Submitted to *Applied and Environmental Microbiology* on 30 Aug 2012

Phylogenetically novel LuxI/LuxR-type Quorum Sensing Systems Isolated Using a Metagenomic Approach

Eri Nasuno,[1,2] Nobutada Kimura,[1*] Masaki J. Fujita,[3] Cindy H. Nakatsu,[4] Yoichi Kamagata,[1,5] and Satoshi Hanada[1,2]

Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan,[1] Graduate school of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8572, Japan,[2] Creative Research Institution, Hokkaido University, 3-1-1 Minato-cho, Hakodate, Hokkaido 041-8611, Japan,[3] Department of Agronomy, Purdue University, 915 West State St., West Lafayette, Indiana 47907-2054, USA,[4] Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),2-17 Tsukisamu-Higashi, Toyohira, Sapporo, Hokkaido 062-8517, Japan[5]

Running title: New QS systems isolated using a metagenomic approach

Journal section: Microbial ecology

*Corresponding author. Mailing address: Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Central 6, AIST Tsukuba Center, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan. Phone: 81–29–861–8767. Fax: 81–29–861–6587.

E-mail:n-kimura@aist.go.jp

ABSTRACT

A great deal of research has been done to understand bacterial cell-to-cell signaling systems but there is still a large gap in our current knowledge because the majority of microorganisms in natural environments do not have cultivated representatives. Metagenomics is one approach to identify novel quorum sensing (QS) systems from uncultured bacteria in environmental samples. In this study, fosmid metagenomic libraries were constructed from a forest soil and an activated sludge from a coke plant, and the target genes were detected using a GFP-based *Escherichia coli* biosensor strain whose fluorescence was screened by spectrophotometry. DNA sequence analysis revealed two pairs of new LuxI family *N*-acyl-L-homoserine lactone (AHL) synthases and LuxR family transcriptional regulators (clones N16 and N52, designated AubI/AubR and AusI/AusR, respectively). Both AubI and AusI produced an identical AHL, *N*-dodecanoyl-L-homoserine lactone (C12-HSL), as determined by NMR and mass spectrometry. Phylogenetic analysis based on amino acid sequences suggested AusI/AusR was from an uncultured member of the β-*Proteobacteria* and AubI/AubR was very deeply branched from previously described LuxI/LuxR homologues in *Proteobacteria* isolates. The phylogenetic position of AubI/AubR indicates that they represent a QS system not acquired recently from the *Proteobacteria* by horizontal gene transfer but share a more ancient ancestry. We demonstrated that metagenomic screening is useful to provide further insight into the phylogenetic diversity of bacterial QS systems by describing two new LuxI/LuxR-type QS systems from uncultured bacteria.

INTRODUCTION

Bacteria interact with one another using chemical molecules as sensing signals.

Detection of the molecules allows bacteria to distinguish between low and high cell population densities, and to control gene expression in response to changes in cell number (46) and local environment (12). This process, referred to as quorum sensing (QS), allows a population of bacteria to coordinately control gene expression. Several types of QS signals have been found, including *N*-acyl-L-homoserine lactone (AHL) in *Proteobacteria* (5, 45, 46).

AHL-producing bacteria have been identified in over 37 genera within the α-, β- and γ-*Proteobacteria* (9, 11, 14, 32) and the *Cyanobacteria* (10). In these bacteria, AHL-dependent QS systems have been shown to regulate many bacterial behaviors, such as virulence (8, 48), and biofilm formation (25, 29) mainly in response to cell densities. Therefore, AHL-dependent QS systems are now recognized for playing important roles in the regulation of bacterial behavior.

The AHL-based QS systems usually contain a *luxI* gene homologue, responsible for synthesis of AHLs, and a *luxR* gene homologue, an AHL-dependent transcriptional regulator (19). The LuxI/LuxR-type QS systems have been experimentally identified and studied in more than 70 different species in the phylum *Proteobacteria* (6, 15). In addition, genome sequencing of a number of cultured δ-*Proteobacteria* (30, 39) and a yet-to-be cultured bacterium belonging to the phylum *Nitrospirae* (40) indicates that they may harbor putative LuxI/LuxR-type QS systems. It has been speculated that half the bacterial phyla (26 candidate phyla) do not have cultivated representatives although at least 52 bacterial phyla have been identified from 16S rRNA gene sequences in environmental samples (33). These results suggest that not only α-, β- and γ-*Proteobacteria* but a diverse array of bacteria including yet uncultivated bacterial phyla likely possess the LuxI/LuxR-type QS systems, but these possible QS systems have not been shown to be

functional. More recently, a bacterial LuxI/LuxR-like QS system found in a methanogenic archaeon, *Methanosaeta harundinacea* 6Ac was shown to be involved in regulating cell assembly and carbon metabolic flux (49).

Better understanding of QS systems will provide us with greater insight into the complex interaction mechanisms used widely among Bacteria and even Archaea in the environment. This research has been limited by the lack of information on community members without cultivated representatives. However, by using a non-cultivation based metagenomic approach new AHL synthase genes have been identified from uncultured organisms (21, 47). Screening of metagenomic libraries constructed from Alaskan soil using the reporter activity of green fluorescence protein (GFP) led to the discovery of a novel LuxI/LuxR-type QS system with low similarity to the known homologues found in γ-*Proteobacteria* (47). Moreover, metagenomic libraries constructed from activated sludge and soil resulted in the isolation of three new LuxI/LuxR-type QS systems producing previously unknown AHLs most closely related to those previously found in α-, β- and γ-*Proteobacteria* (21). These results demonstrated that metagenomic approaches are useful for the discovery of novel QS systems from uncultured bacteria.

In this study, we used a metagenomic approach to find novel LuxI/LuxR-type QS systems in uncultured bacteria belonging to phyla other than the previously studied *Proteobacteria*, such as the phylum *Nitrospirae*. The metagenomic libraries were constructed from activated sludge of a coke plant wastewater treatment system containing various organic pollutants (22, 44), and forest soil as a source of phylogenetically diverse microbes. An intercellular screening method was used and GFP fluorescence of the biosensor was detected using a spectrofluorometer. Structural analyses using ^{1}H NMR, ^{13}C NMR and FAB-MS spectra were used to identify the QS signals produced and novelty

of QS systems were determined by phylogenetic analysis.

MATERIALS AND METHODS

Bacterial strains and culture conditions. Metagenomic DNA was cloned into a fosmid CopyControl™ pCC1FOS, high-copy-inducible fosmid cloning vector (Epicentre, Madison, WI), and transferred into *Escherichia coli* TransforMAX 300, EPI300 (Epicentre). *E. coli* DH5α served as the host for subcloning using plasmids pUC19 and pUC118. *E. coli* EPI300 and DH5α were cultured in Luria-Bertani (LB) medium at 37°C, and *E. coli* strain JB525-MT102 (pJBA132) (2) was cultured at 30°C. When necessary antibiotics were supplied in the following concentrations: 12.5 μg ml^{-1} chloramphenicol (Cm), and 20 μg ml^{-1} tetracycline (Tc).

Soil and activated sludge samples. Activated sludge was collected from the aeration tank of a coke plant wastewater treatment facility in Japan (41) and the soil sample was collected from a forest in the National Institute of Advanced Industrial Science and Technology (AIST; Tsukuba city, Japan), located at 36° 3' 48.7614"N and 140° 7' 46.9878"E, on November 22, 2007 (soil temperature, 15°C at 5 cm depth). The soil sample was sieved (2 mm mesh size) to remove fine roots, leaves and other organic debris. After sampling, the soil and the activated sludge were frozen and stored at -80°C until DNA could be extracted.

Construction of metagenomic libraries. Metagenomic DNA from soil and sludge was extracted as described previously using sodium dodecyl sulfate and proteinase K (23). Extracted DNA was purified, fractionated (around 40 kb) and ligated into pCC1FOS for fosmid cloning following the methods of Rondon *et al*. (35). Fosmid libraries from sludge and soil were constructed using the CopyControl™ Fosmid Library Production Kit

(Epicentre) according to manufacturer's protocols. The transformants were selected on LB agar plates containing 12.5 μg ml^{-1} chloramphenicol (LB/Cm) (22). The presence of recombinant fosmids and polymorphisms of the insert DNA were examined by agarose gel electrophoresis of EcoRI digested purified fosmids from randomly selected *E. coli* transformants.

Approximately 103,700 colonies from the LB plates were inoculated into 500 μl of LB/Cm (using ~1,080 96-well plates) and grown at 37°C with shaking overnight. Then overnight cultures, in groups of 48 clones per well, were dispensed into 96-well microtiter plates (Becton Dickinson) (total of 23 96-well plates, one plate for every 4,608 clones). These plates were stored at -80°C and called the 96-well format libraries.

Screening for biosensor-inducing clones from metagenomic libraries. To isolate metagenomic clones containing putative synthase genes for QS signals, the *luxI-luxR*-GFP indicator system-*E. coli* JB525-MT102 harboring the *gfp* plasmid pJBA132 was used as the biosensor strain. Strain *E. coli* JB525-MT102 produces an unstable GFP in response to *N*-butyryl-L-homoserine lactone (C4-HSL) at a concentration of 1,000 nM, whereas lower levels are needed for *N*-3-oxohexanoyl-L-homoserine lactone (3-oxo-C6-HSL) at 1 nM, *N*-hexanoyl-L-homoserine lactone (C6-HSL) at 10 nM, *N*-octanoyl-L-homoserine lactone (C8-HSL) at 10 nM, and *N*-3-oxododecanoyl-L-homoserine lactone (3-oxo-C12-HSL) at 10 nM (2).

Clone mixtures (~0.1 μl) from each well of the 96-well format libraries were inoculated into a set of 96-well plates containing 100 μl of LB/Cm and incubated at 37°C for 18 h without agitation. Then 400 μl of fresh LB/Cm was added to each well and the plates were incubated at 37°C for 30 min with vigorous agitation at 230 rpm. Subsequently, 0.5 μl of CopyControl Induction Solution (Epicentre) was added to each

well. The solution increased the fosmid copy numbers in *E. coli* to help promote protein expression in the cells. Cultures were grown with agitation (230 rpm) at 37°C for 5 h and then cells were removed by centrifugation (1,109 × g, 10 min, 4°C). Supernatant (100 µl) from each well was transferred into fresh 96-well microtiter plates. Then 100 µl of the reporter, strain JB525-MT102 grown overnight and diluted 5-fold, was added in equivalent numbers into each well of the 96-well microtiter plates and mixed with the supernatants (average pH 6.6). These plates were incubated at 30°C for 4 h without agitation to induce detectable GFP expression from the reporter cells. After incubation, reporter cells were collected by centrifugation (1,109 × g, 10 min, 4°C) and resuspended in 50 µl of sterile water to eliminate background autofluorescence of LB medium. A SpectraMax Gemini XS Microplate Spectrofluorometer (Molecular Devices, Sunnyvale, CA) was used to measure GFP fluorescence of the cell suspension at an excitation wavelength of 485 nm and an emission wavelength of 538 nm. Between well comparisons were made by normalizing relative fluorescence unit (RFU) measurements to the OD_{600} of the reporter strain *E. coli* JB525-MT102 added to each well. The wells with fluorescence intensity greater than that of the negative control based on *t*-test ($P < 0.01$) were selected for further analyses. An *E. coli* strain harboring fosmid pCC1FOS without any inserted DNA was used as the negative control in this screening. To isolate individual biosensor-activating clones from wells with fluorescence, the 48 clone mixtures were diluted and cultured on LB/Cm agar plates at 37°C. After incubation, 96 colonies were randomly picked, cultured and screened for GFP fluorescence intensity greater than the control.

Sequence analysis. The genes responsible for production of the biosensor-activating molecules were subcloned using a series of standard recombinant

DNA techniques that included BamHI or KpnI restriction enzyme digestion, cloning and transformation (37). Clones were sequenced using a 454 Life Sciences pyrosequencer. Sequences were assembled using the Roche Newbler program. The program ORF finder from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) was used to identify open reading frames (ORF) that started with ATG, CTG, or TTG, and terminated with TAG, TGA or TAA. To annotate the fosmids ORFs with putative genes longer than 1,000 bp were translated and compared to the NCBI database using the BLASTP program (1). The nucleotide sequences of inserts from clones N16 and N52 reported in this study were deposited in the GenBank database with the accession numbers AB649138 and AB649139, respectively.

Phylogenetic analysis. The phylogenetic positions of two pairs of LuxI/LuxR homologues, AubI/AubR and AusI/AusR, were determined through alignment against amino acid sequences of LuxI and LuxR families (73 sequences in total, respectively) whose functions have been experimentally determined, together with homologous but functionally undetermined sequences (27 sequences of LuxI and LuxR homologues) of the top BLASTP hits (*E*-values < $1e^{-11}$) against the NCBI non-redundant protein sequences database. In total, 102 LuxI family sequences including AubI and AusI and 102 LuxR family sequences including AubR and AusR were independently aligned, and used to construct phylogenetic trees. Amino acid sequences were aligned using ClustalW (43), and any gap-containing columns in the alignment were removed. Construction of phylogenetic trees using the neighbor-joining (36) and maximum likelihood methods was performed in the MEGA program (42), and bootstrap analyses were performed with 1,000 replicates in each algorithm.

Extraction, purification and structure analysis of the QS signals in clones N16

193 **and N52**. Clones N16 and N52 were cultured in LB medium (20 L) at 30°C for 3 days
194 with shaking (180 rpm). The active molecule was solid phase extracted from the culture
195 broth using a column packed with Diaion HP-20 (Mitsubishi Chemical Co., Tokyo, Japan),
196 then eluted from the resin with ethanol. Concentrated extract was subjected to solvent
197 partitioning between water and CH_2Cl_2, and the CH_2Cl_2 fraction was chromatographed on
198 a silica gel column (CH_2Cl_2:MeOH = 20:1). Fractions containing QS signal, confirmed
199 using the biosensor assay, were combined and further purified by silica gel column
200 chromatography resulting in a pure QS signal compound. The purified QS signals were
201 subjected to reversed-phase analytical high-performance liquid chromatography (HPLC)
202 using a Cosmosil5 C_{18}-MS-II column (4.6 by 250 mm, NacalaiTesque, Kyoto, Japan) and
203 an elution gradient of 30 to 100% MeOH at a 0.6 ml min^{-1} flow rate for 30 min. Detection
204 was made at 220 nm with a retention time of 27.1 min. Fast atom bombardment (FAB)
205 mass spectra were measured on a JEOL JMS-700 MStation mass spectrometer using
206 glycerol as a matrix. Optical rotation was determined on a Jasco DIP-1000 digital
207 polarimeter in MeOH. Nuclear magnetic resonance (NMR) spectra were recorded on a
208 JEOL JNM-A500 NMR spectrometer at 500 MHz for ^{1}H, 125 MHz for ^{13}C in $CDCl_3$.
209 Chemical shifts of ^{1}H and ^{13}C in NMR spectra were determined relative to solvent
210 reference peaks δ_H 7.24 and δ_C 79.0 for $CDCl_3$.
211
212 **RESULTS AND DISCUSSION**
213 **Screening of biosensor-inducing clones**. Metagenomic libraries constructed using
214 DNA from activated sludge and soil samples were composed of 4×10^4 and 6×10^4 clones,
215 respectively. The average insert size was estimated to be 40 kb by restriction enzyme
216 analysis of randomly selected clones. The predicted length of total cloned DNA was 1.6

Gb in the activated sludge library and 2.4 Gb in the soil library. Approximately 103,700 metagenomic fosmid clones in 96-well microtiter plate format libraries were inspected using a microplate spectrofluorometer to identify clones that induced GFP expression of the biosensor strain *E. coli* JB525-MT102. Of all investigated clones, three clones induced GFP expression significantly greater than the biosensor strain after 4 h of incubation (Fig. 1). These clones were isolated from the soil metagenomic library (clone N16) ($P = 0.005$) and the activated sludge metagenomic library (clone N43 and N52) (*P*-values 0.001 and 0.009, respectively). Sequence analysis of clone N43 indicated there were no homologues of AHL synthases therefore further analyses were not performed at this time.

Sequence analysis of clone N16. Sequence analysis revealed fosmid pN16 contained a 34,076 bp insert that was assembled using over 16,000 sequences resulting in 121 × coverage. The insert was comprised of 28 ORFs (Fig. 2A and Table S1). A 5-kb region at the 3' end of the insert contained *luxI* and *luxR* homologues (*orf26* and *orf25*, respectively) oriented in convergent transcriptional directions (Fig. 2A). The highest amino acid sequence similarity to the LuxI homologue (ORF26), designated AubI (237 amino acids), was 42% to a putative LuxI in *Geobacter uraniireducens* Rf4 (39) (Fig. 2A, Table S1). The whole genome of the δ-*Proteobacteria, G. uraniireducens* Rf4, has been sequenced but the function of *luxI* and *luxR* homologues present in its genome has not been experimentally tested. The AubI protein also had 40% similarity to the QS signal synthase protein found in the nearly completed genome sequence of uncultivated *Leptospirillum* group II bacteria (40). The highest protein sequence similarity to a characterized LuxI homologue was 38% to CviI in *Chromobacterium violaceum* ATCC 31532 (28) belonging to the β-*Proteobacteria*. The transcriptional regulatory gene (*aubR*) was adjacent to the *aubI* gene. There was 33% amino acid similarity between AubR and the LuxR family

241 transcriptional regulator in the genome sequence of *Geobacter* sp. FRC-32. The deduced
242 amino acid sequence of *aubR* (267 amino acids) had low similarity to CviR (31%), one of
243 the most studied LuxR family transcriptional regulator, from *C. violaceum* (28). The low
244 similarities of the AubI and AubR to any previously described LuxI/LuxR homologues
245 found in the *Proteobacteria* suggest that they represent a new evolutionary branch of QS
246 systems.
247 Furthermore, genes related to two-component sensor histidine kinase (*orf27*) and
248 signal transduction histidine kinase (*orf28*) were found next to the *aubI/aubR* genes. The
249 presence of these histidine kinases may indicate the presence of a second QS system in
250 this microbe. A membrane-bound hybrid sensor kinase protein LuxN used instead of
251 LuxIR to recognize AHL was first described from the luminous marine bacterium *Vibrio*
252 *harveyi* (18). Both ORF27 and ORF28 have no amino acid sequence similarities to LuxN
253 homologues but a LuxI synthase annotated as a signal transduction histidine kinase has
254 been found in the methanogenic archaeon (49). Therefore, the presence of histidine
255 kinases are being noted for potential future studies of alternative or additional QS systems.
256 There were no phylogenetic marker genes in the pN16 insert DNA that could be used
257 to identify the potential source of the DNA. However, 23 out of 28 ORFs (from *orf1* to
258 *orf16* and *orf18* to *orf24*) showed high amino acid sequence similarity (56% to 89%) and
259 gene arrangement to uncultivated *Candidatus* Nitrospira defluvii (24) and the remaining
260 five ORFs containing AubI and AubR had low similarity (31% to 42%) to previously
261 described proteins from the phyla *Chloroflexi*, *Actinobacteria* and *Proteobacteria*. It is
262 possible that the DNA fragment of fosmid pN16 originated from a bacterium belonging to
263 the phylum *Nitrospirae* including *Ca*. Nitrospira defluvii or *Leptospirillum* group II
264 bacteria. In addition, a new LuxIR-type QS system has been recently identified in the

265 methanogenic archaeon (49). These facts strongly indicate that bacterial phyla other than
266 the *Proteobacteria*, and even other archaeon may possess LuxIR-type QS systems that are
267 similar to the system in *Proteobacteria*.

268 **Sequence analysis of clone N52**. The nucleotide sequence of fosmid pN52 was
269 37,553 bp long and was assembled using over 11,000 sequences resulting in 78×
270 coverage (Fig. 2B and Table S2). The insert was composed of at least 35 intact ORFs and
271 a truncated ORF. The deduced gene product of *ausI* (*orf4*, 202 amino acids) from clone
272 N52 was 57% similar to the LuxI family protein $LuxI_{QS6-1}$ of a metagenomic clone from
273 activated sludge (21) (Fig. 2B and Table S2). The highest amino acid sequence similarity
274 to a characterized LuxI homologue was to PpuI (40%) from *Pseudomonas putida*. The
275 transcriptional regulatory gene, designated *ausR* (*orf3*), was adjacent to the *ausI* gene in
276 the same transcriptional orientation. The deduced gene product AusR (255 amino acids)
277 was 37% similar to the LuxR family transcriptional regulatory protein $LuxR_{QS10-1}$ in a
278 metagenomic clone from forest soil (21). Among experimentally characterized LuxR
279 homologues, BraR in *Burkholderia kururiensis* had the highest similarity (35%) to AusR.

280 Of the 36 ORFs on pN52, 32 had significant similarity to proteins from
281 β-*Proteobacteria*, such as *Alcaligenes faecalis* and *Delftia* sp. (Table S2). A large gene
282 cluster (*orf11* to *30*) upstream of *ausI* and *ausR* was highly similar to genes responsible for
283 the complete metabolism of phenol to TCA-cycle intermediates via the *meta*-cleavage
284 pathway (4, 50). This suggests that these sequences were potentially obtained from a
285 phenol-degrading bacterium, which is highly possible because of the organic pollutants in
286 the coke plant wastewater treatment system where it originated. The effect of AHLs on
287 phenol degradation rates in an activated sludge has been reported (44). Genes related to
288 recombination (*orf1*) and transposition (*orf34-36*) were found flanking the phenol

289 catabolic gene cluster and *ausI/ausR*. These mobile genetic elements indicate that this
290 region may have been acquired by horizontal gene transfer making it difficult to confirm
291 the bacterial source of this fosmid.

292 **Comparison between *luxI* and *lux* box-like element sequences in clones N16 and**
293 **N52.** Comparison of sequences of the two metagenomic fragments from pN16 and pN52
294 indicated that these DNA fragments were derived from different organisms. The deduced
295 gene product of *aubI* from fosmid pN16 was 30% similar to the deduced gene product of
296 *ausI* from fosmid pN52. The G+C content in pN16 (56%) was lower than in pN52 (65%).
297 Promoter prediction software BPROM (Softberry, Mt. Kisco, NY, USA) identified
298 possible σ^{70} promoters in the upstream regions of both AHL-synthesizing enzyme genes,
299 *aubI* and *ausI* (Fig. 3). In addition, a 20-bp length inverted repeat sequence was found
300 close to the -35 box of each σ^{70} promoter. This inverted repeat is typically considered to be
301 a *lux* box-like element where a LuxR regulator binds to the LuxI/LuxR-type QS systems
302 (3, 13, 38).
303 The arrangement of the structural and regulatory genes on the two-fosmid clones was
304 similar but the orientation of transcription differed. The *luxI* genes (*ausI* and *aubI*) were
305 adjacent to their putative *luxR* regulatory genes (*ausR* and *aubR*). But the *aubI* and *aubR*
306 genes were in convergent orientation whereas *ausI* and *ausR* were tandem. Adjacent *luxI*
307 and *luxR* genes have been frequently observed not only in cultivated bacteria but also
308 uncultured bacteria (7, 19, 21, 47). The putative LuxR homologues found in this study
309 would likely form a complex with the AHL produced by the LuxI homologues. The
310 complexes would bind to the *lux* box-like element found upstream of *luxI* to regulate its
311 expression in the same manner as in previously studied LuxI/LuxR-type QS systems.

312 **Phylogenetic analysis of the AubI/AubR and AusI/AusR**. Phylogenetic analysis
313 using neighbor-joining trees generated from the LuxI and LuxR sequences revealed that
314 AusI and AusR sequences from clone N52 clustered with two metagenomic LuxIR
315 sequences ($LuxIR_{QS6-1}$ and $LuxIR_{QS10-1}$) within the β- and γ-*Proteobacteria* (Fig. 4). In
316 contrast, AubI and AubR sequences from clone N16 were found to be distant from any
317 LuxI or LuxR homologues whose function to synthesize AHL has been demonstrated.
318 Putative LuxI/LuxR homologues found in the δ-*Proteobacteria* were also distant from
319 other well-studied *Proteobacteria*, and formed a distinct clade. However, AubI and AubR
320 were more deeply branched than the delta-proteobacterial clade, and also clearly distinct
321 from a putative homologue found in the *Nitrospirae*, *Leptospirillum* group II bacteria (Fig.
322 4). The LuxI and LuxR trees also indicated that the majority of proteins from the α-, β- and
323 γ-*Proteobacteria* did not make distinct clades (Fig. 4) that reflect their phylogeny based
324 on 16S rRNA gene sequences. The phylogenetic positions of AusI/AusR and AubI/AubR
325 were similar in neighbor-joining trees and maximum likelihood trees (Fig. S1).
326 The phylogenetic analysis based on amino acid sequences indicated that AubI/AubR
327 was clearly distant from those in the *Proteobacteria* and deeply branched in the
328 phylogenetic tree. There has been a hypothesis that *luxI*/*luxR* genes frequently spread by
329 horizontal gene transfer within the α-, β- and γ-*Proteobacteria* (20, 26). The remote
330 phylogenetic position of AubI/AubR suggests that the genes were not recently transferred
331 from the *Proteobacteria* to the *Nitrospirae*. This critical finding also indicates that the QS
332 systems did not originate within the *Proteobacteria* but likely evolved prior to the
333 divergence of the phyla *Proteobacteria* and *Nitrospirae*. Whereas, the *ausI* and *ausR*
334 genes for the other positive clone, N52, were also novel but fell within the
335 β-*Proteobacteria*.

Characterization of quorum sensing signals from clones N16 and N52. Structural analyses of the QS signal identified from fosmid clones N16 and N52 using ^{1}H NMR, ^{13}C NMR and FAB-MS spectra revealed that both of them direct the synthesis of an AHL signal, C12-HSL (Fig. S2 and S3). The proton NMR spectrum of the active molecule showed typical AHL signals (Fig. S2A and S3A). There were no olefinic protons observed in the spectrum, suggesting that the acyl-chain was saturated. In the ^{13}C NMR spectrum there were only two signals around the carbonyl carbon region indicating the absence of a keto-group on the acyl-chain (Fig. S2B and S3B). This speculation was supported by 2D NMR data including COSY, HMQC, and HMBC (data not shown). FAB-MS analysis clearly determined the chain length to be C12 based on ion peaks at *m/z* 282.2 (negative mode) (Fig. S2C and S3C). Both ^{1}H and ^{13}C NMR spectra of the synthetic C12-HSL could be superimposed with data of the compound from these metagenomic clones. Additionally, analytical reversed-phase HPLC and NMR analysis confirmed its identity with authentic C12-HSL with a retention time of 27.1 min. Finally, comparison of the optical rotation value with that of a synthetic standard confirmed its absolute stereochemistry was a 2*S* configuration. The *aubI* gene from clone N16 and *ausI* from clone N52 have been subcloned and structural analyses of each sub-clone showed similar production of C12-HSL (data not shown). Although sequence similarity was low between AubI and AusI, their QS signals were identical to C12-HSL (Fig. S2 and S3).

The deduced amino acid sequences from clone N16 supported our hypothesis that it originated from an uncultured bacterium belonging to the phylum *Nitrospirae* (40). However, the regulatory gene arrangements and structure of the QS signal produced were similar to the LuxIR-type QS system mediated by an AHL (C12-HSL). In conclusion, these results demonstrate that metagenomics can be used to discover LuxI/LuxR

360 homologues from yet-to-be cultivated bacterial phyla in order to gain a greater
361 understanding of bacterial interactions in natural ecosystems.
362
363 **REFERENCES**

364 1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and**
365 **D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein
366 database search programs. Nucleic Acids Res.**25:**3389-3402.
367 2. **Andersen, J. B., A. Heydorn, M. Hentzer, L. Eberl, O. Geisenberger, B. B.**
368 **Christensen, S. Molin, and M. Givskov.** 2001. *gfp*-based *N*-acyl
369 homoserine-lactone sensor systems for detection of bacterial communication. Appl.
370 Environ. Microbiol.**67:**575-585.
371 3. **Antunes L. C. M., R. B. R. Ferreira, C. P. Lostroh, and E. P. Greenberg.** 2008. A
372 mutational analysis defines *Vibrio fischeri* LuxR binding sites. J. Bacteriol.
373 **13:**4392-4397.
374 4. **Arai, H., T. Ohishi, M. Y. Chang, and T. Kudo.** 2000. Arrangement and regulation
375 of the genes for meta-pathway enzymes required for degradation of phenol in
376 *Comamonas testosteroni* TA441. Microbiology **146:**1707-1715.
377 5. **Bassler, B. L., M. Wright, and M. R. Silverman.** 1994. Multiple signalling systems
378 controlling expression of luminescence in *Vibrio harveyi*: sequence and function of
379 genes encoding a second sensory pathway. Mol. Microbiol.**13:**273-286.
380 6. **Boyer, M., and F. Wisniewski-Dyé.** 2009. Cell-cell signalling in bacteria: Not
381 simply a matter of quorum. FEMS Microbiol.Ecol.**70:**1-19.
382 7. **Brint, J. M., and D. E. Ohman.** 1995. Synthesis of multiple exoproducts in
383 *Pseudomonas aeruginosa* is under the control of RhlR-RhlI, another set of regulators

in strain PAO1 with homology to the autoinducer-responsive LuxR-LuxI family. J. Bacteriol.**177:**7155-7163.

8. **Burr, T., A. M. L. Barnard, M. J. Corbett, C. L. Pemberton, N. J. L. Simpson, and G. P. C. Salmond.** 2006. Identification of the central quorum sensing regulator of virulence in the enteric phytopathogen, *Erwinia carotovora*: the VirR repressor. Mol. Microbiol.**59:**113-125.
9. **Burton, E. O., H. W. Read, M. C. Pellitteri, and W. J. Hickey.** 2005. Identification of acyl-homoserine lactone signal molecules produced by *Nitrosomonas europaea* strain Schmidt. Appl. Environ. Microbiol.**71:**4906-4909.
10. **Chong, G., O. Kimyon, S. A. Rice, S. Kjelleberg, and M. Manefield.** 2012. The presence and role of bacterial quorum sensing in activated sludge. Microbiol. Biotech.**5**:621-633.
11. **d'Angelo-Picard, C., D. Faure, I. Penot, and Y. Dessaux.** 2005. Diversity of *N*-acyl homoserine lactone-producing and -degrading bacteria in soil and tobacco rhizosphere. Environ. Microbiol.**7:**1796-1808.
12. **Decho, A. W., R. S. Norman, and P. T. Visscher.** 2010. Quorum sensing in natural environments: emerging views from microbial mats. Trends in Microbiol.**18:**73-80.
13. **Devine, J. H., C. Countryman, and T. O. Baldwin.** 1988. Nucleotide sequence of the *luxR* and *luxI* genes and structure of the primary regulatory region of the *lux* regulon of *Vibrio fischeri* ATCC 7744. Biochemistry **27:**837-842.
14. **Dunphy, G., C. Miyamoto, and E. Meighen.** 1997. A homoserine lactone autoinducer regulates virulence of an insect-pathogenic bacterium, *Xenorhabdus nematophilus* (Enterobacteriaceae). J. Bacteriol.**179:**5288-5291.
15. **Eberl, L., and K. Riedel.** 2011. Mining quorum sensing regulated proteins-Role of

bacterial cell-to-cell communication in global gene regulation as assessed by proteomics. Proteomics **11:**3070-3085.

16. **Engebrecht, J. A., and M. Silverman.** 1987. Nucleotide sequence of the regulatory locus controlling expression of bacterial genes for bioluminescence. Nucleic Acids Res.**15:**10455-10467.
17. **Flavier, A. B., L. M. Ganova-Raeva, M. A. Schell, and T. P. Denny.** 1997. Hierarchical autoinduction in *Ralstonia solanacearum*: control of acyl-homoserine lactone production by a novel autoregulatory system responsive to 3-hydroxypalmitic acid methyl ester. J. Bacteriol.**179:**7089-7097.
18. **Freeman, J. A., B. N. Lilley, and B. L. Bassler.** 2000. A genetic analysis of the functions of LuxN: a two-component hybrid sensor kinase that regulates quorum sensing in *Vibrio harveyi*. Mol. Microbiol.**35:**139-149.
19. **Fuqua, W. C., S. C. Winans, and E. P. Greenberg.** 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J. Bacteriol.**176:**269-275.
20. **Gray, K. M., and J. R. Garey.** 2001. The evolution of bacterial LuxI and LuxR quorum sensing regulators. Microbiology **147:**2379-2387.
21. **Hao, Y., S. C. Winans, B. R. Glick, and T. C. Charles.** 2010. Identification and characterization of new LuxR/LuxI-type quorum sensing systems from metagenomic libraries. Environ. Microbiol.**12:**105-117.
22. **Kimura, N., K. Sakai, and K. Nakamura.** 2010. Isolation and Characterization of a 4-Nitrotoluene-Oxidizing Enzyme from Activated Sludge by a Metagenomic Approach. Microbes Environ.**25:**133-139.
23. **Kimura, N., Y. Shinozaki, T. H. Lee, and Y. Yonezawa.** 2003. The microbial

community in a 2,4-dinitrophenol-digesting reactor as revealed by 16S rDNA gene analysis. J. Biosci. Bioeng.**96:**70-75.

24. **Lücker, S., M. Wagner, F. Maixner, E. Pelletier, H. Koch, B. Vacherie, T. Rattei, J. S. S. Damsté, E. Spieck, and D. Le Paslier.** 2010. A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. Proc. Nat. Acad.Sci. USA.**107:**13479-13484.
25. **Labbate, M., S. Y. Queck, K. S. Koh, S. A. Rice, M. Givskov, and S. Kjelleberg.** 2004. Quorum sensing-controlled biofilm development in *Serratia liquefaciens* MG1. J. Bacteriol.**186:**692-698.
26. **Lerat, E., and N. A. Moran.** 2004. The evolutionary history of quorum-sensing systems in bacteria. Mol. Biol. Evol.**21:**903-913.
27. **Lewenza, S., B. Conway, E. Greenberg, and P. A. Sokol.** 1999. Quorum sensing in *Burkholderia cepacia*: identification of the LuxRI homologs CepRI. J. Bacteriol.**181:**748-756.
28. **McClean, K. H., M. K. Winson, L. Fish, A. Taylor, S. R. Chhabra, M. Camara, M. Daykin, J. H. Lamb, S. Swift, B. W. Bycroft, G. S. Stewart, and P. Williams.** 1997. Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of *N*-acylhomoserine lactones. Microbiology **143:**3703-3711.
29. **McLean, R. J., M. Whiteley, D. J. Stickler, and W. C. Fuqua.** 1997. Evidence of autoinducer activity in naturally occurring biofilms. FEMS Microbiol. Lett.**154:**259-263.
30. **Nakazawa, H., A. Arakaki, S. Narita-Yamada, I. Yashiro, K. Jinno, N. Aoki, A. Tsuruyama, Y. Okamura, S. Tanikawa, and N. Fujita.** 2009. Whole genome

sequence of *Desulfovibrio magneticus* strain RS-1 revealed common gene clusters in magnetotactic bacteria. Genome Res.**19:**1801-1808.

31. **Passador, L., J. M. Cook, M. J. Gambello, L. Rust, and B. H. Iglewski.** 1993. Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication. Science **260:**1127-1130.
32. **Pinto, U. M., E. de Souza Viana, M. L. Martins, and M. C. D. Vanetti.** 2007. Detection of acylated homoserine lactones in gram-negative proteolytic psychrotrophic bacteria isolated from cooled raw milk. Food control **18:**1322-1327.
33. **Rappe, M. S., and S. J. Giovannoni.** 2003. The uncultured microbial majority. Annual Rev. Microbiol.**57:**369-394.
34. **Rivas, M., M. Seeger, D. S. Holmes, and E. Jedlicki.** 2005. A Lux-like Quorum Sensing System in the Extreme Acidophile, *Acidithiobacillus ferrooxidans*. Biol. Res.**38:**283-297.
35. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol.**66:**2541-2547.
36. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol.**4:**406-425.
37. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular Cloning : A Laboratory Manual, 2nd ed. Cold Spring Harbor, NY.
38. **Schuster, M., M. L. Urbanowski, and E. P. Greenberg.** 2004. Promoter specificity in *Pseudomonas aeruginosa* quorum sensing revealed by DNA binding of purified

480 LasR. Proc. Nat. Acad. Sci. USA.**101:**15833-15839.

481 39. **Shelobolina, E. S., H. A. Vrionis, R. H. Findlay, and D. R. Lovley.** 2008.
482 *Geobacter uraniireducens* sp. nov., isolated from subsurface sediment undergoing
483 uranium bioremediation. Int. J. Syst. Evol. Microbiol.**58:**1075-1078.

484 40. **Simmons, S. L., G. Dibartolo, V. J. Denef, D. S. Goltsman, M. P. Thelen, and J. F.**
485 **Banfield.** 2008. Population genomic analysis of strain variation in *Leptospirillum*
486 group II bacteria involved in acid mine drainage formation. PLoS Biol.**6:**1427-1442.

487 41. **Suenaga, H., T. Ohnuki, and K. Miyazaki.** 2007. Functional screening of a
488 metagenomic library for genes involved in microbial degradation of aromatic
489 compounds. Environ. Microbiol.**9:**2289-2297.

490 42. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular
491 Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol.
492 Evol.**24:**1596-1599.

493 43. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving
494 the sensitivity of progressive multiple sequence alignment through sequence
495 weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids
496 Res.**22:**4673-4680.

497 44. **Valle, A, M. J. Bailey, A. S. Whiteley, and M. Manefield.** 2004.
498 *N*-acyl-L-homoserine lactones (AHLs) affect microbial community composition and
499 function in activated sludge. Environ. Microbiol.**6**:424-433.

500 45. **Whitehead, N. A., A. M. Barnard, H. Slater, N. J. Simpson, and G. P. Salmond.**
501 2001. Quorum-sensing in Gram-negative bacteria. FEMS Microbiol.
502 Rev.**25:**365-404.

503 46. **Williams, P., K. Winzer, W. C. Chan, and M. Camara.** 2007. Look who's talking:

communication and quorum sensing in the bacterial world. Philos. Trans. R. Soc. B.**362:**1119-1134.

47. **Williamson, L. L., B. R. Borlee, P. D. Schloss, C. Guan, H. K. Allen, and J. Handelsman.** 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl. Environ. Microbiol.**71:**6335-6344.
48. **Xu, L., H. Li, C. Vuong, V. Vadyvaloo, J. Wang, Y. Yao, M. Otto, and Q. Gao.** 2006. Role of the *luxS* quorum-sensing system in biofilm formation and virulence of *Staphylococcus epidermidis*. Infect. Immun.**74:**488-496.
49. **Zhang, G., F. Zhang, G. Ding, J. Li, X. Guo, J. Zhu, L. Zhou, S. Cai, X. Liu, Y. Luo, G. Zhang, W. Shi, and X. Dong.** 2012. Acyl homoserine lactone-based quorum sensing in a methanogenic archaeon. The ISME J.**6**:1336-1344.
50. **Zhu, C., L. Zhang, and L. Zhao.** 2008. Molecular cloning, genetic organization of gene cluster encoding phenol hydroxylase and catechol 2, 3-dioxygenase in *Alcaligenes faecalis* IS-46. World J. Microbiol. Biotech.**24:**1687-1695.

FIGURE LEGENDS

FIG. 1. GFP fluorescence intensity of the biosensor induced by QS signal-producing clones. The biosensor detects small diffusible signal molecules that induce QS. The extracellular QS signals bind and form a complex with the LuxR transcriptional regulator. The complex binds to the promoter and activates the expression of GFP. Error bars represent the standard deviation of the mean for three replicates.

FIG. 2. Physical map and gene organization of fosmids pN16 (A) and pN52 (B). The arrows in the physical map indicate the size, location, and direction of transcription of the

ORFs. Black arrows represent *luxI* family genes; grey arrows indicate *luxR* family genes. Predicted protein functions of ORFs are listed below the map.

FIG. 3. Analysis of *lux* box-like elements. (A) Location of *lux* box-like elements upstream of *aubI* and *ausI*. (B) Comparison of the elements with known *lux* box-like elements from *Pseudomonas aeruginosa* RhlI and LasI (7, 31), *Ralstonia solanacearum* SolI (17), *Aliivibrio fischeri* LuxI (16), *Acidithiobacillus ferrooxidans* AfeI (34), *Burkholderia cepacia* CepI (27). Sequences with more than 60% identity are shaded.

FIG. 4. Phylogenetic trees of (A) LuxI and (B) LuxR protein sequence homologues obtained from *Proteobacteria* and *Nitrospirae*. Sequences obtained from metagenomic clones are indicated in bold type. AubI, AubR, AusI and AusR characterized in this study are enclosed in red boxes. Species belonging to the α-, β-, γ- and δ-*Proteobacteria* are grouped within brackets. The bold brackets represent two groups of phyla. The *Leptospirillum* sequence was used as the outgroup. Trees were constructed using the neighbor-joining algorithm with bootstrap values obtained from 1,000 replicates. Nodes with bootstrap value more than 50% are labeled. The scale bars represent 0.1 and 0.2 substitutions per amino acid site, respectively. Red branches represent sequences that have been experimentally tested. Abbreviations for bacterial genus names: *Ab*, *Acidithiobacillus*; *Ae*, *Aeromonas*; *Ag*, *Agrobacterium*; *Al*, *Aliivibrio*, *Az*, *Azospirillum*; *B*, *Burkholderia*; *C*, *Chromobacterium*; *D*, *Desulfovibrio*; *E*, *Erwinia*; *G*, *Geobacter*; *Ga*, *Gallionella*; *M*, *Mesorhizobium*; *Mb*, *Methylobacterium*; *Ni*, *Nitrosospira*; *Nc*, *Nitrococcus*; *P*, *Pseudomonas*; *Po*, *Polymorphum*; *R*, *Rhizobium*; *Ra*, *Ralstonia*; *Rb*, *Rhodobacter*; *Rp*, *Rhodopseudomonas*; *S*, *Sinorhizobium*; *Se*, *Serratia*; *V*, *Vibrio*; *Y*,

Yersinia. Accession numbers for all LuxI and LuxR homologues used in this analysis are listed in Table S3.

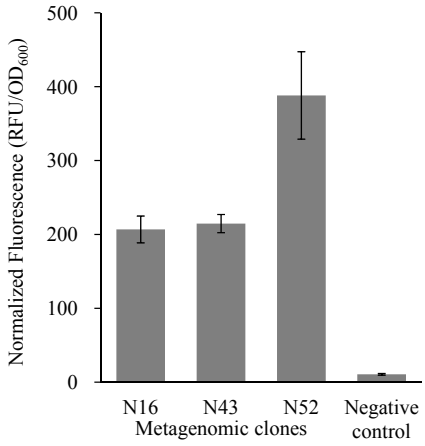
Figure S1. Phylogenetic trees of the protein sequences of LuxI (A) and LuxR (B) homologues constructed using maximum likelihood. Sequences obtained from metagenomic clones are indicated in bold type. AubI, AubR, AusI and AusR characterized in this study are enclosed in red boxes. Species belonging to the α-, β-, γ- and δ-*Proteobacteria* are grouped within brackets. The bold brackets represent two groups of phyla. The *Leptospirillum* sequence was used as the outgroup. Branch length and bootstrap values were obtained from 50 replicates using the maximum likelihood algorithm. Nodes with bootstrap values more than 50% are labeled. The scale bar represents 0.2 substitutions per amino acid site. The red branches are sequences with experimental evidence. Abbreviations for bacterial genus names: *Ab*, *Acidithiobacillus*; *Ae*, *Aeromonas*; *Ag*, *Agrobacterium*; *Al*, *Aliivibrio*, *Az*, *Azospirillum*; *B*, *Burkholderia*; *C*, *Chromobacterium*; *D*, *Desulfovibrio*; *E*, *Erwinia*; *G*, *Geobacter*; *Ga*, *Gallionella*; *M*, *Mesorhizobium*; *Mb*, *Methylobacterium*; *Ni*, *Nitrosospira*; *Nc*, *Nitrococcus*; *P*, *Pseudomonas*; *Po*, *Polymorphum*; *R*, *Rhizobium*; *Ra*, *Ralstonia*; *Rb*, *Rhodobacter*; *Rp*, *Rhodopseudomonas*; *S*, *Sinorhizobium*; *Se*, *Serratia*; *V*, *Vibrio*; *Y*, *Yersinia.* Accession numbers for all LuxI and LuxR homologues used in this analysis are listed in Table S3.

Figure S2. ^{1}H (A) and ^{13}C (B) NMR spectra and the negative-mode FAB-MS spectrum (C) of QS signal produced by clone N16.

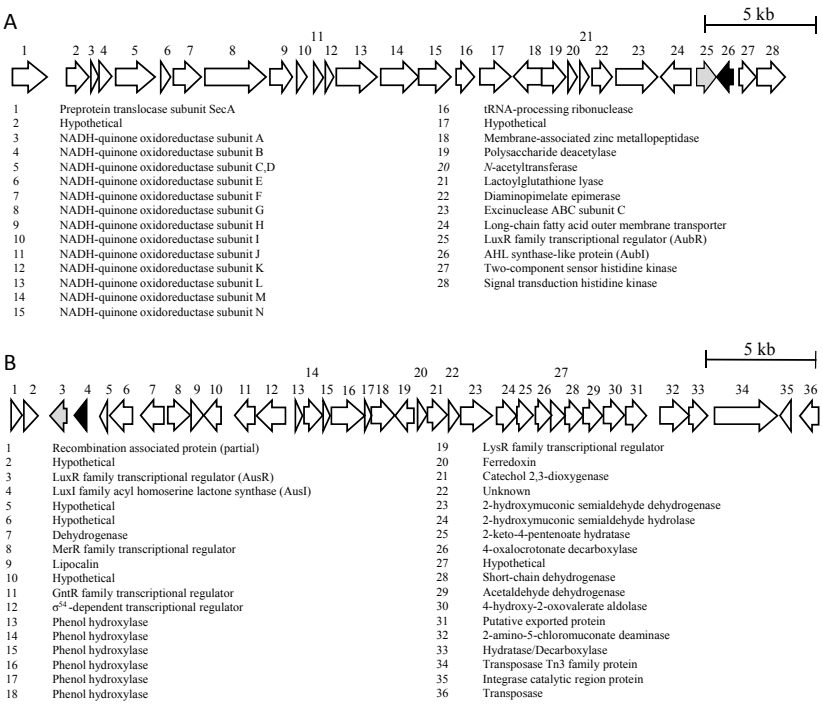
Figure S3. ^{1}H (A) and ^{13}C (B) NMR spectra and the negative-mode FAB-MS spectrum (C)

576 of QS signal produced by clone N52.

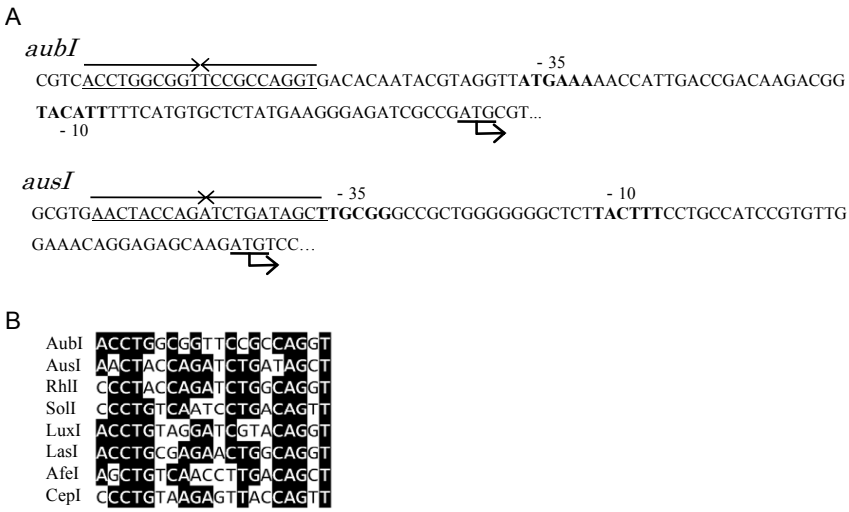
Nasuno *et al*. Fig. 1



Nasuno *et al*. Fig. 2



Nasuno *et al*. Fig. 3



Nasuno *et al*. Fig. 4