Journal of Advanced Research 19 (2019) 39–48

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original article

Novel insights into plant-associated archaea and their functioning in arugula (*Eruca sativa* Mill.)

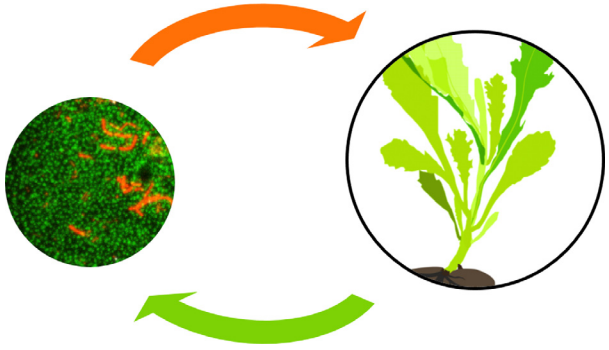
Julian Taffner, Tomislav Cernava, Armin Erlacher, Gabriele Berg *

Institute of Environmental Biotechnology, Graz University of Technology, Petersgasse 12, 8010 Graz, Austria

HIGHLIGHTS

- The first insights into archaea associated with a traditional plant are provided.
- Archaea showed habitat-specific colonization of arugula.
- The functional capacities of plant-associated archaea were revealed.
- Indications of archaea-host interactions were found.
- A basis is provided for developments that will benefit plant and human health.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 December 2018
Revised 24 April 2019
Accepted 24 April 2019
Available online 30 April 2019

Keywords:
Archaea
Eruca sativa Mill.
Brassicaceae
Metagenomics
Microbiome
Holobiont

ABSTRACT

A plant's microbiota has various implications for the plant's health and performance; however, the roles of many microbial lineages, particularly *Archaea*, have not been explored in detail. In the present study, analysis of archaea-specific 16S rRNA gene fragments and shotgun-sequenced metagenomes was combined with visualization techniques to obtain the first insights into the archaeome of a common salad plant, arugula (*Eruca sativa* Mill.). The archaeal communities associated with the soil, rhizosphere and phyllosphere were distinct, but a high proportion of community members were shared among all analysed habitats. Soil habitats exhibited the highest diversity of *Archaea*, followed by the rhizosphere and the phyllosphere. The archaeal community was dominated by *Thaumarchaeota* and *Euryarchaeota*, with the most abundant taxa assigned to *Candidatus Nitrososcosmicus*, species of the 'Soil Crenarchaeotic Group' and, interestingly, *Methanosarcina*. Moreover, a large number of archaea-assigned sequences remained unassigned at lower taxonomic levels. Overall, analysis of shotgun-sequenced total-community DNA revealed a more diverse archaeome. Differences were evident at the class level and at higher taxonomic resolutions when compared to results from the 16S rRNA gene fragment amplicon library. Functional assessments primarily revealed archaeal genes related to response to stress (especially oxidative stress), CO_2 fixation, and glycogen degradation. Microscopic visualizations of fluorescently labelled archaea in the phyllosphere revealed small scattered colonies, while archaea in the rhizosphere were found to be embedded within large bacterial biofilms. Altogether, *Archaea* were identified as a rather small but niche-specific component of the microbiomes of the widespread leafy green plant arugula.

Introduction

Leafy green plants have become a central element of a healthy diet, mainly due to their high fibre content but also due to the

Peer review under responsibility of Cairo University.
* Corresponding author.
E-mail address: gabriele.berg@tugraz.at (G. Berg).

https://doi.org/10.1016/j.jare.2019.04.008
2090-1232/

various micronutrients they contain. In modern diets, arugula (*Eruca sativa* Mill.) stands out due to its peppery, pungent taste that stems from various glucosinolates and other sulphur-containing compounds in the edible parts [1]. In addition to their flavour, isothiocyanates that are formed during the degradation of glucosinolates are thought to be involved in cancer prevention [2]. Arugula belongs to the *Brassicaceae* family and is commonly known as rucola (or garden rocket); it originated in the Mediterranean and has been cultivated at least since Roman and ancient Egyptian times [3]. In traditional medicine, arugula is used as a medicinal plant to treat disorders of the digestive system and has several other medical indications as well as aphrodisiac properties [4]. Moreover, various cultivars are broadly accepted in Western cuisine, where they are used in their raw form in various types of salads. Arugula has also been associated with *Salmonella* Thompson outbreaks causing severe illnesses in humans [5]. Therefore, it is important to understand the entire arugula microbiome because its structure, network and function as well as its colonization stability are crucial factors affecting outbreaks and the functioning of the holobiont [6]. To date, various important plant species-specific microbial key players have been identified; however, the focus of most studies is on bacterial and fungal communities. For example, the bacteriome of various *Brassicaceae* plants, including *E. sativa*, was previously identified [7,8]. It was shown that the phyllosphere of arugula harbours higher proportions of antibiotic-resistant bacteria than its rhizosphere and the surrounding soil [8]. However, details related to archaeal communities associated with *Brassicaceae* plants and their functioning still remain largely unknown.

Archaea have been identified as interactive components of complex microbiomes, such as those in the environment or associated with the gastrointestinal tract of animals, the gut and skin of humans and even the rhizosphere and endosphere of plants [9–12]. However, the function of archaea and their structural interactions with their host and other microorganisms remain mostly unclear, mainly due to methodological limitations. On plant hosts, *Archaea* have been found to colonize the rhizosphere and the endosphere at high abundances, whereas the phyllosphere is less colonized [13,14]. These different colonization patterns are influenced by different abiotic conditions but also by biotic factors, such as competition and interactions with bacteria and fungi [11,15]. Recent studies on the natural vegetation of alpine bogs revealed that the plant genotype also influences colonization by *Archaea*. On bog vegetation *Archaea* were further found to have the potential to interact with plants. These potential interactions based on functions such as plant growth promotion through auxin biosynthesis, nutrient supply, and protection against abiotic stress were identified by metagenomic mining [12]. Although factors influencing archaeal functionality in rice roots have been analysed [13], the interactions of archaea with cultivated plants remain mostly unclear. However, due to the ubiquitous occurrence of archaea and their important functions in healthy natural vegetation, *Archaea* presumably play a role in cultivated plants.

The objective of our study was to analyse the colonization patterns of *Archaea* with respect to micro-niche specificity on a widespread leafy green plant and to further increase our understanding of their role and functionality in plants in general. In addition, we aimed to fill gaps in our understanding of the microbiome-host interactions between *Archaea* and plants. Therefore, we analysed the specific archaeal communities of each habitat (soil, rhizosphere, and phyllosphere) of *E. sativa* grown under non-intensive horticultural conditions. Samples were obtained from home gardens (Graz, Austria) and analysed with a complementary approach combining metagenomics and targeted sequencing of the V4 region of the 16S rRNA gene fragment.

Material and methods

Sampling of arugula plants and isolation of total-community DNA

Arugula plants were grown in garden soil (hereafter referred to as bulk soil) in a suburban region of Graz (Austria; approx. 47°4′13″N, 15°28′19″E). Plants were watered by above-ground irrigation with a water hose. Plants were harvested in their final stage of leaf development in July. The plant leaves and their short stalks (edible plant parts) are called the phyllosphere throughout this paper when referring to the microbial habitat. In addition to the phyllosphere samples, rhizosphere samples were collected separately from the same plants, and bulk soil was included as a reference material. For each of the sample types, five equal specimens were obtained. All samples were stored on ice and immediately processed after arrival at a nearby laboratory. To homogenize the samples, 5 g of plant material or bulk soil per replicate was physically disrupted with a sterile mortar and pestle, re-suspended in 10 mL of 0.85% NaCl, transferred into two 2 mL Eppendorf tubes and subsequently centrifuged (16500*g*, 20 min, 4 °C). The obtained pellet was used to isolate the total-community DNA with a FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, USA) according to the manufacturer's instructions. All DNA extracts were stored at −80 °C until further processing.

Preparation of the 16S rRNA gene fragment amplicon library for high-throughput sequencing

Community DNA extracted from the soil, rhizosphere and phyllosphere habitats of arugula plants was subjected to PCR-based barcoding. The approach entailed a nested PCR, with the archaea-specific primers 344f and 915r [16] in the first PCR and the modified primer pair S-D-Arch-0349-a-S-17/S-D-Arch-0519-a-A-16 (hereafter 349f/519r) [17] containing an additional 10 bp primer pad (TATGGTAATT/AGTCAGCCAG) and linker (GT/GG) in the subsequent PCR, as previously described in protocols of the Earth Microbiome Project [18]. In a third PCR, the Golay barcodes were annealed. The first PCRs (20 μL) comprised 4 μL of GC buffer (7.5 mM), 2 μL of bovine serum albumin (BSA) (10 mg/mL), 2 μL of dNTP mix (2 mM), 0.25 μL of Phusion polymerase (New England Biolabs, Frankfurt, Germany; 2 U/μL), 9.55 μL of PCR-grade water, and 0.6 μL each of forward and reverse primers (10 μM). Amplifications were conducted with the following settings: 95 °C for 2 min, followed by 10 cycles of 96 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min, 15 cycles of 94 °C for 25 s, 60 °C for 30 s, and 72 °C for 1 min, and a final elongation step at 72 °C for 10 min. The nested PCRs with primers 349f and 519r were executed in the same way as the previous PCR but with different settings: 95 °C for 5 min, followed by 25 cycles of 95 °C for 40 s, 66 °C for 2 min, and 72 °C for 1 min and a final elongation step at 72 °C for 10 min. A final barcode-annealing PCR was conducted to attach sample-specific Golay barcodes to the primer pads on each forward and reverse primer, with the following settings: 95 °C for 2 min, followed by 10 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s and a final elongation step at 72 °C for 10 min. Amplified PCR products were checked by gel electrophoresis after each PCR, and 1 μL of PCR product from the previous PCR was used as a template for the subsequent PCR. All PCRs were conducted in triplicate, purified with a Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA), and pooled in equimolar concentrations prior to sequencing. The sequencing was then conducted using an Illumina MiSeq Personal Sequencer (GATC Biotech AG, Konstanz, Germany).

Bioinformatic analyses of archaeal 16S rRNA gene fragments

The generated 16S rRNA gene Illumina library reads were analysed and processed by using the open source software package

Quantitative Insights Into Microbial Ecology (QIIME) release 1.9.1 for pre-processing and pre-filtering, and QIIME2 release 2018.2 [19] was used for further analysis following tutorials provided on the QIIME2 homepage (https://docs.qiime2.org/2018.2/). First, the read quality was checked with FastQC, and barcodes were extracted in QIIME 1.9.1. Then, reads and metadata were imported into QIIME2, in which demultiplexing, denoising of truncated reads, and generation of ribosomal sequence variants (RSVs) were conducted using the DADA2 algorithm [20]. The RSVs were then summarized in a feature table and rarefied to a depth of 1000 RSVs. Feature tables subjected to additional filtering were used to calculate metrics of alpha and beta diversity, including the Shannon diversity index, Faith's phylogenetic diversity, evenness, the Jaccard index and the Bray-Curtis distance, with the QIIME2 core diversity metrics. For phylogenetic analysis, the MAFFT script was used to align representative sequences, and FastTree was used to generate a phylogenetic tree. For taxonomic composition analysis, the taxonomy was assigned to representative sequences by using a customized naïve Bayes classifier trained on 16S rRNA gene OTUs clustered at 97% similarities within the Silva 128 database [21]. In addition, 2D principal coordinate analysis (PCoA) plots were constructed using Emperor weighted and unweighted UniFrac distances. The distribution of taxa among the habitats was visualized with Cytoscape 3.3.0 [22] based on habitat-specific core archaeomes, which were identified by using a frequency threshold of 0.8 (present in 4 out of 5 samples). The most abundant sequences, which were showing a low taxonomic resolution, were further assigned by using nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Archaea-targeting functional metagenomics

Shotgun-sequenced datasets were available from a previous study [8] that utilized arugula plants from the same garden (repository IDs were provided in the respective section). The datasets (phyllosphere, rhizosphere, and bulk soil) were used to explore the plant's bacteriome and the functioning of the enterobacterial subpopulation therein. In the present study, functional and taxonomic analyses were performed on the Metagenomic RAST server (MG-RAST; http://www.mg-rast.org). Quality-filtered reads from HiSeq Illumina runs were uploaded to the server and initially processed with default parameters. The reads were filtered for artificial replicate sequences, low-quality sequences, short sequences, and sequences containing ambiguous bases. The filtered sequences were then annotated using hierarchical classification with default parameters: SEED subsystems as an annotation source, a maximum e-value of 10^{-5}, a minimum identity of 60% and a minimum alignment length of 15 measured in aa for protein and bp for RNA databases. Each subsystem within the metagenomes represents a group of sequences encoding for a specific biological process or a structural complex. Furthermore, the metagenomes were screened for functional signatures annotated to *Archaea* within MG-RAST. The functional hits were subsequently exported and normalized to the lowest number of sequences containing predicted proteins with known function from the soil habitat (8,400,892 sequences). Then, the structure and abundance of the functional signatures were visualized using MEtaGenome ANalyzer (MEGAN) 5 [23]. For taxonomic analysis, the structure of the archaeal community was aligned and annotated with the M5nr database as a reference and exported via the MG-RAST application programming interface (API) server (http://www.mg-rast.org).

In situ visualization of Archaea with confocal laser scanning microscopy

Archaeal colonization of the rhizosphere and phyllosphere samples of *E. sativa* was analysed by fluorescent *in situ* hybridization in combination with confocal laser scanning microscopy (FISH/CLSM). The microscope used for the imaging was a Leica TCS SPE confocal laser scanning microscope (Leica Microsystems, Mannheim, Germany) equipped with Leica ACS APO 40.0 × oil CS and Leica ACS APO 63 × oil CS oil immersion objective lenses. Plant tissues were fixed with 4% paraformaldehyde and 1x phosphate-buffered saline (3:1) for 6 h at 4 °C. For each habitat, three fixed replicates were analysed. The samples were then stained by in-tube FISH according to Cardinale and colleagues [24]. The FISH probes ARCH344-Cy5, ARCH1060-Cy5 [25], and ARCH915-Cy5 [26] and an equimolar mixture of Cy3-labelled EUB338, EUB338-II, and EUB338-III [27,28] were used to visualize colonization patterns of *Archaea* and bacteria, respectively (max. extinction/emission in nm, Cy3: 548/562 and Cy5: 650/670). As a positive control for visualization of *Archaea*, a culture of *Candidatus Altiarchaeum hamiconexum* was used. To visualize the structure of the plant, Calcofluor white staining was conducted (Sigma Aldrich, St. Louis, USA) using a stationary laser at a 405 nm wavelength. Further maximum projections of optical z-stack slices were used to generate micrographs of archaeal and bacterial colonization.

Repository deposition of next-generation sequencing data

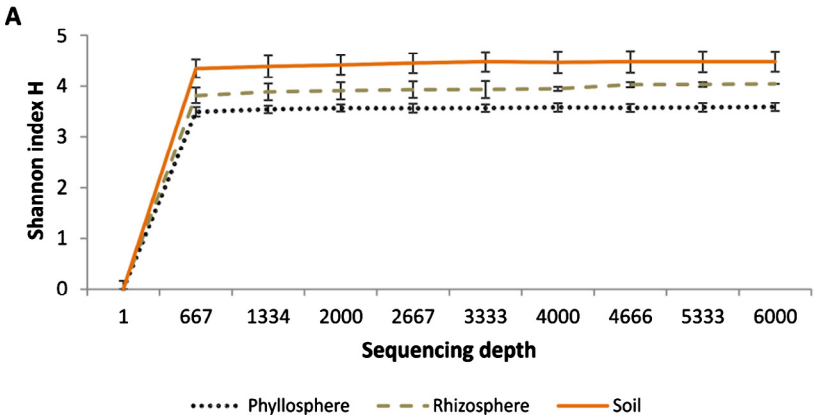
The metagenomes are publicly accessible on the MG-RAST server (http://www.mg-rast.org) under the accession numbers mgm4551355.3 (phyllosphere), mgm4551391.3 (rhizosphere), and mgm4551574.3 (bulk soil). All amplicon sequencing data sets were submitted to the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena) and are accessible under the project accession number PRJEB28404.

Results

Archaeal diversity in arugula plants

Quality-filtering of the 16S rRNA gene fragment dataset of *E. sativa* resulted in 1668 features (RSVs) with a total abundance of 1,040,565. The 16S rRNA amplicon analysis included three habitats: soil (413,055 reads), the rhizosphere (146,847 reads) and the phyllosphere (470,663 reads). Alpha diversity was analysed by phylogenetic and non-phylogenetic based methods; outliers were excluded (n = 4). The highest diversity of *Archaea* was found in the soil habitat (Shannon index (H): 4.44; Faith: 83.06), followed by the rhizosphere (H: 3.95; Faith: 71.63) and the phyllosphere (H: 3.56; Faith: 52.47) (Fig. 1).

Weighted and unweighted PCoA plots with all 15 samples revealed a distinct clustering of archaeal communities belonging to each of the habitats (Fig. 2A and B). In the phyllosphere (77.8%), rhizosphere (73.4%) and soil (74.2%), the archaeal community was dominated by *Thaumarchaeota*. The second most abundant archaeal phylum throughout all habitats was *Euryarchaeota*, totalling 15.1% in the phyllosphere, 20.9% in the rhizosphere, and 15.3% in the soil. *Archaea* assigned to the phyla *Woesearchaeota* (phyllosphere: 0.4%; rhizosphere: 3.0%; soil: 7.1%) and *Bathyarchaeota* (phyllosphere: 0.1%; rhizosphere: 0.2%; soil: 0.1%) were less represented. However, 4.0% of all sequences remained unassigned. Overall, at the genus level, the most abundant feature was assigned to *Candidatus Nitrosocosmicus* (27.8%), as further revealed by BLAST analysis. Other abundant genera were assigned to an uncultured *Thaumarchaeota* strain (14.2%) and to *Methanosarcina thermophila* (8.8%). The most abundant taxa were shared among all three habitats, representing the core archaeome of arugula (Fig. 3). Taxa shared exclusively between two habitats were found only between the soil and the rhizosphere. Unique taxa, i.e., those found exclusively in one habitat, were detected in the



B

Habitat	Shannon index H
Phyllosphere	3.56 ± 0.03
Rhizosphere	3.95 ± 0.07
Soil	4.44 ± 0.05

Fig. 1. Visualization of Shannon index H, as function of sequencing depth of habitats of *E. sativa* (A). The applied method is alpha rarefaction with 10 repeats at 10 different sequencing depths. Displayed habitats are soil (solid line), rhizosphere (dashed line) and phyllosphere (dotted line). Shannon index H values are displayed with their corresponding standard-deviation (B).

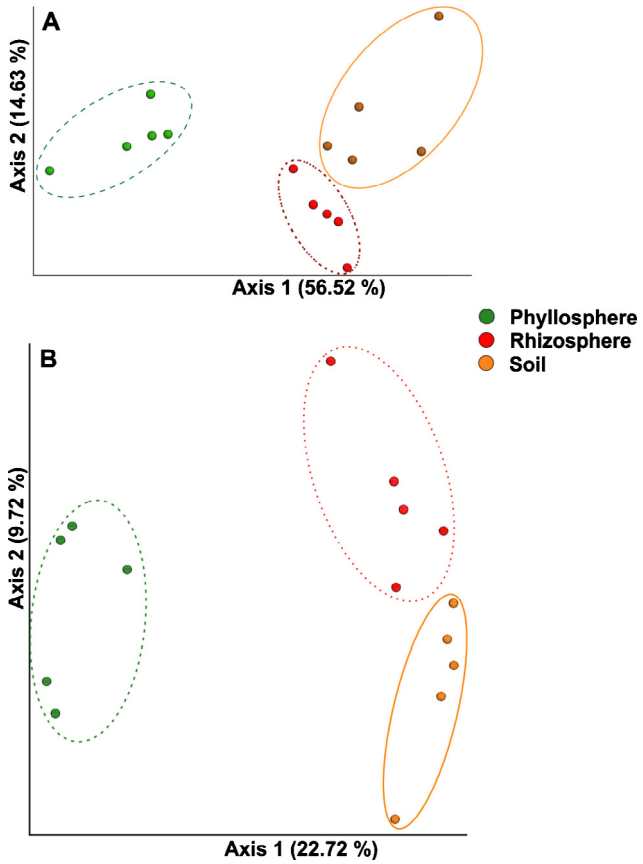


Fig. 2. Comparison of archaeal communities from the soil, rhizosphere and phyllosphere of *Eruca sativa* by principal coordinate analysis (PCoA). Plots were calculated using Emperor weighted UniFrac distances (A) and unweighted UniFrac distances (B). Each dot represents a distinct sample of a habitat: the phyllosphere in green (dashed circle), the rhizosphere in red (dotted circle) and the soil in orange (solid circle). The variation explained by each principal coordinate (PC) is defined on the plot.

soil and (to a lesser extent) in the rhizosphere. All taxa found in the phyllosphere were also present in the other habitats.

The three metagenomes revealed a different taxonomic structure than the 16S rRNA amplicon dataset (Fig. 4). The relative archaeal abundance was highest in the soil, at 0.7% (48,603 sequences) of all prokaryotic sequences (7,396,616 sequences). The relative abundance of *Archaea* was slightly lower in the rhizosphere, at 0.5% (45,140 sequences) of all prokaryotic sequences (9,822,615 sequences). The lowest archaeal abundance was found in the phyllosphere, accounting for 0.1% (5949 sequences) of the total prokaryotic community (8,531,239 sequences). At the phylum level, the distribution of the archaeal community was similar in all habitats, whereas the dominant archaeal group was *Euryarchaeota*, accounting for 67.1–74.5% of all archaea. In contrast to the results from the 16S rRNA gene fragment dataset, *Thaumarchaeota* accounted for only 10.5–15.6% of archaea, followed by the phylum *Crenarchaeota*, which accounted for 13.0–14.8%. At the class level, *Methanomicrobia* and *Halobacteria* were the most abundant taxa, representing 27.0–31.8% and 15.6–18.5% of archaea, respectively, followed by *Thermoprotei* (11.6–12.6%) and unclassified *Thaumarchaeota* (10.5–15.6%). *Thermococci*, *Methanococci*, *Methanobacteria*, *Archaeoglobi* and *Thermoplasmata* were less represented. Similar to the results for the 16S rRNA gene fragment dataset, the relative abundance of unclassified reads was high, ranging from 7.4–8.8% of archaea.

Visualization of archaeal communities of arugula

Archaeal colonization patterns in the phyllosphere and the rhizosphere of *E. sativa* plants was visualized using a FISH/CLSM approach (Fig. 5). In the analysed phyllosphere samples, small archaeal colonies were spatially distant from each other, mainly forming colonies in close proximity to plant stomata (Fig. 5A). The colonies were clearly separated from each other and mostly consisted of fewer than 100 individual cells. In contrast, larger colonies were found in the rhizosphere (Fig. 5B). These colonies were also found to be embedded within large bacterial biofilms. Archaeal colonies were mainly found on lignified plant parts and especially on rotten roots. Bacteria that were labelled with a different fluorescent dye were visualized with the same approach. These bacteria often co-localized with archaeal colonies in the rhizosphere but not in the phyllosphere and were predominant in the plant samples.

Metagenome-based functional analysis of Archaea associated with arugula

From the three normalized metagenomes, functional analysis resulted in 5804 archaeal sequences. These sequences were assigned to certain functional subsystems of the SEED database (Table 1). Most of the sequences were assigned to primary metabolic functions of *Archaea*, such as carbohydrates (4161 hits; 71.7%), including central carbohydrate metabolism (1706 hits;

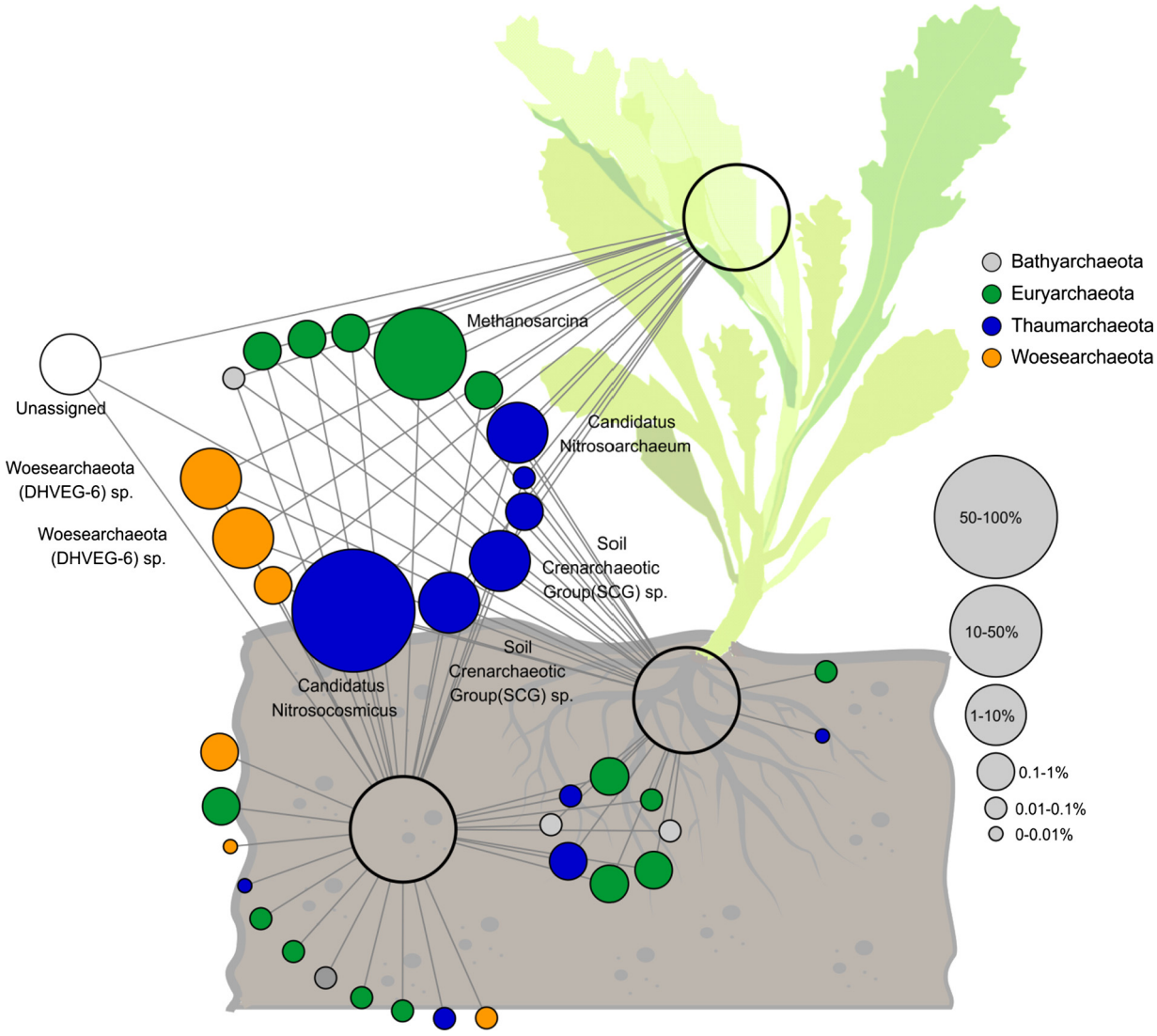


Fig. 3. Feature network of the plant's archaeal communities at the genus level, based on 16S rRNA gene fragment datasets. The datasets were obtained from the soil, rhizosphere and phyllosphere habitats of *Eruca sativa*. For each habitat, a core archaeome was identified with a frequency threshold of 0.8 (4 out of 5 samples). Archaeal phyla are indicated by coloured bubbles: *Bathyarchaeota* in grey; *Euryarchaeota* in green; *Thaumarchaeota* in blue; and *Woesearchaeota* in orange. The size of the bubble represents the relative abundance of the archaeal taxa throughout all habitats.

29.4%) and polysaccharides (2115 hits; 36.4%), and cofactors, vitamins, prosthetic groups and pigments (621 hits; 10.7%). Functions involved in one-carbon metabolism (657 hits; 11.3%) and fermentation (21 hits; 0.4%) were also found. Furthermore, archaeal functions were assigned to subsystems involved in nutrient cycling, such as functional signatures for CO_2 fixation (400 hits; 6.9%), whereas signatures for nitrogen fixation were not detected. Functions assigned to cofactors, vitamins, prosthetic groups and pigments were mainly involved in the pyrimidine deaminase pathway (535 hits; 9.2%). A high proportion of archaeal functions were also assigned to subsystems involved in glycogen degradation (1022 hits; 17.6%) and DNA metabolism, especially DNA replication (472 hits; 8.1%). In contrast, functions involved in RNA metabolism were less represented (86 hits; 1.5%). In addition, functional signatures involved in response to stress, especially oxidative stress (17 hits; 0.3%), and signatures involved in protein degradation (215 hits; 3.7%) were also found.

Furthermore, the habitat specificity of archaeal functions in *E. sativa* was analysed. To do so, the functional distributions of the normalized metagenomes were compared among the habitats. In general, most assigned functions belonged to the soil habitat (48.8%), followed by the rhizosphere (36.2%) and the phyllosphere (15.0%). Functional signatures involved in glycogen degradation, and amino acids and derivatives were found at a higher relative abundance in the soil than in the other habitats, with representations of 40.6%, and 4.0%, respectively (Fig. 6). Additional functions involved in CO_2 fixation and DNA replication were similarly distributed in the soil and the rhizosphere with a representation of 7.4% and 8.8% in the soil and 7.7% and 9.0% in the rhizosphere, respectively, whereas the relative abundance of these functions in the phyllosphere was below 3.7%. Functions involved in fermentation were not found in the phyllosphere. Further, functions involved in stress response and oxidative stress were represented in low relative abundance in the phyllosphere (0.2%), compared to the soil (0.4%) and the rhizosphere (0.2%). The only exceptions included signatures assigned to one-carbon metabolism, more precisely the serine-glyoxylate cycle, the TCA cycle, and the biosynthesis of riboflavin, flavin mononucleotide (FMN) and flavin

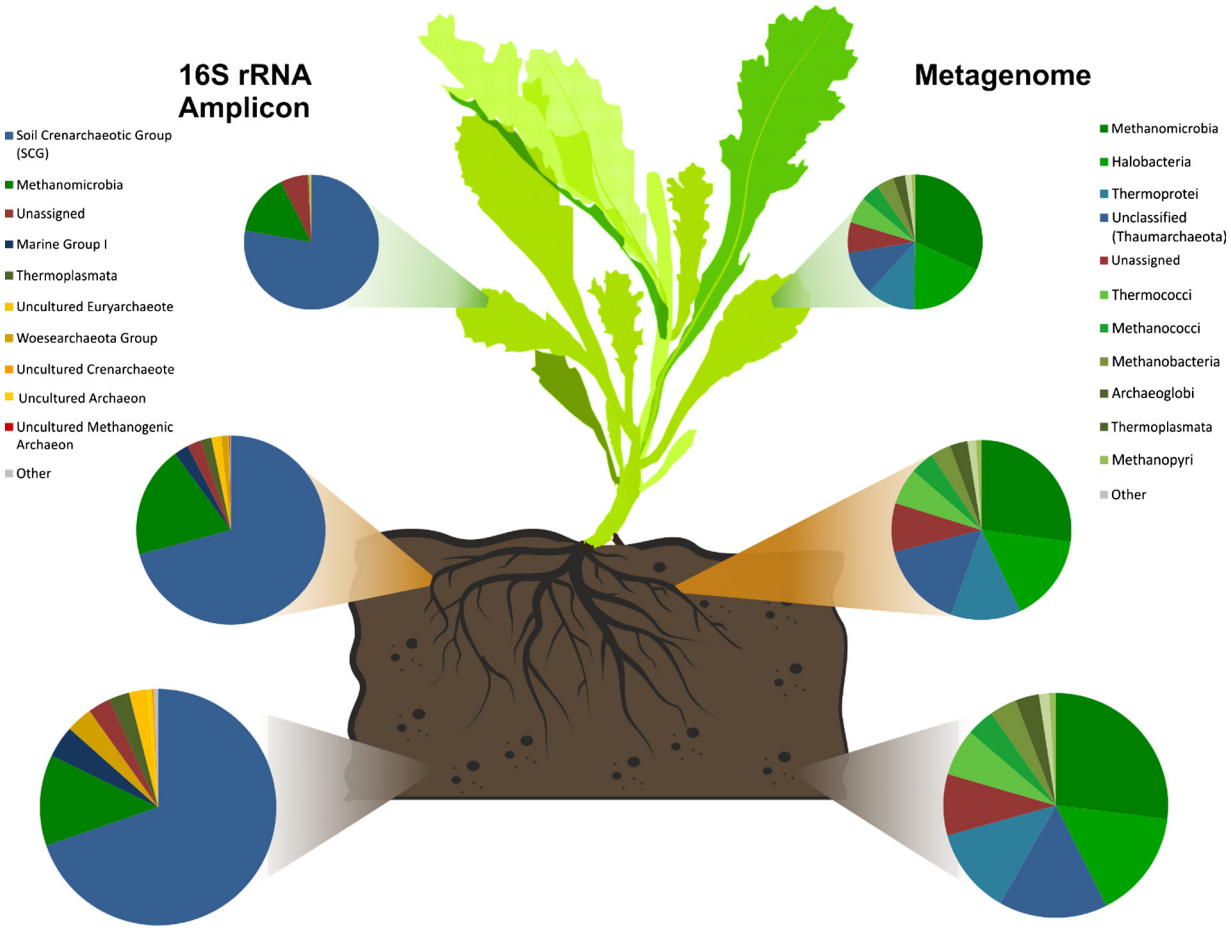


Fig. 4. Taxonomic composition of archaeal communities of *Eruca sativa* revealed by 16S rRNA amplicon and shotgun sequencing-based metagenomics analysis. The archaeal community is described at the class level for each habitat: soil, rhizosphere and phyllosphere. The abundances of archaeal genera are displayed relative to all sequences assigned to *Archaea* in the metagenomics dataset (soil: 48,603 sequences; rhizosphere: 45,140 sequences; phyllosphere: 5949 sequences) as well as relative to all sequences assigned to the 16S rRNA gene fragment dataset (soil: 82,611 sequences; rhizosphere: 31,369 sequences; phyllosphere: 94,133 sequences).

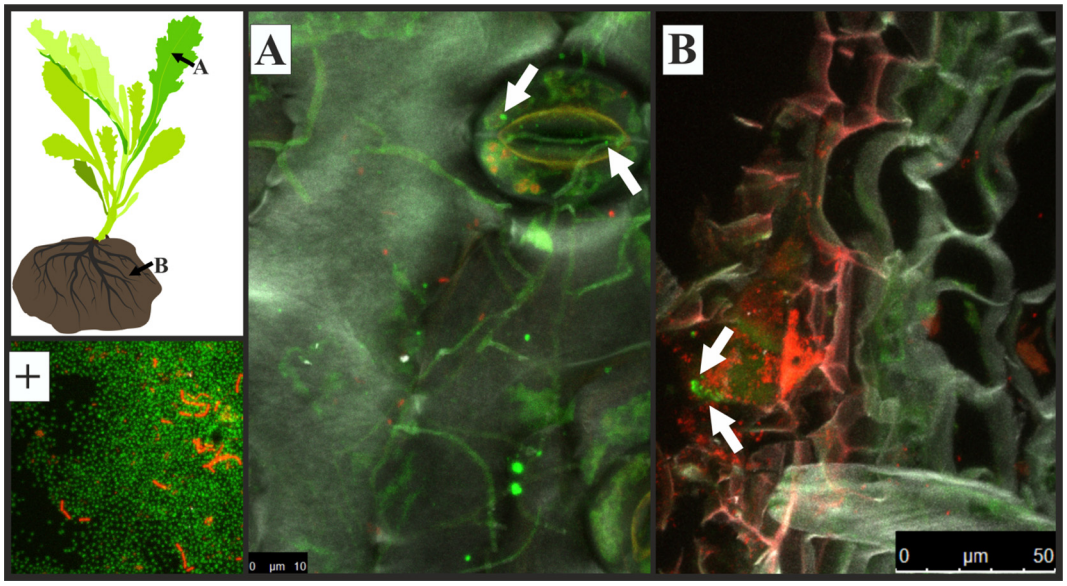


Fig. 5. FISH/CLSM visualization of archaeal colonization patterns in the phyllosphere (A) and rhizosphere (B) of *Eruca sativa*. *Archaea* were stained with the fluorochrome Cy5 and are shown in green and highlighted with white arrows. For better contrast, bacteria were stained with the fluorochrome Cy3 and are shown in red. To visualize the structure of the plant, Calcofluor white staining was conducted. As a positive control for visualization of *Archaea*, a culture of *Candidatus Altiarchaeon hamiconexum* was used (+).

Table 1
List of functional signatures of *Archaea* associated with *E. sativa*. Functional signatures were obtained from three metagenomes of the habitats soil, rhizosphere and phyllosphere of *E. sativa*, annotated using functional subsystems of SEED database, processed with MG-Rast. Total abundances of each signature are separately shown for each habitat.

SEED Level				Habitat		
SEED L1	SEED L2	SEED L3	SEED L4	Soil	Rhizosphere	Phyllosphere
Total archaeal functional hits				**2831**	**2102**	**871**
Carbohydrates	Central			**833**	**626**	**247**
	carbohydrate metabolism	Pyruvate metabolism I: anaplerotic reactions, PEP	Phosphoenolpyruvate carboxylase, archaeal (EC 4.1.1.31)	14	7	0
		Glycolate, glyoxylate interconversions	Phosphoglycolate phosphatase, archaeal type (EC 3.1.3.18)	19	9	1
		TCA Cycle	Archaeal succinyl-CoA ligase [ADP-forming] alpha chain (EC 6.2.1.5)	10	10	2
		TCA Cycle	Archaeal succinyl-CoA ligase [ADP-forming] beta chain (EC 6.2.1.5)	27	17	1
		TCA Cycle	Putative malate dehydrogenase (EC 1.1.1.37), similar to archaeal MJ1425	112	116	89
		Glycolysis and Gluconeogenesis,	2,3-bisphosphoglycerate-independent phosphoglycerate mutase, archaeal type (EC 5.4.2.1)	410	276	120
		Glycolysis and Gluconeogenesis,	Fructose-1,6-bisphosphatase, type V, archaeal (EC 3.1.3.11)	98	87	16
		Glycolysis and Gluconeogenesis,	Fructose-bisphosphate aldolase, archaeal class I (EC 4.1.2.13)	20	9	2
		Glycolysis and Gluconeogenesis,	Glucose-6-phosphate isomerase, archaeal (EC 5.3.1.9)	11	20	4
		Glycolysis and Gluconeogenesis,	NAD(P)-dependent glyceraldehyde 3-phosphate dehydrogenase archaeal (EC 1.2.1.59)	112	75	12
		Entner-Doudoroff Pathway	2,3-bisphosphoglycerate-independent phosphoglycerate mutase, archaeal type (EC 5.4.2.1)	410	276	120
		Glycolysis and Gluconeogenesis	Fructose-bisphosphate aldolase, archaeal class I (EC 4.1.2.13)	20	9	2
	One-carbon			**169**	**243**	**245**
	Metabolism	Serine-glyoxylate cycle	Putative malate dehydrogenase (EC 1.1.1.37), similar to archaeal MJ1425	112	116	89
		Serine-glyoxylate cycle	Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	57	127	156
	Polysaccharides			**1134**	**752**	**229**
		Glycogen metabolism	Glycogen branching enzyme, GH-57-type, archaeal (EC 2.4.1.18)	541	400	152
		Glycogen metabolism	Putative glycogen debranching enzyme, archaeal type, TIGR01561	593	352	77
	CO2 fixation			**210**	**162**	**28**
		Calvin-Benson cycle	Fructose-1,6-bisphosphatase, type V, archaeal (EC 3.1.3.11)	98	87	16
		Calvin-Benson cycle	NAD(P)-dependent glyceraldehyde 3-phosphate dehydrogenase archaeal (EC 1.2.1.59)	112	75	12
	Fermentation			**14**	**7**	**0**
		Fermentations: Mixed acid	Phosphoenolpyruvate carboxylase, archaeal (EC 4.1.1.31)	14	7	0
Stress Response	Oxidative stress			**10**	**5**	**2**
		Glutathione: Biosynthesis and gamma-glutamyl cycle	Glutamate–cysteine ligase archaeal (EC 6.3.2.2)	10	5	2
Protein Metabolism	Protein			**110**	**87**	**18**
	degradation	Proteasome archaeal	Proteasome subunit alpha (EC 3.4.25.1), archaeal	30	27	6
		Proteasome archaeal	Proteasome subunit beta (EC 3.4.25.1), archaeal	63	50	6
		Proteasome archaeal	Proteasome-activating AAA-ATPase (PAN), archaeal	17	10	6
RNA Metabolism				**46**	**36**	**4**
	RNA processing and modification	tRNA nucleotidyltransferase	tRNA nucleotidyltransferase, archaeal type (EC 2.7.7.21) (EC 2.7.7.25)	24	23	2
	Transcription	RNA polymerase archaeal initiation factors	Archaeal transcription factor E	22	13	2
DNA Metabolism				**250**	**190**	**32**
	DNA replication	DNA replication, archaeal	Archaeal DNA polymerase I (EC 2.7.7.7)	106	72	11
		DNA replication, archaeal	Archaeal DNA polymerase II large subunit (EC 2.7.7.7)	113	94	16
		DNA replication, archaeal	Archaeal DNA polymerase II small subunit (EC 2.7.7.7)	31	24	5

(*continued on next page*)

Table 1 (*continued*)

SEED Level				Habitat		
SEED L1	SEED L2	SEED L3	SEED L4	Soil	Rhizosphere	Phyllosphere
Cofactors, Vitamins, Prosthetic Groups, Pigments				**258**	**200**	**163**
	Coenzyme A	Coenzyme A Biosynthesis	Dephospho-CoA kinase archaeal, predicted (EC 2.7.1.24)	10	4	2
		Coenzyme A Biosynthesis	Pantoate kinase, archaeal (EC 2.7.1.-)	6	2	1
		Coenzyme A Biosynthesis	Phosphopantothenate synthetase, archaeal	34	21	4
	Riboflavin, FMN, FAD	Riboflavin, FMN and FAD metabolism	CTP-dependent archaeal riboflavin kinase	1	1	0
		Riboflavin, FMN and FAD metabolism	Pyrimidine deaminase archaeal predicted (EC 3.5.4.26)	207	172	156
Miscellaneous	Miscellaneous	Peptidyl-tRNA hydrolase, archaeal type (EC 3.1.1.29)	Peptidyl-tRNA hydrolase, archaeal type (EC 3.1.1.29)	20	14	12
Amino Acids and Derivatives	Methionine	Methionine Biosynthesis	Archaeal S-adenosylmethionine synthetase (EC 2.5.1.6)	113	65	8

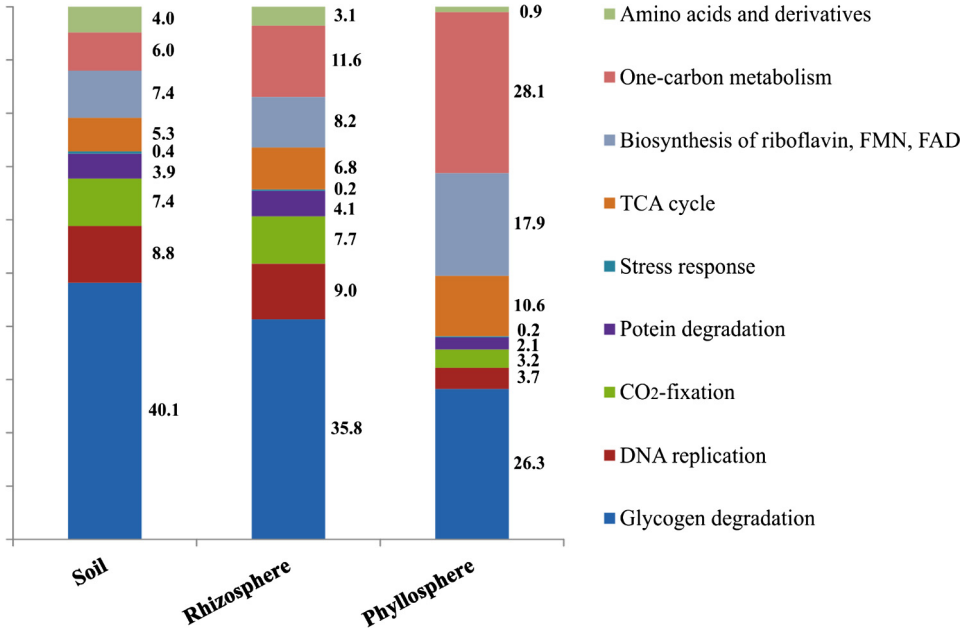


Fig. 6. Comparison of the relative distributions of specific archaeal functions in the soil, rhizosphere and phyllosphere of *Eruca sativa* based on metagenomics datasets. Abundances of the functional signatures are shown as proportion of all functions assigned to *Archaea* in the metagenomics dataset of the corresponding habitat (soil: 2831 total hits; rhizosphere: 2102 total hits; phyllosphere: total 871 hits). The values next to distinct segments indicate their respective percentages in the archaeal fraction.

adenine dinucleotide (FAD), which were more represented in the phyllosphere (28.1%, 10.6%, and 17.9%) than in the soil (6.0%, 5.3%, and 7.4%) and rhizosphere (11.6%, 6.8%, and 8.2%).

Discussion

In the present study, we found various indicators for *Archaea* to be important components of the microbiomes of plants domesticated by humans. In *E. sativa*, *Archaea* exhibited a habitat-specific structure, colonization patterns and functions. Similar to the ways in which biotic and abiotic factors shape bacterial and fungal communities, the archaeome is likely also affected by its environment. Each plant habitat is affected by different environmental conditions, such as low nutrient availability and exposure to environmental changes in the phyllosphere; the availability of root exudates in the plant rhizosphere; and the more stable conditions in the soil. These conditions might be among the factors influencing the habitat-specific diversity of *Archaea*. In the present study, the highest archaeal diversity was found in soil samples, and the lowest diversity was found in the phyllosphere. Furthermore, the composition of the archaeal community at the phylum and class levels was similar among the habitats, which was also observed in the metagenomics analysis. However, the predominant lineages in the metagenomics dataset, namely, *Euryarchaeota* and *Thaumarchaeota*, were inverted in the 16S rRNA gene fragment dataset. Overall, the metagenomics shotgun-sequencing approach revealed a more diverse taxonomy than the 16S rRNA gene fragment amplicons. This bias was described previously and can occur due to differences in database entries and errors during PCR amplification and amplicon sequencing [29]. In general metagenomic sequencing reveals a higher richness than the 16S rRNA approach, whereat the 16S rRNA approach additionaly misses 10% of yet uncharacterized *Archaea*, showing the limitations of the accurate identification of microbes within a microbiome [30,31].

An in-depth analysis of the 16S rRNA gene fragment dataset with a feature network highlighted the habitat-specific colonization of plants by *Archaea*. Soil samples exhibited the greatest number of habitat-specific features. However, the rhizosphere also harboured unique features, whereas the phyllosphere had no unique features. Overall, a large core archaeome was shared among

the habitats, with the most abundant taxa assigned to *Candidatus Nitrososcosmicus*, a member of the ammonium-oxidizing *Archaea* (AOA) that plays an important role in nitrification processes and is expected to possess key genes associated with protection from abiotic stress [32]. *Archaea* assigned to *Methanosarcina*, *Candidatus Nitrosoarchaeum* and *Woesearchaeota* were also abundant in the core archaeome. These lineages were previously found in animal digestive tracts, sediments, and the human gut, respectively [33–35]. Methanogens are strict anaerobes; therefore, the detection of *Methanosarcina* in the phyllosphere might be explained by anaerobic niches in the phyllosphere or, more likely, by contact between the plant and animals, as arugula was grown under field conditions [36]. The visualization of archaeal colonization of *E. sativa* revealed a habitat-specific distribution of the overall population. In the phyllosphere, we found small scattered colonies, whereas in the rhizosphere, *Archaea* formed larger colonies and colonies in close proximity to or even within bacterial biofilms without any obvious zone of inhibition. No negative interactions between archaea and bacteria have been observed to date, suggesting mostly synergistic relationships between the two groups [37]. Moreover, *Archaea* were found to accumulate in nutrient-rich hotspots such as rotten roots, indicating that they play a direct or indirect role in decomposition processes. In Finnish forests, *Archaea* and *Thaumarchaeota* were previously found to be active components of the decaying wood microbiota [38].

The arugula archaeome harboured specific archaeal functions mainly assigned to central carbohydrate metabolism and polysaccharides. We also found functional signatures involved in nutrient cycling such as CO_2 fixation but no signatures involved in nitrogen fixation, although *E. sativa* was mainly colonized by AOA. These findings are in accordance with our previous study on *Archaea* associated with bog vegetation [12]. Arugula has low nutrient requirements, therefore we hypothesize that archaea are not selected by arugula in order to complement the nitrogen balance of the holobiont. In the current study, we also detected functions involved in glycogen degradation, even at higher relative abundances. Glycogen is used by fungi as a main storage unit and is also excreted by them as part of common exudates. This relationship indicates potential interactions with fungi, as fungal exudation rates and fungal colonization were previously shown to be correlated with archaeal abundance [39]. Functions involved in response to stress, especially oxidative stress, were less represented, which might be due to the specific micro-environments of arugula examined. The highest abundances of functional hits for *Archaea* were found in the soil and rhizosphere, whereas the phyllosphere was relatively low in sequences corresponding to archaeal functions. This discrepancy indicates that soil and the rhizosphere are the preferred habitats of *Archaea* and the habitats with the highest archaeal metabolic activity. Functions that were relatively more abundant in the phyllosphere were involved in the serine-glyoxylate cycle and assigned to "serine-pyruvate aminotransferase/archaeal aspartate aminotransferase" (EC 2.6.1.51) and "putative malate dehydrogenase" (EC 1.1.1.37). "Serine-pyruvate aminotransferase" is involved in the glyoxylate cycle, which enables the utilization of simple carbon sources when complex and energetically more valuable carbon sources (*e.g.*, glucose) are absent [40], as is the case in the phyllosphere.

Conclusions

Archaea might show less functional adaptation to agricultural plants such as *E. sativa* than to their wild ancestors due to differences in genotype and the environment. These differences include altered nutrient and energy levels in the soil caused by introducing fertilizers and the accompanying phenotypic changes of plants. Since *Archaea* are adapted to energy deficiency, stress and energy limitations, they might lose their advantage over bacteria in terms of environmental tolerance and subsequently be outcompeted by bacteria, which focus on exploiting energy-rich resources. In summary, *Archaea* are small but potentially important niche-specific components of plant microbiomes, and therefore, we must advance our understanding of plant-associated *Archaea* before they disappear [12] due to our agricultural practices.

Conflict of interest

The authors declare no conflicts of interest.

Compliance with Ethics Requirement

This article does not contain any studies with human or animal subjects.

Acknowledgements

This project was funded by the European Funds for Regional Development (EFRE) and co-supported by the regional government of Styria (Das Land Steiermark, Austria), project code A3-11.P-33/2011-6.

References

[1] Piślewska-Bednarek M, Nakano RT, Hiruma K, Pastorczyk M, Sanchez-Vallet A, Singkaravanit-Ogawa S, et al. Glutathione Transferase U13 functions in pathogen-triggered Glucosinolate metabolism. Plant Physiol 2018;176:538–51.
[2] Halkier BA, Gershenzon J. Biology and bioechmistry of glucosinolates. Annu Rev Plant Biol 2006;57:303–33.
[3] Zeven A, Wet J De. Dictionary of cultivated plants and their regions of diversity: excluding most ornamentals, forest trees and lower plants. Centre for Agricultural Publishing and Documentation; 1982.
[4] Garg G, Sharma V. *Eruca sativa* (L.): botanical description, crop improvement, and medicinal properties. J Herbs Spices Med Plants 2014;20:171–82.
[5] Nygård K, Lassen J, Vold L, Andersson Y, Fisher I, Löfdahl S, et al. Outbreak of *Salmonella* Thompson infections linked to imported rucola lettuce. Foodborne Pathog Dis 2008;5:165–73.
[6] Erlacher A, Cardinale M, Grosch R, Grube M, Berg G. The impact of the pathogen *Rhizoctonia solani* and its beneficial counterpart *Bacillus amyloliquefaciens* on the indigenous lettuce microbiome. Front Microbiol 2014;5:175.
[7] Wassermann B, Rybakova D, Müller C, Berg G. Harnessing the microbiomes of *Brassica* vegetables for health issues. Sci Rep 2017;7:17649.
[8] Cernava T, Erlacher A, Soh J, Sensen CW, Grube M, Berg G. Enterobacteriaceae dominate the core microbiome and contribute to the resistome of arugula (*Eruca sativa* Mill.). Microbiome 2019;7:13.
[9] Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol 2007;73:7059–66.
[10] Probst AJ, Auerbach AK, Moissl-Eichinger C. Archaea on human skin. PLoS ONE 2013;8:e65388.
[11] Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA. Archaea are interactive components of complex microbiomes. Trends Microbiol 2018;26:70–85.
[12] Taffner J, Erlacher A, Bragina A, Berg C, Moissl-Eichinger C, Berg G. What is the role of Archaea in plants? New insights from the vegetation of alpine bogs. MSphere 2018;3:e00122–e218.
[13] Pump J, Pratscher J, Conrad R. Colonization of rice roots with methanogenic archaea controls photosynthesis-derived methane emission. Environ Microbiol 2015;17:2254–60.
[14] Müller H, Berg C, Landa BB, Auerbach A, Moissl-Eichinger C, Berg G. Plant genotype-specific archaeal and bacterial endophytes but similar *Bacillus* antagonists colonize mediterranean olive trees. Front Microbiol 2015;6:1–9.
[15] Bengtson P, Sterngren AE, Rousk J. Archaeal abundance across a pH gradient in an arable soil and its relationship to bacterial and fungal growth rates. Appl Environ Microbiol 2012;78:5906–11.
[16] Casamayor EO, Massana R, Benlloch S, Ovreas L, Diez B, Goddard VJ, et al. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ Microbiol 2002;4:338–48.
[17] Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-

generation sequencing-based diversity studies. Nucleic Acids Res 2013;41. e1 e1.
[18] Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved bacterial 16S rRNA Gene (V4 and V4–5) and fungal Internal Transcribed Spacer marker gene primers for microbial community surveys. MSystems 2016;1:e00009–e15.
[19] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6.
[20] Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581–3.
[21] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:590–6.
[22] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[23] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res 2007;17:377–86.
[24] Cardinale M, Grube M, Erlacher A, Quehenberger J, Berg G. Bacterial networks and co-occurrence relationships in the lettuce root microbiota. Environ Microbiol 2015;17:239–52.
[25] Moissl C, Rudolph C, Rachel R, Koch M, Huber R. In situ growth of the novel SM1 euryarchaeon from a string-of-pearls-like microbial community in its cold biotope, its physical separation and insights into its structure and physiology. Arch Microbiol 2003;180:211–7.
[26] Stahl AD. Development and application of nucleic acid probes. Nucleic Acid Tech Bact Syst 1991:205–48.
[27] Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol 1990;56:1919–25.
[28] Daims H, Brühl A, Amann R, Schleifer K-H, Wagner M. The Domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. Syst Appl Microbiol 1999;22:434–44.
[29] Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. Strengths and limitations of 16S rRNA Gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 2014;9:e93827.
[30] Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol 2016;1:15032.
[31] Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun 2016;469:967–77.
[32] Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of *Candidatus Nitrosocosmicus exaquare*, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. ISME J 2017;11:1142–57.
[33] Miller TL, Wolin MJ. Methanogens in human and animal intestinal Tracts. Syst Appl Microbiol 1986;7:223–9.
[34] Mosier AC, Allen EE, Kim M, Ferriera S, Francis CA. Genome sequence of "*Candidatus Nitrosoarchaeum limnia*" bg20, a low-salinity ammonia-oxidizing archaeon from the San Francisco ay estuary. J Bacteriol 2012;194:2119–20.
[35] Koskinen K, Pausan MR, Perras AK, Beck M, Bang C, Mora M, et al. First insights into the diverse human Archaeome: Specific detection of Archaea in the gastrointestinal tract, lung, and nose and on skin. MBio 2017. doi: https://doi.org/10.1128/mbio.00824-17.
[36] Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. NatureCom 2014;5:5497.
[37] Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C. Microbial syntrophy: interaction for the common good. FEMS Microbiol Rev 2013;37:384–406.
[38] Rinta-Kanto JM, Sinkko H, Rajala T, Al-Soud WA, Sørensen SJ, Tamminen MV, et al. Natural decay process affects the abundance and community structure of Bacteria and Archaea in *Picea abies* logs. FEMS Microbiol Ecol 2016;92:fiw087.
[39] Karlsson AE, Johansson T, Bengtson P. Archaeal abundance in relation to root and fungal exudation rates. Microbiol Ecol 2012;80:305–11.
[40] Lorenz MC, Fink GR. Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryot Cell 2002;1:657–62.