Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposits

Kari E. Dunfield  
*University of Maine*

Gary M. King  
*University of Maine*

Follow this and additional works at: https://digitalcommons.lsu.edu/biosci_pubs

**Recommended Citation**

https://doi.org/10.1128/AEM.70.7.4242-4248.2004

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.
Molecular Analysis of Carbon Monoxide-Oxidizing Bacteria Associated with Recent Hawaiian Volcanic Deposits†

Kari E. Dunfield and Gary M. King*

Darling Marine Center, University of Maine, Walpole, Maine 04573

Received 14 November 2003/Accepted 2 April 2004

Genomic DNA extracts from four sites at Kilauea Volcano were used as templates for PCR amplification of the large subunit (\textit{coxL}) of aerobic carbon monoxide dehydrogenase. The sites included a 42-year-old tephra deposit, a 108-year-old lava flow, a 212-year-old partially vegetated ash-and-tephra deposit, and an approximately 300-year-old forest. PCR primers amplified \textit{coxL} sequences from the OMP clade of CO oxidizers, which includes isolates such as \textit{Oligotropha carboxidovorans}, \textit{Mycobacterium tuberculosis}, and \textit{Pseudomonas thermocarboxydovorans}. PCR products were used to create clone libraries that provide the first insights into the diversity and phylogenetic affiliations of CO oxidizers in situ. On the basis of phylogenetic and statistical analyses, clone libraries for each site were distinct. Although some clone sequences were similar to \textit{coxL} sequences from known organisms, many sequences appeared to represent phylogenetic lineages not previously known to harbor CO oxidizers. On the basis of average nucleotide diversity and average pairwise difference, a forested site supported the most diverse CO-oxidizing populations, while an 1894 lava flow supported the least diverse populations. Neither parameter correlated with previous estimates of atmospheric CO uptake rates, but both parameters correlated positively with estimates of microbial biomass and respiration. Collectively, the results indicate that the CO oxidizer functional group associated with recent volcanic deposits of the remote Hawaiian Islands contains substantial and previously unsuspected diversity.

Aerobic carbon monoxide oxidation occurs ubiquitously, with significant activity reported for a wide range of aquatic and terrestrial systems, plant roots, and macroalgal tissues (10, 14, 20, 21, 26, 27). Nonetheless, past efforts to enrich CO oxidizers have resulted in only 12 species in 11 genera (12, 20, 21). More-recent efforts, including genome studies, have identified at least 25 CO-oxidizing species in 10 genera (12, 13). Although these results suggest that CO oxidizers may be as diverse as the habitats in which they are found, essentially nothing is known about in situ diversity.

Analyses of CO oxidizer diversity have been constrained by several factors. High CO concentrations traditionally used for enrichments inhibit growth of many newly recognized strains (12). Thus, simple enrichment strategies based on high CO concentrations may not be suitable for culture-based diversity surveys. The diversity of known CO oxidizers has also precluded development of signature 16S rDNA gene oligonucleotides analogous to those used for phylogenetically coherent groups, such as methanotrophs and ammonia oxidizers (3, 5, 7, 15, 31, 34).

However, primers developed for amplification of a portion of \textit{coxL}, the large subunit of carbon monoxide dehydrogenase (CODH), have provided a new tool for assessing CO oxidizer diversity (12). \textit{coxL} PCR primers amplify a 1,260- to 1290-bp fragment containing the CODH active site and sites for binding the molybdopterin cytosine dinucleotide (MCD) cofactor (9, 28). Two sets of primers have been developed and evaluated, one for each of two CO-oxidizing clades (12).

One clade, designated OMP, contains \textit{coxL} sequences derived from known CO oxidizers (e.g., \textit{Oligotropha carboxydovorans}, \textit{Mycobacterium tuberculosis}, and \textit{Pseudomonas thermocarboxydovorans}) and newly recognized strains (e.g., \textit{Mycobacterium smegmatis}, \textit{Silicibacter pomeroyi}, and \textit{Stenotrophomonas} strain LUP) (12). The second clade, designated BMS, contains putative CODH sequences derived primarily from newly recognized CO oxidizers (e.g., \textit{Burkholderia} strain LUP, \textit{Mesorhizobium loti}, and \textit{Sinorhizobium mellotii}) and from taxa that possess both BMS and OMP \textit{coxL} genes (\textit{Bradyrhizobium japonicum}, \textit{Burkholderia fungorum} LB400, and \textit{Stappia stellulata} [12]).

We report here analyses of OMP \textit{coxL} diversity based on PCR amplification of DNA extracts from volcanic deposits at four varied sites in and near the Kilauea Volcano caldera. Deposits at all sites consume CO actively, and CO contributes significantly to respiratory reducing equivalent flow at three of the four sites. Phylogenetic and statistical analyses of sequences from clone libraries indicate that distinct and complex CO-oxidizing populations occur at each site. Although some clone sequences could be associated with known CO oxidizers in proteobacterial and \textit{Firmicutes} lineages, most of the sequences appear to have been derived from novel organisms with uncertain phylogenetic affiliations. These results represent the first molecular ecological analyses of CO oxidizers and reveal a substantial and previously unsuspected diversity.

**MATERIALS AND METHODS**

**Site description and sampling.** Surface deposits (0 to 2 cm depth) of four sites in or near Kilauea Volcano caldera were sampled in April 2002. The sites range in age, water availability, plant community development, and organic contents, among other parameters (11). The Halema’uma’u and Caldera Rim sites occur in a region of the caldera with relatively limited water availability, based on water contents (11). Both sites consist of ash and tephra overlying weathered lava flows and support limited patchy growth of pioneer ferns, shrubs, and the Ohia tree.
**RESULTS**

Phylogeny of clone libraries. PCR products of the appropriate length (approximately 1,260 bp) were successfully amplified from volcanic deposit DNA. A total of 109 coxL clones were sequenced. BLAST analysis indicated that 108 of the clone sequences were most similar to OMP coxL genes, while one, HM-12, was most similar to a BMS putative coxL gene. A majority of Forest and Pu’u Pua’i clone sequences clustered with sequences from known CO oxidizers (Fig. 1). Sixty-two percent of Forest clones clustered with known \( \alpha \)-Proteobacteria sequences (e.g., *Bradyrhizobium japonicum* and *Oligotropha* carboxydivorans) and 15% clustered with known \( \beta \)-Proteobacteria (e.g., *Burkholderia fungorum* and *Pseudomonas thiooxidans*). Heterologous coverage, a measure of the representation of a collection of libraries. A bootstrap procedure was used to estimate the significance of differences in coverage versus evolutionary distance between libraries. Differences in coverage as a function of evolutionary distance \((\Delta c)\) were considered significant for \( P \) values of \( \leq 0.05 \).

Arlequin estimated the significance of differences in population pairwise fixation index \((F_{ST})\) values and average pairwise difference among libraries by using analysis of molecular variance (AMOVA), and the software calculated nucleotide diversity, average pairwise differences, and mismatch distributions for each library. \( F_{ST} \) values were derived from the genetic diversity within a library and the total diversity of a pair of pooled libraries (23, 29, 35). Nucleotide diversity was estimated from the number of variable positions for aligned sequences in a given library; average pairwise differences were estimated from comparisons within a library of the number of sequence differences between a given clone and all other clones; and mismatch distributions were derived for each library from the frequency of pairwise differences.

A \( P \)-test was used to determine the significance of covariation between the distribution of unique sequences within libraries and phylogeny (18, 22). Briefly, PAUP 4.0b was used to generate 1,000 random trees from the combined clone libraries. The lengths of maximum parsimony trees for the clones were compared by \( P \)-test to the 95% lower confidence limit of the random tree lengths obtained by maximum parsimony with a heuristic search algorithm.

**Nucleotide sequences accession numbers.** The sequence of the volcanic deposit identified in this study have been deposited in GenBank under accession numbers Y3463248 to Y346356.
FIG. 1. Neighbor-joining analysis (1,000 bootstrap replicates) of inferred amino acid sequences for nonredundant coxL clone sequences and known CO-oxidizing isolates. A total of 109 clone sequences (163 amino acid residues each) from the Forest (FOR), Pu’u Puai (PP), Halema’uma’u (HM), and the Caldera Rim (CR) libraries were analyzed. OMP and BMS clades are indicated based on known coxL gene sequences (13). Phylogenetic lineages are designated based on affiliations with known CO oxidizers. Bootstrap values of <70% are not shown. The Pyrobaculum aerophilum clade is highlighted in red, and the OMP clade is highlighted in blue.
those of the Caldera Rim and Pu’u Puai; the Halema’uma’u library was least diverse (Table 3). AMOVA comparisons revealed significant differences (P < 0.05) among libraries for pairwise fixation indices and average pairwise differences (Table 4). Both values were lowest when Caldera Rim and Halema’uma’u libraries were compared, followed by the Pu’u Puai and Forest comparison. The highest F_ST values and average pairwise differences were observed for comparison of the Halema’uma’u library with that of either Pu’u Puai or Forest.

Frequency distributions of sequence mismatches for Forest and Caldera Rim libraries were multimodal, or ragged (Fig. 2). In contrast, the mismatch distribution for Pu’u Puai was largely unimodal, with the greatest frequency of mismatches occurring between 0 and 6 mismatches (Fig. 2B). The Halema’uma’u library distribution was approximately bimodal, with a prominent peak at approximately 50 mismatches and a smaller secondary peak between 0 and 4 mismatches (Fig. 2B).

Plots of cumulative percentage of the total mismatch for each library revealed unique patterns, the most unique of which occurred for Pu’u Puai, where a large fraction of closely related clone sequences (<10 mismatches) accounted for nearly 70% of total cumulative mismatch (Fig. 3). The cumulative percentage of total mismatch increased approximately exponentially for the Forest library, with the majority of sequences having greater than 70 mismatches. The cumulative percent of total mismatch increased approximately sigmoidally with mismatch level in the Caldera Rim and Halema’uma’u libraries, with the majority of sequences showing greater than 50 mismatches.

**DISCUSSION**

Recent enrichment and genomic studies have shown that aerobic CO oxidation occurs among numerous lineages within the α-, β-, and γ-Proteobacteria and Actinobacteria and Firmicutes (12). However, the extent of CO oxidizer diversity in situ remains largely unknown, since the only information available to date has been limited to culture descriptions. Molecular studies have been precluded until recently due to the lack of suitable CODH or 16S rDNA primers. Nonetheless, insights from enrichments and cultures have indicated that CO oxidizer diversity may parallel that of the wide range of systems in which CO oxidation occurs.

Primers based on sequences derived from a variety of representative CO oxidizers (12) have been used to assess the presence and diversity in volcanic deposits of \( \text{coxL} \). In a clade (OMP) characterized by several well-known carboxydrotrophs (e.g., Oligotropha carboxidovorans and Pseudomonas thermocarboxydovorans) and a number of newly recognized CO oxidizers (12). In addition to documenting greater diversity than previously recognized in known lineages, results presented here also document substantial diversity in yet-unidentified bacterial lineages.

Since phylogenetic analyses of inferred amino acid sequences obtained from a variety of isolates have indicated that the \( \text{coxL} \) phylogeny is congruent with 16S rDNA phylogeny (12), it is possible tentatively to assign clones to major bacterial divisions and, in some cases, subdivisions (Fig. 1; Table 1). Accordingly, about 40% (44 of 109) of the clones overall cluster with Proteobacteria and Actinobacteria and Firmicutes. However, 60% of the clone sequences cannot be associated with bacterial divisions known to harbor CO oxidizers. This finding suggests that aerobic CO oxidation is more widespread among the bacterial domain than previously imagined and that CO oxidizers may occur in common terrestrial divisions such as

\[ \text{Staphylococcus aureus COX, AY307907 (BMS); Strain LUP, AY307900 (BMS) and AY307913 (OMP);} \]

\[ \text{Stappia strain M4, AY307902 (BMS) and AY307916 (OMP);} \]

\[ \text{Bradyrhizobium japonicum (str.) COX, AY307908 (BMS);} \]

\[ \text{Bradyrhizobium japonicum Soda USDA 6, AY307921 (BMS);} \]

\[ \text{Bradyrhizobium strain CPP, AY307900 (BMS) and AY307913 (OMP);} \]

\[ \text{Burkholderia fungorum LB400, AY307901 (BMS), AY307916 (OMP);} \]

\[ \text{Burkholderia strain LUP, AY307907 (BMS);} \]

\[ \text{Mesorhizobium strain NM1B, AY307906 (BMS);} \]

\[ \text{Mycobacterium smegmatis, AY307914 (OMP);} \]

\[ \text{Oligotropha carboxidovorans, X82447 (OMP);} \]

\[ \text{Pseudomonas thermocarboxydovorans, Y77931 (OMP);} \]

\[ \text{Stappia agregata, AY307904 (BMS) and AY307918 (OMP); Stappia strain KB12, AY307898 (BMS) and AY307193 (OMP); Stappia strain KB902, AY307899 (BMS), AY307914 (OMP);} \]

\[ \text{Staphylococcus aureus COX, AY307916 (OMP);} \]

\[ \text{Staphylococcus aureus COX, AY307916 (OMP); and} \]

\[ \text{Xanthobacter strain COX, AY307911 (BMS).} \]
Acidobacterium and Verrucomicrobium (1, 2, 6, 8, 16, 17, 19, 25), representatives of which have been identified in Hawaiian volcanic deposits by 16S rDNA sequence analysis (24; V. Gomez-Alvarez, G. M. King, and K. Nuesselein, unpublished data).

For the Forest and Pu‘u Puai libraries, modest numbers of the clones, 12 and 32%, respectively, clustered with unknown lineages. However, the majority of clones from the Caldera Rim (81%) and Halema‘uma‘u (96%) could not be assigned to a bacterial phylum. This result may reflect differences in successional state and various biological and abiological parameters that distinguish the Caldera Rim and Halema‘uma‘u samples from the Forest and Pu‘u Puai samples. In particular, the latter sites occur within a well-vegetated, relatively moist area of the caldera, while the former occur in a region with reduced water availability and substantially less plant growth (11).

It should also be noted that one of the Halema‘uma‘u clones, HM-12, clustered with BMS putative coxL sequences. This occurrence likely results from the similarity between OMP and BMS sequences in the regions targeted for primer development (12). Nonetheless, the rare occurrence of a BMS-like clone indicates that the primers used here are relatively specific for environmental OMP sequences across a broad phylogenetic range.

Patterns for nucleotide diversity and average pairwise difference generally agree with clone phylogeny. For instance, the lowest values for both parameters are obtained for the Halema‘uma‘u library (Table 3), which also has the most limited phylogenetic distances among clones, while the Forest site supports the most-divergent sequences and has the highest nucleotide diversity and average pairwise difference. However, it is also evident from results for the Pu‘u Puai and Caldera Rim libraries that similar levels of diversity can be obtained from libraries with substantially different phylogenetic compositions (Tables 1 and 3). Therefore, without additional information about population structure, such indices may have limited value.

Paired reciprocal comparisons performed using LIBSHUFF indicate that each of the libraries differs significantly from the others (Table 2), in general agreement with phylogenetic results. This result is true even for sites with similar nucleotide diversity and average pairwise difference values, e.g., Pu‘u Puai and the Caldera Rim (Tables 3 and 4). Thus, LIBSHUFF comparisons, which have thus far been used primarily for 16S rDNA sequence data, appear to provide useful insights for functional genes and to complement other analyses.

However, unlike the case for 16S rDNA sequence analyses, it is not yet evident what level of nucleotide similarity should be used for distinguishing coxL OTUs. For the study described herein, only 100% identical sequences were considered equivalent OTUs. This decision reflects the fact that essentially identical coxL sequences have been obtained from taxonomically distinct cultures (e.g., Mycobacterium bovis, M. tuberculosis, and M. microti) (13). Accordingly, LIBSHUFF data may underestimate the true level of library coverage.

The unique distribution of taxa among sites is supported by several other analyses. Fixation indices (FST) and corrected

---

**TABLE 3. Characteristics of coxL clone libraries**

<table>
<thead>
<tr>
<th>Site</th>
<th>No. of clones</th>
<th>No. of variable positions</th>
<th>Nucleotide diversity</th>
<th>θ(π)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halema‘uma‘u</td>
<td>31</td>
<td>334</td>
<td>0.28 (0.14)a</td>
<td>128.8 (63.2)a</td>
</tr>
<tr>
<td>Caldera Rim</td>
<td>27</td>
<td>339</td>
<td>0.37 (0.18)b</td>
<td>177.9 (87.6)b</td>
</tr>
<tr>
<td>Pu‘u Puai</td>
<td>25</td>
<td>355</td>
<td>0.38 (0.19)c</td>
<td>179.7 (88.7)c</td>
</tr>
<tr>
<td>Forest</td>
<td>26</td>
<td>429</td>
<td>0.46 (0.23)d</td>
<td>221.5 (109.1)d</td>
</tr>
</tbody>
</table>

*Variable positions indicate the number of variable nucleotide sites (total n = 488) within each library. Values with different letters (superscripts) are significantly different (comparisons with Halema‘uma‘u [P ≤ 0.05] and Forest [0.05 ≤ P ≤ 1] by a one-tailed t-test).*

---

**TABLE 4. Corrected average pairwise differences (θ[π], above diagonal) and pairwise fixation indices (FST, below diagonal)**

<table>
<thead>
<tr>
<th>Site</th>
<th>Halema‘uma‘u</th>
<th>Caldera Rim</th>
<th>Pu‘u Puai</th>
<th>Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halema‘uma‘u</td>
<td>11.2</td>
<td>94.1</td>
<td>98.9</td>
<td></td>
</tr>
<tr>
<td>Caldera Rim</td>
<td>0.067</td>
<td>57.2</td>
<td>72.3</td>
<td></td>
</tr>
<tr>
<td>Pu‘u Puai</td>
<td>0.371</td>
<td>0.236</td>
<td>37.6</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>0.357</td>
<td>0.261</td>
<td>0.153</td>
<td></td>
</tr>
</tbody>
</table>

*All values are significant (P ≤ 0.05).*

---

**FIG. 2. Frequency distribution of pairwise mismatches within coxL clone libraries. (A) Caldera Rim (●) and Halema‘uma‘u (○); (B) Forest (●) and Pu‘u Puai (○).**
average pairwise differences derived from paired library comparisons by AMOVA suggest that the most divergent populations occur in sites with distinct plant and water regimes (e.g., Halema‘uma‘u versus Pu‘u Puai or Forest; Table 4), while the least divergent populations occur in sites with more similar plant and water regimes (e.g., Forest and Pu‘u Puai or Halema‘uma‘u and Caldera Rim; Table 4). These observations agree with results from LIBSHUFF and a P-test (17), which show that sequence distributions within libraries covary significantly with phylogenetic position. Collectively, the various statistical analyses suggest that CO oxidizer population structure responds to environmental (moisture and plant growth) and historical (e.g., eruption and date of deposition) differences among sites located within a radius of less than 2.5 km.

Mismatch distributions also varied among libraries (Fig. 2). For the Forest and Caldera Rim samples, mismatch distributions are multimodal, reflecting a relatively even distribution of closely and distantly related taxa. In contrast, mismatch distributions for Pu‘u Puai and Halema‘uma‘u samples are approximately unimodal and bimodal, respectively. The shape of the mismatch curves for the latter sites may result from recent population expansions or constraints on diversity imposed by limited ecosystem complexity. Curves for the former sites may result from increased diversity accompanying greater deposit age and successional development (11, 33). Similar results have been obtained for analyses of the large subunit gene for ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL) (22).

Although distinct CO-oxidizing communities exist at each of the sites, variations among sites in coxL diversity indices (e.g., nucleotide diversity and average pairwise difference) do not correlate \( (r = 0.23) \) with variations in atmospheric CO uptake rates (11). Indeed, atmospheric CO uptake rates are lowest for samples from Forest, the site with the highest diversity. Similarly, patterns of rbcL diversity for the same sites do not correlate with CO and hydrogen uptake, even though these gases likely support most, if not all, of the facultative lithotrophic activity (22). Whether similar relationships exist between the diversity of other functional genes and their activities in situ is uncertain. However, the results presented here indicate that diversity and function may be uncoupled in at least some systems.

In contrast to the weak correlation between coxL diversity and CO uptake, both microbial biomass and respiration correlate strongly \( (r = 0.99; \text{Fig. 4}) \) with average pairwise difference (and nucleotide diversity). One explanation for this correlation is that ecosystem level changes that promote increased microbial abundance and activity also increase opportunities for niche differentiation and diversification of CO oxidizers. Changes in the availability and diversity of organic substrates with successional development may be particularly important since almost all CO oxidizers function preferentially as heterotrophs (20, 21).

Of course, additional analyses, e.g., reverse transcription-PCR, will be necessary to assess the diversity of populations that actively express coxL, and that may be involved with atmospheric CO consumption under in situ conditions. However, organic limitation in soils generally and volcanic deposits specifically (11) may select for numerous taxa that use CO as an energy supplement for growth with heterotrophic substrates.

In conclusion, phylogenetic analyses of coxL sequences derived from genomic extracts of recent Hawaiian volcanic deposits have revealed a substantial and previously unrecognized level of CO oxidizer diversity. Some clone sequences cluster with, yet are distinct from, sequences of known CO oxidizers in Proteobacteria, Actinobacteria, and Firmicutes; the majority of sequences occur in unidentified lineages. Phylogenetic and statistical analyses also show that distinct CO-oxidizing communities occur at sites differing in deposit age and ecosystem development, but differences in community structure are not correlated with estimates of atmospheric CO uptake. Remarkably, the level of CO oxidizer diversity observed in a limited regional scale sampling of a remote island appears to exceed reports of diversity for another major group of trace gas-utilizing bacteria, the methanotrophs, which have been assayed more extensively over a much broader range of habitats.
ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation (BIO-0085495). We thank K. Nanba for technical assistance and W. Gilmartin for use of Hale Mahana.

REFERENCES


