Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph *Nitrosomonas europaea*†

Patrick Chain,[1,2] Jane Lamerdin,[1,2] Frank Larimer,[1,3] Warren Regala,[1,2] Victoria Lao,[1,2]
Miriam Land,[1,3] Loren Hauser,[1,3] Alan Hooper,[4] Martin Klotz,[5] Jeanette Norton,[6]
Luis Sayavedra-Soto,[7] Dave Arciero,[4] Norman Hommes,[7]
Mark Whittaker,[4] and Daniel Arp[7]*

*Joint Genome Institute, Walnut Creek, California 94598[1]; Lawrence Livermore National Laboratory, Livermore, California 94550[2];
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831[3]; University of Minnesota, Minneapolis, Minnesota 55455[4];
University of Louisville, Louisville, Kentucky 40208[5]; Utah State University, Logan, Utah 84322[6]; and
Oregon State University, Corvallis, Oregon 97331[7]*

***Nitrosomonas europaea* (ATCC 19718) is a gram-negative obligate chemolithoautotroph that can derive all its energy and reductant for growth from the oxidation of ammonia to nitrite. *Nitrosomonas europaea* participates in the biogeochemical N cycle in the process of nitrification. Its genome consists of a single circular chromosome of 2,812,094 bp. The GC skew analysis indicates that the genome is divided into two unequal replichores. Genes are distributed evenly around the genome, with ~47% transcribed from one strand and ~53% transcribed from the complementary strand. A total of 2,460 protein-encoding genes emerged from the modeling effort, averaging 1,011 bp in length, with intergenic regions averaging 117 bp. Genes necessary for the catabolism of ammonia, energy and reductant generation, biosynthesis, and CO_2 and NH_3 assimilation were identified. In contrast, genes for catabolism of organic compounds are limited. Genes encoding transporters for inorganic ions were plentiful, whereas genes encoding transporters for organic molecules were scant. Complex repetitive elements constitute ca. 5% of the genome. Among these are 85 predicted insertion sequence elements in eight different families. The strategy of *N. europaea* to accumulate Fe from the environment involves several classes of Fe receptors with more than 20 genes devoted to these receptors. However, genes for the synthesis of only one siderophore, citrate, were identified in the genome. This genome has provided new insights into the growth and metabolism of ammonia-oxidizing bacteria.**

Nitrosomonas europaea is a bacterium that can derive all its energy and reductant for growth from the oxidation of ammonia to nitrite. The cell's demand for carbon has to be met almost entirely by the fixation of carbon dioxide. Additional mineral salts complete the cell's nutritional needs. Although this bacterium can incorporate small amounts of organic compounds into cellular biomass (19, 20, 59, 60), there is an obligate requirement for oxidation of ammonia and assimilation of inorganic nutrients to support growth. As such, this bacterium is a member of a small group of bacteria known as obligate chemolithoautotrophs.

N. europaea and other ammonia-oxidizing bacteria participate in the biogeochemical N cycle in the process of nitrification (the biological conversion of reduced nitrogen in the form of ammonia [NH_3] or ammonium [NH_4^+] to oxidized N in the form of nitrite [NO_2^-], nitrate [NO_3^-], or gaseous forms [NO, N_2O]). The gaseous products of nitrification, NO and N_2O, are "greenhouse gases," whereas the soluble forms are readily leached into ground and surface waters. In contrast to these detrimental consequences of ammonia oxidation, nitrifiers increase N availability to plants, are important to the treatment of wastewater, and have potential for the bioremediation of

sites contaminated with chlorinated aliphatic hydrocarbons (44, 45, 61, 76).

Ammonia-oxidizing bacteria such as *Nitrosomonas* convert NH_3 to NO_2^- by the successive action of ammonia monooxygenase (AMO) and hydroxylamine oxidoreductase (HAO):
$$NH_3 + O_2 + 2H^+ + 2e^- \rightarrow NH_2OH + H_2O \rightarrow NO_2^- + 5H^+ + 4e^-.$$
Two of the four electrons return to the AMO reaction, and two are either reductant for biosynthesis or pass to a terminal electron acceptor (41, 83).

N. europaea is a member of the proteobacteria, a large group of eubacteria of presumed photosynthetic ancestry (73). All terrestrial ammonia-oxidizing bacteria currently identified belong to the β-subdivision of the proteobacteria, including the *Nitrosomonas* and *Nitrosospira* genera, whereas marine strains are found in both the γ-subdivision *Nitrosococcus* genus and the β-subdivision genera (34, 82).

N. europaea is the best studied of the ammonia-oxidizing bacteria at the molecular level. Previous work has focused primarily on unique aspects of this bacterium's growth, namely, its ability to utilize NH_3, and the bioenergetics associated with this process. Thus, the pathway for NH_3 oxidation has received considerable attention at both the biochemical and the genetic levels. AMO is a heteromultimeric enzyme encoded by the *amoCAB* operon that exists in two strikingly similar copies in the *N. europaea* genome (9, 37, 47, 54, 55). HAO is a homotrimer in which each subunit harbors eight *c*-type hemes, one of which binds substrate. Electrons pass from HAO through two tetraheme *c* cytochromes (cytochromes c_{554} and c_{m552}) to

ubiquinone (79). Hence, one theoretical HAO-electron acceptor complex includes 48 *c*-hemes. Similar to the multiple copies of the *amo* operon, the *hao* gene is present in three nearly identical copies (55). Gene sequences for three cytochromes that participate in NH_3 oxidation have also been determined (7, 39), and some of these exist as multiple copies in the genome. In contrast to NH_3 oxidation, other cellular processes have received only limited attention. Analysis of the complete genome sequence has allowed new insights into all aspects of the growth and environmental responses of this ammonia-oxidizing bacterium.

MATERIALS AND METHODS

Construction, isolation, and sequencing of small-insert and large-insert libraries. Genomic DNA, isolated from the *N. europaea* (ATCC 19718), was sequenced by using a conventional whole-genome shotgun strategy (30). Briefly, random 2- to 3-kb DNA fragments were isolated after mechanical shearing. These gel-extracted fragments were concentrated, end repaired, and cloned into pUC18. Double-ended plasmid sequencing reactions were carried out by using PE BigDye Terminator chemistry (Perkin-Elmer, Foster City, Calif.) and sequencing ladders were resolved on PE 377 automated DNA sequencers. Two rounds (near 55,000 reads each) of small-insert library sequencing were performed, generating 17-fold redundancy.

A large-insert (~30 kb) fosmid library was also constructed by *Sau*3AI partial digestion of genomic DNA and cloning into the pFos1 cloning vector (46). End sequencing of approximately 300 fosmid clones (0.05-fold sequence redundancy) generated roughly 3-fold genome scaffold coverage. The fosmids were fingerprinted with *Eco*RI to aid in assembly verification and determination of gap sizes and provided a minimal scaffold used for order and orientation across assembly gaps.

Sequence assembly and gap closure. Sequence traces were processed with PHRED (27, 28) for base calling and assessment of data quality prior to assembly with PHRAP (P. Green, University of Washington) and visualization with CONSED (33). Even with 17-fold redundancy, the assembly consisted of 400 contigs. The previously unknown abundance of repetitive elements within the *N. europaea* genome, a feature which confuses most assembly programs, was found to be responsible for this large number of contigs. Due to the repetitive nature of this genome, many assembly gaps (~70%) were the direct result of misassembling these repeat elements, while the remainder were "true" gaps. Most repeats smaller than 2 kb, along with most of the other "true" gaps, were closed by primer walking on gap-spanning library clones (identified by using linking information from forward and reverse reads). Alternatively, some of the larger gaps, including the larger repeat elements and regions covered only by fosmid clones, were closed by primer walking on PCR products. Remaining physical (uncaptured) gaps, some of which are regions suspected of being lethal in *Escherichia coli* when in high copy number (e.g., the *amoCAB* promoter region), were closed by combinatorial (multiplex) PCR.

Assembly quality. Once a single contig assembly was achieved, the consensus genome sequence was screened for regions of single-clone coverage and low sequence quality. These regions were resequenced, usually either by primer walking on PCR products or on different templates. The finishing phase (gap closure, repeat resolution, and resequencing poor-quality regions) added an additional ~12,000 reads to the assembly. Based on the consensus quality scores generated by PHRAP, we estimate the overall error rate to be substantially less than one error in 10,000 bases.

Sequence analysis and annotation. Gene modeling was performed by using the Critica (3), Glimmer (23), and Generation (http://compbio.ornl.gov /generation/index.shtml) modeling packages; the results were combined, and a BLASTP search of the translations versus GenBank's nonredundant database (NR) was conducted. The alignment of the N terminus of each gene model versus the best NR match was used to pick a preferred gene model. If no BLAST match was returned, the longest model was retained. Gene models that overlapped by more than 10% of their length were flagged, giving preference to genes with a BLAST match. The revised gene-protein set was searched against the KEGG GENES, Pfam, PROSITE, PRINTS, ProDom, and COGs databases, in addition to the BLASTP versus NR databases. From these results, categorizations were developed by using the KEGG and COGs hierarchies. Initial criteria for automated functional assignment required a minimum 50% residue identity over 80% of the length of the match for BLASTP alignments, plus concurring evidence from pattern or profile methods. Putative assignments were made for

TABLE 1. Summary of *N. europaea* genome

Parameter	Value
No. of chromosomes	1
No. of bp	2,812,094
GC content (%)	50.7
Coding density (%)	88.4
No. of predicted protein coding genes	2,460
No. of predicted proteins with putative function	1,863
No. of unknown proteins with matches to other proteins of unknown function	285
No. of unknown proteins unique to *N. europaea*	312
No. of predicted RNA coding genes	47
No. of tRNAs	41
No. of 16S-Ala $tRNA_{TGC}$-Ile $tRNA_{GAT}$-23S-5S tRNAs	1
No. of genes in protein category[a]:	
Energy	103
Transport	285
DNA replication	212
Transcription	36
Translation	136
Regulation and signaling	187
Amino acid metabolism	120
Carbohydrate metabolism	55
Nucleotide metabolism	48
Cofactor metabolism	80
Lipid metabolism	48
Secondary metabolites	48
Cellular processes	481
General function	155

[a] Classifications of genes are based on clusters of orthologous genes (COGs).

identities down to 30%, over 80% of the length. The sequence, as well as the results of automatic annotations, is available online (http://genome.ornl.gov /microbial/neur/embl/).

Nucleotide sequence accession number. The sequence of the complete *N. europaea* strain ATCC 19718 is available under EMBL-EBI accession number AL954747.

RESULTS

Genome organization. The genome of *N. europaea* ATCC 19718 consists of a single circular chromosome of 2,812,094 bp (the general features of the genome are listed in Table 1, and a detailed map is shown in Fig. 1); nucleotide 1 was assigned at the predicted origin of replication. The GC skew analysis also reveals that the genome is clearly divided into two unequal replichores (roughly 1/3 to 2/3). The mechanism and biological reason for this asymmetry is unclear but may be a result of the abnormally large amount of repetitive material within the genome and presumed fluidity of the genome. One clear effect of this bias is an ~1% difference (25.84% versus 24.87%) in G versus C strand composition. Another peculiarity is that the multiple copies of genes participating in ammonia oxidation (*amoCAB* and *hao*), along with a large (7.5 kb) recently duplicated region, are all on the larger replichore (see Fig. 1). Overall, the *N. europaea* genome is 50.7% G+C. Although several spikes are observed, these are not postulated to have been acquired via lateral transfer. However, some of these G+C spikes do correlate with GC skew spikes. Interestingly, some of these spikes also correspond to the regions containing repeated genes and operons such as the *hao* gene region (HAO and associated cytochromes), *amoCAB* (AMO operon), and *tufB* (elongation factor Tu), all of which are themselves associated with several ribosomal proteins (not shown).

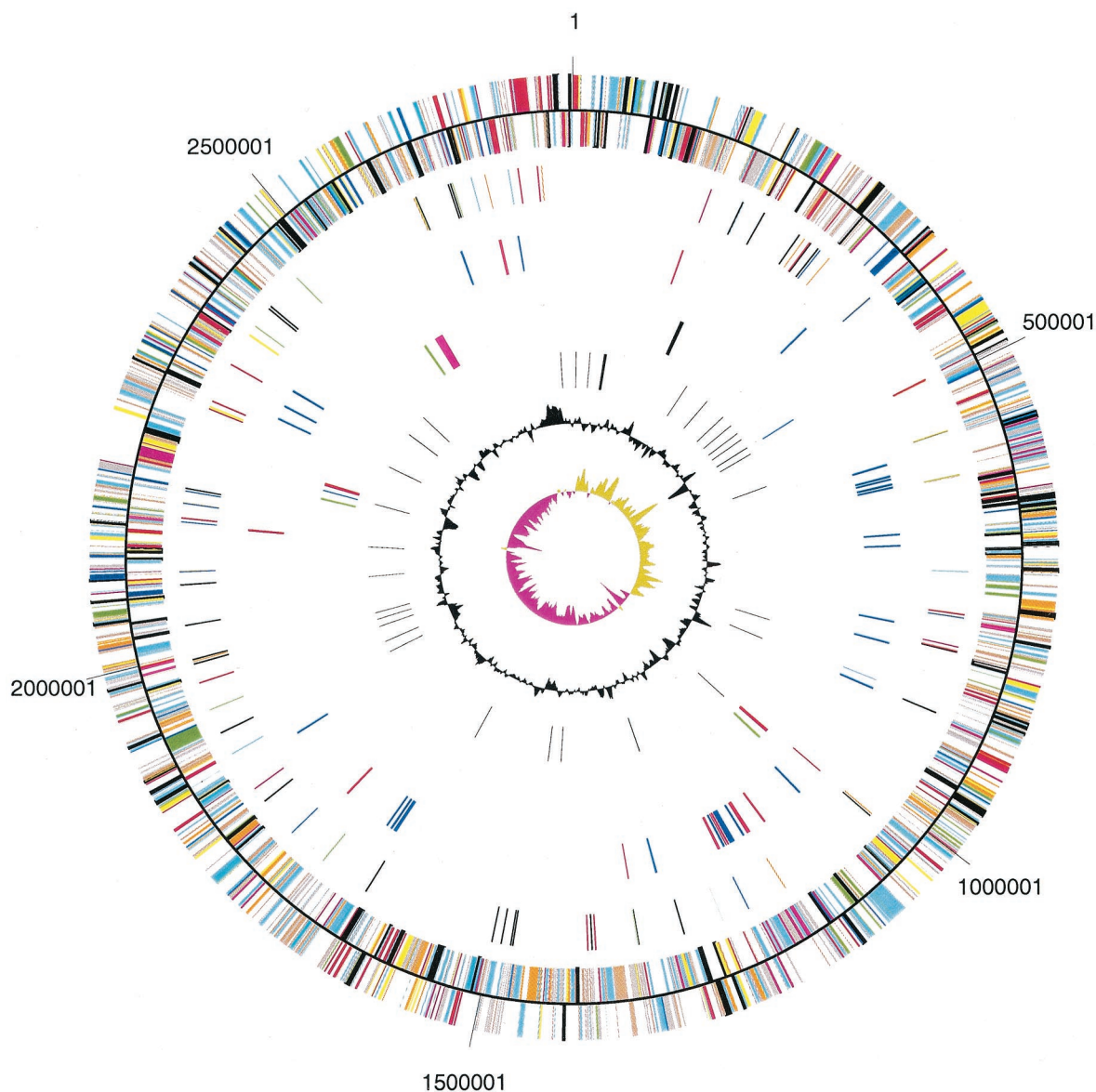Genes are distributed evenly around the genome, with

FIG. 1. Circular representation of the 2,812,094-bp genome of *N. europaea* ATCC 19718. The outer two circles represent protein-encoding and structural-RNA genes, plus and minus strand (green, energy metabolism; red, DNA replication; magenta, transcription; yellow, translation; orange, amino acid metabolism; dark blue, carbohydrate metabolism; pale red, nucleotide metabolism; black, coenzyme metabolism; cyan, lipid metabolism; light blue, cellular processes; brown, general function; gray, hypothetical and conserved hypothetical genes; pale green, structural RNAs). The third circle indicates the major IS element families as follows: black, ISne1 family; orange, ISne2; red, ISne3; green, ISne4; blue, ISne5; cyan, ISne6; magenta, ISne7; yellow, ISne8. The fourth circle indicates siderophore-receptor genes (blue) and pseudogenes (red). The fifth circle indicates *amo* (red), *hao* (green), the duplicated genes for EF-Tu (blue), the 7.5-kb tandem duplication (magenta) and the 339- and 318-bp repeats (black). RNA structural genes are indicated in the sixth circle. The inner two circles are the GC bias and GC skew.

~47% transcribed from the forward strand and ~53% from the complementary strand. A total of 2,460 protein-encoding genes emerged from the modeling effort, averaging 1,011 bp in length, with intergenic regions averaging 117 bp. These open reading frames (ORFs) account for 2,487,261 nucleotides of coding sequence (88.4%). An additional 113 ORFs are fragmentary, frameshifted, or interrupted by insertion sequence (IS) elements; these have been designated pseudogenes. Of the 2,460 putative proteins, 2,147 matched a sequence in the NR database with an e-score of <1e-5; of these, 1,863 have simi-

larity to a protein with a functional assignment, and 285 match a protein of unknown function. An additional 312 (13%) are unique to *N. europaea*. Roughly 75% of the predicted proteins have the potential to be assigned a function. Other searches give similar results: 1,737 proteins match InterPro profiles, 1,967 match a Pfam hmm profile (default threshold), and 1960 can be assigned to a COG group (3BeTs). In addition to protein-encoding genes, we identified 41 tRNAs, representing all 20 amino acids. The only rRNA operon in this strain is of the 16S-Ala tRNA$_{TGC}$-Ile tRNA$_{GAT}$-23S-5S type and contains
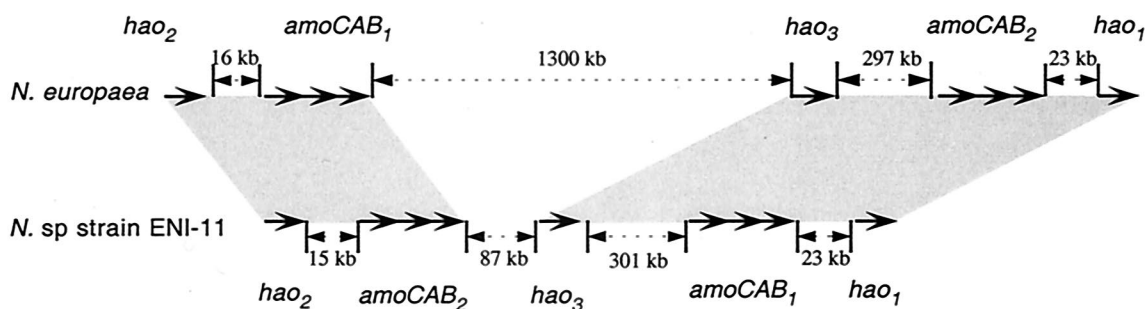
FIG. 2. Arrangement of *hao* and *amo* in *N. europaea* and *Nitrosomonas* sp. strain ENI-11. Arrows indicate the orientation of the genes. The nomenclature of the genes in *N. europaea* is with respect to their *Kpn*I fragment sizes for *hao* (38) and *Eco*RI fragment sizes for *amo* (37). The nomenclature for *Nitrosomonas* sp. strain ENI-11 follows that described by Hirota et al. (36). Distances between regions are indicated in kilobases and are not drawn to proportion. The gray areas indicate the regions similarly arranged.

the typical I-*Ceu*I endonuclease site in the gene for the 23S rRNA.

Gene homolog distribution. The sequence of this third β-proteobacterium makes a fine addition to the repertoire of structural diversity and content observed in the genomes sequenced thus far. A range of organisms (>130) were represented with top BLAST hits to one or more genes. *Ralstonia solanacearum*, one of two β-proteobacterial genome sequences currently in the public databases, was most often the top BLAST hit (31%, 768 of the 2,460 predicted ORFs). *Neisseria meningitidis*, the other sequenced β-proteobacterium, was the top hit with 5% of the genes. A surprising 13% had, as the top hit, *Pseudomonas aeruginosa*, a primarily soil-dwelling γ-proteobacterium. Other proteobacteria, including common soil inhabitants, were often found at the top of BLAST lists. Genes from *E. coli*, *Caulobacter crescentus*, *Sinorhizobium meliloti*, *Vibrio cholerae*, *Xylella fastidiosa*, and *Mesorhizobium loti* were frequent top BLAST hits with matches to 4.0, 2.9, 2.8, 2.6, 2.2, and 1.9% of *N. europaea* genes, respectively. At least 50 genes were also identified with each of the following classes of microorganisms: cyanobacteria and gram-positive actinobacteria and bacilli. The broad distribution of BLAST hits may reflect the few previously completed genome sequences from members of the β-proteobacteria or from any other ammonia-oxidizing bacteria.

Comparative genomics. Hirota et al. (36) conducted pulsed-field gel electrophoresis experiments to localize both copies of *amoCAB* and all three copies of *hao* to a single 487-kb fragment of DNA in *Nitrosomonas* sp. strain ENI-11. Each copy of *amo* was found to be within 15 or 23 kb of a copy of *hao*, based on restriction maps followed by long-range PCR for *Nitrosomonas* sp. strain ENI-11. The genome sequence of *N. europaea* ATCC 19718 reveals nearly identical proximity (15.5 and 23.1 kb) of these two important ammonia oxidation gene clusters to a copy of *hao*. The genes between each *amo* and *hao* cluster identified in *Nitrosomonas* sp. strain ENI-11 are also the same in *N. europaea*. These include the genes encoding threonyl tRNA synthetase (*thrS*), initiation factor 3 (*infC*), ribosomal protein L20 (*rplT*), phenylalanyl tRNA synthetase α and β subunits (*pheS* and *pheT*) in the 15.5-kb intergenic spacer region and the RNA polymerase β and β′ subunits (*rpoBC*) in the 23.1-kb span. The complete 23.1-kb spanning region includes tRNA genes, several ribosomal genes, elonga-

tion factors G (*fusA*) and Tu (*tufB*), and the transcription anti-termination gene *nusG*. Both genomes also have a third *hao* gene copy that is located approximately 300 kb upstream of the *amo*/*hao* cluster with the 23-kb spanning region which suggests two similar *amo* and *hao* gene arrangements in these two *Nitrosomonas* strains (Fig. 2). However, these clusters are separated by a strikingly different spanning region (87 kb in ENI-11 versus 1,300 kb in *N. europaea*). Thus, although the overall arrangement of two *amo*/*hao* gene clusters is conserved, dramatic rearrangements have nonetheless occurred since these two *Nitrosomonas* strains diverged.

Complex repetitive sequences. Like most ammonia-oxidizing bacteria examined, *N. europaea* has multiple copies of the genes coding for AMO, HAO, and cytochrome c_{554} (Table 2). The question of whether other genes, and their associated functions, might also be duplicated has long been speculated. In contrast to several other autotrophs, the genes coding for ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCo)

TABLE 2. Complex repetitive sequences

Sequence group	Size (bp)	Copy no.	% Identity
Gene, operon, and/or region			
amoCAB, ORF4 and ORF5	4,833	2	>99
hao, ORF, c_{554} and c_{m552}	4,303	2	>99
hao, ORF, c_{554}	3,601	1	>99
tufB	1,193	2	>99
aat,′ ORF, *ppsA*, ORF, ORF, *glnS-lysU′*	7,510	2	100
IS elements			
ISne1 (single transposase, several strong database hits)	~1,250	27	>95
ISne2 (single transposase, strong hits to N and C terminus)	~1,012	16	>97
ISne3 (two transposases, similar to IS*401*)[a]	~1,316	13	>98
ISne4 (two transposases, several weak hits to both)	~843	10	>98
ISne5 (two transposases, similar to ISne4)	~815	9	>98
ISne6 (single transposase, several strong database hits)	~1,000	5	>98
ISne7 (single transposase, few weak database hits)	1,050	3	>99
ISne8 (single transposase, several strong database hits)	2,026	2	>99

[a] *Burkholderia cepacia*.

are not duplicated. The ribosomal genes also are not duplicated as they are in many other bacteria. One of the two copies of another duplicated gene (*tufB*), encoding elongation factor Tu, has been found associated with one of the two *amo/hao* gene clusters (with the 23-kb spanning region [see Fig. 1 and 2]).

A perfect tandemly repeated region, a possible example of a recent duplication event, may give us insight into the mechanism of gene duplication and evolution and provides another example that gene duplication in *N. europaea* is not limited to genes involved with ammonia catabolism. This 7.5-kb duplication was found to carry genes for phosphoenolpyruvate synthase (*ppsA*; NE2359 and NE2366), glutaminyl tRNA synthetase (*glnS*; NE2356 and NE2363), along with three conserved hypothetical proteins, as well as portions of the genes for lysyl tRNA synthetase (*lysS*; NE2361) and aspartate aminotransferase (NE2362). These two likely "pseudogenes" have full-length representatives at the left and right borders of this most obvious duplication event. A second, different type of complex tandem repeat is composed of 15 copies of an ~339-bp degenerate repeat (nine forms) interspersed with four identical copies of a second repeat type (317 bp). This 6.3-kb repeat region lies within (and preserves throughout its length) an ORF (NE0161) encoding a large 3,064-amino-acid conserved hypothetical protein with a putative hemolysin-type calcium-binding region.

Besides the abundance of duplicated (presumably) native genomic material, the genome contains 85 predicted IS elements (most are complete elements) representing eight different families not previously described (see Table 2). Although these elements appear randomly distributed around the genome (Fig. 1), closer inspection reveals that there are seven IS-free islands between 100 and 200 kb in size, whereas the genome in its entirety averages one IS per 33 kb. The three least prominent IS families (ISne6, ISne7, and ISne8) do in fact appear randomly distributed, but the IS families with 9 to 27 members appear to have preferentially inserted in the vicinity of other IS elements (or alternatively served as attractants for subsequent IS integrations). The ISne1 and ISne3 classes displayed only a relatively random integration pattern, with half of the members being in close (<2 kb) proximity to other IS elements. Although ISne4 and ISne5 encode similar transposases and share some level of identity, they are not found near one another. Rather, four of the ten ISne4 are proximal to ISne1 integrations, and another two are close to ISne2 sites, whereas five of the nine ISne5 are very close to an ISne2 integration site. In fact, four of these are directly adjacent to ISne2 elements and, surprisingly, in three different orientations, indicating completely independent integration events. ISne2 itself is <2 kb away from other IS elements in 11 of its 16 members. The reason for this striking integration bias is unknown. Once a function is lost due to IS integration, if this mutation is not deleterious to the organism, further integrations in this gene would not be lethal and would be maintained in the population. This mechanism of acquiring integrations could be extended to subsequent IS integrations in the surrounding region if nearby genes participate in the same function as the gene with the first IS integration and are, therefore, rendered functionless by that integration event.

A different class of duplicated genes includes likely members of a recently diverged paralogous family that codes for or regulates production of Fe siderophore receptors. Several classes of siderophore receptors, along with regulatory genes, were identified in the genome. Together, all of the repeated elements within this genome constitute ~5% of the nucleic acid sequence, which ranks as one of the most densely populated bacterial genomes in terms of complex repetitive DNA.

Lithoautotrophy. The use of ammonia as an energy source, an example of lithotrophy, requires the ability to catabolize ammonia and to generate reductant for biosynthesis and to generate a chemiosmotic gradient to drive ATP synthesis (Fig. 3). The genes coding for AMO (*amoCAB*), HAO (*hao*), and cytochromes c_{554} and c_{m552} were previously sequenced (7, 9, 39, 47, 54, 64, 65) and confirmed with the genome sequence. No additional genes were identified that might be involved in the oxidation of ammonia to nitrite. Electrons from hydroxylamine oxidation flow through cytochrome c_{554} and cytochrome c_{m552} into the electron transport chain at the level of ubiquinone (Fig. 3) (79). Genes for a typical ubiquinone-cytochrome *c* oxidoreductase and a cytochrome aa_3-type cytochrome *c* oxidase are present; the soluble monoheme cytochrome c_{552} is thought to mediate between the two enzymes. The gene for cytochrome c_{552} is not located in an energetics-related gene cluster. Contributions to the proton gradient include scalar protons that are products of the HAO reaction and proton translocation by the bc_1 complex and the terminal oxidase. Genes for all subunits of a typical ATP synthase are present. Reduction of NAD^+ requires proton-driven reverse electron flow, presumably through NADH-ubiquinone oxidoreductase which is encoded in a discrete gene cluster. Reduction of $NADP^+$ may be carried out by the proton gradient-dependent $NAD^+/NADP^+$ transhydrogenase also encoded in a gene cluster. The energy cost of reduction of $NADP^+$ is thus very high. Two gene clusters contribute an Na^+-dependent NADH-ubiquinone reductase and an Na^+/H^+ antiporter, which may facilitate some Na^+-driven secondary transporters and also may be significant in marine environments.

The genome reveals that *N. europaea* has a relatively limited number of optional paths to terminal electron acceptors. Only one type of terminal oxidase of the aa_3 family is present. The only other potential terminal oxidase is the soluble, periplasmic copper enzyme noted below. The apparent presence of a membrane-anchored cytochrome c_4 suggests that electrons from ubiquinol may, at times, bypass the bc_1 complex on their way to cytochrome oxidase.

N. europaea is incapable of reducing nitrate but is capable of reducing nitrite with the formation of nitric and nitrous oxide but not dinitrogen (reviewed in reference 41). No full ORFs were identified with strong similarity to known dissimilatory nitrate reductases (EC 1.7.99.4) or nitrous oxide reductases (EC 1.7.99.6), a finding consistent with the biochemical evidence. A cluster of genes encodes periplasmic proteins seemingly related to the transfer of electrons to and reduction of nitrite and/or oxygen (Fig. 4B). The first gene encodes an aerobically expressed, soluble "blue-copper oxidase" (24). The next two genes encode a soluble, monoheme *c* cytochrome and a diheme *c* cytochrome, respectively. The fourth gene in this cluster has the best match to the *aniA* gene of *Neisseria gonorrhoeae* encoding an inducible nitrite reductase (PAN1) (verified by mutational inactivation of nitrite reduction) (56). The
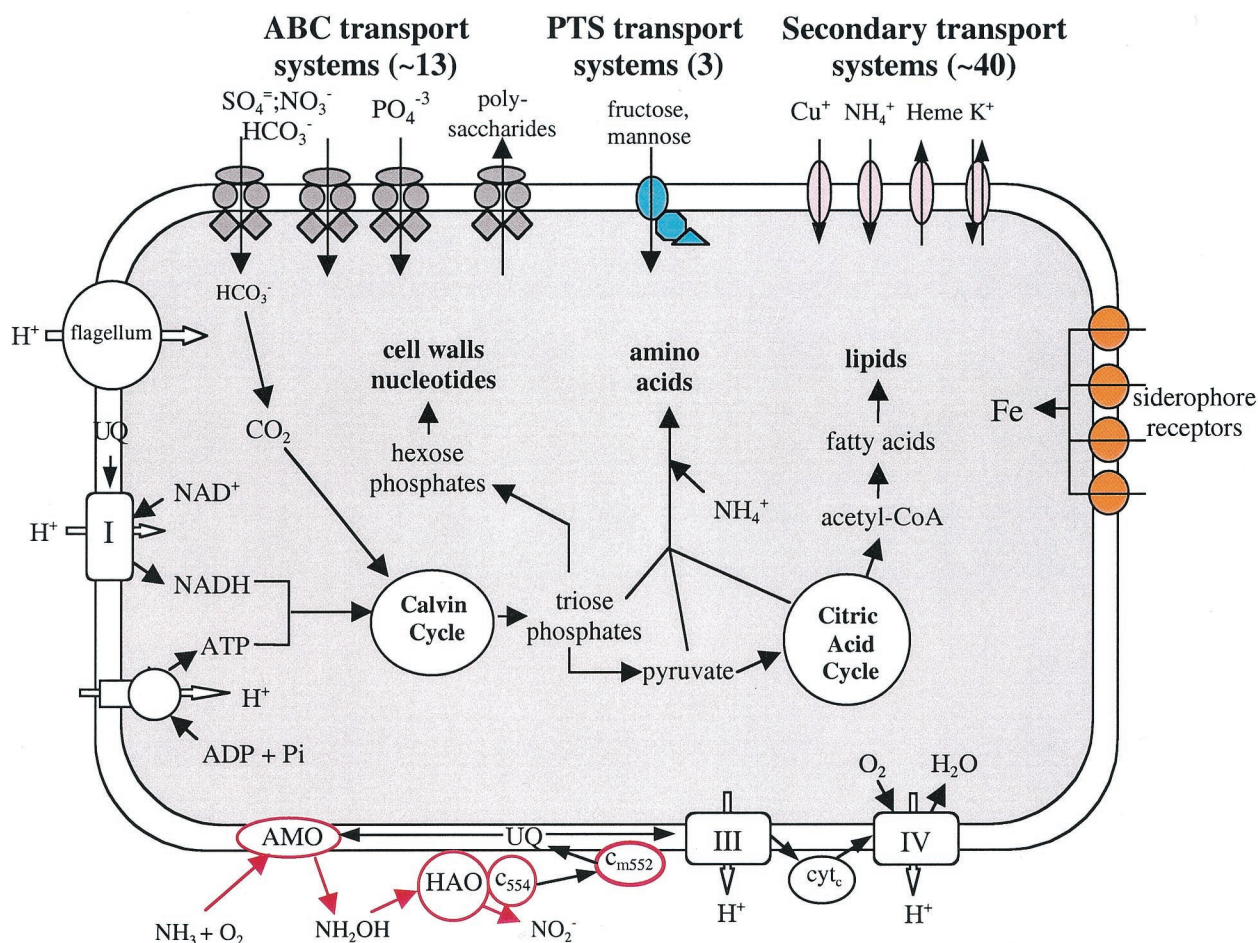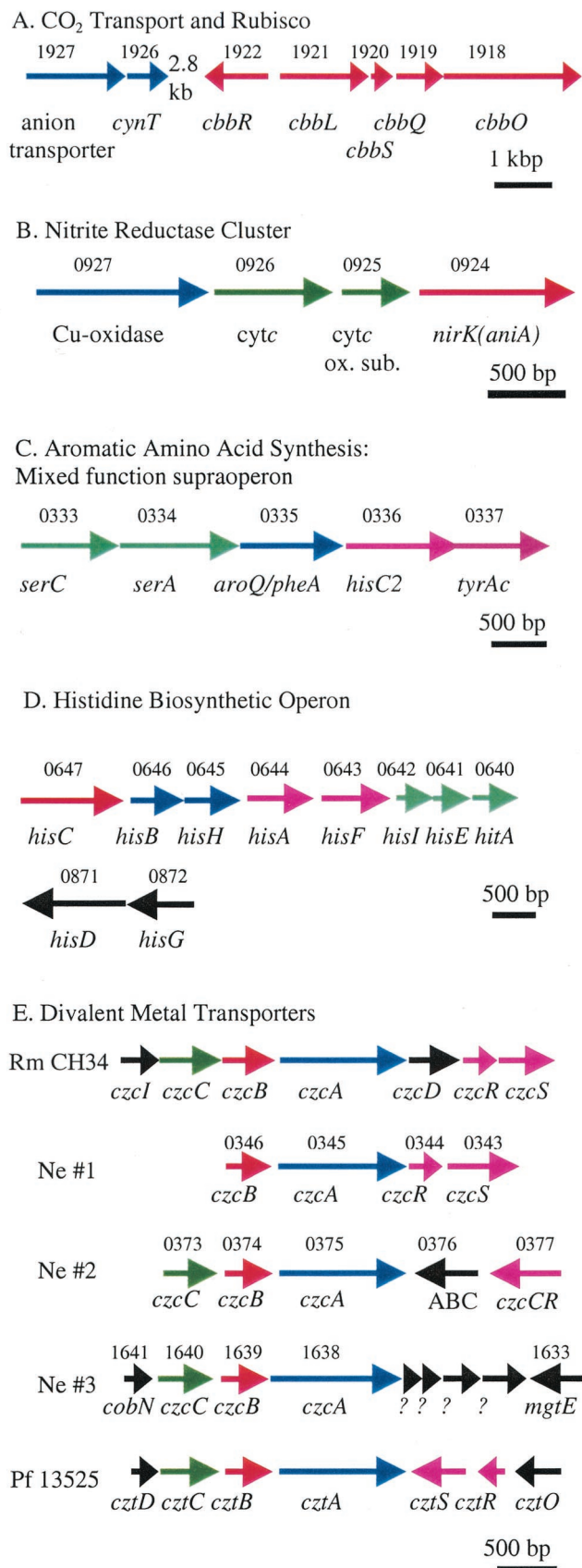
FIG. 3. Diagram of *N. europaea* cell. Processes primarily associated with the generation of a proton gradient are indicated on the bottom, and processes associated with utilization of the gradient are on the other sides. ATP and NADH production are indicated on the left. Transporters are indicated on the top and right. Not all transporters are indicated; the numbers in parentheses indicate the approximate numbers of each type. Major metabolic pathways and biosynthetic pathways are indicated in the center of the cell. The roman numbers refer to the enzyme complex I (NADH-ubiquinone reductase), complex III (ubiquinol-cytochrome *c* reductase), and complex IV (cytochrome *c* oxidase) in the respiratory chain.

putative nitrite reductases from other *Neisseria* spp. and *N. europaea* all have significant similarities to conserved domains from copper-containing nitrite reductases from several denitrifying bacteria but form a separate clade in the phylogeny of *nirK*. A signal sequence of 24 amino acids is predicted. *N. europaea* appears to have a distinct form of *nirK* versus other ammonia oxidizers from the β subdivision (17). Disruption of the putative *nirK* did not lead to a loss of nitrous oxide production but did result in an increased sensitivity to nitrite (5). The NO reducing system is encoded in a *nor* gene cluster (*norCBQD*; NE2003, NE2004, NE2005, and NE2006) with an organization similar to that found in *Pseudomonas* sp. strain G-179 (6). Anaerobic metabolism of *N. europaea* was reported with pyruvate as the reductant and nitrite as the terminal electron acceptor (1). The genome encodes the enzymes necessary for this process as mediated by pyruvate dehydrogenase and possibly the citric acid cycle. Under these conditions, electrons apparently pass to nitrite reductase by way of NADH-ubiquinone reductase. Anaerobically, *Nitrosomonas eutropha* and *N. europaea* are reported to oxidize H_2 and reduce nitrite, although the rate was much higher in *N. eutropha* (12). A gene

for hydrogenase was not identified in the genome of *N. europaea*.

Assimilation of carbon dioxide is initiated by a type I RubisCo (Fig. 4). The genes for this enzyme are most similar to those from *Acidithiobacillus ferroxidans*, another obligate lithoautotroph. A carbonic anhydrase gene (*cynT*) is next to a gene for an anion transporter and only 4.6 kb from the RubisCo genes (Fig. 4A). If this transporter is for carbonate or bicarbonate, then it and carbonic anhydrase would promote accumulation of CO_2, the substrate for RubisCo. Genes for three additional carbonic anhydrases were also identified. CO_2-repressible carbonic anhydrase activity has been observed in *Nitrosomonas* (43). *Ralstonia eutropha* H16, upon the inactivation of the gene encoding carbonic anhydrase, was unable to grow at ambient CO_2 concentration (49). *N. europaea* does not have genes associated with the production of carboxysomes (16).

With the exception of two enzymes, genes for all enzymes to complete the Calvin-Benson-Bassham cycle are present. A gene for sedoheptulose 1,7-bis-phosphatase (EC 3.1.3.37) is absent. However, the fructose 1,6-bis-phosphatase (EC 3.1.3.11)

A. CO₂ Transport and Rubisco

B. Nitrite Reductase Cluster

C. Aromatic Amino Acid Synthesis: Mixed function supraoperon

D. Histidine Biosynthetic Operon

E. Divalent Metal Transporters

encoded by NE0521 may have higher activity with sedoheptulose 1,7-bis-phosphate and may function primarily in its hydrolysis and not in gluconeogenesis (see below). This was found to be the case with the highly similar enzymes from _Ralstonia metallidurans_ (formerly _Alcaligenes eutrophus_) (84) and _Xanthobacter flavus_ (67). The gene encoding NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.13) is also absent and is apparently replaced by a gene for the NADH-dependent enzyme (EC 1.2.1.12) "borrowed" from and also used by gluconeogenesis and glycolysis, as has been observed in other chemoautotrophs (67). Since during growth of this obligate autotroph it is unlikely that there is an extended period of time when both CO_2 fixation and gluconeogenesis are not occurring, there is little advantage in having two separate genes that can be independently regulated. In fact, energy is conserved by having only one enzyme. Metabolic regulation at the enzyme level can still provide the appropriate flux of gluconeogenesis and glycolysis.

A very significant energy savings is achieved by the use of NADH rather than NADPH (generated from NADH in an energy-dependent reaction) to reduce 3-phosphoglycerate. It is interesting that an obligate autotroph, lacking catabolic pathways, had no reason to adopt a system of metabolic regulation dependent on the separate use of NAD^+ or $NADP^+$ as a redox mediator and effector in catabolic or anabolic pathways, respectively, as is seen in the _Eukaryota_. It continues to use both NADH and NADPH for biosynthesis.

Genes for the enzymes common to gluconeogenesis and glycolysis were present. From an energetic point of view it seems likely that the flux through either pathway is limited to biosynthetic requirements or recycling of fixed carbon. As noted above, the gluconeogenic enzyme fructose 1,6-bis-phosphatase (EC 3.1.3.11) likely has a substrate specificity suitable for a role in the hydrolysis of sedoheptulose 1,7-bis-phosphate. The hydrolysis of fructose 1,6-bis-phosphate may be carried out by a pyrophosphate-dependent 6-phosphofructokinase (EC 2.7.1.11), i.e., fructose 1,6-bis-phosphate + P_i → fructose-6-phosphate + PP_i. Notably, the adjacent gene encodes a pyrophosphatase whose action would "pull" gluconeogenesis. The same pyrophosphate-dependent 6-phosphofructokinase, which is reversible, may also catalyze the production of fructose 1,6-bis-phosphate in glycolysis (participating in a pyrophosphate-dependent energy economy). Genes with high sequence similarities to ATP-dependent 6-phosphofructokinases (EC 2.7.1.11) were not found.

Genes for all enzymes of the tricarboxylic acid (TCA) cycle were identified. The predominant theory for the basis of obligate autotrophy holds that the TCA cycle is incomplete in these organisms (68). Indeed, for many autotrophs, including _N. europaea_ (40), α-ketoglutarate dehydrogenase activity was not detected or was very low, in contrast to other activities of the TCA cycle. However, genes for subunits E1, E2, and E3

FIG. 4. _N. europaea_ gene clusters. Several gene clusters described in the text are diagrammed. Each arrow represents a gene in the cluster. The N. europaea gene numbers are above each arrow and the gene names (or other identifiers) are below each arrow. In panel E, gene clusters from _R. metallidurans_ CH34 (Rm CH34) and _P. fluorescens_ 13525 (Pf 13525) are included for comparison.

(which is shared by pyruvate dehydrogenase) are present. It remains to be seen whether the genes for α-ketoglutarate dehydrogenase are expressed.

N. europaea does not synthesize glycogen or β-hydroxybutyrate as storage products but does accumulate polyphosphate when growth is limited by low values of pH (72). The gene for polyphosphate kinase (NE0323) found in the genome has a very high sequence similarity to the equivalent gene in *N. meningitidis*.

Biochemical evidence suggests that ammonia is assimilated via glutamate dehydrogenase (40), and this pathway is consistent with the presence of a gene coding for an NADPH-specific glutamate dehydrogenase. Also consistent with the biochemical evidence (11), a *glnA* homolog, which encodes glutamine synthetase, is present. Although a putative *glnE*, which encodes the glutamine synthetase adenylating enzyme, was found, homologs for *glnD* (PII uridyl transferase), *glnB* (PII), or *glnK* (alternative PII) were not identified. The absence of a gene encoding a PII protein is surprising given the broad distribution of this regulatory protein. Glutamate synthase activity was not observed previously (11). Although a gene with similarity to glutamate synthase is present, the gene is truncated and appears to encode only the domains that transfer ammonia to α-ketoglutarate and accept and transfer electrons from a donor to the reductive amination domain. The domain responsible for the hydrolysis of ammonia from glutamine was not identified in this truncated gene or elsewhere in the genome. The gene profile supports the biochemical evidence that ammonia is assimilated via glutamate dehydrogenase, whereas the role of glutamine synthetase is to produce glutamine.

An ammonium transporter is present, which at low pH may supplement the passive uptake of ammonium. *Nitrosomonas* is reported to assimilate nitrite-N but not nitrate-N (62). The observed assimilation of nitrite-N is presumed to involve the siroheme-containing sulfite or nitrite reductase that is encoded in the genome. It is interesting that this frugal autotroph would expend reducing power on the production of ammonium.

Genes encoding the classical urease (e.g., *ureABC* coding for urea amidohydrolase) are not present. A candidate for urea metabolism in *N. europaea* is a variation of urea amidolyase. However, the gene in *N. europaea* is shorter than the one found in *S. cerevisiae* (a well studied system; [15, 31, 32]) and appears to contain only the carboxylase function of this bifunctional enzyme, and not the hydrolase/amidase function. No good homologs for the hydrolase/amidase were identified. To our knowledge, there are no reports indicating that urea can support the growth of *N. europaea*.

Biosynthesis. Autotrophy requires the ability to synthesize most, and in the case of *N. europaea*, all required small molecules and macromolecules from inorganic constituents. The gene profile of *N. europaea* is consistent with this requirement. In general, most and often all of the genes needed for particular biosynthetic pathways can be identified. Fatty acid and lipid synthesis, production of cofactors and prosthetic groups, nucleic acid synthesis, and amino acid synthesis can all be accounted for based on the gene profile. For example, similarity searches revealed many genes for the synthesis of purines and pyrimidines. For purine synthesis, genes were present for de novo synthesis of adenyl and guanisyl phosphates and their deoxy derivatives. Of the 23 steps required for the synthesis of

ATP, dATP, GTP, and dGTP from ribose-5-phosphate, genes for all of the steps were identified. In the case of pyrimidine synthesis, complete pathways for synthesis from carbamoyl phosphate to UTP, dTTP, CTP, and dCTP were identified.

The genes for enzymes needed to synthesize fatty acids up to hexadecanate from acetyl coenzyme A (acetyl-CoA) were identified. *N. europaea* has three 3-ketoacyl-ACP synthases that catalyze the initial condensation reaction between acetyl-CoA and malonyl-CoA. Both synthase I (encoded by *fabB*) and synthase II (encoded by *fabF*) can elongate saturated fatty acids; however, only synthase I can catalyze the synthesis of unsaturated fatty acids. Synthase III (encoded by *fabH*) is involved with branched-chain fatty acid biosynthesis. As in other bacteria, several genes involved in fatty acid synthesis in *N. europaea* are arranged in an operon containing *fabF*, *acpP* (acyl carrier protein), *fabG* (3-ketoacyl-ACP reductase), *fabD* (malonyl-CoA ACP transacylase), *fabH*, and *plsX* (undefined role in fatty acid biosynthesis). No match for enoyl-ACP hydratase was found in the genome. However, a 3-hydroxymyristoyl/2-hydroxydecanoyl ACP hydratase was found (encoded by *fabZ*) that has a broad substrate range, including both short-chain and saturated and unsaturated long-chain fatty acids (35). Genes for the synthesis of squalene from dimethylallyl-PP and isopentenyl-PP were identified. In addition, the gene for the branch point for hopane synthesis from squalene, squalene-hopene synthase, is present. However, the gene for the previous enzymatic step, squalene epoxidase, was not found.

Genes for amino acid synthesis are among the most conserved pathways and allow for comparisons of gene organization among organisms. Genes for the synthesis of aromatic amino acids and histidine illustrate this point (Fig. 4C and D). The synthesis of Phe, Tyr, and Trp from phosphoenolpyruvate and erythrose-4-phosphate is via the shikimic acid pathway. Genes for all enzymes required for chorismate synthesis and the tryptophan branch were identified. Tryptophan is synthesized from chorismate via anthralinate. All elements of the pathway were identified and are found in three clusters in the genome. Anthranilate synthase component I (EC 4.1.3.27; encoded by *trpE*) and anthranilate phosphoribosyltransferase (EC 2.4.2.18; encoded by *trpD*) likely encode a bifunctional enzyme. *N*-(5′-Phosphoribosyl) anthranilate isomerase (EC 5.3.1.24) (encoded by *trpF*) and indole-3-glycerolphosphate synthase EC 4.1.1.48 (encoded by *trpC*) also likely encode a bifunctional enzyme. Tryptophan synthase α (EC 4.2.1.20; encoded by *trpA*) and β (EC 4.2.1.20; encoded by *trpB*) chains are contiguous.

Phenylalanine could be produced via the nonarogenate branch from chorismate to phenylpyruvate by a dual function chorismate mutase-prephenate dehydratase (P protein) (NE0335). An amino transferase (EC 2.6.1.9; NE0336) would convert phenylpyruvate to phenylalanine. The P protein will apparently also produce free prephenate. An aspartate amino transferase (EC 2.6.1.1) could presumably transform prephenate to arogenate. No aromatic specific aminotransferase (EC 2.6.1.57) was identified, but HisC2 has high similarity to *Pseudomonas stutzeri* aminotransferase with a broad specificity in the biosynthesis of histidine, phenylalanine, or tyrosine. In *N. europaea*, the prephenate dehydratase of the P protein is specific and does not metabolize arogenate (71). Therefore, arogenate is a precursor of tyrosine but not phenylalanine.

This observation was supported by the pathway identified in the genome sequence.

Tyrosine synthesis occurs via prephenate which is produced by the P protein. An aminotransferase then forms arogenate (see above). Arogenate dehydrogenase function may be included in another dual-function enzyme combining arogenate dehydrogenase (EC 1.3.1.43) and prephenate dehydrogenase activity (EC 1.3.1.12) (TyrA$_c$). However, Subramaniam et al. (71) observed that this enzyme was arogenate and NADP$^+$ specific in *N. europaea*. Therefore, arogenate would be an obligatory intermediate in tyrosine synthesis and 4-hydroxyphenylpyruvate would not be formed. Subramaniam et al. (71) have shown that tyrosine is required for the prephenate dehydratase activity and that phenylalanine is an inhibitor, thus assuring a balanced production of phenylalanine and tyrosine. For tyrosine to regulate phenylalanine synthesis, two separate paths must be present (after prephenate). Since the aspartate aminotransferase is probably promiscuous (for phenylpyruvate and prephenate) this requires P protein to be specific for prephenate (versus arogenate). Subramaniam et al. (71) observed that the use of arogenate as a precursor for tyrosine and phenylpyruvate for phenylanine is characteristic of the cyanobacteria and coryneform bacteria.

The histidine biosynthetic operon in *N. europaea* exhibits overall organization (Fig. 4D) similar to that seen in *N. meningitidis* and *E. coli*, with some exceptions that help to place gene fusion events along the evolutionary lines of the proteobacteria. In *N. europaea*, the majority of the *his* genes are contiguous; however, *hisDG* genes are separated from the rest of the operon. The *hisI* and *hisE* genes are not fused in *N. europaea*, but they are adjacent genes whose ORFs overlap. In the enteric bacteria (γ-proteobacteria) examined to date, the *hisIE* gene product is a bifunctional enzyme. Examples of monofunctional enzymes encoded by *hisI* and *hisE* are commonly found in the β subdivisions. This observation would indicate that the *hisIE* gene fusion occurred after the enteric proteobacteria split from these subdivisions. The *hisB* gene of *N. europaea* is predicted to encode imidazole glycerol phosphate dehydratase (EC 4.2.1.19), while the histidinol phosphatase (EC 3.1.3.15) is likely encoded by a separate gene (NE1185) outside of the operon, similarly to that found in *R. solanacearum* (63). This observation confirms that the fusion of these two activities occurred after the evolutionary split separating the γ subdivision from the other subdivisions of the proteobacteria (29). HitA matches with a nucleotide-binding protein, similar to members of the HIT (histidine triad) family. A second putative *hisC* gene (NE0647) encoding an aminotransferase was identified outside of the *his* operon structure. However, this gene may be involved in the biosynthesis of other aromatic amino acids (see above).

The pathways for glycine, serine, and threonine synthesis are identified and are the presumed to be starting materials for the synthesis of the osmoprotectants betaine and glycine betaine. The *betA* gene encoding EC 1.1.99.1 in the osmoregulatory choline-glycine betaine pathway was identified (NE1237), although other genes found in the *bet* operon of proteobacteria were not found in close proximity. A possible ABC transport system for osmoprotectants may be encoded for by a polyamine transport operon similar to *potABCD* in *P. aeruginosa* (NE1870 to NE1873).

Genes for aminoacyl tRNA synthetases for all amino acids are present. Glutamyl-tRNA synthetase is encoded by *gltX* (NE1624). The duplicate *glnS* genes (NE2356 and NE2363) likely encode a specific glutaminyl-tRNA synthetase rather than the indiscriminate glutaminyl/glutamyl-tRNA synthetase. The absence of a complete *gatCAB* operon encoding the Glu-tRNA amidotransferase further supports the glutamine specificity for *glnS*. The presence of both *glnS* and *gltX* may be the result of an earlier gene duplication event and subsequent functional divergence (14).

Catabolism of organic compounds. In contrast to the genes for the biosynthesis of cellular constituents, genes for the catabolism of organic compounds are scant. For example, no genes for the degradation of either purine or pyrimidine nucleosides were identified. Salvage pathways for nucleotides were present and included genes for DNA exo- and endonucleases and ribonucleases. Nucleoside salvage appeared to be limited to uracil and thymidine and was nonexistent for the nucleobases. Likewise, complete pathways for the catabolism of most amino acids were not identified. Where a few genes were present, these genes were most often also required for processes other than catabolism. For a few of the simpler amino acids (e.g., aspartate, serine, and glycine), for which transamination would result in an intermediate in a primary pathway such as the TCA cycle, pathways for catabolism could be envisioned based on the gene profile. Likewise, as discussed above, genes for complete glycolytic pathways and the TCA cycle are present, suggesting that complete oxidation of simple sugars and organic acids should be possible.

A notable exception to the dearth of genes for catabolic enzymes is seen with fatty acid oxidation since genes for all of the enzymes required for fatty acid oxidation are present. As in *E. coli* and other bacteria, many of the activities of fatty acid oxidation are contained in two subunits of a multienzyme complex. The 3-hydroxyacyl-CoA dehydrogenase, enoyl-CoA hydratase, *cis*-Δ 3-*trans*-Δ2-enoyl-CoA isomerase, and the 3-ketoacyl-CoA epimerase are all contained in one subunit (encoded by *fadA*). The 3-keto-CoA thiolase is associated with a second subunit (encoded by *fadB*). As in other bacteria, the *fadA* and *fadB* genes are adjacent. Although *N. europaea* does have phospholipase D, which cleaves the head group from phospholipids, the genes for phospholipase A$_1$ and A$_2$ (enzymes that remove the fatty acids) and phospholipase C were not present, suggesting that *N. europaea* is not able to degrade phospholipids.

Transport systems. *N. europaea* has an array of active (primary and secondary) transporters (i.e., ion-coupled, ATP hydrolysis, or redox-driven transporters). Approximately 285 ORFs in its genome (11.5%) are dedicated to the active transport of molecules across its membranes. Other gram-negative bacteria have 3 to 12% of the ORFs in their genome encoding proteins involved in transport (78). The number of predicted ABC transporters in *N. europaea*, 13, is similar to the numbers in other lithoautotrophic bacteria but smaller than the numbers in facultative bacteria. For example, in *Methanobacterium thermoautotrophicum* ΔH (a lithoautotrophic thermophilic archaebacterium), 10 clusters are predicted to code for ABC transporters (69). In *Aquifex aeolicus* (an obligate chemolithotrophic eubacterium), 13 clusters similar to ABC transporter systems are present (22). In contrast, *E. coli*

(a chemoorganoheterotrophic bacterium) has 51 clusters containing sequences similar to ABC transporters (out of 746 transport and binding gene products from which 382 are characterized; 15.8% of the total ORFs [66]). *Thermotoga maritima* (representing a new genus of unique extremely thermophilic eubacteria growing up to 90°C) has 31 ABC transporters (57). *Halobacterium* sp. strain NRC-1 has at least 27 members of the ABC transporter superfamily.

In *N. europaea* ca. 14% (40 ORFs) of the active transport proteins are dedicated to Fe transport (see below). The 13 putative active transporters that include ATP-binding cassettes (ABC) represent approximately 75 ORFs or 3% of the total ORFs. These ABC transporters contain at least two of the typical three components for ABC transporters (nucleotide-binding domain, membrane-spanning domain, and solute-binding protein). ABC-type multidrug, protein, and/or lipid transporters are predominant. For instance, capsular polysaccharide export and polysaccharide or polyol phosphate-type transporters are present, as is an ABC-type sugar transferase system involved in lipopolysaccharide synthesis with a putative membrane cation efflux permease. This lipopolysaccharide transporter may confer to *N. europaea* the ability to adhere to surfaces. Two ABC-type transporters for anions (e.g., nitrate) could also be inferred. The nitrate transport systems have similarity to the sulfonate and bicarbonate transporters (52). A sulfate permease and an ABC-type sulfate/molybdenum transport system are present. *N. europaea* has two ORFs coding for a potassium uptake system and another for a sodium-translocating NADH dehydrogenase. Two loci for heme-exporting proteins adjacent to cytochrome related genes are also present in the genome.

Systems for the uptake of organic molecules are few in *N. europaea*. ABC transporters for some single molecules are present (i.e., for glutamine and spermidine or putrescine). An amino acid permease and an amino acid transporter are present but are not linked with a nucleotide-binding domain or a membrane-spanning domain, which are characteristic of ABC transporters. Only one sugar transporter is apparent as a phosphotransferase system (PTS) with similarity to fructose or mannose transporters. Two additional PTSs are nitrogen related. The limited number of permease genes for organic molecules may contribute to the obligate nature of lithoautotrophy by *N. europaea*.

The genome also contains loci encoding a Hg scavenger-like transport system, which could be responsible for heavy metal tolerance. Furthermore, several candidate genes encoding divalent cation (Cd, Zn, and Co) transporters have been identified. Because of the high redundancy of encoding genes, the cases of iron and zinc transport are discussed below.

The subject of iron uptake is among the most interesting revelations of the genome sequence. Although genes for siderophore biosynthesis seem to be absent, the genome of *N. europaea* contains 20 likely functional *fecIR* gene tandems, most of which are directly preceded or succeeded by iron siderophore receptor-encoding genes (Fig. 5). Among these are ORFs with significant similarity to known receptors for the uptake of ferrichrome-, pseudobactin-, and pyoverdin-like siderophores and other ferric iron siderophores, whose regulation seems to be linked to the FecI/FecR sigma factor/membrane sensor system of ferric-dicitrate iron accumulation (26). Over-

all, the genome of *N. europaea* appears to contain 22 genes encoding FecI-sigma factor-like proteins (Fig. 5). Our phylogenetic analysis of putative FecI-FecR homologs revealed that an ancestral *fecI* and *fecR* gene tandem has coevolved into two distinct subgroups of *fecI* and *fecR* gene tandems by an early duplication event. While the *fecIR* gene tandems have further coevolved by numerous gene duplication events within their subgroups, these subgroups did not evolve with any preference for the proximity of a particular iron siderophore-sensing receptor gene (see indices in Fig. 5). Whereas all *fecR* genes were found next to a *fecI* gene, only two *fecI* genes were found not in tandem with a *fecR* gene. Because unnecessary redundancy is usually easily lost from prokaryotic genomes, we predict that the identified *fecIR* gene tandems have physiological relevance for *Nitrosomonas*' high need for effective iron acquisition. Additionally, a gene encoding an enterobactin-like siderophore receptor (NE1205) was found in close proximity to a substrate-binding protein (NE1206) and five ORFs that likely comprise an ABC type 2 transport uptake system (NE1207 to NE1211). The overall scenario suggests that *N. europaea* may utilize siderophores produced by other organisms in its environmental consortium while under iron stress, and the possible citrate mechanism may serve as a "last resort" in the event iron is not available from any other source. The advantage of this mechanism would be the ability of the organism to supply its iron requirement without the costly secretion of reduced carbon. The *fecIR* gene tandem NE1217-NE1218 (followed by a TonB-dependent outer membrane receptor) is adjacent to a *copF*-like gene (NE1216) that may encode a Cu^{2+} cation transport ATPase. Copper ion transport is likely facilitated by the product of NE1019, which is highly similar to the CopA copper transport ATPase from *Staphylococcus aureus*. To prevent copper toxicity while facilitating Cu utilization, *N. europaea* likely expresses the three copper resistance proteins A, B, and D (NE0279, NE0280, and NE2058); a copper-binding protein, CopC (NE1491); and an inner membrane copper tolerance protein (NE2389).

In addition to iron and copper sensors and transporters, the *N. europaea* genome contains three separate gene clusters putatively involved in divalent cation (cobalt, zinc and cadmium) transport (Fig. 4E). Cluster 1 (NE0346/5/4/3) is organized similarly to the *czc* gene cluster in *R. metallidurans* CH34, and genes in the second cluster (NE0373/4/5/6/7) are arranged like the *czt* gene cluster in *P. fluorescens* 13525, whereas the third cluster seems to lack the two-component regulatory system found in the other two clusters (Fig. 4E). The third *czc* gene cluster is in proximity to a gene coding for an Mg^{2+} transporter protein, MgtE (NE1633). Based on the similarity to genes encoding heavy metals efflux systems in *Ralstonia* and *Pseudomonas* spp., the *czc* and *czt* clusters may also encode heavy metal efflux systems in *N. europaea*. Because of its proximity to a putative Mg^{2+} uptake protein, the third *czc* cluster may be involved in metal uptake.

Oxidative stress. *N. europaea*, like virtually all other aerobic organisms, is expected to contain enzymes that convert superoxide and hydroperoxides into innocuous products (21, 42, 48). Abundant aerobic bacteria such as *Bacillus subtilis* and *P. aeruginosa* have elaborate and redundant complements of superoxide dismutases (SODs), hydroperoxidases (HP), and "iron management" enzymes, and their expression is regulated
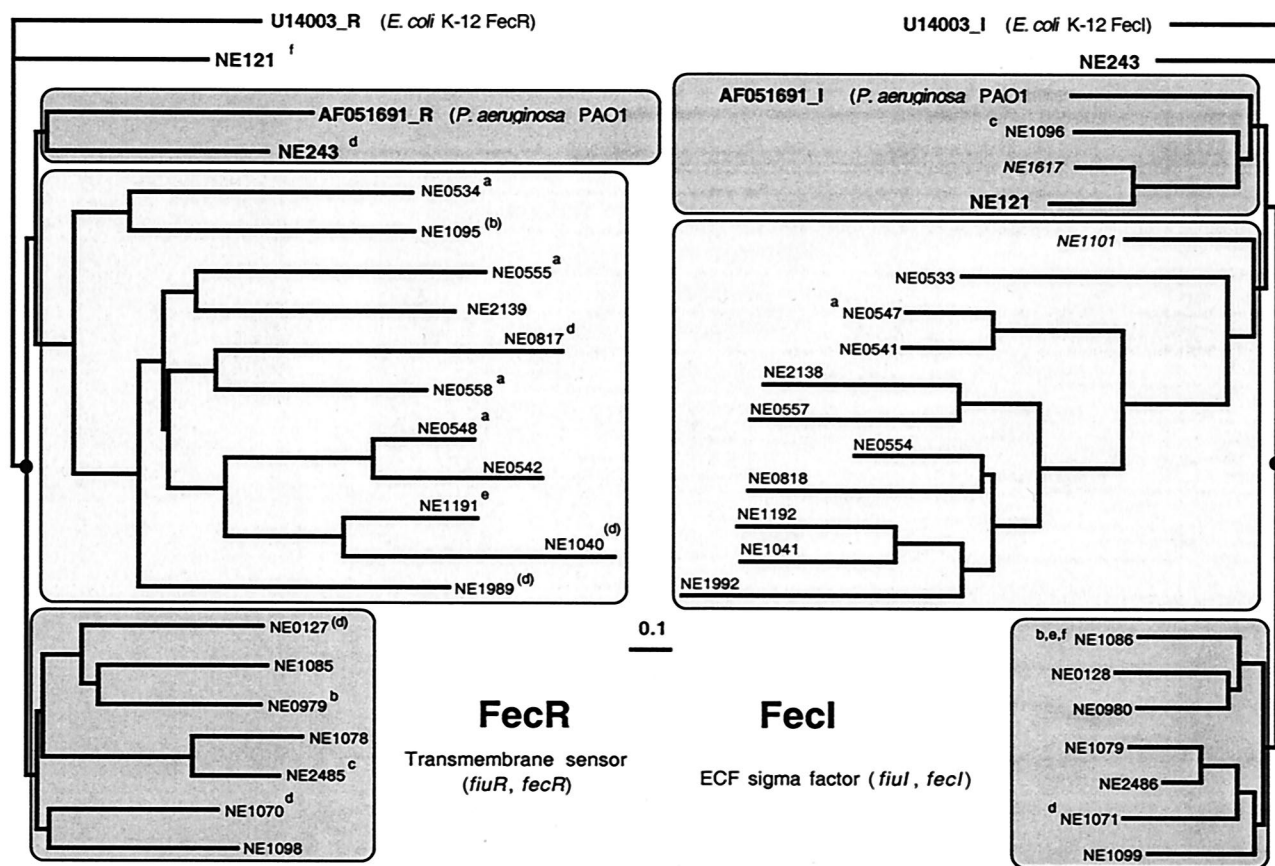
FIG. 5. Phylogeny of FecI and FecR homologous proteins and associated iron siderophore receptors. Distance-neighbor-joining trees were derived from CLUSTALW alignments of FecI- and FecR-homologous protein sequences that were deduced from identified ORFs in the genome of *N. europaea*. Sequences of identified functional FecI (FiuI) and FecR (FiuR) proteins from *E. coli* K-12 (U14003) and *P. aeruginosa* PAO1 (AF051691) were obtained from GenBank and used as a reference for the alignments. Designations of *N. europaea* proteins with high similarity to the FecIR proteins from *E. coli* and *P. aeruginosa* are given in boldface. Designations of FecI proteins without a respective FecR are given in italics. Putative iron siderophore sensing receptors that preceded or succeed a *fecIR* gene tandem in the genome are indicated with superscript letters as follows: a, ferrichrome-iron receptor 3 PCC6803 (D90899); b, ferrichrome-iron receptor *B. bronchioseptica* (U77671); c, hydroxamate ferrisiderophore receptor PAO1 (AF051691); d, unidentified ferric siderophore receptor PAO1 (U33150); e, ferripyoverdine receptor PAO1 (L10210); and f, TonB-dependent outer membrane receptor PAO1 (AE004674). Letters in parentheses indicate that the ORF is probably truncated. An additional gene encoding a ferrienterobactin receptor similar to AF083948 was found not associated with FecIR (data not indicated).

through complex regulatory networks (70). The genome of *N. europaea* contains genes that encode a monofunctional small-subunit catalase (HPII, *katA*), a catalase-peroxidase (HPI, *katG*), a thioredoxin-dependent peroxide reductase (alkyl HP, *ahpC*), and an iron-containing SOD (Fe-SOD, *sodB*). The HPI-encoding gene is preceded by a truncated copy of its N terminus, as found in the genome of *Burkholderia fungorum* (*cepacia*) LB400, also a member of the β-proteobacteria. Most of the HPs are heme-containing enzymes and genes coding for bacterioferritin (Bfr, *bfr*) and bacterioferritin comigratory protein (Bcp) were identified. In contrast, genes encoding a thioredoxin reductase (NADH-peroxiredoxin reductase, AhpF), which is part of the thioredoxin redox couple in many bacteria, glutathione oxidoreductase (*gorA*), and other isozymes of HP or SOD were not found in the genome.

The cytochrome *c* peroxidase of *N. europaea* is a diheme cytochrome. Homologous enzymes require the reduction of one heme before reaction with hydrogen peroxide. In contrast, the diferric form of the *N. europaea* enzyme reacts with hydro-

gen peroxide, making it a relatively better scavenger at a higher cellular redox potential (2). The cytochrome P460 in *N. europaea* has a possible NO-scavenging role (10). The gene encoding cytochrome P460 in *N. europaea* was sequenced previously (8) and is corroborated in the genome sequence.

To detect and defend themselves against oxidative damage, *E. coli* cells sense their cytoplasmic redox state with the OxyR protein, whose expression is upregulated autogenously by hydrogen peroxide concentrations of 50 to 200 nM and which is a key regulator protein in many multigene stress defense networks in most bacteria (70). Surprisingly, the genome of *N. europaea* does not contain a gene similar to known *oxyR* genes. Thus, the *katG* and *ahpC* genes, which are "normally" regulated by OxyR, are likely independently regulated in *N. europaea*, as is the gene encoding the ferric uptake regulator protein (Fur, NE0616, *fur*). The Fur protein, for instance, regulates ferric citrate (FecIR) and ferrichrome (*fhu* operon) transport, exotoxin synthesis, and the expression of HPs (HPI and HPII) in several bacteria (75, 77). Genes encoding other

regulators in the Fur protein family such as Zur, the zinc uptake regulator protein, were found in the genome of *N. europaea* (Zur, NE1722, *zur*). OxyR also controls the expression of the alternate sigma factor for stationary-phase-specific transcription, RpoS, in enterobacteria, whereas *rpoS* gene expression in *P. aeruginosa* is dependent on cell density (80). Surprisingly, the genome of *N. europaea* lacks an *rpoS*-like gene entirely. On the other hand, genes encoding other alternative sigma factors such as RpoN (σ^{54}-NE0062), RpoH (σ^{32}-NE0584), and RpoE (σ^{24}-NE2331) were found in the genome.

Motility, cell division, and signaling. *N. europaea* is motile and can form biofilms; hence, its genome should contain structural and regulatory genes necessary for the synthesis (and its regulation) of flagella, as well as for the correlation of flagellar synthesis and function with environmental cues and challenges. The complement of operons needed for flagellum biosynthesis is complete compared to available information from other bacteria (13, 25). However, the organization of these genes and the operon locations in the genome are remarkably different from those in other bacteria (25, 51). As far as is known, the flagellar master operon *flhDC* (NE2407/2406) is required for the transcriptional initiation of flagellation and chemotaxis both through direct activation and/or derepression of operons and indirectly through control of the FliA protein (NE2491), an alternative sigma factor (σ^{28}). Five classes of methyl-accepting chemotaxis proteins (MCPs) are known in enterobacteria (13, 51), and the *N. europaea* genome contains genes similar to members of three classes: (i) a Tsr-like protein (with HAMP domain; NE1864) that directly senses Ser, Ala, Gly, and aminoisobutyrate (*tsr*); (ii) a Tar-like protein (with PAC, PAS, MA, and HAMP domains; NE1863) that senses Asp and Glu directly and maltose through a periplasmic binding protein and is responsive to Co and Ni (*tar*); and (iii) a Tap-like protein (with MA domain; NE1251) that senses dipeptides through a periplasmic binding protein (*tap*). Additionally, a gene encoding a pseudomonad PilJ-like MCP (with the MA domain; NE1251) was identified. It is not yet known if these *N. europaea* genes are responding to these same attractants. Genes similar to ones known to encode the ribose/glucose/galactose sensor Trg and the redox sensor Aer in enterobacteria were not found. In comparison to other genomes, best conserved is the cluster with the genes encoding the proteins CheA, CheW, and MCP-Tsr (NE1866, NE1865, and NE1864), followed by genes encoding MCP-Tar, CheR, and CheB (NE1863, NE1861, and NE1859), a sequence of chemotaxis genes that is found in the *E. coli* genome. The other two MCP-encoding genes are adjacent to genes encoding additional CheW proteins (NE1250/51 and NE1396/97). In comparison to the *P. aeruginosa* genome that encodes more than 20 MCP-like proteins, chemotactic responsiveness of *N. europaea* mediated through only 4 MCP-like proteins are rather limited. Remarkable also are the fairly distant locations of the operons that contain *fliA* (NE2491), *cheYZ* (NE1923 and NE1924), and *cheA* (NE1866), which are usually in close proximity to or members of the same operon (25). The *N. europaea* genome also lacks loci encoding the FlgM/FliT proteins that regulate FliA availability in enterobacteria through anti-sigma activity (51). Given the slower growth rates, such a complex regulation of flagellation would likely drain more energy from the tight budget than it could save. Because *N. europaea* seems able to respond to the autoinducer

N-acylhomoserine lactone (AHL) and of forming biofilms (4), some sort of integration of these signal cascades can be expected. The flagellar master regulon *flhDC* is also involved in the regulation of virulence factor synthesis and cell division and seems to be a good candidate for functional compensation of the missing stationary-phase regulation via RpoS. Like its β-proteobacterial relative *Neisseria*, *N. europaea* contains only the essential suite of cell division proteins (*ftsI*, *murEF*, *mraY*, *murD*, *ftsW*, *murGCB*, *ddlB*, NE0985 to NE0994; *ftsQAZ*, NE0995 to NE0997; *ftsK*, NE1051; and *minCDE*, NE1831 to NE1829), and it lacks genes for SulA, ZipA, FtsL, and FtsN proteins that are found in γ-proteobacteria (53). Surprisingly, it lacks the functional two-component regulatory systems of LasRI/RhlR, which play key roles in connecting quorum sensing, motility, stationary-phase response, and the synthesis of virulence and stress tolerance factors in many environmental bacteria such as pseudomonads (81). Nevertheless, since the synthesis of autoinducer branches off fatty acid biosynthesis pathways and genes encoding the FAB pathway have been identified in *N. europaea*, we conclude that NE1184 encodes a putative AHL synthase similar to that found in *P. fluorescens* (50). In *P. fluorescens*, this autoinducer synthase produces both short- and long-chain AHLs (three identified). Although *N*-(3-oxohexanoyl)-L-homoserine lactone acted as a signal molecule in cell density regulated recovery from starvation in *N. europaea* (4), the actual sensor molecule(s) produced have not been biochemically characterized.

DISCUSSION

The gene profile reveals the basic array of catalysts necessary for all of biosynthesis starting with ammonia and carbon dioxide. Genes necessary for catabolism of ammonia, energy and reductant generation, biosynthesis, and CO_2 and NH_3 assimilation were identified. In contrast, genes for catabolism of organic compounds are limited. Whereas genes are present for the enzymes to degrade macromolecules to monomers and building blocks (e.g., by proteases and nucleases), pathways for further degradation are not represented in the genome. A likely consequence of this lack of degradation pathways is that this energy-challenged bacterium is expected to secrete rather than to recycle many organic wastes. *N. europaea* also appears to have a limited ability to transport organic molecules from the environment into the cell.

One of the interesting features of this bacterium is that it is obligate with regard to its necessity to use ammonia as a substrate. No other sources of energy, at least as sole sources of energy, have been identified. The genome provides some new insights to this nutritionally limited status. No evidence for other lithotrophic capabilities was apparent in the genome. For example, the presence of genes for a hydrogenase might have indicated that the bacterium could also grow with H_2 as the source of energy and reductant. However, for this ammonia-oxidizing bacterium, genes for the utilization of H_2, CO, Fe, or other inorganic sources of energy were not identified. *N. eutropha* was shown to grow anaerobically with H_2 as the reductant and nitrite as the electron acceptor. Such a capability in *N. europaea* would seem unlikely, given the absence of recognizable hydrogenase genes. As described above, genes for the catabolism of organic compounds are scant. Nonetheless, com-

plete pathways for a few compounds can be written based on the gene profile. At least in the case of fructose and/or mannose, a complete PTS transporter also seems to be present. To date, growth has not been observed on any organic compounds. The presence of the genes for α-ketoglutarate dehydrogenase, the step in the TCA cycle that appears to be absent in many obligate autotrophs, further clouds our understanding of obligate autotrophy. Although the genome provides some insights to this nutritionally limited status, it also raises new questions about why at least some organic compounds are not utilized.

One of the more interesting and unexpected findings to come out of the genome is insight into the strategy of *N. europaea* to accumulate Fe from the environment. These bacteria have a large appetite for Fe given all of the cytochromes they produce. Therefore, it was not surprising to learn that receptors for Fe-siderophores are encoded. The diversity of different classes of receptors and the number of genes devoted to these receptors (>20) was something of a surprise. However, even more surprising was the almost complete absence of biosynthetic pathways for siderophores. An exception is the citrate transporter, which utilizes a product and intermediate of the TCA cycle. Apparently, *N. europaea* relies on other bacteria to produce siderophores and then uses an arsenal of receptors to harvest their products. This reliance on other bacteria for siderophore production is in contrast to the otherwise complete self-reliance of this bacterium, which can extract energy and produce complex cellular constituents from simple inorganic nutrients. It is likely that *N. europaea* seeks to manage the Fe nutrition of its environment as a survival strategy (much like some pseudomonads on plant surfaces or pathogens in hosts).

Only two other complete genomes of β-proteobacteria are currently published, namely, *N. meningitidis* (74) and *R. solanacearum* (63). Although 31% of the genes in *N. europaea* were most similar to genes from *R. solanacearum* based on current BLAST searches, 13% of the genes were most similar to *P. aeruginosa* genes. As more β-proteobacterial genomes are completed, it will be interesting to see how similar their genomes are to *N. europaea* in terms of both gene similarity and genome organization. Of particular interest will be the genome of *A. ferroxidans*, another β-proteobacterium that is also an obligate lithoautotroph, albeit with a dependence on reduced sulfur or ferrous iron rather than ammonia. It also remains to be seen how representative *N. europaea* is of other ammonia-oxidizing bacteria. As noted above, major rearrangements of the locations of the *amo/hao* clusters are already evident between *N. europaea* and *Nitrosomonas* strain ENI-11. In comparing *N. europaea* to other ammonia oxidizers, differences in the number of copies of *amo* and *hao* and in their physiological capabilities (e.g., growth of *N. eutropha* on H_2 and the presence of urease in *Nitrosospira*) suggests that more differences should be expected.

Another interesting feature of the genome is the preponderance of insertion sequences. Both the variety and the number of elements suggest that this genome is prone to loss or accession of genetic elements. For *N. europaea*, it seems most likely that the genome is decreasing in size and the insertion sequences may be assisting in this evolutionary march toward a more compact genome. Two lines of evidence support this

notion. First, the genomes of many other proteobacteria are much larger. If *N. europaea*, like all of the proteobacteria, had its origin as a phototroph, then it has lost this ability. While other proteobacteria that have lost phototrophy have nonetheless gained additional capabilities (and larger genomes), *N. europaea* has arrived at a smaller genome. Similar dramatic gene losses have occurred in other divisions of the gram-negative bacteria such as the sphingobacteria and planctobacteria (18). The second line of evidence comes from thinking about the minimum genome set that is required for lithoautotrophy or photoautotrophy. Certain cyanobacteria (e.g., *Prochlorococcus* spp.) and archaebacteria (e.g., *M. thermoautotrophicum*) reveal that a much smaller genome is sufficient to sustain the autotrophic lifestyle. Perhaps *N. europaea* is continuing its downsizing toward a more compact genome like other catabolically limited bacteria.

As with other whole-genome sequencing projects, this genome has provided remarkable new insights into the growth and metabolism of this organism. Not unexpectedly, the sequence has also raised a number of questions that now await experimental investigation.

REFERENCES

1. **Abeliovich, A., and A. Vonsahk.** 1992. Anaerobic metabolism of *Nitrosomonas europaea*. Arch. Microbiol. **158:**267–270.
2. **Arciero, D., and A. B. Hooper.** 1994. A di-heme cytochrome *c* peroxidase from *Nitrosomonas europaea* catalytically active in both the oxidized and half-reduced states. J. Biol. Chem. **269:**11878–11886.
3. **Badger, J. H., and G. J. Olsen.** 1999. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. **16:**512–524.
4. **Batchelor, S. E., M. Cooper, S. R. Chhabra, L. A. Glover, G. S. Stewart, P. Williams, and J. I. Prosser.** 1997. Cell density-regulated recovery of starved biofilm populations of ammonia-oxidizing bacteria. Appl. Environ. Microbiol. **63:**2281–2286.
5. **Beaumont, H. J., N. G. Hommes, L. A. Sayavedra-Soto, D. J. Arp, D. M. Arciero, A. B. Hooper, H. V. Westerhoff, and R. J. van Spanning.** 2002. Nitrite reductase of *Nitrosomonas europaea* is not essential for production of gaseous nitrogen oxides and confers tolerance to nitrite. J. Bacteriol. **184:**2557–2560.
6. **Bedzyk, L., T. Wang, and R. W. Ye.** 1999. The periplasmic nitrate reductase in *Pseudomonas* sp. strain G-179 catalyzes the first step of denitrification. J. Bacteriol. **181:**2802–2806.
7. **Bergmann, D. J., D. M. Arciero, and A. B. Hooper.** 1994. Organization of the *hao* gene cluster of *Nitrosomonas europaea*: genes for two tetraheme *c* cytochromes. J. Bacteriol. **176:**3148–3153.
8. **Bergmann, D. J., and A. B. Hooper.** 1994. The primary structure of cytochrome P460 of *Nitrosomonas europaea*: presence of a c-heme binding motif. FEBS Lett. **353:**324–326.
9. **Bergmann, D. J., and A. B. Hooper.** 1994. Sequence of the gene, *amoB*, for the 43-kDa polypeptide of ammonia monooxygenase of *Nitrosomonas europaea*. Biochem. Biophys. Res. Commun. **204:**759–762.
10. **Bergmann, D. J., J. A. Zahn, A. B. Hooper, and A. A. DiSpirito.** 1998. Cytochrome P460 genes from the methanotroph *Methylococcus capsulatus* bath. J. Bacteriol. **180:**6440–6445.

11. **Bhandari, B., and D. J. Nicholas.** 1981. Some properties of glutamine synthetase from the nitrifying bacterium *Nitrosomonas europaea*. Aust. J. Biol. Sci. **34:**527–539.

12. **Bock, E.** 1995. Nitrogen loss caused by denitrifying *Nitrosomonas* cells using ammonium or hydrogen as electron donors and nitrite as electron acceptor. Arch. Microbiol. **163:**16–20.

13. **Bren, A., and M. Eisenbach.** 2000. How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. J. Bacteriol. **182:**6865–6873.

14. **Brown, J. R., and W. F. Doolittle.** 1999. Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J. Mol. Evol. **49:**485–495.

15. **Bussereau, F., L. Mallet, L. Gaillon, and M. Jacquet.** 1993. A 12.8-kb segment, on the right arm of chromosome II from *Saccharomyces cerevisiae* including part of the DUR1,2 gene, contains five putative new genes. Yeast **9:**797–806.

16. **Cannon, G. C., C. E. Bradburne, H. C. Aldrich, S. H. Baker, S. Heinhorst, and J. M. Shively.** 2001. Microcompartments in prokaryotes: carboxysomes and related polyhedra. Appl. Environ. Microbiol. **67:**5351–5361.

17. **Casciotti, K. L., and B. B. Ward.** 2001. Dissimilatory nitrite reductase genes from autotrophic ammonia-oxidizing bacteria. Appl. Environ. Microbiol. **67:**2213–2221.

18. **Cavalier-Smith, T.** 2002. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int. J. Syst. Evol. Microbiol. **52:**7–76.

19. **Clark, C., and E. L. Schmidt.** 1966. Effect of mixed culture on *Nitrosomonas europaea* simulated by uptake and utilization of pyruvate. J. Bacteriol. **91:** 367–373.

20. **Clark, C., and E. L. Schmidt.** 1967. Uptake and utilization of amino acids by resting cells of *Nitrosomonas europaea*. J. Bacteriol. **91:**1309–1315.

21. **Costa-Seaver, L., and J. A. Imlay.** 2001. Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in *Escherichia coli*. J. Bacteriol. **183:**7173–7181.

22. **Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.** 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392:** 353–358.

23. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.

24. **DiSpirito, A. A., L. R. Taaffee, J. D. Lipscomb, and A. B. Hooper.** 1985. A "blue" copper oxidase from *Nitrosomonas europaea*. Biochim. Biophys. Acta **827:**320–326.

25. **Ditty, J. L., A. C. Grimm, and C. S. Harwood.** 1998. Identification of a chemotaxis gene region from *Pseudomonas putida*. FEMS Microbiol. Lett. **159:**267–273.

26. **Enz, S., S. Mahren, U. H. Stroeher, and V. Braun.** 2000. Surface signaling in ferric citrate transport gene induction: interaction of the FecA, FecR, and FecI regulatory proteins. J. Bacteriol. **182:**637–646.

27. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.

28. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.

29. **Fani, R., E. Mori, E. Tamburini, and A. Lazcano.** 1998. Evolution of the structure and chromosomal distribution of histidine biosynthetic genes. Orig. Life Evol. Biosph. **28:**555–570.

30. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496– 512.

31. **Genbauffe, F. S., and T. G. Cooper.** 1986. Induction and repression of the urea amidolyase gene in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **6:**3954– 3964.

32. **Genbauffe, F. S., and T. G. Cooper.** 1991. The urea amidolyase (DUR1,2) gene of *Saccharomyces cerevisiae*. DNA Seq. **2:**19–32.

33. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.

34. **Head, I. M., W. D. Hiorns, T. M. Embley, A. J. McCarthy, and J. R. Saunders.** 1993. The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16S ribosomal RNA gene sequences. J. Gen. Microbiol. **139:**1147–1153.

35. **Heath, R. J., and C. O. Rock.** 1996. Roles of the FabA and FabZ β-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. J. Biol. Chem. **271:**27795–27801.

36. **Hirota, R., A. Yamagata, J. Kato, A. Kuroda, T. Ikeda, N. Takiguchi, and H. Ohtake.** 2000. Physical map location of the multicopy genes coding for ammonia monooxygenase and hydroxylamine oxidoreductase in the ammonia-oxidizing bacterium *Nitrosomonas* sp. strain ENI-11. J. Bacteriol. **182:** 825–828.

37. **Hommes, N. G., L. A. Sayavedra-Soto, and D. J. Arp.** 1998. Mutagenesis and expression of *amo*, which codes for ammonia monooxygenase in *Nitrosomonas europaea*. J. Bacteriol. **180:**3353–3359.

38. **Hommes, N. G., L. A. Sayavedra-Soto, and D. J. Arp.** 1996. Mutagenesis of hydroxylamine oxidoreductase in *Nitrosomonas europaea* by transformation and recombination. J. Bacteriol. **178:**3710–3714.

39. **Hommes, N. G., L. A. Sayavedra-Soto, and D. J. Arp.** 1994. Sequence of *hcy*, a gene encoding cytochrome *c-554* in *Nitrosomonas europaea*. Gene **146:**87– 89.

40. **Hooper, A. B., J. Hansen, and R. Bell.** 1967. Characterization of glutamate dehydrogenase from the ammonia-oxidizing chemoautotroph *Nitrosomonas europaea*. J. Biol. Chem. **242:**288–296.

41. **Hooper, A. B., T. Vannelli, D. J. Bergmann, and D. M. Arciero.** 1997. Enzymology of the oxidation of ammonia to nitrite by bacteria. Antonie Leeuwenhoek **71:**59–67.

42. **Howell, M. L., E. Alsabbagh, J. F. Ma, U. A. Ochsner, M. G. Klotz, T. J. Beveridge, K. M. Blumenthal, E. C. Niederhoffer, R. E. Morris, D. Needham, G. E. Dean, M. A. Wani, and D. J. Hassett.** 2000. AnkB, a periplasmic ankyrin-like protein in *Pseudomonas aeruginosa*, is required for optimal catalase B (KatB) activity and resistance to hydrogen peroxide. J. Bacteriol. **182:**4545–4556.

43. **Jahnke, J. S., C. Lyman, and A. B. Hooper.** 1984. Carbonic anhydrase, carbon dioxide levels, and growth of *Nitrosomonas*. Arch. Microbiol. **140:** 291–293.

44. **Juliette, L. Y., M. R. Hyman, and D. J. Arp.** 1993. Mechanism-based inactivation of ammonia monooxygenase in *Nitrosomonas europaea* by allylsulfide. Appl. Environ. Microbiol. **59:**3728–3735.

45. **Keener, W. K., and D. J. Arp.** 1994. Transformations of aromatic compounds by *Nitrosomonas europaea*. Appl. Environ. Microbiol. **60:**1914–1920.

46. **Kim, U. J., H. Shizuya, P. J. de Jong, B. Birren, and M. I. Simon.** 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res. **20:**1083–1085.

47. **Klotz, M. G., J. Alzerreca, and J. M. Norton.** 1997. A gene encoding a membrane protein exists upstream of the *amoA/amoB* genes in ammonia oxidizing bacteria: a third member of the *amo* operon? FEMS Microbiol. Lett. **150:**65–73.

48. **Korshunov, S. S., and J. A. Imlay.** 2002. A potential role for periplasmic superoxide dismutase in blocking the penetration of external superoxide into the cytosol of gram-negative bacteria. Mol. Microbiol. **43:**95–106.

49. **Kusian, B., D. Sultemeyer, and B. Bowien.** 2002. Carbonic anhydrase is essential for growth of *Ralstonia eutropha* at ambient CO_2 concentrations. J. Bacteriol. **184:**5018–5026.

50. **Laue, B. E., Y. Jiang, S. R. Chhabra, S. Jacob, G. S. Stewart, A. Hardman, J. A. Downie, F. O'Gara, and P. Williams.** 2000. The biocontrol strain *Pseudomonas fluorescens* F113 produces the *Rhizobium* small bacteriocin, *N*-(3-hydroxy-7-*cis*-tetradecenoyl)homoserine lactone, via HdtS, a putative novel *N*-acylhomoserine lactone synthase. Microbiology **146:**2469–2480.

51. **Macnab, R. M.** 1992. Genetics and biogenesis of bacterial flagella. Annu. Rev. Genet. **26:**131–158.

52. **Maeda, S., G. D. Price, M. R. Badger, C. Enomoto, and T. Omata.** 2000. Bicarbonate binding activity of the CmpA protein of the cyanobacterium *Synechococcus* sp. strain PCC 7942 involved in active transport of bicarbonate. J. Biol. Chem. **275:**20551–20555.

53. **Margolin, W.** 2000. Themes and variations in prokaryotic cell division. FEMS Microbiol. Rev. **24:**531–548.

54. **McTavish, H., J. A. Fuchs, and A. B. Hooper.** 1993. Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. J. Bacteriol. **175:**2436–2444.

55. **McTavish, H., F. LaQuier, D. Arciero, M. Logan, G. Mundfrom, J. A. Fuchs, and A. B. Hooper.** 1993. Multiple copies of genes coding for electron transport proteins in the bacterium *Nitrosomonas europaea*. J. Bacteriol. **175:** 2445–2447.

56. **Mellies, J., J. Jose, and T. F. Meyer.** 1997. The *Neisseria gonorrhoeae* gene *aniA* encodes an inducible nitrite reductase. Mol. Gen. Genet. **256:**525–532.

57. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, C. M. Fraser, et al.** 1999. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323–329.

58. **Nielsen, A. K., K. Gerdes, and J. C. Murrell.** 1997. Copper-dependent reciprocal transcriptional regulation of methane monooxygenase genes in *Methylococcus capsulatus* and *Methylosinus trichosporium*. Mol. Microbiol. **25:**399–409.

59. **Prosser, J. I.** 1989. Autotrophic nitrification in bacteria. Adv. Microb. Physiol. **30:**125–177.

60. **Prosser, J. I. (ed.).** 1986. Nitrification, vol. 20. IRL Press, Washington, D.C.
61. **Rasche, M. E., M. R. Hyman, and D. J. Arp.** 1991. Factors limiting aliphatic chlorocarbon degradation by *Nitrosomonas europaea*: cometabolic inactivation of ammonia monooxygenase and substrate specificity. Appl. Environ. Microbiol. **57:**2986–2994.
62. **Ritchie, G. A., and D. J. Nicholas.** 1974. The partial characterization of purified nitrite reductase and hydroxylamine oxidase from *Nitrosomonas europaea*. Biochem. J. **138:**471–480.
63. **Salanoubat, M., S. Genin, F. Artiguenave, J. Gouzy, S. Mangenot, M. Arlat, A. Billault, P. Brottier, J. C. Camus, L. Cattolico, M. Chandler, N. Choisne, C. Claudel-Renard, S. Cunnac, N. Demange, C. Gaspin, M. Lavie, A. Moisan, C. Robert, W. Saurin, T. Schiex, P. Siguier, P. Thebault, M. Whalen, P. Wincker, M. Levy, J. Weissenbach, and C. A. Boucher.** 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature **415:**497–502.
64. **Sayavedra-Soto, L. A., N. G. Hommes, and D. J. Arp.** 1994. Characterization of the gene encoding hydroxylamine oxidoreductase in *Nitrosomonas europaea*. J. Bacteriol. **176:**504–510.
65. **Sayavedra-Soto, L. A., N. G. Hommes, S. A. Russell, and D. J. Arp.** 1996. Induction of ammonia monooxygenase and hydroxylamine oxidoreductase mRNAs by ammonium in *Nitrosomonas europaea*. Mol. Microbiol. **20:**541–548.
66. **Serres, M. H., S. Gopal, L. A. Nahum, P. Liang, T. Gaasterland, and M. Riley.** 2001. A functional update of the *Escherichia coli* K-12 genome. Genome Biol. **2:**RESEARCH0035.1–RESEARCH0035.7. [Online.]
67. **Shively, J. M., G. van Keulen, and W. G. Meijer.** 1998. Something from almost nothing: carbon dioxide fixation in chemoautotrophs. Annu. Rev. Microbiol. **52:**191–230.
68. **Smith, A. J., and D. S. Hoare.** 1968. Acetate assimilation by *Nitrobacter agilis* in relation to its "obligate autotrophy." J. Bacteriol. **95:**844–855.
69. **Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve, et al.** 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. J. Bacteriol. **179:**7135–7155.
70. **Storz, G., and J. A. Imlay.** 1999. Oxidative stress. Curr. Opin. Microbiol. **2:**188–194.
71. **Subramaniam, P., R. Bhatnagar, A. Hooper, and R. A. Jensen.** 1994. The dynamic progression of evolved character states for aromatic amino acid biosynthesis in gram-negative bacteria. Microbiology **140:**3431–3440.
72. **Terry, K. R., and A. B. Hooper.** 1970. Polyphosphate and orthophosphate content of *Nitrosomonas europaea* as a function of growth. J. Bacteriol. **103:**199–206.
73. **Teske, A., E. Alm, J. M. Regan, S. Toze, B. E. Rittmann, and D. A. Stahl.** 1994. Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. J. Bacteriol. **176:**6623–6630.
74. **Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter.** 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287:**1809–1815.
75. **Thompson, D. K., A. S. Beliaev, C. S. Giometti, S. L. Tollaksen, T. Khare, D. P. Lies, K. H. Nealson, H. Lim, J. Yates III, C. C. Brandt, J. M. Tiedje, and J. Zhou.** 2002. Transcriptional and proteomic analysis of a ferric uptake regulator (Fur) mutant of *Shewanella oneidensis*: possible involvement of Fur in energy metabolism, transcriptional regulation, and oxidative stress. Appl. Environ. Microbiol. **68:**881–892.
76. **Vannelli, T., M. Logan, D. M. Arciero, and A. B. Hooper.** 1990. Degradation of halogenated aliphatic compounds by the ammonia-oxidizing bacterium *Nitrosomonas europaea*. Appl. Environ. Microbiol. **56:**1169–1171.
77. **Vassinova, N., and D. Kozyrev.** 2000. A method for direct cloning of Fur-regulated genes: identification of seven new Fur-regulated loci in *Escherichia coli*. Microbiology **146**(Pt. 12):3171–3182.
78. **Ward, A., C. Hoyle, S. Palmer, J. O'Reilly, J. Griffith, M. Pos, S. Morrison, B. Poolman, M. Gwynne, and P. Henderson.** 2001. Prokaryote multidrug efflux proteins of the major facilitator superfamily: amplified expression, purification and characterization. J. Mol. Microbiol. Biotechnol. **3:**193–200.
79. **Whittaker, M., D. Bergmann, D. Arciero, and A. B. Hooper.** 2000. Electron transfer during the oxidation of ammonia by the chemolithotrophic bacterium *Nitrosomonas europaea*. Biochim. Biophys. Acta **1459:**346–355.
80. **Winzer, K., C. Falconer, N. C. Garber, S. P. Diggle, M. Camara, and P. Williams.** 2000. The *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL are controlled by quorum sensing and by RpoS. J. Bacteriol. **182:**6401–6411.
81. **Withers, H., S. Swift, and P. Williams.** 2001. Quorum sensing as an integral component of gene regulatory networks in gram-negative bacteria. Curr. Opin. Microbiol. **4:**186–193.
82. **Woese, C. R., E. Stackebrandt, W. G. Weisburg, B. J. Paster, M. T. Madigan, V. J. Fowler, C. M. Hahn, P. Blanz, R. Gupta, K. H. Nealson, and G. E. Fox.** 1984. The phylogeny of purple bacteria: the alpha subdivision. Syst. Appl. Microbiol. **5:**315–326.
83. **Wood, P. M.** 1986. Nitrification as a bacterial energy source, p. 39–62. *In* J. I. Prosser (ed.), Nitrification. Society for General Microbiology/IRL Press, Oxford, England.
84. **Yoo, J. G., and B. Bowien.** 1995. Analysis of the *cbbF* genes from *Alcaligenes eutrophus* that encode fructose-1,6-/sedoheptulose-1,7-bisphosphatase. Curr. Microbiol. **31:**55–61.

Vol. 185, No. 21

AUTHOR'S CORRECTION

Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph *Nitrosomonas europaea*

Patrick Chain, Jane Lamerdin, Frank Larimer, Warren Regala, Victoria Lao,
Miriam Land, Loren Hauser, Alan Hooper, Martin Klotz, Jeanette Norton,
Luis Sayavedra-Soto, Dave Arciero, Norman Hommes,
Mark Whittaker, and Daniel Arp

*Joint Genome Institute, Walnut Creek, California 94598; Lawrence Livermore National Laboratory, Livermore, California 94550;
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831; University of Minnesota, Minneapolis, Minnesota 55455;
University of Louisville, Louisville, Kentucky 40208; Utah State University, Logan, Utah 84322; and
Oregon State University, Corvallis, Oregon 97331*

Volume 185, no. 9, p. 2759–2773, 2003. Page 2767, column 2: The first sentence should read as follows: "Genes for aminoacyl tRNA synthetases are present for all amino acids except asparagine."

Page 2767, column 2: The fourth sentence should read as follows: "A *gatCAB* operon, putatively encoding a Glu-tRNA amidotransferase, was identified, although in this bacterium the enzyme may function to transamidate Asp-tRNA Asn."