

2016

# Metagenomic Signatures of Bacterial Adaptation to Life in the Phyllosphere of a Salt-Secreting Desert Tree

Omri M. Finkel

Tom O. Delmont

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.uri.edu/gsofacpubs>

Terms of Use

All rights reserved under copyright.

---

## Citation/Publisher Attribution

Finkel OM, Delmont TO, Post AF, Belkin S. 2016. Metagenomic signatures of bacterial adaptation to life in the phyllosphere of a salt-secreting desert tree. *Appl Environ Microbiol* 82:2854–2861. doi:10.1128/AEM.00483-16.

Available at: <http://dx.doi.org/10.1128/AEM.00483-16>

This Article is brought to you for free and open access by the Graduate School of Oceanography at DigitalCommons@URI. It has been accepted for inclusion in Graduate School of Oceanography Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).

---

**Authors**

Omri M. Finkel, Tom O. Delmont, Anton F. Post, and Shimshon Belkin

# Metagenomic Signatures of Bacterial Adaptation to Life in the Phyllosphere of a Salt-Secreting Desert Tree

Omri M. Finkel,<sup>a\*</sup> Tom O. Delmont,<sup>b\*</sup> Anton F. Post,<sup>b\*</sup> Shimshon Belkin<sup>a</sup>

Institute of Life Sciences, Hebrew University of Jerusalem, Jerusalem, Israel<sup>a</sup>; Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biology Laboratory, Woods Hole, Massachusetts, USA<sup>b</sup>

## ABSTRACT

The leaves of *Tamarix aphylla*, a globally distributed, salt-secreting desert tree, are dotted with alkaline droplets of high salinity. To successfully inhabit these organic carbon-rich droplets, bacteria need to be adapted to multiple stress factors, including high salinity, high alkalinity, high UV radiation, and periodic desiccation. To identify genes that are important for survival in this harsh habitat, microbial community DNA was extracted from the leaf surfaces of 10 *Tamarix aphylla* trees along a 350-km longitudinal gradient. Shotgun metagenomic sequencing, contig assembly, and binning yielded 17 genome bins, six of which were >80% complete. These genomic bins, representing three phyla (*Proteobacteria*, *Bacteroidetes*, and *Firmicutes*), were closely related to halophilic and alkaliphilic taxa isolated from aquatic and soil environments. Comparison of these genomic bins to the genomes of their closest relatives revealed functional traits characteristic of bacterial populations inhabiting the *Tamarix* phyllosphere, independent of their taxonomic affiliation. These functions, most notably light-sensing genes, are postulated to represent important adaptations toward colonization of this habitat.

## IMPORTANCE

Plant leaves are an extensive and diverse microbial habitat, forming the main interface between solar energy and the terrestrial biosphere. There are hundreds of thousands of plant species in the world, exhibiting a wide range of morphologies, leaf surface chemistries, and ecological ranges. In order to understand the core adaptations of microorganisms to this habitat, it is important to diversify the type of leaves that are studied. This study provides an analysis of the genomic content of the most abundant bacterial inhabitants of the globally distributed, salt-secreting desert tree *Tamarix aphylla*. Draft genomes of these bacteria were assembled, using the culture-independent technique of assembly and binning of metagenomic data. Analysis of the genomes reveals traits that are important for survival in this habitat, most notably, light-sensing and light utilization genes.

Plant leaves provide an extensive and diverse habitat for colonization by microorganisms. The global number of plant species is estimated at ~300,000 (1), and they encompass a wide range of morphologies, leaf surface chemistries, and ecological ranges. Sympatric plant species may harbor different microbial communities (2, 3), and microbial communities in the phyllosphere of conspecific plants differ across habitat (4) and time (5). Within a single plant, the microbial communities can differ between leaves (6) or even between different parts of the same leaf (7). In fact, the core traits of leaves that define microbial habitats are surprisingly few. Leaf surfaces are typically waxy surfaces dotted with stomata and are by definition exposed to solar radiation and to the open air (with the exception of submerged plants). All other characteristics, such as exudate composition, water availability, salinity, pH, or climatic conditions, are plant and/or habitat specific. As such, diversity in habitat characteristics often dictates a corresponding diversity in microbial community composition and a fittingly diminished common core of functional traits. A growing number of studies report deep sequencing of the ribosomal small-subunit (SSU) or 16S rRNA genes of such phyllosphere communities, revealing the relationship of microbial diversity to plant species, environment, climate, and geography (8–11). Unlike most phyllosphere studies, which are often skewed toward temperate climates and food crops, our investigations (4, 12–15) focus on the microbiome of desert trees. More specifically, we study the phyllosphere community on *Tamarix* (salt cedar), a salt-secreting tree adapted to a wide range of water

availabilities (16). *Tamarix* trees are widely distributed. They are found in Europe, Asia, and Africa and, since the late 19th century, have been reported as invasive species in America and Australia. *Tamarix* leaves are 1 to 2 mm long and grow as rosettes tightly wrapped around the pigmented stalk. *Tamarix aphylla* leaves are spartoid, rendering the stalk the main photosynthetic organ. The hygroscopic salt crystals that form on the *Tamarix* leaf and stalk surface adsorb moisture during damp nights and deliquesce into highly saline and alkaline dew droplets. Chemical analyses of droplets collected from *Tamarix aphylla* leaves, the focus of the current study, reveal a harsh environment. NaCl concentrations

Received 16 February 2016 Accepted 29 February 2016

Accepted manuscript posted online 4 March 2016

Citation Finkel OM, Delmont TO, Post AF, Belkin S. 2016. Metagenomic signatures of bacterial adaptation to life in the phyllosphere of a salt-secreting desert tree. *Appl Environ Microbiol* 82:2854–2861. doi:10.1128/AEM.00483-16.

Editor: P. D. Schloss, University of Michigan

Address correspondence to Omri M. Finkel, omri.finkel@mail.huji.ac.il.

\* Present address: Omri M. Finkel, Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA; Tom O. Delmont, Department of Medicine, University of Chicago, Chicago, Illinois, USA; Anton F. Post, Coastal Resource Center, University of Rhode Island, Narragansett, Rhode Island, USA.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.00483-16>.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

are high ( $>100$  g/liter), sulfate concentrations exceed 23 mg/liter (12), and the pH ranges between 8.5 and 10.5 (17). Organic substrates, however, are plentiful; dissolved organic carbon concentrations have been reported (12) to be higher than 3 g/liter and ca. 1 to 3 mg/g leaf.

We previously reported the existence of diverse microbial communities on *Tamarix aphylla* leaves from different habitats, including the Mediterranean and Dead Sea regions in Israel and the Sonoran desert in the United States (4, 12–14). As expected from the physical characteristics of *Tamarix* leaf surfaces, the taxonomic composition of their resident bacterial community is very different from that found on the leaves of other plant species (see Fig. S1 in the supplemental material). Analysis of hypervariable regions of the small ribosomal subunit gene revealed thousands of bacterial genera accompanied by dozens of fungal taxa. Both geographic location and tree species were determinants of microbial community structure, with the former being significantly more dominant. *Tamarix* leaves in the Mediterranean region were dominated by *Halomonadaceae*, whereas trees from the Dead Sea and Sonoran desert area were dominated by *Bacillaceae* (4, 13). In light of these differences in community composition within the same tree species and between *Tamarix* and other host species, we set out to characterize the functional traits of bacteria inhabiting the *Tamarix* phyllosphere.

In order to gain insight into the functional characteristics of the major bacterial inhabitants of the *Tamarix* phyllosphere, we performed shotgun metagenomic sequencing of 10 leaf samples collected from *Tamarix aphylla* trees from five sites. Metagenomic reads were assembled, and contigs were binned into taxonomically coherent clusters that represent the major bacterial taxa in the *Tamarix aphylla* phyllosphere. The process of assembly and binning provides a useful *in silico* technique to separate bacterial from eukaryotic DNA. It also allows not only collection of a census of taxa and their functions in the environment but also making the crucial link between taxa and functions, identifying entire metabolic pathways and their corresponding gene families/operons. This link, provided by genome bins of the major bacterial components of this habitat, helps us understand the evolutionary processes that shaped the community adapted to reside on *Tamarix* leaves. It also enables the functional comparison of the two dominant groups, *Halomonadaceae* and *Bacillaceae*. We hypothesized that the genome bins extracted from the data will indicate adaptations to the stressful conditions found on the *Tamarix* leaf surface: high salinity, high alkalinity, daily desiccation, and UV exposure. We further hypothesized that the *Tamarix* epiphytes will be adapted to an aerobic heterotrophic lifestyle, due to the facts that they are located atop primary producing oxygenic organs and that dissolved organic carbon is available in ample quantities on *Tamarix* leaves. However, the main focus of this study is the exploration of hitherto-unknown and unique genomic characteristics of this bacterial habitat and how it differs from other hypersaline environments.

## MATERIALS AND METHODS

**Leaf sampling and processing.** Duplicate leaf samples were collected from five sites across Israel (see Fig. S2 in the supplemental material): by the Mediterranean (MM) (32.561428°N, 34.909674°E), north of the Dead Sea (EB) (31.843883°N, 35.513255°E), south of the Dead Sea (NK) (30.942444°N, 35.367059°E), in the Negev Desert highlands (SB) (30.871124°N, 34.78456°E), and near the northern tip of the Gulf of

Aqaba, the city of Eilat, in the southernmost part of the country (E) (29.581368°N, 34.965274°E). Two trees were sampled in each site, with the exception of MM, where only one tree was found and was sampled from two opposing sides of the tree. Leaves were collected on 21 and 22 March 2013 between 11:00 a.m. and 4:00 p.m. from different parts of each tree at random and were stored in sterile paper envelopes; the samples were returned to the laboratory and processed within 2 to 5 h of sampling. The leaves were placed inside 50-ml sterile plastic test tubes (Falcon) and immediately immersed in sterile phosphate-buffered saline (PBS) medium (10 g of leaf/40 ml of PBS, pH 7.4). Bacteria were dislodged from the leaves using a sonication tub (Transistor/ultrasonic T7; L&R Manufacturing Co.) for 2 min at medium intensity and by vortexing six times for 10 s at 5-min intervals. The leaf wash (LW) was separated from the leaf debris by decanting and kept for analysis. Leaf washes were filtered onto a 0.22- $\mu$ m-pore-size membrane filter (Millipore), and microbial community DNA was extracted with the PowerSoil microbial DNA extraction kit (MoBio, Carlsbad, CA). For chemical analysis, an additional 1 g of leaves was immersed in distilled water, vortexed 3 times for 15 s, and filtered. The filtrate was measured for electrical conductivity (EC) (a proxy for salinity), using a conductivity meter (S30 SevenEasy Conductivity; Mettler, Toledo, OH). pH was determined using a pH meter equipped with a combination glass electrode (model 420, Orion; ThermoOrion). Climatic data for each site were retrieved from the Israeli Meteorological Service (IMS) database (<http://www.ims.gov.il/IMSEng/CLIMATE>), which has weather stations within a 50-km radius of each site.

**DNA amplification and sequencing.** A detailed description of the amplicon sequencing pipeline used here has previously been published (18). Briefly, the V6 region of the 16S rRNA gene (60 to 65 nucleotides [nt] in length) was PCR amplified in triplicate for each sample using a mix of four forward primers, 967F (5'-CTAACCGANGAACCTYACC-3', 5'-ATACGCGARGAACCTTACC-3', 5'-CNACGCGAAGAACCCTTANC-3', and 5'-CAACGCGMARAACCTTACC-3'), and one degenerate reverse primer, 1046R (5'-CGACRRCCATGCANCACCT-3'), fused to barcodes compatible with the Illumina HiSeq1000 sequencing platform bridge adapters as described previously (18, 19). The 10 amplicon libraries were sequenced as part of a 96-sample barcoded run on a single HiSeq lane. Metagenomic libraries were generated from each DNA sample using the Ovation ultralow kit (NuGen) with an input of 100 ng of DNA and 8 amplification cycles. Overlapping ( $2 \times 100$  nt with  $\sim 40$  nt of overlap) metagenomic DNA libraries were run on Pippin prep to precisely select the desired length for DNA fragments to be used for sequencing on a HiSeq platform (Illumina). Two separate sequencing runs were performed, each combining all 10 samples on a single HiSeq lane. Paired-end reads were quality filtered and merged using the “merge-illumina-pairs” custom script (18) with default parameters. The script is available online at <http://github.com/meren/illumina-utils>.

**16S ribosomal gene analysis.** Sequences were processed automatically using the VAMPS Illumina annotation pipeline (20). The pipeline implements sequence quality trimming and filtering, based on perfect overlap of paired-end reads, which eliminates the bulk of sequencing errors (18). After quality filtering, the reads are assigned taxonomy using the Global Assignment of Sequence Taxonomy (GAST) (21) against the SILVA 102 database (22).

**Metagenomic analysis.** A total of  $2.3 \times 10^8$  overlapping paired-end reads (165 nt on average) distributed over 20 samples (5 sites with 2 biological replicates and 2 technical replicates;  $1.9 \times 10^9 \pm 4.3 \times 10^8$  bp per replicate) were annotated using MG-RAST (23) to evaluate the overall diversity and functionality of the *Tamarix* phyllosphere. These reads were assembled using CLC genomics workbench V.6 (CLCbio, Aarhus, Denmark). We required a minimum of 97% sequence identity over the full read length for both assembly and coverage estimates. First, a coassembly of data from the 20 samples was performed, processed, and visualized using the platform *anvi'o* v. 1.2.3 (24). The display of all samples revealed that most bacterial populations were site specific (Fig. 1). Thus, we also performed site-specific metagenomic coassemblies (2 biological replicates



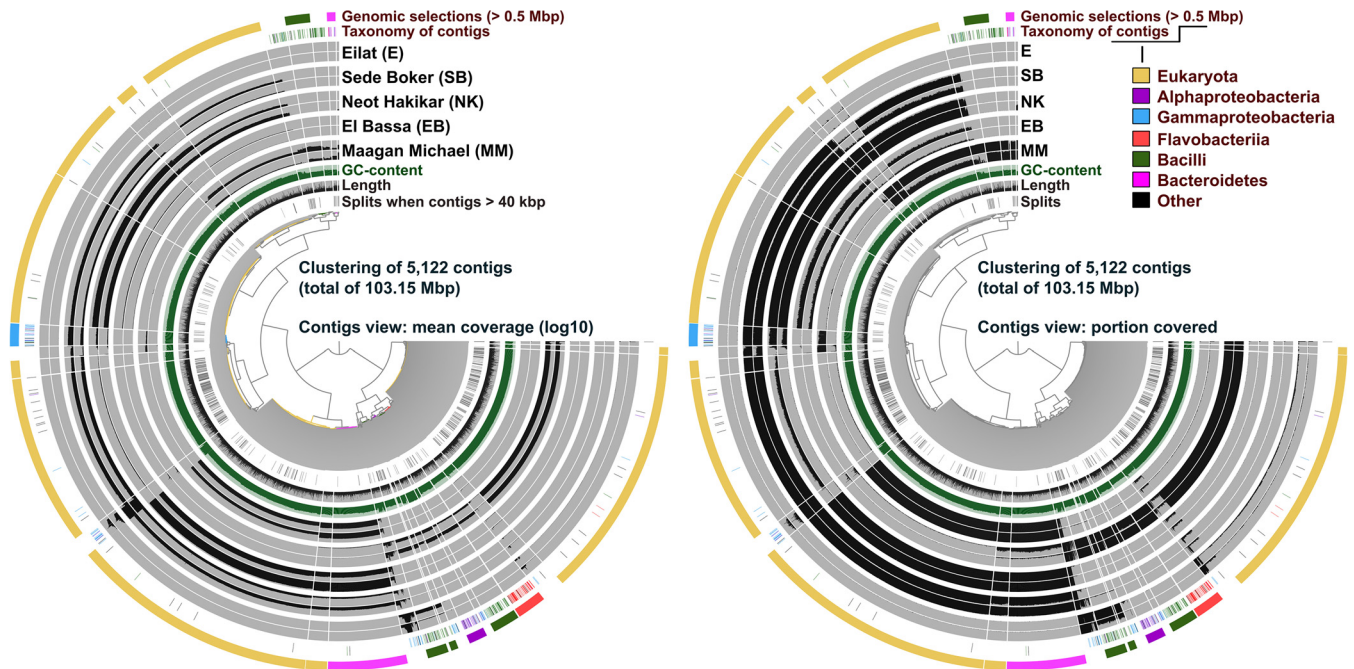


FIG 1 Clustering of 5,122 contigs (minimum length, 5 kbp; total of 103.15 Mbp) coassembled from all phyllosphere metagenomic data ( $n = 20$ ) based on tetranucleotide frequencies and coverage values. Contigs longer than 40 kbp were split into pieces of 20 kbp (see splits layer) in order to optimize the view of well-assembled genomes. Length, GC content, and taxonomy (when detected) are displayed for each contig as independent layers. Taxonomy was inferred using myRAST function `svr_assign_to_dna_using_figfams`. The mean coverage (left panel) and portion covered (right panel) of each contig are displayed across biological samples ( $n = 10$ ). Finally, 19 genomic selections were made for clusters larger than 0.5 Mbp (outer layer) and colored based on RAST taxonomy results.

and 2 technical replicates per site) in order to minimize data complexity and optimize identification of draft genomes. Contigs were clustered according to tetranucleotide frequency (TNF) using hierarchical clustering (`hclust` function in `stats` package for R with a Euclidean distance metric), and bins were manually curated by comparing TNF clusters to read coverage data. For each site, contigs with coverage patterns significantly different from the rest of the bin were removed. In addition, raw metagenomics reads were mapped to the genomic bins to assess their relative abundance across sites. Reads were assigned taxonomy using BLAST (25).

Genomic bins were annotated as genomes using the RAST API (<http://rast.nmpdr.org>) (26). As the focus of this study was the bacterial fraction of the microbial community, eukaryotic genomic bins were omitted from downstream analysis. Eukaryotic bins were identified based on (i) the presence of eukaryotic genes and (ii) poor annotation (2 to 4% subsystem coverage, compared to 30 to 60% for bacterial bins and no significant BLAST hit for most contigs). Some eukaryotic genomic bins could also be identified by their low gene density, but not in all cases, as some fungi have gene densities that resemble those of bacteria (27).

The completeness and level of cross contamination of the resulting genomic bins were estimated using `checkM` (28) (see Table S1 in the supplemental material). To identify genomic elements enriched in the *Tamarix* phyllosphere, genomic bins were compared against a reference data set of 40 genomes. These genomes were selected from the RAST database by the criterion of having a similarity score of  $>450$  ("View closest neighbors" on The Seed Viewer) to at least one genomic bin in our data set. Genomes were compared by (i) using RAST functional profiles and (ii) generating *de novo* orthologous clusters from the data set, thus including hypothetical proteins in the profile (`Get_homologues` [29]). A phylogenetic analysis of reference genomes and genomic bins was performed using 9 concatenated ribosomal proteins that were shared across all genomes. Concatenated proteins were aligned using `MUSCLE` (30), and a maximum-likelihood (ML) bootstrapped tree was generated using `MEGA6` (31). `MEGA6` was also used to visualize the cluster dendrogram generated by `Get_homologues`. Genomic bins are named here using the

acronym of the site from which they were assembled and binned (see Fig. S2 in the supplemental material).

**Accession numbers.** All nucleotide data from this study were submitted to the NCBI SRA database under the project accession number [SRP068421](https://www.ncbi.nlm.nih.gov/sra/PRJNA308562) and BioProject accession number [PRJNA308562](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA308562). The VAMPS database (<https://vampls.mbl.edu>) provides public access to quality-filtered V6 reads under the project name APF\_TAM-Bv6 (*Tamarix* Phyllosphere Time Series). The present publication includes only samples MM1, MM2, EB1, EB2, NK1, NK2, SB1, SB2, E1, and E2 within this data set. Unassembled metagenomics reads and annotations are available online at <http://metagenomics.anl.gov> under the project titled "*Tamarix aphylla* phyllosphere metagenome." GenBank and nucleic acid fasta files of genome bins can be accessed at <https://figshare.com/s/550709df3331c43838c>.

## RESULTS

**Bacterial community composition and diversity.** Microbial community composition was determined by (i) estimating the relative abundance of bacterial taxa from amplicon sequencing of the V6 region of the 16S rRNA gene and (ii) extracting 16S rRNA gene fragments from both technical replicates of shotgun metagenomics data. Taxonomy derived from 16S rRNA gene reads extracted from shotgun metagenomic data typically has a lower resolution, as different reads cover different regions in the gene, some of them highly conserved across phyla. However, it provides a comparison of relative abundances across domains. An average of 25,000 16S/18S rRNA gene reads were detected in each data set querying against the MG-RAST M5RNA database. Bacteria and fungi accounted for  $36\% \pm 23\%$  and for  $16\% \pm 20\%$  of reads, respectively. Animal-derived reads accounted for about a fifth of the combined 16S/18S rRNA gene reads in the data (Fig. 2). The

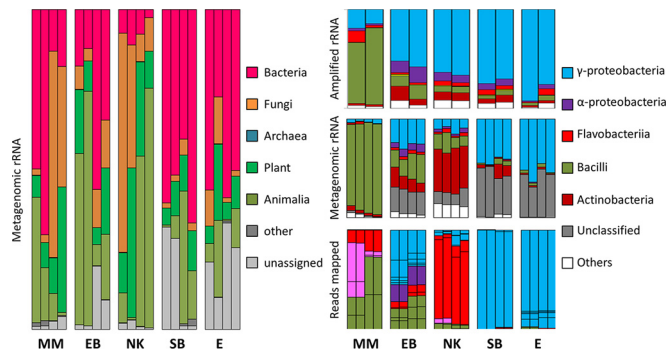


FIG 2 Relative abundance of SSU reads. Left, cross domain relative abundance based on rRNA gene (SSU) read counts in metagenomic data sets. Right, comparison between bacterial SSU profiles from amplicon (top panels) and shotgun metagenomic (middle panels) data sets, normalized to the number of bacterial SSU copies in the metagenomic data set. The bottom panel shows the proportion of reads mapped from each metagenomic data set to genomic bins.

bacterial community composition was generally uniform within a site (Fig. 2, right panel).

Bacterial communities at the desert sites (EB, NK, SB, and E) were dominated by *Halomonadaceae*, whereas the site from the Mediterranean shore (MM) was dominated by *Bacillaceae* (Fig. 2 and 3). *Gammaproteobacteria* and *Bacilli*, the corresponding classes, were negatively correlated (Fig. 3, inset; Pearson correlation,  $r = -0.80$ ;  $P = 0.006$ ). While this observation stands in apparent contrast to a previous study where an opposite trend was observed (4), the contrast between *Halomonadaceae*- and *Bacillaceae*-dominated communities is consistent with previous observations. Taxonomy assignments of 16S rRNA gene amplicons were in general agreement with those derived from metagenome reads (Pearson correlation,  $r = 0.98$  [ $P < 1 \times 10^{-6}$ ] for *Bacilli*;  $r = 0.78$  [ $P = 0.007$ ] for *Gammaproteobacteria*). The 16S rRNA gene data set from the first metagenomic technical replicate contained reads belonging to *Clostridia* and *Bacteroidia*. These reads, being absent from the second replicate, were suspected to be contamination that was introduced during the sequencing process and were omitted from further analyses.

**Assembly and binning of 17 bacterial genomic bins.** An initial assembly and binning step with a combined assembly of all data sets (Fig. 1), demonstrated that (i) the majority of contigs originated from eukaryotic genomes and (ii) the read coverage of most of the bacterial genomic bins can be mapped to a single site. Therefore, the data were reassembled, assembling and binning samples from each site separately. This approach yielded 17 bacterial genomic bins, varying in size from  $\sim 200$  kb to  $\sim 3.5$  Mb. These bins ranged in completeness from 47% to 99% and in contamination from 0 to 23% (see Table S1 in the supplemental material). Phylogenetic placement was assessed by aligning the concatenated sequence of 9 ribosomal proteins that were found as single-copy genes in each of the genomic bins. Thirty-eight reference genomes from the RAST database were added to the alignment by selecting genomes with a similarity score of  $>450$  (“View closest neighbor” option in the SEED Viewer) to at least one of the genomic bins, and a maximum-likelihood phylogeny was constructed (Fig. 4, right panel). The 17 genomic bins belong to 5 classes: *Gammaproteobacteria*, *Alphaproteobacteria*, *Flavobacteria*, *Bacteroidetes*, and *Bacilli*. The taxonomy of the genomic bins reflects the 16S rRNA gene-derived taxonomic composition, albeit with one anomaly:

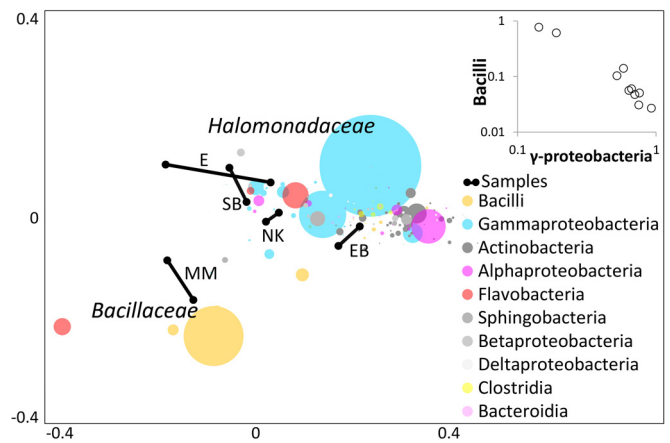


FIG 3 Nonmetric multidimensional scaling (NMDS) ordination of Bray-Curtis similarity matrix based on V6 data at the family taxonomic level. The spatial proximity of samples (black dots) on the graph is proportional to the similarity in community composition. Bacterial families are shown as bubbles, positioned according to their relative abundance among samples. Bubbles are colored according to class and sized by relative abundance. Classes in gray scale were not represented in genomic bins. The inset graph displays the negative correlation between the abundances at the different sampling sites of the two major classes, *Bacilli* and *Gammaproteobacteria*.

an absence of genomic bins belonging to *Actinobacteria*. *Halomonadaceae* and *Bacillaceae*, the two most dominant families according to this and previous 16S rRNA gene surveys, were represented by five bins each (Fig. 4). Five other bins appear to have originated from insect endosymbionts (MM6, MM8, SB2, and E2). Two additional bins, representing *Rhodobacteraceae* and *Flavobacteria*, originated from site EB. Sites MM and EB contributed most of the genomic diversity, with 5 genomic bins originating from the assembly of each. Site NK, on the other hand, did not contribute to any of the bins. This site also had the smallest proportion of bacterial DNA (Fig. 2, left panel). The genomic bins originating from site MM were classified as *Bacteroidetes* (MM6 and MM8) and *Bacillus* (MM5). The taxonomy of bins MM7 and MM2 could not be determined due to the absence of informative marker genes. However, the overall gene content of MM7 places it within *Bacilli*. Bins originating from site EB were classified as *Rhodobacteraceae* (EB4), *Oceanospirillales* (EB6 and EB8), *Flavobacteria* (EB1), and *Bacillus* (EB2, EB3). Site SB contributed two bins: one *Oceanospirillales* (SB1) and one unresolved bin closely related to MM8. Site E contributed three bins: one *Oceanospirillales* (E3, a close relative of SB1 and EB8), one *Bacilli* (E1, a close relative of EB3), and one unresolved (E2, a close relative of MM8 and SB2). Comparative genome analysis was performed within each class to detect genes and gene families that are enriched in the *Tamarix* phyllosphere. All predicted protein sequences were grouped into ortholog clusters, resulting in 38,869 unique clusters across the 17 bins and the 38 reference genomes. Genomes were hierarchically clustered according to gene content (Fig. 4, left panel). The phylogenetic and gene content trees were in close agreement with each other, except for a notable difference with regard to bins MM8, SB2, E2, and MM6. The phylogenetic analysis places MM8, SB2, and E2 as an ancestral clade of *Flavobacteria* and MM6 as a distant relative of “*Candidatus* Amoebophilus asiaticus,” an endosymbiont belonging to *Bacteroidetes*. The orthologous gene contents of all four bins, however, were relatively close to each other and to

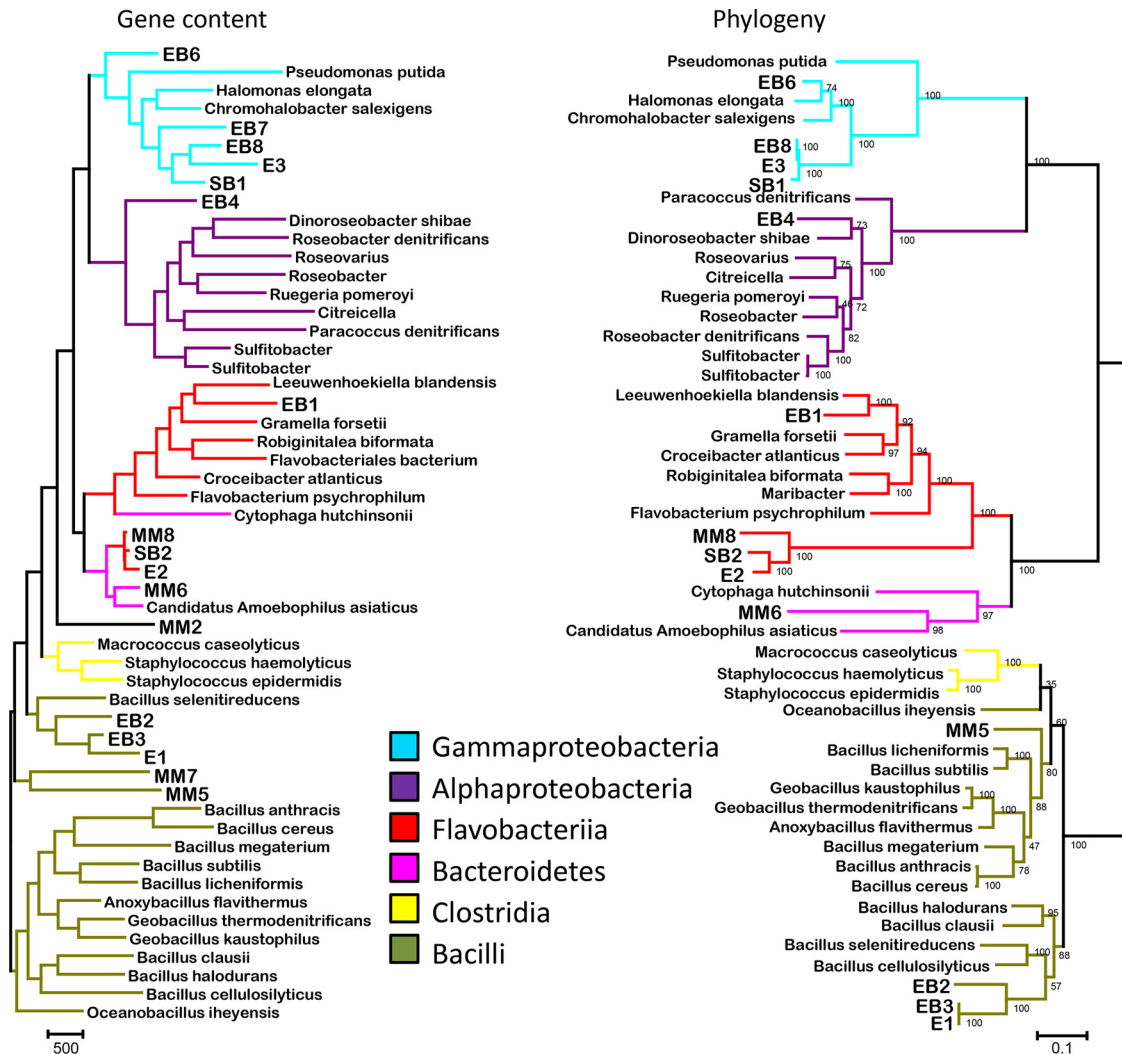


FIG 4 Comparative view of gene content tree (based on Get\_homologue analysis) (left) and phylogenetic tree (based on a concatenated alignment of 9 ribosomal proteins) (right). Both trees include genomic bins and reference genomes used for comparative genomics. Branches are colored by class. Bootstrap values are shown on the nodes of the phylogenetic tree.

“*Candidatus Amoebophilus asiaticus*.” This suggests that (i) these are genomes of endosymbionts probably associated with an animal/protist inhabitant of *Tamarix* leaves and (ii) despite considerable phylogenetic divergence, they show convergent adaptation to a similar lifestyle. A great number of orthologs ( $n = 9,145$ ) within this pool were not present in any of the *Tamarix* phyllosphere genomic bins. This was partly due to the fact that the genomic bins were only partially completed. Conversely, 1,163 of the orthologs were found exclusively in *Tamarix* phyllosphere genomic bins.

**Contrasting metabolic lifestyles of *Halomonadaceae* and *Bacillaceae*.** The KO annotations of *Halomonadaceae* and *Bacillaceae* were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway reconstruction tool. Using this tool, the metabolic maps from four categories (carbohydrate metabolism, energy metabolism, xenobiotic biodegradation, and membrane transport) were used to reconstruct the primary metabolic features of the genomes (Fig. 5). This analysis suggests that *Bacillaceae* in this habitat correspond to a more generalistic/opportunistic lifestyle, with the ability to scavenge and utilize a wide

variety of carbon compounds. *Halomonadaceae*, on the other hand, display a less versatile and more specialized metabolic repertoire but are potentially able to metabolize plant-derived aromatic compounds such as salicylate and vanillate.

**Adaptations to the ionic environment of *Tamarix* leaves.** In a habitat such as the *Tamarix* phyllosphere, with a high concentration of  $\text{Na}^+$  ions and a low concentration of  $\text{H}^+$  ions, it is expected that  $\text{Na}^+/\text{H}^+$  antiporter will be a prevalent mechanism (32, 33). Indeed,  $\text{Na}^+/\text{H}^+$  antiporter genes are present in multiple copies in all genomic bins of the free-living bacteria (and absent in all four bins identified as insect endosymbionts). In addition to the need to acidify the cell, one important but conflicting challenge faced by alkaliphilic bacteria is a bioenergetic one: ATP synthase requires an  $\text{H}^+$  gradient opposite to the one found in a high-pH environment. While it is not known how alkaliphilic *Bacilli* overcome this hindrance, several mutations in the *a* and *c* subunits of the  $F_0$  domain of the ATP synthase of alkaliphilic *Bacilli* have been shown to be important for functioning at a high pH (34, 35). Interestingly, the  $F_0$  domain found in genomic bins EB3 and E1



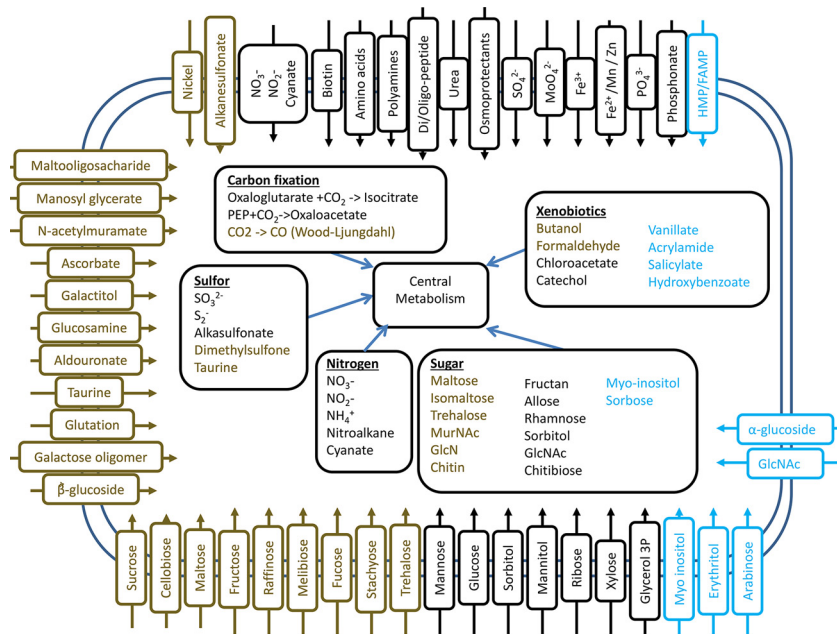


FIG 5 Metabolic diagram of the main inhabitants of the *Tamarix* phyllosphere, based on mapping of the KEGG profile onto metabolic maps. Compounds appearing in the diagram are source compounds that can be linked to the central metabolic modules. Membrane transport includes ATP binding cassette (ABC) transporters and phosphotransferase systems (PTS). For ABC transporters, it was required that at least two of the three components are present; for PTS with more than one component, both were required to be present in order to be included in the model. Brown, Gram positive; blue, Gram negative; black, core.

has only one of these alkaliphilic residues (subunit a, K180) (see Fig. S3 in the supplemental material). It also has a typical thermophilic residue (subunit 3, S22) (see Fig. S3 in the supplemental material), placing it as a hitherto-unknown intermediary form between neutrophilic, alkaliphilic, and thermophilic forms of this gene.

**Photoperception is a core trait of *Tamarix* phyllosphere bacteria.** To identify functions of potential importance for survival in this habitat, we created clusters of orthologous groups of proteins (COGs) from the combined data set of genomic bins and reference genomes. We rationalized that core functions important for survival in the *Tamarix* phyllosphere would not be confined to a single taxonomic class. Thus, we compiled a list of COGs that were found in at least three taxonomic classes within the genomic bins but in no more than one of the four taxonomic classes within the reference genomes (*Alphaproteobacteria*, *Gammaproteobacteria*, *Flavobacteria*, and *Firmicutes*). This approach yielded eight *Tamarix*-enriched COGs (Fig. 6). Only one COG, annotated as hypothetical, was found in all four taxonomic classes in the *Tamarix* phyllosphere while present only in *Flavobacteria* among the reference group. This COG is also represented in three different sampling sites. This COG shares sequence similarity with methyltransferases and ATP binding proteins and is commonly found downstream from a hypothetical kinase. Two COGs, found in *Firmicutes* reference genomes only, were identified here in members of the *Gammaproteobacteria* and *Flavobacteria* as well as in *Firmicutes*. One is an Mn-containing catalase, and the other is KdgT, which is involved in the transport and degradation of pectin, a component of plant cell walls. The third group of COGs was found in both proteobacterial classes and in *Firmicutes*. This group includes two hypothetical proteins, as well as cobrinic synthase and cyanate hydratase.

A COG homologous to a rhodopsin gene was found in genome bins from all three Gram-negative classes and from two separate

sampling sites (EB and E). Interestingly, the rhodopsins found here belong to the newly described family of NQ rhodopsins, implicated as sodium pumps (36) (see Fig. S4 in the supplemental material). Genes for photoactive yellow protein (PYP), another light-sensing mechanism, were found in two proteobacterial genome bins from two separate sites (SB1 and EB4) but were completely absent from reference genomes. In bin SB1, the gene was found in tandem with a coenzyme A (CoA) ligase gene, in a syntentic formation to the genomes of *Idiomarina loihiensis* and *Halorhodospira halophila*. This gene is necessary for the PYP chromophore biosynthesis (37). Photoperception also appears to be a defining trait among Gram-positive *Tamarix* genome bins: blue-

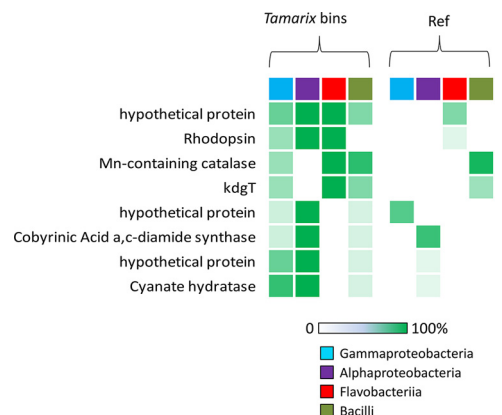


FIG 6 Functions unique to *Tamarix*. Functions that are ubiquitous in *Tamarix* bins and are restricted to one class or less in closely related reference genomes are shown. The functions shown include only those that are shared by at least three separate taxonomic classes among genome bins, allowing occurrence in only one taxonomic class among the reference genomes. Color is proportional to the prevalence of the gene within each class.



light photoreceptor genes were found in three *Bacillus* bins but were absent from all reference genomes.

In a complementary approach, we compiled a list of operons unique to *Tamarix* by finding genes that are absent in reference genomes and colocalized on contigs of genomic bins (allowing up to 2 open reading frames [ORFs] between them). In *Bacilli*, this method identified 27 unique contiguous gene clusters (see Data Set S2 in the supplemental material), among them the genes for the blue light photoreceptor mentioned above and the *gvp* operon, encoding gas vesicle proteins in three *Bacillus* genome bins from sites EB and E.

In *Gammaproteobacteria*, we identified 65 unique contiguous gene clusters that were found in two or more *Tamarix* genome bins (see Data Set S3 in the supplemental material). One cluster, represented in three genome bins, contains genes for lycopene cyclase, phytoene synthase, and phytoene dehydrogenase, involved in the biosynthesis of retinal (the rhodopsin cofactor). Another notable cluster, represented in four genome bins, contains 6 genes, among them two copper resistance genes and genes *yagS* and *yagR*, which are involved in the degradation of aromatic compounds.

**Anoxygenic phototrophy.** The metabolic profiles of the genomic bins found in this study clearly indicated an aerobic, heterotrophic lifestyle, with one exception: bin EB4, which is taxonomically related to the genus *Dinoroseobacter* (Fig. 4). Members of the *Roseobacteria* lineage are known to have symbiotic relationships with a variety of aquatic photosynthetic organisms, including higher plants (38). This bin contains genes encoding an anoxygenic photosynthetic apparatus, providing evidence that light may be harvested for energy by phyllosphere bacteria. A comparison of the gene content of EB4 to that of *Dinoroseobacter shibae* (see Data Set S1 and Fig. S5 in the supplemental material) reveals that 32 of the 44 phototrophy genes are shared among these two genomes. Forty of the photosynthetic apparatus assembly genes are found in a single genomic cluster in *D. shibae*. There is a large degree of synteny between this region and a collection of five contigs from EB4 (see Fig. S5 in the supplemental material). The cytochrome *c* subunit gene is missing from the middle of a contig and is replaced by a *pufX* gene, which is not found in *D. shibae*.

## DISCUSSION

The environmental challenge faced by the *Tamarix* bacterial epibionts is more akin to marine and alkaline soil environments than to other phyllosphere habitats. However, there are some unique characteristics of the phyllosphere that would be common to all plant surfaces, including those of *Tamarix*. One such trait, as is apparent from the ubiquity of light-sensing and light protection genes, is the exposure to solar radiation. The ability to sense and utilize light by phyllosphere microbes is emerging as a general trend (39–41). Indeed, the prevalence of light-sensing and light utilization mechanisms, such as rhodopsin, PYP, and, notably, anoxygenic photosystem genes, reiterates the contention that light is an important resource and signal in the leaf surface habitat. The anoxygenic photosystem, however, does not appear to be coupled with the ability to fix carbon, as no copies of RuBisCo were found in the data set. It is highly likely that this photosynthetic apparatus is used for ATP production, a process that, as explained above, poses a special challenge in this environment.

All four *Bacillus* and *Oceanobacillus* genomic bins contain a gas vesicle operon. Gas vesicle proteins (Gvps) are known in *Bacilli*, predominantly in aquatic species, as the only known use for gas

vesicles is buoyancy. Among the related *Bacillus* genomes used here for comparison, only *Bacillus megaterium* contained buoyancy genes. Li and Cannon (42) described the *Gvp* operon in *B. megaterium*, referring to it as a soil organism. However, *B. megaterium* was also found in marine habitats or as an endophyte (43). The presence of this gene cassette in all *Bacillus* genomic bins testifies to this operon serving a yet-unidentified adaptive role in the *Tamarix* phyllosphere.

Despite the limiting conditions on *Tamarix* leaves, the communities surveyed in this and previous studies (4, 13) cluster into two distinct types: *Halomonadaceae* dominated and *Bacillaceae* dominated. In previous studies, conducted during the summer, *Halomonadaceae* dominated the temperate, more humid sites and *Bacillaceae* dominated the desert sites. The present study, conducted toward the end of winter, revealed a surprising opposite trend, with *Halomonadaceae* dominating the desert sites and *Bacillaceae* the Mediterranean site. This observation, which warrants further investigation, suggests that *Halomonadaceae* have a relatively limited range of temperature and humidity in which they can flourish and that their optimal conditions are encountered in the desert during the winter but in the Mediterranean region during summer. The use of genomic binning enabled a direct comparison of the dominant members of these two groups, revealing contrasting life history strategies between the generalist and opportunistic lifestyle of the *Bacillaceae* and the specialized lifestyle of the *Halomonadaceae*. While this observation does not contradict the hypothesis that *Halomonadaceae* are limited by abiotic conditions, it also offers an alternative hypothesis that would explain the stark metabolic differences between these two groups. This alternative hypothesis would suggest that physiological changes undergone by the host tree under different temperature and humidity regimens may result in different leaf exudate compositions, which give rise to different microbial communities.

## ACKNOWLEDGMENT

We are indebted to Kasia Hammer of the Josephine Bay Paul Center in Comparative Molecular Biology and Evolution at the Marine Biology Laboratory for her excellent and dedicated technical support.

## FUNDING INFORMATION

This research was supported by United States-Israel Binational Science Foundation grant 2010262 to S.B. and A.F.P. Travel grants were awarded to O.M.F. by the United States-Israel Binational Science Foundation (Prof. Rahamimoff Travel Grant for Young Scientists) and by the Bat-sheva de Rothschild Fund (Aharon and Ephraim Katzir Travel Fellowship).

## REFERENCES

1. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? *PLoS Biol* 9:e1001127. <http://dx.doi.org/10.1371/journal.pbio.1001127>.
2. Redford A, Bowers RA, Knight R, Linhart Y, Fierer N. 2010. The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves. *Environ Microbiol* 12:2885–2893. <http://dx.doi.org/10.1111/j.1462-2920.2010.02258.x>.
3. Lambais MR, Crowley DE, Cury JC, Bull RC, Rodrigues RR. 2006. Bacterial diversity in tree canopies of the Atlantic forest. *Science* 312:1917. <http://dx.doi.org/10.1126/science.1124696>.
4. Finkel OM, Burch AY, Lindow SE, Post AF, Belkin S. 2011. Geographical location determines the population structure in phyllosphere microbial communities of a salt-excreting desert tree. *Appl Environ Microbiol* 77:7647–7655. <http://dx.doi.org/10.1128/AEM.05565-11>.
5. Maignien L, DeForce EA, Chafee ME, Eren AM, Simmons SL. 2014.

- Ecological succession and stochastic variation in the assembly of *Arabidopsis thaliana* phyllosphere communities. *mBio* 5:e00682-13. <http://dx.doi.org/10.1128/mBio.00682-13>.
6. Copeland JK, Yuan L, Layghifrad M, Wang PW, Guttman DS. 2015. Seasonal community succession of the phyllosphere microbiome. *Mol Plant Microbe Interact* 28:274–285. <http://dx.doi.org/10.1094/MPMI-10-14-0331-FI>.
  7. Remus-Emsermann MNP, Lückner S, Müller DB, Potthoff E, Daims H, Vorhold JA. 2014. Spatial distribution analyses of natural phyllosphere-colonizing bacteria on *Arabidopsis thaliana* revealed by fluorescence in situ hybridization. *Environ Microbiol* 16:2329–2340. <http://dx.doi.org/10.1111/1462-2920.12482>.
  8. Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlappbach R. 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc Natl Acad Sci U S A* 106:16428–16433. <http://dx.doi.org/10.1073/pnas.0905240106>.
  9. Müller T, Ruppel S. 2014. Progress in cultivation-independent phyllosphere microbiology. *FEMS Microbiol Ecol* 87:2–17. <http://dx.doi.org/10.1111/1574-6941.12198>.
  10. Rastogi G, Coaker GL, Leveau JH. 2013. New insights into the structure and function of phyllosphere microbiota through high-throughput molecular approaches. *FEMS Microbiol Lett* 348:1–10. <http://dx.doi.org/10.1111/1574-6968.12225>.
  11. Vorholt JA. 2012. Microbial life in the phyllosphere. *Nat Rev Microbiol* 10:828–840. <http://dx.doi.org/10.1038/nrmicro2910>.
  12. Qvit-Raz N, Jurkevitch E, Belkin S. 2008. Drop-size soda lakes: transient microbial habitats on a salt-secreting desert tree. *Genetics* 178:1615–1622. <http://dx.doi.org/10.1534/genetics.107.082164>.
  13. Finkel OM, Burch AY, Elad T, Huse SM, Lindow SE, Post AF, Belkin S. 2012. Distance-decay relationships partially determine diversity patterns of phyllosphere bacteria on *Tamarix* trees across the Sonoran Desert. *Appl Environ Microbiol* 78:6187–6193. <http://dx.doi.org/10.1128/AEM.00888-12>.
  14. Qvit-Raz N, Finkel OM, Al-Deeb TM, Malkawi HI, Hindiyeh MY, Jurkevitch E, Belkin S. 2012. Biogeographical diversity of leaf-associated microbial communities from salt-secreting *Tamarix* trees of the Dead Sea region. *Res Microbiol* 163:142–150. <http://dx.doi.org/10.1016/j.resmic.2011.11.006>.
  15. Burch AY, Finkel OM, Cho JK, Belkin S, Lindow S. 2013. Diverse microhabitats experienced by *Halomonas variabilis* on salt-secreting leaves. *Appl Environ Microbiol* 79:845–852. <http://dx.doi.org/10.1128/AEM.02791-12>.
  16. Waisel Y. 1961. Ecological studies on *Tamarix aphylla* (L.) Karst. III. The salt economy. *Plant Soil* 13:356–364.
  17. Waisel Y. 1991. The glands of *Tamarix aphylla*: a system for salt recretion or for carbon concentration? *Physiol Plant* 83:506–510.
  18. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS One* 8:e66643. <http://dx.doi.org/10.1371/journal.pone.0066643>.
  19. Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, Sogin ML, Vanreusel A. 2014. Host-specificity among abundant and rare taxa in the sponge microbiome. *ISME J* 8:1198–1209. <http://dx.doi.org/10.1038/ismej.2013.227>.
  20. Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM, Sogin ML. 2014. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics* 15:41. <http://dx.doi.org/10.1186/1471-2105-15-41>.
  21. Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, Sogin ML. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* 4:e1000255. <http://dx.doi.org/10.1371/journal.pgen.1000255>.
  22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <http://dx.doi.org/10.1093/nar/gks1219>.
  23. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. <http://dx.doi.org/10.1186/1471-2105-9-386>.
  24. Eren AM, Esen ÖC, Quince CC, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi: an advanced genomics and visualization platform for omics data. *PeerJ* 3:e1319. <http://dx.doi.org/10.7717/peerj.1319>.
  25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
  26. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702. <http://dx.doi.org/10.1093/nar/gki866>.
  27. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. 2005. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 15:1620–1631. <http://dx.doi.org/10.1101/gr.3767105>.
  28. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <http://dx.doi.org/10.1101/gr.186072.114>.
  29. Contreras-Moreira B, Vinusa P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. <http://dx.doi.org/10.1128/AEM.02411-13>.
  30. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <http://dx.doi.org/10.1093/nar/gkh340>.
  31. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.
  32. Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* 28:4317–4331. <http://dx.doi.org/10.1093/nar/28.21.4317>.
  33. Takami H, Takaki Y, Uchiyama I. 2002. Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res* 30:3927–3935. <http://dx.doi.org/10.1093/nar/gkf526>.
  34. Krulwich TA, Sachs G, Padan E. 2011. Molecular aspects of bacterial pH sensing and homeostasis. *Nat Rev Microbiol* 9:330–343. <http://dx.doi.org/10.1038/nrmicro2549>.
  35. McMillan DGG, Keis S, Dimroth P, Cook GM. 2007. A specific adaptation in the a subunit of thermoalkaliphilic F<sub>1</sub>F<sub>0</sub>-ATP synthase enables ATP synthesis at high pH but not at neutral pH values. *J Biol Chem* 282:17395–17404. <http://dx.doi.org/10.1074/jbc.M611709200>.
  36. Kwon SK, Kim BK, Song JY, Kwak MJ, Lee CH, Yoon JH, Oh TK, Kim JF. 2013. Genomic makeup of the marine flavobacterium *Nonlabens (Donghaeana) dokdonensis* and identification of a novel class of rhodopsins. *Genome Biol Evol* 5:187–199. <http://dx.doi.org/10.1093/gbe/evs134>.
  37. Kort D, Hoff WD, Van West M, Kroon SM, Hoffer SM, Vlieg KH, Crielaand W, Van Beeumen JJ, Helingwerf KJ. 1996. The xanthopsins: a new family of eubacterial blue-light photoreceptors. *EMBO J* 15:3209–3218.
  38. Wagner-Döbler I, Biebl H. 2006. Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol* 60:255–280. <http://dx.doi.org/10.1146/annurev.micro.60.080805.142115>.
  39. Atamna-Ismael N, Finkel OM, Glaser F, Sharon I, Schneider R, Post AF, Spudich JL, von Mering C, Vorholt JA, Iluz D, Béja O, Belkin S. 2012. Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environ Microbiol* 14:140–146. <http://dx.doi.org/10.1111/j.1462-2920.2011.02554.x>.
  40. Atamna-Ismael N, Finkel OM, Glaser F, von Mering C, Vorholt JA, Koblížek M, Belkin S, Béja O. 2012. Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environ Microbiol Rep* 4:209–216. <http://dx.doi.org/10.1111/j.1758-2229.2011.00323.x>.
  41. Stiefel P, Zambelli T, Vorholt J. 2013. Isolation of optically targeted single bacteria by application of fluidic force microscopy to aerobic anoxygenic phototrophs from the phyllosphere. *Appl Environ Microbiol* 79:4895–4905. <http://dx.doi.org/10.1128/AEM.01087-13>.
  42. Li N, Cannon MC. 1998. Gas vesicle genes identified in *Bacillus megaterium* and functional expression in *Escherichia coli*. *J Bacteriol* 180:2450–2458.
  43. Liu X, Zhao H, Chen S. 2006. Colonization of maize and rice plants by strain *Bacillus megaterium* C4. *Curr Microbiol* 52:186–190. <http://dx.doi.org/10.1007/s00284-005-0162-3>.