Origin of microbial biomineralization and magnetotaxis during the Archean

Wei Lin[a,b,1,2], Greig A. Paterson[a,2], Qiyun Zhu[c], Yinzhao Wang[a,b], Evguenia Kopylova[d], Ying Li[e], Rob Knight[d,f], Dennis A. Bazylinski[g], Rixiang Zhu[h], Joseph L. Kirschvink[i,j,1], and Yongxin Pan[a,b,1]

[a]Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China; [b]France–China Bio-Mineralization and Nano-Structures Laboratory, Chinese Academy of Sciences, Beijing 100029, China; [c]Genomic Medicine, J. Craig Venter Institute, La Jolla, CA 92037; [d]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92037; [e]College of Biological Sciences, China Agricultural University, Beijing 100193, China; [f]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92037; [g]School of Life Sciences, University of Nevada, Las Vegas, NV 89154-4004; [h]State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China; [i]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125; and [j]Earth–Life Science Institute, Tokyo Institute of Technology, Meguro, Tokyo 152-8551, Japan

Edited by Donald E. Canfield, Institute of Biology and Nordic Center for Earth Evolution, University of Southern Denmark, Odense M, Denmark, and approved January 10, 2017 (received for review September 3, 2016)

Microbes that synthesize minerals, a process known as microbial biomineralization, contributed substantially to the evolution of current planetary environments through numerous important geochemical processes. Despite its geological significance, the origin and evolution of microbial biomineralization remain poorly understood. Through combined metagenomic and phylogenetic analyses of deep-branching magnetotactic bacteria from the *Nitrospirae* phylum, and using a Bayesian molecular clock-dating method, we show here that the gene cluster responsible for biomineralization of magnetosomes, and the arrangement of magnetosome chain(s) within cells, both originated before or near the Archean divergence between the *Nitrospirae* and *Proteobacteria*. This phylogenetic divergence occurred well before the Great Oxygenation Event. Magnetotaxis likely evolved due to environmental pressures conferring an evolutionary advantage to navigation via the geomagnetic field. Earth's dynamo must therefore have been sufficiently strong to sustain microbial magnetotaxis in the Archean, suggesting that magnetotaxis coevolved with the geodynamo over geological time.

Archean | microbial biomineralization | magnetotaxis | magnetotactic bacteria | geodynamo

Magnetotactic bacteria (MTB) are a polyphyletic group of microorganisms that biomineralize intracellular nano-sized magnetosomes of magnetite (Fe_3O_4) and/or greigite (Fe_3S_4) (1). Magnetosomes are normally organized into chain-like structures to facilitate the navigation of MTB using Earth's magnetic field, a behavior known as magnetotaxis (2). Because of their ubiquitous distribution, MTB play key roles in global iron, nitrogen, sulfur, and carbon cycling (3). Magnetosome crystals can be preserved as magnetofossils after MTB die and lyse, and these crystals can be used as reliable biomarkers and are major contributors to sedimentary paleomagnetic records (4–7). Evidence suggests that magnetofossils as old as ∼1.9 Ga may be preserved (4).

Molecular and genetic studies have identified a group of clustered genes that control magnetosome biomineralization in MTB (8–10). However, the origin and evolution of these magnetosome gene clusters (MGCs) remain controversial, and several scenarios have been posited to explain the patchy distribution of magnetosome formation over a broad phylogenetic range. Such scenarios include multiple evolutionary origins (11), extensive horizontal gene transfers (12), and vertical gene transfer (13, 14). MTB in the *Nitrospirae* phylum represent one of the deep-branching MTB groups (15, 16). Therefore, analysis of MGCs from *Nitrospirae* MTB could yield insights into the origin and evolution of magnetosome biomineralization and magnetotaxis. However, no *Nitrospirae* MTB has been successfully cultured, and only three draft genomes from this phylum are currently available (17, 18). Recent advances in high-throughput sequencing combined with improving computational methods are enabling the successful recovery of high-quality population genomes directly from metagenomic data (19, 20). Here, we assess whether the phylogenies of key magnetosome genes from MTB of the *Nitrospirae* and *Proteobacteria* phyla are consistent with their taxonomic phylogeny using a metagenomic approach to acquire the population genome and MGC-containing contigs from environmental *Nitrospirae* MTB.

Results

We sampled MTB from two freshwater locations in China: the city moat of Xi'an in Shaanxi province (HCH) and Lake Miyun near Beijing (MY). MTB belonging to the *Nitrospirae* phylum were identified in both samples (Fig. S1). Recovered metagenomic DNA was sequenced and assembled into contigs, which generated ∼13 Mb of 3,042 contigs ≥1 kb for HCH and ∼32 Mb of 7,893

Significance

A wide range of organisms sense Earth's magnetic field for navigation. For some organisms, like magnetotactic bacteria, magnetic particles form inside cells and act like a compass. However, the origin of magnetotactic behavior remains a mystery. We report that magnetotaxis evolved in bacteria during the Archean, before or near the divergence between the *Nitrospirae* and *Proteobacteria* phyla, suggesting that magnetotactic bacteria are one of the earliest magnetic-sensing and biomineralizing organisms on Earth. The early origin for magnetotaxis would have provided evolutionary advantages in coping with environmental challenges faced by microorganisms on early Earth. The persistence of magnetotaxis in separate lineages implies the temporal continuity of geomagnetic field, and this biological evidence provides a constraint on the evolution of the geodynamo.

Author contributions: W.L., J.L.K., and Y.P. designed research; W.L. and Y.W. performed research; Q.Z. contributed new reagents/analytic tools; W.L. and Y.W. collected samples; W.L., G.A.P., and Q.Z. analyzed data; and W.L., G.A.P., Q.Z., E.K., Y.L., R.K., D.A.B., R.Z., J.L.K., and Y.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The draft genome of *Candidatus* Magnetominusculus xianensis strain HCH-1 reported in this paper has been deposited in the DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank database (accession no. LNQR00000000; the version described in this paper is LNQR01000000). The magnetosome gene cluster-containing contigs reported in this paper have been deposited in the DDBJ/EMBL/GenBank database (accession nos. KU221504–KU221507).

[1]To whom correspondence may be addressed. Email: weilin0408@gmail.com, kirschvink@caltech.edu, or yxpan@mail.iggcas.ac.cn.

[2]W.L. and G.A.P. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1614654114/-/DCSupplemental.

contigs ≥1 kb for MY (Table S1). In HCH, only a single population belonging to the phylum *Nitrospirae* was found and comprised ~19% of the sampled MTB community (Fig. S1). The population genome of this *Nitrospirae* was recovered using a combination of similarity- and composition-based approaches (Table 1). Only a single copy of an rRNA operon was identified, which had a 16S rRNA gene sequence with <92% similarity to the genera *Candidatus* Magnetobacterium (17, 18) and *Candidatus* Magnetoovum (18). This MTB population was named as "*Candidatus* Magnetominusculus xianensis" strain HCH-1 (HCH-1 hereafter). One contig (18,138 bp) of HCH-1 contained a nearly complete MGC including homologs to known magnetosome genes (Fig. 1*A* and Table S2). Despite their phylogenetic distance, the gene content and gene order of the MGC from HCH-1 and available *Nitrospirae* MGCs were highly conserved (>76%, Fig. 1*A*).

To explore the evolutionary history of the MGCs between the *Nitrospirae* and *Proteobacteria* phyla, we performed a phylogenetic analysis of five core magnetosome proteins (MamABEKP) from HCH-1 and published MTB strains. These proteins were selected because they were the only bidirectional best hits between the complete MGC of the *Nitrospirae* MTB strain "*Candidatus* Magnetobacterium casensis" (Mcas) and that of representative *Proteobacteria* MTB. These five proteins from the *Nitrospirae* phylum form a monophyletic group (Fig. S2). Branches with >75% bootstrap values have similar topology and agree with those in the tree based on concatenated magnetosome proteins (Fig. 1*B*). This is consistent with the pattern in the phylogenomic tree that is generated based on a concatenated alignment of up to 400 highly conserved proteins (Fig. 1*B*), suggesting that these magnetosome genes coevolved via vertical transmission with the genome. Furthermore, comparison of magnetosome genes to three housekeeping genes (*recA*, *gyrB*, and *pyrH*) revealed that the codon use bias was not significantly different ($P > 0.05$ by t test) between magnetosome genes and the vertically transmitted housekeeping genes. Similarly, the number of synonymous substitutions per synonymous site was also not significantly different ($P > 0.05$ by t test; Fig. 1*C*). The results of the codon use test are consistent with that of the phylogenetic-based analysis, which rules out uncertainty in the reliability of the codon analysis alone (21), and strongly suggest that the magnetosome genes were inherited through vertical transfer (22). Together, these results indicate that magnetosome biomineralization and magnetotaxis is an ancient metabolic process and was present before the separation of the *Nitrospirae* and *Proteobacteria* phyla, or transferred undetectably early between the base of *Nitrospirae* and *Proteobacteria* soon after divergence.

Table 1. General genomic features of the genome of *Candidatus* Magnetominusculus xianensis strain HCH-1

Parameter	*Ca.* Magnetominusculus xianensis strain HCH-1
Total genome size, Mb	3.593273
GC content, %	45.40
No. of contigs	152
N50, kb	45.767
Maximum contig length, kb	150.216
No. of coding sequences (CDS)	3,415
No. of RNAs	47
No. of copies of 5S rRNA	1
No. of copies of 16S rRNA	1
No. of copies of 23S rRNA	1
Estimated completeness, %	98.18
Estimated contamination, %	0.91

Considering the deep-branching lineage of *Nitrospirae* MTB and their tightly packed magnetosome genes (Fig. 1*A*), the content and order of magnetosome genes in *Nitrospirae* are likely conserved and represent an ancestral MGC. To test this, we compared the implicit phylogenetic pattern of the core magnetosome genes to all other genes in HCH-1 and available *Nitrospirae* MTB genomes. The magnetosome genes are clustered toward the majority of other genes (Fig. 2*A*), suggesting that the evolutionary history of magnetosome genes in *Nitrospirae* is not distinct from the genomic background. Analysis of metagenomic data from MY yielded four contigs containing nearly complete *Nitrospirae* MGCs (Fig. 2*B*). Of these, contig MY-22 (44,819 bp) was >99.99% identical to the MGC of previously identified Mcas isolated from the same lake (17), indicating the high-quality assembly of metagenomic contigs in this study. All four MGCs have a high level of conservation with HCH-1, Mcas, "*Candidatus* Magnetobacterium bavaricum" (Mbav), and "*Candidatus* Magnetoovum chiemensis" strain CS-04 (Mchi) (Fig. 2*B*). Phylogenetic trees of concatenated magnetosome proteins (Fig. 2*B*) and each of MamABEKP (Fig. S3) from the *Nitrospirae* MGCs form a monophyletic group with consistent topology, which further indicates the coevolution of magnetosome genes and their antiquity and suggests the origin of magnetotaxis is earlier than previously thought (13). Considering that all known *Nitrospirae* MTB biomineralize bullet-shaped magnetite magnetosomes, a crystal morphology that has not yet been observed in magnetite as a result of abiotic processes (3), our results indicate that bullet-shaped magnetite may be the first type of magnetosome (13) and may represent reliable microbial fossils.

There are very few geological constraints on the timing of the separation of the major clades within the Bacterial domain. Earlier evidence, based on organic biomarkers and the following phylogenetic analysis, suggested an Archean origin of the *Proteobacteria* (23) and the divergence of the *Nitrospirae* and *Proteobacteria* before ~2.7 Ga (24). However, the biomarkers upon which those inferences were based are now known to be contaminants (25). A global phylogenomic reconstruction of the evolution of gene families suggests that the *Proteobacteria* phylum diverged during the Archean Eon (26). We calculate the age of divergence for the *Nitrospirae* and *Proteobacteria* by phylogenomic and Bayesian relaxed molecular clock analyses (Figs. S4 and S5). Two different relaxed clock models (log normal autocorrelated clock and uncorrelated gamma multipliers clock) with two different combinations of time calibrations were implemented (*SI Materials and Methods*). Results are consistent across clock modes and time calibrations, and all analyses suggest that the *Nitrospirae* and *Proteobacteria* phyla diverged before 3.0 Ga (~3.38–3.21 Ga) (Fig. S4). Hence, MTB likely evolved in the mid-Archean, and, as previously suggested (4), may be one of the earliest biomineralizing and magnetotactic organisms on Earth.

Discussion

Molecular O_2 is abundant in the atmosphere (~21%) and oceans on the present Earth. The early Earth, however, was characterized by a reducing system, and the ocean chemistry in the Archean was different from today [e.g., with a scarcity of molecular O_2 and abundant dissolved Fe(II) (~40–120 μmol/L) (27, 28)]. Multiple lines of evidence have suggested that, through volcanic processes, Archean oceans likely had sufficient nutrients (e.g., CO_2, H_2, SO_2, H_2S, NO, NO_3^-, NO_2^-, NH_4^+, etc.) to sustain microorganisms with anaerobic or microaerobic metabolisms (29, 30). All known present-day MTB are microaerophilic and anaerobic, and, according to their metabolic and genomic analyses, it seems that the nutrients available in Archean oceans could support the survival and growth of MTB. For example, many extant MTB are chemolithoautotrophic and have the ability to fix CO_2 either via the Calvin–Benson–Bassham cycle, the reverse tricarboxylic acid cycle, or the reductive acetyl-CoA (Wood–Ljungdahl) pathway (1, 17, 18).

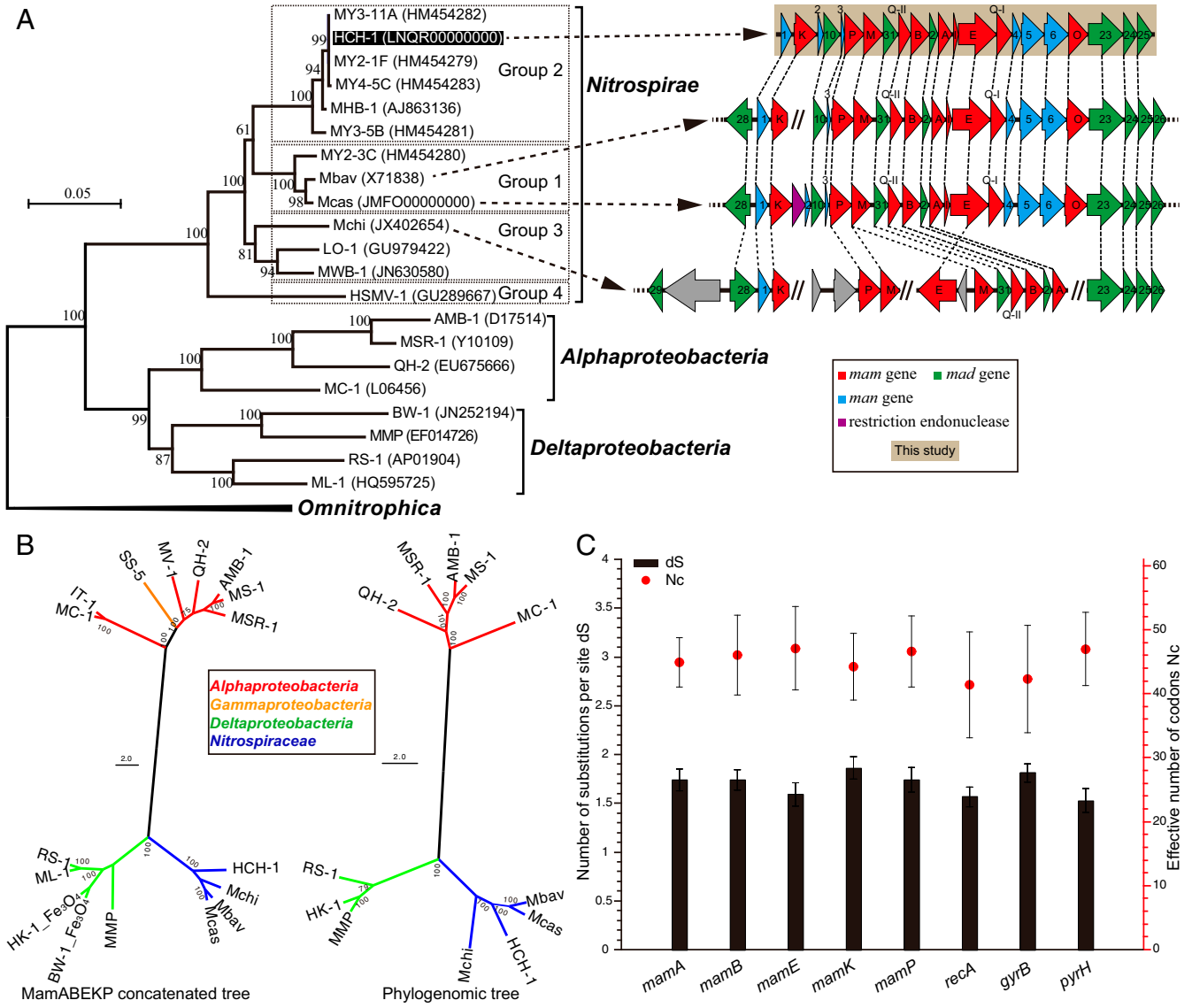

Fig. 1. (*A*) Maximum-likelihood phylogenetic tree of 16S rRNA gene sequences showing the relationship between *Candidatus* Magnetominusculus xianensis strain HCH-1 (HCH-1) and related magnetotactic bacteria. Bootstrap values at nodes are percentages of 100 replicates. On the right-hand side, comparison of magnetosome gene clusters between HCH-1 and other reported *Nitrospirae* MTB. (*B*) Bootstrap consensus trees based on concatenated protein alignment of five magnetosome proteins (MamABEKP) and phylogenomic tree based on concatenated ubiquitous amino acid sequences. Only bootstrap values of more than 75% are shown, and branches of the trees are proportionally transformed. MSR-1, *Magnetospirillum gryphiswaldense* MSR-1; AMB-1, *Magnetospirillum magneticum* AMB-1; QH-2, *Magnetospira* sp. QH-2; MC-1, *Magnetococcus marinus* MC-1; BW-1, "*Candidatus* Desulfamplus magnetomortis BW-1"; MMP, "*Candidatus* Magnetoglobus multicellularis"; RS-1, *Desulfovibrio magneticus* RS-1; ML-1, alkaliphilic magnetotactic strain ML-1; MV-1, *Magnetovibrio blakemorei* MV-1; IT-1, "*Candidatus* Magnetofaba australis strain IT-1"; SS-5, Gammaproteobacteria magnetotactic strain SS-5. (*C*) Comparison of sequence divergence of five magnetosome genes and three housekeeping genes (*recA*, *gyrB*, and *pyrH*). *Nc*, the codon use bias; dS, the number of synonymous substitutions per synonymous site.

They are also capable of reducing NO_3^-, NO_2^-, and NO through the denitrification pathway and capable of oxidizing H_2S through the sulfur oxidation pathway (e.g., refs. 17, 18, and 31–33). The temperature of Archean ocean water is still debated, with estimates ranging from values comparable to modern tropical waters (26–35 °C) (34) to as warm as 55–85 °C (35). Thermophilic MTB, living at temperatures of 32–63 °C, have been discovered in present-day hot springs (36), supporting the possibility of these bacteria surviving in a hot Archean ocean.

Extant MTB respond to vertical redox gradients in the water or sediment columns, particularly that of O_2 and H_2S, presumably exploiting these gradients to maintain their optimal positions within their microenvironments (1, 37). Archean oceans are traditionally thought to have been uniformly anoxic and therefore devoid of the redox gradients similar to today. Under such conditions, abundant ferrous iron would have been prevalent throughout the entire Archean water column, except in the euphotic zone near the surface where iron-oxidizing photosynthetic bacteria would presumably rapidly remove it from solution (38). The resultant ferric precipitates resulted in the production of the banded iron formations, yielding an "upside-down" biosphere (39), in which the sediments were rich in ferric iron but overlain by water rich in dissolved ferrous iron. This might have produced enough of a vertical redox gradient to provide MTB an environmental niche. Recently, however, there is growing evidence that early oceans might not be uniformly anoxic but were redox stratified, likely in

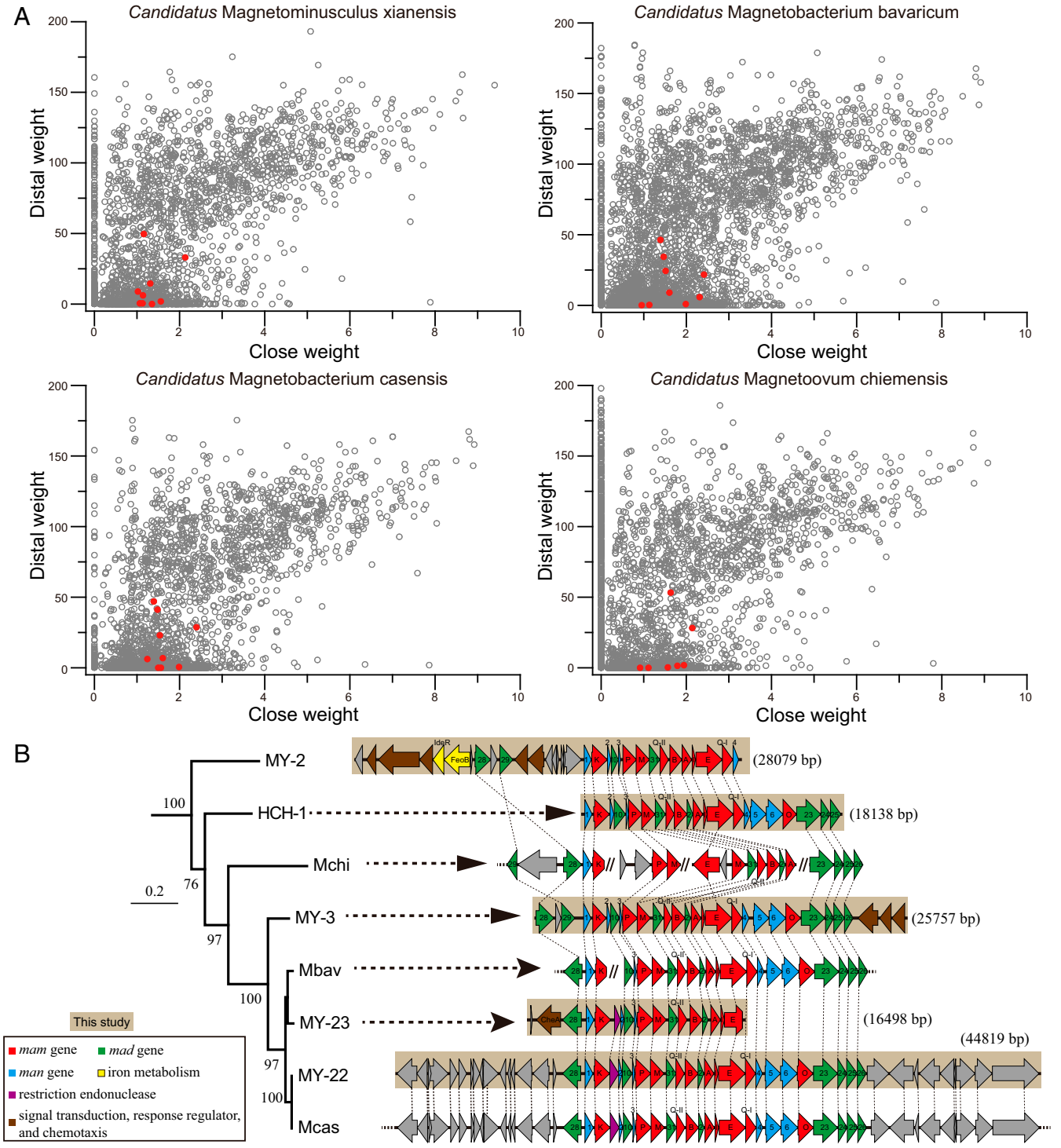

Fig. 2. (*A*) Implicit phylogenetic pattern of magnetosome core genes (*mamABEIKMOPQ*) of four *Nitrospirae* MTB in their genomic background. Each point represents a protein-coding gene in the genome, with magnetosome core genes highlighted in red. The *x* axis represents the sum of the relative bit scores of all hits within the phylum *Nitrospirae*, and the *y* axis represents that outside the phylum *Nitrospirae*. Typically, genes that were vertically transmitted have a moderate-to-high *x* value, representing its homologs in closely related taxonomic groups. Genes with a horizontal origin outside this phylum have a zero-to-low *x* value along with a moderate-to-high *y* value, representing homologs in distant taxonomic groups instead of close ones. Genes without traceable evolutionary history are located close to the origin, suggesting a lack of homologs in the database. (*B*) Maximum-likelihood tree based on a concatenation of MamABEP and schematic comparison of the *Nitrospirae* MGCs recovered from metagenomic data of MY with those of HCH-1, Mcas, Mbav, and Mchi.

shallow water, and that "oxygen oases" may have existed as far back as 2.8–3.2 Ga (40–42). These redox-stratified waters could also provide potential habitable environments for the origin and evolution of ancient MTB.

Typically, bacteria use a 3D "run-and-tumble" search strategy for finding their preferred microenvironments. For MTB, magnetotaxis reduces this 3D search to an optimized 1D search along geomagnetic field lines in chemically stratified water or sediment

columns (2). Previous observations that magnetofossil concentrations in marine sediments plummet during weak-field intervals surrounding geomagnetic reversals implies that magnetotaxis confers a selective advantage in field strengths of ~6 μT or greater (43). Recent analysis of reliable Archean paleointensity data suggests that field strengths of 20–50 μT are observed (44), indicating that the Archean geodynamo was sufficient to support magnetotaxis. An Archean origin of magnetotaxis, and its persistence in multiple lineages since their divergence during Archean time, implies both temporal continuity of Earth's dynamo and persistent environmental stratification.

The present study suggests an ancient origin of MTB in the Archean Eon that is much earlier than previously reported. MTB therefore represent one of the earliest magnetic-sensing and biomineralizing organisms on Earth. The Archean origin of MTB indicates that magnetotaxis is an evolutionarily ancient characteristic and that evolved in response to gradients in the Archean environment.

Materials and Methods

Site Description and MTB Sample Preparation. Surface sediments (depths of 5–10 cm) were sampled from the following two freshwater locations in China: the city moat of Xi'an in Shaanxi province (HCH) (34.25287, 108.92187), and Lake Miyun near Beijing (MY) (40.48874, 117.00714). The collected sediments were transferred to 600-mL plastic flasks, transported to the laboratory (in Beijing), and incubated at room temperature without disturbance. MTB cells from HCH were magnetically enriched using a double-ended open magnetic separation apparatus known as the "MTB trap" (45). The collected cells were washed twice and resuspended in sterile distilled water. MTB samples from MY used in this work were collected by Lin et al. (46) and had been stored at −80 °C. Genomic DNA was extracted from the enriched MTB by using the TIANamp Bacteria DNA Kit (Tiangen) following the manufacturer's instructions.

16S rRNA Gene Sequences Generation and Analysis. 16S rRNA gene sequences were amplified using the bacteria/archaeal universal primers 515F/806R that target the V4 region (47). The purified PCR products were cloned into the pMD19-T vector (TaKaRa) and chemically DH5a competent cells (Tiangen), following the manufacturers' instructions. Clones were randomly selected and were sequenced using an ABI 3730 genetic analyzer (Beijing Genomics Institute, Beijing, China). Sequences were processed, aligned, and clustered into operational taxonomic units (OTUs) using the QIIME pipeline (48).

Shotgun Metagenomic Sequencing and Data Analysis. The metagenomic DNA was amplified using multiple displacement amplification. Shotgun sequencing of metagenomic DNA was performed using Illumina HiSeq 2000 using the pair-end 125 × 125 library with a 600-bp inset size (Beijing Genomics Institute, Beijing, China). Metagenomic reads were assembled into contigs using Velvet, version 1.2.10, assembler (49). Resulting contigs were filtered by a minimal length cutoff of 1 kb. For details, see *SI Materials and Methods*.

Population Genome Binning of a Magnetotactic *Nitrospirae* from HCH. Because only a single population of *Nitrospirae* MTB with high relative abundance was identified in the metagenome of HCH, its population genome was recovered from the assemblies using a combination of similarity- [MEGAN, version 5 (50)] and composition-based [CLARK, version 1.1.2 (51)] approaches. The quality and accuracy of the acquired population genome were assessed using CheckM (52), using the lineage-specific workflow. For details, see *SI Materials and Methods*.

Implicit Phylogenomic Analysis of *Nitrospirae* MTB Genes. The global implicit phylogenetic pattern of the magnetotactic *Nitrospirae* genomes of HCH-1, Mcas, Mbav, and Mchi was inferred using HGTector 0.2.0 (53), a sequence similarity-based HGT prediction pipeline. For details, see *SI Materials and Methods*.

Identification of *Nitrospirae* MGC-Containing Contigs. To detect *Nitrospirae* MGC-containing contigs, magnetosome protein sequences from Mcas (17), Mchi (18), and Mbav (18) were used as queries in tBLASTn analyses against the assembled contigs of each sample. All matches (E value ≤ 1e-5) were then manually verified.

Phylogenetic Analysis of Magnetosome Proteins. Magnetosome gene orthologs between the complete MGC of Mcas and MGCs of representative MTB populations including HCH-1, MSR-1, AMB-1, QH-2, MV-1, MC-1, IT-1, BW-1, MMP, and RS-1 were calculated by bidirectional best-hit analysis through the SEED viewer (54). The amino acid sequences of magnetosome proteins were aligned by MUSCLE algorithms (55) using MEGA, version 6.06 (56), and poorly conserved regions were trimmed by using the Gblocks method (57). Appropriate protein models of substitution were selected using the Find Best DNA/Protein Models implemented in MEGA, version 6.06 (56), and maximum-likelihood phylogenetic trees were constructed using MEGA, version 6.06, with a GTR and Gamma model (56) for 16S rRNA genes and RAxML, version 8.0.19, with a LG and Gamma model (58) for magnetosome proteins. Phylogenomic tree construction was performed using PhyloPhlAn (59), which uses USEARCH (60) and MUSCLE (55) to extract up to 400 conserved ubiquitous proteins (Table S3) coded in genomes and perform individual protein alignments. The universally conserved and phylogenetically discriminative positions in each protein alignment were then concatenated into a single long sequence through PhyloPhlAn. PhyML, version 3.0 (initial tree: BioNJ; tree topology search: NNIs) (61), was used to generate a maximum-likelihood tree. Confidence in phylogenetic results was assessed using the 100 bootstrap resampling approaches.

Sequence Divergence. The average number of synonymous substitutions per synonymous site (dS) of the five magnetosome genes and three housekeeping genes (*recA*, *gyrB*, and *pyrH*) were calculated using MEGA, version 6.06 (56). The effective number of codons (N_c) was calculated with CodonW tool (version 1.4.4 at codonw.sourceforge.net/) (62). N_c varies between 21 for maximum codon bias and 61 for minimum codon bias. MTB involved in this analysis were AMB-1, MSR-1, MC-1, QH-2, RS-1, MMP, Mbav, Mcas, Mchi, and HCH-1.

Divergence Time Estimation. Sequence data from 64 genomes were used to estimate the divergence time between phyla *Nitrospirae* and *Proteobacteria*. A subset of up to 400 core proteins was extracted and aligned using PhyloPhlAn (59). Maximum likelihood tree and bootstrap values were performed using RAxML, version 8.0.19, with a VT and Gamma model (58). Divergence times were estimated using PhyloBayes, version 4.1c (63). For details, see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Longfei Wu for valuable comments and suggestions. W.L. and Y.P. acknowledge financial support from National Natural Science Foundation of China (NSFC) Grants 41330104, 41621004, and 41374074. W.L. acknowledges support from the Youth Innovation Promotion Association of the Chinese Academy of Sciences. G.A.P. acknowledges funding from NSFC Grants 41374072 and 41574063. D.A.B. is supported by US National Science Foundation Grant EAR-1423939. J.L.K. is supported by US National Aeronautics and Space Administration Exobiology Grant EXO14_2-0176.

1. Bazylinski D, Lefèvre C, Schüler D (2013) Magnetotactic bacteria. *The Prokaryotes*, eds Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F (Springer, Berlin), pp 453–494.
2. Frankel RB, Bazylinski DA, Johnson MS, Taylor BL (1997) Magneto-aerotaxis in marine coccoid bacteria. *Biophys J* 73(2):994–1000.
3. Lin W, Bazylinski DA, Xiao T, Wu L-F, Pan Y (2014) Life with compass: Diversity and biogeography of magnetotactic bacteria. *Environ Microbiol* 16(9):2646–2658.
4. Chang SBR, Kirschvink JL (1989) Magnetofossils, the magnetization of sediments, and the evolution of magnetite biomineralization. *Annu Rev Earth Planet Sci* 17:169–195.
5. Roberts AP, et al. (2013) Magnetic properties of pelagic marine carbonates. *Earth Sci Rev* 127:111–139.
6. Kopp RE, Kirschvink JL (2008) The identification and biogeochemical interpretation of fossil magnetotactic bacteria. *Earth Sci Rev* 86(1-4):42–61.
7. Pan Y, et al. (2005) The detection of bacterial magnetite in recent sediments of Lake Chiemsee (southern Germany). *Earth Planet Sci Lett* 232(1-2):109–123.
8. Grünberg K, Wawer C, Tebo BM, Schüler D (2001) A large gene cluster encoding several magnetosome proteins is conserved in different species of magnetotactic bacteria. *Appl Environ Microbiol* 67(10):4573–4582.
9. Murat D, Quinlan A, Vali H, Komeili A (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc Natl Acad Sci USA* 107(12):5593–5598.
10. Lohsse A, et al. (2011) Functional analysis of the magnetosome island in *Magnetospirillum gryphiswaldense*: The *mamAB* operon is sufficient for magnetite biomineralization. *PLoS One* 6(10):e25561.
11. Delong EF, Frankel RB, Bazylinski DA (1993) Multiple evolutionary origins of magnetotaxis in bacteria. *Science* 259(5096):803–806.

12. Jogler C, Schüler D (2009) Genomics, genetics, and cell biology of magnetosome formation. *Annu Rev Microbiol* 63(1):501–521.
13. Lefèvre CT, et al. (2013) Monophyletic origin of magnetotaxis and the first magnetosomes. *Environ Microbiol* 15(8):2267–2274.
14. Zeytuni N, et al. (2015) MamA as a model protein for structure-based insight into the evolutionary origins of magnetotactic bacteria. *PLoS One* 10(6):e0130394.
15. Jogler C, et al. (2011) Conservation of proteobacterial magnetosome genes and structures in an uncultivated member of the deep-branching *Nitrospira* phylum. *Proc Natl Acad Sci USA* 108(3):1134–1139.
16. Spring S, et al. (1993) Dominating role of an unusual magnetotactic bacterium in the microaerobic zone of a freshwater sediment. *Appl Environ Microbiol* 59(8):2397–2403.
17. Lin W, et al. (2014) Genomic insights into the uncultured genus "*Candidatus* Magnetobacterium" in the phylum *Nitrospirae*. *ISME J* 8(12):2463–2477.
18. Kolinko S, Richter M, Glöckner FO, Brachmann A, Schüler D (2016) Single-cell genomics of uncultivated deep-branching magnetotactic bacteria reveals a conserved set of magnetosome genes. *Environ Microbiol* 18(1):21–37.
19. Evans PN, et al. (2015) Methane metabolism in the archaeal phylum *Bathyarchaeota* revealed by genome-centric metagenomics. *Science* 350(6259):434–438.
20. Brown CT, et al. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523(7559):208–211.
21. Friedman R, Ely B (2012) Codon usage methods for horizontal gene transfer detection generate an abundance of false positive and false negative results. *Curr Microbiol* 65(5):639–642.
22. Wallau GL, Ortiz MF, Loreto EL (2012) Horizontal transposon transfer in Eukarya: Detection, bias, and perspectives. *Genome Biol Evol* 4(8):689–699.
23. Sheridan PP, Freeman KH, Brenchley JE (2003) Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiol J* 20(1):1–14.
24. Brocks JJ, Summons RE (2003) Sedimentary hydrocarbons, biomarkers for early life. *Treatise on Geochemistry*, eds Holland HD, Turekian KK (Pergamon, Oxford), pp 63–115.
25. French KL, et al. (2015) Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc Natl Acad Sci USA* 112(19):5915–5920.
26. David LA, Alm EJ (2011) Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* 469(7328):93–96.
27. Canfield DE (2005) The early history of atmospheric oxygen: Homage to Robert M. Garrels. *Annu Rev Earth Planet Sci* 33(1):1–36.
28. Habicht KS, Gade M, Thamdrup B, Berg P, Canfield DE (2002) Calibration of sulfate levels in the archean ocean. *Science* 298(5602):2372–2374.
29. Canfield DE, Rosing MT, Bjerrum C (2006) Early anaerobic metabolisms. *Philos Trans R Soc Lond B Biol Sci* 361(1474):1819–1834, discussion 1835–1836.
30. Buick R (2012) Geobiology of the Archean Eon. *Fundamentals of Geobiology*, eds Knoll AH, Canfield DE, Konhauser KO (Wiley, Chichester, UK), pp 351–370.
31. Williams TJ, Zhang CL, Scott JH, Bazylinski DA (2006) Evidence for autotrophy via the reverse tricarboxylic acid cycle in the marine magnetotactic coccus strain MC-1. *Appl Environ Microbiol* 72(2):1322–1329.
32. Ji B, et al. (2014) Comparative genomic analysis provides insights into the evolution and niche adaptation of marine *Magnetospira* sp. QH-2 strain. *Environ Microbiol* 16(2):525–544.
33. Schübbe S, et al. (2009) Complete genome sequence of the chemolithoautotrophic marine magnetotactic coccus strain MC-1. *Appl Environ Microbiol* 75(14):4835–4852.
34. Blake RE, Chang SJ, Lepland A (2010) Phosphate oxygen isotopic evidence for a temperate and biologically active Archaean ocean. *Nature* 464(7291):1029–1032.
35. Knauth LP (2005) Temperature and salinity history of the Precambrian ocean: Implications for the course of microbial evolution. *Palaeogeogr Palaeoclimatol Palaeoecol* 219(1-2):53–69.
36. Bazylinski DA, Lefèvre CT (2013) Magnetotactic bacteria from extreme environments. *Life (Basel)* 3(2):295–307.
37. Spring S, Bazylinski DA (2006) Magnetotactic bacteria. *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, ed Dworkin M (Springer, New York), pp 842–862.
38. Kappler A, Pasquero C, Konhauser KO, Newman DK (2005) Deposition of banded iron formations by anoxygenic phototrophic Fe(II)-oxidizing bacteria. *Geology* 33(11):865–868.
39. Walker JCG (1987) Was the Archaean biosphere upside down? *Nature* 329(6141):710–712.
40. Planavsky NJ, et al. (2014) Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nat Geosci* 7(4):283–286.
41. Riding R, Fralick P, Liang L (2014) Identification of an Archean marine oxygen oasis. *Precambrian Res* 251(9):232–237.
42. Satkoski AM, Beukes NJ, Li W, Beard BL, Johnson CM (2015) A redox-stratified ocean 3.2 billion years ago. *Earth Planet Sci Lett* 430:43–53.
43. Kirschvink JL (1982) Paleomagnetic evidence for fossil biogenic magnetite in western Crete. *Earth Planet Sci Lett* 59(2):388–392.
44. Biggin AJ, et al. (2015) Palaeomagnetic field intensity variations suggest Mesoproterozoic inner-core nucleation. *Nature* 526(7572):245–248.
45. Jogler C, et al. (2009) Toward cloning of the magnetotactic metagenome: Identification of magnetosome island gene clusters in uncultivated magnetotactic bacteria from different aquatic sediments. *Appl Environ Microbiol* 75(12):3972–3979.
46. Lin W, Jogler C, Schüler D, Pan Y (2011) Metagenomic analysis reveals unexpected subgenomic diversity of magnetotactic bacteria within the phylum *Nitrospirae*. *Appl Environ Microbiol* 77(1):323–326.
47. Caporaso JG, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6(8):1621–1624.
48. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336.
49. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
50. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17(3):377–386.
51. Ounit R, Wanamaker S, Close TJ, Lonardi S (2015) CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* 16:236.
52. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7):1043–1055.
53. Zhu Q, Kosoy M, Dittmar K (2014) HGTector: An automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics* 15(1):717.
54. Overbeek R, et al. (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42(Database issue, D1):D206–D214.
55. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
56. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729.
57. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540–552.
58. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
59. Segata N, Börnigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304.
60. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.
61. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
62. Peden JF (1999) Analysis of codon usage. PhD dissertation (University of Nottingham, Nottingham, UK).
63. Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25(17):2286–2288.
64. Surget-Groba Y, Montoya-Burgos JI (2010) Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* 20(10):1432–1440.
65. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
66. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12(1):59–60.
67. Bekker A, et al. (2004) Dating the rise of atmospheric oxygen. *Nature* 427(6970):117–120.
68. Crowe SA, et al. (2013) Atmospheric oxygenation three billion years ago. *Nature* 501(7468):535–538.
69. Hofmann HJ (1976) Precambrian microflora, Belcher Islands, Canada; significance and systematics. *J Paleontol* 50(6):1040–1073.
70. Shih PM, Hemp J, Ward LM, Matzke NJ, Fischer WW (2017) Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* 15(1):19–29.
71. Blank CE, Sánchez-Baracaldo P (2010) Timing of morphological and ecological innovations in the cyanobacteria—a key to understanding the rise in atmospheric oxygen. *Geobiology* 8(1):1–23.

Supporting Information

Lin et al. 10.1073/pnas.1614654114

SI Materials and Methods

Shotgun Metagenomic Sequencing and Data Analysis. To obtain sufficient DNA for shotgun metagenomic sequencing, multiple displacement amplification was performed using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare) following the manufacturer's instructions. Briefly, 1 μL of DNA was used as the template and was mixed with 9 μL of sample buffer. The mixed DNA was heated at 95 °C for 3 min and cooled to 4 °C, before incubation at 30 °C for 90 min with 1 μL of enzyme mixture and 9 μL of reaction buffer. To terminate the reaction, the sample was heated at 65 °C for 10 min. For each sample, nine amplifications were pooled to reduce potential bias. These were purified using TIANquik Maxi Purification Kit (Tiangen).

Shotgun sequencing of metagenomic DNA was performed using Illumina HiSeq 2000 using the pair-end 125 × 125 library with a 600-bp inset size (Beijing Genomics Institute, Beijing, China). The entire dataset of two samples is ∼5.55 Gb. Illumina reads were trimmed to remove the adapter sequences and low-quality bases, after which 86–88% of paired reads were retained for each sample. Trimmed, paired-end reads were assembled using a multiple *k*-mer–based assemblies (64). Briefly, metagenomic reads of each sample were individually assembled into contigs using the Velvet, version 1.2.10, assembler (49) with a range of *k*-mers (41, 51, 61, 71, 81, and 91). The different assembles were subsequently merged, and the duplicated and suboptimal contigs were removed through CD-HIT-EST (65) using a sequence identity threshold of 0.95 and a word length of 8 to get the final assembly for each sample. Resulting contigs were filtered by a minimal length cutoff of 1 kb.

Population Genome Binning of a Magnetotactic *Nitrospirae* from HCH. Contigs of sample HCH were sorted using BLASTn alignment against the NCBI genomes database (version May 2015) together with previously sequenced MTB draft genomes of Mcas (17), Mchi (18), and Mbav (18). BLASTn alignment hits with *E* values larger than 1×10^{-5} were filtered, and the taxonomical level of each contig was determined by the lowest common ancestor algorithm implemented in MEGAN, version 5 (50). All contigs binned to known *Nitrospirae* MTB species of Mcas, Mbav, and Mchi were selected. Due to the incomplete nature of available magnetotactic *Nitrospirae* draft genomes, the remaining contigs were further classified using CLARK, version 1.1.2 (51), based on reduced sets of *k*-mers by comparison with available genomes or draft genomes of MTB strains. The measure of conservation of gene content and gene order of MGCs between HCH-1 and three available *Nitrospirae* MTB (Mcas, Mbav, and Mchi) is the ratio between the number of genes located in conserved content and order and the total number of bidirectional best-hits genes between MGCs of HCH-1 and three *Nitrospirae* MTB.

Implicit Phylogenomic Analysis of *Nitrospirae* MTB Genes. The global implicit phylogenetic pattern of the magnetotactic *Nitrospirae* genomes of HCH-1, Mcas, Mbav, and Mchi was inferred using HGTector 0.2.0 (53). Protein sequence similarity search was performed using DIAMOND 0.9.7 (66) against a database (generated by HGTector) that contains one representative per species from all available nonredundant RefSeq prokaryotic proteomes (October 2015), plus the MTB proteomes reconstructed in this study. Quality cutoffs for valid hits were *E* value ≤ 1e-20, percentage identity ≥ 30%, and query coverage ≥ 50%. For each protein-coding gene, the top 250 highest-scoring hits from different species were retained. For each hit, a "relative bit score" was calculated as the original bit score of the hit divided by the bit score of the query sequence aligned against itself. The overall distribution pattern of all genes in a genome was visualized by plotting the sum of the bit scores of hits within phylum *Nitrospirae* against that outside this phylum per gene.

Divergence Time Estimation. Molecular-dating analyses were performed using PhyloBayes, version 4.1c (63). The CAT-GTR model was implemented for amino acid replacement, and analyses were run under either the log-normal autocorrelated relaxed clock (-ln) or the uncorrelated gamma multipliers (-ugam). For each condition, two replicate chains with 20,000 generations were run. Dates were assessed by running the readdiv with the first 20% of generations removed as burn-in for each analysis. Two different combinations of age constraints were used for the divergence time estimation. For the first combination of age constraints, the minimum age of the root of Oxyphotobacteria (oxygenic Cyanobacteria) was set at 2.32 Ga (the rise in atmospheric oxygen) (67), and the maximum age was set at 3.0 Ga (40, 68). For the second combination, a minimum age of 1.9 Ga (the first widely accepted fossil oxygenic Cyanobacteria) (69) and a maximum age of 2.32 Ga (postdating the rise of oxygen according to ref. 70) were implemented as the oxyphotobacterial root. In addition, for the second combination another time constraint, the divergence time between Oxyphotobacteria and Melainabacteria, was included, which was set from 2.5 Ga (70) to 3.8 Ga (the end of late heavy bombardment). For all analyses, the age calibration for the last common ancestor of all taxa used in this study was set between 2.32 and 3.8 Ga (71).

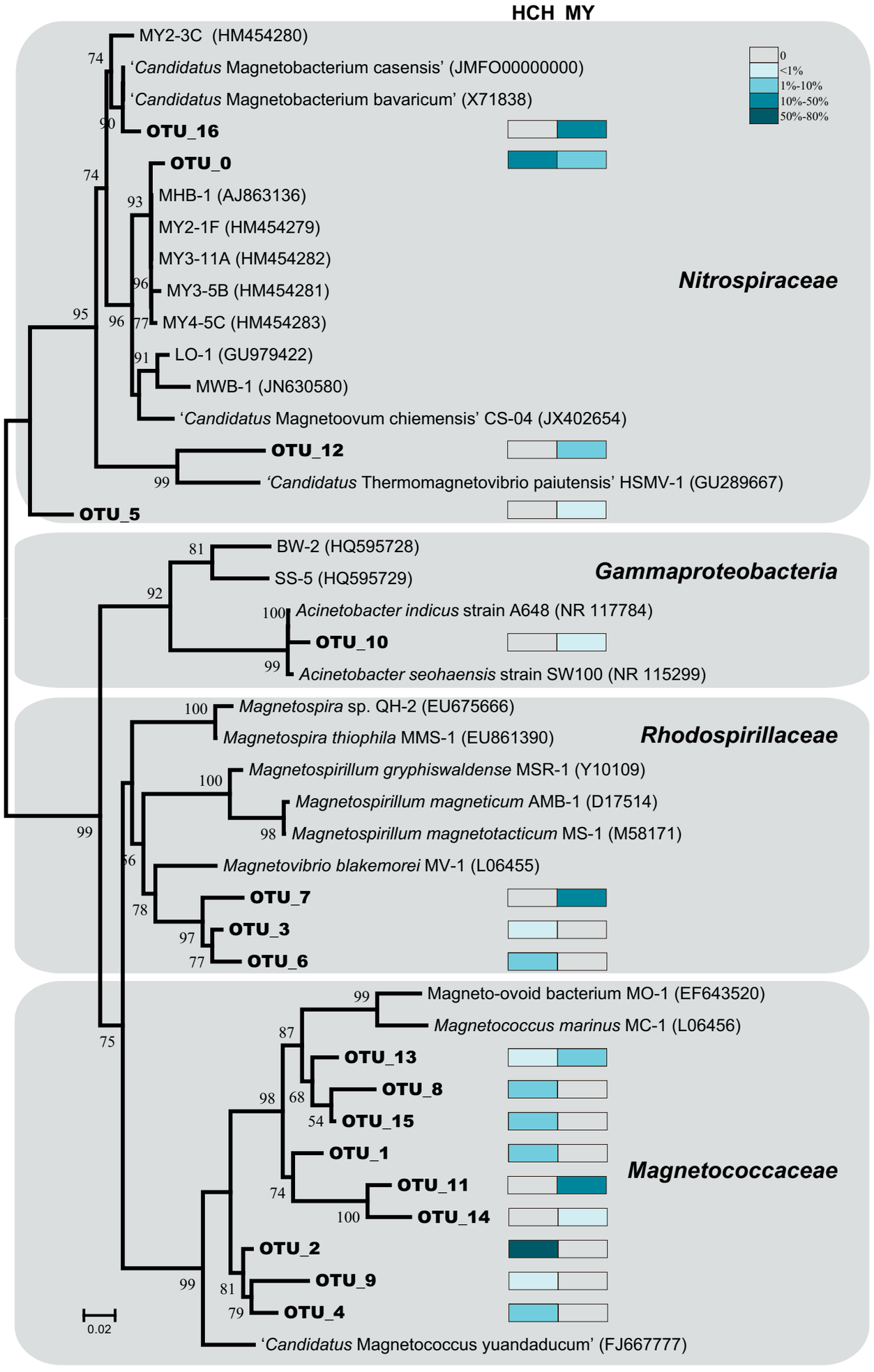

Fig. S1. Phylogenetic tree of operational taxonomic units (OTUs at 97% threshold similarity) for 16S rRNA gene clone libraries of MTB communities from the city moat of Xi'an in Shaanxi province (HCH) and Lake Miyun near Beijing (MY). The evolutionary history was inferred by using the maximum-likelihood method based on the Kimura two-parameter model with 100 bootstraps. On the right-hand side, a heatmap shows the relative abundance and distribution of each OTU from this study.

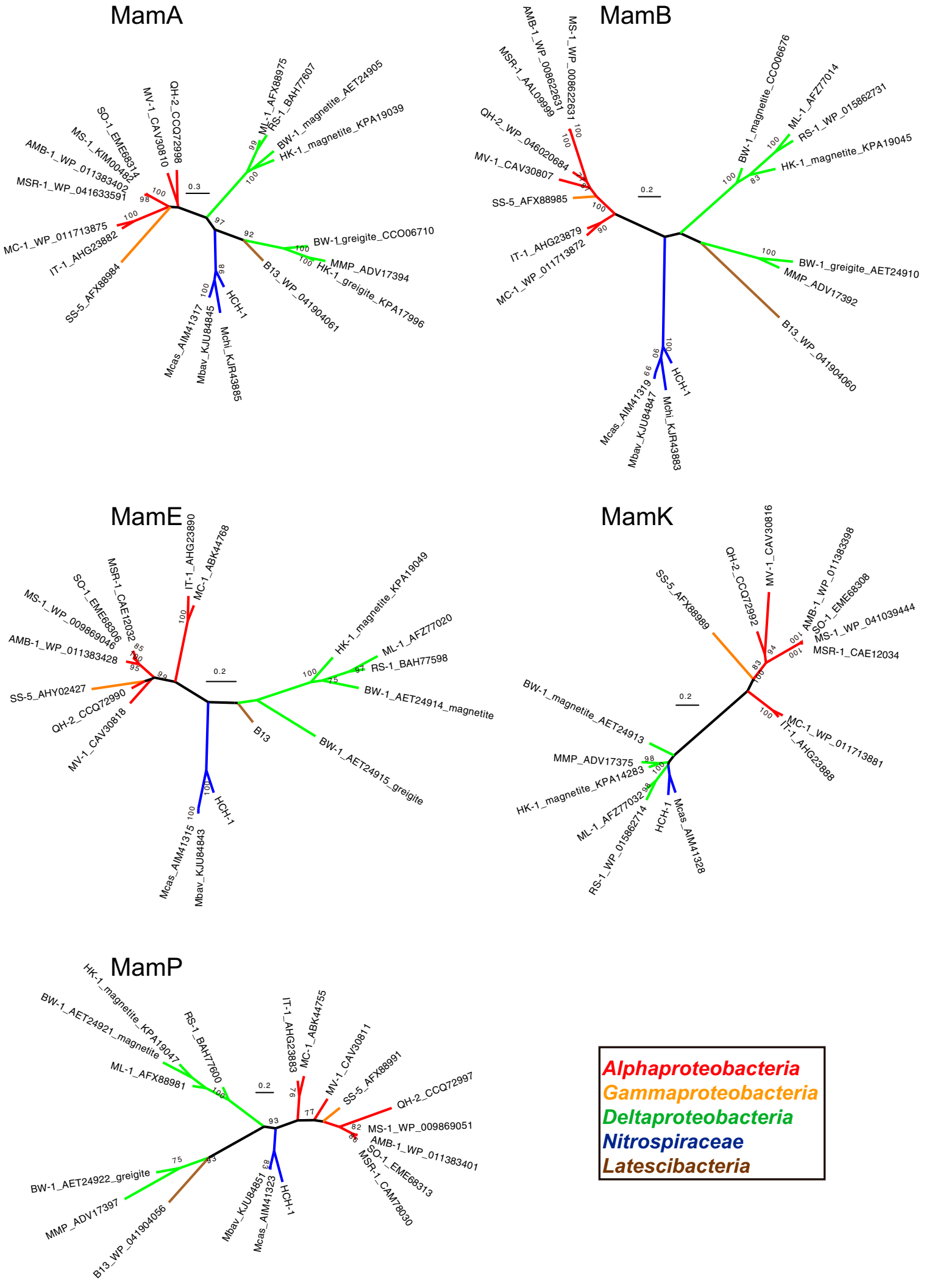

Fig. S2. Bootstrap consensus trees of five magnetosome proteins (MamABEKP) based on the maximum-likelihood method. Only full-length protein sequences were included in this analysis. Bootstrap values are expressed as percentages, and only values of more than 75% are shown. MSR-1, *Magnetospirillum gryphiswaldense* MSR-1; AMB-1, *Magnetospirillum magneticum* AMB-1; SO-1, *Magnetospirillum* sp. SO-1; QH-2, *Magnetospira* sp. QH-2; MC-1, *Magnetococcus marinus* MC-1; BW-1, *Candidatus* Desulfamplus magnetomortis BW-1; MMP, *Ca*. Magnetoglobus multicellularis; RS-1, *Desulfovibrio magneticus* RS-1; ML-1, alkaliphilic magnetotactic strain ML-1; MV-1, *Magnetovibrio blakemorei* MV-1; IT-1, *Ca.* Magnetofaba australis strain IT-1; SS-5, Gammaproteobacteria magnetotactic strain SS-5; HK-1, *Ca*. Magnetomorum sp. HK-1; B13, *Latescibacteria bacterium* SCGC AAA252-B13.

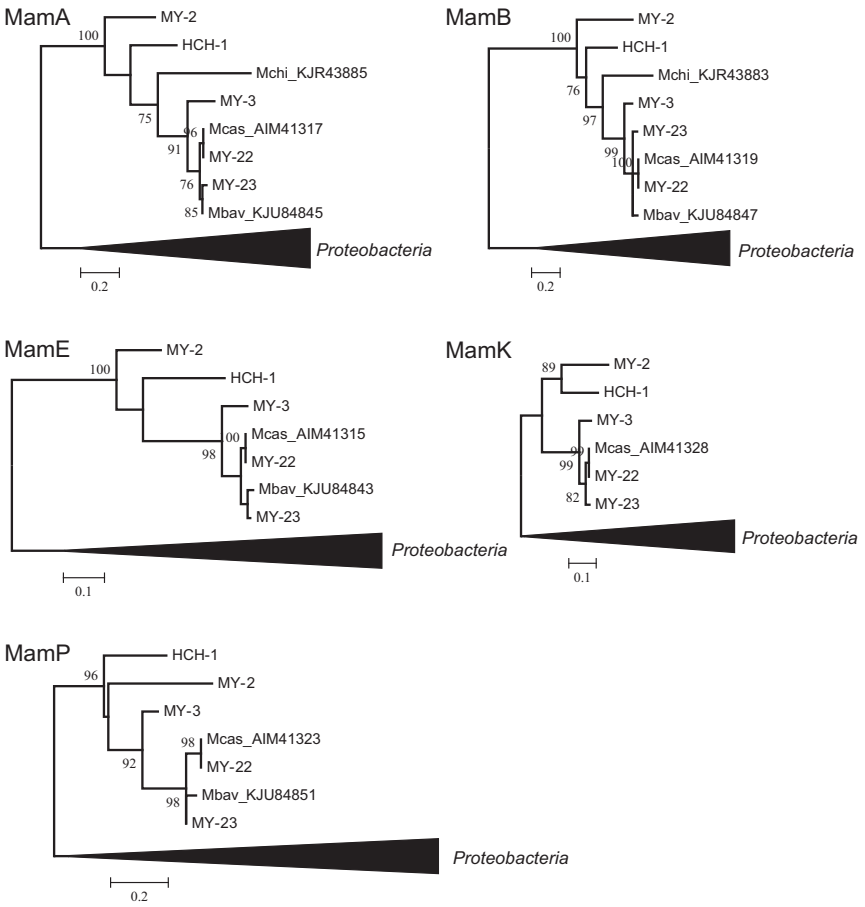

Fig. S3. Bootstrap consensus trees of magnetosome proteins MamABEKP from the four additional *Nitrospirae* MGCs of MY and those full-length proteins from all available *Nitrospirae* MTB based on the maximum-likelihood method. Bootstrap values are expressed as percentages, and only values of more than 75% are shown.

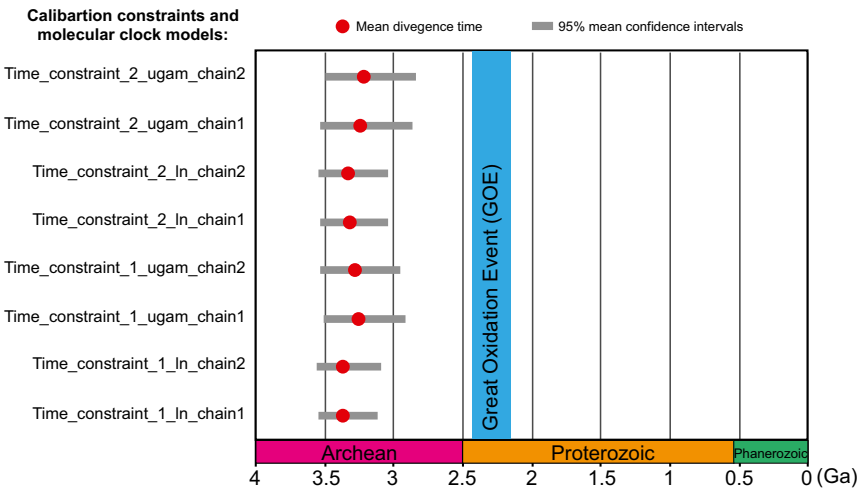

Fig. S4. Summary of mean divergence dates for the *Nitrospirae* and *Proteobacteria* phyla estimated using Bayesian relaxed molecular-clock analyses with two different time constraints and two different molecular clock models (see *SI Materials and Methods* for details). Two replicated chains were run for each condition. The input phylogenomic tree used here is shown in Fig. S5.

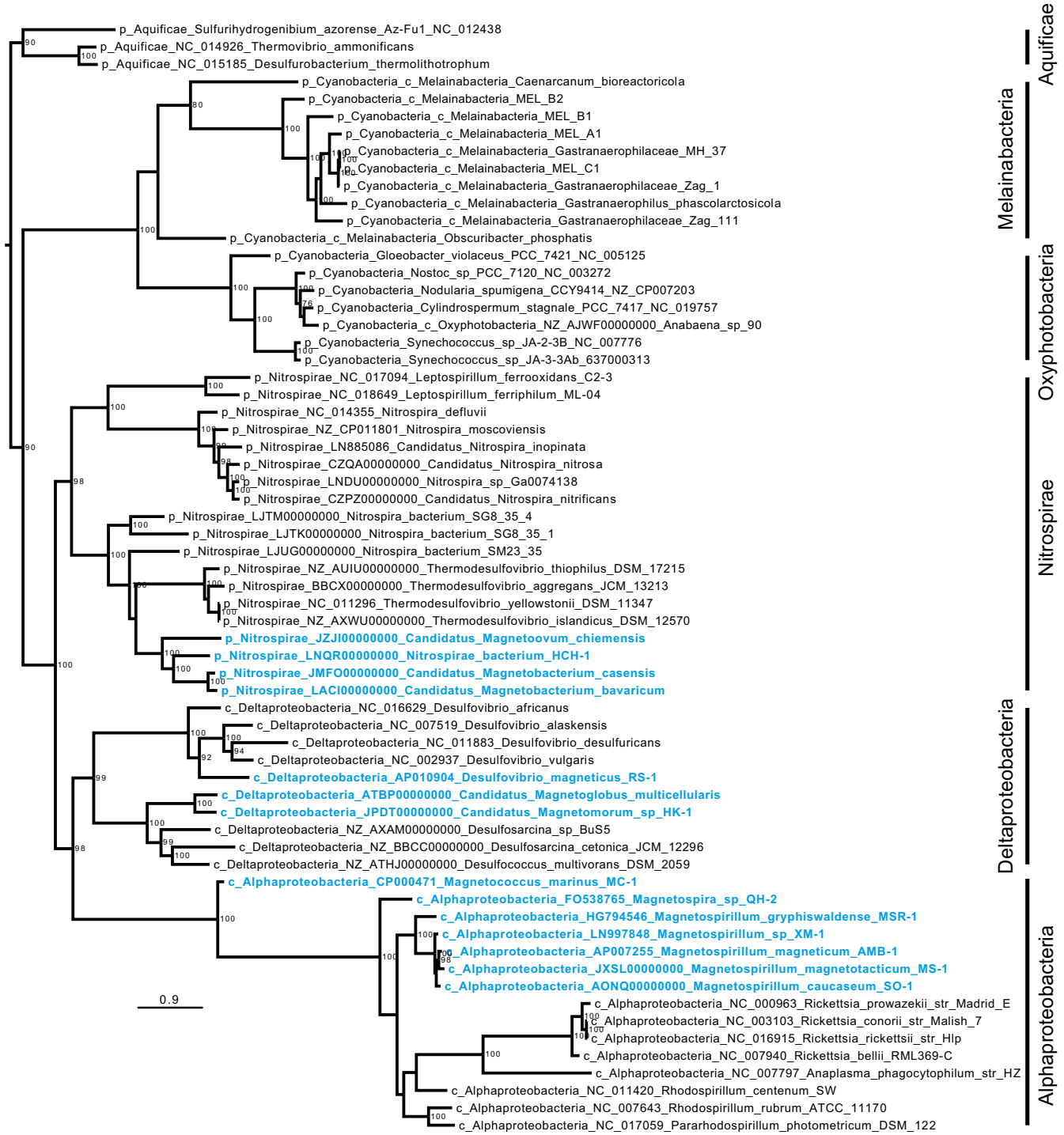

Fig. S5. Phylogenomic maximum-likelihood tree of 64 bacterial genomes. Bootstrap values are expressed as percentages, and only values of >75% are shown. Magnetotactic bacteria are displayed in blue. The Aquificae strains were used as outgroup.

Other Supporting Information Files

Table S1 (DOCX)
Table S2 (DOCX)
Table S3 (DOCX)