Intracellular Screen To Identify Metagenomic Clones That Induce or Inhibit a Quorum-Sensing Biosensor

Lynn L. Williamson,1,2 Bradley R. Borlee,2 Patrick D. Schloss,2 Changhui Guan,2 Heather K. Allen,2,3 and Jo Handelsman2*

Department of Bacteriology,1 Department of Plant Pathology,2 and Microbiology Doctoral Training Program,3 University of Wisconsin—Madison, Madison, Wisconsin 53706

Received 7 February 2005/Accepted 18 May 2005

The goal of this study was to design and evaluate a rapid screen to identify metagenomic clones that produce biologically active small molecules. We built metagenomic libraries with DNA from soil on the floodplain of the Tanana River in Alaska. We extracted DNA directly from the soil and cloned it into fosmid and bacterial artificial chromosome vectors, constructing eight metagenomic libraries that contain 53,000 clones with inserts ranging from 1 to 190 kb. To identify clones of interest, we designed a high throughput “intracellular” screen, designated METREX, in which metagenomic DNA is in a host cell containing a biosensor for compounds that induce bacterial quorum sensing. If the metagenomic clone produces a quorum-sensing inducer, the cell produces green fluorescent protein (GFP) and can be identified by fluorescence microscopy or captured by fluorescence-activated cell sorting. Our initial screen identified 11 clones that induce and two that inhibit expression of GFP. The intracellular screen detected quorum-sensing inducers among metagenomic clones that a traditional overlay screen would not. One inducing clone carries a LuxI homologue that directs the synthesis of an N-acyl homoserine lactone quorum-sensing signal molecule. The LuxI homologue has 62% amino acid sequence identity to its closest match in GenBank, AmfI from Pseudomonas fluorescens, and is on a 78-kb insert that contains 67 open reading frames. Another inducing clone carries a gene with homology to homocitrate synthase. Our results demonstrate the power of an intracellular screen to identify functionally active clones and biologically active small molecules in metagenomic libraries.

Soil is densely populated with microorganisms that produce many small bioactive molecules, including antibiotics, anticancer compounds, immunosuppressive agents, insecticides, and other molecules that are important in nature and useful to humans (22, 45, 46, 48). Most of the known molecules of these classes are derived from the cultured microorganisms in soil, which represent <1% of the soil community (63). Soil is likely to contain >6,000 species per gram (7) and therefore probably contains more extensive functional diversity than is represented by cultured microorganisms (6, 24, 25).

Metagenomics, or the analysis of the collective genomes of an assemblage of organisms (22, 52), provides a strategy to find bioactive compounds from among uncultured organisms. Metagenomic libraries are constructed from DNA extracted directly from an environmental sample. Functional profiling of metagenomic libraries, in which all clones in a library that express a certain function are identified and then sequenced, may provide insight into the genomic context of the genes encoding the function and the phylogeny of the organisms that contain these genes (24, 50, 54). This approach has recovered novel biocatalysts and bioactive compounds, and has expanded our knowledge of uncultured microorganisms (for reviews, see references 8, 25, and 50). However, the outstanding challenges in functional metagenomics include screening the massive libraries needed to represent all of the genomes in an environmental sample and overcoming barriers to heterologous gene expression in the host species harboring the library. These challenges are significant barriers in identifying small molecules such as antibiotics and signal compounds (24, 26, 38, 60, 65). Typically, these have been identified in screens involving overlays with indicator organisms, but these screens are not sufficiently sensitive, only detect exported molecules, and are too laborious to apply easily to libraries containing millions of clones.

Our focus is on discovering small molecules in members of soil communities that have not been well studied. The site of interest in the present study is a nonpermafrost soil supporting a balsam poplar stand on a floodplain island in the Tanana River in central Alaska. The average daily soil temperature remains below 12°C (42). Unlike microorganisms in many soils of temperate climates, microorganisms in the extreme terrestrial environments of Alaska have not been well studied. This extreme environment might be home to microorganisms that produce novel biologically active molecules, as indicated by the novelty of organisms in other cold environments. For example, a study of a Siberian tundra soil using culture-independent methods revealed 16S rRNA sequences that diverge deeply from known sequences and may represent previously undescribed divisions of bacteria (69). Moreover, substantial metabolic activity has been measured in permafrost sediments (51), an arctic glacier (56), and mountain snow (55).

To develop an intracellular screen for small molecules, we sought a system that responds to diverse molecules and is found in diverse organisms and environments. Quorum sensing satisfied these requirements. It is a process of cell-cell communication mediated by small signal molecules that enables bac-
teria to sense their own cell density. Quorum sensing has emerged as a central aspect of microbial life, and therefore diverse, as-yet-uncultured organisms are likely to produce compounds that regulate quorum sensing. Structural variation in quorum-sensing inducers provides specificity for regulation of many ecologically important functions, which include luminescence (12, 27), virulence (30, 32, 47, 49, 66), and antibiotic production (2, 43, 67). Molecules that have been shown to inhibit quorum sensing include furanones (18), lactonases (10), acylases (33), and long-chain acyl-homoserine lactones (HSLs) (41).

Quorum sensing was discovered in *Vibrio fischeri*, in which LuxI directs the synthesis of the acyl-HSL signal molecule (12, 44). When the signal molecule reaches sufficient concentration in the cell with the transcriptional activator, LuxR, it activates transcription of target genes, including luxI, in an amplification loop (15). Diverse small molecules other than acyl-HSL molecules activate or antagonize lux-based biosensors (9, 16, 20, 28, 40). Of particular interest is the finding that sublethal concentrations of antibiotics such as rifampin and erythromycin induce quorum-sensing-regulated gene expression (20). Based on these features, we chose the lux-luxR system as the basis for an intracellular screen.

Here we describe construction and screening of metagenomic libraries from Alaskan soil microorganisms. We present and evaluate an intracellular screen, designated METREX, for metagenomic libraries in which the biosensor that detects active clones is inside the same cell as the metagenomic DNA.

### MATERIALS AND METHODS

#### Soil.
Soil was collected from islands, BP1 and BP3, in the National Science Foundation Long-Term Ecological Research site at Bonanza Creek Experimental Forest near Fairbanks, Alaska, in October 2001, May 2002, and October 2002. Organic and mineral layers from soil cores were treated as separate samples. The textures, organic matter contents, and pHs of representative soil cores were determined by standard methods at the Soil and Plant Analysis Lab, University of Wisconsin—Madison (see Table 2). Soil was put through a 5-mm sieve to remove roots and large particulates and stored at 4°C until use.

#### Bacterial strains and culture conditions.
The bacterial strains and plasmids used in the present study are listed in Table 1. Reporter strain JB525 (*Escherichia coli* MT102 (pJBA132), referred to as reporter strain “JB525”) was grown in a yeast extract-mannitol (AMA) (34) broth medium at 28°C. PAO1, *V. fischeri* CHROMOBACTERIUM violaceum CV026, and *P. aeruginosa* PAO1, were grown in Luria-Bertani (LB) medium at 37°C, except for reporter strain JB525, which was grown in LB amended with 5 g of NaCl/liter. *Agrobacterium tumefaciens* KYC55 (pJZ410, pJZ384, pJZ372) was grown in 5 ml of broth medium at 28°C. *Pseudomonas putida* F117 (pKR-C12) (57), and *Pseudomonas putida* IsoF (41), were grown in a Trypticase Soy Broth (DHSa) (Promega, Madison, WI) with 5 ml of transfection buffer (68) in centrifuge tubes and subjected to two freeze-thaw cycles.

#### DNA extraction and preparation.
Two procedures were used to extract DNA: direct lysis within the soil matrix (direct extraction) and lysis of cells that were separated from the soil matrix by differential centrifugation prior to lysis (cell separation). The direct extraction procedure, used for libraries AK1 through AK4, was based on the method of Rondon et al. with modifications as follows (52). The cell separation procedure, used for libraries AK5 through AK8, was performed as previously described (37, 64). For the direct extraction procedure used for libraries AK1 through AK4, 100 g of soil was mixed with 150 ml of extraction buffer (68) in centrifuge tubes and subjected to two freeze-thaw cycles.

#### Construction of plasmid libraries.
Soil was put through a 5-mm sieve to remove roots and large particulates and stored at 4°C until use. Soils were sieved through 5-mm sieves, ground to 0.2-mm size, and used for library construction, and *E. coli* used in the present study are listed in Table 1.

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristic(s)</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biosensor strains and plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em> KYC55 (pJZ410, pJZ384, pJZ372)</td>
<td>KYC55: R10 derivative lacking pTiR10, pJZ410, pJZ384, pJZ372: T7 expression system to express TraR</td>
<td>5, 70</td>
</tr>
<tr>
<td><em>Agrobacterium tumefaciens</em> NT1 (pTiC58ΔaccR)</td>
<td>Acyl-HSL producer</td>
<td>3</td>
</tr>
<tr>
<td><em>Chromobacterium violaceum</em> CV026</td>
<td>Acyl-HSL-negative mutant of ATCC 31532; most sensitive to N-hexanoyl-L-HSL</td>
<td>41</td>
</tr>
<tr>
<td>Escherichia coli MT102 (pJBA132), referred to as reporter strain “JB525”</td>
<td>MT102: araD139 (ara-lev)7697 Δlac thi hsrR, pJBA132: pME6031 carrying luxR-PaerP-BSII-gfp(ASV)-Tc-Tc; Tet’; based on components of the <em>V. fischeri lux</em> quorum-sensing system, it is most sensitive to N-(3-oxohexanoyl)-l-HSL</td>
<td>1</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em> PAO1</td>
<td>Acyl-HSL producer</td>
<td>Pseudomonas Genetic Stock Center</td>
</tr>
<tr>
<td><em>Pseudomonas putida</em> F117 (pKR-C12)</td>
<td>F117: acyl-HSL-negative derivative of IsoF, PpuI-, pKR-C12: pBBR1MC8-S carrying Paaac-GFP(ASV)-Pana-luxR; Gm’; based on components of the <em>P. aeruginosa las</em> quorum-sensing system, it is most sensitive to N-(3-oxodecanoic)-l-HSL and other long-chain acyl-HSL molecules</td>
<td>57</td>
</tr>
<tr>
<td><em>Pseudomonas putida</em> IsoF</td>
<td>Acyl-HSL producer</td>
<td>57</td>
</tr>
<tr>
<td>E. coli hosts for library construction or cloning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHLs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TransforMAX 300 EPI300 Vectors</td>
<td>recA1 and endA1 cloning strain</td>
<td>Lab stock</td>
</tr>
<tr>
<td>trfA’ host for inducible copy number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pSuperBAC</td>
<td>High-copy-inducible BAC cloning vector; Cm’</td>
<td>Epicentre</td>
</tr>
<tr>
<td>pCC1BAC</td>
<td>High-copy-inducible BAC cloning vector; Cm’</td>
<td>Epicentre</td>
</tr>
<tr>
<td>pCC1FOS</td>
<td>High-copy-inducible fosmid cloning vector; Cm’</td>
<td>Epicentre</td>
</tr>
<tr>
<td>pGEM-T</td>
<td>Cloning vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pGEM-3Zf(+)</td>
<td>Cloning vector</td>
<td>Promega</td>
</tr>
</tbody>
</table>

*a* Tet’, tetracycline resistance; Gm’, gentamicin resistance; Cm’, chloramphenicol resistance.
Library Construction. We constructed four bacterial artificial chromosome (BAC) libraries (AK1, -2, -4, and -5) and a fosmid library (AK6) in CopyControl pCC1FOS (Epicenter) (Table 2). The pSuperBAC was isolated and purified with QIAprep Miniprep (QIAGEN) and precipitated with isopropanol (53). From this preparation, 50 ng of DNA larger than 50 kb was excised and extracted from the gel with GELase or AgarACE as described above.

DNA ligation reactions contained an insert/vector ratio of 1:10 for pSuperBAC or 1:1 for CopyControl pCC1BAC. Vectors and insert were heated to 55°C for 10 min and then cooled at room temperature 15 min prior to adding 3 U of T4 DNA ligase and buffer (Promega). Ligation reactions were incubated overnight at 16°C. Transformation was performed as described previously (17) and plated on LB medium containing Cm at 10 μg ml⁻¹. Libraries were evaluated by restriction digest analysis of randomly selected clones using NotI. Colonies were pooled by scraping them from plates into LB containing Cm and mixing them with an equal volume of 14% dimethoxy sulfone, followed by storage at −80°C. To construct AK6, purified DNA (2.5 μg) was enzymatically treated to prepare 5' phosphorylated blunt ends and ligated by using a CopyControl Fosmid Library Production Kit (Epicenter). After packaging into lambda phages, infected cells were plated and stored as described above for the BAC libraries.

Intracellular Screen. In this screen, plasmid DNA from pooled metagenomic clones (Table 2) was transferred into cells containing the biosensor plasmid pJBA132 (provided by M. Givskov) (Fig. 1), and clones were detected by fluorescence-activated cell sorting (FACS) or fluorescence microscopy. A pool of all clones from libraries AK1 through AK8 was prepared and then incubated in SOB (0.5% yeast extract, 2.0% tryptone, 0.05% NaCl, 10 mM MgCl₂, 2.5 mM KCl) for 2 h at 37°C, and the plasmid DNA was isolated by using QIAprep Miniprep (QIAGEN) and precipitated with isopropanol (53). From this preparation, 50 ng of DNA was transformed into reporter strain JB525 (E. coli MT102 containing pJBA132) using the conditions provided in the description of library construction. After 1 h of recovery in SOB at 37°C with shaking, 1:300 dilutions were prepared by using 10 μl of recovery mix in 3 ml total volume of SOB containing Cm and Tet and 50 and 80 mM, respectively, N-(3-oxoheaxanoyl)-1-HSL. The same host carrying the empty vector served as a negative control. The dilutions were incubated at 28°C for 5 h and sorted into green fluorescent protein (GFP)-expressing and non-GFP-expressing fractions by using an EPICS Elite flow cytometer/cell sorter (Coulter Corp., Fullerton, CA) as previously described (11). Non-GFP-expressing cells were collected into 1 ml of SOB containing Cm and Tet and cultured on plates containing the background. The plasmids (BAC/fosmid plasmid and biosensor plasmid) were isolated, retransformed into E. coli TransforMax EP1000, and plated on LB medium containing Cm to select for retransformants containing the BAC or fosmid plasmid. Resulting colonies were tested on LB medium containing Tet to confirm that the biosensor plasmid was expressed and non-GFP-expressing fractions by using an EPICS Elite flow cytometer/cell sorter (Coulter Corp., Fullerton, CA) as previously described (11). Non-GFP-expressing cells were collected into 1 ml of SOB containing Cm and Tet and cultured on plates containing the background. The plasmids (BAC/fosmid plasmid and biosensor plasmid) were isolated, retransformed into E. coli TransforMax EP1000, and plated on LB medium containing Cm to select for retransformants containing the BAC or fosmid plasmid. Resulting colonies were tested on LB medium containing Tet to verify and eliminate colonies that contained both the biosensor and BAC/fosmid plasmids. Isolated BAC/fosmid plasmids were retransformed into a fresh re-
porter strain JB525, and the phenotype of the transformants was assessed to determine whether the cloned DNA was responsible for the phenotype. Plasmids isolated from the original colonies and from the retransformed colonies were digested with NotI, and the restriction digest patterns were compared by pulsed-field gel electrophoresis.

**Specificity of biosensor.** Genomic DNA was isolated with the Easy-DNA kit (Invitrogen, Carlsbad, CA) from *P. aeruginosa* PAO1, *P. putida* IsoR, and *A. tumefaciens* NT1. Primers were designed to amplify lasI, pupI, and traI based on published sequences (the accession numbers are AE004572 [59], AY115588 [58], and L22207 [29], respectively) (las forward, 5'-ATC GTA CAA ATT GTT CGG-3'; las reverse, 5'-TGA CGC CAG TCG CTG-3'; pup forward, 5'-CTI ATC TCC ATT GAT GGC C-3'; pup reverse, 5'-ATG GCA GTT ATT CTT GAT-3'; tra forward, 5'-CGG ATA CTC ACA GTT TCG-3'; tra reverse, 5'-GGC GGC ACT CCT GAT GGC-3'). PCR amplification of acyl-HSL synthesis genes (50-μl reaction volume) used 100 ng of genomic DNA, a 200 mM concentration of each deoxynucleoside triphosphate, 2.5 U of Taq polymerase, and 20 pmol of each forward and reverse primer. Reactions were performed in an Eppendorf Mastercycler gradient thermocycler (Eppendorf, Hamburg, Germany) using a miniprep method (Agencourt Bioscience Corp, Beverly, MA). Purified plasmid DNA from *E. coli* DH5α (Promega), which was linearized by using BamHI (Promega). Fragments greater than 2 kb were extracted from the gel and ligated into pGEM-T Easy (Promega), which was transformed by using BamHI (Promega). Plasmid DNA from *E. coli* DH5α transformants was prepared by the SprintPrep miniprep method (Agencourt Bioscience Corp, Beverly, MA). Purified plasmids were each sequenced by using the universal sequencing primers M13f and M13r at the University of Wisconsin DNA sequencing facility using BigDye Terminator.

**Chemical characterization of culture supernatants.** Extracts for thin-layer chromatography (TLC) analysis were prepared from 50-ml cultures grown to stationary phase at 37°C. Bacteria were removed by centrifugation, and the supernatants were extracted twice with 50 ml of acidified ethyl acetate (0.1 ml of glacial acetic acid per liter). Extracts were evaporated to a final volume of ca. 1.5 ml and stored at −20°C. Then, 1-μl samples were applied to C18 reversed-phase TLC plates (200-μm layer; J. T. Baker, Philipsburg, NJ) and developed with 60:40 (vol/vol) methanol-water. Solvents were evaporated, and the plates were analyzed for the presence of acyl-HSLs with *A. tumefaciens* indicator strain JKC55 (pZ410) (pZ284) (pZ272). The acyl-HSL indicator was grown to stationary phase, and 5 ml of culture was added to 95 ml of AMA amended with 7.5 g of agar/liter and 50 µg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)/ml and then maintained at 30°C. After the plates were incubated for 24 h, colonies were screened for fluorescence by using a fluorescence microscope as described above.

**Overaly assay to detect biactive compounds produced by metagenomic clones.** To compare the sensitivity of the intracellular screen to an overlay screen, active clones detected by the intracellular screen were plated at a density sufficient to visualize single colonies on LB medium, incubated overnight at 37°C, and overlaid with 6 ml of soft agar (LB with 0.005 g of agarose ml −1) that had been inoculated with 0.05 ml of an overnight culture of reporter strain JB525. The same host carrying the empty vector served as a negative control. After 24 h of incubation at 37°C, colonies were screened for fluorescence by using a fluorescence microscope as described above.

**Overaly assay to detect biactive compounds produced by cultured isolates.** To compare the sensitivity of the intracellular screen to an overlay screen of cultured isolates, dilutions of Alaskan soil samples comparable to those used to make the metagenomic libraries were cultured on 1/10-strength tryptic soy broth agar (i.e., 3.0 g of tryptic soy broth/15% agar) at 28°C for 1 week. Reporter strain JB525 was applied as an overlay to three of these plates, which each contained about 300 colonies (total of 1,000 colonies).

**Specificity of quorum-sensing inducing activity.** All assays were performed with isolated metagenomic DNA plasmids maintained in *E. coli* TransforMax EPI300. To determine the specificity of the active molecule with different reporter strains, overnight cultures of QS1 and QS1-C were cross-streaked with the reporter strain JB525 (E. coli MT102 (pJBA132)), *P. putida* F117 (pKR-C12), and *C. violaceum* CV026, and activity was measured after 24 h of incubation at 28°C. *E. coli* TransforMax EPI300 containing pSuperBAC was used as a negative control.

**Chemical characterization of culture supernatants.** Extracts for thin-layer chromatography (TLC) analysis were prepared from 50-ml cultures grown to stationary phase at 37°C. Bacteria were removed by centrifugation, and the supernatants were extracted twice with 50 ml of acidified ethyl acetate (0.1 ml of glacial acetic acid per liter). Extracts were evaporated to a final volume of ca. 1.5 ml and stored at −20°C. Then, 1-μl samples were applied to C18 reversed-phase TLC plates (200-μm layer; J. T. Baker, Phillipsburg, NJ) and developed with 60:40 (vol/vol) methanol-water. Solvents were evaporated, and the plates were analyzed for the presence of acyl-HSLs with *A. tumefaciens* indicator strain JKC55 (pZ410) (pZ284) (pZ272). The acyl-HSL indicator was grown to stationary phase, and 5 ml of culture was added to 95 ml of AMA amended with 7.5 g of agar/liter and 50 µg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)/ml and then maintained at 30°C. After the plates were incubated for 24 h, colonies were screened for fluorescence by using a fluorescence microscope as described above.

**Overaly assay to detect biactive compounds produced by cultured isolates.** To compare the sensitivity of the intracellular screen to an overlay screen of cultured isolates, dilutions of Alaskan soil samples comparable to those used to make the metagenomic libraries were cultured on 1/10-strength tryptic soy broth agar (i.e., 3.0 g of tryptic soy broth/15% agar) at 28°C for 1 week. Reporter strain JB525 was applied as an overlay to three of these plates, which each contained about 300 colonies (total of 1,000 colonies).
tor (v. 3.1; Applied Biosystems, Foster City, CA). Sequence reads were assembled using the STADEN computer program (v. 1.40; http://staden.sourceforge.net/). Contigs and regions initially covered by only one sequence read were joined by sequencing PCR fragments until the average sequencing coverage was >7-fold and every region had at least 2-fold coverage. Putative open reading frames (ORFs) were identified by using ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and were annotated by using ASAP (19). Plasmids from subclones containing metagenomic DNA were isolated, transformed into the reporter strain, plated, and screened by fluorescence microscopy to identify active subclones.

Inserts in plasmids pQS11 (from metagenomic clone QS11 that conferred fluorescence on JB525) and pLhNh (from metagenomic clone INH6 that is a putative inhibitor of fluorescence) were sequenced by primer walking from the cloning vector. ORFs were identified as described above. The pQS11 ORF of interest that had homology to isopropylmalate/homocitrate synthase was PCR amplified and cloned into pGEM-T (Promega). The ligated product was screened for GFP-inducing activity as described above.

**Phylogenetic analysis.** Phylogenetic trees of each ORF from clones QS1, QS11, and INH6 were constructed to determine their function and closest known relative (13, 35). We obtained homologues of each ORF by performing BlastP searches of 185 finished archaeal and bacterial genomes. Those matches where the BLAST score ratio (BSR; i.e., the ratio of the BlastP score for the database match to the score of the metagenomic ORF) was more than 0.30 were retained for further analysis. This is approximately the same as 30% identity over the full length of the protein sequence. We considered any ORF without a homologue to be a hypothetical ORF. For each ORF, we used CLUSTALW to align the homologous protein sequences and constructed maximum-parsimony trees using PAUP* (61).

The methods used to construct the LuxI-LuxR phylogenetic tree have previously been described (21, 36). Briefly, the QS1 LuxI and LuxR protein sequences were used to perform a BlastP search against the nonredundant NCBI protein database. Protein sequences from completed genomes and those from members of the Enterobacteriaceae were selected for further analysis. The accession numbers for the sequences used in this analysis are available online (http://www.plantpath.wisc.edu/fac/joh/LuxIR-Accessions.htm). The LuxI and LuxR protein sequences were concatenated (LuxIR) and aligned by using CLUSTALW (4). Any gap-containing columns in the alignment were removed from the analysis using ARB (39). Bootstrapped maximum-parsimony trees were constructed in PAUP* version 4.0 using 1,000 replicates (61). We used MEGA version 3.0 to construct bootstrapped neighbor joining trees with 1,000 replicates and a gamma distance correction (31). Tetrancleotide frequency analysis of the insert was performed by using TETRA and its database of finished genomes (62).

**Nucleotide sequence accession numbers.** The DNA sequence of the QS1 78-kb fragment has been deposited in GenBank under accession number AY688432. The DNA sequences of the QS11 and INH6 fragments have been deposited in GenBank under accession numbers DQ008471 and DQ008472, respectively.

## RESULTS

The goal of the present study was to develop a rapid method to identify metagenomic clones that produced biologically active small molecules. We constructed large libraries of metagenomic DNA from soil and developed an intracellular screen to identify clones that induce or inhibit a bacterial quorum sensing biosensor.

### Library construction.

We constructed eight soil libraries with DNA extracted from the organic and mineral layers of soil cores from Alaska. Restriction digest analysis of 223 randomly selected clones using NotI revealed inserts ranging from 1 to 190 kb. Collectively, the libraries contain 653 Mb of DNA (Table 2).

### Biosensor.

The biosensor plasmid used in our intracellular screen is plasmid pJBA132, in which luxR-P lux I, from V. fischertei is fused to gfp(ASV) (1). The E. coli host strain does not produce detectable quorum-sensing inducing compounds, and thus the GFP is expressed either when exogenous inducer is applied or a gene (or genes) directing the synthesis of an inducer is introduced into the cell. To test the specificity of the biosensor, we introduced lasI, ppuI, and tral from P. aeruginosa, P. putida, and A. tumefaciens, respectively, into the biosensor strain. The biosensor was activated by the products of each of the three genes, indicating that it detects a range of acyl-HSL compounds.

### Intracellular screen for inducers.

We inspected approximately 180,000 colonies for GFP expression by fluorescence microscopy and found three clones, QS1, QS2, and QS3, that fluoresced after 12 h of incubation (Table 3). Restriction digest and pulsed-field gel electrophoresis revealed that the clones were siblings containing a 78-kb insert. The same clone was rediscovered in subsequent screens. Ten additional clones, with inserts ranging from 3.5 to 42 kb, fluoresced after 24 h of incubation on LBA containing 3 nM N-(3-oxohexanoyl)-\(\beta\)-HSL at 37°C.

### Intracellular screen for inhibitors.

We used FACS to identify cells in which the presence of the metagenomic DNA results in inhibition of GFP expression induced by exogenous N-(3-oxoheaxanoyl)-\(\beta\)-HSL. In one screen for inhibitors of GFP expression, we included 50 nM N-(3-oxoheaxanoyl)-\(\beta\)-HSL and detected GFP expression on a continuum from dark to bright. In a second screen, using 80 nM N-(3-oxoheaxanoyl)-\(\beta\)-HSL, we detected better separation of the “dark” and “bright” phenotypes. In both screens for inhibition, approximately 10,000 cells with the “dark” phenotype were collected. To sort potential clones producing inhibitors from dead cells, the cells collected from FACS were plated onto LBA containing Cm and Tet and incubated overnight. Two clones, INH6 and INH11, had a confirmable phenotype and were selected for further study. Sequence information for pINH6 and pQS11 inserts is reported below.

Plasmids from subclones of pQS1 were transformed into the reporter strain and screened by fluorescence microscopy. Clones QS1-C, a subclone of the plasmid in QS1, induced intense GFP fluorescence after 12 h of incubation on LBA at 37°C; B, colonies fluoresced after 24 h of incubation on LBA at 37°C; C, colonies did not fluoresce after 24 h of incubation on LBA containing 3 nM N-(3-oxoheaxanoyl)-\(\beta\)-HSL at 37°C.

### Comparison of intracellular screen with overlay assay.

To determine whether the intracellular nature of the screen provided more sensitivity in detecting active clones, we tested all of the clones identified with the intracellular biosensor in a...
typical “intercellular” or overlay screen. Among the active clones tested, only QS1 induced GFP expression when the biosensor was used as an overlay. The remaining clones (QS4 through 6, QS8 through 11, QS15, QS35, and QS55) did not activate the overlay-based reporter, indicating that the intracellular feature of our screen enhances its sensitivity and enables it to detect clones that would be missed in an overlay assay.

Overlay assay to detect bioactive compounds produced by cultured isolates. To compare the frequency of quorum-sensing inducing activity in metagenomic libraries with that in cultured bacteria from the same site, we tested a collection of approximately 1,000 bacterial cultures from Alaskan soil. Among these isolates, which likely contain approximately 10 times as much DNA as the metagenomic libraries, only one activated the biosensor in the overlay assay.

Nucleotide sequence of pQS1. We sequenced the DNA insert of pQS1 to characterize the gene responsible for the quorum-sensing phenotype and characterize its genomic environment and to reveal phylogenetic information that might indicate the affiliation of the organism from which the DNA originated (13, 35). Plasmid pQS1 contains an insert of 77,858 bp and 64 complete ORFs (Fig. 2A). A 5-kb region at the 3’ end of the insert contains luxI and luxR homologues (Fig. 2A, green arrows) that are oriented toward each other and whose products reliably cluster with the Pseudomonas sp. group of the LuxI and LuxR homologues (Fig. 3). The LuxI homologue has 62% amino acid sequence identity to its closest match in GenBank, AmfI from Pseudomonas fluorescens. Three genes encoding putative transposases separate the luxI and luxR homologues.

The remainder of the pQS1 insert contains two large regions (27.4 and 24.5 kb, Fig. 2A, shaded in blue) in which all ORFs are most similar to enterics (blue), Chromobacterium violaceum (purple), Pseudomonas syringae (red), indeterminate (turquoise), or unknown (black). The box around the luxI represents the region found in the subclone of pQS1 (pQS1-C) that retained the quorum-sensing activity.

FIG. 2. Gene maps of pQS1 (A), pQS11 (B), and pINH6 (C) (distances in kb). ORFs are color coded as follows: gray, conserved hypothetical (BSR > 0.30); black, hypothetical (BSR < 0.30); yellow, ORFs with nonhypothetical BLAST matches; red, genes related to transposases; blue, genes related to acyl carrier proteins; green, luxI and luxR (A), homocitrate/isopropylmalate synthase (B), and putative cytochrome (C). The bars above the ORFs represent regions that are most similar to enterics (blue), Chromobacterium violaceum (purple), Pseudomonas syringae (red), indeterminate (turquoise), or unknown (black). The box around the luxI represents the region found in the subclone of pQS1 (pQS1-C) that retained the quorum-sensing activity.
FIG. 3. Phylogenetic tree of the concatenated LuxI and LuxR protein sequence from pQS1 and its homologues. The topology of the tree was determined by using 1,000 bootstrap replicates each for neighbor joining with a gamma distance correction and using maximum parsimony. The branch lengths were taken from the neighbor-joining analysis. Numbers above and below each node represent the bootstrap values obtained by using neighbor-joining and maximum-parsimony analysis, respectively. Nodes lacking a number represent those that could not be supported by at least 50% of the bootstrap replicates. Accession numbers for each LuxI and LuxR homolog represented in the tree are available at http://www.plantpath.wisc.edu/fac/joh/LuxIR-Accessions.htm.
Tetranucleotide frequency analysis of the ORFs in the insert contradicted the phylogenetic analysis of the ORFs. Comparison of tetranucleotide frequencies predicted that the insert was most similar to *Pseudomonas syringae* pv. tomato (correlation = 0.88). The tetranucleotide frequency correlation with the *Y. pestis* genomes was low (correlation = 0.57). When the analysis was repeated using the regions that were most similar to members of the Enterobacteriaceae, the *P. syringae* genome still had the highest correlation score (0.88), and correlations to the *Yersinia* sp. genomes were low (all 0.57).

The G+C content of the pQS1 insert was also analyzed. The two Enterobacteriaceae fragments are 55.4 and 50.8% G+C, respectively. The *C. violaceum* fragment and the region between the second Enterobacteriaceae fragment and the *C. violaceum* fragment are both 51% G+C. Finally, the G+C content of the *P. syringae* (i.e., the luxI and luxR) fragment is 48.3%. For comparison, the G+C content of the sequenced *P. syringae* (accession number NC_004578) and *Y. pestis* genomes (accession numbers NC_003143, NC_004088, and NC_005810) are 58 and 48%, respectively.

**Nucleotide sequence of pQS11**. Plasmid pQS11 contains a 3.5-kb insert with four ORFs. Two ORFs are hypothetical proteins, one is a fragment of a gene with homology to a transcriptional activator, and the other is similar to either 2-isopropylmalate synthase or homocitrate synthase. Protein sequence analysis of the transcriptional activator and the synthase was inconclusive because the phylogeny of the homologues for both ORFs demonstrates no consistent meaningful phylogenetic relationships. The most similar synthase homolog is from *Desulfovibrio vulgaris* (55% identity, BSR = 0.55; GenBank accession no. YP_012192). The ORF containing the synthase gene was cloned into pGEM-T, and the resulting clone induced GFP expression in the biosensor strain.

**Nucleotide sequence of pINH6**. Plasmid pINH6 carries an insert of 3.8 kb containing five ORFs. Of the five ORFs, three are hypothetical proteins, one is 53% identical to a conserved hypothetical ORF from *Streptomyces coelicolor* (NP_626248), and one is 32% identical to a putative cytochrome from *Leptospira interrogans* (YP_000668 or NP_713687, BSR = 0.23). The mechanism of inhibition of quorum sensing by this clone is under investigation.

**Diffusion and specificity of active compounds from clones QS1 and QS1-C**. The signal molecule produced by QS1 and QS1-C diffused in agar, as indicated both by cross-streak assays, as well as when culture was spotted on a lawn of reporter strain and incubated overnight. Both clones induced high levels of expression in reporter strain JB525 and *C. violaceum* CV026, but not *P. putida* F117(pKR-C12), indicating that the clones produce compounds with similar specificity to short-chain acyl-HSLs and not to long-chain acyl-HSL molecules.

**Chemical characterization of active compounds**. The active compound(s) in QS1 and QS1-C were compared to synthetic acyl-HSLs in TLC and were detected with the overlay of the *A. tumefaciens* biosensor. The active material from QS1 and QS1-C produced teardrop-shaped patterns that were indistinguishable from the *N-(3-oxohexanoyl)-l-HSL*. Its pattern and mobility (Rf = 0.68), coupled with its biological specificity, indicate that the active signal molecule from clone QS1 is *N-(3-oxohexanoyl)-l-HSL*.

**DISCUSSION**

One of the challenges of functional metagenomics is detecting rare active clones in libraries containing many members. The potential for poor expression of genes derived from organisms distantly related to the host species harboring the library makes the development of highly sensitive assays imperative. In an attempt to address both of these issues, we developed a rapid, sensitive intracellular screen in which the sensor for activity is in the same cell as the metagenomic DNA. We designate this screen “METREX,” for metabolite-regulated expression. We demonstrated the power of METREX in identifying metagenomic clones that produce biologically active small molecules by screening Alaskan soil microbial DNA libraries. METREX identified 11 clones that induce expression of the biosensor and two that inhibit expression. The METREX screen, conducted in host species other than *E. coli* may detect additional novel quorum-sensing genes. METREX shares certain features with the recently described screen, SIGEX (65). Whereas SIGEX is a trap for promoters that are responsive to certain metabolites, METREX detects metabolites that might activate such promoters (26, 65).

One clone identified by METREX, designated QS1, produces a biologically active small molecule that induces GFP expression. Based on its chemical properties and specificity, the signal molecule produced by QS1 appears to be an acyl-HSL, and most resembles *N-(3-oxohexanoyl)-l-HSL*. However, the gene that directs synthesis of this inducer differs from known homologues of luxI based on both its nucleotide sequence and the genomic organization of the flanking DNA. Based on the similarity of the 67 ORFs carried by pQS1 to other γ-Proteobacteria sequences and the clustering of its LuxI and LuxR homologues in a family comprised exclusively of γ-Proteobacteria sequences, it seems likely that this fragment originated from a member of the γ-Proteobacteria. Our results are consistent with studies that show acyl-HSL systems are restricted to Proteobacteria (21, 36). Further study of the 53 ORFs for which no function can be inferred based on sequence alone may provide insight into the lifestyle of the organism from which this DNA fragment originated.

Two hypotheses have been presented previously to account for the clustering of LuxI and LuxR homologues from *P. syringae* with enteric γ-Proteobacteria. The first suggests that bacteria similar to *P. syringae* obtained their LuxI and LuxR homologues via horizontal gene transfer from an enteric γ-Proteobacterium (21). The second hypothesis posits two LuxI and LuxR families, one of which contains only γ-Proteobacteria and likely arose before the divergence of *Pseudomonas* species (36). Interestingly, the majority of the pQS1 ORFs have symmetry with those of the *Y. pestis* genome (an enteric γ-Proteobacterium), yet the LuxIR homologues cluster with PsyIR from *P. syringae* and AmfIR from *P. fluorescens*. The tetranucleotide frequency in pQS1 is most similar to *P. syringae*. Although QS1 does not contain enough information to support either hypothesis, the presence of transposases within the LuxIR region suggests that horizontal gene transfer may have occurred.
The intracellular METREX screen detects clones that would not be detected in an overlay screen. Inhibitors and possibly inducers of the biosensor can be identified with fluorescence-activated cell sorting, which can sort 60,000 cells/min (11) and quickly identify rare active clones. Inhibitors of quorum sensing are difficult to identify in the overlay because the bright fluorescence of the majority of the colonies obscures the dark ones. Some inducers of quorum sensing would not be detectable in an overlay assay because the active compounds are not detectable outside of the producing cell. For example, ten clones that were detected by the intracellular screen did not induce GFP expression in reporter colonies when tested in an overlay assay.

An unexpected result was that the estimated frequency of quorum-sensing inducers detected in the metagenomic libraries with the intracellular screen was substantially higher than among cultured bacteria from the same soil. We found only one active colony among 1,000 colonies that were isolated from the soil, whereas we found 11 active clones in metagenomic libraries from the same soil that contained approximately 130 genome equivalents (650 Mb of DNA divided by an average genome size of 5 Mb). Although we do not know whether the difference is due to the screen (intracellular versus overlay), the source of the activity (uncultured versus cultured bacteria), or to other factors such as redundancy in culturing, the metagenomic libraries screened intracellularly appear to be richer sources of quorum sensing-inducing activity than are the cultured bacteria screened with the overlay assay.

Since the metagenomic DNA in the libraries is only a small sample of the total soil metagenome and the cultured isolates are only a small sample of the population of microbes, the frequency of quorum-sensing inducers found in either screen does not necessarily represent the proportion of quorum-sensing inducers in Alaskan soil. Both screens indicate that a small proportion of microorganisms produce detectable signals that activate the quorum-sensing biosensor. The frequency of quorum-sensing activators in soil is not known. Elasri et al. found that among Pseudomonas species, acyl-HSL production is common among plant-associated bacteria but uncommon among soilborne bacteria (14).

Functional metagenomics is a powerful approach for discovering new genes and gene products because it does not require prior knowledge of the gene sequence. Despite all of its limitations due to barriers of heterologous gene expression, the potential for discovery with functional metagenomics is significant, as demonstrated by the present study. With METREX, we found genes that would not have been detected with a sequence-based approach. pQS11 and pINH6, for example, do not contain genes related to any that are known to encode inducers or inhibitors of quorum sensing, and only a functional analysis could have discovered their activity. As we overcome some of the limitations of functional metagenomics, screens such as METREX will become more fruitful and may help realize the promise of metagenomics for discovery of small molecules.

ACKNOWLEDGMENTS

We thank Michael Givskov for the reporter strain JB525, Ronald Binder for the acyl-HSL standards, Leo Eberl for F117(pKR-C12), and Steve Winans for the A. tumefaciens acyl-HSL indicator strain. We also thank Zakee Sabree for designing Fig. 1 and Nicole T. Perna for facilitating our use of ASAP.

P.D.S. was partially supported by a USDA soil biology postdoctoral fellowship. This study was supported by grants from the National Science Foundation Microbial Observatory program (MCB-0132085), The Howard Hughes Medical Institute Professors Program, the Biotechnology Research and Development Corporation, and Hatch Project 4534 from the College of Agricultural and Life Sciences of the University of Wisconsin–Madison.

REFERENCES
