The structure and function of soil archaea across biomes

Robert Starke[1,*], José A. Siles[2], Maysa Lima Parente Fernandes[1], Kay Schallert[3], Dirk Benndorf[4], Cesar Plaza[5], Nico Jehmlich[6], Manuel Delgado-Baquerizo[7], Felipe Bastida[2]

[1]Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083, 14220 Praha 4, Czech Republic
[2]CEBAS-CSIC, Campus Universitario de Espinardo, Murcia, E-30100, Spain
[3]Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany
[4]Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany
[5]Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, Serrano 115 bis, 28006 Madrid, Spain
[6]Helmholtz-Centre for Environmental Research GmbH – UFZ, Department of Molecular Systems Biology, Permoserstraße 15, 04318 Leipzig, Germany
[7]Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Sevilla, Spain

*corresponding author: robert.starke@biomed.cas.cz

Classification: Biological Sciences (Agricultural Sciences, Ecology)
Keywords: soil; 16S rRNA gene amplicon sequencing; metaproteomics; archaea; climate; vegetation

The authors declare no conflict of interest.

Abstract

We lack a predictive understanding of the environmental drivers determining the structure and function of archaeal communities as well as the proteome associated with these important soil organisms. Here, we characterized the structure (by 16S rRNA gene sequencing) and function (by metaproteomics) of archaea from 32 soil samples across terrestrial ecosystems with contrasting climate and vegetation types. Our multi-"omics" approach unveiled that genes from *Nitrosophaerales* and *Thermoplasmata* dominated soils collected from four continents, and that archaea comprise 2.3±0.3% of microbial proteins in these soils. Aridity positively correlated with the proportion of *Nitrosophaerales* genes and the number of archaeal proteins. The interaction of climate x vegetation shaped the functional profile of the archaeal community. Our study provides novel insights into the structure and function of soil archaea across climates, and highlights that these communities may be influenced by increasing global aridity.

Cross-biome studies on soil microbes, their environmental drivers and their contribution to ecosystem services have mainly focused on bacteria and fungi. However, archaea are ubiquitous in terrestrial environments [1,2] and have key roles in global carbon (C) (e.g., methanogenesis or CO_2 fixation) and nitrogen (N) (e.g., N_2 fixation or oxidation of ammonia) cycles [3]. Archaea make up to 10% of prokaryotes in soil based on the sequencing of the 16S rRNA gene from extracted DNA [2,4] but were of higher relative abundance in extreme habitats, i.e. with high acidity and low temperatures [5]. The actual contribution of soil archaea to ecosystem functioning can be better understood by complementary "omic" approaches including metaproteomics. Thus, The identification of proteins offers important advantages because proteins are a more realistic surrogates of functionality [6]. Here, we describe the environmental drivers shaping the structure and functionality of the archaeal soil community. We applied 16S rRNA gene amplicon sequencing of extracted DNA and metaproteomics to a subset of 32 soils collected from four continents (**Figure S1**) with different climate and vegetation types that belonged to a larger set of soil samples [7,8] with the aim to extract the amplicon sequences on OTU level and meta-proteins associated with soil archaea (DNA and protein extraction, measurement, bioinformatic approaches and analysis of environmental parameters are described in the supplementary material). Spearman correlations of the proportions of archaeal classes or orders from amplicon sequencing and total archaeal proteins were performed against 11 environmental variables including important biological, chemical, and physical parameters (**Table S1**). Regression models were calculated for the variables with significant Spearman correlations. Permutational multivariate analysis of variance (PERMANOVA) to elucidate the impact of climate and vegetation types was computed at the class or order level for taxonomy and at the class level of cluster of orthologous groups (COGs) for functionality. Even though deeper taxonomic levels are possible to assess in the amplicon data, the low abundance of archaeal proteins made the use of deeper functional levels unfeasible, which is why we used broad levels for both structure and function. Similar to bacterial and fungal communities [7,9,10], we expect climate and vegetation to impact the composition of the archaeal community via changes in C and N stocks, pH or plant cover.

The archaeal community was dominated by the thaumarchaeote *Nitrososphaerales* and the euryarchaeote *Thermoplasmata* making up more than 80% of all archaeal sequences

regardless of climate or vegetation types (**Figure 1**), which is in agreement with previous studies on soil archaea [4,11,12]. PERMANOVA revealed that neither climate (R^2 = 0.039, F = 1.243, *P* = 0.272) nor vegetation type (R^2 = 0.099, F = 1.546, *P* = 0.192) had a significant influence on the community structure. However, we were able to identify multiple significant associations between the proportion of archaeal taxa and individual environmental factors. For example, we found a negative correlation between the proportion of *Nitrososphaerales* and the content of organic C associated with free-light fractions (i.e., relatively available forms of C in soils), which suggests that this group of archaea often affiliated with slow growth rates might prefer low C environments. The proportion of *Nitrososphaerales* negatively correlated with aridity index and was therefore especially abundant in the most arid ecosystems (**Figure 1**). Increasing aridity was reported to promote shifts in the soil niches by nutrient depletion, soil salinization and N-losses [13] and the structure of archaeal communities [4,14,15]. We also found that the proportion of taxa from the phylum *Euryarchaeota* previously reported to be more abundant in extreme environments [16,17] were similar in dryland and mesic climates, and unaffected by any environmental variables.

An average (± standard error) of 7,751±600 meta-proteins were identified from 59,007±3,084 spectra per sample (**Table S1**). Most proteins were affiliated with bacteria, dominated by *Proteobacteria* (62.2±0.9%), *Actinobacteria* (10.6±0.7%), and *Firmicutes* (8.6±0.7%) (**Table S2**), consistent with other large-scale genomic soil surveys [18,19]. In total, 416 meta-proteins associated with soil archaea (**Table S3**). We focused on the functionality of these proteins as little is known about archaea and only a few archaeal protein coding sequences are present among all sequences in UniProtKB/SwissProt. However, general functionality of COGs should be equally well described as these processes are essential to both archaea and bacteria. The proportion of archaeal proteins ranged between 0.6% and 6.8% of all soil proteins with an average (± standard error) of 2.3±0.3%, in the range of the 16S rRNA gene abundance reported for soil archaea of up to 10% [2,4]. Methyl-coenzyme M reductase (*MCRA*, relative spectral abundance = 2.3%), Tyrosine--tRNA ligase (*SYY*, relative spectral abundance = 1.7%) and DNA protection during starvation (*DPS*, relative spectral abundance = 1.7%) were the most common archaeal proteins found across global biomes. *MCRA* is central to methanogenic pathways [24] performed by strictly anaerobic archaea who convert a restricted number of substrates to methane [25]. The relatively large proportion of *MCRA* proteins in soil highlights the potential influence of soil archaea in controlling methane production across climates. Vegetation, but not climate (drylands vs. mesic), was significantly shaping the proportion of proteins assigned to archaea (Tukey's HSD-test). We found a higher relative abundance of archaeal proteins in shrublands than in forests and grasslands (**Figure 2a**). These results suggested that changes in land use that increase the proportion of shrubs (e.g., via shrub encroachment) can have direct impact on the number of archaea in microbial communities. Our results further showed that aridity helps to explain the distribution of archaeal proteins across soils from contrasting ecosystems (**Figure 2b & 2c**). In fact, the increase of the number of genes from *Nitrososphaerales* and of archaeal proteins with increasing aridity aligns with the loss in biodiversity of bacteria and fungi with increasing aridity [26]. Further, PERMANOVA showed that only the interaction of climate x vegetation (R^2 = 0.145, F = 2.450, *P* = 0.019) but not climate (R^2 = 0.025, F = 0.840, *P* = 0.515) or vegetation type (R^2 = 0.061, F = 1.030, *P* = 0.403) had a significant impact on the functionality of archaea as estimated with metaproteomics. Particularly in dryland environments

with forest vegetation (n = 5), proteins related to protein biosynthesis and glycolysis were most abundant while proteins related to one-C metabolism and gluconeogenesis were specific for grassland vegetation (n = 5) (**Figure 2d**). Otherwise in mesic environments, proteins related to transport of hydrogen and sodium ions were most abundant with forest vegetation (n = 15) while proteins for biosynthesis were highly abundant with grassland vegetation (n = 2). It is still unclear what the differential translation of proteins means for the archaeal community specifically and the microbial community but only the combination of climate and vegetation shaped their proportions.

In summary, our results constitute a first step to unveil the environmental drivers of the structure (16S rRNA gene sequencing) and function (metaproteomics) of the soil archaeal community across biomes. We observed that climatic features such as aridity might influence the proportion of dominant archaeal groups and of archaeal proteins, highlighting the impact of climate on the archaeal community. Our work indicates that the inclusion of archaea in future research of ecosystem functioning has critical implications to understand how these ecosystems respond to global change. Admittedly, our proteomic results could be biased by less stringent search parameters (10% FDR). The FDR concept for protein identification was originally established for pure culture proteomics [27], allowing to compare different mass spectrometers and database search algorithms with a defined threshold of 1% [28]. However, searches against large databases, such as the database used in this study, not only require long computation times but also decrease the number of identified proteins due to the overestimation of the FDR [29]. In fact, the limitation of the target-decoy controlled FDR approach in combination with large databases was responsible for missing valuable protein identifications [30], which makes FDR's higher than 1% common in metaproteomic approaches [31–34]. Therefore, using 10% FDR in combination with all known protein coding sequences from UniProtKB/SwissProt was a feasible way to obtain archaeal taxonomical and functional information in our study without having ad hoc metagenomes that would allow for a more stringent search.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO), and FEDER funds, through the project AGL2017–85755-R MINECO/AEI/FEDER; i-LINK+ 2018 (LINKA20069) from CSIC, and funds from "Fundación Séneca" from Murcia Province (19896/GERM/15). We would like to thank the researchers involved in the CLIMIFUN project for the help with soil sampling. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 702057. M.D-B. is supported by a Ramón y Cajal grant from the Spanish Government (agreement no. RYC2018-025483-I). KS was supported by the German Federal Ministry of Education and Research (de.NBI network, (project MetaProtServ, grant no. de-NBI-039031L0103). The Czech Science Foundation (20-02022Y) supported RS and MLPF.

Compliance with ethical standards

The authors declare that they have no conflict of interest.

References

[1] S. Timonen, M. Bomberg, Archaea in dry soil environments, in: Phytochem. Rev., 2009.

doi:10.1007/s11101-009-9137-5.
[2] E.F. DeLong, N.R. Pace, Environmental diversity of bacteria and archaea, Syst. Biol. (2001). doi:10.1080/106351501750435040.
[3] S. Leininger, T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster, C. Schleper, Archaea predominate among ammonia-oxidizing prokaryotes in soils, Nature. (2006). doi:10.1038/nature04983.
[4] S.T. Bates, D. Berg-Lyons, J.G. Caporaso, W.A. Walters, R. Knight, N. Fierer, Examining the global distribution of dominant archaeal populations in soil, ISME J. (2011). doi:10.1038/ismej.2010.171.
[5] A.A. Korzhenkov, S. V. Toshchakov, R. Bargiela, H. Gibbard, M. Ferrer, A. V. Teplyuk, D.L. Jones, I. V. Kublanov, P.N. Golyshin, O. V. Golyshina, Archaea dominate the microbial community in an ecosystem with low-to-moderate temperature and extreme acidity, Microbiome. (2019). doi:10.1186/s40168-019-0623-8.
[6] R. Starke, N. Jehmlich, F. Bastida, Using proteins to study how microbes contribute to soil ecosystem services: The current state and future perspectives of soil metaproteomics, J. Proteomics. (2018). doi:10.1016/j.jprot.2018.11.011.
[7] M. Delgado-Baquerizo, R.D. Bardgett, P.M. Vitousek, F.T. Maestre, M.A. Williams, D.J. Eldridge, H. Lambers, S. Neuhauser, A. Gallardo, L. García-Velázquez, O.E. Sala, S.R. Abades, F.D. Alfaro, A.A. Berhe, M.A. Bowker, C.M. Currier, N.A. Cutler, S.C. Hart, P.E. Hayes, Z.Y. Hseu, M. Kirchmair, V.M. Peña-Ramírez, C.A. Pérez, S.C. Reed, F. Santos, C. Siebe, B.W. Sullivan, L. Weber-Grullon, N. Fierer, Changes in belowground biodiversity during ecosystem development, Proc. Natl. Acad. Sci. U. S. A. (2019). doi:10.1073/pnas.1818400116.
[8] F. Bastida, C. García, N. Fierer, D.J. Eldridge, M.A. Bowker, S. Abades, F.D. Alfaro, A. Asefaw Berhe, N.A. Cutler, A. Gallardo, L. García-Velázquez, S.C. Hart, P.E. Hayes, T. Hernández, Z.Y. Hseu, N. Jehmlich, M. Kirchmair, H. Lambers, S. Neuhauser, V.M. Peña-Ramírez, C.A. Pérez, S.C. Reed, F. Santos, C. Siebe, B.W. Sullivan, P. Trivedi, A. Vera, M.A. Williams, J. Luis Moreno, M. Delgado-Baquerizo, Global ecological predictors of the soil priming effect, Nat. Commun. (2019). doi:10.1038/s41467-019-11472-7.
[9] T.W. Crowther, J. van den Hoogen, J. Wan, M.A. Mayes, A.D. Keiser, L. Mo, C. Averill, D.S. Maynard, The global soil community and its influence on biogeochemistry, Science (80-.). (2019). doi:10.1126/science.aav0550.
[10] M. Bahram, F. Hildebrand, S.K. Forslund, J.L. Anderson, N.A. Soudzilovskaia, P.M. Bodegom, J. Bengtsson-Palme, S. Anslan, L.P. Coelho, H. Harend, J. Huerta-Cepas, M.H. Medema, M.R. Maltz, S. Mundra, P.A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome, Nature. (2018). doi:10.1038/s41586-018-0386-6.
[11] J.W. Neilson, K. Califf, C. Cardona, A. Copeland, W. van Treuren, K.L. Josephson, R. Knight, J.A. Gilbert, J. Quade, J.G. Caporaso, R.M. Maier, Significant Impacts of Increasing Aridity on the Arid Soil Microbiome, MSystems. (2017). doi:10.1128/msystems.00195-16.
[12] R. Isoda, S. Hara, T. Tahvanainen, Y. Hashidoko, Comparison of archaeal communities in mineral soils at a boreal forest in Finland and a cold-temperate forest in Japan, Microbes Environ. (2017). doi:10.1264/jsme2.ME17100.
[13] M. Huang, L. Chai, D. Jiang, M. Zhang, Y. Zhao, Y. Huang, Increasing aridity affects soil

archaeal communities by mediating soil niches in semi-arid regions, Sci. Total Environ. (2019). doi:10.1016/j.scitotenv.2018.07.305.

[14] B. Chaban, S.Y.M. Ng, K.F. Jarrell, Archaeal habitats - From the extreme to the ordinary, Can. J. Microbiol. (2006). doi:10.1139/w05-147.

[15] C. Gubry-Rangin, B. Hai, C. Quince, M. Engel, B.C. Thomson, P. James, M. Schloter, R.I. Griffiths, J.I. Prosser, G.W. Nicol, Niche specialization of terrestrial archaeal ammonia oxidizers, Proc. Natl. Acad. Sci. U. S. A. (2011). doi:10.1073/pnas.1109000108.

[16] P. Cao, L.M. Zhang, J.P. Shen, Y.M. Zheng, H.J. Di, J.Z. He, Distribution and diversity of archaeal communities in selected Chinese soils, FEMS Microbiol. Ecol. (2012). doi:10.1111/j.1574-6941.2011.01280.x.

[17] K. Takai, T. Komatsu, F. Inagaki, K. Horikoshi, Distribution of Archaea in a Black Smoker Chimney Structure, Appl. Environ. Microbiol. (2001). doi:10.1128/AEM.67.8.3618-3629.2001.

[18] M. Bahram, F. Hildebrand, S.K. Forslund, J.L. Anderson, N.A. Soudzilovskaia, P.M. Bodegom, J. Bengtsson-Palme, S. Anslan, L.P. Coelho, H. Harend, J. Huerta-Cepas, M.H. Medema, M.R. Maltz, S. Mundra, P.A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome, Nature. (2018). doi:10.1038/s41586-018-0386-6.

[19] M. Delgado-Baquerizo, A.M. Oliverio, T.E. Brewer, A. Benavent-González, D.J. Eldridge, R.D. Bardgett, F.T. Maestre, B.K. Singh, N. Fierer, A global atlas of the dominant bacteria found in soil, Science (80-.). (2018). doi:10.1126/science.aap9516.

[20] M.C. Portillo, J.W. Leff, C.L. Lauber, N. Fierer, Cell size distributions of soil bacterial and archaeal taxa, Appl. Environ. Microbiol. (2013). doi:10.1128/AEM.02710-13.

[21] F. Bastida, I.F. Torres, J.L. Moreno, P. Baldrian, S. Ondoño, A. Ruiz-Navarro, T. Hernández, H.H. Richnow, R. Starke, C. García, N. Jehmlich, The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils, Mol. Ecol. 25 (2016). doi:10.1111/mec.13783.

[22] F. Bastida, I.F. Torres, M. Andrés-Abellán, P. Baldrian, R. López-Mondéjar, T. Větrovský, H.H. Richnow, R. Starke, S. Ondoño, C. García, F.R. López-Serrano, N. Jehmlich, Differential sensitivity of total and active soil microbial communities to drought and forest management, Glob. Chang. Biol. 23 (2017). doi:10.1111/gcb.13790.

[23] R. Starke, F. Bastida, J. Abadía, C. García, E. Nicolás, N. Jehmlich, Ecological and functional adaptations to water management in a semiarid agroecosystem: A soil metaproteomics approach, Sci. Rep. 7 (2017). doi:10.1038/s41598-017-09973-w.

[24] Z. Lyu, C.W. Chou, H. Shi, L. Wang, R. Ghebreab, D. Phillips, Y. Yan, E.C. Duin, W.B. Whitman, Assembly of methyl coenzyme M reductase in the methanogenic archaeon Methanococcus maripaludis, J. Bacteriol. (2018). doi:10.1128/JB.00746-17.

[25] R. Hedderich, W.B. Whitman, Physiology and Biochemistry of the Methane-Producing Archaea, in: The Prokaryotes, 2006. doi:10.1007/0-387-30742-7_34.

[26] F.T. Maestre, R. Salguero-Gómez, J.L. Quero, It is getting hotter in here: Determining and projecting the impacts of global environmental change on drylands, Philos. Trans. R. Soc. B Biol. Sci. (2012). doi:10.1098/rstb.2011.0323.

[27] J.E. Elias, S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat. Methods. (2007).

doi:10.1038/nmeth1019.
[28] K. Barnouin, Guidelines for experimental design and data analysis of proteomic mass spectrometry-based experiments, Amino Acids. (2011). doi:10.1007/s00726-010-0750-9.
[29] R. Heyer, K. Schallert, R. Zoun, B. Becher, G. Saake, D. Benndorf, Challenges and perspectives of metaproteomic data analysis, J. Biotechnol. (2017). doi:10.1016/j.jbiotec.2017.06.1201.
[30] T. Muth, C.A. Kolmeder, J. Salojärvi, S. Keskitalo, M. Varjosalo, F.J. Verdam, S.S. Rensen, U. Reichl, W.M. de Vos, E. Rapp, L. Martens, Navigating through metaproteomics data: A logbook of database searching, Proteomics. (2015). doi:10.1002/pmic.201400560.
[31] T. Schneider, K.M. Keiblinger, E. Schmid, K. Sterflinger-Gleixner, G. Ellersdorfer, B. Roschitzki, A. Richter, L. Eberl, S. Zechmeister-Boltenstern, K. Riedel, Who is who in litter decomposition Metaproteomics reveals major microbial players and their biogeochemical functions, ISME J. (2012). doi:10.1038/ismej.2012.11.
[32] F. Bastida, T. Hernández, C. García, Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods, J. Proteomics. (2014). doi:10.1016/j.jprot.2014.02.006.
[33] K.M. Keiblinger, I.C. Wilhartitz, T. Schneider, B. Roschitzki, E. Schmid, L. Eberl, K. Riedel, S. Zechmeister-Boltenstern, Soil metaproteomics - Comparative evaluation of protein extraction protocols, Soil Biol. Biochem. (2012). doi:10.1016/j.soilbio.2012.05.014.
[34] S.M. Figarska, S. Gustafsson, J. Sundström, J. Ärnlöv, A. Mälarstig, S. Elmståhl, T. Fall, L. Lind, E. Ingelsson, Associations of circulating protein levels with lipid fractions in the general population, Arterioscler. Thromb. Vasc. Biol. (2018). doi:10.1161/ATVBAHA.118.311440.

Figures

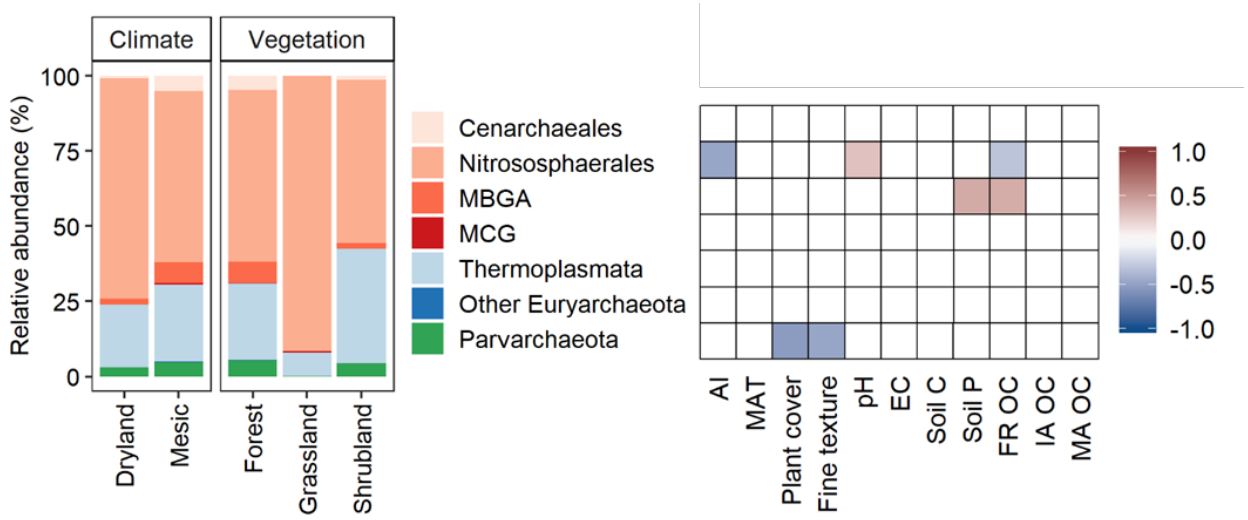

Figure 1: Relative abundance of archaeal taxa in the 16S rRNA gene amplicon sequencing as affected by climate and vegetation, with taxa from the phylum *Thaumarchaeota* in shades of red, *Euryarchaeota* in shades of blue and *Parvarchaeota* in shades of green, and their significant Spearman correlation with environmental variables ($P < 0.05$). MA OC, IA OC, and FR OC stand for mineral-associated, intra-aggregate and free organic carbon fraction, respectively, P for phosphorus, C for carbon, MAT for mean annual temperature, and AI for aridity index.

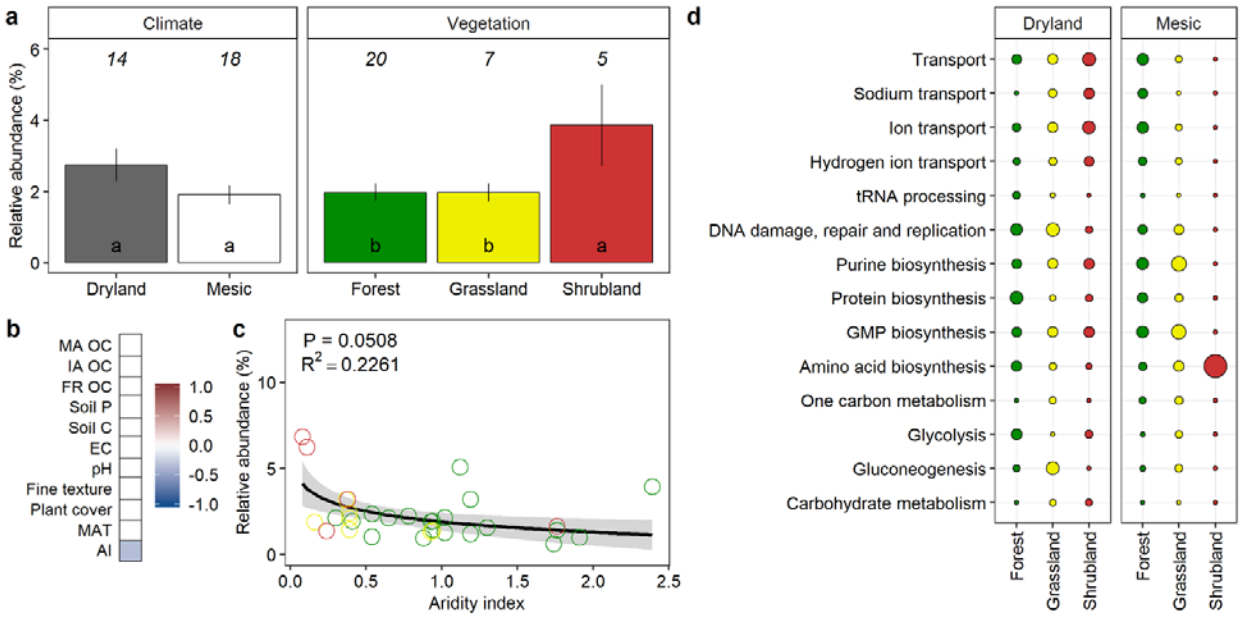

Figure 2: The relative abundance of archaeal proteins across climate and vegetation types as average with standard errors (a). Italic numbers represent the sample size for each variable. Data followed by the same letter is not statistically different according to the HSD test ($P < 0.05$). Spearman correlations of the relative abundance of archaeal proteins with environmental variables ($P < 0.05$) (b). Relationship between relative abundance of archaeal proteins and aridity index with colors representing the different vegetation types (c). The functional profile of archaeal proteins on the class level of cluster of orthologous groups across the interaction of climate x vegetation types with bubble size as relative abundance (d). MA OC, IA OC, and FR OC stand for mineral-associated, intra-aggregate and free organic carbon fraction, respectively, P for phosphorus, C for carbon, MAT for mean annual temperature, and AI for aridity index.