The unique coding sequence of *pmoCAB* operon from type Ia methanotrophs simultaneously optimizes transcription and translation

Juan C. Villada¹, Maria F. Duran¹, and Patrick K. H. Lee^{1*}

¹School of Energy and Environment, City University of Hong Kong, Hong Kong SAR

*Correspondence: B5423, Yeung Kin Man Academic Building, School of Energy and Environment, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China; E-mail: patrick.kh.lee@cityu.edu.hk; Tel: (852) 3442-4625; Fax: (852) 3442-0688.

Understanding the interplay between genotype and phenotype is a fundamental goal of functional genomics. Methane oxidation is a microbial phenotype with global-scale significance as part of the carbon biogeochemical cycle, and is a sink for greenhouse gas. Microorganisms that oxidize methane (methanotrophs) are taxonomically diverse and widespread around the globe. Recent reports have suggested that type Ia methanotrophs are the most prevalent methane-oxidizing bacteria in different environments. In methanotrophic bacteria, complete methane oxidation is encoded in four operons (*pmoCAB*, *mmoXYZBCD*, *mxoFI*, and *xoxF*), but how evolution has shaped these genes to execute methane oxidation remains poorly understood. Here, we used a genomic meta-analysis to investigate the coding sequences that encode methane oxidation. By analyzing isolate and metagenome-assembled genomes from phylogenetically and geographically diverse sources, we detected an anomalous nucleotide composition bias in the coding sequences of particulate methane monooxygenase genes (*pmoCAB*) from type Ia methanotrophs around the globe. We found that this was a highly conserved sequence that optimizes codon usage in order to maximize translation efficiency and accuracy, while minimizing the synthesis cost of transcripts and proteins. We show that among the seven types of methanotrophs, only type Ia methanotrophs possess a unique coding sequence of the *pmoCAB* operon that is under positive selection for optimal resource allocation and efficient synthesis of transcripts and proteins in environmental counter gradients with high oxygen and low methane concentrations. This adaptive trait possibly enables type Ia methanotrophs to respond robustly to fluctuating methane availability and explains their global prevalence.

Keywords *Methane monooxygenase · Codon usage · Translation efficiency · Synthesis cost · Resource allocation*

Introduction

Microbial methane oxidation plays a number of fundamental roles in the global ecosystem¹. Methane-oxidizing microorganisms can mitigate methane emissions by acting as methane sinks^{2,3} and thereby reduce the contribution of methane to climate change⁴. Microbial oxidation of methane also provides an entry point for methane into the global food web, and can serve as a primary carbon source for large trophic systems^{5,6}. Methane oxidation is a globally distributed phenotype expressed in microorganisms from diverse taxonomic groups. Based on a range of phenotypic and phylogenetic features, methanotrophic bacteria are categorized into seven major types⁷: Ia, Ib, Ic, IIa, IIb, III, and the candidate division NC10. The distribution of marker gene sequences for the major methanotroph types suggests they are differentially prevalent across environments⁷. In bacteria, methane oxidation involves two enzymatic steps where methane is first converted to methanol by a soluble (sMMO) or particulate (pMMO) methane monooxygenase, then methanol is oxidized to formaldehyde by a pyrroloquinoline quinone-containing methanol dehydrogenase that can be calcium-dependent (Mxa) or lanthanide-dependent (Xox). The resulting formaldehyde can be directed to energy production or biomass synthesis⁸. sMMO is encoded in the six-gene operon *mmoXYZBCD*, while pMMO is encoded in the three-gene operon *pmoCAB*. Of the methanol dehydrogenases, Mxa is encoded in the two-gene operon *mxoFI*, and Xox by the gene *xoxF*.

In evolutionary history, methane oxidation appeared at around the same time as oxygenic photosynthesis, nitrogen fixation, nitrification and denitrification⁹, and it is possible that the emergence of methanotrophy occurred soon after the last universal common ancestor¹⁰. Hence, evolution has likely shaped methanotrophs with many as yet undiscovered properties¹⁰. One unexplored question of fundamental importance to our understanding of methanotrophy is how the genes that encode the methane oxidation metabolic module have been shaped by evolution to efficiently execute methanotrophy. In the three cellular domains and in viruses, evolution has selected gene sequences that perform cellular functions beyond just encoding the amino acid composition of proteins. For example, gene sequences have nucleotide composition and codon usage that direct mRNA folding^{11,12},

transcript abundance¹³, mRNA degradation¹⁴, RNA toxicity¹⁵, protein synthesis¹⁶, cotranslational protein folding¹⁷, and promote the interaction of peptides with the Signal Recognition Particle¹⁸. Gene sequences can also affect the cellular economy of protein synthesis¹⁹, reduce the metabolic burden of nucleotide synthesis by incorporating less expensive nucleotides²⁰, and allocate resources required for transcription²¹ and translation²².

In this study, we performed an in-depth genomic meta-analysis to investigate the coding sequences of the genes that encode the methane oxidation metabolic module in bacteria. We found that evolution has shaped the *pmoCAB* operon of type Ia methanotrophs with a unique coding sequence that optimizes resource allocation by reducing the biosynthetic costs of transcription and translation while ensuring translation efficiency and accuracy. This study provides novel insights into the molecular biology and evolution of methanotrophic bacteria, and extends our understanding of the mechanisms developed by nature to sustain metabolism and life on Earth.

Results

Meta-analysis of methanotroph genomes. To investigate the coding sequence of genes encoding the methane oxidation metabolic module, we analyzed 67 methanotrophic bacterial genomes including 49 isolate genomes and 18 metagenome-assembled genomes (MAGs) (Additional file 1: Figure S1, Additional file 2). These genomes originated from different sampling locations around the globe (Additional file 1: Figure S1a) and were phylogenetically diverse (Additional file 1: Figure S1b), comprising 19 genera from five families (Additional file 2) and with GC contents ranging from 40% (*Methylobacterium thermophilum* Kam1) to 65% (*Methylobacterium trichosporium* OB3b). The 67 methanotroph genomes were categorized into the seven major types⁷: Ia ($n = 36$), Ib ($n = 6$), Ic ($n = 3$), IIa ($n = 11$), IIb ($n = 5$), III ($n = 4$), and NC10 ($n = 2$).

Type Ia *pmoCAB* coding sequences have an anomalous nucleotide composition bias. We measured the distribution of nucleotide composition bias of all coding sequences in each genome (Additional file 1: Figure S1b). The guanine + cytosine (GC) content of every coding sequence was calculated, as well

Fig. 1 Nucleotide composition bias of *pmoCAB* coding sequences. An extended analysis of all coding sequences together with controls is shown in Additional file 1: Figure S2.

as the GC content in the third nucleotide of each codon (GC_3). In genomes of methanotroph types Ia, Ib, IIa, IIb, and NC10, the GC_3 content of coding sequences tended to be higher than the GC, while in type III genomes GC_3 tended to be lower than GC (Additional file 1: Figure S1b). The distributions in type Ic genomes were variable, and the non-methanotrophic bacterium *Bacteroides ovatus*, included as an outgroup, had no difference between GC_3 and GC.

Next, we analyzed the nucleotide composition bias of the 12 coding sequences in the four operons forming the methane oxidation metabolic module (Additional file 1: Figure S2). All coding sequences from all seven types showed patterns of relative GC/ GC_3 bias consistent with the whole genomes (Additional file 1: Figure S1), except for *pmoCAB* coding sequences of type Ia methanotrophs where GC_3 content was lower than GC (Fig. 1). The *pxmABC* and *hao* coding sequences were selected as controls as they appear in most of the methanotroph genomes and are not part of the methane oxidation metabolic module. As expected, these coding sequences showed a pattern of relative GC/ GC_3 bias consistent with the whole genomes (Additional file 1: Figure S2). Similar to the type Ia *pmoCAB*, the type III coding sequences tended to have lower GC_3 content compared to GC (Additional file 1: Figure S2), consistent with the whole-genome pattern in type III (Additional file 1: Figure S1).

The anomalous nucleotide composition of type Ia *pmoCAB* coding sequences provides an atypical codon usage bias. The anomalous nucleotide composition of type Ia *pmoCAB* coding sequences led us to hypothesize that this may generate an atypical bias in codon usage. To test whether the codon usage differs among coding sequences in the methane oxidation metabolic module of each methanotroph, we first calculated the relative synonymous codon usage (RSCU) in every coding sequence (controls were also included). On ordination of the RSCU values of type Ia coding sequences (Additional file 1: Figure S3a), *pmoCAB* coding sequences clustered separately (Euclidean distance PERMANOVA, $F = 36.55$, $p < 0.01$) and the effect of source operon in the separation was strong ($R^2 = 0.37$). A reduced effect of source operon was observed in the clustering pattern for coding sequences of types Ib ($F = 3.82$, $p < 0.01$, $R^2 = 0.21$) and IIa ($F = 12.67$, $p < 0.01$, $R^2 = 0.24$), while weak effects ($R^2 < 0.2$) or non-significant probabilities ($p \geq 0.01$) were observed for types Ic, IIb, III and NC10 (Additional file 1: Figure S3).

This result indicates that the type Ia *pmoCAB* coding sequences have an atypical codon usage bias compared to other coding sequences in the methanotrophy module. We hypothesized that this may be an adaptation to allow for higher expression of these genes. To test this, we calculated a Codon Adaptation Index (CAI) that compared the codon usage of a coding sequence to the codon usage of a reference set of sequences from the same genome. The first CAI, defined as CAI_{genome} , used

the codon frequency of all coding sequences in the genome as a reference, while the second CAI, defined as $CAI_{ribosome}$, used the codon frequency of only ribosomal protein genes as a reference. $CAI_{ribosome}$ thus served as a proxy to represent the codon usage of highly expressed genes. The comparison between the CAI_{genome} and $CAI_{ribosome}$ showed that *mxoF* coding sequences tended to have similar codon usage to both reference sets in all types of methanotrophs (Additional file 1: Figure S2a), which was also observed for *mxoI* and *soxF* coding sequences in types Ib and Ic. Type Ia *mxoI* coding sequences tended to deviate from the codon bias of ribosomal protein genes and towards the genome-scale codon usage pattern, while the opposite was observed for type IIa and IIb *mxoI* coding sequences. Type Ia *soxF* coding sequences appeared in the top percentile ranks of $CAI_{ribosome}$ and were less similar to the genomic codon usage. A similar but much stronger pattern was observed in type Ia *pmoCAB* coding sequences where they appeared in the bottom 10% of CAI_{genome} values, but at much higher percentile ranks for $CAI_{ribosome}$ (Additional file 1: Figure S2a). The results for all coding sequences in the methane oxidation module and controls are given in Additional file 1: Figure S4a. Overall, the extreme deviation of type Ia *pmoCAB* coding sequences from the genomic norm shows that they have a codon usage more similar to that of the ribosomal protein-coding sequences.

This deviation of codon usage from the genomic signature could be due to a higher codon usage bias in *pmoCAB* coding sequences. To test this, we computed the effective number of codons (ENC) of every coding sequence in all type Ia methanotrophs. ENC measures the number of synonymous codons used to encode amino acids. Thus, the ENC is 61 if a coding sequence uses all codons equally to produce all twenty standard amino acids, whereas the ENC is 20 if only one codon is used per amino acid (i.e. a high codon usage bias). Contrary to our hypothesis, the ENC of type Ia *pmoCAB* coding sequences was not significantly different (two-tailed Wilcoxon signed rank-test, 99% confidence level) to that of *mxoFI* and *soxF* coding sequences (Additional file 1: Figure S2b), while the ENC of *pmoCAB* coding sequences was significantly lower than coding sequences of ribosomal proteins and whole genomes (Additional file 1: Figure S4b). Overall, this indicates that even though the codons in coding sequences of *pmoCAB* and ribosomal proteins tend to be similar (Fig. 2a), the codon usage bias of *pmoCAB* coding sequences is higher. We hypothesized that if the type Ia *pmoCAB* coding sequences have a codon usage bias similar to other coding sequences in the methane oxidation metabolic module (Fig. 2b) but their codon usage frequencies are divergent (Additional file 1: Figure S2a), there should be a clear association between the level of codon bias (Fig. 2b) and the anomalous nucleotide composition bias (Fig. 1). We found that type Ia *pmoCAB* coding sequences increase their level of codon bias at the expense of decreasing the GC_3 content (Fig. 2c), which contrasts with all other methanotroph types and coding sequences tested (Additional file 1: Figure S5). The *pmoCAB* coding sequences of type III show similar patterns to type Ia, but this is expected based on the whole-genome nucleotide bias (Additional file 1: Figure S1).

Codon biases in methanotrophy operons optimize translation efficiency. The anomalous nucleotide composition bias of type Ia *pmoCAB* coding sequences suggests that there is a fitness advantage to either the particular codons selected or to having a lower codon diversity. We hypothesized that this advantage may lie in increased translation efficiency. To evaluate this, we first analyzed the encoded tRNA pool of all isolate genomes. MAGs were not included in this analysis as their tRNA pool could be partial due to incomplete genome assembly. We found

Fig. 2 Analysis of the codon usage bias. **a** Analysis of the Codon Adaptation Index (CAI). CAI indexes were converted to percentile ranks based on the relative distribution of CAI in each methanotroph. Dashed lines with varying slopes delineate the variations between the percentile ranks of both indexes. **b** Analysis of the distribution of effective number of codons (ENC) values of coding sequences from type Ia methanotrophs. The two-tailed Wilcoxon signed rank-test was applied to test the difference of ENC values between the *pmoCAB* and other coding sequences. Non-significant result of the tests is denoted by *ns* ($p > 0.01$); $**p \leq 0.01$; $***p \leq 0.001$; $****p \leq 0.0001$. **c** Analysis of the ENC as a function of GC_3 with the line representing the fit of a linear model between ENC and GC_3 .

that the main difference among methanotrophs is that the type Ia genomes lack tRNAs with cytosine in the 3'-end of the anticodon (Additional file 1: Figure S6). Taking wobble-pairing into account, we queried whether the codon usage bias is an optimization strategy in which type Ia *pmoCAB* codon usage has been selected to better match available tRNA isoacceptors and thereby improve translation efficiency by recruiting the most abundant tRNAs. To address this question, we estimated the translation efficiency of coding sequences using the tRNA adaptation index (tAI). The tAI considers exact and wobble-pairing to estimate how efficiently a given coding sequence would be translated based on its codon usage and the genomic tRNA pool. We found that even though the preference for individual codons is different between type Ia *pmoCAB*, *mxal*, and *xoxF* coding sequences, their translation efficiency is similarly ranked around the top 10th percentile in each genome; the only coding sequence showing significant difference to *pmoC* is *mxal* (*t*-test, 99% confidence level) (Fig. 3a). We included a control analysis to test the difference between the *pmoC* coding sequence and controls (Fig. 3a) and found that in this case the percentile rank of tAI is significantly different (*t*-test, 99% confidence level). The tAI results of other coding sequences and methanotrophs can be found in Additional file 1: Figure S7.

Overall, these results suggest that for type Ia methanotrophs, the translation efficiency of *pmoCAB*, *mxal*, and *xoxF* coding sequences are similar and these coding sequences have been optimized in parallel through different codon usage biases. A possible evolutionary explanation for these observations is that the distinct codon usage bias of each coding sequence causes these genes to segregate between different tRNA isoacceptors, allowing simultaneous translation of the coding sequences required for methane oxidation while avoiding competition for the same tRNAs. To investigate this, we reconstructed the protein synthesis network accounting for the distribution of tRNAs among coding

sequences of the metabolic module (Fig. 3b). We also included in the network the median RSCU value of each codon per coding sequence of type Ia methanotrophs (Additional file 1: Figure S8-S9), so that we can account for the codon preference and its interaction with tRNAs through different pairing mechanisms. The median codon usage biases of the whole genome and coding sequences of ribosomal proteins ($n = 1,246$, median ribosomal proteins per genome = 55, standard deviation = 3.75) were integrated in the network to compare the links between codons and tRNAs outside the methane oxidation module (details on the tRNA pool of each genome are given in Additional file 1: Figure S10). The resulting networks (Fig. 3b; see Additional file 1: Figure S11 for networks for all amino acids) did not support the hypothesis that the codons favored by each operon provided differential access to tRNA isoacceptors. We also analyzed how access to different tRNA isoacceptors changes with variations in individual codon preference of coding sequences (Fig. 3c). We found that the different codon usage biases do not provide the operons with access to significantly different fractions of the tRNA pool (Chi-square test for equality of proportions, $p = 0.87$). Taken together, these results suggest that the codon usage bias of type Ia *pmoCAB* coding sequences does not provide a fitness advantage either in translation efficiency or in differential access to different tRNA isoacceptors when compared against the *mxal* and *xoxF* coding sequences. Despite the differences in codon usage between operons from the methane metabolic module, simultaneous translation of these operons would still require competition for the same tRNA isoacceptors.

Type Ia *pmoCAB* coding sequences optimize translation accuracy.

We hypothesized that the divergent codon bias of *pmoCAB* coding sequences could be a strategy to optimize translation accuracy. This strategy prevents the incorporation of incorrect amino acids in the nascent protein^{23,24} by reducing the binding probability of near-cognate tRNA isoacceptors (i.e. tRNAs that differ by one base from the correct anticodon) for the ribosomal A-site²⁵. Such an optimization for accuracy can be achieved at the coding sequence level through specialization of the codon-anticodon interaction in which an amino acid is encoded by a frequent and optimal codon that is recognized by an abundant cognate tRNA²⁶. We hypothesized that if translation accuracy is a selective pressure acting on a coding sequence, the positive selection for an optimal codon (or the purifying selection acting on non-optimal codons) should increase as that codon's amino acid is increasingly used in the encoded protein.

To test this, we first calculated the amino acid usage of each protein (Additional file 1: Figure S12) and mapped the RSCU with the highest median per amino acid (Additional file 1: Figure S8) to the median frequency of amino acid usage. Finally, we used a linear regression to analyze the relationship between the codon preference in coding sequences with the amino acid usage frequency in their proteins. Consistent with our translational accuracy hypothesis, we found that among all methanotrophs and all coding sequences of the methane oxidation metabolic module, only the type Ia *pmoCAB* coding sequences showed a significant linear relationship ($p < 0.01$, Fig. 3d, Additional file 1: Figure S13). This is especially notable for the two amino acids with the highest codon degeneracy (i.e. the 6-fold Leu and Arg).

Minimizing protein synthesis cost through favoring of prebiotic amino acids.

The use of CAI, ENC, RSCU and tAI enabled us to investigate the role of synonymous codons in optimization for translation efficiency and accuracy. However, as these measures account for amino acid usage biases, optimization strategies at the protein sequence level can be masked. We

Fig. 3 Strategies for optimal translation. **a** The tRNA adaptation index (tAI) was used to estimate the translation efficiency of coding sequences. tAI values were converted to percentile ranks based on the relative distribution of tAI in each methanotroph. Coding sequences were ordered by the mean of the tAI percentile rank and the two-tailed *t*-test was applied to test whether the percentile rank of tAI is significantly different between the *pmoC* coding sequence (highest mean) and other coding sequence. ns ($p > 0.01$); *** $p \leq 0.001$. **b** Network of the interaction between codons and tRNA in type Ia methanotrophs. The integrated network was composed by median RSCU values, median number of tRNA copies, and the codon-anticodon pairing rules. The two toy examples illustrate the expected networks when *pmoCAB* coding sequences have a codon usage bias that (I) grants an advantage (marked by an arrow) in the competition for the tRNA pool or (II) does not grant any advantage. The observed network for threonine is shown in (III) and all other amino acids are shown in Additional file 1: Figure S11. **c** Quantitative analysis of the codon-tRNA interaction network. The figure shows how many copies of tRNA a gene can access according to a RSCU cut-off varying from 0.0 to 2.0 with a step size of 0.01. The reference line at RSCU = 1.0 indicates the change from the region of non-preference codon usage (RSCU < 1.0) to the region of high-preference for individual codons (RSCU > 1.0). The colored lines show the fit of a polynomial surface using local fitting. The Chi-squared test of proportions was applied to test whether the proportion of tRNAs available to coding sequences in the methane oxidation metabolic module is different between RSCU = 0 and RSCU = 2. **d** The median usage of each amino acid in each translated coding sequence as a function of the median RSCU value of the codon exhibiting the highest preference for each amino acid for type Ia methanotrophs. The *p*-value corresponds to the significance of the linear regression. **e** Usage of prebiotic (cheap) and modern (expensive) amino acids in the proteins of the methane oxidation metabolic module of type Ia methanotrophs. The *t*-test was used to evaluate the difference between the two groups.

hypothesized that if translation optimization at the protein sequence level is achieved through reducing the synthesis cost of proteins^{27,28}, there should be an increased usage of the prebiotic (cheaper) amino acids²⁹ in the type Ia pMMO sequence. To test this, we first visually inspected the occurrence of prebiotic and modern amino acids in the linear regression between RSCU and amino acid usage (Fig. 3d), then quantitatively compared the median composition of cheaper and expensive amino acids in each protein (Fig. 3e). We observed that usage of prebiotic amino acids is significantly higher than modern (expensive) amino acids in pMMO, sMMO and Xox, while Mxa showed a non-significant difference (*t*-test, 99% confidence level; Fig. 3e).

The codon usage of *pmoCAB* coding sequences encodes an effective resource allocation strategy to synthesize highly demanded transcripts. While the codon usage and nucleotide composition biases of coding sequences in the type Ia methane oxidation module does not seem to provide an advantage in translation efficiency, it is possible that they provide an efficiency advantage at the level of transcription. Previous reports on methanotrophic species^{30–32} and communities^{33–35} have shown that *pmoCAB* coding sequences are consistently highly expressed regardless of the experimental conditions, and the demand on cellular resources to transcribe *pmoCAB* coding sequences could be significant. To test whether the biases confer increased transcriptional efficiency, the transcriptomic datasets of three different type Ia methanotrophs obtained under different experimental conditions were examined. For illustra-

tion purposes, the results of *Methylobacter tundripaludum* 31/32 grown with methane as sole carbon source and supplemented with lanthanum³² are described here. Detailed results of the other methanotrophs and experimental conditions can be found in Additional file 1: Figure S14-S16.

To test the transcription optimization hypothesis, we first analyzed the purine and pyrimidine content of coding sequences in each genome and found that purine-derived residues occurred more frequently (52%) than pyrimidine nucleotides (Additional file 1: Figure S14a). We postulated that, if *pmoCAB* coding sequences are optimized for efficient transcription, they should have a higher pyrimidine than purine content, which reduces the mRNA synthesis cost as pyrimidines have fewer carbon and nitrogen atoms per molecule (Additional file 1: Figure S14a). We found that the *pmoCAB* were the most expressed transcripts and they have a higher pyrimidine than purine content (Additional file 1: Figure S14b), contrasting the *xoxF* and *mxoFI* coding sequences (Additional file 1: Figure S14b) and the genome-scale frequency (Additional file 1: Figure S14a). We calculated the mean pyrimidine content of 1,000 random combinations of three coding sequences (Additional file 1: Figure S14c) and found that < 3.5% of the combinations had pyrimidine contents higher than the *pmoCAB* coding sequences.

To understand the effect of *pmoCAB* nucleotide frequency bias on cellular ribonucleotide composition, we analyzed the effect on the total ribonucleotide composition if we removed all the transcripts of an individual operon from the transcriptome. We calculated the total ribonucleotide composition by

Fig. 4 Strategies for optimal transcription in type Ia methanotrophs. a, b Elemental composition of transcribed coding sequences in *M. tundripaludum* 31/32. The main panel shows the number of element atoms per codon in each transcribed coding sequence with the dashed lines representing the mean of all transcribed coding sequences in the genome. The top and right panels show the mean number of element atoms per codon of the *pmoCAB* transcribed coding sequences and the mean number of element atoms per codon of 1,000 random combinations of three coding sequences. c A summary of the carbon and oxygen atom demands per codon of transcribed coding sequences. The difference (Δ) was calculated using the genome-scale mean demand as the reference (dashed line). *M. buryatense* 5G possesses genes encoding both the pMMO and sMMO. An extended analysis for each methanotroph is shown in Additional file 1: Figure S14, S15 and S16.

summing the ribonucleotide composition of all transcribed coding sequences in the transcriptome. We found that removing *mxoF* and *mxoF* transcripts had negligible effect on the total ribonucleotide composition (Additional file 1: Figure S14d-f). However, removing *pmoCAB* transcripts caused a shift of 1% in the purine/pyrimidine balance, a shift of > 2% in the GC content, and a shift of 2% in uracil content (Additional file 1: Figure S14d-f).

If the nucleotide composition bias of *pmoCAB* coding sequences is an adaptation to reduce transcript synthesis cost, it would be expected that the atomic element investment per codon is reduced in *pmoCAB* transcribed coding sequences compared to others in the methane oxidation metabolic module and to the whole-genome average. To evaluate this, we calculated the atomic element composition of each transcribed coding sequence in the metabolic module and of all transcribed coding sequences in the genome based on their adenine, guanine, cytosine and uracil composition. We found that *pmoCAB* transcripts have fewer carbon, hydrogen, and nitrogen atoms per codon compared to other transcripts from the methane metabolic module and the genome-scale mean (Fig. 4a and 4b). However, although the codon usage of *pmoCAB* transcripts requires reduced investment of carbon, nitrogen, and hydrogen, this comes at the expense of increased oxygen demand, with the oxygen demand exceeding the genome-scale mean (Fig. 4b). We selected 1,000 random combinations of three transcribed coding sequences from the whole genome, calculated the per-codon element compositions, and compared them to the *pmoCAB* transcribed coding sequences demand (Fig. 4a and 4b). We found that < 3% of the combinations of transcribed coding sequences yielded a mean element composition below that of the *pmoCAB* transcribed coding sequences for nitrogen, carbon and hydrogen or above for oxygen. This pattern was even more pronounced for *Methylomicrobium*

alcaliphilum 20Z (< 0.4%; Additional file 1: Figure S15g-h) and *Methylomicrobium buryatense* 5G (< 1.5%; Additional file 1: Figure S16g-h).

Based on these results, an intriguing new question emerged: is the element composition of type Ia *pmoCAB* transcripts a reflection of the low methane and high oxygen composition of habitats where these methanotrophs are commonly found¹⁰? To assess this question, we calculated the distance (Δ) between the mean carbon and oxygen atom counts per codon for methane oxidation metabolic module transcripts and all transcripts from each genome. We found that the *pmoCAB* transcripts are the only set of transcripts in the methane oxidation metabolic module that had low carbon and high oxygen composition per codon (Fig. 4c). Finally, we assessed whether the elemental composition of the *pmoCAB* operon is a result of a genome-wide trend for reducing the atoms per codon as a function of mRNA abundance. Rejecting this hypothesis, we found only weak correlations between elemental content and mRNA abundance (Additional file 1: Figure S14i, S15i and S16i). Together, these results suggest an optimization strategy unique to type Ia *pmoCAB* coding sequences whereby highly demanded transcripts use fewer carbon, nitrogen, and hydrogen atoms at the expense of more oxygen atoms per codon. Given the high oxygen requirement for synthesis of *pmoCAB* transcripts, we hypothesized that type Ia genomes may harbor a genetic regulatory mechanism adjusting *pmoCAB* transcription in response to oxygen availability. Supporting this, we identified proteins containing the PAS domain, which can potentially act as oxygen sensors³⁶, located near the *pmoCAB* coding sequences in the type Ia genomes of genera *Methylosarcina*, *Methylomonas*, *Methylomicrobium*, *Methylomarinum* (about ten coding sequences upstream), and *Methylobacter tundripaludum* 31/32 (five coding sequences downstream) (Additional file 1: Figure S17).

Fig. 5 Schematic representation of the molecular characteristics of type Ia methanotrophs. a A cartoon depicting the environmental counter-gradient of oxygen and methane in terrestrial and aquatic settings and the potential coupling with the molecular demands of oxygen and carbon atoms per codon of the type Ia *pmoCAB* coding sequences. The demands are relative to the genome-scale mean. b The three levels of molecular optimization encoded in the coding sequences of the methane oxidation metabolic module of type Ia methanotrophs.

Discussion

This study provides genomic evidence that the *pmoCAB* coding sequences of type Ia methanotrophs possess a unique adaptive trait manifesting as a strong nucleotide bias (Fig. 1) that fine-tune codon usage (Fig. 2) and can optimize methane oxidation through maximizing translation efficiency and accuracy (Fig. 3), while minimizing protein (Fig. 3) and transcript synthesis costs (Fig. 4). The discovery of this unique coding sequence was enabled by meta-analysis of a large number of isolate genomes and MAGs from diverse taxonomy and geography (Additional file 1: Figure S1). This finding illustrates a sophisticated adaptive linkage between molecular genotype and phenotype for methane oxidation.

From an environmental point of view, this adaptive trait would enable type Ia methanotrophs to be competitive and efficient in oxidizing methane, which is a gas with limited solubility and mass transfer in liquid³⁷ and often found at low concentrations in the aerobic zone of the oxygen/methane counter-gradient¹⁰. The high oxygen and low carbon demands of the type Ia *pmoCAB* transcribed coding sequences (Fig. 4), relative to other transcribed coding sequences in the methane oxidation metabolic module and the rest of the transcriptome, reflect the efficiency of these strictly aerobic³⁸ methanotrophs in colonizing and performing methane oxidation in environmental zones with high oxygen and low methane concentrations^{35,39,40} (Fig. 5a). Past reports on the molecular optimization of genetic sequences have shown that transcribed regions of genes can have fine-tuned nucleotide compositions that reduce their biosynthetic cost²⁰. For example, DNA coding sequences are found to reduce nitrogen atoms per codon when nitrogen is scarce in the environment⁴¹, and a large proportion of bacterial genomes present a dual optimization strategy to reduce per-codon nitrogen demand while increasing translation efficiency⁴². Extending these findings, we show that type Ia *pmoCAB* coding sequences possess an adaptation that minimizes not only nitrogen but also carbon and hydrogen per transcribed codon, while reducing the metabolic burden of erroneous biosynthesis²⁵ through optimal translation efficiency and accuracy. This optimization strategy likely increases the fitness of type Ia methanotrophs in the competition for methane and contributes to the prevalence of type Ia methanotrophs across diverse environments⁷. However, the transcript optimization strategy found in type Ia *pmoCAB* coding sequences results in an increased demand for oxygen atoms per codon. The high demand for oxygen to synthesize *pmoCAB* transcripts could be part of a mechanism to modulate cellular metabolism based on oxygen availability in the environment and to tightly regulate the first enzymatic step of the methane oxidation metabolic module at the transcription level. This is supported by the presence of genes encoding the PAS domain proteins, potentially acting as oxygen sensors³⁶, flanking the *pmoCAB* coding sequences in the genomes of several type Ia methanotrophs (Additional file 1: Figure S17).

In the second enzymatic step of the methane oxidation metabolic module, methanol is oxidized to formaldehyde by the methanol dehydrogenase encoded by either the *mxoFI* or *xoxF* genes. We show here that the type Ia *pmoCAB*, *mxoFI*, and *xoxF* coding sequences have similar translation efficiency (Fig. 3a), similar access to tRNA isoacceptors (Fig. 3b and 3c), and a similar level of codon bias (Fig. 2b); however, their preference for individual codons (Additional file 1: Figure S3a) and translation accuracy (Fig. 3d) is different. Considering only the genes for the second enzymatic step, the difference in codon usage between the *mxoFI* and *xoxF* coding sequences reduces the *xoxF* transcripts' demand for carbon, nitrogen, and hydrogen atoms per codon (Fig. 4). We also show that protein synthesis has been optimized through the frequent incorporation of prebiotic (cheap) amino acids in the pMMO and Xox protein sequences, though not in Mxa (Fig. 3e). Together, these results indicate although the *mxoFI* and *xoxF* coding sequences share similar translation efficiency, the *xoxF* coding sequence reduces transcript and protein synthesis costs (Fig. 5b), suggesting that Xox has evolved to be the predominantly expressed methanol dehydrogenase in environments where oxygen concentration is high and methane concentration is low. Supporting this, recent experimental reports have found that Xox is the preferentially expressed methanol dehydrogenase in isolates of type Ia methanotrophs⁴³, while in complex microbial communities, the *mxoFI* coding sequences of type Ia methanotrophs are preferentially expressed when methane and oxygen are in high concentrations,

but the *xoxF* coding sequence is specifically expressed when the methane concentration is low⁴⁴. Overall, the results in this study show that the pMMO-Xox configuration of the methane oxidation module should be the most efficient molecular strategy to catalyze the sequential oxidation of methane and methanol (Fig. 5b). This finding complements recent investigations in isolates⁴³ and communities⁴⁴ by providing a molecular basis to explain the prevalent expression of *pmoCAB* and *xoxF* coding sequences.

While the genes encoding pMMO have been proposed to have evolved independently in separate genomes following an ancient speciation event^{45–47}, it has also been recently suggested that methane oxidation genes may have been acquired by horizontal gene transfer^{48,49}. The evolutionary history of the optimized *pmoCAB* coding sequences in type Ia methanotrophs warrants further investigation. Furthermore, hypothesis-driven investigations focusing on the implications of the *pmoCAB* coding sequence in the cellular and environmental performance of methanotrophs will be required. Fundamentally, our results emphasize the type Ia *pmoCAB* coding sequences have been shaped by natural selection and are not the result of random genetic drift even at the level of synonymous codons. The ecological role of methanotrophs as the primary producers in complex communities^{5,6} can exert strong evolutionary pressures on pMMO. Efficient conversion of methane to methanol not only serves as the first enzymatic step for methanotrophs, but also provides methanol as the carbon source for methylotrophic bacteria³². The availability of methanol can also support methane oxidation in methanotrophs that have a low affinity for methane when methane concentration is at atmospheric levels⁵⁰. Consequently, the unique adaptations of the type Ia *pmoCAB* coding sequences are likely the result of molecular evolution at the organismal, community, and perhaps even planetary scale⁵¹.

Conclusions

By conducting an in-depth meta-analysis of isolate genomes and MAGs from diverse taxonomy and geography we show that, among the seven types of methanotrophs, only type Ia methanotrophs possess a unique coding sequence of the *pmoCAB* operon that is under positive selection for optimal resource allocation and efficient synthesis of transcripts and proteins in environmental counter gradients with high oxygen and low methane concentrations. The coding sequence is conserved in type Ia methanotrophs inhabiting diverse environments around the globe, implying that natural selection rather than neutral evolution governs the molecular evolution of methane oxidation in type Ia methanotrophs. Overall, this study provides genomic evidence of an intriguing adaptive trait in the *pmoCAB* coding sequences of type Ia methanotrophs and extends our knowledge of the evolutionary landscape explored by nature to sustain metabolism in highly diverse environments on Earth.

Materials and Methods

Detailed description of the methods used in this study is available in Additional file 1: Supplemental Methods.

Methanotroph genomes. Of the 67 methanotroph genomes analyzed in this study, 49 are isolates and 18 are metagenome-assembled genomes (MAGs) (see Additional file 2 for details). This collection included most of the isolate genomes and relevant methanotrophic metagenomic datasets available in the public domain as of October 2017. Genome annotations were compiled from the JGI/IMG⁵² and NCBI/RefSeq⁵³ databases. Functional annotations were absent or ambiguous for the genomes of the type Ia methanotrophs *Methylobacter tundripaludum* SV96, *Methylococcaceae bacterium* Sn10-6, and *Methylophilum alcaliphilum*

20Z; type Ib *Methylococcus capsulatus* Bath and *Methylococcus capsulatus* Texas; type IIa *Methylocystis parvus* OBBP and *Methylocystis* sp. SC2; and type III *Methylacidiphilum infernorum* V4. The amino acid sequences from their translated coding sequences were further annotated with KEGG/BlastKOALA⁵⁴. Phylogeny of the genomes was assigned by PhyloPhlAn⁵⁵ based on the whole genome.

Metagenomic analysis pipeline. Briefly, raw reads of publicly available metagenomic datasets were obtained from NCBI. Quality control of raw reads was performed with *illumina-utils*⁵⁶. Contigs were co-assembled with MEGAHIT⁵⁷ and binning was performed with MaxBin⁵⁸. The quality of MAGs was assessed with CheckM⁵⁹ and MAGs were categorized according to the quality standards defined in the Minimum Information about a Metagenome-Assembled Genome of bacteria⁶⁰. Annotation of MAGs was conducted with Prokka⁶¹ and eggNOG-mapper⁶².

Sequence analysis. Nucleotide composition bias (GC and GC₃ content) were analyzed with *seqinR*⁶³ in R. Codon usage metrics were formulated in R according to the equations described for RSCU⁶⁴ and CAI⁶⁴. tAI was calculated with *codonR*⁶⁵. The tRNA datasets required for reproduction of tAI calculations can be found in Additional file 1: Figure S10. ENC⁶⁶ was calculated with *chips* in EMBOSS⁶⁷.

Network analysis. The network dataset was generated using in-house scripts. The network matrix was imported and edited with Cytoscape⁶⁸.

Transcriptomic analysis. Normalized mRNA abundance data from RNA-Seq experiments were obtained from three publicly available independent studies conducted in the type Ia methanotrophs *Methylococcus buryatense* 5G³¹, *Methylococcus alcaliphilum* 20Z³⁰ and *Methylobacter tundripaludum* 31/32³².

Code and data availability. Scripts required to reproduce all the results and figures can be obtained from https://github.com/PLeeLab/methane_oxidation_genetic_trait/.

Declarations

Acknowledgements This research was supported by the Research Grants Council of Hong Kong through Project 11206514. JCV acknowledges support provided by the Hong Kong PhD Fellowship Scheme (HKPFS). We thank David Wilkins for critiquing the manuscript.

Competing interests The authors declare no competing interests.

References

- [1] Joshua F. Dean, Jack J. Middelburg, Thomas Röckmann, Rien Aerts, Luke G. Blauw, Matthias Egger, Mike S. M. Jetten, Annie E. E. Jong, Ove H. Meisel, Olivia Rasigraf, Caroline P. Slomp, Michiel H. Zandt, and A. J. Dolman. Methane feedbacks to the global climate system in a warmer world. *Reviews of Geophysics*, 56:207–250, 2018.
- [2] Kirsten Oswald, Jana Milucka, Andreas Brand, Philipp Hach, Sten Littmann, Bernhard Wehrli, Marcel M.M. Kuypers, and Carsten J. Schubert. Aerobic gammaproteobacterial methanotrophs mitigate methane emissions from oxic and anoxic lake waters. *Limnology and Oceanography*, 61:S101–S118, 2016.
- [3] Alexander B. Michaud, John E. Dore, Amanda M. Achberger, Brent C. Christner, Andrew C. Mitchell, Mark L. Skidmore, Trista J. Vick-Majors, and John C. Priscu. Microbial oxidation as a methane sink beneath the West Antarctic Ice Sheet. *Nature Geoscience*, 10:582–586, 2017.
- [4] Christian Knoblauch, Christian Beer, Susanne Liebner, Mikhail N. Grigoriev, and Eva-Maria Pfeiffer. Methane production as key to the greenhouse gas budget of thawing permafrost. *Nature Climate Change*, 8:309–312, 2018.
- [5] B Phillips, D Brankovits, F Alvarez, H Niemann, JW Pohlman, KW Becker, MB Leigh, MC Leewis, MF Lehmann, and TM Iliffe. Methane- and dissolved organic carbon-fueled microbial loop supports a tropical subterranean estuary ecosystem. *Nature Communications*, 8:1835, 2017.
- [6] Clemens Karwautz, Günter Kus, Michael Stöckl, Thomas R Neu, and Tillmann Lueders. Microbial megacities fueled by methane oxidation in a mineral spring cave. *The ISME Journal*, 2017.
- [7] Claudia Knief. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. *Frontiers in Microbiology*, 6:1346, 2015.
- [8] Alan A DiSpirito, Jeremy D Semrau, Colin J Murrell, Warren H Gallagher, Christopher Dennison, and Stéphane Vuilleumier. Methanobactin and the link between copper and bacterial methane oxidation. *Microbiology and Molecular Biology Reviews*, 80:387–409, 2016.
- [9] Eli K. Moore, Benjamin I. Jelen, Donato Giovannelli, Hagai Raanan, and Paul G. Falkowski. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geoscience*, 10:629–636, 2017.
- [10] Ludmila Chistoserdova and Marina G. Kalyuzhnaya. Current trends in methylophily. *Trends in Microbiology*, 26:703–714, 2018.
- [11] Luba Katz and Christopher B. Burge. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Research*, 13:2042–2051, 2003.
- [12] Kajetan Bentele, Paul Saffert, Robert Rauscher, Zoya Ignatova, and Nils Blüthgen. Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology*, 9:675–675, 2013.
- [13] Zachary R. Newman, Janet M. Young, Nicholas T. Ingolia, and Gregory M. Barton. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proceedings of the National Academy of Sciences*, 113:E1362–E1371, 2016.
- [14] Gavin Hanson and Jeff Collier. Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, 19:20, 2017.
- [15] Pragya Mittal, James Brindle, Julie Stephen, Joshua B. Plotkin, and Grzegorz Kudla. Codon usage influences fitness through RNA toxicity. *Proceedings of the National Academy of Sciences*, 115:201810022, 2018.
- [16] Grégory Boël, Reka Letso, Helen Neely, W. Nicholson Price, Kam-Ho Wong, Min Su, Jon D. Luff, Mayank Valecha, John K. Everett, Thomas B. Acton, Rong Xiao, Gaetano T. Montelione, Daniel P. Aalberts, and John F. Hunt. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature*, 529:358, 2016.
- [17] Chien-Hung Yu, Yunkun Dang, Zhipeng Zhou, Cheng Wu, Fangzhou Zhao, Matthew S. Sachs, and Yi Liu. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Molecular Cell*, 59:744–754, 2015.
- [18] Sebastian Pechmann, Justin W Chartron, and Judith Frydman. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. *Nature Structural and Molecular Biology*, 21:1100–1105, 2014.
- [19] Idan Frumkin, Dvir Schirman, Aviv Rotman, Fangfei Li, Liron Zahavi, Ernest Mordret, Omer Asraf, Song Wu, Sasha F. Levy, and Yitzhak Pilpel. Gene architectures that minimize cost of gene expression. *Molecular Cell*, 65:142–153, 2017.
- [20] Wei-Hua Chen, Guanting Lu, Peer Bork, Songnian Hu, and Martin J. Lercher. Energy efficiency trade-offs drive nucleotide usage in transcribed regions. *Nature Communications*, 7:11334, 2016.
- [21] Zhipeng Zhou, Yunkun Dang, Mian Zhou, Lin Li, Chien-hung Yu, Jingjing Fu, She Chen, and Yi Liu. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences*, 113:E6117–E6125, 2016.
- [22] Idan Frumkin, Marc J. Lajoie, Christopher J. Gregg, Gil Hornung, George M. Church, and Yitzhak Pilpel. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proceedings of the National Academy of Sciences*, 115:E4940–E4949, 2018.
- [23] Eduardo Rocha. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, 14:2279–2286, 2004.
- [24] Mor Meyerovich, Gideon Mamou, and Sigal Ben-Yehuda. Visualizing high error levels during gene expression in living bacterial cells. *Proceedings of the National Academy of Sciences*, 107:11543–11548, 2010.
- [25] Hila Gingold and Yitzhak Pilpel. Determinants of translation efficiency and accuracy. *Molecular Systems Biology*, 7:481–481, 2011.
- [26] Tessa E.F. Quax, Nico J. Claassens, Dieter Söll, and John van der Oost. Codon bias as a means to fine-tune gene expression. *Molecular Cell*, 59:149–161, 2015.
- [27] Ning Li, Jie Lv, and Deng-Ke Niu. Low contents of carbon and nitrogen in highly abundant proteins: evidence of selection for the economy of atomic composition. *Journal of Molecular Evolution*, 68:248–255, 2009.
- [28] Daniel R. Smith and Matthew R. Chapman. Economical evolution: microbes reduce the synthetic cost of extracellular proteins. *mBio*, 1:e00131–10, 2010.
- [29] Liam M. Longo, Jihun Lee, and Michael Blaber. Simplified protein design biased for prebiotic amino acids yields a foldable, halophilic protein. *Proceedings of the National Academy of Sciences*, 110:2135–2139, 2013.

- [30] MG Kalyuzhnaya, S Yang, ON Rozova, NE Smalley, J Clubb, A Lamb, Nagana GA Gowda, D Raftery, Y Fu, F Bringle, S Vuilleumier, DAC Beck, YA Trotsenko, VN Khmelenina, and ME Lidstrom. Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nature Communications*, 4:2785, 2013.
- [31] Andrea Torre, Aisha Metivier, Frances Chu, Lieve M. L. Laurens, David A. C. Beck, Philip T. Pienkos, Mary E. Lidstrom, and Marina G. Kalyuzhnaya. Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in *Methylobacterium buryatense* strain 5G(B1). *Microbial Cell Factories*, 14:1–15, 2015.
- [32] Sascha MB Krause, Timothy Johnson, Yasodara Karunaratne, Yanfen Fu, David AC Beck, Ludmila Chistoserdova, and Mary E Lidstrom. Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions. *Proceedings of the National Academy of Sciences*, 114:358–363, 2017.
- [33] Ryan A Lesniewski, Sunit Jain, Karthik Anantharaman, Patrick D Schloss, and Gregory J Dick. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *The ISME Journal*, 6:2257, 2012.
- [34] Achim Quaiser, Xavier Bodi, Alexis Dufresne, Delphine Naquin, André-Jean Francez, Alexandra Dheilly, Sophie Coudouel, Mathieu Pedrot, and Philippe Vandenkoornhuyse. Unraveling the stratification of an iron-oxidizing microbial mat by metatranscriptomics. *PLoS ONE*, 9:e102561, 2014.
- [35] Yuanfeng Cai, Yan Zheng, Paul L. E. Bodelier, Ralf Conrad, and Zhongjun Jia. Conventional methanotrophs are responsible for atmospheric methane oxidation in paddy soils. *Nature Communications*, 7:11728, 2016.
- [36] B L Taylor and I B Zhulin. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, 63:479–506, 1999.
- [37] Chad A Haynes and Ramon Gonzalez. Rethinking biological activation of methane and conversion to liquid fuels. *Nature Chemical Biology*, 10:331–339, 2014.
- [38] Ian R. McDonald, Levente Bodrossy, Yin Chen, and J. Colin Murrell. Molecular ecology techniques for the study of aerobic methanotrophs. *Applied and Environmental Microbiology*, 74:1305–1315, 2008.
- [39] S. Wajih A. Naqvi, Phyllis Lam, Gayatree Narvenkar, Amit Sarkar, Hema Naik, Anil Pratihary, Damodar M. Shenoy, Mangesh Gauns, Siby Kurian, Samir Damare, Manon Duret, Gaute Lavik, and Marcel M. K. Kuypers. Methane stimulates massive nitrogen loss from freshwater reservoirs in India. *Nature Communications*, 9:1265, 2018.
- [40] Garrett J Smith, Jordan C Angle, Lindsey M Solden, Mikayla A Borton, Timothy H Morin, Rebecca A Daly, Michael D Johnston, Kay C Stefanik, Richard Wolfe, Bohrer Gil, and Kelly C Wrighton. Members of the genus *Methylobacter* are inferred to account for the majority of aerobic methane oxidation in oxic soils from a freshwater wetland. *mBio*, 9, 2018.
- [41] Emily A. Seward and Steven Kelly. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biology*, 17:226, 2016.
- [42] Emily A. Seward and Steven Kelly. Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria. *Genome Biology*, 19:102, 2018.
- [43] Frances Chu and Mary E Lidstrom. XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph *Methylobacterium buryatense*. *Journal of Bacteriology*, 198:1317–1325, 2016.
- [44] Zheng Yu, David AC Beck, and Ludmila Chistoserdova. Natural selection in synthetic communities highlights the roles of *Methylococcaceae* and *Methylophilaceae* and suggests differential roles for alternative methanol dehydrogenases in methane consumption. *Frontiers in Microbiology*, 8:2392, 2017.
- [45] Andrew J. Holmes, Andria Costello, Mary E. Lidstrom, and J. Colin Murrell. Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiology Letters*, 132:203–208, 1995.
- [46] M G Klotz and J M Norton. Multiple copies of ammonia monooxygenase (*amo*) operons have evolved under biased AT/GC mutational pressure in ammonia-oxidizing autotrophic bacteria. *FEMS Microbiology Letters*, 168:303–311, 1998.
- [47] Patricia L Tavormina, Victoria J Orphan, Marina G Kalyuzhnaya, Mike Jetten, and Martin G Klotz. A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environmental Microbiology Reports*, 3:91–100, 2011.
- [48] Craig D. Osborne and Victoria S. Haritos. Horizontal gene transfer of three co-inherited methane monooxygenase systems gave rise to methanotrophy in the Proteobacteria. *Molecular Phylogenetics and Evolution*, 2018.
- [49] Roshan Khadka, Lindsay Clothier, Lin Wang, Chee Kent Lim, Martin G. Klotz, and Peter F. Dunfield. Evolutionary history of copper membrane monooxygenases. *Frontiers in Microbiology*, 9:2493, 2018.
- [50] J Benstead, GM King, and HG Williams. Methanol promotes atmospheric methane oxidation by methanotrophic cultures and soils. *Applied and Environmental Microbiology*, 64:1091–8, 1998.
- [51] Joshua E. Goldford and Daniel Segrè. Modern views of ancient metabolic networks. *Current Opinion in Systems Biology*, 2018.
- [52] Victor M Markowitz, I.-Min A Chen, Ken Chu, Ernest Szeto, Krishna Palaniappan, Manoj Pillay, Anna Ratner, Jinghua Huang, Ioanna Pagani, Susannah Tringe, Marcel Huntemann, Konstantinos Billis, Neha Varghese, Kristin Tennesen, Konstantinos Mavromatis, Amrita Pati, Natalia N Ivanova, and Nikos C Kyrpides. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Research*, 42, 2014.
- [53] Nuala A. O’Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciuffo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wracko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O’Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, and Kim D. Pruitt. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44:D733–D745, 2016.
- [54] Minoru Kanehisa, Yoko Sato, and Kanae Morishima. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, 428:726–731, 2016.
- [55] Nicola Segata, Daniela Börnigen, Xochitl C. Morgan, and Curtis Huttenhower. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*, 4, 2013.
- [56] A. Murat Eren, Joseph H. Vineis, Hilary G. Morrison, and Mitchell L. Sogin. A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS ONE*, 8:e66643, 2013.
- [57] Dinghua Li, Chi-Man Liu, Ruihang Luo, Kunihiko Sadakane, and Tak-Wah Lam. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31:1674–1676, 2015.
- [58] Yu-Wei Wu, Blake A Simmons, and Steven W Singer. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32, 2016.
- [59] Donovan H Parks, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and Gene W Tyson. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25, 2015.
- [60] Robert M Bowers, Nikos C Kyrpides, Ramunas Stepanauskas, Miranda Harmon-Smith, Devin Doud, T B K Reddy, Frederik Schulz, Jessica Jarett, Adam R Rivers, Emiley A Elloe-Fadrosh, Susannah G Tringe, Natalia N Ivanova, Alex Copeland, Alicia Clum, Eric D Becraft, Rex R Malmstrom, Bruce Birren, Mircea Podar, Peer Bork, George M Weinstock, George M Garrity, Jeremy A Dodsworth, Shibu Yooseph, Granger Sutton, Frank O Glöckner, Jack A Gilbert, William C Nelson, Steven J Hallam, Sean P Jungbluth, Thijs J G Ettema, Scott Tighe, Konstantinos T Konstantinidis, Wen-Tso Liu, Brett J Baker, Thomas Rattei, Jonathan A Eisen, Brian Hedlund, Katherine D McMahon, Noah Fierer, Rob Knight, Rob Finn, Guy Cochrane, Ilene Karsch-Mizrachi, Gene W Tyson, Christian Rinke, Nikos C Kyrpides, Lynn Schriml, George M Garrity, Philip Hugenholtz, Granger Sutton, Pelin Yilmaz, Folker Meyer, Frank O Glöckner, Jack A Gilbert, Rob Knight, Rob Finn, Guy Cochrane, Ilene Karsch-Mizrachi, Alla Lapidus, Folker Meyer, Pelin Yilmaz, Donovan H Parks, A M Eren, Lynn Schriml, Jillian F Banfield, Philip Hugenholtz, and Tanja Woyke. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35:725, 2017.
- [61] Torsten Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30, 2014.
- [62] Jaime Huerta-Cepas, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, 2017.
- [63] Delphine Charif and Jean R. Lobry. Biological and medical physics. *Biomedical Engineering*, pages 207–232, 2007.
- [64] Paul M Sharp and Wen-Hsiung Li. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15, 1987.
- [65] Mario dos Reis, Renos Savva, and Lorenz Wernisch. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research*, 32, 2004.
- [66] Frank Wright. The ‘effective number of codons’ used in a gene. *Gene*, 87:23–29, 1990.
- [67] Peter Rice, Ian Longden, and Alan Bleasby. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, 16:276–277, 2000.
- [68] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2003.