Genomic insights into the ecological role and evolution of a novel Thermoplasmata order, “Candidatus Sysuiplasmatales”

Yang Yuan\textsuperscript{a}, Jun Liu\textsuperscript{b}, Tao-Tao Yang\textsuperscript{a}, Shao-Ming Gao\textsuperscript{a}, Bin Liao\textsuperscript{a}, Li-Nan Huang\textsuperscript{a}\#

\textsuperscript{a}School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, PR China
\textsuperscript{b}State Key Laboratory of Agricultural Microbiology, College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, PR China

Running title: Metabolism and evolution of \textit{Ca.} Sysuiplasmatales

#Address correspondence to Li-Nan Huang, eseshln@mail.sysu.edu.cn

\textbf{Keywords}: \textit{Ca.} Sysuiplasmatales; Acid mine drainage; Facultatively anaerobic heterotrophic lifestyle; Genome-resolved metagenomics; Ancestral state reconstruction
Abstract

Recent omics studies have provided invaluable insights into the metabolic potential, adaptation and evolution of novel archaeal lineages from a variety of extreme environments. We have utilized a genome-resolved metagenomic approach to recover eight medium- to high-quality metagenome-assembled genomes (MAGs) that likely represent a new order (“Candidatus Sysuiplasmatales”) within Thermoplasmata from mine tailings and acid mine drainage (AMD) sediments sampled from two copper mines in South China. 16S rRNA gene based analyses revealed a narrow habitat range for these uncultured archaea limiting to AMD and hot spring-related environments. Metabolic reconstruction indicated a facultatively anaerobic heterotrophic lifestyle. This may allow the archaea to adapt to oxygen fluctuations and is thus in marked contrast to the majority of lineages in the domain Archaea which typically show obligately anaerobic metabolisms. Notably, “Ca. Sysuiplasmatales” could conserve energy through degradation of fatty acids, amino acid metabolism and oxidation of reduced inorganic sulfur compounds (RISCs), suggesting that they may contribute to acid generation in the extreme mine environments. Unlike its closely related Methanomassiliicoccales and “Ca. Gimiplasmatales”, “Ca. Sysuiplasmatales” lack the capacity to perform methanogenesis and carbon fixation. Ancestral state reconstruction indicated that “Ca. Sysuiplasmatales” and its closely related Methanomassiliicoccales, “Ca. Gimiplasmatales”, and the SG8-5 and the RBG-16-68-12 orders originated from a facultatively anaerobic ancestor capable of carbon fixation via the bacterial-type H_4F Wood–Ljungdahl pathway (WLP). Their metabolic divergence might be attributed to different evolutionary paths.

Importance
A wide array of archaea populate Earth’s extreme environments thereby they may play important roles in mediating biogeochemical processes such as iron and sulfur cycling. However, our knowledge of archaeal biology and evolution is still limited considering the uncultured majority of archaeal diversity. For instance, most order-level lineages except Thermoplasmatales, Aciduliprofundales and Methanomassiliicoccales within Thermoplasmata do not have cultured representatives. Here, we report the discovery and genomic characterization of a novel order, namely “Ca. Sysuiplasmatales”, within Thermoplasmata in the extremely acidic mine environments. “Ca. Sysuiplasmatales” are inferred to be facultatively anaerobic heterotrophs and likely contribute to acid generation through the oxidation of RISCs. The physiological divergence between “Ca. Sysuiplasmatales” and its closely related Thermoplasmata lineages may be attributed to different evolutionary paths. These results expand our knowledge of archaea in the extreme mine ecosystem.

Introduction

Archaea as a distinct domain of life have been shown to inhabit diverse habitats including extreme (such as hot springs, hypersaline environments, acid mine drainage (AMD), etc.) and non-extreme environments (such as sea water, soils, and sediments) (1-6), where they may play crucial roles in biogeochemical cycles (7, 8). However, the majority of archaeal diversity remains uncultured, limiting our understanding of their metabolic capacities and evolutionary history (7-9). New cultivation-independent genomic approaches especially metagenomics and single-cell genomics have enabled reconstruction of near-complete or even closed genomes for uncultured archaea directly from the environment (10). This has led to the discovery of many candidate archaeal phyla and dramatically expanded the archaeal tree of life (7, 8). Subsequent
analyses of the metabolic capabilities, origin, and evolutionary history of these uncultured microorganisms on the basis of the retrieved genomes have significantly advanced our understanding of archaeal biology (8).

Thermoplasmata is a globally distributed and ecologically significant class-level archaeal lineage (11). Due to their wide range of lifestyles and specific genomic and cellular features, this unique archaea group could serve as an excellent model for the study of environmental adaptation and evolutionary processes within the Archaea (12). For instance, all members of the Thermoplasmatales order have extremely acidic pH optima for growth (among the lowest known, with Picrophilus spp. showing the lowest pH value of 0.7), and their cells (except for Picrophilus spp.) are typically pleomorphic as a consequence of the lack of an intact cell wall (13). Thus, characterization of these cell-wall deficient archaea may provide important insights into the acid tolerance mechanisms of microbial life (14). Meanwhile, while Archaea as a domain of life has been implicated to have evolved and diversified primarily in anoxic habitats, aerobic metabolism has been found in a handful of archaeal lineages including some members of Sulfolobales and Thermoplasmatales (15), which may have coevolved with the acidity of their habitats after the appearance of oxygenic photosynthesis (16). The details of the evolutionary transition from an anaerobic metabolism to the aerobic lifestyle and the process of geological–biological feedbacks associated with aerobic archaeal lineages remain largely unresolved, however. The co-existence of both anaerobic and aerobic members within the Thermoplasmata renders it as an interesting target for investigating this issue. However, the number of Thermoplasmata orders characterized thus far is limited, hampering our understanding of the metabolic potential, ecology and evolutionary history of this important archaeal class.

The extreme AMD environment represents an important model system for the study of
microbial community structure, function and evolution (17, 18). Early environmental surveys based on 16S rRNA sequencing and more recent metagenomics explorations have revealed the existence of novel archaea lineages (2, 19), including the ‘alphabet plasmas’ within the Thermoplasmatales (20), in AMD ecosystems. Here, we report discovery and genomic characterization of a novel order (“Ca. Sysuiplasmatales”) within Thermoplasmata based on eight metagenome-assembled genomes (MAGs) retrieved from acidic mine tailings and AMD sediments collected from two mine sites in South China. Reconstruction of metabolic pathways has resolved physiology and potential ecological roles of these uncultured archaea and ancestral state reconstruction revealed evolutionary paths leading to extant divergent Thermoplasmata lineages including “Ca. Sysuiplasmatales”.

Results

Metagenomic discovery, phylogeny and environmental distribution of the novel Thermoplasmata order, “Ca. Sysuiplasmatales”. A total of 303 Gbp of quality-controlled metagenomic data were generated from four extremely acidic mine tailings samples and two AMD sediments collected from two copper mines in South China (Dataset S2). Metagenomic assembly and binning recovered eight unique Thermoplasmata genomes (metagenome assembled genomes, MAGs). Genome sizes ranged from 1.96 to 2.29 Mbp, with estimated completeness ranging between 86% and 99% and an estimated contamination of 0.8%, suggesting that these MAGs are well curated and high-quality (Table 1). Phylogenetic analysis based on 122 concatenated archaeal marker genes and 16 concatenated ribosomal proteins revealed that these archaeal MAGs likely represent one distinct lineage within class Thermoplasmata and grouped phylogenetically closely related to Methanomassiliicoccales,
“Candidatus Gimiplasmatales”, the SG8-5 and the RBG-16-68-12 orders (Figs. 1, S2). Pairwise 16S rRNA gene identity values between the members in the order “Ca. Sysuiplasmatales” and other members that affiliated with its closely related Methanomassiliicoccales, “Candidatus Gimiplasmatales”, the SG8-5 and the RBG-16-68-12 orders were lower than 89% (Dataset S3), which indicated that they are belong to one novel Thermoplasmata order (21). Furthermore, genome-identity values (average amino acid identity and OrthoANI) of “Ca. Sysuiplasmatales” compared with its four closely related orders were as low as values observed among its four closely related orders (Fig. S3). Therefore, we propose one novel order within class Thermoplasmata. Comparison of 16S rRNA gene sequences of the novel order with the NCBI GenBank database only retrieved four 16S rRNA gene sequences from AMD and hot spring-related environments (Fig. S1; Dataset S4) (22-24).

Description of the taxa. The eight MAGs recovered represent three species within a single genus from a novel order affiliated with Thermoplasmata. Taking into account the consensus statement regarding nomenclature of uncultivated prokaryotes (25), we propose the following taxonomic names for these eight MAGs.

Description of “Ca. Sysuiplasmatales” ord. nov. (Sy.su.i.plas.ma.ta'les. N.L. neut. n. Sysiuplasma a Candidatus prokaryote name; -ales ending to denote a order; N.L. fem. pl. n. Sysiuplasmatales the Sysiuplasma order).

Description of “Ca. Sysuiplasmataceae” fam. nov. (Sy.su.i.plas.ma.ta.ce'ae. N.L. neut. n. Sysiuplasma a Candidatus prokaryote name; -aceae ending to denote a family; N.L. fem. pl. n. Sysuiplasmataceae the Candidatus Sysuiplasma family).

Description of “Ca. Sysuiplasma” gen. nov. (Sy.su.i.plas'ma. Gr. neut. n. plasma anything
formed or molded; N.L. neut. n. Sysuiplasma a form named after SYSU, acronym of Sun Yat Sen University).

Description of “Ca. Sysuiplasma acidicola” sp. nov. (a.ci.di’co.la. L. masc. n. acidus, sour; L. masc./fem. suff. -cola (from L. masc./fem. n. incola), dweller, inhabitant; N.L. n. acidicola, an inhabitant of an acidic environment). “Ca. Sysuiplasma acidicola” includes YP4-bin.258, TUT18-bin.7 and TUT19-bin.121 and the type material is TUT18-bin.7.

Description of “Ca. Sysuiplasma superficiale” sp. nov. (su.per.fi.ci.a’le. L. neut. adj. superficiale belonging to the surface). “Ca. Sysuiplasma superficiale” includes TUT19-bin.139 and YP2-bin.285 and the type material is YP2-bin.285.

Description of “Ca. Sysuiplasma jiujiangense” sp. nov. (jiu.jiang.en’se. N.L. neut. adj. jiujiangense, pertaining to Jiujiang, Jiangxi Province, China). “Ca. Sysuiplasma jiujiangense” includes TUT18-bin.177, UT3-bin.92 and libTUT1-bin.22 and the type material is UT3-bin.92.

Metabolic potential. In order to reveal the metabolic potentials of “Ca. Sysuiplasmatales”, we used multiple functional database comparisons for annotation (Dataset S5). Comparative analysis was then performed to determine metabolic differences between “Ca. Sysuiplasmatales” and its closely related Methanomassiliicoccales, “Ca. Gimiplasmatales”, and the SG8-5 and the RBG-16-68-12 orders (Fig. 3; Datasets S6, S8).

Carbon metabolism. For all MAGs of “Ca. Sysuiplasmatales”, genes encoding carbohydrate-active enzymes (CAZymes) were found (Fig. 2; Dataset S5). Among these CAZymes, genes encoding starch-degradation and glycogen debranching enzymes were
prevalent (Dataset S5). The generated glucose could be further oxidized via Embden-Meyerhof-Parnas glycolysis and TCA cycle, suggesting that “Ca. Sysiulasmatales” could conserve energy by glycolysis and TCA cycle. Moreover, “Ca. Sysiulasmatales” coded for phosphoenolpyruvate carboxylase (ppc), suggesting phosphoenolpyruvate might be a central intermediate connecting glycolysis with TCA cycle. Furthermore, “Ca. Sysiulasma acidicola” had the genetic potential for conversion of pyruvate into acetyl-CoA through pyruvate dehydrogenase (pdh) and pyruvate ferredoxin/flavodoxin oxidoreductase (por) under oxic and anoxic conditions, respectively. Notably, all “Ca. Sysiulasmatales” MAGs could perform acetate assimilation through AMP-forming acetyl-CoA synthetase (acs), and the generated acetyl-CoA could be used for the biosynthesis of carbohydrates, proteins, lipids and energy metabolism (11, 26). This further suggests a heterotrophic lifestyle for “Ca. Sysiulasmatales” through acetate assimilation, which is similar to its closely related members from the Methanomassiliicoccales, “Ca. Gimiulasmatales”, and the SG8-5 and the RBG-16-68-12 orders (11) (Fig. 3; Dataset S6). Surprisingly, unlike its closely related members in the Methanomassiliicoccales, “Ca. Gimiulasmatales” and the RBG-16-68-12 order, “Ca. Sysiulasmatales” do not harbor the bacterial-type carbon monoxide dehydrogenase/acetyl-CoA synthase complex (CODH/ACS), which are key enzymes for anaerobic carbon fixation via the WLP (27), suggesting that these new archaea may not have the capability of carbon fixation via WLP. Furthermore, the eight MAGs of “Ca. Sysiulasmatales” coded for L-lactate dehydrogenase complex protein LldG but lacked LldE and LldF, suggesting a deficiency in lactate fermentation (28). In addition, fatty acid β-oxidization potential was present in all MAGs of “Ca. Sysiulasmatales”, thus “Ca. Sysiulasmatales” were inferred to conserve energy via degradation of fatty acids.
**Sulfur metabolism.** Thiosulfate ($S_2O_3^{2-}$) and hydrogen sulfide ($H_2S$), which are generated during leaching of metal sulfides (29), represent two kinds of RISCs in acidic mining environments. Metabolic reconstruction suggested that “Ca. Sysuiplasmatales” could conserve energy through oxidation of RISCs (Fig. 2). For example, “Ca. Sysuiplasmatales” could utilize adenyllylsulfate reductase (*arpAB*) and sulfate adenyllytransferase (*sat*) to produce sulfate through oxidation of sulfite, which might be generated through transformation of thiosulfate via thiosulfate/3-mercaptopyruvate sulfurtransferase (*sseA*) (30). Furthermore, “Ca. Sysuiplasma jiuijiangense” coded for sulfite reductase (ferredoxin), which could reduce sulfite to hydrogen sulfide. The generated hydrogen sulfide might be oxidized to polysulfide via sulfide-quinone oxidoreductase (*sqr*), which functions under anaerobic conditions in most cases (31).

**Oxidative phosphorylation.** Complete pathways for oxidative phosphorylation (including NADH-quinone oxidoreductase, succinate dehydrogenase, the cytochrome bc1 complex, *aa3*-type/*bd*-type cytochrome c oxidase and V/A-type ATPase) were identified in “Ca. Sysuiplasmatales” (Fig. 2; Dataset S5). All MAGs harbored the *aa3*-type cytochrome c oxidase, which is a low-affinity $O_2$ terminal oxidase working under oxic conditions, whereas only “Ca. Sysuiplasma superficiale” harbored *bd*-type cytochrome c oxidase, which is a high-affinity $O_2$ terminal oxidase that could work under oxygen-limiting conditions (32). This suggested that “Ca. Sysuiplasma superficiale” could be more versatile and thrive in microaerobic and aerobic environments, and “Ca. Sysuiplasma acidicola” and “Ca. Sysuiplasma jiuijiangense” could thrive in aerobic environments. Additionally, all MAGs of this novel order harbored genes encoding the aerobic molybdenum-dependent CO dehydrogenase (*coxLMS*) for CO oxidization (33), further suggesting the proposed aerobic or microaerobic lifestyle for “Ca. Sysuiplasmatales”. In
combination with the above-mentioned \( \text{sqr, por} \) and \( \text{pdh} \) genes and the below-mentioned \( \text{trx, sod} \) and \( \text{prxQ} \) genes, these results collectively suggested a facultatively anaerobic lifestyle for “\( \text{Ca. Sysuiplasmatales} \)”. Furthermore, among Methanomassiliicoccales, “\( \text{Ca. Gimiplasmatales} \)”, the SG8-5 and the RBG-16-68-12 orders, some members affiliated with the RBG-16-68-12 order harbored the \( \text{aa}_3 \)-type cytochrome c oxidase (Fig. 3), suggesting the capacity for aerobic respiration, which was consistent with previous findings (34).

**Cell membrane biosynthesis.** Isoprenoids are necessary for all living organisms and involved in several vital functions such as compartmentalization and protection (35). As found in other archaea, all MAGs of “\( \text{Ca. Sysuiplasmatales} \)” possessed complete mevalonate pathway (MVA) for biosynthesis of IPP and DMAPP (isopentenyl diphosphate and dimethylallyl diphosphate), which are precursors for isoprenoid biosynthesis (35, 36), thus indicating that “\( \text{Ca. Sysuiplasmatales} \)” could synthesize isoprenoids de novo. Additionally, “\( \text{Ca. Sysuiplasmatales} \)” harbored mevalonate 5-kinase and mevalonate 5-phosphate decarboxylase of the classical archaeal MVA pathway, and did not harbor mevalonate 3-phosphate 5-kinase and mevalonate 3,5-bisphosphate decarboxylase of the novel MVA pathway (37), which has been found to be employed by some extreme acidophiles (such as members of Thermoplasmatales) to efficiently produce isoprenoids (37). Sequence alignment of mevalonate 5-phosphate decarboxylase of “\( \text{Ca. Sysuiplasmatales} \)” against classical MVA pathway decarboxylases showed retention of the invariant Asp/Lys/Arg catalytic triad required for decarboxylation (Fig. S4), and retention of one of the two ATP binding residues in the mevalonate 5-phosphate decarboxylases of “\( \text{Ca. Sysuiplasmatales} \)”, suggesting dual function (decarboxylase and kinase) of mevalonate 5-phosphate decarboxylases of “\( \text{Ca. Sysuiplasmatales} \)”. As mevalonate 5-phosphate decarboxylase...
loses the kinase function at low pH, we propose that mevalonate 5-phosphate decarboxylases of “Ca. Sysuiplasmatales” may produce isoprenoids in the near-neutral cytoplasmic environments which could be maintained by the acid tolerance mechanisms including the potassium-transporting ATPase system (kdpABC) and proton buffer molecules (Dataset S5).

**Amino acid biosynthesis and metabolism.** “Ca. Sysuiplasmatales” harbored biosynthetic pathways for some amino acids such as alanine and aspartate (Fig. 2; Dataset S5). These organisms could also obtain amino acids via oligopeptide and amino acid transporters and diverse peptidases. Then, amino acids might be oxidized to acetyl-CoA via aminotransferases and 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase (kor) (34) (Fig. 2), and the reduced ferredoxin could be regenerated concomitantly. It was reported that the MBG-D Single Amplified Genome (SAG) could utilize amino acid metabolism to conserve energy (38). Thus, we proposed that “Ca. Sysuiplasmatales” could also conduct energy conservation through amino acid metabolism, similar to the predicted metabolism for “Ca. Lunaplasma lacustris” belonging to its closely related RBG-16-68-12 order (34).

**Nucleotide biosynthesis.** Although “Ca. Sysuiplasmatales” didn’t harbored complete pentose phosphate pathways, all MAGs possessed the complete metabolisms for the nonoxidative pentose phosphate pathway (Fig. 2), which could be used to generate phosphoribosyl pyrophosphate (PRPP) (2, 11). PRPP is a common precursor for de novo nucleotide biosynthesis (11). In combination with the presence of complete de novo nucleotide biosynthesis in “Ca. Sysuiplasmatales” (Fig. 2; Dataset S5), these results showed that “Ca. Sysuiplasmatales” had the genetic potential for nucleotide de novo biosynthesis (11, 39), which
might help these organisms adapt to the acidic habitat given that free nucleotides might be easily degradable in acidic environments (2).

Electron transfer. An electron-transferring flavoprotein complex (fixABCX) and butyryl-CoA dehydrogenase (bcd) were present in all MAGs of “Ca. Sysuiplasmatales”, thus “Ca. Sysuiplasmatales” might utilize FixABCX to transfer electrons released from acyl-CoA to ferredoxin given that this cytoplasmic electron-bifurcating complex was reported to couple the regeneration of reduced ferredoxin and oxidization of butyryl-CoA to crotonyl-CoA in Clostridium kluyveri (40). Moreover, all MAGs of “Ca. Sysuiplasmatales” harbored the membrane-bound respiratory complex NADH-ubiquinone oxidoreductase (Nuo) and hydrogenase (mbh), which could re-oxidize reduced ferredoxin to create a transmembrane proton motive force. Then ATP could be generated via the V/A-type ATPase. The absence of the NuoEFG subunits encoding the NADH-binding module further confirmed the potential use of reduced ferredoxin as electron donor (41, 42). Furthermore, all MAGs of “Ca. Sysuiplasmatales” coded for heterodisulfide reductases (hdrABC2) (Fig. 3), which were ancient enzymes belonging to the prokaryotic common ancestor (43). This ancient enzyme complex might be involved in the oxidation of inorganic sulfur compounds, the reduction of ferric iron and sulfate, and electron transfer (44, 45), not just methanogenic and methanotrophic pathways (46, 47).

Environmental adaptation. As “Ca. Sysuiplasmatales” inhabit the acidic mine environments, they must deal with the extreme environmental stresses including acid stress, heavy metals toxicity, and oxidative stress (Fig. 2; Dataset S5). For acid stress, all MAGs of “Ca. Sysuiplasmatales” encoded the complete potassium-transporting ATPase system (kdpABC),
which could help these organisms prevent inward deflection of protons partially by generating an inside-positive membrane potential (48). Meanwhile, “Ca. Sysuiplasmatales” could produce proton buffer molecules such as arginine and use Na⁺/H⁺ antiporters to excrete excess H⁺ to maintain the cytoplasmic pH near neutral (48-50). For heavy metals toxicity, all MAGs of “Ca. Sysuiplasmatales” harbored P-type Cu⁺ transporter (CopA) and YP2-bin.285 encoded the cobalt-zinc-cadmium efflux system protein, suggesting that efflux of metal ions may be an important strategy to deal with heavy metal stresses. Another strategy employed may be to reduce metal or metalloids ions to reduce toxicity and then export them. Taking arsenic resistance as an example, “Ca. Sysuiplasmatales” could reduce As⁵⁺ to As³⁺ and export them through the arsABC operons (11, 48). Additionally, “Ca. Sysuiplasmatales” could reduce Hg²⁺ to less toxic Hg⁰ via mercuric reductase (merA) and then export them (48). For oxidative stress, all MAGs of “Ca. Sysuiplasmatales” encoded thiorexin reductase (trxR), peroxiredoxins (prxQ) and superoxide dismutase (sod), which were employed widely by AMD microorganisms to deal with oxidative stress (51). Notably, three MAGs of “Ca. Sysuiplasmatales” harbored rubreythrin (rbr) and rubredoxin (rbo) comprising the oxidative stress protection system, which was used to deal with oxidative stress in anaerobes (52, 53), suggesting that these organisms were facultative anaerobes and exposed to severe oxidative stress. These strategies have been used by other acidophiles to prevent these environmental stresses (48).

The evolutionary history of Thermoplasmata. To infer evolutionary histories of Thermoplasmata, the Dollo parsimony method in COUNT was implemented to identify the gene gain and loss events by mapping the gene orthologous groups to the phylogenomic tree of 122 concatenated archaeal marker genes. As the focus of this study was to resolve the ecological
roles of “Ca. Sysuiplasmatales” and its closely related Methanomassiliicoccales, “Ca. Gimiplasmatales”, and the SG8-5 and the RBG-16-68-12 orders and reveal the evolutionary paths leading to their different ecological strategies, we focused on tracking the evolutionary history of metabolic pathways among these five orders. The common ancestor of these five orders was inferred to encode 3,052 orthologous genes (Fig. 4; Dataset S7), including several terminal oxidases (the aa3-type cytochrome c oxidase and the bd-type cytochrome c oxidase subunit I), pyruvate dehydrogenase (pdh) and pyruvate ferredoxin/flavodoxin oxidoreductase (por), sulfide-quinone oxidoreductase, ruberythrin (rbr) and rubredoxin (rbo), suggesting a facultatively anaerobic lifestyle for this ancestral organism (Dataset S7). Moreover, the key bacterial-type H4F WLP genes encoding the CODH/ACS complex were gained at the branch leading to this common ancestor, indicating that this ancestral organism may be capable of carbon fixation via WLP (11, 54). From this common ancestor, a large number of loss events and a flux of gene family content of considerably smaller magnitude occurred at the two branches leading to node 28 and node 37, respectively (Fig. 4). For node 28, some important genes were lost, including those encoding the aa3-type cytochrome c oxidase subunit II and the bd-type cytochrome c oxidase subunit I, potentially shaping the distribution of aerobic metabolism in the SG8-5 and the RBG-16-68-12 orders. Moreover, the genes encoding the key CODH/ACS complexes for the bacterial-type H4F WLP were lost at the branch leading to node 30 which is the last common ancestors (LACAs) of the RBG-16-68-12 organisms except for GCA_004377185.1 (Fig. 4), and the bacterial-type H4F WLP has been discovered in GCA_004377185.1 and some members of Methanomassiliicoccales and “Ca. Gimiplasmatales” (11). The rooted phylogenetic tree of CODH/ACS complexes showed that the sequences of Methanomassiliicoccales, “Ca. Gimiplasmatales” and the RBG-16-68-12 order formed a robust
monophyletic clade within the bacterial lineages, suggesting that these genes may be horizontally transferred from bacteria (Fig. S6). Additionally, the aa3-type cytochrome c oxidase subunit II was gained along the branch leading to node 33 (Fig. S5). From node 37, likewise, a large number of loss events and rare gene gain and loss events occurred at the two branches leading to node 38 and node 44, respectively. For node 38, some important genes such as the aa3-type cytochrome c oxidase and the bd-type cytochrome c oxidase subunit I were lost, suggesting an anaerobic lifestyle for “Ca. Gimiplasmatales”. In combination with retention of the CODH/ACS complexes in node 38 (Dataset S7), these results suggested the capacity of carbon fixation via WLP for “Ca. Gimiplasmatales”, which was reported by Hu et al. (11) and Zinke et al. (34). Afterwards, two large events of gene flux occurred along the two branches leading to node 45 and node 79, and the extent of gene gain was considerably smaller than that of gene loss, which might shape genome contents of extant organisms of these two orders. Take the last common ancestor of Methanomassiliicoccales (node 45) as an example, some genes encoding the key methyl-coenzyme M reductase complex associated with methane metabolism were gained; in the meanwhile, some important genes such as the the aa3-type cytochrome c oxidase and the bd-type cytochrome c oxidase subunit I were lost (Dataset S7). For node 79, some important genes such as those encoding the CODH/ACS complexes and pyruvate carboxylase were lost, while some other important genes such as those encoding aerobic terminal oxidases (the aa3-type cytochrome c oxidase and the bd-type cytochrome c oxidase subunit I) were retained (Dataset S7), which might contribute greatly to metabolic differentiation between “Ca. Sysuiplasmatales” and other orders. The finding was consistent with the capacity of aerobic respiration and absence of capacity of carbon fixation via WLP for “Ca. Sysuiplasmatales” (Fig. 3). Furthermore, “Ca. Sysuiplasmatales” couldn’t connect glycolysis to TCA cycle through catalyzing carboxylation of
pyruvate to oxalacetic acid by pyruvate carboxylase, which might be easily explained by the above-mentioned gene loss events. Finally, according to the evolutionary path for the genome content of Thermoplasmata, genome contents of most orders might have been shaped by genome streamlining.

Discussion

We proposed the novel order “Ca. Sysuiplasmatales” within class Thermoplasmata based on robust phylogenomic analyses and comparative analysis of their genome-identity values with Methanomassiliicoccales, “Ca. Gimiplasmatales”, and the SG8-5 and the RBG-16-68-12 orders. Unlike Thermoplasmatales, another order from Thermoplasmata whose members have been frequently detected in acid mine environments and exhibit a global distribution (55), “Ca. Sysuiplasmatales” is only distributed in a limited number of AMD and hot spring-related environments (Fig. S1; Dataset S4). All the recovered 16S rRNA sequences of “Ca. Sysuiplasmatales” from NCBI GenBank belong to the same genus as identified in the current study. Although more “Ca. Sysuiplasmatales” clades might remain to be discovered, such efforts may be hampered by their low abundance in nature. It is well accepted that PCR-based 16S rRNA gene surveys often overlook rare taxa in natural microbial communities. Our discovery of “Ca. Sysuiplasmatales” indicates that, while decades of molecular surveys have significantly expanded our knowledge of microbial diversity in AMD environments, novel microbial lineages still remain undiscovered. Technical advances such as environmental genomics hold promise in exploration of this microbial dark matter.

Prediction of metabolic capabilities revealed possible ecological roles of “Ca. Sysuiplasmatales” and the remarkable differences in the lifestyle and metabolic potentials
between “Ca. Sysuiplasmatales” and its closely related clades within Thermoplasmata. The genetic determinants suggest a facultative anaerobic heterotrophic lifestyles capable of sustaining organisms in the acidic environment. This is important because shifts between oxic and anoxic conditions usually occur in mine tailings (56) and surface sediments (57). Among the four orders that are phylogenetically closely related to “Ca. Sysuiplasmatales”, only the RBG-16-68-s12 order harbor the capacity of aerobic respiration. In the domain Archaea, aerobic metabolism has been found in a limited number of lineages including the deep-sea Thaumarchaeota, microaerophiles within the Micrarchaeota and Parvarchaeota, the haloarchaeota in evaporite deposits and brine lakes, and specific members of Thermoplasmatales and Thermoproteales and aerobic and facultatively anaerobic organisms within Sulfolobales (15, 58). Thus, the discovery of aerobic capacity in “Ca. Sysuiplasmatales” and the RBG-16-68-12 order (34) extend the phylogenetic range of this ‘unusual’ feature in Archaea. Notably, “Ca. Sysuiplasmatales” have the genetic potential for utilization and degradation of complex carbon sources including starch and glycogen. Such capacities have also been found in uncultured archaea affiliated with Micrarchaeota and Parvarchaeota populating acid mine environments (2). These findings are reasonable as the sampled habitats are open environments where inputs of external carbon sources are highly possible. Moreover, “Ca. Sysuiplasmatales” could conserve energy through degradation of fatty acids, amino acid metabolism and oxidation of RISCs. While the oxidation of RISCs represents an important strategy to conserve energy in AMD and associated environments, the release of protons from such oxidative reactions could significantly accelerate the acidification processes in situ (20). As rare taxa have been implicated to execute key ecological functions in AMD systems (49, 59), the less abundant “Ca. Sysuiplasmatales” might also act as transcriptionally active taxa and thus contribute to AMD generation via sulfur
oxidation. Notably, all recovered “Ca. Sysuiplasmatales” genomes do not encode the capability of carbon fixation via WLP, although some members of Methanomassiliicoccales, “Ca. Gimiplasmatales” and the RBG-16-68-12 order have been found to possess the bacterial-type \( \text{H}_4\text{F WLP} \) and thus may be capable of energy conservation through inorganic carbon assimilation via the bacterial-type \( \text{H}_4\text{F WLP} \). Additionally, the absence of methanogenesis capacity in the “Ca. Sysuiplasmatales” genomes provides further evidence that Methanomassiliicoccales are the only known anaerobic methanogens in class Thermoplasmata (34), which may be explained by the acquisition of genes encoding the key methyl-coenzyme M reductase complex along the branch leading to the LACA of Methanomassiliicoccales. Finally, we revealed the evolutionary paths leading to the physiological divergence among “Ca. Sysuiplasmatales” and its closely related four orders within Thermoplasmata. The LACA of these five orders might acquire bacterial-type CODH/ACS complexes genes through the ancient interdomain HGT event, strengthening the viewpoint that gene acquisitions from bacteria provided the key innovations in the evolution of Archaea (60). Subsequently, several key loss events of terminal oxidases and CODH/ACS complexes genes shaped the distribution of aerobic metabolism and WLP in these five orders respectively, further supporting the importance of genome streamlining processes in shaping genome contents in Archaea (61).

**Concluding remarks**

Our analyses have revealed the environmental distribution and potential ecological roles of “Ca. Sysuiplasmatales” and provided important insights into the evolution of this novel order and its evolutionary relationships with other closely related orders within class Thermoplasmata. Future cultivation attempts are needed to confirm our metagenomics-based inference of metabolic capacities. Such efforts could be challenging as members of the Thermoplasmata are...
often difficult to isolate. Meanwhile, in situ activity measurements, particularly when combined with multi-comics analyses, could also provide additional support to functions predicted by gene information. Ultimately, the continuing discovery, genomic characterization and subsequent functional validation of novel acidophilic prokaryotic lineages will further our knowledge of the microbial diversity and ecology in the model AMD system.

Materials and Methods

Sampling, DNA extraction and sequencing. The tailings and sediments samples (0-10 cm depth) were collected from the Chengmenshan Copper Mine (29°41’ N, 115°49’ E, samples libTUT1, TUT18, TUT19 and UT3) and Yongping Copper Mine (28°15’ N, 117°50’ E, samples YP2 and YP4), China, respectively. Samples were collected into sterile 50-mL centrifuge tubes and kept on dry ice for transportation to the laboratory. Physicochemical characteristics were determined as described previously (62, 63). All tailings and sediment samples were characterized with low pH (2.6-3.1) and high levels of sulfate (18-111 g kg⁻¹), total iron (4.3-36.1 g kg⁻¹) and heavy metals (Pb, Zn, Cu, Cd) (Dataset S1), which are typical of AMD environments (64, 65). Community genomic DNA of the sediments and tailings samples was extracted using the FastDNA Spin kit (MP Biomedicals, Irvine, CA) according to the manufacturer’s instructions and the method previously described by Tan et al. (66), respectively. Quality-checked DNA was sent to the Guangdong Magigene Company for construction of standard 300-bp fragment libraries, which were subsequently sequenced using 150 bp paired-end sequencing strategy on an Illumina HiSeq 4000 platform.

Metagenomic analysis. Raw metagenomic reads were preprocessed through in-house perl
scripts and Sickle with the parameters “-q 30 -l 50” (67). Then, the high-quality metagenomic reads for each sample were assembled individually using SPAdes (version 3.11.0) with the parameters ‘-k 21, 33, 55, 77, 99 --meta’ or ‘-k 21, 33, 55, 77, 99,127 --meta’ to obtain the scaffolds, which were binned based on sequence composition and scaffold coverage using MetaBAT (version 2.12.1) (68), MaxBin (version 2.2.2) (69) and Concoct (version 0.4.0) (70) with default parameters. The consensus binning software DASTool (version 1.1.2) (71) was utilized to combine the resulting bins to obtain the preliminary genome bins. The phylogenomic analysis based on the concatenated alignment of 122 archaeal-specific conserved marker genes generated by GTDB-Tk (72) was performed using IQtree (version 1.6.10) (73) to determine the phylogenetic placement of these genome bins, and eight of them which formed a distinct clade within Thermoplasmata and grouped phylogenetically closely related to Methanomassiliicoccales, “Candidatus Gimiplasmatales”, the SG8-5 and the RBG-16-68-12 orders were selected. They were re-assembled based on the mapped reads using BBMap (74) as previously reported (48), and were manually curated to remove contaminations. Their genome completeness, contamination, and strain heterogeneity were assessed using CheckM (version 1.1.3) (75). These eight curated genomes were retained for further analyses. The relative abundance of each MAG was determined by calculating the relative abundance of its rpS3 as previously reported (76). For each sample, all rpS3 proteins were identified by AMPHORA2 and clustered by USEARCH with 99% identity (77). Then we selected the representative scaffold in protein cluster of each rpS3 as the reads mapping target for abundance calculation, which was performed by BBMap with the parameters “minid = 0.97, local = t”. The relative abundance of each bin was determined as the coverage of the corresponding rpS3 divided by the total coverage of all rpS3 in the community.
Functional annotation and metabolic reconstruction of MAGs. The open reading frames (ORFs) in scaffolds from each MAG were determined using Prodigal (version 2.6.3) with the “-p single” option, and the generated ORFs were annotated through comparisons with the KEGG, EggNOG, NCBI-nr and Pfam databases using DIAMOND with E-value ≤ 1e-5 (78). The repertoires of Carbohydrate-active enzymes (CAZymes) were investigated through the dbCAN2 meta server (79). Ribosomal RNA genes and transfer RNA genes were predicted using Barrnap (https://github.com/Victorian-Bioinformatics-Consortium/barrnap) and tRNAscan-SE 2.0 web server, respectively (80). In addition, the assignment of functional domain to all proteins was conducted through the EBI InterProScan server (81). Metabolic pathways were constructed for the eight novel order MAGs based on these gene annotation results.

Phylogenomic and phylogenetic analyses. Phylogenetic analysis of 16S rRNA genes was conducted using sequences from all novel order genomes and other archaeal genomes, as well as novel order related sequences from NCBI GenBank. Sequences were aligned using the SINA alignment algorithm on the SILVA web interface, and then the alignment was filtered to remove columns with more than 95% gaps using TrimAL v1.4 (82). The maximum likelihood tree was constructed using IQtree (version 1.6.10) with the parameters “GTR+F+R4 -alrt 1000 -bb 1000” (73). Phylogenomic analysis was performed with 16 ribosomal proteins (83) and 122 archaeal-specific conserved marker genes identified by GTDB-Tk (72), respectively. The MAGs with less than 8 of the 16 ribosomal proteins were removed. The sixteen ribosomal proteins were aligned individually using MUSCLE (version 3.8.31) with default parameters (84), and subsequently were filtered using TrimAL v1.4 with the parameters “-gt 0.95 -cons 50” (82). Then, the filtered
alignments were concatenated in order and the phylogenomic tree was inferred based on the concatenated alignment using IQtree (version 1.6.10) with the parameters "LG+R10 -alrt 1000 -bb 1000" (73). Meanwhile, the concatenated alignment of 122 archaeal-specific conserved marker genes was generated using GTDB-Tk and the phylogenomic tree was constructed based on the concatenated alignment using IQtree (version 1.6.10) with the parameters "LG+F+R10 -alrt 1000 -bb 1000". For phylogenetic analysis of CODH/ACS complexes genes, reference datasets were derived from Adam et al. (27) and Hu et al. (11). Combined with the AcsE-CdhDE genes detected in this study, sequences were aligned and trimmed using MUSCLE and TrimAL with the same parameters as above. The maximum-likelihood phylogenetic tree was constructed using IQtree (version 1.6.10) with the parameters "LG+F+R8 -alrt 1000 -bb 1000". The phylogenetic tree was rooted according to ref. 63. All trees were uploaded to iTOL for visualization and formatting (85).

**Comparative genomics.** Representative genomes of Thermoplasmata in the GTDB database were downloaded (86). The values of average amino acid identity and OrthoANI among the genomes of this novel order and its closely related Methanomassiliicoccales, “Candidatus Gimiplasmatales”, the SG8-5 and the RBG-16-68-12 order were calculated by CompareM (https://github.com/dparks1134/CompareM) and OrthoANIu (87), respectively. For COUNT analysis, only genomes from Thermoplasmata with completeness >85% were taken into consideration, and 78 draft genomes from GTDB and eight novel order MAGs were kept for further analyses. An all-against-all genome BLAST was employed to generate the reciprocal best BLAST hits (rBBHs) based on the thresholds E-value < 1e-10 and local amino acid identity ≥ 25%. Then these protein pairs were globally aligned using the Needleman-Wunsch algorithm in
EMBOSS v6.5.7 based on the threshold ≥ 25% global amino acid identity (60), and MCL (-I 1.4) was employed to yield protein clusters based on rBBHs (88), thereby generating 33527 gene families with 21577 orphans. The genome tree (based on 122 archaeal-specific conserved marker genes) for COUNT was constructed as described above and evolutionary histories of these archaeal lineages were inferred using COUNT v9.1106 with Dollo parsimony (89).

**Data availability.** All sequencing data generated in this study have been deposited in the NCBI database under BioProject number PRJNA719487. In particular, the raw metagenomic sequences have been deposited in the Sequence Read Archive under accession numbers SRR14316976, SRR14318494, SRR14320033 and SRR14318403, and the eight described MAGs have been deposited in the GenBank database under accession numbers JAGVSJ000000000, JAHBML000000000, JAHDZZ000000000, JAHEAA000000000, JAHEAB000000000, JAHEAC000000000, JAHEAD000000000, JAHEAE000000000.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (nos. 31570500, 31870111, and 40930212).

**References**


**Figure and Table legends**

**Fig. 1.** Phylogenomic tree of novel order based on 122 archaeal-specific conserved marker genes, orthoANI and 16S rRNA gene similarity values. White spaces refer to cases where no 16S rRNA gene was identified or the length of the 16S rRNA gene was too short to be used for 16S rRNA sequence similarity calculation. The tree was constructed based on the concatenated alignment using IQ-TREE with 1000 ultrafast bootstrapping iterations. The concatenated alignment was generated by GTDB-Tk. Support values more than 75% are shown using black solid circles. Crenarchaeota was selected as the outgroup.

**Fig. 2.** Metabolic potentials of novel order. Genes identified in “Ca. Sysuiplasma acidicola”, “Ca.
Sysuiplasma superficiale” and “Ca. Sysuiplasma jiujiangense” are represented by light green, orange and dark green solid circles, respectively. The copy numbers of each gene in each genome are listed in Dataset S5.

Fig. 3. Occurrence of key proteins of interest in the MAGs of novel order, Methanomassiliicoccales, “Ca. Gimiplasmatales”, the RBG-16-68-12 order and the SG8-5 order. The MAGs were grouped according to their phylogenetic relationships and the gene names were grouped by functional categories. The copy numbers of each gene in each genome were provided in Dataset S6. The genes composed of multiple subunits were marked as present if half or more than half of the subunits were identified.

Fig. 4. Ancestral genome content reconstruction of Thermoplasmata including novel order. The numbers of gene gain and loss events were marked accordingly on the nodes and tips of the phylogenomic tree. The COG category information of the gained and lost genes for the key nodes was shown using the pie chart. Some key gene gain and lost events were also marked. The complete topology of the phylogenomic tree and ancestral state reconstruction results were shown in Fig. S5. A list of gained and lost genes for the nodes and tips were shown in Dataset S7.

Table 1. Genomic information of the Thermoplasmata novel order MAGs retrieved from metagenomes.
### TABLE 1 Genomic information of the Thermoplasmata new order MAGs retrieved from metagenomes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size (bp)</td>
<td>1,972,931</td>
<td>1,960,374</td>
<td>1,981,594</td>
</tr>
<tr>
<td>No. of scaffolds</td>
<td>35</td>
<td>66</td>
<td>73</td>
</tr>
<tr>
<td>N50 (bp)</td>
<td>847 39</td>
<td>490 42</td>
<td>471 86</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>50.93</td>
<td>50.95</td>
<td>51.05</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.01</td>
<td>99.01</td>
<td>97.25</td>
</tr>
<tr>
<td>Contamination (%)</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>No. of predicted genes</td>
<td>2,000</td>
<td>2,012</td>
<td>2,083</td>
</tr>
<tr>
<td>No. of genes annotated by KEGG</td>
<td>974 (48.7%)</td>
<td>967 (48.1%)</td>
<td>979 (48.1%)</td>
</tr>
<tr>
<td>No. of genes annotated by COG</td>
<td>1538 (76.8%)</td>
<td>1511 (76.1%)</td>
<td>1577 (77.3%)</td>
</tr>
<tr>
<td>No. of genes annotated by Pfam*</td>
<td>1363 (68.4%)</td>
<td>1375 (68.4%)</td>
<td>1395 (68.5%)</td>
</tr>
<tr>
<td>No. of 16S rRNAs</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>No. of tRNAs</td>
<td>118</td>
<td>75</td>
<td>128</td>
</tr>
<tr>
<td>No. of CRISPR loci*</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Relative abundance (%)</td>
<td>0.07</td>
<td>0.06</td>
<td>0.28</td>
</tr>
</tbody>
</table>

*The completeness and contamination of these new order genomes were estimated by CheckM.

* Functional annotation was against different databases via Diamond (e-value ≤ 10^-5).

* CRISPR loci of these genomes were annotated using CRT.