the ivory module, although many genes in both modules have functions of binding to ferromagnetic, paramagnetic or diamagnetic metals. Furthermore, some genes in the blue module have functions of specifically binding to ADP and UTP, whereas more genes in both modules have functions of binding to ATP, GTP and cAMP. Thirdly, two genes in the blue modules have functions of binding to NADP⁺ or NADPH specifically, but a gene in the ivory module has functions of binding to NAD⁺. In addition, two genes in the blue module have functions of binding to small ribosomal subunit and RNA polymerase, and meanwhile two genes in the ivory module have functions of binding to 5S rRNA, an opponent of large ribosomal subunit, or mRNA.

There are significant differences in enzyme activities between the blue and ivory modules. Although two modules share 22 enzyme activities, the blue module has 48 unique enzyme activities, while the ivory module has 41 unique enzyme activities. Several kinase activities such as lipid kinase activity and

riboflavin kinase activity were identified among these enzyme activities. Differences between the blue and ivory modules suggested energy conversion rate or related molecular movement might lead to boundaries between the blue and ivory modules. UTP can be biosynthesized from UDP by Nucleoside Diphosphate Kinase after using phosphate group from ATP. Conversion between NAD⁺ and NADP⁺ is catalyzed by NAD⁺ kinase or NADP phosphatase. An additional phosphate group is added or removed by these two enzymes. Small and large ribosomal subunits have different functions during the process of translation. The small subunit programs protein synthesis via binding to mRNA by hydrogen bonds, while the large one takes care of formation of peptides bonds.

GSEA-Pre v3 analysis was performed for closely correlated modules of the blue and ivory modules. Several similarities were found between the blue-close and ivory-close modules. For example, a common GO term between the thistle1 and royalblue modules is cytoplasm (GO:0005737), and this term is also shared among the steelblue, salmon4, and red modules. A common KEGG, i.e., two-component system (ko02020), was not only identified between the darkorange2 and royalblue modules but also among the steelblue, salmon4, red modules. In spite of these similarities, an obvious difference was also found between the blue-close and ivory-close modules using KEYWORDS classification. All blue-close modules shared a keyword, that is 3D-structure, whereas all ivory-close module shared not only the 3D-structure keyword but also an ATP-binding keyword. This finding highlights the importance of energy in defining boundaries of several co-expression modules as well.

Cytoscape input files of these co-expression modules were also created by the WGCNA, and the blue and ivory modules were visualized using Cytoscape as well. The evaluation of node properties showed that known magnetosome-related genes are likely ranked similarly by any method of CytoHubba. For example, *amb1018* was ranked 979th while *amb1005* was 981th out of 1345 genes by MCC method in ascending order in the blue module. Meanwhile, *amb1027* was ranked 1053th whereas *amb0964* ranked 1042th by Closeness method in the ivory module. We identified 104 genes who are nearest neighbors of all known magnetosome-related genes in the CtrA regulon in the blue and ivory modules totally (Table 3). We also compared this list with the aforementioned list of functional similar genes, and consequently identified 4 genes commonly shared between two lists. They are *dnaA*, *plsX*, *amb0273*, and *amb3001*. Both *dnaA* and *amb0273* were also identified as commonly existing genes when the list of nearest neighbor nodes of magnetosome-related genes in PPI was compared with the same list of functional similar genes. They have functions in energy conversion or membrane structure. Finally, we got a list of 10 potential members of magnetosome-related genes based on the list of functional similar genes. They are *dnaA*, *lolD*, *plsX*, *amb1551*, *amb0674*, *amb3712*, *amb4015*, *amb0273*, and *amb4397*, *amb3001*. These genes are enriched in ATP binding (GO:0005524), plasma membrane (GO:0005886), and integral component of membrane (GO:0016021).

Table 3 Nearest neighbor nodes of known magnetosome-related genes in the CtrA regulon in WGCNA modules. They were identified by at least 3 methods provided by CytoHubba.

Known Genes	Nearest Neighbors of Known Genes
<i>mamJ</i> (<i>amb0964</i>)	<i>amb0081, amb0087, amb0137, amb0151, amb0163, amb0255, amb0257, amb0273, amb0353, amb0403, amb0444, amb0465, amb0492, dnaA, amb0677, amb0710, lipB, amb0736, amb0750, amb0756, amb0901, amb0903, ftsZm</i>
(<i>amb1015</i>)	<i>amb0904, amb0929, amb0934, amb1031, amb1069, amb1074, amb1105, amb1113, amb1175, amb1282, amb1330, amb1358, amb1391, amb1445, amb1488, amb1489, amb1522, amb1531, amb1601, map, amb1640, amb1652, amb1719, amb1777, mamU</i>
(<i>amb0977</i>)	<i>amb1857, amb1860, amb1869, amb1889, amb1967, amb1994, amb2026, amb2099, amb2179, amb2264, lipA, plsX, amb2426, amb2653, accC, amb2805, amb2925, amb2962, amb2975, amb2983, pxpB, amb3001, amb3006, msrP, rpsH, amb3273, mamU</i>
(<i>amb2502</i>)	<i>amb3283, amb3284, amb3305, tnpB, amb3341, parE, amb3379, amb3449, miaA, amb3573, amb3580, amb3581, amb3688, amb3691, rsmA, amb3774, amb3814, dapE, amb3891, amb3928, accD, rplS, rlmH, amb4101, amb4256, amb4259, terL, amb4282, mamW</i>
(<i>amb0981</i>)	<i>amb4317, amb4415, ald, amb4510</i>
<i>mamX</i> (<i>amb1017</i>)	
<i>mamY</i> (<i>amb1018</i>)	
<i>mamS</i> (<i>amb0975</i>)	
<i>mamQ2</i> (<i>amb1005</i>)	
<i>mms5</i> (<i>amb1027</i>)	
<i>magA</i> (<i>amb3990</i>)	
<i>mamF-like</i> (<i>amb0412</i>)	
<i>mamD-like</i> (<i>amb0400</i>)	

Six methods include MCC, DMNC, Closeness, Radiality, Betweenness, and Stress.

A known magnetosome-related gene, *mamU* (*amb0977*), in the CtrA regulon was identified as one of top 10 hub genes by the closeness, radiality, stress, and betweenness methods in the blue module. Furthermore, another magnetosome-related gene *mamO* (*amb0969*) outside the CtrA regulon was also one of top 10 hub genes identified by the betweenness method in the ivory module. Most magnetosome-related genes were not hub nodes in these two co-expression modules. We found that there was no significant GO term for top 10 hub genes in the blue and ivory modules, which were identified by at least 3 methods of CytoHubba. However, a term, ATP-binding (GO:0005524), was commonly found between two genes in the blue module and a gene in the ivory module. Some GO terms were comparable between the blue and ivory modules, such as cell wall (GO:0005618) and cell wall organization (GO:0071555), DNA

binding (GO:0003677) and DNA repair (GO:0006281), or fatty acid biosynthetic process (GO:0006633) and polysaccharide biosynthetic process (GO:0000271).

Discussion

Different mutations in MTB genomes have been described in various articles. A large deletion in the magnetosome island lead to a nonmagnetic mutant of *Magnetospirillum gryphiswaldense* (Schübbe et al. 2003). Frequent rearrangements in the genomic islands were also detected during serial subcultivation of this bacteria species in the laboratory (Ullrich et al. 2005). More mutations were also identified in other bacteria, such as *Magnetospirillum magneticum* AMB-1 (Komeili et al. 2006). Since the discovery of new magnetotactic bacteria and magnetosome phenotypes, it is far enough to understand mechanisms of magnetosome biogenesis and related gene functions. Microarray bioinformatics analysis allow us to study thousands of gene expressions simultaneously. This analysis can produce vast amount of information, which may shed light on various aspects of biological entities and mechanisms. In our study, we analyzed a microarray dataset of *Magnetospirillum magneticum* AMB-1 and identify hub nodes of both the protein-protein interaction network and the co-expression networks using several tools such as R, STRING, WGCNA, and Cytoscape. These analyses were proved to be useful in cancer research, and they are also worthwhile when bacterial are research species.

A small number of microarray datasets for magnetotactic bacteria are collected in the GEO database at the moment. Consequently, we can only compare samples of CtrA mutants with wild-type bacteria. If more datasets are available, much more analyses would be performed to reveal complex mechanisms and roles of differential expressed genes in magnetotactic bacteria. Furthermore, high resolution co-expression networks can be created when more sample data are analyzed using WGCNA. Therefore, a better study result would be obtained and more unknown magnetosome-related genes would be identified. Nevertheless, we did our best to find potential genes relating to magnetosome in the CtrA regulon using an existing dataset in GEO.

Microbiologists are still looking for new genes relevant to magnetosome and are trying to unveil detailed mechanisms of magnetosome formation. Our studies might provide useful information for seeking new genes that play roles in magnetosome biogenesis. GO terms those these genes enriched in highlight the importance of energy and membrane in magnetosome formation. Multiple biological networks built in our study could be used to display, compute, and model the interactions of intracellular proteins and their genes. They may shed new light on cellular and molecular mechanisms in magnetotactic bacteria. A better understanding of these mechanisms could have important implications on the practice of biology for both bacterial research and other researches. For example, both biologists and material scientists want to understand fundamental mechanisms of magnetosome biogenesis, since material scientists would like to simulate this biological process for the synthesis of magnetic artificial nanoparticles (Jacinto et al. 2021). They may wonder what proteins in magnetotactic bacteria could help produce high-quality nanomaterials. It is crucial to understand basic mechanism for answering this question. Our studies might offer several clues to this question. The proteins encoded by potential and known

magnetosome-related genes might help synthesis of magnetic artificial nanoparticles if appropriate conditions are available.

Although several researches revealed various factors affecting co-expression module or gene expression, few researches focused on biological mechanisms for co-expression module boundaries. Multispecies symbioses can substantially change co-expression network structure of plant and fungi (Palakurty et al. 2018). Stress conditions can lead to metabolic tightening and expansion of the co-expression network in plant (Fait et al. 2020). According to Wagner's estimates, energy constraints can affect the evolution of gene expression in yeasts (Wagner 2005). Our study suggests energy could be a key factor for regulating gene expression and consequently several co-expression module structures in *Magnetospirillum magneticum* AMB-1. Not only GO terms that differential expression genes between CtrA mutants and wild-type bacteria are enriched in but also multiple comparisons between the blue and ivory co-expression modules gave us this impression. Like other biological processes, magnetosome biogenesis requires extra energy to modulate expression of genes encoding magnetosome proteins. How to import energy carriers into membrane vesicles and then convert them to other formats properly is essential for the success of magnetosome formation and its excellent magnetic properties. This study identified 10 potential members of magnetosome-related genes in the CtrA regulon. Together with known information for magnetosome biogenesis, it may help reveal more mechanisms for bacterial biomineralization.

Conclusion

In conclusion, we have identified 10 genes potentially associated with magnetosome formation by differential expression analysis for CtrA mutants, GO and KEGG pathway enrichment analysis for diverse groups of genes, protein-protein interaction network analysis for differential expressed genes, weighted gene co-expression construction analysis for all genes in *Magnetospirillum magneticum* AMB-1, and nearest neighbor search analysis using rank similarity based on CytoHubba results. Moreover, we have also found the clues on the importance of energy in regulating co-expression module structures by comparing two key co-expression modules, i.e., the blue and ivory modules, which most known magnetosome-related genes belong to. Additionally, we found that many magnetosome-related genes are not hub genes in networks, and therefore they may develop new abilities to interact with other gene encoding proteins during evolution, including their functions in magnetosome biogenesis. Although our research is in its early stages, the results of this study still provide new insights into magnetosome biomineralization.

Abbreviations

Magnetotactic bacteria (MTB)

Declarations

Acknowledgments This work was supported by the National Key R&D Program of China (Grant No. 2017YFA0104302), the National Natural Science Foundation of China (Grant No. 61821002, 51832001).

Funding

This work was supported by the National Key R&D Program of China (Grant No.2017YFA0104302), the National Natural Science Foundation of China (Grant No. 61821002, 51832001).

Conflicts of interests

The authors declare that they have no conflict of interest.

Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The dataset analyzed during the current study is available in the NCBI GEO database.

Code availability

Not applicable

Author contributions

Yizi Yang analyzed the data and wrote the draft of this manuscript. Ning Gu designed the study and is also a research supervisor to support this research.

References

1. Arakaki A, Tanaka M, Matsunaga T (2018) Molecular Mechanism of Magnetic Crystal Formation in Magnetotactic Bacteria.. In: In: Matsunaga T, Tanaka T, Kisailus D (eds) Biological Magnetic Materials and Applications. Springer, Singapore, pp 23–51. doi:10.1007/978-981-10-8069-2_2
2. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4:2. doi:10.1186/1471-2105-4-2

3. Barrett T et al (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 41:D991–995. doi:10.1093/nar/gks1193
4. Carvalho SM, de Jong A, Kloosterman TG, Kuipers OP, Saraiva LM (2017) The *Staphylococcus aureus* alpha-Acetolactate Synthase ALS Confers Resistance to Nitrosative. *Stress Front Microbiol* 8:1273. doi:10.3389/fmicb.2017.01273
5. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 Suppl 4S11. doi:10.1186/1752-0509-8-S4-S11
6. Davis S, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23:1846–1847. doi:10.1093/bioinformatics/btm254
7. Deng K, Wu L (2018) Bioinformatics analysis of esophageal squamous cell carcinoma genome microarray based on GEO database *International Journal of Science* 5
8. Fait A, Batushansky A, Shrestha V, Yobi A, Angelovici R (2020) Can metabolic tightening and expansion of co-expression network play a role in stress response and tolerance? *Plant Sci* 293:110409. doi:10.1016/j.plantsci.2020.110409
9. Frankel RB, Bazylnski DA (2006) How magnetotactic bacteria make magnetosomes queue up. *Trends Microbiol* 14:329–331. doi:10.1016/j.tim.2006.06.004
10. Gentleman R, Carey V, Morgan M, Falcon S (2005) Biobase: base functions for Bioconductor R package version 1. doi: 10.18129/B9.bioc.Biobase
11. Greene SE, Brilli M, Biondi EG, Komeili A (2012) Analysis of the CtrA pathway in *Magnetospirillum* reveals an ancestral role in motility in alphaproteobacteria. *J Bacteriol* 194:2973–2986. doi:10.1128/JB.00170-12
12. Jacinto MJ, Silva VC, Valladao DMS, Souto RS (2021) Biosynthesis of magnetic iron oxide nanoparticles: a review *Biotechnol Lett* 43:1–12. doi:10.1007/s10529-020-03047-0
13. Kolde R, Kolde MR (2015) Package 'pheatmap' R Package 1
14. Komeili A, Li Z, Newman DK, Jensen GJ (2006) Magnetosomes are cell membrane invaginations organized by the actin-like protein. *MamK Science* 311:242–245. doi:10.1126/science.1123231
15. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi:10.1186/1471-2105-9-559
16. Lefevre CT et al (2013) Comparative genomic analysis of magnetotactic bacteria from the Deltaproteobacteria provides new insights into magnetite and greigite magnetosome genes required for magnetotaxis. *Environ Microbiol* 15:2712–2735. doi:10.1111/1462-2920.12128
17. Lin W et al (2014) Genomic insights into the uncultured genus 'Candidatus Magnetobacterium' in the phylum Nitrospirae. *ISME J* 8:2463–2477. doi:10.1038/ismej.2014.94
18. Lin W, Zhang W, Zhao X, Roberts AP, Paterson GA, Bazylnski DA, Pan Y (2018) Genomic expansion of magnetotactic bacteria reveals an early common origin of magnetotaxis with lineage-specific evolution. *ISME J* 12:1508–1519. doi:10.1038/s41396-018-0098-9

19. Lohsse A et al (2014) Genetic dissection of the mamAB and mms6 operons reveals a gene set essential for magnetosome biogenesis in *Magnetospirillum gryphiswaldense*. *J Bacteriol* 196:2658–2669. doi:10.1128/JB.01716-14
20. Murat D, Quinlan A, Vali H, Komeili A (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc Natl Acad Sci U S A* 107:5593–5598. doi:10.1073/pnas.0914439107
21. Palakurty SX, Stinchcombe JR, Afkhami ME (2018) Cooperation and coexpression: How coexpression networks shift in response to multiple mutualists. *Mol Ecol* 27:1860–1873. doi:10.1111/mec.14550
22. Raschdorf O, Forstner Y, Kolinko I, Uebe R, Plitzko JM, Schuler D (2016) Genetic and Ultrastructural Analysis Reveals the Key Players and Initial Steps of Bacterial Magnetosome Membrane. *Biogenesis PLoS Genet* 12:e1006101. doi:10.1371/journal.pgen.1006101
23. Schübbe S et al (2003) Characterization of a Spontaneous Nonmagnetic Mutant of *Magnetospirillum gryphiswaldense* Reveals a Large Deletion Comprising a Putative Magnetosome. *Island J Bacteriol* 185:5779–5790. doi:10.1128/jb.185.19.5779-5790.2003
24. Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. doi:10.1101/gr.1239303
25. Smyth GK (2005) limma: Linear Models for Microarray Data. *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, New York, NY, pp 397–420. doi:10.18129/B9.bioc.limma
26. Szklarczyk D et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607–D613. doi:10.1093/nar/gky1131
27. Taoka A, Fukumori Y (2018) Structure and Function of Aligned Magnetic Crystals in Magnetotactic Bacteria.. In: In: Matsunaga T, Tanaka T, Kisailus D (eds) *Biological Magnetic Materials and Applications*. Springer, Singapore, pp 3–22. doi:10.1007/978-981-10-8069-2_1
28. Uebe R et al (2011) The cation diffusion facilitator proteins MamB and MamM of *Magnetospirillum gryphiswaldense* have distinct and complex functions, and are involved in magnetite biomineralization and magnetosome membrane assembly. *Mol Microbiol* 82:818–835. doi:10.1111/j.1365-2958.2011.07863.x
29. Uebe R, Schuler D (2016) Magnetosome biogenesis in magnetotactic bacteria. *Nat Rev Microbiol* 14:621–637. doi:10.1038/nrmicro.2016.99
30. Ullrich S, Kube M, Schubbe S, Reinhardt R, Schuler D (2005) A hypervariable 130-kilobase genomic region of *Magnetospirillum gryphiswaldense* comprises a magnetosome island which undergoes frequent rearrangements during stationary growth. *J Bacteriol* 187:7176–7184. doi:10.1128/JB.187.21.7176-7184.2005
31. Wagner A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol* 22:1365–1374. doi:10.1093/molbev/msi126

32. Wickham H (2011) ggplot2 Wiley Interdisciplinary Reviews. Comput Stat 3:180–185.
doi:10.1002/wics.147

33. Wickham H (2012) reshape2: Flexibly reshape data: a reboot of the reshape package R package
version 1

Figures

Figure 1

Identification of DEGs. (a) Density plot. (b) Volcano plot of the 806 DEGs and the 6 unannotated ORFs with $FDR < 0.05$ and $|\log FC| > 1$. The red dots represent 428 up-regulated genes and 6 unannotated ORFs while the blue represent 378 down-regulated genes.

Figure 2

Protein-protein interaction network of all DEGs in AMB-1 identified by STRING.

Figure 3

Hypothesis of gaining ability for the evolution of magnetosome-related genes. Red means a hypothesized evolutionary event.

Figure 4

Construction of WGCNA co-expression modules. (a) The dendrogram of the eigengenes; (b) The heatmap of the eigengene adjacency. The heatmap shows the relatedness of the 20 co-expression modules identified by WGCNA except the grey module. Red means positive correlation, blue means negative correlation. The stronger the color, the larger the correlation magnitude.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FileS1.xlsx](#)