MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1986–1999 doi: 10.1111/mec.13607

Molecular adaptation of ammonia monooxygenase during independent pH specialization in Thaumarchaeota

DANIEL J. MACQUEEN and CÉCILE GUBRY-RANGIN

Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK

Abstract

Microbes are abundant in nature and often highly adapted to local conditions. While great progress has been made in understanding the ecological factors driving their distribution in complex environments, the underpinning molecular-evolutionary mechanisms are rarely dissected. Therefore, we scrutinized the coupling of environmental and molecular adaptation in Thaumarchaeota, an abundant archaeal phylum with a key role in ammonia oxidation. These microbes are adapted to a diverse spectrum of environmental conditions, with pH being a key factor shaping their contemporary distribution and evolutionary diversification. We integrated high-throughput sequencing data spanning a broad representation of ammonia-oxidizing terrestrial lineages with codon modelling analyses, testing the hypothesis that ammonia monooxygenase subunit A (AmoA) – a highly conserved membrane protein crucial for ammonia oxidation and classical marker in microbial ecology – underwent adaptation during specialization to extreme pH environments. While purifying selection has been an important factor limiting AmoA evolution, we identified episodic shifts in selective pressure at the base of two phylogenetically distant lineages that independently adapted to acidic conditions and subsequently gained lasting ecological success. This involved nonconvergent selective mechanisms (positive selection vs. selection acting on variants fixed during an episode of relaxed selection) leading to unique sets of amino acid substitutions that remained fixed across the radiation of both acidophilic lineages, highlighting persistent adaptive value in acidic environments. Our data demonstrates distinct trajectories of AmoA evolution despite convergent phenotypic adaptation, suggesting that microbial environmental specialization can be associated with diverse signals of molecular adaptation, even for marker genes employed routinely by microbial ecologists.

Keywords: ammonia oxidation, archaea, environmental adaptation, evolutionary ecology, microbial evolution, natural selection

Received 30 November 2015; revision received 24 February 2016; accepted 2 March 2016

Introduction

Molecular evolutionary analyses are widely used in ecology and evolution to understand the genetic diversity of organisms. Such studies have been enormously valuable in describing adaptive processes in cultivated model microorganisms such as *Escherichia coli* (e.g. Barrick *et al*. 2009), *Pseudomonas fluorescens* (e.g. McDonald *et al*. 2009) and *Burkholderia cenocepacia* (e.g. Traverse *et al*. 2013), as well as pathogens isolated during disease outbreaks (e.g. Kennemann *et al*. 2011; Snitkin *et al*. 2011; He *et al*. 2013). However, this field remains in its infancy for non-pathogenic prokaryotes that abound in natural environments, especially uncultivated microbes. Adaptation to environmental change occurs through evolutionary processes (i.e. mutation, natural selection, genetic drift,

Correspondence: Daniel J. Macqueen, E-mail: daniel.macqueen@abdn.ac.uk; and Cécile Gubry-Rangin, E-mail: c.rangin@abdn.ac.uk

recombination and lateral gene transfer) that modify phenotypes via changes at the molecular level, especially protein functions and/or regulation of gene expression. Therefore, to understand the origins of adaptation in noncultivated and nonpathogenic microbes, including the phenotypic specializations that ultimately determine natural distributions and crucial ecosystem functions, it becomes crucial to infer the underpinning molecular evolutionary processes. Several studies have used such approaches, for example to study marine microbes (Tai *et al.* 2011; Luo *et al.* 2015), the gut microbiome (Schloissnig *et al.* 2013) and microbes from acid mine drainage systems (Allen *et al.* 2007), but generally, detailed molecular evolutionary analysis of prokaryotes in complex environments, e.g. agricultural and grassland soils, are rare (e.g. Thakur *et al.* 2013; Li *et al.* 2014).

Among the widely distributed nonpathogenic prokaryotes, Thaumarchaeota have been highly studied over the last decade, owing largely to their key nitrification function (which is central to the global nitrogen cycle) and massive abundance in marine and terrestrial ecosystems. This microbial phylum is phylogenetically diverse (Gubry-Rangin *et al.* 2011, 2015; Pester *et al.* 2012; Vico Oton *et al.* 2015) and specialized to a number of environmental conditions (Hatzenpichler 2012). Most work on Thaumarchaeota has focused on questions relating to ecophysiology and ecological distribution. On these lines, and similarly to other microbes (Fierer & Jackson 2006), pH has been shown to be a central factor driving thaumarchaeotal terrestrial distribution and adaptation (Nicol *et al.* 2008; Gubry-Rangin *et al.* 2011; Vico Oton *et al.* 2015), with several highly successful lineages known to be physiologically specialized to acid or alkaline conditions. Interestingly, a strong link between pH and thaumarchaeotal diversity extends across vast evolutionary time, as reconstructed rates of pH adaptation and lineage diversification are tightly coupled over the phylum's entire evolutionary history – stretching over half a billion years (Gubry-Rangin *et al.* 2015).

Most studies analysing the environmental distribution of Thaumarchaeota have focused on a conserved gene coding the ammonia monooxygenase subunit A (*amoA*). Its protein product (AmoA) forms part of the ammonia monooxygenase protein complex required for ammonia oxidation and energy production in archaea and bacteria (Bédard & Knowles 1989; Hyman & Arp 1992; McTavish *et al.* 1993). Despite the huge contemporary application of *amoA* as a marker gene in diverse environments (Rotthauwe *et al.* 1997; Pester *et al.* 2011; Vico Oton *et al.* 2015) – evidenced by >750 papers in PubMed with the terms 'amoA' in the title/abstract (along with >120 000 microbial *amoA* sequences within the NCBI nucleotide database) – it remains unknown whether AmoA undergoes adaptation to the broad range of environments inhabited by ammonia-oxidizing bacteria and archaea. In this respect, it is critical to note that the evolution of proteins located within the lipid cytoplasmic membrane, including AmoA (Walker *et al.* 2010), is expected to be strongly influenced by extracellular pH. In archaea, the lipid membrane plays a crucial role in maintaining cellular homeostasis and acts as the main barrier between the environment and cellular cytoplasm (see López de Saro *et al.* 2015). The main membrane lipids are glycerol dibiphytanyl glycerol tetraethers (GDGTs) and archaea alter the conformation of these molecules, particularly the number of cyclopentane rings, in response to the environment. This modifies lipid packing density, influencing membrane thermal stability and ionic permeability (López de Saro *et al.* 2015). The number of GDGT cyclopentane rings correlates inversely with pH (e.g. Macalady *et al.* 2004; Pearson *et al.* 2008; Boyd *et al.* 2011, 2013; Xie *et al.* 2015), which is thought to reduce the influx of protons in acidic environments (López de Saro *et al.* 2015). Therefore, as a study premise, we postulate that modifications to lipid membrane conformation required under different pH conditions necessitates adaptations in membrane localized proteins, including ammonia monooxygenase, to maintain essential protein functions and protein–protein interactions, along with normal protein folding. In the case of AmoA, while much of this protein comprises helixes within the lipid membrane, a large region is further localized within the pseudo-periplasm (Lieberman & Rosenzweig 2005; Walker *et al.* 2010), which is likely even more directly influenced by environmental pH, given that the only barrier to the external environment, the S-layer cell wall, is unlikely to markedly buffer external pH (Sára & Sleytr 2000).

Therefore, our study objective was to test the hypothesis that AmoA undergoes molecular adaptation during pH specialization – particularly to low pH environments – in terrestrial Thaumarchaeota. This was achieved using an extensive representation of *amoA* sequence diversity generated by high-throughput 454-sequencing of contemporary soil environments covering a large pH spectrum. This data set was used in comprehensive probabilistic codon analyses within a phylogenetic framework, allowing a detailed dissection of evolutionary changes in AmoA during major independent adaptation events leading to highly successful pH-adapted thaumarchaeotal lineages.

Material and methods

Data used in evolutionary analyses

Sampling of thaumarchaeotal a*moA* sequences and generation of phylogenetic trees used in this study have

been described elsewhere (Gubry-Rangin *et al.* 2011, 2015). Briefly, high-throughput sequencing was performed on 47 UK soils targeting the archaeal *amoA* gene (Gubry-Rangin *et al.* 2011) and a resultant set of 370 sequences, broadly representative of phylum-scale diversity, was used to perform a joint Bayesian phylogenetic and relaxed molecular clock analysis after exclusion of confounding effects linked to recombination and mutational saturation (Gubry-Rangin *et al.* 2015). An associated 370-sequence in-frame alignment representing 89% of the *amoA* coding sequence (missing 8 and 14 codons from the respective 5′ and 3′ of the alignment) was used for evolutionary analyses, along with an associated phylogenetic tree employed in evolutionary analyses. Both the alignment and tree are provided in Data S1 (Supporting information).

Codon analyses and tests of selective pressures

All analyses of selective pressures acting on the *amoA* coding sequence were performed using models and packages available within HyPhy (Pond *et al.* 2005) and the associated web-server Datamonkey (Delport *et al.* 2010). We used HyPhy to build likelihood functions for two models where d_N and d_S was either free or fixed across branches of the phylogeny, which was done under the MG94 codon model (Muse & Gaut 1994) crossed with the standard general time reversible (GTR) nucleotide substitution model. The fit of the data under the two likelihood functions was compared using a likelihood ratio test. We separately estimated d_N and d_S for every branch/node in the 370-sequence *amoA* phylogeny by crossing MG94 with the best-fitting of all possible GTR models, using parametric bootstrapping with 200 replicates to establish variance around parameter estimates. The BUSTED test (Murrell *et al.* 2015) was run using the full 370-sequence alignment, setting foreground lineages as described in the main text. The BS-REL and RELAX tests (Kosakovsky Pond *et al.* 2011; Wertheim *et al.* 2015) were run using a subset of 73 sequences from the 370-sequence codon alignment. This was done by manually pruning tips within the phylogenetic tree using the editor within the Mega 5.0 package (Tamura *et al.* 2013) and therefore without removing any major thaumarchaeotal groups or modifying the backbone structure of the tree. The 73-sequence alignment and associated phylogenetic tree is provided as supporting information (Data S2, Supporting information). Finally, AmoA amino acid sequences were reconstructed for every node in the thaumarchaeotal phylogeny using the 370-sequence in-frame alignment and associated tree topology. This was done by translating codon sequences reconstructed within HyPhy using a joint maximum-likelihood approach (Pupko *et al.* 2000) after a local fitting (i.e. to every branch) of the MG94 codon model crossed with the GTR model, implementing the CF3 × 4 option to estimate codon frequencies. This model was selected to ensure that ancestral sequence reconstructions were exactly congruent with the BS-REL analysis, which implements the same (i.e. MG94xGTR-CF3 × 4) model during calculation.

Modelling of AmoA protein structure

There is no published 3-dimensional structure for the α-subunit ammonia monooxygenase encoded by *amoA*. Therefore, the functional implications of residues of interest were inferred with respect to a predicted AmoA secondary structure, modelled using the TMHMM algorithm (Krogh *et al.* 2001) and rendered in TMRPres2D (Spyropoulos *et al.* 2004).

Results

Linking pH adaptation to molecular evolution of AmoA

Phylogenetic reconstruction of *amoA* thaumarchaeotal sequences representing extensive mesophilic terrestrial diversity and diverse pH conditions has been recently used to demonstrate that pH adaptation occurred during thaumarchaeotal evolution, leading to several pH-adapted lineages (Gubry-Rangin *et al.* 2011, 2015). Three pH-adapted lineages are particularly important in terms of their high abundance in contemporary terrestrial ecosystems, i.e. the phylogenetically-distant acidophilic lineages called C11 (within *Nitrososphaera*) and C14/C15 (within *Nitrosotalea*) and the alkalinophilic lineage called C1/2 (within *Nitrososphaera*) (see Fig. 1). In this study, we focus on these three major pH-adapted lineages and hypothesize that important adaptations involved in pH specialization arose in their common ancestors. Therefore, in a phylogenetic framework, we place *a priori* significance on the ancestral branches where pH adaptation first occurred and make the following predictions: (i) because AmoA function is essential for energy production, it is therefore highly conserved and a predominant signal of purifying selection should exist across the thaumarchaeotal phylogeny, (ii) but despite this, AmoA underwent episodic periods of adaptation concurrent to major pH specialization events, and (iii) amino acid changes fixed along branches where pH adaptation first occurred and have then subsequently remained invariant in pH-adapted lineages are particularly strong candidates to represent substitutions of persistent adaptive value in pH-specialized lineages.

These predictions were tested using codon-based modelling analyses employing 370 near full-length

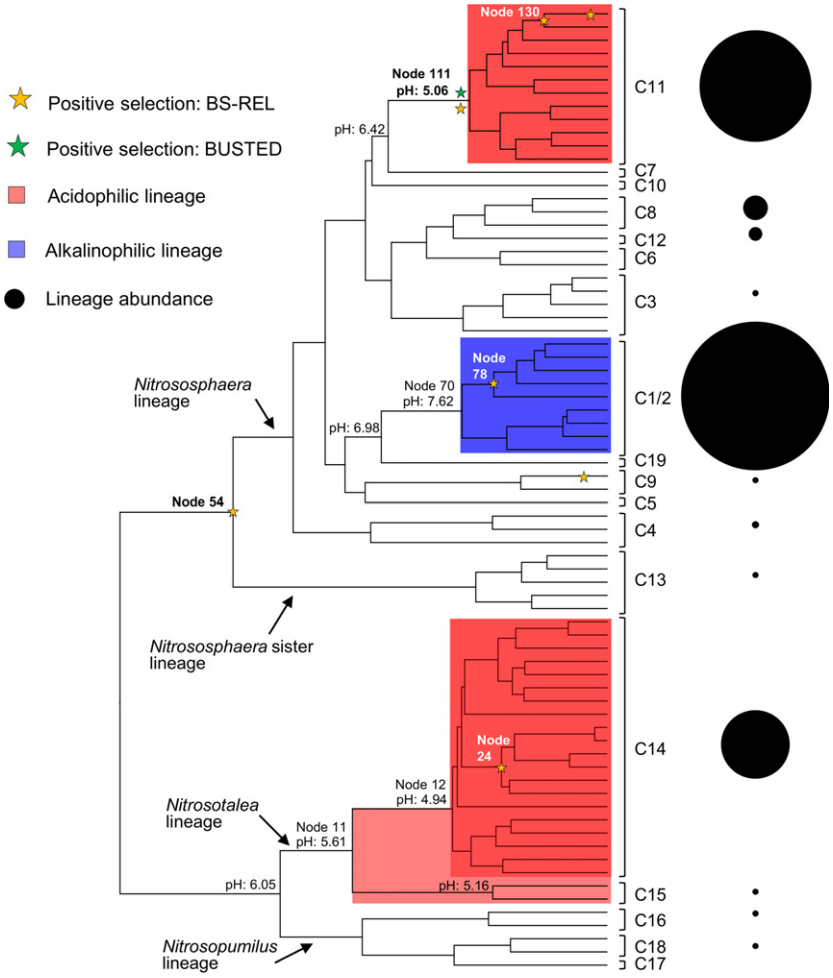

Fig. 1 Evolution of pH-adapted thaumarchaeotal lineages and results of tests for positive selection. The black circles indicate the relative abundance of the different thaumarchaeotal lineages (after Gubry-Rangin *et al.* 2011) and highlight the ecological dominance of the three pH-adapted lineages of interest in this study. Stars indicate nodes where positive selection was inferred following analyses using two tests (BS-REL and BUSTED) (also see Tables 1 and 2). Reconstruction of pH preferences is also shown for nodes at the base of the three abundant pH-adapted lineages (after Gubry-Rangin *et al.* 2015).

amoA protein-coding sequences (591 bp, 197 codons) broadly representative of mesophilic ammonia-oxidizing thaumarchaeotal diversity at the phylum level and devoid of detectable recombination events (after Gubry-Rangin *et al.* 2015). These analyses were done using maximum likelihood (ML) within the HyPhy package (Pond *et al.* 2005) along a robust Bayesian phylogeny derived from *amoA* sequences (Gubry-Rangin *et al.* 2015), which shows high congruence with an independent marker gene, 16S rRNA (Vico Oton *et al.* 2015).

General description of selective constraints operating on AmoA

As predicted, purifying selection has been a predominant selective force operating on AmoA during thaumarchaeotal evolution. By enforcing a single evolutionary model to estimate the overall rate of nonsynonymous (d_N) and synonymous (d_S) substitutions across the thaumarchaeotal phylogeny, we observed d_N/d_S (ω) = 0.050 [95% confidence interval (CI_{95}): 0.047, 0.053], indicative of strong purifying selection to remove coding variants that are presumably deleterious. Nevertheless, marked amino acid variation was still observed within the *amoA* codon alignment. Moreover, an unconstrained model allowing separate d_N and d_S estimates for each branch of the tree had massively better fit to the data vs. a model where d_N and d_S were constrained to be equal across branches (unconstrained model: Log$L$ = −65 357, AIC value = 135 138; vs. constrained model: Log$L$ = −74 155, AIC value = 148 316; likelihood ratio test [LRT] statistic: 17.6, $P < 0.00001$). Therefore, despite the strong background of purifying selection, there is clear evidence of heterogeneity in the historical pattern of selective pressure acting on the *amoA* coding region, likely representing variation both among lineages and/or sites of the AmoA protein.

Tests for episodic positive selection linked to pH adaptation

As a step towards understanding the heterogeneity of AmoA evolution across different regions of the thaumarchaeotal phylogeny, two related branch-site tests

were performed in a random effects likelihood framework to identify episodic periods of positive selection (Kosakovsky Pond *et al.* 2011; Murrell *et al.* 2015). The first approach ('BS-REL': Kosakovsky Pond *et al.* 2011) tests every branch in the tree for positive selection, using a subsequent Holm-Bonferroni correction to control for multiple comparisons (Kosakovsky Pond *et al.* 2011). Due to computing constraints, this approach had to be done on a subrepresentation of the 370-sequence codon alignment (n = 73 sequences), still fully covering the diversity of all major thaumarchaeotal groups – in particular, the 'tip-pruning' approach used (see methods) did not affect the topology associated with nodes leading into the major pH-adapted lineages. The second approach ('BUSTED': Murrell *et al.* 2015) allows branches of interest to be assigned *a priori* then tested for positive selection and can accommodate larger alignments – hence, it was ran on the full 370-sequence codon alignment. Four nodes leading into the abundant pH specialized-lineages (Gubry-Rangin *et al.* 2015) were defined as *a priori* candidates for positive selection, allowing significance to be assigned without need for multiple comparison correction in the BS-REL test. These were the reconstructed branches leading to C1/2, C11, C14/C15 and C14 alone (Fig. 1). The rationale for considering C14/C15 and C14 ancestor nodes separately comes from the reconstruction of pH adaptation (see Gubry-Rangin *et al.* 2015): while adaptation to low pH started in the ancestor to the C14/C15 lineage, C14 went on to become specialized to even lower pH, i.e. pH adaptation leading into the abundant C14 lineages spans two sequential branches (Fig. 1).

The BS-REL test provided evidence of positive selection in the ancestral branch leading into C11 ($P < 0.001$), with a small number of sites evolving under strong positive selection (proportion: 3%, mean $\omega = 10\,000$) and most others under strict purifying selection (proportion: 96%, mean $\omega = 0.02$) (Fig. 1, Table 1). The BUSTED test also detected positive selection at the same node (LRT statistic: 7.66, $P = 0.029$), again highlighting a small number of sites under strong positive selection (Table 2). Conversely, both BS-REL and BUSTED failed to detect positive selection at the ancestral branches leading to C14/C15 and C14 (BS-REL: $P = 1$ in both cases, BUSTED: $P = 0.914$ and 0.836, respectively) (Fig. 1, Tables 1 and 2). Interestingly, in contrast with C11, the BS-REL results imply that a significant fraction of codons belong to the ω^N class and hence evolved under relaxed constraints in the ancestors to both C14/C15 and C14 (Table 1). Positive selec-

Table 1 Results of the branch-site test (BS-REL) for positive selection on AmoA during Thaumarchaeota evolution and pH adaptation

Node/ branch	Mean ω	ω^-	Pr [ω = ω^-]	ω^N	Pr [ω = ω^N]	ω^+	Pr [ω = ω^+]	LRT	*P*-value	Corrected *P*-value	Comment
Node111	0.08	0.00	0.96	0.00	0.01	10000.00*	0.03	8.66	0.0010	N/A	Ancestor C11
Node11	0.07	0.00	0.87	0.80	0.13	0.65	0.00	0.00	1.0000	N/A	Ancestor C14/C15
Node12	0.51	0.50	0.13	0.50	0.73	0.53	0.14	0.00	1.0000	N/A	Ancestor C14
Node70	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.0000	N/A	Ancestor C1/C2
Node130	0.22	0.00	0.96	0.00	0.03	10000.00*	0.01	14.50	0.0000	0.01	Within C11
Node24	10.00*	0.00	0.99	0.00	0.00	10000.00*	0.01	13.40	0.0010	0.02	Within C14
C9_S14_116_2	0.06	0.00	0.96	0.00	0.02	1188.10	0.02	13.30	0.0010	0.02	Within C9
Node54	0.18	0.00	0.90	0.00	0.00	73.01	0.09	13.20	0.0010	0.02	Basal node in tree
C11_S0_2_59	0.04	0.00	0.99	0.00	0.00	279.49	0.01	11.90	0.0010	0.04	Within C11
Node78	10.00*	0.00	0.98	0.00	0.00	10000.00*	0.01	11.50	0.0010	0.05	Within C1/2

Mean ω is the average ω inferred under the MG94 codon x REV nucleotide substitution model allowing variation among lineages, but not sites; ω^- is a ML estimate of a first rate class where $\omega \leq 1$ (purifying selection); Pr[ω = ω^-] is a ML estimate of the proportion of sites fitting the ω^- class; ω^N is a ML estimate of a second rate class where $\omega\text{-} \leq \omega \leq 1$; Pr[ω = ω^N] is a ML estimate of the proportion of sites fitting the ω^N class; ω^+ is a ML estimate of a rate class where $\omega^N \leq \omega$ and otherwise unconstrained; Pr[ω = ω^+] is a ML estimate of the proportion of sites fitting the ω^+ class; LRT is the result of a likelihood ratio test of a null hypothesis where $\omega^+ = 1$ (null) vs. ω^+ unrestricted (alternative).

Underlined nodes/data were set as *a priori* candidates for positive selection and hence not subjected to multiple comparison correction. After systematic testing of all remaining nodes, the remaining (i.e. not underlined) shown nodes were inferred to be under positive selection after application of Holm-Bonferroni correction.

*ω values highlighted as either 10.00^A or $10\,000.00^A$ are unrealistic ω estimates that have been capped and are associated with a lack of synonymous substitutions on the tested branches, leading to negligible estimates of synonymous substitution rate (i.e. $d_S = 0$) during the HyPhy calculation. Despite being unreliable indicators of ω, such values have no impact on the statistical inference of positive selection with the BS-REL test.

Table 2 Results of BUSTED analysis for positive selection on AmoA during Thaumarchaeota evolution and pH adaptation

Test branch	Model	log*L*	Parameters	AICc	ω^1	ω^2	ω^3	LRT statistic	*P*-value
C11 ancestor	Alternative	−31995.0	761	65528.3	0.028 (90%)	0.028 (7.6%)	779.00 (2.4%)	7.04	0.029
	Null	−31998.5	760	65533.3	0.000 (92%)	0.00 (0.39%)	1.00 (7.9%)		
C14 ancestor	Alternative	−30037.7	761	61613.7	0.048 (79%)	0.022 (15%)	2.65 (6.0%)	0.18	0.91
	Null	−30037.8	760	61611.8	0.00 (63%)	0.00 (22%)	1.00 (15%)		
C14/C15 ancestor	Alternative	−31987.2	761	65512.7	1.0 (84%)	1.00 (15%)	100.00 (1.5%)	3.60	0.16
	Null	−31989.0	760	65514.2	1.00 (7.7%)	0.90 (0.0%)	1.00 (92%)		
C1/2 ancestor	Alternative	−31993.8	761	65526.0	0.0017 (100%)	0.036 (0.0%)	1.04 (0.0%)	N/A	1.00

In each analysis, all other branches in the phylogeny were set as the background to the test branch. The alternative model allows the test and background branches to each have three modelled classes of ω: ω^1, ω^2 and ω^3; the estimated proportion of sites within each class is shown in parentheses. The null model forces ω^3 to be equal to 1 (i.e. neutral evolution). Rejection of the null model in favour of the alternative model provides evidence of positive selection on the test branch. The AICc (corrected Akaike information criterion) is shown for comparison of model fit. See (Murrell *et al.* 2015) for more details on the BUSTED approach. For the C1/2 ancestor there were no sites in ω3 under the alternative model, so there is no evidence for positive selection.

tion was not detected for the ancestral branch leading into the alkalinophilic C1/2 lineage by either BS-REL or BUSTED ($P = 1$ and 1) (Tables 1 and 2).

To add to the above hypothesis tests, and for exploratory analysis, we considered all remaining branches with corrected *P*-values <0.05 in the BS-REL (uncorrected $P < 0.001$) as additional candidates for positive selection (Fig. 1, Table 1). This led to the inference of positive selection at six further nodes in the phylogeny (Fig. 1). Interestingly, represented among these was the deepest branch in the tree (corrected $P = 0.02$), which separates *Nitrososphaera* lineages from the ancestor to *Nitrosotalea* and *Nitrosopumilus* lineages and roughly correlates with the earliest pH adaptation events in thaumarchaeotal evolution and the onset of lineage diversification (Gubry-Rangin *et al.* 2015). At this node, a notable fraction of *amoA* codons were inferred to have experienced positive selection (proportion: 9%, mean $\omega = 73.01$), with all others being subject to strong purifying selection (Fig. 1, Table 1). Among the remaining five nodes inferred to have evolved under positive selection, four are from pH-adapted lineages, including three each from C11 and C14/15 and one from C1/C2. Despite this predominance of pH-adapted lineages, this finding does not deviate from random expectations as approximately half the lineages in the analysis come from the three abundant pH-adapted lineages ($P > 0.05$, Fishers exact test).

Tests for relaxed selection linked to pH adaptation

The BS-REL or BUSTED tests are not designed to identify instances where adaptive functions evolve by purifying selection acting on genetic variation arising during episodes of relaxed selection, often termed the Dykhuizen–Hartl effect (Kimura 1983; Hughes 2008; e.g. Macqueen *et al.* 2010). Therefore, to further dissect out the evolutionary signal in our data, we modelled d_N and d_S separately for each branch along the 370-sequence phylogeny using parametric bootstrapping (200 iterations) to establish variation in parameter estimates (Fig. 2). Instances of AmoA evolution involving the Dykhuizen–Hartl effect should be accompanied by elevated ω values in ancestral branches leading to pH-adapted lineages, with low ω values in subsequent daughter branches, indicative of purifying selection on associated functional changes. Unsurprisingly, while most branches had very low ω estimates, significant heterogeneity in ω was observed (Fig. 2). These branches were located across the tree without obvious bias towards pH-adapted lineages, with low ω values observed for the ancestral branches leading to C11, C14/15 and C1/2. However, we observed $\omega = 1.36$ (CI_{95}: 1.04, 1.94) for the ancestral branch leading to C14, consistent with a scenario involving either relaxed or positive selection or a combination of both.

As branch estimates of ω cannot easily distinguish between such possibilities, the 73-sequence *amoA* codon alignment was employed in another BS-REL framework called 'RELAX' that aims to detect relaxed selection for *a priori* defined branches and can reliably distinguish relaxation in selective pressure from intensification of purifying or positive selection (Wertheim *et al.* 2015). RELAX examines whether selection intensity increases relative to a defined background state, recaptured by a modelled parameter *K* (Wertheim *et al.* 2015). When *K* is >1 and <1 respectively, selection is inferred to have been intensified (through purifying or positive selection) or relaxed in the defined test branch. A LRT is then done comparing models where *K* is fixed to be

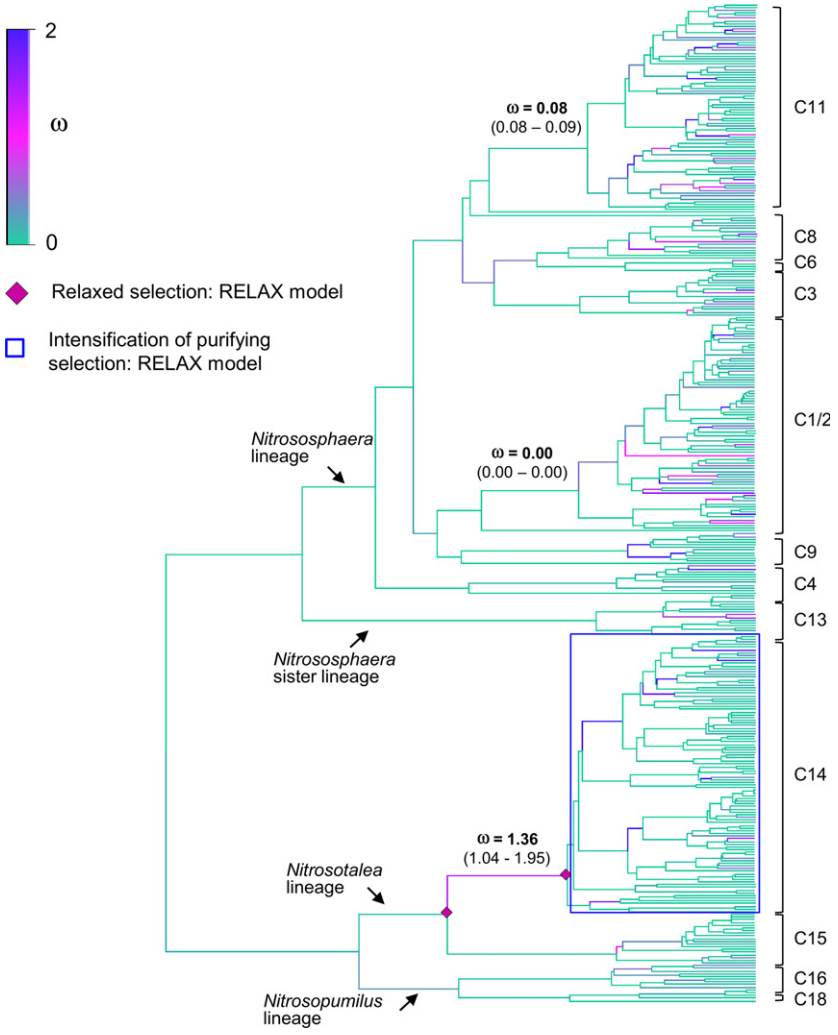

Fig. 2 Evidence for heterogeneity in selective pressures operating on the *amoA* coding sequence during thaumarchaeotal evolution, done by modelling ω for every branch in the 370 sequence phylogeny. Estimates of ω and 95% confidence intervals are shown for nodes ancestral to the three abundant pH-adapted lineages. The evidence gained from the RELAX test of an episodic relaxation in selective pressure in the ancestors of the C14 lineage that was followed by an immediate and persistent increase in purifying selection (see Table 3) is also highlighted.

equal across the tree (null model), or allowed to differ between test and background branches (alternative model). For the ancestral branch leading into C14, there was strong evidence in favour of a major relaxation in selective pressure ($K < 0.001$, $P < 0.00001$; Table 3; branch shown on Fig. 2) relative to the background rate across the tree. The RELAX test also indicated a relaxation in selective pressure in the ancestral branch leading into C14/15 ($K = 0.33$, $P = 0.0008$, Table 3, Fig. 2), again relative to the tree-wide background. The Dykhuizen–Hartl effect should lead to a scenario where purifying selection was then intensified in descendent C14 lineages, after the initial relaxation of selection in their ancestor. This prediction was scrutinized using RELAX by setting the two ancestral nodes to C14/15 and C14 as background branches and all internal C14 branches as test branches. This led to a highly significant result suggesting a strong intensification in selective pressure in C14 descendent lineages compared to their direct ancestors ($K = 4.57$, $P = 0.0001$; Fig. 2). When interpreting this result it is informative to consider a RELAX model where three ω rate classes were separately estimated for the C14/15 ancestor nodes and the C14 descendant branches (see Table 3). In this respect, 99.8% of codons from the C14 descendant branches belong to rate classes where ω = 0.027 indicative of strong purifying selection, whereas 15% of codons from the C14/15 ancestor nodes belong to a class where ω = 1.09, approximating neutral evolution, while no codons fit to a rate class suggesting positive selection. Thus, our hypothesis for the Dykhuizen–Hartl effect is strongly supported.

Unsurprisingly (considering the BS-REL results), there was no evidence for relaxed selection along the ancestral branch leading into the acidophilic C11 clade ($K = 0.63$, $P = 0.21$, Table 3), while in the alkalinophilic C1/2 cluster, there was evidence of an intensification in selective pressure relative to the background (rest of the tree) ($K = 49.1$, $P = 0.006$), which is likely to have resulted from an increase in purifying selection (Table 3).

Table 3 Results of the RELAX test for relaxed selective pressure associated with major pH adaptation events during Thaumarchaeota evolution

Test branch	Background branches	Model	Parameters	log*L*	AICc	*K*	LRT statistic	*P*-value
C14 ancestor	Rest of tree	Alternative	164	−12947.6	26227.0	0.001*	25.6	0.00001
		Null	163	−12960.4	26250.5			
C14/15 ancestor	Rest of tree	Alternative	164	−12954.7	26241.3	0.330†	11.3	0.00080
		Null	163	−12960.4	26250.5			
C14 descendent lineages	C14 ancestor and C14/15 ancestor	Alternative	164	−12832.3	26221.3	4.570‡	16.4	0.00001
		Null	163	−12840.5	26235.6			
C11 ancestor	Rest of tree	Alternative	164	−12959.9	26251.6	0.630	1.6	0.21000
		Null	163	−12960.6	26251.2			
Cl/2 ancestor	Rest of tree	Alternative	164	−12957.0	26245.9	49.060§	7.3	0.00700
		Null	163	−12960.6	26251.1			

*–§(shown below) provides data from a RELAX model (the 'partitioned descriptive model', see Wertheim *et al.* 2015) where three ω distributions were estimated separately for the test and background branches. The data are an estimate of ω in three rate classes, ω^1, ω^2 and ω^3. The estimated proportion of sites fitting each class is given in parentheses.
*Background branches: $\omega^1 = 0.000726$ (pr. 96.0%); $\omega^2 = 0.899$ (pr. 3.4%); $\omega^3 = 1190$ (pr. 0.29%).
Test branches: $\omega^1 = 1.00$ (pr. 17.0%); $\omega^2 = 1.00$ (pr. 24.0%); $\omega^3 = 1.00$ (pr. 59.0%).
†Background branches: $\omega^1 = 0.000107$ (pr. 96%); $\omega^2 = 1.00$ (pr. 3.4%); $\omega^3 = 1090$ (pr. 0.28%).
Test branches: $\omega^1 = 0.0182$ (pr. 89%); $\omega^2 = 0.741$ (pr. 1.0%); $\omega^3 = 1.70$ (pr. 9.9%).
‡Background branches: $\omega^1 = 0.00$ (pr. 85.0%); $\omega^2 = 0.854$ (pr. 0.0%); $\omega^3 = 1.09$ (pr. 15.0%).
Test branches: $\omega^1 = 0.0273$ (pr. 95.0%); $\omega^2 = 0.0273$ (pr. 4.4%); $\omega^3 = 1000$ (pr. 0.2%).
§Background branches: $\omega^1 = 0.000839$ (pr. 96%); $\omega^2 = 0.966$ (pr. 3.6%); $\omega^3 = 746$ (pr. 0.28%).
Test branches: $\omega^1 = 0.00290$ (pr. 100%); $\omega^2 = 0.00358$ (pr. 0.0%); $\omega^3 = 10\ 000$ (pr. 0.0%).

AmoA protein evolution during ancestral pH adaptation events

Ancestral reconstruction of all nodes in the 370-sequence *amoA* phylogeny was used to explore protein-level evolution associated with pH adaptation, focusing on amino acid changes arising during the distinct selective regimes at the stem of the three major pH-adapted lineages, with notable changes for the two major acidophilic lineages (Figs 3 and 4). The reconstructions were done at the codon level using a model fully congruent with the BS-REL test (see methods). According to this approach, eight amino-acid replacements were fixed along the ancestral *Nitrososphaera* C11 branch which subsequently remained invariant throughout the radiation of C11 lineage (Fig. 3a, b), which has spanned hundreds of millions of years (Gubry-Rangin *et al.* 2015). According to models predicting AmoA structure within the lipid membrane, these sites are invariably localized within the transmembrane region, often clustered near the membrane boundary (Fig. 3d). While seven replacements led to a functionally similar hydrophobic amino acid, one change within the last transmembrane region, Gly to Ser, modified the residue property radically (Fig. 3c). Among the conservative changes, it is notable that three independent replacements of isoleucine to valine occurred in the C11 ancestor.

More extreme evolution of the AmoA protein was associated with acidophilic specialization in *Nitrosotalea*, including 10 amino acid replacements in the ancestor to C14/C15 and nine along the ancestral branch to C14 that have remained fixed in these lineages (Fig. 4 a, b). In contrast with C11, many involved a radical amino acid change. However, as observed in the C11 ancestor, several conservative changes occurred involving a range of nonpolar/hydrophobic residues (with no obvious drive towards a specific amino acid) likewise within the predicted membrane region (Fig. 4a, c). Interestingly, four amino acid substitutions in the C14/C15 ancestor were within the pseudo-periplasmic region, with two radical changes away from residues with nonpolar side chains to Asp, which has an acidic side chain (Fig. 4a, c). It is also notable that changes both in C14/15 ancestor and C14 ancestor involved repeated substitutions that replaced residues that cannot undergo phosphorylation with Ser, Thr or Tyr (i.e. all of which can undergo phosphorylation) (Fig. 4a, c). These changes were predicted to be located within the membrane, pseudo-periplasmic and cytoplasmic regions. Reciprocally, there were no substitutions away from Ser, Thr and Tyr leading into C14/15 or C14 (Fig. 4c). All the amino acid replacements arising in the C14 ancestor were localized within the lipid membrane, including a cluster at the boundary between the first transmembrane domain and the pseudo-periplasm.

Therefore, in addition to the nature of selection associated with AmoA evolution in the two abundant aci-

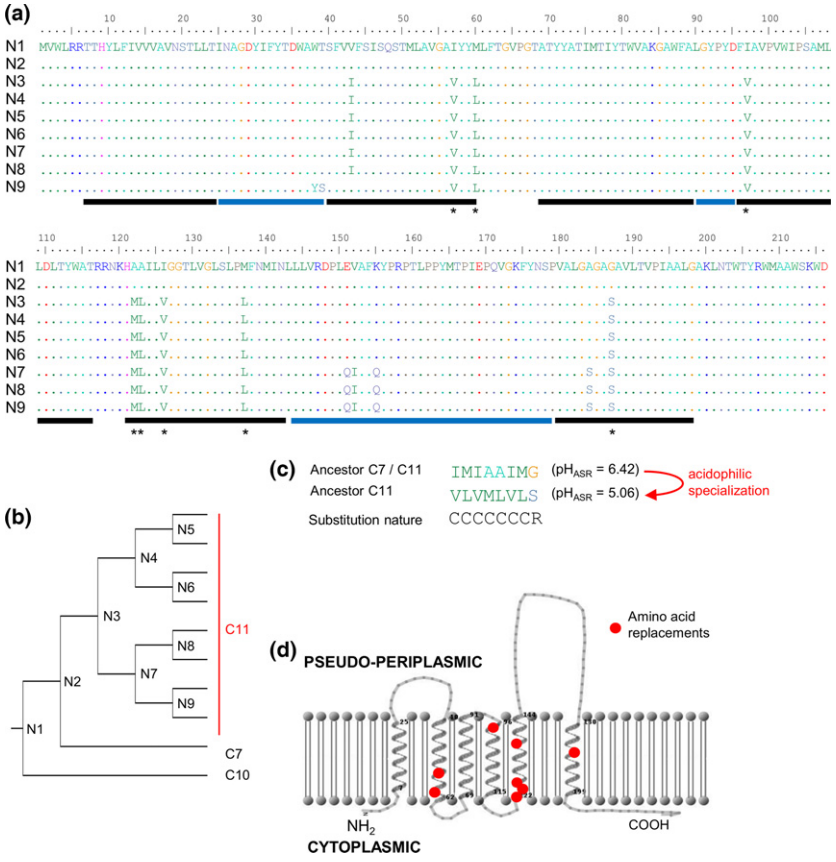

Fig. 3 Ancestral reconstruction of sequence changes fixed in the AmoA protein during the evolutionary transition to acidophily in the C11 lineage of *Nitrososphaera*. An alignment of AmoA primary sequences is shown for a range of reconstructed C11 lineages (including the common ancestor) as well as for lineages immediately ancestral to C11 (a). Fixed changes arising during the transition to acidophily are highlighted by asterisks. Coloured bars highlight the position of amino acids according to predicted membrane modelling (blue = pseudo-periplasmic, black = transmembrane, none = cytoplasmic). Also shown is a phylogeny of the highlighted lineages (b), along with the nature of amino acid substitution (C = conservative, R = radical) (c). Finally, amino-acid changes inferred to have persistent adaptive value in C11 lineages are shown on a predicted 2D membrane model of AmoA (d).

dophilic lineages, the position and implications of amino acid changes are evidently quite distinct. In contrast with C11 and C14/15, ancestral reconstruction suggested that no changes in the AmoA protein occurred during the evolutionary transition to alkalinophily along the ancestral branch leading to C1/2 (not shown).

Discussion

There is currently a dearth of information on the mechanisms underlying environmental adaptation in Thaumarchaeota, which are hugely abundant on Earth and have vital ecosystem functions (Leininger *et al.* 2006; Wuchter *et al.* 2006). The same is true generally, in that there exists a large gap in understanding about the processes linking ecology to molecular evolution in environmental microbes. Accordingly, our study has exploited approaches commonly used to infer patterns of molecular evolution to reveal mechanisms of past adaptation in Thaumarchaeota, testing the hypothesis that ancestral phenotypic adaptation was coupled to adaptive changes in a gene crucial for ammonia oxidation.

While our findings are consistent with the single past study that considered codon-level evolution of AmoA in Thaumarchaeota, which revealed persistent purifying selection across a broad habitat range at the whole gene level (Biller *et al.* 2012), our study markedly extends this work. Specifically, we aimed to identify variation in the rates and types of evolution operating across different thaumarchaeotal lineages and in different regions of the AmoA protein. Moreover, we directly tested if phenotypic adaptation to environmental pH was recaptured in AmoA molecular evolution. As far as we are aware, such detailed attempts to test for an association between ecological adaptation and molecular evolution is unique in Thaumarchaeota and rare for environmental prokaryotes generally.

Our main hypothesis was supported by evidence of episodic shifts in AmoA evolution during independent adaptation events leading to radiations into acidic environments. Considering the function of AmoA, which is to catalyse the aerobic oxidation of ammonia into hydroxylamine (Vajrala *et al.* 2013), it would be speculative to assume that changes in this protein were essential for initial evolutionary entrance to acidic environments. However, considering its vital role for energy production, the protein can be assumed to have been required for the cell to function properly, hence allowing long-term persistence at low pH. By extrapolation

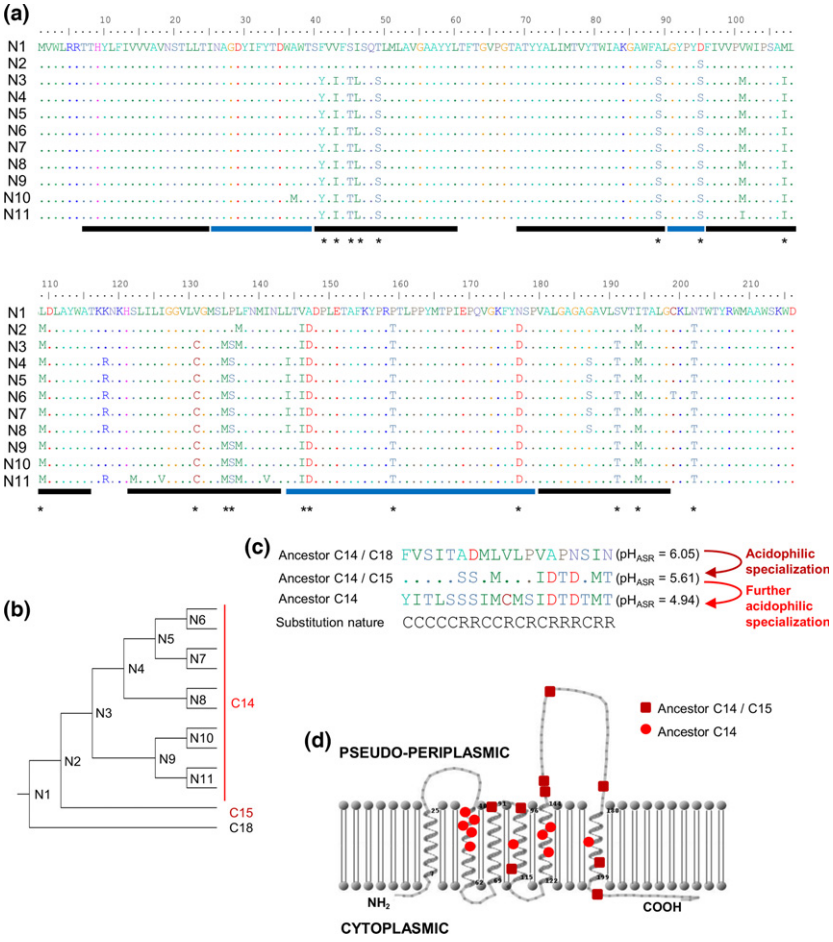

Fig. 4 Reconstruction of AmoA protein sequence evolution during a major evolutionary transition to acidophily in the C14 lineage of *Nitrosotalea*. All details are as provided in the Fig. 3 except with respect to the evolution of the C14 lineage.

from other extremophilic archaea (see references within the introduction), the C11 and C14/15 ancestors likely had to increase the number of GDGT cyclopentane rings within their lipid membranes to reduce proton influx at low pH and maintain cell homeostasis. In addition, if we assume that the ancestors of these major acidophilic lineages predominantly catalysed ammonia (NH_3) and not ammonium (NH_4^+) (Lehtovirta-Morley *et al.* 2016), AmoA function would have been further challenged by a reduction in the availability of its substrate at low pH (NH_3 is 1000 times less available at pH 4 than at pH 8). Therefore, our results probably recapture evolutionary changes necessary to maintain AmoA function under radically altered (both intracellular and extracellular) conditions moving from neutral to acidic pH. Such changes in membrane conformation and substrate availability would not be expected under the increased alkaline conditions faced by the C1/2 ancestor, which is consistent with the evident lack of AmoA evolution during this major pH adaptation event. However, our findings are currently limited, as the importance of residue changes in the ancestral C11 and C14/C15 lineages will remain elusive until experimental functional tests can be performed on ancestrally reconstructed AmoA proteins (e.g. Dean & Thornton 2007), which is yet to be achieved in Thaumarchaeota.

Nonetheless, our data suggest that two independent evolutionary transitions leading to abundant acidophilic Thaumarchaeota lineages were associated with nonconvergent modes of AmoA evolution. In the branch leading to C14/C15, extensive AmoA evolution occurred during a period where selective pressure was relaxed. While it is impossible to determine what underlies this relaxation in selection, a plausible possibility is a population bottleneck that reduced effective population size and hence the efficiency of natural selection (e.g. Lynch 2007). One mechanism by which population size could have been reduced is the existence of an initial period where a small pool of poorly adapted C14/15 ancestors survived at moderately low pH. Under this scenario, neutral or mild deleterious genetic changes fixed under inefficient purifying selection, such as those observed in AmoA, gained adaptive value at low pH, a scenario akin to the so-called Dykhuizen–Hartl effect (Kimura 1983; Hughes 2008). Consistent with this hypothesis, it has been previously suggested that episodes of relaxed

selection might be an important route by which microbes increase metabolic versatility in extreme environments and are ultimately able to adapt (e.g. Konstantinidis *et al.* 2009; Li *et al.* 2014; further considered later in this discussion). In any case, this hypothesis predicts that other protein-coding genes would have experienced a similar relaxation along the same ancestral C14/15 branches, which can be tested in the future. By contrast, positive selection acting in a background of very strong purifying selection was evidently the driving force behind a more limited episode of AmoA evolution in the ancestral C11 node. This indicates that the C11 ancestors likely had a large effective population size, allowing efficient fixation of adaptive mutations by natural selection during pH adaptation.

In addition to nonconvergent evolutionary mechanisms, the functional nature of protein evolution during pH adaptation was distinct comparing the C14/15 and C11 radiations. While conservative substitutions of hydrophobic amino acids are a feature in both cases, pH adaptation in C14/15 was more strongly coupled to a directional drive towards an increase in Ser, Thr and Tyr. While no role for phosphorylation has yet been ascribed to AmoA, the importance of this post-translational modification is well-established in prokaryotes (Cozzone 1988). Reversible protein phosphorylation occurs through conserved protein kinases that transfer phosphate groups from ATP onto acceptor amino acids – this modification serves to regulate protein structure and protein–protein interactions and has particularly important roles in conserved signalling pathways. Several Ser/Thr/Tyr kinases are conserved in archaeal genomes, including transmembrane kinases (Krupa & Srinivasan 2005) and Ser/Thr are known to undergo phosphorylation in these microbes (e.g. Skorko 1984). Therefore, we speculate that the increased fixation of residues involved in phosphorylation during initial pH adaptation in the C14/C15 lineage acted to modify the interaction of AmoA with other proteins, perhaps under different conditions regulated by protein kinases. The reason why such changes were required more in C14/15 than C11 remains a mystery, but might be explained by differences in genetic background or AmoA function comparing the distantly related acidophilic lineages.

It is also important to note that adaptive mechanisms associated with microbial adaptation are likely coded within multiple genes and may be driven by lateral gene transfer. Hence, future genome and transcriptome-wide analyses of environmental samples will be required to fully understand the molecular-evolutionary mechanisms underlying pH specialization in Thaumarchaeota. Such approaches are currently difficult considering the limited cultivation of these organisms. In this sense, it is important to note that two strains within *Nitrosotalea* C14 lineage have recently been successfully isolated (Lehtovirta-Morley *et al.* 2014); thus, there now exists an urgent need to cultivate a C11 strain, allowing the physiological and genomic characteristics of each pH-adapted group to be compared at depth.

Finally, when considering how our findings fit within the broader field of microbial ecology and evolution, past research has similarly shown that the environment can shape patterns of natural selection and evolution in microorganisms. However, the approaches used and their overall goals have been distinct to our attempts to reconstruct the selective regimes underlying the evolution of an important functional protein during environmental adaptation. For example, using the powerful approach of studying microbial evolution over experimental generations (e.g. Barrick *et al.* 2009; McDonald *et al.* 2009), Tenaillon *et al.* 2012 gained genome-wide insights into processes allowing *E. coli* to adapt to temperature, revealing importance for convergent mutations at the level of genes and gene functions, but not specific amino acids. This is partly consistent with our conclusions of convergent adaptation of AmoA to low pH in distant thaumarchaeotal lineages, which involved distinct amino acid changes. Nonetheless, it is difficult to directly compare such contrasting approaches, particularly as C11 and C14 are separated by hundreds of millions of years, where extensive convergence would be more surprising. Moreover, past studies identifying mutations over hundreds/thousands of generations are not easy to compare with phylogenetic reconstructions of environmentally adaptive substitutions that have been fixed over hundreds of millions of years. In terms of past studies that have explored the link between environmental and molecular adaptation in microbes over large evolutionary timescales, most have aimed to understand broad evolutionary trends, for example: how temperature drives evolutionary rate in archaea (Groussin & Gouy 2011), how salinity is linked to the evolution of base composition (e.g. Luo *et al.* 2015), which evolutionary processes are linked to adaptation to the deep ocean at the metagenomic scale (Konstantinidis *et al.* 2009) and how evolutionary rate differs across microbial communities comparing extreme and more normal habitats (Li *et al.* 2014). A common theme of such studies consistent with our data is frequent evidence for the role of relaxed selection in microbial evolution, especially during evolutionary transitions to different environments (e.g. Konstantinidis *et al.* 2009; Li *et al.* 2014; Luo *et al.* 2015). While these past studies are extremely valuable, we feel there is an ongoing need to complement understanding of broad evolutionary trends with in depth reconstructions of gene evolution and environmental adaptation, as performed here.

In conclusion, we hope our study will stimulate further work integrating approaches and data sets produced by microbial ecologists with the latest computational models in molecular evolution. Such an approach has offered insights into mechanisms of physiological adaptation in environmental Thaumarcheota. It is highly likely that similar opportunities exist to define the molecular-evolutionary underpinnings of ecological success and ecosystem functions in numerous other microbe lineages.

Acknowledgements

This work was funded by Natural Environment Research Council Fellowship NE/J019151/1 and by institutional funding from within the University of Aberdeen.

References

Allen EE, Tyson GW, Whitaker RJ, Detter JC, Richardson PM, Banfield JF (2007) Genome dynamics in a natural archaeal population. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1883–1888.

Barrick JE, Yu DS, Yoon SH *et al.* (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, **461**, 1243–1249.

Bédard C, Knowles R (1989) Physiology, biochemistry, and specific inhibitors of CH_4, NH_4^+, and CO oxidation by methanotrophs and nitrifiers. *Microbial Reviews*, **53**, 68–84.

Biller SJ, Mosier AC, Wells GF, Francis CA (2012) Global biodiversity of aquatic ammonia-oxidizing archaea is partitioned by habitat. *Frontiers in Microbiology*, **3**, 252.

Boyd ES, Pearson A, Pi Y *et al.* (2011) Temperature and pH controls on glycerol dibiphytanyl glycerol tetraether lipid composition in the hyperthermophilic crenarchaeon *Acidilobus sulfurireducens*. *Extremophiles*, **5**, 59–65.

Boyd ES, Hamilton TL, Wang J, He L, Zhang CL (2013) The role of tetraether lipid composition in the adaptation of thermophilic archaea to acidity. *Frontiers in Microbiology*, **4**, 62.

Cozzone AJ (1988) Protein Phosphorylation in Prokaryotes. *Annual Review of Microbiology*, **42**, 97–125.

Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nature Reviews Genetics*, **8**, 675–688.

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, **26**, 2455–2457.

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 626–631.

Groussin M, Gouy M (2011) Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Molecular Biology and Evolution.*, **28**, 2661–2674.

Gubry-Rangin C, Hai B, Quince C *et al.* (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 21206–21211.

Gubry-Rangin C, Kratsch C, Williams TA *et al.* (2015) Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 9370–9375.

Hatzenpichler R (2012) Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Applied and Environmental Microbiology*, **78**, 7501–7510.

He M, Miyajima F, Roberts P *et al.* (2013) Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nature Genetics*, **45**, 109–113.

Hughes AL (2008) The origin of adaptive phenotypes. *Proceedings of the National Academy of Sciences of the United States of America*, **36**, 13193–13194.

Hyman MR, Arp DJ (1992) $^{14}C_2H_2$- and $^{14}CO_2$-labeling studies of the *de novo* synthesis of polypeptides by *Nitrosomonas europaea* during recovery from acetylene and light inactivation of ammonia monooxygenase. *The Journal of Biological Chemistry*, **267**, 1534–1545.

Kennemann L, Didelot X, Aebischer T *et al.* (2011) *Helicobacter pylori* genome evolution during human infection. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5033–5038.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Konstantinidis KT, Braff J, Karl DM, DeLong EF (2009) Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Applied and Environmental Microbiology*, **75**, 5345–5355.

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delport W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*, **28**, 3033–3043.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, **305**, 567–580.

Krupa A, Srinivasan N (2005) Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. *BMC Genomics*, **6**, 129.

Lehtovirta-Morley LE, Ge C, Ross J, Yao H, Nicol GW, Prosser JI (2014) Characterisation of terrestrial acidophilic archaeal ammonia oxidisers and their inhibition and stimulation by organic compounds. *FEMS Microbiology Ecology*, **89**, 542–552.

Lehtovirta-Morley LE, Sayavedra-Soto LA, Galois N *et al.* (2016) Identifying potential mechanisms enabling acidophily in the ammonia-oxidising archaeon '*Candidatus Nitrosotalea devanaterra*'. *Applied and Environmental Microbiology*, Advance Online Publication, doi:10.1128/AEM.04031-15.

Leininger S, Urich T, Schloter M *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, **442**, 806–809.

Li SJ, Hua ZS, Huang LN *et al.* (2014) Microbial communities evolve faster in extreme environments. *Scientific Reports*, **4**, 6205.

Lieberman RL, Rosenzweig AC (2005) Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature*, **434**, 177–182.

López de Saro FJ, Díaz-Maldonado H, Amils R (2015) Microbial evolution: the view from the acidophiles. In: *Microbial*

Evolution Under Extreme Conditions (ed. Bakermans C), pp. 19–30. De Gruyter, Berlin, Germany.

Luo H, Thompson LR, Stingl U, Hughes AL (2015) Selection maintains low genomic GC content in marine SAR11 lineages. *Molecular Biology and Evolution*, **32**, 2738–2748.

Lynch M (2007) *The Origins of Genome Architechture*. Sinauer Associates Inc, Sunderland, Massachusetts.

Macalady JL, Vestling MM, Baumler D, Boekelheide N, Kaspar CW, Banfield JF (2004) Tetraether-linked membrane monolayers in *Ferroplasma* spp: a key to survival in acid. *Extremophiles*, **8**, 411–419.

Macqueen DJ, Delbridge ML, Manthri S, Johnston IA (2010) A newly classified vertebrate calpain protease, directly ancestral to CAPN1 and 2, episodically evolved a restricted physiological function in placental mammals. *Molecular Biology and Evolution*, **27**, 1886–1902.

McDonald MJ, Gehrig SM, Meintjes PL, Zhang XX, Rainey PB (2009) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics*, **183**, 1041–1053.

McTavish H, Fuchs JA, Hooper AB (1993) Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *Journal of Bacteriology*, **175**, 2436–2444.

Murrell B, Weaver S, Smith MD *et al.* (2015) Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, **32**, 1365–1371.

Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, **11**, 715–724.

Nicol GW, Leininger S, Schleper C, Prosser JI (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia-oxidizing archaea and bacteria. *Environmental Microbiology*, **10**, 2966–2978.

Pearson A, Pi Y, Zhao W *et al.* (2008) Factors controlling the distribution of archaeal tetraethers in terrestrial hot springs. *Applied and Environmental Microbiology*, **74**, 3523–3532.

Pester M, Schleper C, Wagner M (2011) The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Current Opinion in Microbiology*, **14**, 300–306.

Pester M, Rattei T, Flechl S *et al.* (2012) *amoA*-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions. *Environmental Microbiology*, **14**, 525–539.

Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**, 676–679.

Pupko T, Pe'er I, Shamir R, Graur D (2000) A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Molecular Biology and Evolution*, **17**, 890–896.

Rotthauwe JH, Witzel KP, Liesack W (1997) The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Applied and Environmental Microbiology*, **63**, 4704–4712.

Sára M, Sleytr UB (2000) S-Layer proteins. *Journal of Bacteriology*, **182**, 859–868.

Schloissnig S, Arumugam M, Sunagawa S *et al.* (2013) Genomic variation landscape of the human gut microbiome. *Nature*, **493**, 45–50.

Skorko R (1984) Protein phosphorylation in the archaebacterium *Sulfolobus acidocaldarius*. *European Journal of Biochemistry*, **145**, 617–622.

Snitkin ES, Zelazny AM, Montero CI *et al.* (2011) Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 13758–13763.

Spyropoulos IC, Liakopoulos TD, Bagos PG, Hamodrakas SJ (2004) TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics*, **20**, 3258–3260.

Tai V, Poon AF, Paulsen IT, Palenik B (2011) Selection in coastal Synechococcus (cyanobacteria) populations evaluated from environmental metagenomes. *PLoS ONE*, **6**, e24249.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2759.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL *et al.* (2012) The molecular diversity of adaptive convergence. *Science*, **335**, 457–461.

Thakur S, Normand P, Daubin V, Tisa LS, Sen A (2013) Contrasted evolutionary constraints on secreted and non-secreted proteomes of selected Actinobacteria. *BMC Genomics*, **14**, 474.

Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS (2013) Tangled bank of experimentally evolved *Burkholderia* biofilms reflects selection during chronic infections. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 250–259.

Vajrala N, Martens-Habbena W, Sayavedra-Soto LA *et al.* (2013) Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1006–1011.

Vico Oton E, Quince C, Nicol GW, Prosser JI, Gubry-Rangin C (2015) Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *The ISME Journal*, **10**, 85–96.

Walker CB, de la Torre JR, Klotz MG *et al.* (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8818–8823.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, **32**, 820–832.

Wuchter C, Abbas B, Coolen MJ *et al.* (2006) Archaeal nitrification in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12317–12322.

Xie W, Zhang CL, Wang J *et al.* (2015) Distribution of ether lipids and composition of the archaeal community in terrestrial geothermal springs: impact of environmental variables. *Environmental Microbiology*, **17**, 1600–1614.

Both authors designed and performed the research, analysed data and equally contributed to the writing of the manuscript.

Data accessibility

Codon sequence alignments and phylogenetic trees used in evolutionary analyses of selective pressure including tests for positive or relaxed selection: provided as Supporting Information on the Molecular Ecology Website (Data S1 and S2, Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1. AmoA 370-sequence codon alignment and phylogenetic tree.

Data S2. AmoA_73-sequence codon alignment and phylogenetic tree.