RubisCO Gene Clusters Found in a Metagenome Microarray from Acid Mine Drainage

Xue Guo, a,b Huaqun Yin, a,b Jing Cong, a Zhimin Dai, a,b Yili Liang, a,b Xueduan Liu, a,b

The enzyme responsible for carbon dioxide fixation in the Calvin cycle, ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), is always detected as a phylogenetic marker to analyze the distribution and activity of autotrophic bacteria. However, such an approach provides no indication as to the significance of genomic content and organization. Horizontal transfers of RubisCO genes occurring in eubacteria and plastids may seriously affect the credibility of this approach. Here, we presented a new method to analyze the diversity and genomic content of RubisCO genes in acid mine drainage (AMD). A metagenome microarray containing 7,776 large-insertion fosmids was constructed to quickly screen genome fragments containing RubisCO. Form I large-subunit genes (cbbL). Forty-six cbbL-containing fosmids were detected, and six fosmids were fully sequenced. To evaluate the reliability of the metagenome microarray and understand the microbial community in AMD, the diversities of cbbL and the 16S rRNA gene were analyzed. Fosmid sequences revealed that the form I RubisCO gene cluster could be subdivided into form IA and IB RubisCO gene clusters in AMD, because of significant divergences in molecular phylogenetics and conservative genomic organization. Interestingly, the form I RubisCO gene cluster coexisted with the form II RubisCO gene cluster in one fosmid genomic fragment. Phylogenetic analyses revealed that horizontal transfers of RubisCO genes may occur widely in AMD, which makes the evolutionary history of RubisCO difficult to reconcile with organismal phylogeny.
TABLE 1 Main features and RubisCO-associated cluster present in the analyzed fosmids

| Fosmid   | Length (bp) | %GC content | Coverage (%) | RubisCO-associated genes and characteristics
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DX-8J-22</td>
<td>32,275</td>
<td>58.37</td>
<td>37</td>
<td>CbbR: A. ferrooxidans ATCC 23270, 99%; CbbL: A. ferrooxidans ATCC 23270, 100%; CbbS: A. ferrooxidans ATCC 23270, 100%; CsoS2: A. ferrooxidans ATCC 53993, 99%; CsoS3: A. ferrooxidans ATCC 23270, 100%; CsaC: A. ferrooxidans ATCC 23270, 100%; CsaE: A. ferrooxidans ATCC 23270, 98%; CsaE: A. ferrooxidans ATCC 23270, 100%; CsoQ: A. ferrooxidans ATCC 23270, 100%</td>
</tr>
<tr>
<td>DX-4H-17</td>
<td>37,399</td>
<td>64.53</td>
<td>32</td>
<td>CbbR: Beggiatia sp. PS, 53%; CbbL: Acidithiobacillus sp. P2, 90%; CsoS: A. ferrovarans SS3, 69%; CsoS2: H. neapolitanus c2, 45%; CsoS3: H. neapolitanus c2, 58%; CsaA: T. denitrificans ATCC 25259, 86%; CsaC: A. ferruvorans SS3, 74%; CsaE: A. caldus ATCC 51756, 95%; CsoS: A. ferruvorans SS3, 88%; CsoS3: A. ferruvorans SS3, 57%</td>
</tr>
<tr>
<td>DX-4K-26</td>
<td>31,477</td>
<td>61.01</td>
<td>38</td>
<td>CbbR: A. ferruvorans SS3, 78%; CbbL: A. ferruvorans SS3, 93%; CsoS: A. ferruvorans SS3, 56%; CsoS3: A. ferruvorans SS3, 57%</td>
</tr>
<tr>
<td>DX-1A-14</td>
<td>32,546</td>
<td>64.00</td>
<td>36</td>
<td>CbbR: Beggiatia sp. PS, 53%; CbbL: Acidithiobacillus sp. P2, 90%; CsoS: A. ferruvorans SS3, 69%; CsoS2: H. neapolitanus c2, 45%; CsoS3: H. neapolitanus c2, 58%</td>
</tr>
<tr>
<td>DX-3D-09</td>
<td>32,174</td>
<td>66.73</td>
<td>37</td>
<td>CbbM: L. cholodnii SP-6, 90%; CbbQ: L. cholodnii SP-6, 82%; CsoS: L. cholodnii SP-6, 99%; CsoS: A. ferruvorans, 92%; CbbS: T. denitrificans ATCC 25259, 89%; CbbQ: C. metallidurans CH34, 91%; CsoS: T. denitrificans ATCC 25259, 89%</td>
</tr>
<tr>
<td>DX-7F-24</td>
<td>40,978</td>
<td>65.85</td>
<td>30</td>
<td>CbbM: L. cholodnii SP-6, 90%; CbbQ: L. cholodnii SP-6, 83%; CsoS: L. cholodnii SP-6, 70%</td>
</tr>
</tbody>
</table>

* Subunit present, closest strain hit, and percent similarity are shown.

insertion clones. A metagenome microarray was constructed as described previously (18). Each clone was incubated in a shaking incubator at 37°C and 170 rpm in the presence of chloramphenicol (12.5 μg/ml) and an inducer (1 μl/ml) (Epitope). Cells were harvested the next day, and the fosmid DNA was extracted using a QIAprep Spin Miniprep kit (Qiagen, Germany) according to the manufacturer’s protocol. The fosmid DNAs were stored in a final concentration of 40 ng/μl. Each clone was incubated in a shaking incubator at 37°C. A metagenome microarray was constructed as described previously (18). Each clone was incubated in a shaking incubator at 37°C and 170 rpm in the presence of chloramphenicol (12.5 μg/ml) and an inducer (1 μl/ml) (Epitope). Cells were harvested the next day, and the fosmid DNA was extracted using a QIAprep Spin Miniprep kit (Qiagen, Germany) according to the manufacturer’s protocol. The fosmid DNAs were stored in a final concentration of 40 ng/μl. Each clone was incubated in a shaking incubator at 37°C. Each clone was incubated in a shaking incubator at 37°C. A metagenome microarray was constructed as described previously (18). Each clone was incubated in a shaking incubator at 37°C and 170 rpm in the presence of chloramphenicol (12.5 μg/ml) and an inducer (1 μl/ml) (Epitope). Cells were harvested the next day, and the fosmid DNA was extracted using a QIAprep Spin Miniprep kit (Qiagen, Germany) according to the manufacturer’s protocol. The fosmid DNAs were stored in a final concentration of 40 ng/μl. Each clone was incubated in a shaking incubator at 37°C.
and cbbL genes were ligated into vector pGM-T (Tiangen Biotech, China) and transformed into competent E. coli DH5α according to the manufacturer’s protocol. Plasmids from the 16S rRNA gene and cbbL libraries were subsequently extracted using a TIANprep miniplasmid kit (Tiangen Biotech, China). Clones containing 16S rRNA and putative cbbL genes were screened by restriction fragment length polymorphism (RFLP) withMspI and HinfI restriction endonucleases (MBI Fermentas). Restriction fragments were analyzed in 3% (wt/vol) agarose gels in 1× TAE buffer. Selected unique plasmids were sequenced bidirectionally with the vector-specific primers SP6 reverse and T7 promoter. Sequences were edited manually with the DNAstar package (Madison, WI).

Phylogenetic analysis. Phylogenetic analysis of 16S rRNA and cbbL gene sequences was performed on sequences screened from 16S rRNA gene and cbbL libraries, and representatives from metagenome microarrays (see Fig. 2 and 3). 16S rRNA gene sequences were aligned with sequences from the GenBank database with the CLUSTAL program (26), and phylogenetic trees were constructed by the use of Molecular Evolutionary Genetics Analysis 4.0 software (MEGA, version 4.0) (27). Deduced amino acid sequences for cbbL clones and representatives of 46 cbbL-containing fosmids were aligned with known sequences from the GenBank database with the CLUSTAL program (26), and cbbL gene trees were also constructed by using MEGA 4.0 (27).

Nucleotide sequence accession numbers. Sequences determined and annotated in this study are available in GenBank under accession numbers JQ815894 to JQ815896 and JX308284 to JX308286. The incomplete cbbL sequences in GenBank have been compiled under accession numbers JX297619 to JX297625 and JQ815897 to JQ815942, and 16S rRNA gene sequences have been compiled under accession numbers JX297607 to JX297618.

RESULTS
Metagenomic library and metagenome microarray hybridization. Samples were collected from AMD in Dexing Copper Mine, China. The environmental DNA was extracted and used to construct a fosmid library. The library collection contained in total 7,776 large-insert clones. The genomic fragments cloned in the fosmids had a size of between 30 and 45 kb.

To quickly screen Rubisco genes from the metagenomic library, we constructed a metagenome microarray using the fosmid library. After hybridization and checking by PCR amplification, 46 cbbL-containing fosmids were screened out. And we fully sequenced 21 fosmids that were selected from the 46 cbbL-containing fosmids. All of the six fosmids showed evidence of coding either a form I or form II Rubisco large subunit. The genomic fragments in the fosmids had a size between 31.4 and 40.9 kb, and all could be assembled in one single contig (Table 1). The total annotations of 46 genomic fragments are presented in the supplemental material.

Rubisco genes in the fosmids. The most interesting genomic fragment was found in fosmid DX-3D-09, in which we identified two Rubisco clusters, one belonging to form I and the other belonging to form II (Fig. 1). The subunits CbbL and CbbS are essential for the catalytic activity of the enzyme in form I Rubisco, while CbbM is the sole catalytic subunit in the form II Rubisco enzyme. The catalytic subunit cbbM was complete, as this fosmid contained complete subunits cbbL and cbbS. In addition to this fosmid, complete subunits cbbL and cbbS were also identified in DX-8J-22, DX-4H-17, DX-4K-26, and DX-1A-14, while subunit cbbM was also found in DX-7F-24. In all six of the fosmids, additional subunits in the Rubisco gene cluster were identified, providing more reliable evidence of the presence of a functional Rubisco enzyme.

Form I Rubisco. Among the six fosmids, five ORFs were identified as form I large-subunit cbbL from the identification of the conserved catalytic sequence motif KDDE (28) and the small-subunit cbbS located in the immediate downstream region, providing more evidence to distinguish cbbL from cbbM. Comparing Rubisco gene clusters in the six fosmids with different bacterial genomic islands, we could subdivide genes in the form I Rubisco gene cluster into two types: (i) the form IA Rubisco gene cluster and (ii) the form IB Rubisco gene cluster (Fig. 1). Both forms contained catalytic subunits cbbL and cbbS. The form IA Rubisco gene cluster harbored subunits cbbQ and cbbO that are believed to be required in the processes of regulation and posttranslational modification of the Rubisco enzyme (29). The genomic organization of the form IA Rubisco gene cluster is L→S→Q→O. The organization is rather conservative in different organisms. The organization of the form IB Rubisco gene cluster is significantly different from that of the form IA Rubisco gene cluster. Although accessory subunits cbbQ and cbbO may also occur in form IB Rubisco gene clusters, genes encoding the carboxysome are inserted between catalytic subunits cbbL and cbbM (Fig. 1). This is the most conspicuous difference between form IA and IB Rubisco gene clusters.

Form IA gene cluster. The only representative of the form IA Rubisco gene cluster was found in fosmid DX-3D-09 (Fig. 1). It shows high overall similarity to other known form IA Rubisco gene clusters. The large-subunit CbbL encoded by fosmid DX-3D-09 has its best hit within Betaproteobacteria, and it is very (93%) similar to the CbbL found in Thiobacillus denitrificans. In addition, the organization of cbbS, cbbQ, and cbbO genes associated with this cbbL gene is L→S→Q→O, typical of the form IA Rubisco gene cluster. The subunits CbbS, CbbQ, and CbbO have their best hits within Betaproteobacteria. These subunits showed highest similarities with Thiobacillus denitrificans ATCC 25259 (89%), Capriovivus metallidurans CH34 (91%), and Thiobacillus denitrificans ATCC 25259 (80%), respectively (Table 1). The organization of the form I Rubisco gene cluster in fosmid DX-3D-09 is very similar to that of Thiobacillus denitrificans ATCC 25259, but the overall organization of fosmid DX-3D-09 is most like that of Capriovivus metallidurans, a strain within the class Betaproteobacteria. Interestingly, the same fosmid contains another cluster of Rubisco genes with a form II large-subunit cbbM (see below).

Form IB gene cluster. Among the other fosmids, four fosmids (DX-1A-14, DX-4H-17, DX-4K-26, and DX-8J-22) contained the form IB Rubisco gene cluster. In the four fosmids, a gene encoding transcriptional regulator CbbR (30) was located in the immediate upstream of cbbL in a divergent orientation. The CbbL protein encoded by fosmid DX-4H-17 (see Table S2 in the supplemental material) has its best hit within Actinobacteria, and it is most (90%) similar to Acidithiomicrobiaceae sp. In the form IB Rubisco gene cluster of fosmid DX-4H-17, the subunit cbbR has its best hit with Beiggiota sp. PS (53%), cbbS has its best hit with Acidithiothiobacillus ferrooxidans SS3 (69%), and cbbO has its best hit with Acidithiothiobacillus caldus ATCC 51756 (64%) (Table 1). However, the above-mentioned three bacteria are within the class Gammaproteobacteria. In addition, we could not find the subunit cbbQ in the surrounding region of the form IB Rubisco gene cluster, but a 6-ORF operon encoding carboxysome proteins was identified in the immediate downstream of cbbLS (Fig. 1). In Acidithiothiobacillus ferrooxidans ATCC 23270 (31) and Synechococcus sp.
WH 8102, the Rubisco cluster is located near genes related to the carboxysome (32). In the 6-ORF operon, all of genes are oriented in the direction that is congruent with cbbLS. The gene csoS2 is thought to encode a carboxysome shell protein with a high molecular weight. csoS3, a gene encoding carbonic anhydrase, was immediately downstream of csoS2. Carbonic anhydrase is sequestered with the Rubisco enzyme in the carboxysome to convert bicarbonate to the Rubisco substrate CO2, as it enters the microcompartment from the cytosol (33). The four remaining genes (cscA, cscB, cscD, and cscE) were shown to encode proteins that are components of the carboxysome shell. All 6 of the genes have their best hits within Gammaproteobacteria. The closest strain hits for the 6 genes are summarized in Table 1. The overall organization of the fosmid DX-4H-17 has the highest similarity hits to the genomic island of Acidithiobacillus caldus Sm-1. As for fosmid DX-8J-22, all genes, including those encoding Rubisco enzymes, are highly similar to those of Acidithiobacillus ferrooxidans (see Table S1 in the supplemental material), and the overall organization of the genes is identical to that of Acidithiobacillus ferrooxidans. So, there is no doubt that the genomic fragment cloned in fosmid DX-8J-22 belongs to Acidithiobacillus ferrooxidans.

The remaining fosmids DX-4K-26 and DX-1A-14 had truncated form IB Rubisco gene clusters. Given the conservative arrangement of carboxysome genes (34), it is possible that a complete operon encoding carboxysome is located in the immediate downstream of cbbLS in the organisms to which these fosmid sequences belong. Fosmid DX-4K-26 contained the subunits cbbR, cbbL, cbbS, and csoS2 (Fig. 1). All of the subunits are most similar to Acidithiobacillus ferrivorans SS3 within the class Gammaproteobacteria. However, the other genes in the further upstream region of these subunits have their best hits within Betaproteobacteria (see Table S3 in the supplemental material). Although none of the individual gene products were most similar to those of Sideroxydans lithotrophicus, the overall organization of genes in fosmid

FIG 1 Genes and similarity comparison of environmental fosmids from the AMD metagenomic library containing Rubisco genes. Rubisco cluster genes are highlighted in different shades. Specific additional genes mentioned in the text are also indicated. The two closest genome fragments of cultivated microbes Acidithiobacillus ferrooxidans ATCC 23270 and Leptothrix cholodnii SP-6 are also shown.
DX-4K-26 is most like that of genes in *Sideroxydans lithotrophicus* ES-1. The organization of the form IB Rubisco gene cluster in fosmid DX-1A-14 is very similar to that of fosmid DX-4H-17. Also, fosmid DX-1A-14 *cbbL* is most (90%) similar to *Acidithio- microbium* sp. (see Table S4 in the supplemental material).

**Form II Rubisco.** Form II Rubisco genes were found in fosmid DX-3D-09 and DX-7F-24 (see Table S6). The form II Rubisco gene clusters in the two fosmids contained subunits *cbbM*, *cbbQ*, and *cbbO* oriented in the same direction. The three subunits in the two fosmids are highly similar (Table 1). The CbbM protein encoded by fosmid DX-3D-09 is 29% similar to CbbL encoded by the same fosmid. CbbM has the conserved catalytic sequence motif GGDFIKNDE, which differentiates form II Rubisco from other members of the Rubisco superfamily (including form I Rubisco) (35). The subunit genes *cbbM*, *cbbQ*, and *cbbO* have their best hits with *Leptothrix cholodnii* SP-6 (90%, 82%, and 69%, respectively) (see Table S5). Aligning the CbbQ and CbbO amino acid sequences of form I and II Rubisco clusters in fosmid DX-3D-09, respectively, we found that the two CbbQs are 68% similar to each other and the two CbbOs are 40% similar to each other. These results suggest that the *cbbQ* and *cbbO* genes in the two clusters may have originated from horizontal gene transfer rather than ancient gene duplication.

**Phylogenetic analysis based on 16S rRNA gene and cbbL sequences.** Rubisco genes substitute for 16S rRNA gene sequences in determining phylogenetic relationships in many microbial communities. To test this assertion and determine the phylogenetic affiliation in AMD, 16S rRNA gene sequence recovery was compared with *cbbL* sequence recovery in the same sample. 16S rRNA gene sequences were designated beginning with “AMD- DX-,” followed by their number in the clone library. The *cbbL* sequences from *cbbL* clone library were designated beginning with “DX-R,” followed by the clone number in the library, and the *cbbL* sequences from metagenome microarray were designated beginning with “AMD,” followed by the clone number in the metagenomic library (Fig. 2 and 3). The phylogenetic tree constructed for *cbbL* sequences strongly conflicts with the 16S rRNA gene phylogenetic tree. According to the 16S rRNA gene phylogenetic tree, 16S rRNA genes of microorganisms in AMD could be assigned into five bacterial lineages (*Alpha-, Beta-, and Gammaproteobacteria, Actinobacteria, and Nitrospira*) (Fig. 2). However, the *cbbL* genes formed two phyletic groups based on the types of Rubisco
gene cluster. Some bacteria such as *Acidithiobacillus ferrivorans* contain both form IA and IB RubisCO gene clusters, but the *cbbL* genes in the two clusters were assigned into different phyletic groups. Comparing the phylogenetic trees of the 16S rRNA gene and *cbbL*, more incongruences were found. *Nitrobacter winogradskyi* Nb-255 belongs to the *Alphaproteobacteria* group in the 16S rRNA gene phylogenetic tree, and *Thiomonas intermedia* K12 is a member of the *Betaproteobacteria*, while they form a phyletic cluster in the *cbbL* phylogenetic tree. On the other hand, *Halorhodospira halophila* and *Halothiobacillus neapolitanus* were assigned into the *Gammaproteobacteria* group and were closely related to each other in the 16S rRNA gene phylogenetic tree, but they were separately divided into form IB and IA phyletic clusters in the *cbbL* phylogenetic tree. Similar contradictions occurred between *Thiobacillus denitrificans* and *Thiomas intermedia*. These results revealed that the evolution of the *cbbL* gene obviously could not reconcile with the organismal phylogeny in AMD.

Focusing on the distribution of *cbbL* sequences from the library and metagenome microarray, we could find that each monophyletic group formed by experimental sequences contained *cbbL* sequences from both the *cbbL* library and the metagenome microarray (Fig. 3). This suggested that the depth of the metagenomic library is sufficient for analysis of RubisCO gene diversity in the AMD environment.

**DISCUSSION**

The rapid screening of target genes from unculturable bacteria can be achieved by the combination of the metagenomic library and the microarray. Here, we report the first case of RubisCO-associated genomic fragments isolated from AMD using the metagenomic microarray. Six RubisCO genes containing genomic fragments were obtained from microarray screening and sequencing. None of the genomic fragments, with the exception of DX-8J-22, could be assigned into the known bacteria, but they showed the diversity of RubisCO gene clusters in AMD.

Interestingly, two RubisCO gene clusters that separately code
for form I and II RubisCO were found in the same fosmid sequence. Previous studies suggested that form I RubisCO can fix carbon dioxide at lower levels of CO₂ and that form II RubisCO is activated at higher levels of CO₂ (36). The form I RubisCO gene cluster was assigned into form IA and IB RubisCO gene clusters. A primary divergence in the gene organization of the form IB RubisCO gene clusters is the insertion of genes encoding carboxysome, which are lacking in form IA RubisCO gene clusters. Carboxysome is a polyhedral bacterial microcompartment that contains the carbon-concentrating mechanism (CCM) (37). In carboxysomes, carbonic anhydrase catalyzes bicarbonate into CO₂ in the vicinity of RubisCO to enhance CO₂ fixation (34). These results suggested that extremophiles in AMD can control carbon dioxide fixation by various mechanisms, including regulation of the expression of different gene clusters and improvement of the CO₂ concentration.

Previous work indicated that discrepancies occurred widely between phylogenies based on RubisCO genes and those based on 16S rRNA genes, but the evidence was much weaker, providing no indication as to the significance of genomic content and organization. The data from the genomic fragments presented here revealed that phylogenetic conflicts occurred widely between the cbbL genes and surrounding genomic fragments. Genes further up- and downstream of the cbbL in fosmid DX-4H-17 showed more congruence to Proteobacteria than to Actinobacteria, within which the cbbL had the best hit. It is possible that these phylogenetic conflicts may result from horizontal gene transfer rather than ancient gene duplication. In fact, extreme environmental conditions such as AMD have a great effect on the occurrence of horizontal gene transfer. The case of fosmid DX-4K-26 is slightly different from those of the above-mentioned two fosmids. All subunits encoded by the RubisCO gene cluster in fosmid DX-4K-26 are most similar to those of Acidithiobacillus ferrivorans SS3. However, it is not enough to determine that the genomic fragment originates from A. ferrivorans SS3. Because all other genes upstream of the RubisCO gene cluster have their best hits within Betaproteobacteria, we deduced that the fragment belongs to Betaproteobacteria. In fact, the RubisCO gene cluster may be acquired by one horizontal gene transfer event. A similar phenomenon occurring in Rhodobactercapsulatus has been reported (38). These data revealed that horizontal transfers of RubisCO genes may occur widely in AMD, which makes the RubisCO gene cluster difficult to reconcile with further up- and downstream genomic fragments. Besides, horizontal transfers of RubisCO genes are significant in ecology, since they may change the microbial carbon dioxide fixation ability, which contributes to niche differentiation in AMD microbial communities.

The cbbL phylogenetic tree revealed the interesting phenomenon that form IA and IB RubisCO gene clusters contain not only the respective genomic organizations but also phylogenetic differences. Previous works showed that some bacteria such as Acidithiobacillus caldus ATCC 51756 (39) contain only the form IB RubisCO gene cluster, some bacteria such as Cupriavidus metallidurans CH34 (40) harbor only the form IA RubisCO gene cluster, and some bacteria such as Acidithiobacillus ferrooxidans ATCC 23270 (31) contain RubisCO gene clusters of both forms (41). Nevertheless, cbbL genes preferentially form two phyletic groups based on the forms of the gene cluster rather than on species in the cbbL phylogenetic tree. Thus, we believe that the cbbL genes in form IA and IB RubisCO gene clusters are significantly divergent in molecular evolution and phylogenetics, though whether the assertion is widely applicable in all organisms needs further phylogenetic analysis. Comparing the 16S rRNA gene phylogenetic tree with the cbbL phylogenetic tree, many conflicts were found. Some bacteria that are close in the 16S rRNA gene phylogenetic tree, such as Thiobacillus denitrificans ATCC 25259 and Thiornonas intermedia K12, were divided into form IA and IB phyletic clusters, which suggested that they contain two different form I RubisCO gene clusters, while some bacteria that are distant in the 16S rRNA gene phylogenetic tree form a cluster with a high bootstrap value in the cbbL phylogenetic tree. This is strong evidence of horizontal gene transfer of RubisCO genes. It is unreliable to study the phylogeny based on cbbL sequences without the knowledge of the genetic content and organization of the RubisCO gene cluster in microbes.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program (no. 2010CB630901) and the National Natural Science Foundation of China (no. 30770051).

REFERENCES

nity structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43.