Evolution and Functional Characterization of the RH50 Gene from the Ammonia-Oxidizing Bacterium *Nitrosomonas europaea* †

Baya Cherif-Zahar, 1,2 Anne Durand, 3 Ingo Schmidt, 4 Nabila Hamdaoui, 1,2 Ivan Matic, 2,5 Mike Merrick, 6 and Giorgio Matassi* 1,4

INSERM, U845, Paris F-75015, France 1; Université Paris Descartes, Faculté de Médecine René Descartes, Paris F-75015, France 2; Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom 3; Department of Microbiology, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany 4; INSERM, U571, Paris F-75015, France 5; and Institut Jacques Monod, CNRS-UMR 7592, Université Paris 6 et Paris 7, 2 Place Jussieu, 75251 Cedex 05, Paris, France 6

Received 10 July 2007/Accepted 4 September 2007

The family of ammonia and ammonium channel proteins comprises the Amt proteins, which are present in all three domains of life with the notable exception of vertebrates, and the homologous Rh proteins (Rh50 and Rh30) that have been described thus far only in eukaryotes. The existence of an RH50 gene in bacteria was first revealed by the genome sequencing of the ammonia-oxidizing bacterium *Nitrosomonas europaea*. Here we have used a phylogenetic approach to study the evolution of the *N. europaea* RH50 gene, and we show that this gene, probably as a component of an integron cassette, has been transferred to the *N. europaea* genome by horizontal gene transfer. In addition, by functionally characterizing the RH50Ne protein and the corresponding knockout mutant, we determined that NeRh50 can mediate ammonium uptake. The RH50Ne gene may thus have replaced functionally the *AMT* gene, which is missing in the genome of *N. europaea* and may be regarded as a case of nonorthologous gene displacement.

Since the first description of rhesus (Rh) antigens in 1940 by Landsteiner and Wiener (45), the Rh blood group has become, after the ABO group, the most clinically significant in transfusion medicine. It is the most polymorphic of human blood groups, consisting of at least 45 independent antigens, and is consequently also widely used in human population genetics studies. In humans, Rh antigens are carried by two erythrocyte membrane proteins, named RhD and RhCE, and are also referred to as Rh30 because of their apparent molecular mass of 30 to 32 kDa. These proteins are coded by two very similar paralogous genes (ca. 96% identical at the nucleotide level), located in tandem on chromosome 1p34-36. The human genome also codes for three other members of the Rh family, the Rh50 transmembrane glycoproteins (Rh50A/RhAG, Rh50B/RhBG, and Rh50C/RhCG) that have an apparent molecular mass of 50 to 58 kDa (6). The Rh50A protein is erythroid specific, like Rh30, and is associated with Rh50 in a multiprotein complex in the red blood cell membrane (10). Rh50A expression is required for Rh blood group antigen expression at the red blood cell membrane (14), and its lack of expression results in the Rh-null phenotype (68). In mammals, the non-erythroid Rh50B and Rh50C proteins are expressed in the kidney, liver, and gastrointestinal tract (79).

The clinical importance of the Rh30 proteins has tended to overshadow the status of the Rh50 proteins. However, RH50 genes have a much longer evolutionary history than RH30 and the latter are likely to have derived by duplication from an RH50-like ancestor (38, 52). Indeed, while RH50 genes are present in the genome of the basal deuterostome animals sequenced thus far, the sea urchin Strongylocentrotus purpuratus (echinoderms), Ciona intestinalis and Ciona savignyi (tunicates), and amphioxus (cephalochordates), RH30 genes are absent in these species. Moreover, RH50 and RH30 genes are both present in teleost fish, amphibians, and mammals. Molecular evolutionary analyses have shown that in mammals Rh50 proteins evolved at a lower rate than Rh30 (38, 52), which is in line with the general trend of a higher evolutionary rate of newly duplicated genes.

Rh50 proteins share a low, albeit significant, sequence similarity (20 to 25% identity) with ammonium transport proteins of the Amt family (48, 53) although, as expected, such similarity is barely detectable for the fast-evolving Rh30 proteins. We use the term ammonium to refer to both the protonated (NH4+) and the unprotonated (NH3) forms and the term ammonia to refer specifically to NH3.

Furthermore, expression of the human Rh50A and Rh50C (also named RhGK) can restore the growth on ammonium of a yeast mutant lacking the three endogenous *AMT* genes (49). Hence, it is now clear that Rh and Amt proteins are homologous (i.e., they are derived from a common ancestor by vertical descent), and they are assigned to the same protein family (Pfam PF00909). Molecular phylogenetic analyses (39, 72) indicated that Amt proteins can be partitioned into two clusters, named Amt-α and Amt-β (39), although the functional significance, if any, of this division remains unclear.

The biochemical function of Amt proteins as ammonium as ammonium (NH4+/NH3) channels is clearly established in bacteria, fungi, and plants (74, 82), and the first evidence for the biological role of Amt in animals has recently been obtained in that ammo-
nium channel expression is essential for brain development and function in the larva of the ascidian *Ciona intestinalis* (51). In contrast, the biochemical function of Rh50 is a subject of a controversy centered on whether the Rh50 substrate is ammonia or CO₂ (43, 82). However, it has also been suggested that Rh proteins may be relatively nonspecific channels for neutral small molecules (10).

The motivation of the present study stems from our interest in the study of the evolutionary history of the Amt/Rh protein family, which spans the entire tree of life. To date, three different evolutionary scenarios have had to be taken into account. First, bacterial and archaeal genomes code only for AMT genes, RH50 genes being absent. Second, both AMT and RH50 genes coexist in some genomes. This is found in a range of eukaryotes from unicellular protists, to dictyostelids, to chon-

When studying the evolutionary history of the Amt/Rh protein family, it is important to note that the RH50 protein can be found in both prokaryotes and eukaryotes. The study of the evolutionary history of the Amt/Rh protein family has been crucial in understanding the function and distribution of these proteins across different taxa.

**Materials and Methods**

**Phylogenetic analyses.** The genome of *N. europaea* is available at the JGI-IMG resource. Fifteen datasets of homologous proteins were obtained using the *N. europaea* sequences (JGI locus tags Ne0441 to Ne0457) as a query in BLAST searches (2) against the bacterial genome sequences currently available at the JGI-IMG resource (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi), the National Center for Biotechnology Information (NCBI) Microbial Genomes Resource (http://www.ncbi.nlm.nih.gov/nuccore/). and the TIGR-CMR Comprehensive Microbial Resource (http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi). The genome of *Candidatus Kuenenia stuttgartiensis* was BLAST searched at http://www.ncbi.nlm.nih.gov/nuccore. The phylogenetic analysis was performed by using PHYML 2.4.4 (24). Nonparametric bootstrap analysis (100 replicates, as implemented in PHYML) and an approximate likelihood ratio test (aLRT, based on a Shimodaira-Hasegawa-like procedure, http://atgc.lirmm.fr) (4) were used as branch support measures.

The strains and plasmids used in the present study are described in Table 1. N. europaea cells were grown in liquid medium (29) containing 25 mM (NH₄)₂SO₄ in the dark at 30°C on a rotary shaker (200 rpm) and reached an optical density at 600 nm (OD 600) of 0.1 in approximately 3 to 4 days. For growth of N. europaea on solid medium, the liquid medium was modified by adding 1% agar R2A (Difco Laboratories, Detroit, MI) and by replacing the phosphate buffer with 50 mM N-tris(hydroxy-methyl)methyl-2-aminoethane sulfonic acid buffer (pH 7.8) (28). N. europaea cells were streaked onto an autoclaved Nytran membrane (Schleicher & Schuell), which was laid out on solid medium and then incubated at 30°C; the membrane

FIG. 1. Molecular phylogenetic evidence for HGT. Maximum-likelihood trees were computed by using PHYML 2.4.4 from datasets comprising 38-taxa (238 amino acid positions), 34-taxa (297 amino acid positions), and 38-taxa (316 nucleotide sites, first and second codon positions) in Ne0448 (A), Ne0445 (B), and Ne0446 (C), respectively. The evolutionary models used were RtREV+Γ, WAG+Γ, and HKY+Γ in Ne0448, Ne0445, and Ne0446, respectively. Support values at the nodes correspond to 100 maximum-likelihood bootstrap replicates (only values ≥75% are shown in red) and aLRT statistics (only values ≥80% are shown in blue). Scale bars indicate the estimated number of substitutions per site. Different symbols were used to visualize the five proteobacteria subdivisions: α, black diamonds; β, blue squares (except for nitrosomonads identified by red squares); γ, green dots; δ, cyan upward triangles; and ε, pink downward triangles. (A) Rh50 (Ne0448) phylogeny. The ABO Nitrosomonas and Nitrospira cluster together and with the other two bacteria “Ca. Kuenenia” and Acidobacteria. The four bacterial proteins are positioned basal to the Rh50 subtree and are distinct from the Amt subfamily. These results provide no evidence either in favor or against HGT. (B) Uroporphyrinogen III decarboxylase (Ne0445) phylogeny. An example of no HGT is that the three nitrosomonads are grouped together and cluster with the other β-proteobacteria. The β-γ proteobacteria clade is recovered as well. (C) 3-demethylubiquinone-9 3-methyltransferase (Ne0446) phylogeny. The clustering of Ne0446 with Geobacter (β-proteobacteria) and Chlorobium (Chlorobi) (99% bootstrap, 99% aLRT) provides evidence of HGT. In contrast, the phylogeny of the Nitrosospira ortholog is congruent with the species tree (clustering with the β-proteobacteria). The same topology was also obtained under the GTR model of sequence evolution (data not shown). No ortholog of Ne0446 is present in N. eutropha. The arrow indicates the branch leading to the subtree comprising Ne0446 orthologs. (D) Summary of the phylogenetic analyses. The RH50α (gene Ne0448) and its neighbors on the chromosome are schematically represented by boxes (not to scale) that show the sense of transcription on each strand. Gene boxes are color coded as follows: cyan (no HGT), red (evidence for HGT), pink (hint for HGT), and yellow (integrases/transposase genes). The genes are Ne0441 (leucyl aminopeptidase, cluster of orthologous groups [COG], COG0260), Ne0442 (DNA polymerase III chi subunit, COG2927), Ne0443 (hypothetical protein; could not be analyzed; see Results), Ne0444 (valyl-tRNA synthetase, COG0525), Ne0445 (uroporphyrinogen III decarboxylase, COG0407), Ne0446 (hypothetical protein homologous to the 3-demethylubiquinone-9 3-methyltransferase, Pfam PF00983), Ne0447 (hypothetical protein homologous to 3-methyladenine DNA glycosylase I, COG2818), Ne0448 (Rh50, INTERPRO IPR002526), Ne0449 (aspartate/glutamate racemase, COG0796), Ne0450 (integron integrase Intl, PF00589, IPR11946, COG582), Ne0451 (integrase catalytic core, PF00665, IPR01584), Ne0452 (transposase IS911, IPR002514), Ne0453 (patatin-like phospholipase, PF01734), Ne0454 (integrase catalytic core, PF00589), Ne0455 (prokaryotic DksA/TraR C4-type zinc finger, PF01258), Ne0456 (esterase/lipase/thioesterase, COG0596), and Ne0457 (hypothetical protein, COG5316).
was transferred to new medium every 4 days. Cell suspensions were grown in the dark at 30°C under constant stirring (800 rpm). In all experiments, kanamycin was used at a concentration of 10 μg ml⁻¹.

**Cloning of the RH50 gene.** Genomic DNA was prepared from 10 ml of a 4-day culture of *N. europaea* (OD₆₀₀ = 0.1) by using a Wizard genomic preparation kit (Promega). The RH50 gene was amplified by PCR using the DsNaNzyme EXT DNA polymerase (Finnzymes) between primers P1 and P2 (see Table S1 in the supplemental material and for all other primers). A standard reaction was carried out using P2 and P3 primers and inserted into the plasmid pRL448 (kindly provided by Tim Fulford). The cells were grown in a chemostat from which cells (∼5 x 10⁶ cells ml⁻¹) were inoculated in mineral medium (63) and grown in the dark at 30°C. The experiments were performed in 300-ml Erlenmeyer flask containing 75 ml of mineral medium at different concentrations of ammonium (1.0, 2.0, and 10 mM ammonium chloride, pH 7.4; n = 4). At regular time intervals samples were taken, and the ammonia, nitrate, and protein concentrations were measured as previously described (9, 63). We define lag phase as the time required at 10 mM NH₄⁺ in the medium to oxidize the first 200 μM at 2 mM NH₄⁺ the first 200 μM, and at 1 mM NH₄⁺ the first 100 μM.

**Western blotting.** Yeast membranes were prepared as described elsewhere (www.stke.org/cgi/content/full/sigtrans;2005/275/pl3). Yeast cells were grown in 50 ml of yeast nutrient broth (YNB) plus 0.2% (wt/vol) glutamate at 30°C to an OD₆₀₀ of 1.0 and then harvested by centrifugation at 700 × g for 5 min at 4°C. The cell pellet was resuspended in 1 ml of 50 mM Tris-HCl (pH 7.5) supplemented with a mixture of protease inhibitors (Roche). The cells were then homogenized into a 2-ml tube, and 300 μl of chilled glass beads (size range, 425 to 600 μm; Sigma) was added. The cells were lysed by vigorous vortex mixing seven times for 30 s and then separated by treatment at 1 min on ice. The lysate was centrifuged at 700 × g for 20 min at 4°C. The supernatant was spun at 700 × g for 10 min at 4°C. The membrane and cytosolic fractions were separated by a centrifugation step at 150,000 × g for 2 h at 4°C. The cytosolic fraction was kept for analysis, and the membrane pellet was resuspended in 50 mM Tris-HCl (pH 7.5) with 1% Triton X-100 and centrifuged as described above. The resulting supernatant (membrane) was kept for Western blot analysis. Western blotting was performed as previously described (17), and proteins were detected with anti-His antibodies (QIAGEN) or anti-AmpB antibodies.

**E. coli transport assays (unwashed).** *E. coli* strains were grown in Luria medium. For growth in nitrogen-limiting conditions, a modified M9 medium was used (M9Gln) that contained 0.2% glucose as a carbon source and glutamine (which replaced ammonium) at 200 μM ml⁻¹. The unwashed assays were performed with Rh50expressing cells (GT1000 cells transformed with pAD7 and pAD9) and E. coli AmpB-expressing cells (GT1000 transformed with plasmid pTOE vector). Experiments were carried out at room temperature as described previously (32).

**S. cerevisiae** transport assays (washed). [¹⁴C]Methylammonium ([¹⁴C]MA) uptake assays (adapted from a previously reported method [80]) were performed with cells expressing RS60 (31019b cells transformed with plasmid pAD10), E. coli AmpB cells (31019b transformed with pTF14), and control cells (2334ac and 31019b cells transformed with empty vector pDR195). Cells were grown overnight in YNB minimal medium (Difco Laboratories) supplemented with 3% (wt/vol) glucose and 0.2% (wt/vol) glutamate. When the cells exceeded an OD₆₀₀ of 0.5, they were washed and resuspended in 3% (wt/vol) glucose buffered with 10 mM phosphate (pH 6.0 or 7.0) or 50 mM HEPEs (pH 8.0). In the assay, a final concentration of 500 μM MA was used. Briefly, 500 μl of 5 mM MA containing 0.725 μCi of [¹⁴C]MA (Amersham Biosciences) was added to 4.5 ml of the growth medium to oxidize the first 200 μM at 2 mM NH₄⁺ the first 200 μM, and at 1 mM NH₄⁺ the first 100 μM.

---

**TABLE 1. Strains and plasmids**

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Description*</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli GT1000</td>
<td>rbs lacZ::IS gyrA hisCΔ Δgkn-ampB</td>
<td>17</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2334ac</td>
<td>MATa ura3</td>
<td>50</td>
</tr>
<tr>
<td>31019b</td>
<td>MATa ura3 mepΔΔ mep2ΔΔ::LEU2 mep3ΔΔ::Km MX2</td>
<td>50</td>
</tr>
<tr>
<td>N. europaea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATCC 19178</td>
<td>Wild type</td>
<td>ATCC</td>
</tr>
<tr>
<td>RH50Ne-KO</td>
<td>RH50::Km insertion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAD7</td>
<td>RH50-His₆₅ cloned into pESV2</td>
<td>This study</td>
</tr>
<tr>
<td>pAD8</td>
<td>RH50 cloned into pESV2</td>
<td>This study</td>
</tr>
<tr>
<td>pAD9</td>
<td>RH50-His₆₅ cloned into pDR195</td>
<td>This study</td>
</tr>
<tr>
<td>pAD10</td>
<td>RH50 cloned into pDR195</td>
<td>This study</td>
</tr>
<tr>
<td>pDR195</td>
<td>Yeast expression vector (URA₄₅; Amp')</td>
<td></td>
</tr>
<tr>
<td>pESV2</td>
<td>amtBΔ (Δgkn derivative of pESV1)</td>
<td>30</td>
</tr>
<tr>
<td>pTF5E</td>
<td>glnK-ampB15 wild-type AmtBₑₑₑ</td>
<td>31</td>
</tr>
<tr>
<td>pTF14</td>
<td>Wild-type AmtBₑₑₑ into pDR195</td>
<td></td>
</tr>
<tr>
<td>pCR2.1-TOPO</td>
<td>Cloning vector (lacZα Km')</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pQB1IN1</td>
<td>Expression vector (GFP Amp'; ori plC; ori)</td>
<td>Q-Biogen</td>
</tr>
<tr>
<td>pRL448</td>
<td>pRL139 derivative (Amp' Km')</td>
<td>21</td>
</tr>
<tr>
<td>pKD13</td>
<td>Template plasmid (oriRy Km')</td>
<td>18</td>
</tr>
<tr>
<td>pNe1</td>
<td>RH50::Km cloned into pCR2.1-TOPO</td>
<td>This study</td>
</tr>
<tr>
<td>pNe2</td>
<td>RH50::Km cloned into pQB1IN1</td>
<td>This study</td>
</tr>
<tr>
<td>pNe3</td>
<td>pNe2 RH50::Km</td>
<td>This study</td>
</tr>
<tr>
<td>pNe4</td>
<td>pRL448 RH50::Km</td>
<td>This study</td>
</tr>
</tbody>
</table>

*Amplification; Km', kanamycin resistance.  
**ATCC, American Type Culture Collection.
RESULTS

Evidence for HGT of RH50<sub>N</sub>. We have BLAST searched about 700 bacterial and 34 archaeal completed and draft genome sequences for RH50 homologs, and we identified RH50-like sequences in four bacterial species: two β-proteobacterial AOB, <i>N. europaea</i> ATCC 19718 (Nitrosomonas cluster; JGI-IMG locus tag Ne0448/NCBI Entrez protein accession no. NP_840553), and *Nitrosospira multiformis* ATCC 25196 (Nitrosospira cluster; locus tag Nmul_A0516/YP_411216), the acidobacterium *Acidobacterium bacterium* Ellin345 (locus tag Acid345_3596/YP_592671), and the planctomycete “Ca. Kuenenia stuttgartiensis” (CAJ711226). The existence of RH50 homologs was found in both AOB in the RH50 were reported previously (13, 43). In addition, we noted that AMT genes are missing in two of the AOB genomes, whereas *Acidobacterium bacterium* and “Ca. Kuenenia” possess two (Acid345_3596/YP_593520 and Acid345_3596/YP_590566) and four (CAJ71754, CAJ71757, CAJ71760, and CAJ74453) AMT genes, respectively.

The RH50 gene was found to be missing in the genomes of other members of the corresponding clades. Based on BLAST searches against the currently available genome sequences (JGI-IMG), two other AOB, *N. europaea* C71 (β-proteobacteria, Nitrosomonas cluster) and *Nitrosococcus oceani* ATCC 19707 (γ-proteobacteria), appeared to lack both RH50 and AMT genes. “Ca. Kuenenia stuttgartiensis” belongs to the phylum Planctomycetes, and BLAST searches have been carried out against the two other genomes currently available in this phylum (JGI-IMG), namely, *Rhodopirellula baltica* SH1 (complete) and Blastopirellula marina DSM 3645 (draft), indicating that they lack RH50 genes while possessing three and four AMT genes, respectively. Together with “Acidobacterium bacterium,” two other genome sequences are currently available in the phylum Acidobacteria: Solibacter usitatus Ellin6076 (JGI-IMG) and “Acidobacterium capsulatum” (proposed name) ATCC 51196 (TIGR). Again, BLAST searches revealed the lack of RH50 and the presence of at least two AMT genes in *A. capsulatum* and *S. usitatus*. Interestingly, the genomes of *A. bacterium* and “Ca. Kuenenia stuttgartiensis” also possess two and four copies of the AMT gene, respectively. To our knowledge, this is the first case described thus far of coexistence of AMT and RH50 genes in bacterial genomes.

The presence of RH50 genes in only four bacterial species could be accounted for by a large number of independent gene losses in all of the bacterial lineages known to date. However, this scenario is classically regarded as highly unlikely. The alternative and most parsimonious explanation is that the bacterial RH50 genes have been acquired by HGT (also described as lateral gene transfer), an evolutionary process whereby genetic material is exchanged between distantly related species (20, 41, 56). Provided enough taxon sampling is available, phylogenetic analysis remains the most powerful method to detect the likely occurrence of an HGT event (41, 70). Briefly, a known phylogenetic species tree, taken as a reference, is compared to the tree inferred from the gene under study and incongruent topologies are regarded as evidence of HGT. However, it has been stressed that any kind of evidence of HGT is “always a probabilistic one and rarely direct” (41).

To explore the HGT hypothesis for NeRH50, we carried out phylogenetic analyses on a data set comprising a subset of 29 Rh50 protein sequences, as well as four Amt-a and five Amt-β proteins, using a maximum-likelihood approach. The data set was arbitrarily chosen in order to render the analyses computationally tractable while including the largest possible taxon sampling for eukaryotic Rh50 proteins. Our results show that the Rh50 proteins from *N. europaea* and *N. multiformis* are closely related, as expected (87% bootstrap, 88% aLRT support; see Materials and Methods) and that, together with their homologs from *A. bacterium* and “Ca. Kuenenia stuttgartiensis,” they are distinct (100% bootstrap, 100% aLRT) from the AOB subfamily (Fig. 1A).

The four bacterial Rh50 proteins were clearly separated from the eukaryotic homologs and are positioned at the base of the RH50 subfamily clade. Therefore, in this case no comparison between species tree and gene tree can be done, and thus phylogenetic evidence for RH50<sub>HGT</sub> cannot be obtained, since the basal positioning of the bacterial Rh50 proteins is to be expected, whether they are “true” bacterial or “HGT-acquired” sequences that have accumulated enough divergence. Consequently, we sought evidence of HGT by analyzing the 16 genes (Fig. 1D) that are neighbors of RH50 (Ne0448) on the *N. europaea* chromosome (locus tags Ne0441 to Ne0457).

No phylogenetic analysis could be carried out in the case of Ne0443 (that encodes a hypothetical protein) since no hits were found by BLAST search, except for one sequence in *N. eutropha* (locus tag Neut_0601). For each of the remaining 15 genes we carried out phylogenetic analyses in order to detect incongruence between the species tree, the proteobacterial tree in this case, and the gene tree. The current phylogeny of the five subdivisions of proteobacteria (based on 16S and 23S rRNA, whole-genome comparisons, and molecular signatures) give δ and ε subdivisions as early branching (which one of them is more basal is not clear) and largely support the (α (β, γ)) topology (25, 71).

Figure 1B shows an example of congruence between species tree and gene tree (i.e., no HGT) in the case of the uroporphyrinogen decarboxylase. The three orthologs from the AOB *N. europaea* (Ne0445), *Nitrosomonas eutropha*, and *N. multiformis* cluster together (Fig. 1B, red squares) and are confi-
dently (99% bootstrap, 99% aLRT) located within the β-proteobacteria (Fig. 1B, blue squares). In contrast, the phylogeny of the 3-demethylubiquinone-9 3-methyltransferase gene (Fig. 1C) shows on the one hand the incongruent clustering of Ne0446 with Geobacter spp. (β-proteobacteria) and with Chlorobium lanimica (Chlorobi) (99% bootstrap, 99% aLRT) and on the other hand the congruent clustering of the Nitrosospira ortholog (no ortholog present in N. eutropha) with the β-proteobacteria. These results suggest that Ne0446 entered the N. europaea genome via an HGT event. Incidentally, the Chlorobi bacte-
roides C. ferrooxidans name), “C. ferrooxidans” (proposed name), C. phaeobacte-
roides, C. tepidum, and Porphyromonas gingivalis) and could not
find homologs. To infer this phylogeny, we have used the first
and second codon positions of the genes, since the analysis
based on amino acids did not provide sufficient resolution (not
shown). The results of the phylogenetic analyses of the other
13 neighboring genes are given in the supplemental material
(all datasets and multiple alignments are available upon re-
quest).

Supporting evidence for the HGT hypothesis is given by the
presence of three genes adjacent to and upstream from the
RH50 (Ne0448) gene on the same strand, which code for an
integron-integrase intI (Ne0450, INTERPRO IPR11946), a
catalytic-core integrase (Ne0451, IPR001584; another such in-
tegrase is coded by Ne0454 on the other strand), and an IS911
transposase (Ne0452), all proteins being involved in processes
of DNA recombination and exchange. We carried out phylo-
genetic analyses of these four proteins, and we detected evi-
dence of HGT in the case of Ne0454, which clusters with
α-proteobacteria, but no evidence of HGT for Ne0450, Ne0451,
and Ne0452 (see Fig. S3 in the supplemental mate-
rial).

Three additional neighboring genes are likely HGT candi-
dates, although with various levels of confidence. Significant
evidence was obtained for Ne0447: no homologs were found in
nitrosomonads other than N. europaea, and BLAST searches
identified homologs only in α- and γ-proteobacteria, with
which Ne0447 clusters (see Fig. S2 in the supplemental mate-
rial). In the case of the prokaryotic DksA/TraR transcriptional
regulator, Ne0455, N. multiformis and N. eu-
trophus orthologs clustered with the γ-proteobacteria, while the β-proteobacte-
rial homologs clustered with the α-proteobacteria (see Fig. S4
in the supplemental material). In the phylogeny of the aspar-
gtate-glutamate racemase, Ne0449 and its N. eu-
trophus ortholog cluster with two γ-proteobacteria and a δ-proteobacterium
(92% aLRT, but no bootstrap support), thereby suggesting the
occurrence of an HGT event (see Fig. S2 in the supplemental mate-
rial). Evidence that may indicate HGT was also obtained for
the Ne0456 gene, which is separated by the α- and β-pro-
teobacterial orthologs (81% aLRT, but no bootstrap support)
(see Fig. S4 in the supplemental material). Finally, we obtained
no evidence for HGT in the case of Ne0441, Ne0442, Ne0444,
Ne0453, and Ne0457 (see Fig. S1, S2, and S4 in the supple-
mental material).

The results of the analyses on the 16 gene neighbors of
RH50Ne (Ne0448), located in the chromosomal region span-
ning about 20 kb of the N. europaea chromosome, are sche-
matically summarized in Fig. 1D. Overall, our results suggest
that there are two regions that are candidates for HGT. The
first one consists of the Ne0455 gene and possibly of Ne0456;
adjacent to this region is Ne0454, which codes for an integrase.
The second region comprises at least Ne0446, Ne0451, and
Ne0449. We deduce by inference that Ne0448 (RH50Ne), which
is embedded in this region, is likely to have undergone HGT.

Functional characterization of RH50Ne. Given that there are
no AMT genes in the N. europaea genome, we sought to de-
terminethe whether the RH50 gene could have functionally re-
placed AMT. To test this hypothesis, we used two strategies.
First, we cloned RH50Ne in suitable vectors and expressed it in
E. coli and S. cerevisiae deletion mutants deficient in ammo-
nium uptake. Second, we constructed a knockout RH50Ne
(RH50Ne-KO) and compared its growth on ammonium and
its ability to transport MA to that of the wild type.
Plasmid constructs were made that encoded both the wild-
type type A, B, and a derivative His-tagged (Hisα) at the C
terminus. The latter construct facilitated detection of the
protein by Western blotting with an anti-His antibody. These
constructs were used to express Rh50Ne in the ammonia chan-
nel mutant strains E. coli GT1000 (∆amtB) and S. cerevisiae
strain 31019b (∆mep1-3). NeRh50 was corrected locally to the
membrane both in E. coli (Fig. 2A, lanes 5 to 7) and in
yeast (Fig. 2B, lane 6), as determined by Western blot analysis
of cell fractions. In the case of yeast in particular the Rh50Ne
signal was much stronger in the membrane fraction than in the
whole-cell extract, but the material loaded in Fig. 2B, lane 6, is

FIG. 2. Expression of Rh50Ne in E. coli and S. cerevisiae. (A) De-
tection and localization of Rh50Ne when expressed in E. coli. A West-
ern blot was performed with anti-His antibody. Lane 1, purified E. coli
AmpT-NeAmpB-His6; lanes 2 to 4, GT1000(pAD8) expressing NeRh50; lanes 5
to 7, GT1000(pAD9) expressing Rh50Ne-Hisα; lanes 2 and 5, whole-cell
extract; lanes 3 and 6, cytoplasmic fraction; lanes 4 and 7, membrane
fraction; lane 8, molecular mass markers (indicated in kilodaltons). (B) De-
tection and localization of Rh50Ne when expressed in S. cerevi-
siae. Lanes 1 to 3, S. cerevisiae strain 31019b(pTF14) expressing E. coli
AmtB detected using anti-AmtB antibody; lanes 4 to 6, S. cerevisiae
strain 31019b(pAD9) expressing Rh50Ne detected using anti-His anti-
body; lanes 1 and 4, whole-cell extract; lanes 2 and 5, cytoplasmic
fraction; lanes 3 and 6, membrane fraction. All lanes were loaded with
5 μg of protein, and consequently the relative concentrations of mem-
brane proteins are greater in the membrane fraction than in the whole-
cell extract. Lane 8, molecular mass markers (indicated in kilodaltons).
concentrated relative to that in lane 4. The apparent molecular mass of the predominant molecular species of Rh50<sub>Ne</sub> was ~36 kDa in E. coli and ~30 kDa in yeast. This could predictably correspond to a folded form of the monomer, which has a theoretical molecular mass of ~43 kDa. A small percentage of the Rh50<sub>Ne</sub> expressed in E. coli ran as a higher-molecular-mass species that could reflect a trimeric form.

We assessed the transport activity of Rh50<sub>Ne</sub> with or without the His tag by measuring the influx of the ammonium analogue [14C]MA at different pH values in E. coli and yeast cells using washed and unwashed assays, respectively. The washed assay measures MA influx into cells and its subsequent assimilation into methylglutamine by glutamine synthetase (32), whereas the unwashed assay measures all MA influx into the cell, including that which is not converted to methylglutamine. No significant transport activity was detected in E. coli cells either at 20 µM MA (pH 7.0) or at 250 µM (pH 7.0 and 8.0) (data not shown). Similarly, no significant accumulation of [14C]MA was observed in yeast cells grown at 500 µM MA at pH 6.0, 7.0, or 8.0 (data not shown). Therefore, we used an alternative approach to assess Rh50<sub>Ne</sub> ammonium channel activity by analyzing whether the expression of the protein (with or without a His tag) could restore the ammonium-dependent growth phenotype of the S. cerevisiae 31019b strain, and we compared it with the corresponding activity of the E. coli AmtB protein (AmtB<sub>Ec</sub>) expressed in the same strain.

The transformed yeast cells were plated on nitrogen-free medium supplemented with 0.5, 1.0, or 3.0 mM ammonium chloride at three different pH values (6.0, 6.5, and 7.0). After 5 days of incubation at 30°C, Rh50<sub>Ne</sub> expression restored the growth of the yeast mutant in a pH-dependent manner (Fig. 3). At pH 6.0 growth restoration was only apparent on 1 mM ammonium chloride, but when the pH was raised to 6.5 or 7.0 the restoration of growth was apparent on 0.5 mM ammonium and was significantly improved at 1 mM ammonium. Rh50<sub>Ne</sub> activity was identical with or without His tag. The phenotype observed for Rh50<sub>Ne</sub> was different from that obtained with AmtB<sub>Ec</sub> since expression of the latter restored growth at all pHs and at all of the ammonium chloride concentrations tested (Fig. 3E).

We also studied the ammonium transport properties of Rh50<sub>Ne</sub> in N. europaea. To this end, we inactivated the RH50<sub>Ne</sub> gene by insertional mutagenesis using a kanamycin resistance cassette. In a first set of experiments we compared the growth of N. europaea wild-type and RH50<sub>Ne</sub>-KO strains at different concentrations of ammonium (1.0, 2.0, and 10 mM ammonium chloride, pH 7.4; n = 4). Growth was monitored by measuring ammonia oxidation and nitrite formation. At low ammonium concentrations the lag phase of the KO mutant was significantly longer than that of the wild type (115 ± 8 min versus 85 ± 5 min at 1 mM and 85 ± 6 min versus 71 ± 2 min at 2 mM, respectively). We also compared the growth of the wild type and the KO mutant at various concentrations of CO<sub>2</sub> (0.04, 0.1, 0.5, and 1.0%) in a chemostat culture, and we did not detect any effect on the growth rate, growth yield, or ammonia oxidation activity (data not shown).

Finally, we compared MA influx in N. europaea wild-type and RH50<sub>Ne</sub>-KO strains by performing unwashed transport assays, again using [14C]MA as a radiotracer. The MA uptake activities were examined during the first 4 min, and the results (Fig. 4) demonstrate that the MA uptake rate of the mutant cells (58.6 ± 18.2 nmol mg of protein<sup>-1</sup> min<sup>-1</sup>) was about five times lower than that of wild-type cells (252.7 ± 78.9 nmol mg protein<sup>-1</sup> min<sup>-1</sup>). From these data, we conclude that the Rh50 protein is involved in ammonium uptake in N. europaea.

FIG. 3. RH50<sub>Ne</sub> restores growth in the S. cerevisiae Δmepl-3 mutant. Growth was determined after 5 days at 30°C on minimal YNB medium containing ammonium chloride at 0.5, 1.0, or 3.0 mM. Experiments were carried out with media adjusted to pH 6.0, 6.5, or 7.0. The strains tested were 23344c(pDR195) (mepl<sup>-</sup>) (A), 31019b(pDR195) (Δmepl-3) (B), 31019b(pAD10) (Rh50<sub>Ne</sub>) (C), 31019b(pAD9) (Rh50<sub>Ne</sub>-His<sub>a</sub>) (D), and 31019b(pTF14) (AmtB<sub>Ec</sub>) (E).

FIG. 4. [14C]MA uptake in N. europaea wild type and RH50 KO mutant. The MA concentration in the uptake assay was adjusted to 10 mM, and the cell number was 5 × 10<sup>6</sup> cells ml<sup>-1</sup>. The results are means ± the standard deviation (n = 6). Symbols: ●, N. europaea wild type; □, RH50<sub>Ne</sub>-KO mutant.
DISCUSSION

HGT of RH50 to N. europaea. Not only does evolution at the molecular level proceed via the inheritance of genetic material by vertical descent from a common ancestor, but organisms can also inherit genetic information by nonvertical descent due to the ability of DNA to be carried across species by bacteriophage and plasmids, to be exchanged by transformation, and to be absorbed from the environment and integrated in the genome. These “alternative” processes are collectively referred to as horizontal (or lateral) gene transfer, which is now recognized as a major force in the evolution of at least prokaryotic genomes (20, 41, 56).

We used a phylogenetic approach to assess the likelihood of HGT of the RH50 gene in N. europaea. This analysis was not very informative since, together with the other three bacterial Rh50 proteins, Rh50$_{\text{Ne}}$ occupies a basal position within the Rh50 clade (Fig. 1A). Therefore, we could not determine whether the genes encoding these four proteins are truly bacterial or HGT-acquired sequences that have accumulated divergence. Similar basal positioning was reported for the Pros- thecoc bacter BtubA and BtubB tubulin HGT candidate genes (34). However, we have generated four kinds of indirect evidence supporting the HGT of RH50$_{\text{Ne}}$, which may also hint at HGT as being responsible for the spread of the RH50 gene to N. multiformis, “Ca. Kuenenia stuttgartiensis,” and A. ba- cterium.

First, only four species out of more than 700 bacterial genomes possess a copy of an RH50 gene. Second, the RH50 gene is missing in the sequenced genomes of two AOB (N. eutropha, and N. oceanii), two planctomycetes (R. baltica and B. marina), and two acidobacteria (A. capsulatum and S. usitatus). Third, and most importantly, we have shown that RH50$_{\text{Ne}}$ (Ne0448) is embedded in a region comprising three genes (Ne0446, Ne0447, and Ne0449), all of which, according to our phylogenetic analyses, are candidate horizontally transferred genes (Fig. 1D). Fourth, this HGT candidate region is adjacent to three genes coding for an integron integrase IntI (Ne0450), a catalytic-core integrase (Ne0451), and an IS911 transposase (Ne0452) (Fig. 1D). In particular, integron intI genes are involved in the formation of integron cassettes which are believed to “function as a general gene-capture system in bacterial adaptation” and thus are regarded as key players in HGT events (60). Moreover, the second copy of the IntI integrase gene in the N. europaea genome (Ne2189) was shown to be able to promote excision and integration of resistance gene cassettes (46). The region spanning Ne0446 to Ne0450 may thus correspond to an integron cassette, and we infer that the N. europaea RH50 gene entered the genome via HGT.

HGT of RH50$_{\text{Ne}}$: hypothetical scenarios. Two scenarios can be envisaged to account for the presence of the RH50 gene in N. europaea (and possibly in N. multiformis). We may assume that a single HGT event has occurred in the common ancestor to the AOB β-proteobacteria. If such were the case, we ought to deduce that, for some as-yet-unexplained reasons, a gene loss event has occurred specifically in the N. eutropha lineage (the same AOB cluster as N. europaea). Alternatively, a single HGT event may have occurred in the N. europaea lineage. The first scenario would be favored if we were to demonstrate that the N. multiformis RH50 gene has also been acquired by HGT. Phylogenetic analyses are currently under way to test the hypothesis of HGT for the RH50 genes from Nitrososira multi- formis, “Ca. Kuenenia stuttgartiensis,” and Acidobacteria bac- terium and will be presented elsewhere.

The phylogeny of the bacterial Rh50 proteins (Fig. 1A) contains interesting and puzzling information nonetheless. We note that the branches separating the AOB (N. europaea and N. multiformis) from Acidobacteria sp. and “Ca. Kuenenia” are much shorter than expected given that those species belong to three different phyla. In trying to understand the evolution of the bacterial RH50 genes, we reason as if three independent HGT events had occurred in the three lineages leading to AOB β-proteobacteria, “Ca. Kuenenia,” and Acidobacteria. Short branches may result from exceptionally slow evolution rates (be it adaptive or due to regional or genome-wide mutational biases) of the four RH50 genes, which is highly unlikely (even more so if these genes were truly bacterial). Alternatively and most likely, these short branches may indicate recent HGT events. However, such an explanation contrasts with the clear separation between the bacterial RH50 genes and their eukaryotic homologs. To account for this apparent incongruence, we propose a working hypothesis whereby the donor RH50, likely of eukaryotic origin, may have first entered an “intermediate” prokaryote and then may have been transferred to the three lineages analyzed here. However, the currently available taxon sampling for bacterial RH50 genes, which is expected to expand as more genome data become available, does not allow us to draw any conclusion in this respect. Overall, taking also into account the potential role of the IntI integrase in the spread of RH50, we envisage a eukaryote–prokaryote–prokaryote HGT scenario as a possibility, even though this phenomenon has not been described in the literature thus far.

To date, the vast majority of the reported cases of the trans- fer of genetic material among the three domains of life concern reciprocal exchanges from bacteria to archaea and from prokaryotes to eukaryotes (20, 41). In contrast, evidence for DNA transfer from eukaryotes to prokaryotes is much rarer and is mainly restricted to symbiotic or parasitic relationships (41). Following the hypothesis of a eukaryotic origin for the RH50 gene in N. europaea, this HGT event may be one of the rare cases of eukaryotic versus bacterial transfer.

HGT of RH50$_{\text{Ne}}$: a case of nonorthologous gene displace- ment. Koonin et al. proposed to classify HGT events into a minimum of three categories “with respect to the relationships between the horizontally acquired gene and homologous genes (if any) preexisting in the recipient lineage” (41). In one category, a phylogenetically “distant” ortholog is acquired by HGT, followed by xenologous gene displacement (xeno = foreign), i.e., the elimination of the preexisting gene. In the other two categories, HGT drives the acquisition of either an unrelated gene, which is absent in other members of the same clade, or a distantly related paralog. Again, the acquisition of the new gene may result in the loss of the resident gene; in such a case the phenomenon is dubbed nonorthologous gene displacement (40).

The presence of a copy of the RH50 gene in only four bacterial species out of the more than 700 prokaryotic genomes (bacteria and archaea) we have analyzed, among which AMTs are widely distributed, suggests that RH50 is likely to have arisen by a gene duplication event from an AMT-like
eukaryotic ancestor. The acquisition of RH50 by *N. europaea* may be included in the HGT category of xenologous gene displacement (41), if we follow the most recent definition of this phenomenon given by Koonin (42) and consider AMT and RH50 as members of the same orthologous cluster (COG0004). However, if we consider AMT and RH50 as paralogs, then the HGT of RH50 was to be regarded as a non-orthologous gene displacement event (40). The preexisting *N. europaea* AMT gene was displaced (replaced) by the HGT-acquired RH50; it remains open to question whether this replacement has taken place via a loss-and-regain strategy (following consecutive and opposite adaptive selection pressures) or else the two genes have coexisted in *N. europaea* for a given time span (20). However, the presence of both AMT and RH50 in the genomes of *Ca. Kuenenia* and Acidobacteria (see above) may favor the second hypothesis.

**RH50Ne mediates ammonium uptake and may thus have replaced AMT functionally.** The proposal of Soupene et al. (65) that Amt proteins act as NH₃ channels has been supported by X-ray crystal structures of *E. coli*AmtB (36, 83) and *Archaeoglobus fulgidus* Amt1 (3) and by in vivo studies on *E. coli* AmtB (32). However, studies on other Amt proteins (from *E. coli* and at least four AMT orthologous genes (37). The green alga *Chlamydomonas reinhardii* has two RH50 genes (CrRH1 and CrRh2) and at least four AMT genes (37). Expression of both CrRh1 mRNA and CrRh1 protein increase when the available CO₂ is increased from 0.035 to 3% (66), and RNA interference silencing of CrRh1 protein in vivo showed no evidence of a CO₂-dependent growth effect in a *RH50NeKO* mutant. Hence, we presently have no evidence for its extension to RH50 homologs in other species. The hypotheses on the function of RH50 proteins, as ammonium or CO₂ channels (67) or as nonspecific channels for neutral small molecules (10) are of course not mutually exclusive since it is well known that homologous proteins may have the same functions, similar but also different functions, or multiple functions in different organisms (22).

In *N. europaea* ammonia appears to play a critical role serving as a signal leading to regulation of gene transcription (61, 75, 76), as a metabolic substrate, and as the sole energy source for ammonia oxidation in aerobic conditions. Based on the *N. europaea* genome sequence, Chain et al. predicted that ammonium entry into the cell might occur via the Rh50 protein (13). However, evidence has also been provided for the existence in *N. europaea* of an active membrane-potential-driven transport mechanism (13), which might enable ammonium to reach an internal (cytoplasm plus periplasm) concentration as high as 1 M (62) and allow the organism to cope with the very low concentrations of ammonium present in the environment (8).

Our results indicate that the Rh50Ne protein is involved in ammonium uptake; its depletion in the corresponding KO mutant results in an extended lag phase on low ammonium and in markedly reduced MA uptake compared to the wild type (Fig. 4). Modeling of the human Rh50A protein based on the X-ray crystal structure of the *E. coli* AmtB protein has shown that Rh50A is likely to adopt a structure similar to that of AmtB, and Rh50A is expected to have a channel architecture very similar to that of AmtB (11, 15). Indeed, we recently solved the X-ray crystal structure of Rh50Ne (47a) and showed it to be a trimeric protein with a channel architecture very similar to that predicted by our previous homology modeling of Rh50A (15). Hence, Rh proteins appear to have many of the essential characteristics to facilitate ammonium uptake, and our empirical data with Rh50Ne (Fig. 3 and 4) support this proposition.

Restoration of ammonium-dependent growth to a yeast Δ*mep* mutant by Rh50Ne is more effective as the pH increases (Fig. 3), a result that is compatible either with facilitated transport of NH₃ through the channel or with a NH₃/H⁺ exchanger. However, one feature of Amt proteins is a periplasmic vestibule that has been proposed to contain a binding site for the ammonium ion, defined by three conserved residues (Phe103, Trp148, and Ser219 in AmtB) (15, 47a), and if Rh50Ne were functioning as an NH₃ channel, then the absence of this putative ammonium ion binding site would be consistent with the pH-dependent growth that we observe when Rh50Ne is expressed in the yeast Δ*mep* mutant.

The failure of Rh50Ne to mediate MA uptake is not unprecedented, and similar observations have been made with human Rh50A expressed in *S. cerevisiae* (49, 80). Although it is conceivable that at least some Rh50 proteins might mediate CO₂ uptake, our experiments to characterize Rh50Ne in vivo showed no evidence of a CO₂-dependent growth effect in a RH50NeKO mutant. Hence, we presently have no evidence that Rh50Ne functions as a CO₂ channel.

Overall, we have provided evidence indicating that the Rh50 protein mediates ammonium uptake in *N. europaea* and thus may have replaced functionally a preexisting Amt protein (though the reasons for the loss of the AMT gene are unknown). It has been shown that ammonium is a crucial factor for the differentiation of the ecological niche of AOB (8). When the ammonium concentration in the environment is low, AOB, given their low growth rate, suffer badly from the competition of other microbes for energy sources. AOB appear to have developed a survival strategy at the molecular level by regulating transcription levels. Indeed, under starvation conditions (NH₃ plus CO₂ removal), while ca. 68% of *N. europaea* genes are downregulated (at least twofold) and ammonia monooxygenase and hydroxylamine oxidoreductase transcrip-
tion levels are maintained, 10 genes, mainly related to oxidative stress, are upregulated by more than twofold, and transcription of \( \text{Rhs}^{50}_{\text{Ne}} \) increases nearly twofold (76). \( \text{Rhs}^{50}_{\text{Ne}} \) may therefore be a key player (as a channel and/or a sensor of ammonium) in the survival strategy of \( N. \text{europea} \), which has adapted to long periods of starvation while preparing for the uptake and oxidation of ammonium as soon as it becomes available (8).

**ACKNOWLEDGMENTS**

G.M. expresses the warmest thanks to Dick D’Ari, without whom this work would not have been completed. B.C.-Z. thanks Aleksander Edelman for supporting this study. We thank Tim Fulford for helpful advice on the execution of the yeast experiments and for the construction of plasmid pTF14 and Jasmin Wickinghoff for excellent technical assistance. We thank Norman Hommes for the gift of \( N. \text{europea} \) ATCC 9093 and the \( \text{Rhs}^{50}_{\text{Ne}} \) plasmid for helpful advice and discussion. We thank Hervé Philippe for very helpful discussions and suggestions.

This work was supported by a grant from the Biotechnology and Biological Sciences Research Council to M.M. and A.D.

**ADDENDUM**

Evidence that \( \text{Rhs}^{50}_{\text{Ne}} \) mediates ammonium transport has also been provided by Weidinger et al. (77).

**REFERENCES**


32. Landsteiner, K., and A. S. Wiener. 1940. An agglutinable factor in human