Louisiana State University
LSU Scholarly Repository

Faculty Publications | Department of Biological Sciences

9-1-2005

Analysis of the distribution and diversity in recent Hawaiian volcanic deposits of a putative carbon monoxide dehydrogenase large subunit gene

Kari E. Dunfield
University of Maine

Gary M. King
University of Maine

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

Recommended Citation

Dunfield, K., & King, G. (2005). Analysis of the distribution and diversity in recent Hawaiian volcanic deposits of a putative carbon monoxide dehydrogenase large subunit gene. *Environmental Microbiology, 7* (9), 1405-1412. https://doi.org/10.1111/j.1462-2920.2005.00827.x

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Environmental Microbiology (2005) 7(9), 1405–1412 doi:10.1111/j.1462-2920.2005.00827.x

Analysis of the distribution and diversity in recent Hawaiian volcanic deposits of a putative carbon monoxide dehydrogenase large subunit gene

Kari E. Dunfield and Gary M. King*
Darling Marine Center, University of Maine, 193 Clark's Cove Road, Walpole, ME 04573, USA.

Summary

A putative carbon monoxide dehydrogenase large subunit gene (BMS putative *cox*L) was amplified from genomic DNA extracts of four recent (42–300 year old) Hawaiian volcanic deposits by polymerase chain reaction (PCR). Sequence databases derived from clone libraries constructed using PCR products were analysed phylogenetically and statistically. These analyses indicated that each of the deposits supported distinct BMS putative *cox*L gene assemblages. Statistical analyses also showed that the youngest deposit (42 years old) contained the least diverse sequences ($P < 0.05$), but that diversity did not vary significantly among three older deposits with ages from about 108–300 years. Although diversity indices did not vary among the older deposits, mismatch analyses suggested population structures increased in complexity with increasing deposit age. At each of the sites, most of the clone sequences appeared to originate from *Proteobacteria* not currently represented in culture or recognized as CO oxidizers.

Introduction

Previous assays of recent Hawaiian volcanic deposits have revealed that atmospheric CO and hydrogen contribute significantly to microbial respiratory energy metabolism on a series of sites ranging from about 42 to 210 years old (King, 2003a). CO- and hydrogen-oxidizing communities at several of these sites have been characterized using molecular approaches based on genes for the large subunits of carbon monoxide dehydrogenase (*cox*L; Dunfield and King, 2004) and ribulose-1,5-bisphosphate carboxylase/oxygenase [*rbc*L (also *cbb*L); Nanba *et al*., 2004]. Results show that the various sites harbour distinct communities (Dunfield and King, 2004), and that several diversity indices vary with microbial biomass and respiration (CO_2 production). Results also indicate that a large fraction of the populations represented in clone libraries appear to be derived from phylogenetically novel lineages (Dunfield and King, 2004).

Initial studies of CO oxidizer distribution and diversity were conducted using primers that amplify 'form I' or 'OMP' *cox*L sequences (King, 2003b,c). These sequences correspond to those of *Oligotropha carboxydovorans*, *Mycobacterium tuberculosis*, *Pseudomonas thermocarboxydovorans* and other 'classic' carboxydotrophs (Meyer and Schlegel, 1983; Morsdorf *et al*., 1992; King, 2003b,c). Primers have also been developed for a second, putative *cox*L gene designated 'form II' or 'BMS'. This second gene occurs in some CO oxidizers that possess OMP *cox*L (e.g. *Burkholderia fungorum* LB400 and *Zavarzinia compransoris*; King, 2003b). It also occurs in numerous α-*Proteobacteria*, especially the *Rhizobiaceae*, which oxidize but do not typically grow on CO, and that do not appear to possess OMP *cox* genes (e.g. *Bradyrhizobium japonicum* USDA 110, *Mesorhizobium loti* and *Sinorhizobium meliloti*; King, 2003b). Several lines of evidence indicate that BMS putative *cox*L genes code for functional CO dehydrogenases. For example, the genomic sequence of *B. japonicum* USDA 110 contains a BMS but not OMP *cox*L gene (King, 2003b). The BMS gene codes for a protein previously isolated as a subunit of an active CO dehydrogenase that promotes growth of *B. japonicum* USDA 110 on CO (M. Lorite, pers. comm.; Lorite *et al*., 2000).

We report here results from analyses of BMS putative *cox*L distribution and sequence diversity based on polymerase chain reaction (PCR) amplification of genomic DNA extracts used previously for OMP *cox*L analysis (Dunfield and King, 2004). Although logistical constrains on the sampling regime and clone library size do not allow a full exploration of diversity at the sites (e.g. Dunbar *et al*., 2002), statistical analyses of the sequence database indicate that each of four sites supports distinct populations. Unlike results from OMP analyses, however, (Dunfield and King, 2004), molecular diversity within libraries (e.g. average pairwise difference) does not correlate with atmospheric CO uptake rates, microbial biomass or respiratory

Received 21 June, 2004; accepted 8 November, 2004. *For correspondence. E-mail gking@maine.edu; Tel. (+1) 207 563 3146 ext. 207; Fax (+1) 207 563 3119.

© 2005 Society for Applied Microbiology and Blackwell Publishing Ltd

activity. This indicates that environmental parameters likely shape differentially populations containing these two genes. Sequence results also indicate that the BMS putative *cox*L clade encompasses substantial diversity, most of which occurs within the α-P*roteobacteria* on recent volcanic deposits.

Results and discussion

A total of 111 sequences obtained from cloned PCR products were similar to known *cox*L genes based on BLAST analyses. Phylogenetic analyses performed using a Neighbour-Joining algorithm with 1000 bootstrap replicates and Jukes-Cantor correction indicated that 108 sequences clustered in a clade containing previously recognized BMS putative *cox*L genes, including those of *Bu. fungorum* LB400, *M. loti*, *Silicibacter pomeroyi* and *Xanthobacter* sp. str. COX (Fig. 1; King, 2003b). Branching patterns were generally supported strongly by bootstrap values > 70%.

Three Halema'uma'u sequences (HM-8, HM-41, HM-42) were distinct from the other *cox*L sequences (Fig. 1), and most similar to the quinoline 2-oxidoreductase sequence (*qor*L) from *Pseudomonas putida*. Like *cox*L, *qor*L, codes for the large subunit of a molybdenum hydroxylase with binding sites for molybdopterin cytosine dinucleotide (Hille, 1996). As these binding sites contain conserved sequences that have been used in designing BMS *cox*L primers, some non-*cox*L sequences may be obtained occasionally in BMS *cox*L clone libraries. However, non-*cox* sequences are phylogenetically distinct and can be readily distinguished from *cox* sequences. Accordingly, the three *qor*L-like sequences from Halema'uma'u were not included in the final statistical analyses and library comparisons. The presence or absence of these sequences did not affect the topology of the remaining *cox*L sequences in a Neighbour-Joining tree (Fig. 1).

BMS putative *cox*L diversity was statistically least ($P < 0.05$) for the youngest site, Pu'u Puai, based on within library diversity indices: nucleotide diversity and average pairwise difference (0.18–0.32 substitutions base^{-1} and 171.5–290.4 differences respectively; Table 1). These parameters did not differ statistically for the remaining sites, and did not correlate (not shown) with previously reported CO_2 production rates (respiration), atmospheric CO or hydrogen uptake rates, or estimates of microbial biomass (King, 2003a).

Within library BMS putative *cox*L diversity estimates differ from those obtained from clone libraries based on OMP *cox*L (Dunfield and King, 2004). For the later, nucleotide diversities and average pairwise differences correlated strongly with microbial biomass and respiration and were least for Halema'uma'u. The lack of a similar pattern for BMS putative *cox*L suggests that populations harbouring BMS and OMP genes respond to different environmental parameters. Differential responses likely reflect the fact that BMS putative *cox*L has only been obtained from proteobacterial isolates thus far, while OMP *cox*L occurs across a broad phylogenetic range including *Proteobacteria*, *Firmicutes*, *Actinobacteria* and likely other lineages (King, 2003a; Dunfield and King, 2004). Known physiological and ecological characteristics of these groups suggest that they may respond differentially to numerous parameters, including organic matter, nitrogen availability, water availability, and pH. For example, molecular analyses have led to the conclusion that *Proteobacteria* tend to respond to increased organic matter availability, while *Acidobacterium* appear more abundant in soils with poorer nutritional status (Smit *et al.*, 2001; Torsvik and Øverås, 2002).

Although nucleotide diversities and average pairwise differences were similar for Forest, Halema'uma'u and Caldera Rim libraries, compositions of these libraries and of Pu'u Puai differed. AMOVA indicated a modest but statistically significant level of differentiation among libraries ($P < 0.05$), with an F_{ST} value of 0.15 overall (Table 2), where F_{ST} provides a measure of within to among library variation (Wright, 1951; Neigel, 2002; Weckstein *et al.*, 2002). F_{ST} values were significant at $P < 0.05$ for all pairwise library comparisons (Table 3), with the lowest values occurring for comparisons among Halema'uma'u, Caldera Rim and Forest libraries ($F_{ST} = 0.06$–0.08) and the greatest values occurring in comparisons with Pu'u Puai ($F_{ST} = 0.21$–0.22).

AMOVA analyses of data sets constructed using isolate sequences with known phylogenetic relationships provide a context for interpreting F_{ST}. Two phylogenetically distinct sequence data sets, one each with OMP *cox*L or BMS

Table 1. Characteristics of putative *cox*L clone libraries sequences for genomic extracts from recent volcanic deposits.

Site	Age	Clones	Variable bases	Nucleotide diversity	θ(π)
Halema'uma'u (I-D)	108	18	795	0.32 (0.16)[a]	290.4 (145.7)[a]
Caldera Rim (I-E)	212	28	680	0.28 (0.13)[a]	256.1 (125.7)[a]
Pu'u Puai (II-B)	43	31	612	0.18 (0.09)[b]	171.5 (84.0)[b]
Forest (II-C)	*c.* 300	34	787	0.29 (0.14)[a]	269.1 (131.2)[a]

Values with different letters are significantly different ($P \leq 0.05$) with a one-tailed *t*-test.
Variable positions represent numbers of variable nucleotide sites (total = 982) within each library.

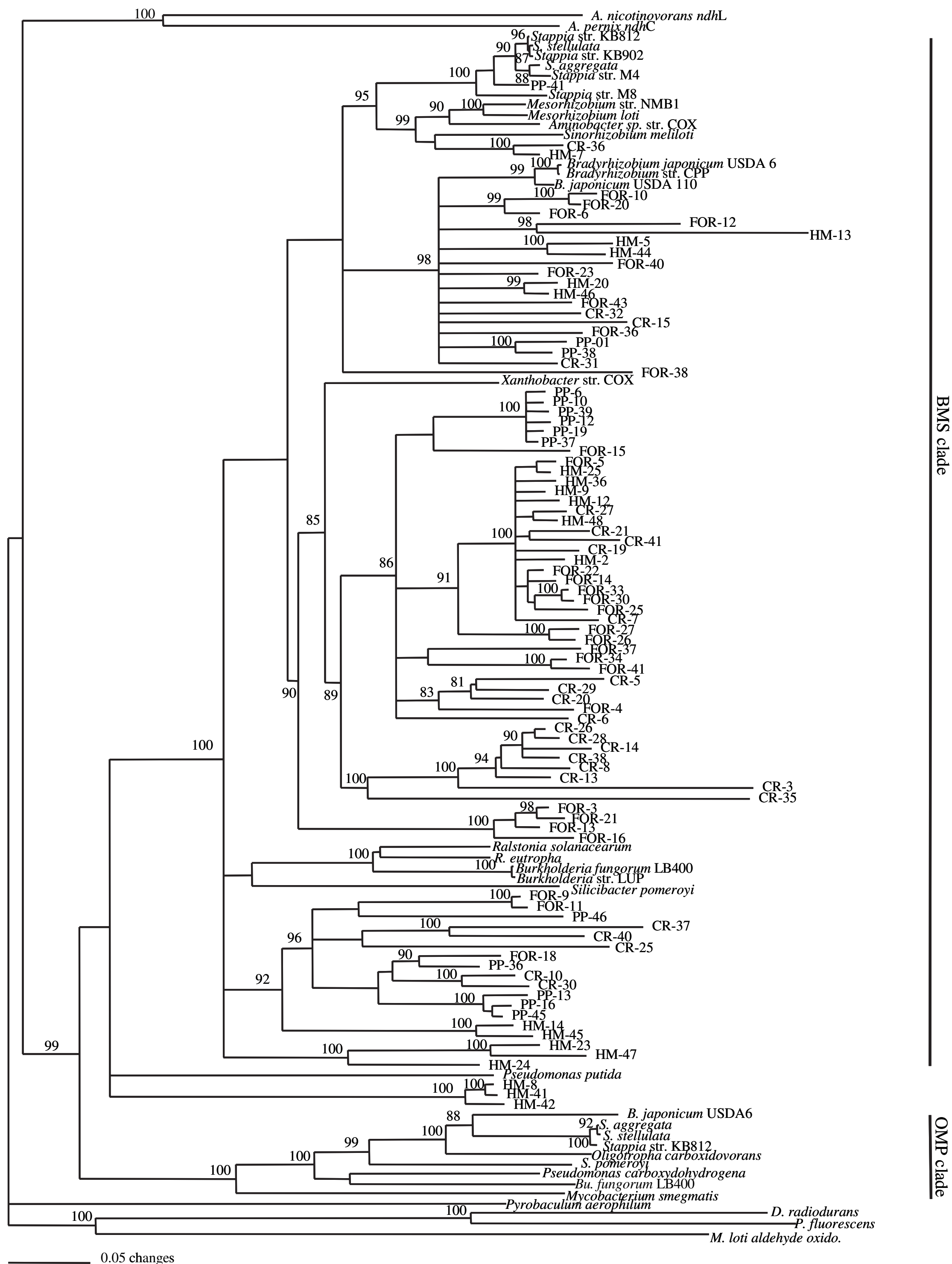

Fig. 1. Neighbour-Joining analysis (1000 bootstrap replicates) of inferred amino acid sequences for authentic and putative *cox*L clone sequences, known CO oxidizing isolates, and inferred amino acid sequences of selected molybdenum hydroxylases: *Aeropyrum pernix* nicotine dehydrogenase, chain C, accession number NP148464; *Arthrobacter nicorinovorans* nicotine dehydrogenase, large subunit, AAK6423; *Deinococcus radiodurans* oxidoreductase, 285554; *M. loti* aldehyde oxidoreductase, NP105651; *Pseudomonas fluorescens* oxidoreductase, COG1319 (contig309gene7610 JGI), and *Pseudomonas putida* quinoline 2-oxidoreductase, CAA66830. 370 amino acids from clone sequences from Forest (FOR), Pu'u Puai (PP), Halema'uma'u (HM), and Caldera Rim (CR) libraries were analysed. Authentic *cox*L gene sequences are indicated in italics, and are differentiated as OMP or BMS gene clades. Clone sequence accession numbers are AY805425-AY805538. Phylogenetic lineages are designated based on affiliation with known CO oxidizers. Bootstrap values < 70% are not shown. *Pyrobaculum aerophilum* putative *cox*L sequence was used as an out-group.

Table 2. Analysis of molecular variance for Pu'u Puai, Halema'uma'u, Caldera Rim and Forest clone libraries as implemented by Arlequin with a Jukes and Cantor distance correction (Schneider *et al*., 2000).

Variation	d.f.	Sum of squares	Variance	% of variation
Among populations	3	2178.3	21.3 V_a	14.46
Within populations	110	13841.9	125.8 V_b	85.54
Total	113	16020.2	147.1	

F_{ST} estimated for population differentiation = 0.147, $P < 0.01$.

putative *cox*L, yielded an F_{ST} of 0.56, while two phylogenetically indistinct data sets comprised of mixed OMP and BMS sequences yielded $F_{ST} = 0.0$. The pooled distinct and indistinct data sets resulted in an F_{ST} of 0.17. Accordingly, the Pu'u Puai clone library appears well differentiated from the other libraries and likely represents an assemblage of distinct populations. In contrast, Forest, Halema'uma'u and Caldera Rim libraries appear less strongly differentiated, most likely resulting from a set of shared phylogenetically similar taxa.

Mismatch analyses also support differentiation among sites. A strongly unimodal peak was evident for the Pu'u Puai mismatch distribution, while the distribution for Halema'uma'u formed a somewhat more dispersed or damped unimodal pattern (Fig. 2). In contrast, the mismatch distribution for the Forest library was more evenly distributed, with 1–2 observations for mismatches from 1 to 100. Caldera Rim mismatch approximated a bimodal distribution with a cluster of observations at low levels of mismatch and a cluster at intermediate levels (Fig. 2). Plots of the cumulative percentage of total mismatch as a function of mismatch level revealed distinct patterns corresponding to the underlying mismatch distributions. For Pu'u Puai and Halema'uma'u, cumulative percentage of mismatch increased sigmoidally; values also increased sigmoidally but in two phases for Caldera Rim (Fig. 3). Conversely, an approximately linear trend was observed for the Forest site library.

Mismatch analyses suggest that BMS putative *cox*L phylogenetic complexity increases with successional age. Unimodal mismatch frequency distributions occurring at relatively low mismatch levels indicate that the two youngest sites, Pu'u Puai and Halema'uma'u, contain the least complex assemblages (Fig. 2). Increasing complexity for the second oldest site, Caldera Rim, and the oldest site, Forest, are supported by a bimodal mismatch frequency distribution for the former and a relatively even mismatch distribution across a broad range of mismatch for the latter.

Trends for BMS putative *cox*L are consistent with results for *rbc*L and OMP *cox*L libraries from the same sites (Dunfield and King, 2004; Nanba *et al*., 2004). In both cases, increasing clone library complexity was associated with increasing successional age and ecosystem development. Similar trends have also been obtained from analyses of other functional groups, e.g. ammonia oxidizers in cultivated and uncultivated soils (Bruns *et al*., 1999), and from analyses of 16S rRNA gene sequences obtained from disturbed and undisturbed terrestrial systems (e.g. Borneman and Triplett, 1997; Kozdroj and van Elsas, 2001; Smit *et al*., 2001; Torsvik and Øverås, 2002). Results from Hawaiian volcanic deposits are distinct, however, in that they reflect the outcome of succession *ab initio*, rather than responses of existing systems to perturbations.

A final analysis, a '*P*-test' (Martin, 2002), compared the mean tree length of 1000 random trees generated from the clone libraries with the length of the most parsimonious tree obtained from a heuristic search using PAUP (4.0b). The latter was significantly less than the 95% confidence limit of the former ($P < 0.05$), and indicated that the sequence distribution within libraries covaried with phylogeny, as has been observed for *rbc*L and OMP *cox*L analyses (Dunfield and King, 2004; Nanba *et al*., 2004).

Several lines of evidence strongly suggest that BMS putative *cox*L genes analysed in this study code for a functionally active form of the large subunit of CO dehydrogenase, and not molybdenum hydroxylases with

Table 3. Corrected average pairwise differences [(π), above diagonal] and pairwise fixation indices (F_{ST}, below diagonal).

	Halema'uma'u	Caldera Rim	Pu'u Puai	Forest
Halema'uma'u	–	25.39	62.34	17.67
Caldera Rim	0.077	–	57.36	19.68
Pu'u Puai	0.220	0.214	–	63.69
Forest	0.050	0.069	0.222	–

All values are significant ($P \leq 0.05$).

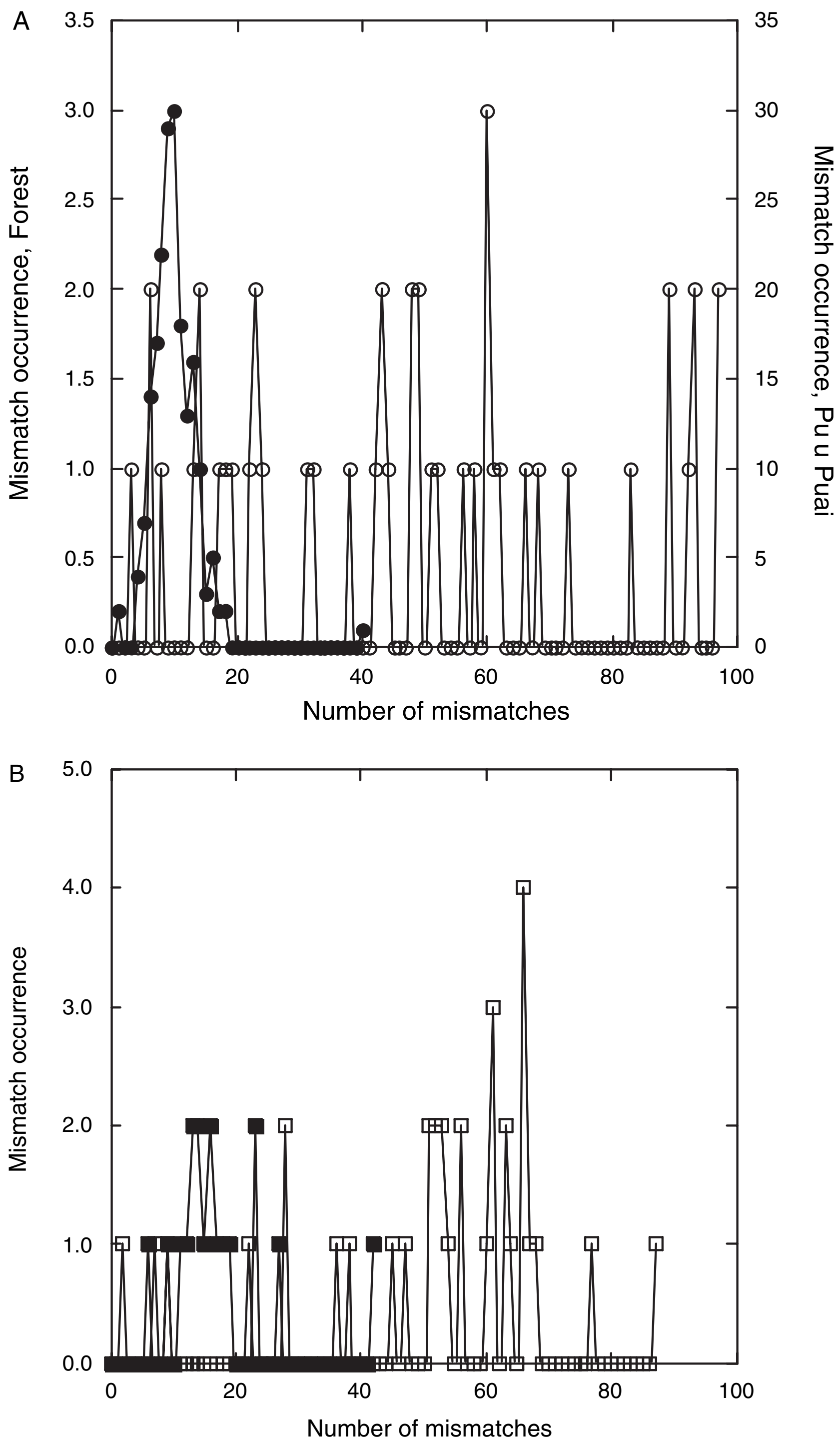

Fig. 2. A. Frequency distribution of pairwise mismatches within *cox*L clone libraries for Pu'u Puai (○) and Forest (●).
B. Frequency distributions for Halema'uma'u (■) and Caldera Rim (□).

unknown function. Perhaps the most significant supporting observations have been derived from *B. japonicum* strain USDA 110, which grows with CO as a sole carbon and energy source (Lorite *et al.*, 2000). The genome sequence of this strain (http://www.kazusa.or.jp/rhizobase/Bradyrhizobium) does not contain OMP (form I) *cox*L; instead, it contains only the BMS (form II) gene that is characterized by a distinct active site motif (Kang and Kim, 1999; Santiago *et al.*, 1999; King, 2003b). In addition, the N-terminal amino acid sequence of *B. japonicum* USDA 110 CO dehydrogenase corresponds identically to a translated genomic sequence, gene blr0336, annotated as a putative large subunit CO dehydrogenase gene (Lorite *et al.*, 2000; Lorite, pers. comm.). The sequence

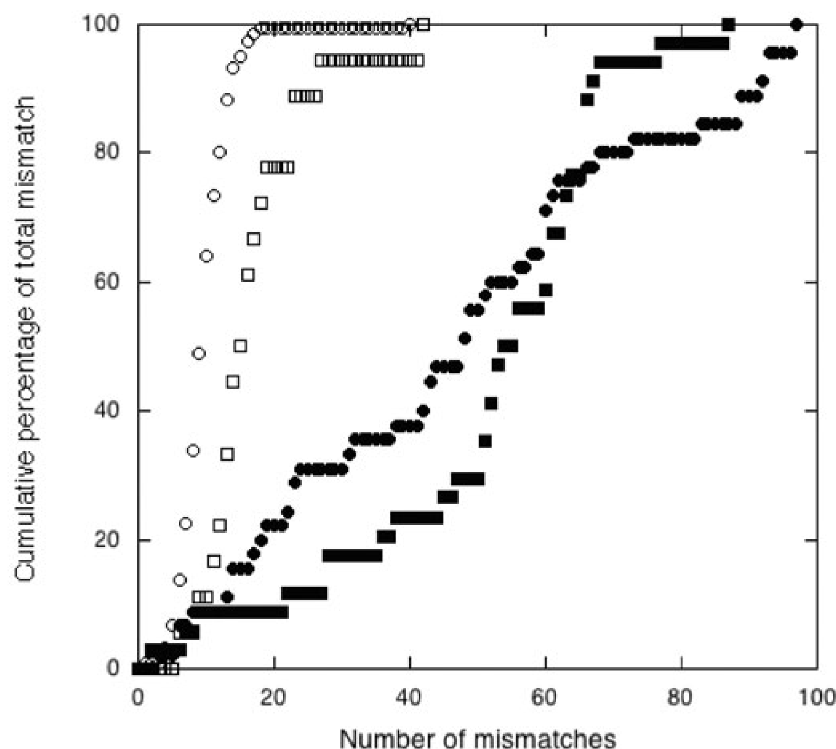

Fig. 3. Cumulative distribution of mismatch as a function of mismatch number for Pu'u Puai (○),Forest (●), Halema'uma'u (□) and Caldera Rim (■).

for blr0336 clusters with other BMS putative *cox*L genes (King, 2003b), all of which are distinct phylogenetically from non-*cox* molybdenum dehydrogenase gene sequences.

Thus, data presented here most likely represent the distribution and diversity of large subunit CO dehydrogenase genes. As such, they complement and expand insights derived from OMP *cox*L about the structure and diversity of CO-oxidizing communities colonizing recent volcanic deposits (Dunfield and King, 2004). Unlike OMP *cox*L libraries, however, the BMS libraries provide a higher degree of resolution for the *Proteobacteria*, since BMS genes have not been reported for isolates from other phyla. Although BMS *cox*L encompasses a more limited spectrum of phylogenetic diversity than OMP *cox*L, libraries for each gene type at each site were characterized by similar levels of nucleotide diversity and average pairwise differences (Tables 1 and 3, Dunfield and King, 2004).

Because some proteobacterial CO oxidizers possess both OMP and BMS *cox* genes (King, 2003b), some of the sequences identified in this study may have been derived from taxa represented in an earlier report of OMP *cox*L libraries (Dunfield and King, 2004). Obviously, the extent to which overlap occurs in the sequence data sets diminishes estimates of total CO oxidizer diversity. Nonetheless, because relatively few OMP sequences in a proteobacterial clade were obtained from analyses of Pu'u Puai, Halema'uma'u or Caldera Rim (Dunfield and King, 2004), most, if not all, BMS sequences from these sites were likely derived from distinct taxa. The greatest potential for overlap exists for the Forest site, for which 77% of the OMP sequences were assigned to the *Proteobacteria*. Although Forest site BMS and OMP phylogenetic structures suggest some distinct differences in clone distribution among the *Proteobacteria*, it is not possible determine which, if any, sequences might belong to the same taxa.

Collectively, results presented here and by Dunfield and King (2004) indicate that the trait for CO oxidation is widespread phylogenetically and occurs in numerous taxa within the *Proteobacteria*. Indeed, based on trends for four remote Hawaiian volcanic deposits, CO oxidizers may prove to be one of the most diverse functional groups that consume or produce trace gases. Future analyses of *cox*L expression using reverse transcriptase PCR will provide insights about the phylogenetic affiliations of this group and differences among sites in populations actively involved in CO dynamics *in situ.*

Experimental procedures

Genomic DNA was extracted from surface deposits (0–2 cm) obtained from four sites in or on Kilauea Volcano: Halema'uma'u (HM); Caldera Rim (CR); Pu'u Puai (PP); and Forest (For). Various aspects of these and similar sites have been reported previously (Nüsslein and Tiedje, 1998; King, 2003a; Dunfield and King, 2004; Nanba *et al.*, 2004). DNA

was extracted from triplicate 10 *g* fresh weight (gfw) samples using a bead-beating method (Dunfield and King, 2004) with an Ultraclean Mega Soil DNA extraction kit (MoBio Laboratories Carlsbad, CA).

Initial PCR reactions were based on primers described by King (2003b,c): BMSf (5′-GGCGGCTT[C/T]GG[C/G]TC[C/G]AAGAT-3′) and O/Br (5′-[C/T]TCGA[T/C]GATCATCGG[A/G]TTGA-3′). All PCR reactions contained standard concentrations of buffer, deoxynucleoside triphosphates (100 μM), primers (0.1 μM) and magnesium ion (3.5 mM). Template concentrations for each site were adjusted to optimize product yields. 'Hot start' PCR was performed in an Eppendorf Mastercycler thermocycler (Brinkmann, Westbury, NY). Triplicate reactions were prepared for each site, and PCR products were visualized with gel electrophoresis and ethidium bromide, and purified using a MoBio (Carlsbad, CA) spin-bind kit. Purified products were used in a subsequent round of PCR based on a forward primer, ASf (5′-GT[C/G]GA[C/T]GCCTATCG[C/T]GG-3′), which targeted the putative *cox*L active site motif, AYRGAGR (King, 2003b). O/Br (above) was used as a reverse primer. This nested design reduced non-specific products and resulted in a 950- to 980-base pair fragment.

For each volcanic deposit, triplicate PCR reactions containing products of the appropriate length were combined and cloned using a TOPO 'TA' cloning kit for sequencing according to the manufacturer's instructions (Invitrogen Life Technologies, Carlsbad, CA). Plasmids were extracted from arbitrarily chosen colonies using a PerfectPrep plasmid extraction kit (Brinkmann, Westbury, NY). One hundred 11 (111) putative *cox*L clones were sequenced by the University of Maine DNA Sequencing facility. Clone sequences identified as BMS putative *cox*L genes by GenBank's BLAST utility were aligned using Clustal X with manual adjustments as necessary. Clone sequences have been deposited in GenBank as AY805425–AY805538.

Phylogenetic and statistical analyses of the resulting sequences were conducted as described by Dunfield and King (2004). Arlequin (Schneider *et al*., 2000) was used for AMOVA estimates of fixation indices (F_{ST}), nucleotide diversities, average pairwise differences and mismatch distributions, and PAUP v.4.0b (Swofford, 2003) was used for phylogenetic analyses. F_{ST} values were calculated from the ratio genetic diversity within a library and the total diversity of pooled libraries (Schneider *et al*., 2000). Nucleotide diversities for aligned sequences in each library were calculated from the number of variable bases at each position; average pairwise differences were estimated from comparisons within a library of the number of sequence differences between a given clone and all other clones; mismatch distributions were derived for each library from the frequency of pairwise differences (Schneider *et al*., 2000).

In addition, two sequence data sets were constructed, one each using aligned BMS putative *cox*L derived from 24 isolates and OMP *cox*L from 20 isolates (e.g. King, 2003b,c; Dunfield and King, 2004). The BMS library consisted of the following sequences (sequence number and accession): *Aminobacter* str. COX (1; AY307907), *B. japonicum* USDA6 (2; AY307909), *Bradyrhizobium* str. CPP (3; AY307900), *Bu. fungorum* LB400 (4; AY307901), *Burkholderia* str LUP (5; AY307906), *Burkholderia* str. HI1 (6), *Carbophilus carboxydus* (7), *Mesorhizobium* str. NMB1 (8; AY307905), *Mesorhizobium amorphae* (9), *Rhizobium ciceri* (10), *Rhizobium etli* (11), *Rhizobium galegae* (12); *Rhizobium tropici* IIB (13), *Sinorhizobium fredi* (14), *Sinorhizobium kostiense* (15), *Stappia aggregata* (16; AY307904), *Sinorhizobium stellulata* (17; AY307908), *Stappia* str. 812 (18; AY307898), *Stappia* str. MIO (19), *Stappia* str. 902 (20; AY307899), *Stappia* str. M4 (21; AY307902), *Stappia* str. M8 (22; AY307903), *Sulfitobacter* str. P10 (23), *Xanthobacter* str COX (24; AY307910). The OMP library consisted of (sequence number and accession): *B. japonicum* USDA6 (25; AY307921), *Bradyrhizobium* str. CPP (26; AY307911), *Bu. fungorum* LB400 (27; AY307914); *Hydrogenophaga pseudoflava* (28; U80806), *Mycobacterium gordonae* (29; AY333109), *Mycobacterium microti* (30; AY333107), *Mycobacterium marinum* (31; AY333108), *Mycobacterium* str. HI1 (32), *Mycobacterium smegmatis* (33; AY307917), *Serratia* str. Uzon1(34), *St. aggregata* (35; AY307918), *S. stellulata* (36; AY307919), *Stappia* str. 812 (37; AY307912), *Stappia* str. MIO (38), *Stappia* str. 902 (39; AY307913), *Stappia* str. M4 (40; AY307915), *Stappia* str. M8 (41; AY307916); *Stappia* str. G15 (42), *Stenotrophomonas* str. COX (43; AY307920), *Z. compransoris* (44).

Two other data sets were constructed, each containing 12 BMS and 10 OMP sequences arbitrarily chosen from the OMP and BMS data sets (Mix1: sequences 4, 7, 8, 9, 10, 12, 13, 17, 18, 20, 21, 24, 27, 28, 29, 32, 33, 35, 39, 40, 41, 43; Mix2: sequences 1, 2, 3, 5, 6, 11, 14, 15, 26, 19, 22, 23, 25, 26, 30, 31, 34, 36, 37, 38, 42, 44). Three separate AMOVA analyses were conducted using the following combinations: the OMP and BMS data sets, the two mixed OMP and BMS data sets and the four data sets respectively. These analyses were equivalent to comparisons of two phylogenetically distinct libraries, two phylogenetically similar libraries, and a combination of two distinct and two indistinct libraries. The results provided a context for interpreting comparisons of the four field site clone libraries.

Acknowledgements

This work was supported in part by funds from the NSF LExEn Program (BIO-0085495) and the Clare S. Darling endowment. Ms K. Roache provided technical support; Drs K. Nanba and C. Giardina assisted with fieldwork. Dr K. Nüsslein and Mr V. Gomez provided helpful feedback and insights. Mr W. Gilmartin permitted Hale Mahana to be used as a field lab and staging centre, and Hawaii Volcanoes National Park provided access to sites.

References

Borneman, J., and Triplett, E.W. (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol* **63:** 2647–2653.

Bruns, M.A., Stephen, J.R., Kowalchuk, G.A., Prosser, J.I., and Paul, E.A. (1999) Comparative diversity of ammonia oxidizer 16S rRNA gene sequences in native, tilled, and successional soils. *Appl Environ Microbiol* **65:** 2994–3000.

Dunbar, J., Barns, S.M., Ticknor, L.O., and Kuske, C.R.

(2002) Empirical and theoretical bacterial diversity in four Arizona soils. *Appl Environ Microbiol* **68:** 3035–3045.

Dunfield, K.E., and King, G.M. (2004) Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposit. *Appl Environ Microbiol* **70:** 4242–4248.

Hille, R. (1996) The mononuclear molybdenum enzymes. *Chem Rev* **96:** 2757–2816.

Kang, B.S., and Kim, Y.M. (1999) Cloning and molecular characterization of the genes for carbon monoxide dehydrogenase and localization of molybdopterin, flavin adenine dinucleotide, and iron-sulfur centers in the enzyme of *Hydrogenophaga pseudoflava*. *J Bacteriol* **181:** 5581–5590.

King, G.M. (2003a) Contributions of atmospheric CO and hydrogen uptake to microbial dynamics on recent Hawaiian volcanic deposits. *Appl Environ Microbiol* **69:** 4067–4075.

King, G.M. (2003b) Molecular and culture-based analyses of aerobic carbon monoxide oxidizer diversity. *Appl Environ Microbiol* **69:** 7257–7265.

King, G.M. (2003c) Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl Environ Microbiol* **69:** 7266–7272.

Kozdroj, J., and van Elsas, J.D. (2001) Structural diversity of microorganisms in chemically perturbed soil assessed by molecular and cytochemical approaches. *J Microbiol Methods* **43:** 197–212.

Lorite, M.J., Tachil, J., Sanjuan, J., Meyer, O., and Bedmar, E.J. (2000) Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Appl Environ Microbiol* **66:** 1871–1876.

Martin, A.P. (2002) Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl Environ Microbiol* **68:** 3673–3682.

Meyer, O., and Schlegel, H.G. (1983) Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu Rev Microbiol* **37:** 277–310.

Morsdorf, G., Frunzke, K., Gadkari, D., and Meyer, O. (1992) Microbial growth on carbon monoxide. *Biodegradation* **3:** 61–82.

Nanba, K., King, G.M., and Dunfield, K. (2004) Analysis of facultative lithotrophs distribution and diversity on volcanic deposits using the large sub-unit of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Appl Environ Microbiol* **70:** 2245–2253.

Neigel, J.E. (2002) Is F_{ST} obsolete? *Conserv Genet* **3:** 167–173.

Nüsslein, K., and Tiedje, J.M. (1998) Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition. *Appl Environ Microbiol* **64:** 1283–1289.

Santiago, B., Schubel, U., Egelseer, C., and Meyer, O. (1999) Sequence analysis, characterization and CO-specific transcription of the *cox* gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. *Gene* **236:** 115–124.

Schneider, S., Roesseli, D., and Excoffier, L. (2000) *Arlequin, Version 2.000, a Software for Population Genetics Data Analysis.* University of Geneva, Switzerland: Genetics and Biometry Laboratory.

Smit, E., Leeflang, P., Gommans, S., van den Broek, J., and Wernars, K. (2001) Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl Environ Microbiol* **67:** 2284–2291.

Swofford, D.L. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Version 4. Sunderland, MA, USA: Sinauer Associates.

Torsvik, V., and Øverås, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5:** 240–245.

Weckstein, J.D., Afton, A.D., Zink, R.M., and Alisauskas, R.T. (2002) Hybridization and population subdivision within and between Ross's Geese and Lesser Snow Geese: a molecular perspective. *Condor* **104:** 432–436.

Wright, S. (1951) The genetic structure of populations. *Ann Eugenics* **15:** 323–354.