Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean

Yong Wang,1* Jiao-Mei Huang,1,2 Guo-Jie Cui,1,2 Takuro Nunoura,3 Yoshihiro Takaki,3,4 Wen-Li Li,1,2 Jun Li,1 Zhao-Ming Gao,1 Ken Takai,4 Ai-Qun Zhang1 and Ramunas Stepanauskas5

1Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan, China.
2University of the Chinese Academy of Sciences, Beijing, China.
3Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, 237-0061, Japan.
4Department of Subsurface Geobiological Analysis and Research, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, 237-0061, Japan.
5Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME, 04544, USA.

Summary
Various lineages of ammonia-oxidizing archaea (AOA) are present in deep waters, but the mechanisms that determine ecotype formation are obscure. We studied 18 high-quality genomes of the marine group I AOA lineages (alpha, gamma and delta) from the Mariana and Ogasawara trenches. The genomes of alpha AOA resembled each other, while those of gamma and delta lineages were more divergent and had even undergone insertion of some phage genes. The instability of the gamma and delta AOA genomes could be partially due to the loss of DNA polymerase B (polB) and methyladenine DNA glycosylase (tag) genes responsible for the repair of point mutations. The alpha AOA genomes harbour genes encoding a thrombospondin-like outer membrane structure that probably serves as a barrier to gene flow. Moreover, the gamma and alpha AOA lineages rely on vitamin B12-independent MetE and B12-dependent MetH, respectively, for methionine synthesis. The delta AOA genome contains genes involved in uptake of sugar and peptide perhaps for heterotrophic lifestyle. Our study provides insights into co-occurrence of cladogenesis and anagenesis in the formation of AOA ecotypes that perform differently in nitrogen and carbon cycling in dark oceans.

Introduction
Nutritional deprivation in deep-sea environments may limit the productivity of an ecosystem. In addition to recalcitrant materials recycled by deep-sea bacteria, organic carbon produced by CO2 fixation can be supplied by ammonia-oxidizing and nitrite-oxidizing prokaryotes (Pachiaidaki et al., 2017). Ammonia-oxidizing archaea (AOA) are ubiquitous in oxic environments on Earth ranging from shallow waters to deep-sea zones and soils (Karner et al., 2001; Leininger et al., 2006; Park et al., 2008). They outcompete ammonia-oxidizing bacteria (AOB) due to their higher affinity for ammonia (Martens-Habbena et al., 2009), which probably permits fitness and wider distribution of AOA in deep oceans compared with AOB (Wang et al., 2017). As a major autotrophic inhabitant of deep oceans, AOA represented by Marine Group I (MGI) Thaumarchaeota are critical for maintaining both the carbon and nitrogen cycles in the water column and in sediments of the full-ocean depth (Kato et al., 1997; Stahl and de la Torre, 2012; Nunoura et al., 2018). Initially assigned to MGI Crenarchaeota, MGI AOA were shown to include at least three lineages: alpha, beta and gamma (Massana et al., 2000). Now affiliated with a new phylum of Thaumarchaeota, at least five lineages (with epsilon and delta introduced as new lineages) of water column AOA have been shown to be distributed at different depths in different environments (Takai et al., 2004; Nunoura et al., 2015). The distribution depth of the beta lineage is less than 500 m, that of the delta lineage is primarily between 500 m and 2000 m, and that of the alpha lineage is largely restricted to >6000 m; the gamma lineage is ubiquitous in the full-ocean depths (Nunoura et al., 2015). The environmental factors and genomic features that determine the depth stratification pattern of these lineages are presently unknown. Complete genomes of MGI AOA were obtained for Nitrosopumilus maritimus and

Received 29 August, 2018; revised 18 December, 2018; accepted 26 December, 2018. *For correspondence. E-mail wangy@idsse.ac.cn; Tel. (86) 898-88381062; Fax (86) 898-8822506.
‘Candidatus Nitrosopelagicus brevis’ (Walker et al., 2010; Santoro et al., 2015), and this paved the way to understanding the metabolism of the model marine MGI AOA species and its lifestyle of autotrophic growth. Using a transmembrane hydroxylamine: ubiquinone reductase module (HURM) complex, AMO and plastocyanin, N. maritimus is able to generate NADH and oxidize ammonia to nitrite. During this process, ammonia serves as an electron donor to generate reducing force for CO₂ fixation by the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle (Walker et al., 2010; Santoro et al., 2015). Ammonia concentration/flux seems to be the limiting factor for the survival of AOA in deep oceans and may also contribute to the diversification of their lineages (Sintes et al., 2013). Nevertheless, alternative sources of ammonia, including urea and glycine, have been suggested by recent genomics studies (León-Zayas et al., 2015; Kerou et al., 2016). In the mesopelagic zone of the Mariana Trench, a considerable percentage of the microbial community is composed of the MGI AOA (Nunoura et al., 2015), but current data provide only a glimpse of the genetic disparities present in the AOA (León-Zayas et al., 2015; Kerou et al., 2016). Without high-quality genomes, the underlying mechanisms that resulted in the divergence of the MGI AOA lineages cannot be completely understood. The MGI AOA lineages might be ecologically distinct populations or ecotypes, but the differences in their metabolic potential have not yet been studied. According to the theory of microbial speciation, new ecotypes are formed by cladogenesis (splitting of a population into ecologically distinct populations) and by anagenesis (adaptive improvements of a single population) (Koeppel et al., 2013). Whether the five AOA lineages evolved into various ecotypes through cladogenesis and/or anagenesis is still a question. Some clues may be provided by comparative genomic analysis of AOA lineages.

In this study, we obtained full-ocean water samples from the Mariana and Ogasawara trenches and used them to investigate the vertical distribution and genomics of AOA lineages. Metagenome assembled genomes (MAGs) and single amplified genomes (SAGs) for three AOA lineages (alpha, gamma and delta) were used in a pan-genomic study, the results of which provided hints regarding the mechanisms that have led to the emergence of ecotypes in the lineages.

## Results

**Ubiquitous ammonia-oxidizing archaea in the Mariana trench**

Water samples from the full-ocean depth were collected from two trenches in the western Pacific Ocean during four cruises (Supporting Information Fig. S1). A total of 75 water samples were obtained from the Mariana Trench, and four were obtained from the Ogasawara Trench (Supporting Information Table S1). Ammonia concentration, ion concentrations and other environmental factors in the waters overlying the Challenger Deep and the Ogasawara Trench have been measured previously (Nunoura et al., 2015; Kawagucci et al., 2018). Tag sequencing of the 16S rRNA gene for the Mariana samples showed that Thaumarchaeota, represented by MGI AOA, was one of the dominant phyla in the communities, as reported previously (Nunoura et al., 2015). Classification of operational taxonomic units (OTUs) at 97% similarity identified the four most abundant MGI AOA OTUs in the Mariana waters: OTU272811, OTU31185, OTU294048 and OTU53278. At depths of 500–6500 m, these four OTUs accounted for up to 17% of the retrieved 16S rRNA genes (Supporting Information Fig. S2). The total percentage of the AOA decreased in the hadal zone >6500 m, where Marinimicrobia (SAR406) dominated the microbial community based on the sequencing of 16S rRNA gene amplicons. From 500 m depth to the bottom of the Challenger Deep, the relative abundance of AOA OTU272811 correlated negatively with that of OTU31185. AOA OTU294048 and OTU53278 seem to be an abyssal lineage. Regarding diversity within the AOA OTUs, further grouping of the tag sequencing reads at 1% dissimilarity revealed that the highest diversity exists within OTU53278 (Supporting Information Table S2). For the subtypes in OTU272811, they were more diversified in upper layers between 500 m and 2000 m (Supporting Information Table S2).

To determine the stratified distribution of the AOA lineages, we obtained 14 metagenomes for waters from the Mariana Trench (Supporting Information Table S3) and then binned 14 MAGs for the MGI AOA. Three metagenome assemblies allowed genome binning of two lineages (Table 1 and Supporting Information Fig. S3). A total of 18 SAGs for MGI AOA were identified among the SAGs from the Ogasawara Trench (Supporting Information Table S4). The complete isolate genomes of N. maritimus and ‘Ca. N. brevis’ contain 85 conserved single-copy genes (CSCGs) that are universally maintained in archaeal genomes (Supporting Information Table S5). Eight of the AOA MAGs also possess the same number of CSCGs (Table 1). Two AOA MAGs harboured potentially contaminant contigs and were less than 80% complete. The AOA SAGs contain at most 77 unique CSCGs, resulting in a completeness of 91.8%.

**Major lineages of AOA in the western Pacific Ocean**

The 16S rRNA genes were extracted from the MAGs, SAGs and the representative amplicon reads of the four OTUs and used to infer phylogenetic relationships with reference sequences from the five known MGI AOA
In the phylogenetic tree, the sample from the Mariana 1000-m depth was grouped with delta MGI AOA; this MAG was abundant in the D11 sample from the Mariana 1000-m depth. Using the 16S rRNA genes, only one of our MAGs was detected as SAGs affiliated with the gamma lineage were obtained, as suggested by the diversity of the gamma AOA obtained from the two trenches. The ANIs between the genomes of the MGI AOA lineage coincided with the similarity between homologous regions of the genomes (Supporting Information Fig. S3). The alpha AOA MAG ma7 and the gamma AOA MAG mg3 contain some lineage-specific regions of the genomes (Supporting Information Table S3 for information on the Mariana Trench (see Supporting Information Table S3 for information on sampling sites).

### Table 1. Summary of AOA MAGs and SAGs.

<table>
<thead>
<tr>
<th>Genome/sample</th>
<th>Depth (m)</th>
<th>AOA type</th>
<th>Name</th>
<th>Genome size (bp)</th>
<th>No. of contigs</th>
<th>No. of CDSs</th>
<th>Total CSCGs</th>
<th>Unique CSCGs</th>
<th>Compl. (%)</th>
<th>Contam. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5C13</td>
<td>505</td>
<td>γ</td>
<td>og1</td>
<td>949 320</td>
<td>27</td>
<td>1119</td>
<td>77</td>
<td>76</td>
<td>90.6</td>
<td>0</td>
</tr>
<tr>
<td>5D11</td>
<td>1000</td>
<td>δ</td>
<td>md1</td>
<td>1 279 113</td>
<td>215</td>
<td>1577</td>
<td>78</td>
<td>75</td>
<td>80.5</td>
<td>6.8</td>
</tr>
<tr>
<td>5F20</td>
<td>2015</td>
<td>γ</td>
<td>og2</td>
<td>1 075 201</td>
<td>19</td>
<td>1264</td>
<td>77</td>
<td>77</td>
<td>91.8</td>
<td>1.2</td>
</tr>
<tr>
<td>5D3</td>
<td>3000</td>
<td>γ</td>
<td>mg1</td>
<td>1 373 839</td>
<td>319</td>
<td>1898</td>
<td>74</td>
<td>60</td>
<td>74.9</td>
<td>22.4</td>
</tr>
<tr>
<td>5N8</td>
<td>5010</td>
<td>γ</td>
<td>og3</td>
<td>884 242</td>
<td>86</td>
<td>1099</td>
<td>55</td>
<td>53</td>
<td>67.1</td>
<td>2.4</td>
</tr>
<tr>
<td>5T1 L11</td>
<td>5080</td>
<td>α</td>
<td>ma1</td>
<td>1 273 336</td>
<td>21</td>
<td>1557</td>
<td>87</td>
<td>85</td>
<td>99.1</td>
<td>0.97</td>
</tr>
<tr>
<td>5D1</td>
<td>5400</td>
<td>α</td>
<td>ma2</td>
<td>1 196 229</td>
<td>15</td>
<td>1477</td>
<td>86</td>
<td>85</td>
<td>98.1</td>
<td>0</td>
</tr>
<tr>
<td>5D1</td>
<td>5400</td>
<td>γ</td>
<td>mg2</td>
<td>883 129</td>
<td>154</td>
<td>1224</td>
<td>81</td>
<td>72</td>
<td>78.9</td>
<td>8.09</td>
</tr>
<tr>
<td>5D17</td>
<td>5900</td>
<td>γ</td>
<td>mg3</td>
<td>1 050 488</td>
<td>37</td>
<td>1235</td>
<td>87</td>
<td>85</td>
<td>99.8</td>
<td>2.91</td>
</tr>
<tr>
<td>5T1 C4</td>
<td>6000</td>
<td>α</td>
<td>ma3</td>
<td>1 134 945</td>
<td>46</td>
<td>1386</td>
<td>86</td>
<td>82</td>
<td>96.1</td>
<td>5.83</td>
</tr>
<tr>
<td>5T1 L9</td>
<td>6890</td>
<td>α</td>
<td>ma4</td>
<td>1 176 403</td>
<td>13</td>
<td>1412</td>
<td>87</td>
<td>85</td>
<td>96.1</td>
<td>0.97</td>
</tr>
<tr>
<td>5T3 L1</td>
<td>7100</td>
<td>α</td>
<td>ma5</td>
<td>1 267 893</td>
<td>16</td>
<td>1585</td>
<td>86</td>
<td>85</td>
<td>98.1</td>
<td>0</td>
</tr>
<tr>
<td>5T3 L15</td>
<td>8150</td>
<td>α</td>
<td>ma6</td>
<td>1 522 689</td>
<td>44</td>
<td>1967</td>
<td>86</td>
<td>85</td>
<td>100</td>
<td>0.97</td>
</tr>
<tr>
<td>5T3 L15</td>
<td>8150</td>
<td>γ</td>
<td>mg4</td>
<td>1 424 723</td>
<td>155</td>
<td>1740</td>
<td>86</td>
<td>83</td>
<td>92</td>
<td>4.21</td>
</tr>
<tr>
<td>5C4</td>
<td>9697</td>
<td>α</td>
<td>oa1</td>
<td>819 148</td>
<td>47</td>
<td>988</td>
<td>46</td>
<td>46</td>
<td>54.1</td>
<td>0</td>
</tr>
<tr>
<td>5T3 L14</td>
<td>10 900</td>
<td>α</td>
<td>ma7</td>
<td>1 202 873</td>
<td>17</td>
<td>1492</td>
<td>86</td>
<td>85</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>5T3 L9</td>
<td>10 890</td>
<td>α</td>
<td>ma8</td>
<td>1 256 128</td>
<td>6</td>
<td>1532</td>
<td>86</td>
<td>85</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>5T3 L9</td>
<td>10 890</td>
<td>γ</td>
<td>mg5</td>
<td>1 009 376</td>
<td>138</td>
<td>1313</td>
<td>86</td>
<td>77</td>
<td>90.3</td>
<td>9.06</td>
</tr>
<tr>
<td>5N. mari</td>
<td>-</td>
<td>α</td>
<td>-</td>
<td>1 645 259</td>
<td>1</td>
<td>1910</td>
<td>86</td>
<td>85</td>
<td>100</td>
<td>0.97</td>
</tr>
<tr>
<td>5Ca. N. br</td>
<td>&lt;500</td>
<td>β</td>
<td>-</td>
<td>1 232 128</td>
<td>1</td>
<td>1434</td>
<td>85</td>
<td>85</td>
<td>99.5</td>
<td>0</td>
</tr>
</tbody>
</table>

a. Metagenome assembled genomes (MAGs) of the AOA from the Mariana Trench (see Supporting Information Table S3 for information on sampling sites).

b. Cultivated AOA.

c. Single amplified genomes (SAGs) for the AOA isolated from the Ogasawara Trench (only >800 Kbp were retained). N. mari: N. maritimus (CP007026); Ca.N.br: Ca. Nitrosopelagicus brevis’ (CP007028); CSCGs: conserved single-copy genes; compl.: completeness; contam.: contamination.

lineages. The AOA in the Mariana and Ogasawara waters were phylogenetically associated with the alpha, gamma and delta lineages (Fig. 1). The most abundant AOA at the depth of 500–2000 m, OTU272811, included the reads of both gamma and delta lineages, indicating that the short 16S rRNA amplicon reads could not be used to distinguish the gamma and delta lineages. In most of our metagenomes, both alpha and gamma AOA were detected (as exemplified in Supporting Information Fig. S3). Note that the contigs of the gamma AOA genome had a much wider range of %GC content and tetranucleotide frequency (TNF) than those of the alpha genome as shown in the binning process of the T3 L15 metagenome (Supporting Information Fig. S3). The alpha lineage was distributed in the abyssal and hadal layers and showed high pairwise similarity of the 16S rRNA sequences in samples from the two trenches. The gamma lineage was identified at all depths and was more diversified, as suggested by the diversity of the gamma AOA OTUs (Supporting Information Table S2). All the SAGs affiliated with the gamma lineage were obtained from waters at a depth of <6000 m, overcoming the problem of misassembly of genomic fragments from different cells. Using the 16S rRNA genes, only one of our MAGs was grouped with delta MGI AOA; this MAG was abundant in the D11 sample from the Mariana 1000-m depth. In the phylogenetic tree, ‘Ca. N. brevis’ was also placed in the beta group composed of species inhabiting <500-m waters. With the exception of the epsilon lineage, all of the MGI AOA lineages inhabiting the water column have complete genomes, SAGs or MAGs at present (Walker et al., 2010; Santoro et al., 2015).

### Phylogenetic relationships and pangenomics of the MGI AOA lineages

Average nucleotide identity (ANI) values for the genomes of the alpha AOA, gamma AOA, delta AOA, N. maritimus (alpha lineage) and ‘Ca. N. brevis’ (beta lineage) were calculated. Within the alpha MGI AOA lineage, pairwise ANIs averaged 99.2% with a small standard deviation of 0.41. The SAG oa1 from the Ogasawara Trench also highly resembled the alpha AOA MAGs from the Mariana Trench (Fig. 2A). The pairwise ANI values for gamma MGI AOA were scattered over a wide range (85%–98%), largely accounting for the divergence of the gamma MGI AOA obtained from the two trenches. The ANIs between gamma and delta MGI AOA genomes were narrowly distributed, ranging from 80% to 83% and were at least 10% higher than those between the gamma and alpha lineages. This indicates that there is a closer relationship between the gamma and delta lineages than between the gamma and alpha lineages. The pairwise ANIs between the lineages coincided with the similarity between homologous regions of the genomes (Supporting Information Fig. S4). The alpha AOA MAG ma7 and the N. maritimus genome share regions of high similarity (>80%), while the gamma AOA MAG mg3 contains some lineage-specific
regions that are not present in other lineages. The genetic distance between the MGI AOA lineages was further displayed using a maximum likelihood (ML) tree built with 24 conserved proteins (Fig. 2B). Two clades formed in the tree, one for alpha and another composed of the remaining lineages. The short distance between the new alpha MGI AOA dwelling in the two trenches indicates their homogenous genetic background. This presents a strong contrast to the relatively larger distances between the gamma AOA. In this tree, *N. maritimus* was separated from these deep-sea alpha AOA, reflecting genomic heterogeneity within the alpha lineage. The delta AOA and ‘*Ca. N. brevis*’ remained in the clade with the gamma lineage but were distantly associated with the gamma lineage (Fig. 2B).

MAGs and SAGs from the alpha and gamma lineages were aligned to display conserved gene clusters and divergent genomic contents. The size of the conserved gene cluster was much larger in the alpha AOA pangenome than in the gamma (Fig. 3), supporting the genomic homogeneity of the alpha lineage. Moreover, the ratio of genes with unknown function was significantly higher in the gamma lineage than in the alpha (chi-square test; $P = 0.012$). The MAGs ma6 and ma8 in the alpha lineage harbour a large number of genomic regions that are specific in the two samples from >10 000 m depth. This is ascribed to the fact that the highest number of singleton gene clusters (352 in ma6 and 380 in ma8) was observed in the two alpha MAGs. Most of the genes that evolved at the bottom of the Challenger Deep are associated with unknown functions. The SAG og1 also contains some specific gene clusters that were perhaps obtained independently in the Ogasawara Trench (Fig. 3). In the pangenome of the gamma lineage, a large fraction of the gene clusters is shared by two but not all of the MAGs or SAGs; this is attributable to the genomic heterogeneity of this lineage, as indicated in Fig. 3B. The MAGs mg3 and mg4 include some singleton gene clusters in the gamma pangenome. In addition, most of the singleton genes of the MAG mg4 remain hypothetical.

**Prediction of the metabolic profiles of AOA lineages**

Metabolic pathways were depicted for the AOA lineages according to the gene annotation results of the MAGs and SAGs (Fig. 4). All the MGI AOA genomes from this study contain the genes necessary for the essential metabolism of AOA, including genes encoding the proteins essential for ammonia oxidation, the HP/HB cycle, respiration complexes and sulfate assimilation (Fig. 4).
The functional genes amoABCX for ammonia oxidation were detected in the AOA MAGs and in some of the SAGs. The deduced AmoA proteins were distributed into three phylogenetic clades (Supporting Information Fig. S5). Those of the gamma and delta lineages were grouped with the low-ammonia cluster (LAC) Ba and Bb types, respectively, while that of the alpha lineage was located in a new branch of type E within the high-ammonia cluster (HAC) (Sintes et al., 2013). AOA depend on ammonia permease to import ammonia that is present in very low concentrations; as expected, the ammonia transporter gene (amt) was identified in most of the AOA MAGs and SAGs. All the MGI AOA lineages from the Mariana and Ogasawara trenches are likely able to obtain ammonia by catalysing the cleavage of carbamoyl phosphate, urea, glycine and probably cyanide, as we identified the genes encoding urease, aminomethyltransferase (GcvT) and nitrilase (Fig. 4). GcvT catalyses the cleavage of glycine to CO₂ and NH₃ with the involvement of lipoylprotein (Supporting Information Fig. S6) (Okamura-Ikeda et al., 1987), which is rarely reported in Archaea. The genes that mediate the lipoylprotein cycle and glycine cleavage were not specialized to any of the AOA lineages; instead, they were present in some MAGs of the alpha AOA (oa1, ma1, ma2, ma4, ma6 and ma7), the delta AOA (md1) and the gamma AOA (mg1). Serine may be derived from the cleavage of peptides by peptidase S8, the gene for which was frequently detected in genomes of the alpha AOA. Choline, betaine and glycine imported via cross-membrane transporters may undergo further degradation via the lipoylprotein cycle in some of the gamma and delta MGI AOA (Supporting Information Fig. S6). However, none of the MGI AOA lineages is able to catalyse the full pathway of choline/betaine degradation as their genomes do not contain some of the necessary genes.
By comparing the KEGG genes in the MAGs and SAGs, we examined lineage-specific genes for the deep-sea AOA lineages (Fig. 5). Remarkably, the gamma and delta AOA genomes lack the genes encoding the two subunits of DNA polymerase B (PolB), which is critical for the repair of deaminated adenine and cytosine bases (Jozwiakowski et al., 2014; Abellon-Ruiz et al., 2016). There was no tag gene encoding a 3-methyladenine DNA glycosylase in these gamma and delta genomes. DNA point mutations caused by 3-methyladenine can be repaired by the enzymes encoded by tag and MPG (Wyatt et al., 1999). The identification of the MPG gene in all of the AOA genomes might to some extent compensate for the loss of the tag gene in the gamma and delta lineages (Fig. 5). The ogg1 gene, which functions in the excision of 8-oxoguanine (a type of point mutation involving guanine), was identified in all of the AOA lineages. Homologous recombination and double-strand breaks might be amended by the RadA (Seitz et al., 1998) and Mre/Her proteins (Zhang et al., 2008) respectively. The relevant genes were present in all the AOA SAGs and MAGs. We were unable to find the Pop4 gene (K03538) in the gamma and delta genomes. As a ribonuclease, POP4 can facilitate pre-tRNA cleavage by forming a complex with RPR2 protein that probably plays a major role in pre-tRNA processing (K03540) (Esakova and Krasilnikov, 2010). The genomes of the alpha AOA contain both Pop4 and Rpr2 genes. The gamma and delta lineages probably depend solely on RPR2 for pre-tRNA cleavage since most of the MAGs possess the gene encoding RPR2.

The genes that function in vitamin B12 synthesis were identified in almost all the AOA MAGs. The protein encoded by the methionine synthesis gene metH requires B12 as a coenzyme, whereas methionine synthesis by the protein encoded by metE is B12-independent (Pejchal and Ludwig, 2005). Surprisingly, the gamma AOA appear to be independent of B12 for methionine synthase, as the metE gene was identified in all their MAGs. In contrast, all of the alpha AOA in the study harboured at least one metH gene (Fig. 5).

Four of the alpha AOA MAGs lack ilvA and ilvD genes for the synthesis of branched-chain amino acids (Fig. 5). In contrast, all the alpha AOA genomes in the study possess the pepA gene, which encodes a protein that is involved in the cleavage of leucine from periplasmic peptides. The released leucines are likely imported directly into the cytoplasm with the assistance of a permease encoded by the livM gene (K01998 in Fig. 5).

Two types of ATP-binding cassette transporter genes that encode proteins responsible for the binding of phosphate compounds, phnD and pstS, were identified. phnD (K02044), which encodes a protein involved in phosphate binding, was significantly more abundant in the gamma AOA than in the alpha AOA genomes (two-tailed t-test; P = 0.004). However, pstS (K02040) for phosphate binding is present in all the alpha AOA genomes. It appears that these deep-sea AOA lineages are diversified in their utilization of phosphate sources.

The delta AOA MAG in particular was found to carry genes that are probably related to bioluminescence (Lee and Meighen, 2000). These include luxC (gene ID: 2569_3; acyl-CoA reductase), luxE (gene ID: 2569_2; long-chain fatty acid-CoA ligase/acyl-protein synthetase)
and \textit{fadD} (gene ID: 2569_1; acyl-CoA synthetase). These genes were also present in an SAG of a deep-sea MGI AOA isolate (NCBI Bioproject accession number PRJNA248555). The sequence identity of the homologues ranged from 51% to 59%, suggesting persistence of the \textit{lux} genes in some of deep-sea MGI AOA species.

The delta AOA MAG contains four sets of \textit{gtsABC} genes that encode proteins responsible for the uptake of glucose (Supporting Information Table S7), indicating a mixotrophic lifestyle for the delta MGI AOA. The \textit{gts} genes were probably obtained through horizontal gene transfer (HGT) from a \textit{Candidatus} Pelagibacter that possessed the homologues with the highest sequence identity (80%). The presence of bacterial genes in the delta AOA genome appears to be a result of horizontal gene transfer between the deep-sea bacteria and archaea (López-García et al., 2015). Our estimate was that about 11.7% of all the CDSs of the delta AOA MAG were obtained by HGT (Supporting Information Table S8). Considering that 56% of the potentially transferred CDSs were located at the ends of the contigs, some of them were probably a result of misassembly. In contrast, only about 1% and 0.25% of the CDSs were estimated to be transferred through HGT in the MAGs and SAGs of the gamma and alpha AOA respectively.

\textbf{Different adaptation strategies as indicated by AOA comparative genomics}

The MGI AOA lineages in this study adapted to their environments through different strategies with respect to their genomic features. The alpha AOA MAGs possess the \textit{cheY} gene, which encodes a protein that serves as a regulator of the chemotaxis response to environmental changes (Parkinson, 2003), but the gamma and delta genomes lack this gene (Fig. 5). In the alpha AOA genomes, the neighbouring gene of \textit{cheY} encodes the type II secretion system protein E, whereas methyl-accepting chemotaxis \textit{mcp} gene and flagella-encoding genes are absent. The MCP protein initiates the signal transfer to CheY, which triggers the subsequent movement (Parkinson, 2003). Therefore, the \textit{cheY}, \textit{mcp} and flagella-encoding genes are typically neighbours, as seen in the shallow-water AOA genomes. Hence, the \textit{cheY} gene in the alpha lineage probably encodes a signal response regulator that functions in signal transduction to stimulate the secretion system rather than motility. The \textit{beta}, \textit{gamma} and delta AOA lineages lack the \textit{cheY} gene (Fig. 5). Likewise, an \textit{ompA} gene (K02368) was identified in all of the alpha MGI AOA genomes but not in the other lineages. Several other genes related to peptidoglycan metabolism (K00012, K02472, K16881 and K02563) may also distinguish the alpha AOA from the other lineages (Fig. 5). Almost all of the deep-sea MGI AOA in this study can probably generate di-myo-inositol-phosphate (DMIP) as an osmolyte to deal with hydrostatic pressure (Cario et al., 2015), since the functional gene \textit{ipct} involved in DMIP synthesis was uniquely found in these deep-sea MGI AOA genomes. The alpha AOA genomes in this study probably do not encode the universal stress protein UspA, which is critical for survival in the stationary phase (Freestone et al., 1997). Expression of UspA is activated by nutrient starvation and by osmotic shock (Nyström and Neidhardt, 1992). In contrast, approximately a dozen
**uspA** genes were identified in the gamma, delta and shallow-water alpha AOA genomes. Numerous phage genes appear to have been integrated into the AOA genomes of the gamma and delta lineages. Phage proteins such as site-specific phage integrase, sigma factor, capsid assembly protein and phage DNA polymerase (89% identity with the homologue of phage SPO1) were present among the predicted proteins of the gamma and delta AOA. Most of the phage proteins are most similar to those of deep-sea phages isolated from the Mediterranean Sea (López-Pérez et al., 2017). The phage genes were not present at the ends of contigs and therefore seem not to be a result of misassembly.

**Discussion**

In this study, we obtained MAGs and SAGs from two trenches in the western Pacific. Based on the genomic features and phylogenetic trees of the alpha and gamma AOA lineages, we argue that many ecotypes exist in these lineages. It is likely that different strategies have contributed to the development of the ecotypes. Pangenic analysis revealed extreme genomic stability of the alpha AOA that is probably a result of adaptation to the hadal environment. The alpha lineage in this study was found to be an abyssal and hadal ecotype, but the factors that have determined its niche specification are unclear.

The alpha lineage uniquely contains genes related to outer-membrane protein A glycoprotein (OmpA) and thrombospondin, but its role in the archaea remains unknown. The thrombospondin protein contains five Ca²⁺-binding domains (Supporting Information Fig. S7A) that may regulate the cellular structure for adhesion and evade macrophage attach (Dai et al., 2017). Unlike the genes encoding known thrombospidins, the homologues of the alpha AOA genes contain several short insertions (Supporting Information Fig. S7B). Previous studies of the gram-negative bacterium *Pseudomonas aeruginosa* revealed the importance of thrombospidin in virulence-related functions such as adhesion and protection.
revealed the presence of lineage-specific genes emerged in the hadal alpha AOA genomes. These genes are probably crucial for the survival of the alpha AOA and experienced high selection pressure by the extreme environment. The high stability of the alpha AOA genomes, together with the presence of some niche-specific novel genes, suggests anagenesis of the alpha AOA ecotypes. By following the deep-sea water current, the alpha AOA could spread in global deep waters. In the present study, the MGI AOA lineages from two trenches of the western Pacific resembled each other in phylogenetic status and predicted metabolism. Recently, alpha AOA SAGs were obtained from the Puerto Rico Trench (Leon-Zayas et al., 2015); these SAGs showed >98% ANI to the MAGs from this study. The gamma AOA display high genomic plasticity that may endow them with fitness at almost full-ocean depths. For the delta AOA, its genome harbours some genes for organic carbon uptake and assimilation. The acquisition of these genes possibly thorough horizontal gene transfer was likely driven by the unstable ammonia concentration in the mesopelagic layer brought about by various types of sinking organic matter. It is difficult to estimate the rate of horizontal gene transfer events in a genome. The recently acquired genes show abnormal %GC content, TNF and codon usage pattern (Lawrence and Ochman, 2002; Techtmann et al., 2012). In our binning results, the gamma and delta AOA contigs are more diversified in %GC and TNF than the alpha contigs, indicating that the gamma and delta AOA genomes are more dynamic due to acquisition of larger amounts of foreign DNA. However, the delta AOA genome could perhaps reflect a broader ecological potential, a topic for future study. The instability of the gamma and delta AOA genomes provides strong evidence for the involvement of cladogenesis in the formation of ecotypes showing high genetic diversity. Given the full-depth distribution of the gamma lineage, we propose that different ecotypes have formed in the gamma lineage across the depth due to allopatric speciation.

The genomic comparison conducted in this study revealed the presence of lineage-specific genes and pathways that reflect the genomic divergence of the lineages under different stresses and to some extent explain their vertical distribution in the deep waters. Lack of PolB, which is essential for DNA template stability, is believed to provide the delta and gamma AOA with more genetic variants due to the resulting accumulation of point mutations caused by the deamination of single-stranded DNA; such mutations are a prerequisite for the emergence of various ecotypes at different depths and in different microenvironments. We cannot exclude the possibility that unknown genes that encode proteins with functions similar to those of PolB are present in our MAGs for the gamma and delta AOA. Another DNA repair system, MutSL, which is specialized for the repair of point mutations (Busch and DiRuggiero, 2010), was also not detected in gamma AOA or in the other lineages in this study. Together with the absence of the tag gene for repair of 3-methyl-adenine mutations, the genomes of the gamma and delta AOA lineages might have experienced extensive genomic changes. Furthermore, phage insertions, which may have brought new genes into their genomes, are present. The loss of the Pop4-encoding gene in the gamma and delta AOA is likely to have resulted in low translation efficiency. Moreover, the lack of extracellular structures (such as thrombospondin in the alpha AOA) in the gamma and delta AOA might facilitate gene flow among members of the microbial community. The cellular structures and the molecular machinery present in the gamma and delta AOA lineages make the formation of ecotypes through cladogenesis an easy process. The occurrence of genomic variants has also permitted a wide distribution of the gamma ecotypes in the full-ocean depths. In contrast, the alpha AOA lineage was restricted in the hadal zone and maintained homogenous genomes. The low rates of gene flow and mutation of the alpha AOA genomes likely hindered anagenesis and sympatric speciation of the alpha AOA in the trenches. In this study, the gamma and delta AOA were shown to employ the MetE, which acts much more slowly than the MetH, for methionine synthesis. The lack of a B12-dependent metH gene could impose a burden on the translational machinery in the gamma and delta AOA (Danchin and Brah, 2017). If true, this may be an indicator of an overall low metabolic level of the gamma and delta AOA. UspA, which was shown in this study to be encoded by the genomes of the gamma and delta AOA, may regulate these ecotypes into dormancy under starvation conditions. Overall, the metabolic incompleteness of some MGI AOA ecotypes in this study can be explained by the ‘black queen hypothesis’ (Morris et al., 2012). The loss of some genes, such as those involved in glycine degradation, might benefit AOA ecotypes with reduced genomic content. The products generated by the leaky functions would be supplied by helpers in the community.

The MGI AOA lineages conduct a critical nitrification process in deep-sea waters. The low ammonia concentration in the abyssal layer is a limiting factor for the chemolithoautotrophic life mode of MGI AOA; ammonia is often present in deep waters at levels below the limit of detection (Nunoura et al., 2015). Hence, ammonia permease was highly expressed by the AOA, perhaps to endow them with a high affinity for ammonia in the water, as shown in this study. A previous study suggests that
enrichment of organic matter by suspended sediments increases the ammonia flux of the hadal zone and subsequent ammonia oxidation by the alpha AOA (Nunoura et al., 2015). In this study, we did not detect genes that support the prevalence of members of the alpha lineage with a heterotrophic lifestyle in the hadal depths. Recently, AOA were argued to be chemolithomixotrophic (Mussmann et al., 2011; Hatzenpichler, 2012). However, recent work has shown that α-keto organic acids such as pyruvate are simply used to detoxify H₂O₂ (Kim et al., 2016). Our results indicate that some AOA lineages have evolved to utilize versatile carbon sources and have probably developed heterotrophic lifestyles. Furthermore, amino acids such as serine and glycine might be used as ammonia, nitrogen and carbon sources by these deep-sea AOA. In previous studies, detrital material was suggested to be degraded by deep-sea microorganisms (Lloyd et al., 2013). With the flourishing of the MGI AOA lineages in the deep oceans, biogenic nitrite flux therein was sustained by AOA. Such a nitrification process is perhaps a further driving force for inorganic carbon fixation by nitrite oxidizers (Pachiaidaki et al., 2017). In return, nitrite oxidizers would probably increase the supply of ammonia in the water by their own degradation of urea, cyanate and amino acids (Pachiaidaki et al., 2017).

Overall, based on the results obtained in this study, we propose that there is an association between stratification pattern, ecotype formation scheme and genomic differences of the AOA lineages in the deep-sea layers. Our study also suggests that these AOA lineages contribute to the carbon and nitrogen cycles in different ways. In situ filtration of the collected water samples in this study made it possible to conduct a metatranscriptomic study of different AOA lineages, which may permit a full evaluation of their contribution in production of organic carbon for the deep-sea ecosystem.

**Experimental procedures**

**Sample collection**

Water samples were collected from the Ogasawara Trench by R/V Kairei in December 2011 and from the southern Mariana Trench by R/V DY37II, Tansuo07 and Tansuo03 between June 2016 and March 2017 (Supporting Information Fig. S1 and Table S1). During the R/V Kairei KR11-11 cruise, ROV ABISMO was used by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) to collect 5-L samples of water in individual Niskin bottles from depths ranging from 505 m to 9697 m. For water collection from the Mariana Trench, Niskin bottles on a conductivity, temperature and depth (CTD) unit and landers were used to collect 70-L water samples from depths between 500 m and 6500 m. The waters were filtered through 0.22 μm polycarbonate membranes on board (diameter 45 mm; Whatman, Clifton, NJ, USA). For samples obtained below 6500 m, an in situ automatic apparatus was employed to filter microbes in 80 l-300 l of bottom waters (1 m above the seafloor) through a 0.22 μm membrane (diameter 142 mm; Whatman, Clifton, NJ, USA) (Supporting Information Table S1 and Fig. S8A). The filtration was driven by an electromagnetic pump controlled by the landers. Immediately after the filtration, ~300 ml RNAlater (Ambion, Carlsbad, CA, USA) was injected automatically by the pump into the filtration chamber to preserve the RNA in the microbial cells (Supporting Information Fig. S8B). The RNAlater injection was also controlled by the hadal lander (Supporting Information Fig. S8). For the single-cell genomics studies, water samples from the Ogasawara Trench were stored as described previously (Nunoura et al., 2013). The samples were maintained at −80 °C on board until further processing.

**16S rRNA amplicons and metagenomic sequencing**

Genomic DNA was extracted from the filtration membranes using a Powersoil DNA isolation kit (Qiagen, Hilden, Germany). 16S rRNA genes were amplified using the universal primers 341F (5′-CCTAYGGGRBGCASCAG-3′) and 802R (5′-TACNVGGTTATCTAATCC-3′) (Claesson et al., 2010; Klindworth et al., 2013), which target the V3-V4 region of 16S rRNA genes. The PCR conditions for 16S rRNA amplification were as follows: initial denaturation at 98 °C for 10 s; 30 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 15 s and extension at 72 °C for 30 s; and a final extension at 72 °C for 5 min. Two PCR replicates were performed for each sample. After library preparation using a TruSeq Nano DNA LT kit (Illumina, San Diego, CA, USA), the DNA samples and 16S rRNA gene amplicons were sequenced on an Illumina MiSeq platform via 2 × 300 bp paired-end sequencing (Illumina). The amplicons were less than 500 bp in length and thus could be fully sequenced using the MiSeq platform.

**Separation of AOA OTUs for different samples**

The raw reads for 16S rRNA gene amplicons were separated based on the 6-nt barcodes assigned to the different samples. Low-quality reads were removed by NGSQC Toolkit (v2.3.3) (Patel and Jain, 2012). The quality-filtered paired-end reads were assembled using PEAR (v0.9.5) (Zhang et al., 2014). The sequenced amplicons were then analysed via the QIIME pipeline with the SILVA rRNA database version 128 as a reference (Caporaso et al., 2010). The reads were grouped into OTUs at 97% similarity. The OTUs in the classification of MGI were extracted for further phylogenetic examination with AOA.
references. The percentages of the AOA OTUs in different replicates and at different depths were averaged for different depth ranges, and the standard deviation of the averaged values was calculated.

**SAG library construction and sequence analyses**

Single-cell sorting and whole genome multiple displacement amplification (MDA) for the Ogasawara Trench samples were conducted in the Bigelow Laboratory Single Cell Genomics Center (SCGC) as described previously (Swan et al., 2011). One SAG library in a 384-well plate was constructed for each sample. PCR screening using primer sets for archaeal and bacterial SSU rRNA genes (Swan et al., 2011) was conducted in the Bigelow Lab and at JAMSTEC. To prepare a template for Sanger sequencing in JAMSTEC, 10 cycles of PCR amplification were conducted to add M13 sequencing primers (forward, 5’-GTTTTCCCAGTCACGAC-3’ and reverse, 5’-CAGGAAACAGCTATGAC-3’) to SSU rRNA primers and metabolic genes. The amplified SSU rRNA and metabolic genes were then directly sequenced using M13 primers on a 3730xl genetic analyser (Thermo Fisher Scientific, Waltham, MA, USA). In the Bigelow Lab, the sequencing of selected SAGs was performed using the Illumina MiSeq platform as previously described (Swan et al., 2011).

**De novo genome assembly with SPAdes** (v3.6) (Nurk et al., 2013) and assembly QC were performed at SCGC as previously described (Stepanauskas et al., 2017). This pipeline was evaluated for assembly errors using five microbial benchmark cultures with diverse genomic complexity and %GC content of DNA; the results indicated 60% average genome recovery, an absence of nontarget or undefined bases, and low frequencies of nonassemblies, indels, and mismatches (Stepanauskas et al., 2017). Functional annotation of the obtained SAGs was performed using a combination of Prokka (Seemann, 2014) and UniProt (Kiefer et al., 2009). For taxonomic assignments, 16S rRNA gene regions were identified using local alignments provided by Blastn (Altschul et al., 1997) against the SILVA reference database (Pruesse et al., 2007). Identified regions were annotated using Blastn hits for each region against CREST’S (Lanzén et al., 2012) curated SILVA reference database (Pruesse et al., 2007). Taxonomic assignment was based on application of a reimplementation of CREST’S LCA algorithm to the top 2% (sorted by bit score) of hits per region.

**Binning and analyses of MAGs**

The quality-filtered metagenomics data were assembled using SPAdes with default parameters for individual water samples (Nurk et al., 2013). Coverage of the contigs was calculated by searching the reads that could be mapped to the contigs using Bowtie 2 with default settings (Langmead and Salzberg, 2012). Genome binning of AOA lineages was performed by separation of contigs with consistent coverage level and %GC content followed by an examination of TNF by CA analysis (Albertsen et al., 2013). MAGs in the contigs were predicted by Prodigal (v2.6.2) (Hyatt et al., 2012). The 162 Pfam models (Rinke et al., 2013) were searched in 29 genomes from 13 archaeal phyla to obtain CSCGs for assessment of completeness of the AOA genomes. A total of 87 archaeal CSCGs were collected from the Pfam models (Supporting Information Table S2). These CSCGs could be identified in at least 27 genomes. According to our survey, known complete AOA genomes have 86 or 87 CSCGs. The CSCGs in the AOA bins were counted to provide a preliminary evaluation of the completeness of the MAGs. A further evaluation of the AOA genomes was conducted using CheckM (v1.0.5) (Parks et al., 2015). The CDSs were annotated against the NCBI nr and KEGG databases using Blastx with an e-value cutoff of 1E-5. In the BLASTP search results, the CDSs possibly obtained through HGT were identified using information of the homologues with the highest BLASTP score. A name list was created for the MGI AOA. If the species of the homologue was out of the list and the similarity was higher than >50%, a case of HGT was predicted for the CDS.

**Pangenomics analysis**

ANI between the genomes was calculated in www.ezbiocloud.net. A pangenomic comparison was conducted using Anvio-4 according to the tutorial (Eren et al., 2015). The gene annotation of the pangenome was performed using BLASTP against the COG database released in December 2014. Proteins with a minimal similarity of 35% to known proteins in the KEGG database were collected to display genomic disparity. For the conserved hypothetical genes, functions were inferred from 3-dimensional structural patterns predicted by SWISS-MODEL (Waterhouse et al., 2018).

**Construction of phylogenetic trees**

The 16S rRNA genes of AOA were extracted from metagenomes, SAGs and MAGs using rRNA_HMM (Huang et al., 2009). They were used to build an ML phylogenetic tree with representative reads of major AOA OTUs and reference sequences. The rRNA sequences were first aligned using MAFFT L-INS-i (v7.294b) (Alva et al., 2016) and then adjusted with trimAl (v1.4) (Capella-Gutierrez et al., 2009). The trees were inferred from 1492 aligned positions by the ML algorithm using raxmlGUI v1.5.

© 2018 Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology
(Silvestro and Michalak, 2012) (GTRGAMMA model). To obtain bootstrap values, 1000 replicates were performed. An ML phylogenetic tree was also constructed using 24 protein sequences. The archaeal conserved proteins were aligned using MAFFT and concatenated, which resulted in 3825 alignment positions. PROTGAMMA +BLOSUM62 was selected as the substitution model for the construction of a phylogenetic tree with bootstrap values calculated based on 1000 replicates.

Availability of data

The reads of 16S amplicons were deposited in the NCBI SRA database under the accession numbers SRR6466501, SRR6466502, SRR6466503 and SRR6466504. The MAGs for AOA lineages were deposited in the BIGD database (Big Data Center Members, 2017) under the accession numbers SAMCO21084-SAMCO21097. The BIGD accession numbers of the four SAGs listed in Table 1 are SAMCO26833-SAMCO26836.

Acknowledgements

We are thankful to the team members aboard the R/V DY37II, Tansuo01, Tansuo03 and Kairei KR11-11 for their invaluable efforts during the sampling cruises. We are also grateful to J. Chen and D.S. Cai for the operation of landers and for in situ sampling work. This study was supported by the National Science Foundation of China (No. 41476104 and No. 31460001), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB06040101) and the National Key Research and Development Program of China (2016YFC0302500) to Y. Wang and for AOA lineages were deposited in the BIGD database (Big Data Center Members. (2017) The BIG data center: from deposition to integration to translation. Nucleic Acids Res 45: D18–D24.


