

Review

Applying Genome-Resolved Metagenomics to Deconvolute the Halophilic Microbiome

Gherman Uritskiy and Jocelyne DiRuggiero *

Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA; guritsk1@jhu.edu

* Correspondence to: jdiruggiero@jhu.edu

Received: 15 February 2019; Accepted: 11 March 2019; Published: 14 March 2019

Abstract: In the past decades, the study of microbial life through shotgun metagenomic sequencing has rapidly expanded our understanding of environmental, synthetic, and clinical microbial communities. Here, we review how shotgun metagenomics has affected the field of halophilic microbial ecology, including functional potential reconstruction, virus–host interactions, pathway selection, strain dispersal, and novel genome discoveries. However, there still remain pitfalls and limitations from conventional metagenomic analysis being applied to halophilic microbial communities. Deconvolution of halophilic metagenomes has been difficult due to the high G + C content of these microbiomes and their high intraspecific diversity, which has made both metagenomic assembly and binning a challenge. Halophiles are also underrepresented in public genome databases, which in turn slows progress. With this in mind, this review proposes experimental and analytical strategies to overcome the challenges specific to the halophilic microbiome, from experimental designs to data acquisition and the computational analysis of metagenomic sequences. Finally, we speculate about the potential applications of other next-generation sequencing technologies in halophilic communities. RNA sequencing, long-read technologies, and chromosome conformation assays, not initially intended for microbiomes, are becoming available in the study of microbial communities. Together with recent analytical advancements, these new methods and technologies have the potential to rapidly advance the field of halophile research.

Keywords: Extremophiles; halophilic microorganisms; hypersaline habitats; metagenomics; microbiome; shotgun sequencing; genome assembly and binning; functional annotation

1. Introduction

Microbial life is one of the most diverse and bioenergetically dominant forces in the earth's ecosphere [1], making microbiome research a critical component of modern ecology. The unparalleled taxonomic and functional diversity of microbial communities has allowed them to populate all locations on the planet [2,3], including environments unfit for colonization by other life forms. In hypersaline environments, unique environmental pressures have forced microbiota to evolve with specific survival adaptations, resulting in highly resilient communities that push the boundaries of life's limit (Figure 1). Halophiles have been found to play important roles in soil bioenergetic processes [4] and food storage and preservation [5,6], and have also been detected in the human gut microbiota [7]. Additionally, studying halophilic life forms has revealed many fundamental aspects of life's survival limits and strategies, including the potential to endure the harsh environments we are most likely to find on other planets [8,9]. Prior to the introduction of high-throughput sequencing, our understanding of halophile genomics was limited to studying cultured organisms [10,11]. While next-generation sequencing technologies have become commonplace in

microbiology, the halophile field lacks a critical analysis of prospects and potential applications of these technologies in halophilic microbiomes.

In this review, we discuss key aspects of halophile community composition and function that metagenomics has revealed and provide examples of studies in various hypersaline environments for a perspective on analytical progress. We then examine the advantages and limitations of applying shotgun metagenomic sequencing in uncovering the structure and function of halophilic microbiomes. We outline the factors and characteristics that make the deconvolution of halophilic metagenomes a major challenge and propose analytical adjustments to be made when investigating these complex communities. Both experimental design and computation analysis approaches that are appropriate in halophilic metagenomics are summarized. Finally, we discuss novel sequencing technologies that show promise in further propelling the halophile metagenomic field.

2. Shotgun Sequencing in Metagenomics

Rapid developments in high-throughput DNA sequencing technologies since the early 2000s have propelled our understanding of not only single-organism genetics, but also microbiome community structure and function [12]. Marker gene (particularly the *16S rRNA* gene) amplicon sequencing has revealed the taxonomic composition of a given community through sequencing a small target of the community's DNA. In contrast, whole-metagenomic sequencing (WMGS) theoretically allows for reconstruction of the entire microbial community's DNA content. This has led to a number of important findings in microbiome research [13–15], as biologists have been able to thoroughly investigate microbial communities at the genetic level without the need for culturing [16].

However, while sequencing technologies are rapidly developing, producing complete genomes of all the microorganisms found in a community is currently unattainable due to low sequencing coverage of the less abundant organisms. Additionally, sequence repeats and regions of homology between organisms limits genome recovery from short-read data, resulting in incomplete assemblies. Instead, long contiguous pieces (contigs) of genomes are produced, ranging in length from 1 Kbp to 1 Mbp [17,18]. These contigs then need to be grouped based on the genome they belong to, a process known as binning. It is only recently that binning has become reliable enough to produce reasonably high-quality metagenome-assembled genomes (MAGs). The ability to produce high-quality MAGs has in turn led to the discovery of thousands of novel organisms and has thus enabled many breakthroughs in characterizing the taxonomic and functional components of microbiomes [19–21].

Shotgun metagenomics offers tremendous advantages in recovering taxonomic and functional potential components of microbial communities, but sequencing costs deter some researchers from deploying this approach in their studies. The high average read coverage required for the assembly of a genome from shotgun reads [22] presents a major challenge for the assembly of less-abundant organisms in a metagenomic context. These highly diverse but underrepresented taxa often constitute significant proportions of microbial communities and play important roles in biome functioning [23]. Despite these challenges, WMGS carries tremendous benefits, empowering researchers to study previously unknown aspects of microbiomes. In particular, WMGS allows for the reconstruction of a given community's gene content, which has enabled ecologists to predict the functional potential of entire communities. This new angle of microbiome analysis has enabled the prediction of metabolic processes potentially present in communities and the study of community natural selection at the functional level [24,25]. The possibility of studying the functional potential of any organism in a community means that our understanding of microbial genetics, dynamics, evolution, and function is no longer limited to cultured organisms. In many fields, such as human microbiome research, this has hailed a new era for research [26,27].

2.1. Halophilic Microbiome Research Powered by Shotgun Metagenomics

Numerous breakthroughs in halophilic microbiome research have been enabled by WMGS [11] (Table 1). This sequencing approach reveals the taxonomic structure of microbiomes in high-salt environments with significantly less taxonomy-based biases than conventional ribosomal amplicon

sequencing. Indeed, in conventional *16S rDNA* amplicon sequencing, primer choices can have a substantial impact on taxonomic distribution, and it is difficult to reliably amplify multiple domains of life, e.g., Bacteria and Archaea, with the same primer set [28]. While WMGS still has biases associated with G + C content, taxonomic annotation of shotgun reads usually results in more accurate and robust taxonomic profiles than amplicon sequencing [29]. This is particularly important in high-salt environments, where both Archaea and Bacteria are found in high abundance. For example, shotgun sequencing has provided more comprehensive taxonomic profiles of an endolithic halite community (Figure 1B) and the discovery that a unique algae was present in this community, in addition to Halobacteria, Cyanobacteria, and other heterotrophic bacteria [30]. In the study of a hypersaline lake (Figure 1D), the use of shotgun sequencing revealed the functional redundancy between taxonomically dissimilar communities constituted of both bacteria and archaea along a salinity gradient [31]. WMGS also provides DNA sequences that are not targeted by *16S rDNA* amplification, including eukaryotic genomes, DNA viruses, and extrachromosomal DNA, such as plasmids. For example, in a study investigating the community composition of saltern ponds (Figure 1A) along a salinity gradient, the use of metagenomics allowed access to both the cellular and viral components of the community within the same sequencing datasets, revealing increased virus abundance at higher salt concentrations [32].

The reconstruction of viral genomes from hypersaline environments [33] using WMGS has resulted in improved characterization of this major component of halophilic microbiomes. Viruses take on the vital role of predators in many microbiomes and contribute to nutrient turnover with their lytic activity [34,35]. While nonshotgun approaches have been used previously to characterize halophilic metaviromes [36,37], high-throughput sequencing has empowered a more streamlined and unbiased recovery and annotation of viral sequences from various types of high-salt environments (Table 1). For example, an investigation of the metavirome in deep-sea haloclines (Figure 1E) through nontargeted shotgun sequencing revealed the stratification of virus lineages along the salinity gradient of the haloclines, likely associated with their host specificity [38]. In WMGS from solar salterns (Figure 1A), perfect alignments between the CRISPR spacers of microorganisms and viral sequences have been used together with di- and trinucleotide frequencies to predict and validate host specificity among halophilic phages across several locations [39]. Another study looking at halophilic *Cyanobacteria* in endolithic communities (Figure 1B) used virus sequences encoded in CRISPR arrays as a high-sensitivity strain signature, which allowed for the tracking of strain dispersal in the region [40].

As previously mentioned, one of the biggest strengths of WMGS is the ability to reconstruct the functional potential of a microbial community. With WMGS, hypersaline water [8,41], soil [4], and endolithic [42] microbiomes have been characterized in terms of their metabolic function, particularly their ability to use a wide range of energy sources. In particular, building on previous culture-dependent methods, systematic functional analysis of halophilic metagenomes has led to major improvements in our understanding of halophile osmotic adaptation and evolution [43]. For example, longitudinal analysis of halite endolith (Figure 1B) microbiota after a heavy rainfall revealed metaproteome adaptations to the temporarily decreased salt concentrations [42]. Functional annotation of longitudinal studies of halophiles from saltern, hypersaline lake, and salt mineral environments has also led to the characterization of horizontal gene transfers, evolutionary dynamics, and functional adaptations across time and space [41,42,44,45]. Functional potential profiling has also uncovered selective pressures and community functional dynamics that were not possible to investigate through taxonomy alone due to high functional redundancy. For example, the investigation of metagenomes from hypersaline soils (Figure 1F) has allowed researchers to uncover core differences in the functioning of their communities compared to more homogeneous aquatic hypersaline environments, which stems from nutrient scarcity, limited mobility, and niche stratification [4]. In a metagenomic study of phototropic hypersaline microbial mats (Figure 1C), functional annotation and pathway quantitation led to a better understanding of energy and nutrient capture and cycling between layers of the mats [46]. In particular, identification of MAGs with complementary parts of nitrogen and sulfur metabolism pathways suggested a dependence on the

metabolite exchange between community members. A functional potential investigation of microbial communities of solar saltern ponds (Figure 1A) revealed a higher prevalence of DNA replication and repair machinery in communities found in saturated brine compared to subsaturated saline environments [32]. With WMGS analysis rapidly improving and halophile databases rapidly growing [47], more breakthroughs will follow.

Another major aspect of metagenomics facilitated by WMGS is the reconstruction of novel individual genomes of halophiles. This is particularly important because extreme halophiles, and extremophiles in general, have been difficult to isolate due to specific growth condition requirements, symbiotic relationships, and cross-species functional pathways [48]. The binning of metagenomics assemblies has enabled researchers to recover hundreds of halophilic MAGs in the past decade [47], with many belonging to previously unknown orders, or even phyla [49]. For example, metagenomic binning of WMGS data from Lake Tyrel resulted in the recovery of near-complete genomes from a new clade of Nanohaloarchaea [50]. Similarly, metagenomic binning of solar saltern metagenomes uncovered several novel lineages of Euryarchaeota, Nanohaloarchaea, and Gammaproteobacteria. Functional annotation of these novel lineages allowed researchers to infer their metabolic functions within the microbiome [51]. In a halite endolith (Figure 1B) longitudinal study following a rare rain, community composition at the strain level was interrogated by genome-resolved metagenomics, leading to a general model of fine-scale taxonomic rearrangement of microbial communities following acute perturbations [42]. In addition to these individual discoveries, the rapidly increasing number of annotated reference halophile genomes allows for more accurate taxonomic and functional annotation in halophilic microbiomes, propelling the field in a positive-feedback loop [47].

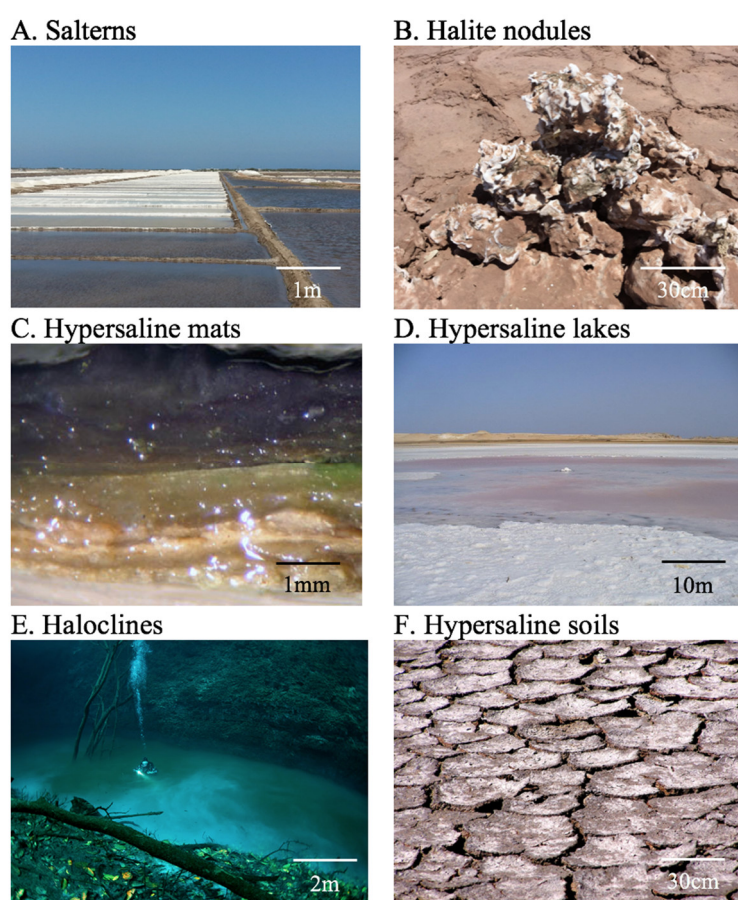


Figure 1. Photographs of commonly studied hypersaline environments: (A) saltern flats, (B) halite nodules, (C) hypersaline microbial mats, (D) hypersaline lakes, (E) underwater haloclines, and (F) hypersaline soils. * Sources for images (free-to-use sources):

https://commons.wikimedia.org/wiki/File:Salterns,_salt_making_fields,_tamil_nadu_-_panoramio.jpg,
https://en.wikipedia.org/wiki/Phototrophic_biofilm#/media/File:Microbial_mat_section.jpg,
https://commons.wikimedia.org/wiki/File:Saline_Lake_at_Ras_Mohamed_National_Park.jpg,
<https://commons.wikimedia.org/wiki/File:Halocline.png>, <https://pxhere.com/en/photo/1132612>.

Table 1. Studies that have contributed novel aspects of halophilic microbial communities through whole-metagenomic sequencing (WMGS) in hypersaline environments (list is not exhaustive). MAG: metagenome-assembled genome.

Environment	Longitudinal Dynamics	MAG Discovery	Functional Potential	Virus Analysis
Hypersaline lakes	Andrade [52], Tschitschko [45], Podell [53]	Narasimarao [50]	Vavourakis [54], Naghoni [31]	Emerson [55], Tschitschko [45], Ramos-Barbero [56]
Salterns	Plominsky [2]	Ramos-Barbero [57], Ghai [51]	Plominsky [32], Ghai [51]	Moller [39], Di Meglio [58]
Hypersaline microbial mats	Mobberley [46], Berlanga [59]	Mobberley [46]	Mobberley [46], Ruvindy [60], Wong [61]	White [62]
Haloclines	N/A	Speth [63]	Guan [64], Pachiadaki [65]	Antunes [38]
Halite endoliths	Uritskiy [42], Finstad [40]	Finstad [40], Uritskiy [42]	Crits-Christoph [66], Uritskiy [42]	Crits-Christoph [66]
Hypersaline soils	Narayan [67]	Vera-Gargallo [4]	Vera-Gargallo [4], Pandit [68]	NA

2.2. Limitations of Shotgun Metagenomics in Halophile Research

In contrast to human and synthetic microbiomes, the reconstruction of environmental metagenomes has been complicated by their sheer diversity and microdiversity. This is especially true in high-salt environments, which often host microbial communities with low taxonomic diversity but very high intraspecific diversity and characteristically high G + C content [69,70]. The presence of a large number of highly similar strains presents major challenges for deconvoluting their DNA content during metagenomic assembly and binning. This is particularly problematic in many halophiles that have genomic island regions of high inter-strain variability stemming from horizontal gene transfer [71,72]. On the other hand, the high G + C content of many dominant halophiles reduces the fraction of unique sequences in the samples [57,73], posing another challenge at the assembly stage. For example, halophilic endolith communities are typically dominated by *Halobacteria* and *Salinibacter*, but their high strain diversity and G + C content (over 60%) leads to relatively poor assembly and MAG quality [33]. In contrast, other community members that are less abundant and have low G + C content, such as *Cyanobacteria*, *Actinobacteria*, and *Gammaproteobacteria*, have yielded high-quality MAGs [42].

Due to the previously mentioned difficulties in culturing a diversity of halophiles, there are a relatively small number of genomes available. In 2018, there were just 942 complete halophile genomes available in NCBI databases [47], a tiny number in the era of high-throughput sequencing, which thus far has yielded over 200,000 prokaryotic complete genomes [74]. This leaves MAG extraction from environmental sequencing data the primary method for obtaining genomes of halophilic organisms, which has been difficult because of their metagenomic properties. In a negative feedback loop, this in turn has further stalled the progress of halophilic microbiome research, as the lack of available reference genomes has made taxonomic and functional annotation difficult. As WMGS becomes commonplace in microbiome research, it is crucial that the halophile field takes full advantage of the new technology and the use of newly available bioinformatic tools to further its understanding of microbial community assembly and function. Since 2014–2015, improvements in analytical methods and assembly software such as metaSPAdes [75], binning software such as

metaBAT2 [76], and processing pipelines such as metaWRAP [19] have allowed for effective deconvolution of WMGS data from even the most complex microbiomes. These new analytical methods will greatly benefit the halophile research field, if applied effectively.

3. Experimental Design Considerations for Sequencing Halophilic Metagenomes

Obtaining MAG-level resolution in a metagenome enables more accurate and meaningful functional pathway and taxonomic annotation and allows for detailed analysis of specific members of the community. With this in mind, the end goal of many microbiome studies is accurate and complete binning of sequence data. There are two general approaches to metagenomic sequencing and analysis for this purpose: (1) co-assembly of multiple shallowly sequenced samples or (2) individual processing of a few deeply sequenced samples. Both approaches have their benefits and limitations, depending on the microbiome that is sequenced and the biological question to answer.

In the first approach, samples are sequenced with relatively low-read coverage, and reads from all samples are combined during metagenomic assembly (Figure 2A). In research projects that demand a large number of samples, such as longitudinal studies, this results in low sequencing costs per sample, while also producing high-quality MAGs from the co-assembly by leveraging differential abundances of the contigs across samples [19,76]. The taxonomic and functional composition of individual samples can be investigated by linking the taxonomic and functional annotations of each contig with its abundance in each sample, allowing for easy comparisons between large numbers of samples [42,44]. Finally, co-assembling data from multiple samples enhances the recovery of genomes from low-abundance organisms, which is not possible from individual samples due to low coverage [50]. However, the use of co-assembly in metagenomics comes with significant drawbacks [57], including the high computational costs of co-assembling large data and the high level of microdiversity introduced by each new biological replicate. This latter point might be counterintuitive, but it leads to poor assemblies of very abundant taxa because accumulated mismatches from strain heterogeneity complicate the De Bruijn graph during assembly. This is particularly problematic with halophilic microbiomes, which are often dominated by highly diverse groups of Euryarchaeota and Bacteroidetes [49]. The high population microdiversity of these taxa is exacerbated when using multiple biological replicates, which results in poor, fragmented, or chimeric assemblies [57]. This in turn translates into poor-quality MAGs. However, when a broad capture of community diversity across many samples is the intent of the study, these limitations should then be considered in data interpretation.

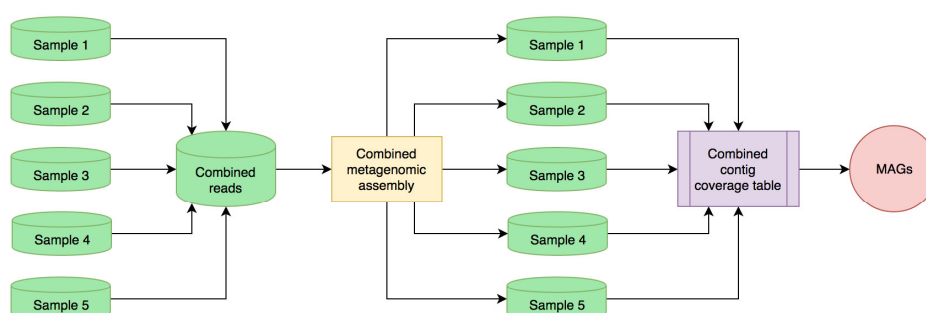
An alternative approach to co-assembly is to sequence a small number of samples with deep coverage and process them individually (Figure 2B). Because of the reduced microdiversity, individual assemblies produce larger contigs, given a comparable sequencing depth [77]. After binning each sample separately, MAGs can be combined into a single set through dereplication, removing duplicate MAGs that share a high nucleotide identity [78]. As with the co-assembly approach, differential contig coverage across samples may be used to improve binning results [40]. While this method is superior in highly heterogeneous communities such as halophilic microbiomes, it comes with a major increase in sequencing cost per sample. For most metagenomes, a meaningful assembly ($N50 > 5$ Kbp) requires 25–50 Gbp of sequencing data per sample, which limits the number of samples that can be multiplexed on a sequencing run. In turn, the limited replication reduces the effectiveness of binning, which leverages differential coverage of contigs across many samples to increase binning accuracy [79]. For many studies that require a large number of replicates, such as longitudinal studies, the cost of this approach may become prohibitively expensive.

An additional consideration in choosing a strategy for metagenomic sequencing and analysis is that of intersample community diversity. Communities in aquatic biomes, such as hypersaline lakes or brine ponds, are often more homogenous, harboring the same microorganisms with different relative abundances at different sampling locations. Under those conditions, a co-assembly strategy for metagenomics, as discussed above, is often preferred [44,50,80]. In contrast, in terrestrial microbiomes with limited dispersal, such as halite nodules in salars of the Atacama Desert, which contain unique taxonomic compositions, an individual assembly approach is more advantageous

[30,40]. Hybrid approaches are also possible in many cases, as binning of the individual and grouped assemblies may be combined and dereplicated to obtain the most robust MAGs of both rare and abundant species [81]. Regardless of the experimental design, it is critical to process samples, generate libraries, and sequence samples together to avoid batch effects [82]. If more than one flow cell is required to achieve the desired read depth, it is usually better to sequence the pooled libraries on several flow cells than to sequence each sample on its own flow cell [82]. For library preparation, it is recommended to use protocols that produce minimal G + C biases in coverage, particularly in halophilic communities that have high G + C content variation in their metagenomes [83,84].

The take-home message is that, when conducting a halophile metagenomic study, it is especially important to design a sampling and sequencing scheme with statistical questions in mind. Because of the high strain-level diversity typically found in halophilic microbiomes, an experimental design should avoid adding unnecessary replicates into the study, as each added biological replicate will introduce more microdiversity into the data, further complicating the assembly and binning stages of the analysis [57]. In practical terms, unless the intent of the study is to capture maximum diversity, the experimental design should include the minimum number of biological replicates that will allow for the intended statistical analysis downstream.

A. Co-assembly binning



B. Individual binning

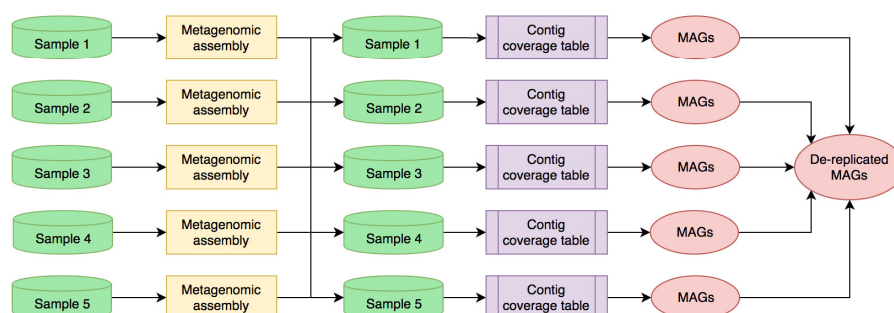


Figure 2. Flowcharts showing two common experimental designs and analysis workflows: (A) co-assembly and (B) individual sample processing and binning.

3.1. Best Bioinformatics Practices for Halophilic Metagenome Analysis

When processing halophilic metagenome sequencing data, it is important to adjust existing pipelines to accommodate for high intraspecific diversity, G + C content diversity, and underrepresentation in most sequence databases. While this section does not provide step-by-step instructions for bioinformatics analysis, it outlines core considerations and adjustments that should be made when processing halophilic metagenomes. Automated metagenomic analysis pipelines such as metaWRAP [19] or SqueezeM [85] may be used to streamline and simplify analysis: However, pipelines that are specifically designed for animal microbiomes, such as gut microbiota, should be

avoided. Indeed, these latter pipelines rely strongly on pre-existing taxonomic and functional databases of closely related organisms, as the majority of organisms found in host-associated microbiomes have been sequenced and characterized.

The preprocessing of WMGS data, which typically includes read trimming, duplicate read removal, and metagenomic assembly, is standard for most types of metagenomes. We encourage testing a variety of software and comparing the results to evaluation programs such as FastQC [86] (for read quality) and MetaQUAST [87] (for assembly quality), as some methods may be more suited to specific microbial community types [88]. For metagenomic assembly, metaSPAdes [75] is currently considered to be the best overall, while MegaHIT [89] is a better solution when resources are a limiting factor, as it is significantly faster and requires less memory [90]. Thanks to recent improvements in assembly software, it is no longer necessary to subsample reads during this stage, as contig quality no longer drops off with increased read depth [90]. However, higher-quality assemblies of abundant organisms can be achieved through individual or grouped sample assembly, as described above.

In contrast to assembly, the annotation of halophilic metagenomes for taxonomies and functions can be somewhat compromised because halophiles have extremely limited representation in standard-distribution taxonomic databases [91,92], which introduces significant biases in sequence annotation. As of 2018, there were only 942 published complete halophilic genomes available in NCBI [47], the main database used as a reference in most taxonomic and functional annotation software. Regarding methods for taxonomic profiling, general alignment-based methods such as MegaBLAST [93] are usually too specific for annotating non-assembled halophilic DNA sequences because they rely on high sequence similarity and skew the annotation toward taxa that are better-represented in the database. To produce more balanced taxonomic annotations given the limited databases, it is recommended to assign taxonomies to assembled contigs based on the genes that they carry and then infer the taxonomy of reads based on their alignment with the contigs. If the intent is to obtain the most accurate taxonomic distribution profile of the community, extracting and annotating marker genes (such as *16S rRNA* genes) with EMIRGE is usually the best alternative [94], as rRNA gene databases are more established and encompass greater taxonomic diversity [95].

Functional annotation (the functional categorization of genes) in halophile metagenomes is also severely limited by existing databases, especially compared to human microbiomes. Because many halophilic genes are not annotated in NCBI databases, metagenome-inclusive custom or specific databases are preferred because they contain a greater variety of noncultured organisms. In particular, services such as the “Integrated Microbial Genomes and Microbiomes” systems from JGI [96] include taxonomic and functional annotation models that are trained on user-submitted metagenomic data, including high-quality MAGs. The annotation sensitivity resulting from using the newest metagenomic data is extremely valuable for both functional and taxonomic annotation in relatively understudied systems, such as halophilic microbiomes. Regardless of the database being used, it is important to regularly update to the most recent release, as new organisms are constantly being sequenced. Annotation pipelines geared toward human microbiomes such as HUMANN2 [97] should be avoided, as they rely on the presence of closely related organisms in databases.

For many metagenomic studies, an important objective is the genome-resolved description of the microbiome of interest, since the analysis of individual MAGs opens up many avenues for more accurate and meaningful functional pathway annotations and strain-level comparative metagenomics. To that extent, the success of metagenomic binning of assemblies depends greatly on software choice, as binning programs perform differently with various data types [19]. Additionally, many popular binning software programs, such as metaBAT1, are trained on gut microbiome data [76], potentially limiting their efficacy in complex halophilic communities. Furthermore, benchmarking of such algorithms is often done on real or synthetic gut microbial communities [88]. Because of this, it is recommended to bin the metagenomic assembly with a variety of the most recent binning software, such as metaBAT2 [76] and CONCOCT [98], and to use a binning consolidation tool, such as metaWRAP or DAS_Tool, to produce the best final bin set [19,99]. When estimating the read coverage of the contigs in a given sample to be fed into the binning algorithms, it is important

to remember that they represent collapsed averages of a number of strains. Given the high intraspecific diversity of halophilic microbiomes [57], more accurate abundance estimations could potentially be obtained with slightly relaxed read alignment parameters, allowing for more approximate matches.

Considering the overwhelming number of metagenomic bioinformatics tools coming out each year, it is difficult to keep up to date with the best analytical methods. In general, we advise testing and benchmarking multiple software programs for each analytical step to determine the best option, as many conventionally used software programs behave unpredictably with halophilic sequence data. For annotation, emphasis should be placed on high sensitivity rather than high precision, given the database limitations.

3.2. The Future of Halophilic Metagenomics

Beyond shotgun sequencing of a microbiome's DNA content, there exist a number of other sequencing technologies that have become available and may further our understanding of halophilic ecosystems. Studies applying these technologies to more developed microbial fields, such as human gut microbiomes, have shown their great promise and their potential applications in halophilic microbial communities in the near future.

Conventional Illumina sequencing is limited to short DNA fragments (50 bp–250 bp), as errors accumulate rapidly at higher read lengths. However, read length, together with sequencing coverage, is undoubtedly a major limiting factor for metagenomics sequence assembly. Longer reads result in more accurate assembly and reduced chimeras, while they improve the contiguity of the assembly by allowing the assembly of repetitive DNA elements [100]. Recent sequencing technologies (minION from Oxford Nanopore and SMRT from PacBio sequencing) produce longer DNA fragments compared to Illumina. PacBio is able to consistently produce long reads (N50 up to 10 Kbp) with a relatively high degree of accuracy [101,102], while Nanopore sequencing produces even longer reads (N50 up to 100 Kbp), but with some sacrifices in accuracy [103,104]. Read lengths from these technologies enable not only the sequencing of complete ribosomal genes for improved taxonomic annotation, but also a significant improvement in the accuracy of metagenomics assembly and binning [102,105]. In highly diverse halophilic communities, long reads can help assemble ambiguous regions resulting from taxonomic heterogeneity, drastically improving the quality of the metagenome assembly [105]. Pseudo-single-cell technology from 10X Genomics, which tags each read with a barcode unique to the cell it came from, also shows great promise in halophilic microbiome deconvolution, as it is able to produce strain-specific synthetic long reads originating from single cells [106]. With reported maximum read lengths of over 1 Mbp from Nanopore, long-read technology is rapidly approaching the point where sequencing complete genomes in a single read is theoretically possible [107]. When this becomes reality, it will propel the field of metagenomics into a new post-assembly era. However, the recovery of less abundant taxa will remain a concern given the relatively low throughput of these methods.

Chromosome conformation capture with Hi-C is another technology that shows great promise in the field of halophilic metagenomics. A Hi-C assay crosslinks DNA based on spatial proximity: The chimeric segments resulting from the crosslink events are then sequenced, revealing sections of DNA that are proximal to each other. Conventionally used to indirectly measure the proximity between sections of a genome, Hi-C was successfully applied in 2017 to microbiomes to improve binning predictions [108]. Considering the difficulty of binning halophilic metagenomes due to their heterogeneity, Hi-C could significantly improve halophile MAG extraction. Hi-C-based binning also enables the recovery of extrachromosomal elements such as viral and plasmid DNA, which so far has been difficult to accomplish [109]. Hi-C can also be used to produce DNA proximity maps in individual MAGs for the study of chromatin conformation in prokaryotes at the metagenomic and single-cell scale [109].

Finally, genome-resolved metatranscriptomics (the analysis of a microbial community's RNA content) has been widely used in a variety of microbiomes to interrogate microbial transcriptional activities [26,110]. Metatranscriptomics has been used in halophile research to characterize carbon

cycling in saline soils [111] and has been extensively used to characterize activity in other soil microbiomes [112,113]. However, it remains a largely underdeployed tool in many other high-salt systems, partly due to the difficulty in depleting ribosomal sequences in archaeal RNA. Another major deterrent has been the difficulty in standardizing transcript expression to the abundance of each individual organism in a sample. In other words, if a transcript is more abundant in a given sample, it can be difficult to determine if the organism carrying it is more abundant in that sample, or if it is truly highly expressed. However, with rapid improvements in genome-resolved metagenomic analysis of halophile communities, it is possible that the metatranscriptomic problem can be simplified down to more conventional transcriptome analysis by investigating the transcriptomes of individual MAGs.

4. Conclusion

Successful applications of whole metagenomics in halophilic communities has already led to numerous breakthroughs in our understanding of their functional composition, virus–host interactions, and strain diversity and dispersal, and has allowed for the genome extraction of previously unknown halophiles. However, the genomic qualities and composition characteristics of halophilic communities have made them difficult to deconvolute in a metagenomic context, limiting the information that can be extracted from halophilic shotgun metagenomes. Combined with relatively low numbers of cultures of halophiles, this has led to their underrepresentation in existing taxonomical and functional databases, which has further complicated analysis. While in silico deconvolution of halophilic metagenomes is a challenge, it can be accomplished with analysis workflows that account for the specific characteristics of halophile communities. With proper tuning, rapidly advancing sequencing technology has the potential to reconstruct the complete nucleic acid content of halophilic communities, allowing the halophile field to focus on microbial functional activity and interactions.

Acknowledgments: This work was supported by grants NNX15AP18G from NASA and DEB1556574 from the NSF. We thank Sarah Preheim for useful suggestions in the writing of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Graham, E.B.; Knelman, J.E.; Schindlbacher, A.; Siciliano, S.; Breulmann, M.; Yannarell, A.; Beman, J.M.; Abell, G.; Philippot, L.; Prosser, J.; et al. Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes?. *Front. Microbiol.* **2016**, *7*, 111.
2. Kallmeyer, J.; Pockalny, R.; Adhikari, R.R.; Smith, D.C.; D'Hondt, S. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci USA* **2012**, *109*, 16213–16216.
3. Whitman, W.B.; Coleman, D.C.; Wiebe, W.J. Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* **1998**, *95*, 6578–6583.
4. Vera-Gargallo, B.; Ventosa, A. Metagenomic Insights into the Phylogenetic and Metabolic Diversity of the Prokaryotic Community Dwelling in Hypersaline Soils from the Odiel Saltmarshes (SW Spain). *Genes* **2018**, *9*, 152.
5. Gibtan, A.; Park, K.; Woo, M.; Shin, J.-K.; Lee, D.-W.; Sohn, J.H.; Song, M.; Roh, S.W.; Lee, S.-J.; Lee, H.-S. Diversity of Extremely Halophilic Archaeal and Bacterial Communities from Commercial Salts. *Front. Microbiol.* **2017**, *8*, 631.
6. Henriot, O.; Fourmentin, J.; Delincé, B.; Mahillon, J. Exploring the diversity of extremely halophilic archaea in food-grade salts. *International Journal of Food Microbiology* **2014**, *191*, 36–44.
7. Seck, E.H.; Dufour, J.-C.; Raoult, D.; Lagier, J.-C. Halophilic & halotolerant prokaryotes in humans. *Future Microbiology* **2018**, *13*, 799–812.
8. Oren, A. Halophilic archaea on Earth and in space: Growth and survival under extreme conditions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2014**, *372*, 20140194.
9. Ma, Y.; Galinski, E.A.; Grant, W.D.; Oren, A.; Ventosa, A. Halophiles 2010: Life in Saline Environments. *Applied and Environmental Microbiology* **2010**, *76*, 6971–6981.

10. Rinke, C.; Schwientek, P.; Sczyrba, A.; Ivanova, N.N.; Anderson, I.J.; Cheng, J.-F.; Darling, A.; Malfatti, S.; Swan, B.K.; Gies, E.A.; et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **2013**, *499*, 431–437.
11. Hedlund, B.P.; Dodsworth, J.A.; Murugapiran, S.K.; Rinke, C.; Woyke, T. Impact of single-cell genomics and metagenomics on the emerging view of extremophile “microbial dark matter.” *Extremophiles* **2014**, *18*, 865–875.
12. Riesenfeld, C.S.; Schloss, P.D.; Handelsman, J. METAGENOMICS: Genomic Analysis of Microbial Communities. *Annu. Rev. Genet.* **2004**, *38*, 525–552.
13. Ranjan, R.; Rani, A.; Metwally, A.; McGee, H.S.; Perkins, D.L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications* **2015**, *469*, 967–977.
14. Tessler, M.; Neumann, J.S.; Afshinnikoo, E.; Pineda, M.; Hersch, R.; Velho, L.F.M.; Segovia, B.T.; Lansac-Toha, F.A.; Lemke, M.; DeSalle, R.; et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci Rep.* **2017**, *7*, 6589.
15. Riesenfeld, C.S.; Schloss, P.D.; Handelsman, J. METAGENOMICS: Genomic Analysis of Microbial Communities. *Annu. Rev. Genet.* **2004**, *38*, 525–552.
16. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Corrigendum: Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **2017**, *35*, 1211.
17. Ghurye, J.S.; Cepeda-Espinoza, V.; Pop, M. Metagenomic Assembly: Overview, Challenges and Applications. *Yale J. Biol. Med.* **2016**, *89*, 353–362.
18. Olson, N.D.; Treangen, T.J.; Hill, C.M.; Cepeda-Espinoza, V.; Ghurye, J.; Koren, S.; Pop, M. Metagenomic assembly through the lens of validation: Recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Brief Bioinform.* **2017**, doi:10.1093/bib/bbx098.
19. Uritskiy, G.V.; DiRuggiero, J.; Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **2018**, *6*, 158.
20. Tully, B.J.; Graham, E.D.; Heidelberg, J.F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **2018**, *5*, 170203.
21. Sangwan, N.; Xia, F.; Gilbert, J.A. Recovering complete and draft population genomes from metagenome datasets. *Microbiome* **2016**, *4*, 197.
22. Sims, D.; Sudbery, I.; Iltott, N.E.; Heger, A.; Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* **2014**, *15*, 121–132.
23. Zaheer, R.; Noyes, N.; Polo, R.O.; Cook, S.R.; Marinier, E.; Van Domselaar, G.; Belk, K.E.; Morley, P.S.; McAllister, T.A. Impact of sequencing depth on the characterization of the microbiome and resistome. *Sci Rep.* **2018**, *8*, 5890.
24. Sharifi, F.; Ye, Y. From Gene Annotation to Function Prediction for Metagenomics. *Alcohol* **2017**, *1611*, 27–34.
25. Zhang, Y.; Wang, J.; Qi, J.; Zhao, F.; He, S.; Wei, S. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Sci Rep.* **2013**, *3*, 1843.
26. Wang, W.-L.; Xu, S.-Y.; Ren, Z.-G.; Tao, L.; Jiang, J.-W.; Zheng, S.-S. Application of metagenomics in the human gut microbiome. *WJG* **2015**, *21*, 803–814.
27. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **2017**, *35*, 833–844.
28. Poretzky, R.; Rodríguez-R, L.M.; Luo, C.; Tsementzi, D.; Konstantinidis, K.T. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE* **2014**, *9*, e93827.
29. White, J.R.; Nagarajan, N.; Pop, M. Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. *PLoS Comput Biol* **2009**, *5*, e1000352.
30. Crits-Christoph, A.; Gelsinger, D.R.; Ma, B.; Wierzbos, J.; Ravel, J.; Davila, A.; Casero, M.C.; DiRuggiero, J. Functional interactions of archaea, bacteria, and viruses in a hypersaline endolithic community. *Environmental Microbiology* **2016**, *18*, 2064–2077.
31. Naghoni, A.; Emtiazi, G.; Amoozegar, M.A.; Cretoiu, M.S.; Stal, L.J.; Etemadifar, Z.; Fazeli, S.A.S.; Bolhuis, H. Microbial diversity in the hypersaline Lake Meyghan, Iran. *Sci Rep.* **2017**, *7*, 11522.
32. Plominsky, A.M.; Henríquez-Castillo, C.; Delherbe, N.; Podell, S.; Ramirez-Flandes, S.; Ugalde, J.A.; Santibañez, J.F.; Engh, G.V.D.; Hanselmann, K.; Ulloa, O.; et al. Distinctive Archaeal Composition of an

- Artisanal Crystallizer Pond and Functional Insights Into Salt-Saturated Hypersaline Environment Adaptation. *Front. Microbiol.* **2018**, *9*, 1800.
33. Roux, S.; Enault, F.; Ravet, V.; Colombet, J.; Bettarel, Y.; Auguet, J.; Bouvier, T.; Lucas-Staat, S.; Vellet, A.; Prangishvili, D.; et al. Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environmental Microbiology* **2015**, *18*, 889–903.
 34. Pedrós-Alió, C.; Calderón-Paz, J.I.; MacLean, M.H.; Medina, G.; Marrasé, C.; Gasol, J.M.; Guixa-Boixereu, N. The microbial food web along salinity gradients. *FEMS Microbiology Ecology* **2000**, *32*, 143–155.
 35. Guixa-Boixereu, N.; Calderón-Paz, J.; Heldal, M.; Bratbak, G.; Pedrós-Alió, C. Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat. Microb. Ecol.* **1996**, *11*, 215–227.
 36. Santos, F.; Meyerdierks, A.; Peña, A.; Amann, R.; Antón, J.; Rosselló-Mora, R.; Rosselló-Mora, R.; Rosselló-Móra, R. Metagenomic approach to the study of halophages: The environmental halophage 1. *Environmental Microbiology* **2007**, *9*, 1711–1723.
 37. Santos, F.; Yarza, P.; Parro, V.; Briones, C.; Antón, J. The metavirome of a hypersaline environment. *Environ Microbiol* **2010**, *12*, 2965–2976.
 38. Antunes, A.; Alam, I.; Simões, M.F.; Daniels, C.; Ferreira, A.J.; Siam, R.; El-Dorri, H.; Bajic, V.B. First Insights into the Viral Communities of the Deep-sea Anoxic Brines of the Red Sea. *Genomics, Proteomics & Bioinformatics* **2015**, *13*, 304–309.
 39. Moller, A.G.; Liang, C. Determining virus-host interactions and glycerol metabolism profiles in geographically diverse solar salterns with metagenomics. *PeerJ* **2017**, *5*, e2844.
 40. Finstad, K.M.; Probst, A.J.; Thomas, B.C.; Andersen, G.L.; Demergasso, C.; Echeverría, A.; Amundson, R.G.; Banfield, J.F. Microbial Community Structure and the Persistence of Cyanobacterial Populations in Salt Crusts of the Hyperarid Atacama Desert from Genome-Resolved Metagenomics. *Front. Microbiol.* **2017**, *8*, 1435.
 41. Kimbrel, J.A.; Ballor, N.; Wu, Y.-W.; David, M.M.; Hazen, T.C.; Simmons, B.A.; Singer, S.W.; Jansson, J.K. Microbial Community Structure and Functional Potential Along a Hypersaline Gradient. *Front. Microbiol.* **2018**, *9*, 1492.
 42. Uritskiy, G.; Getsin, S.; Munn, A.; Gomez-Silva, B.; Davila, A.; Glass, B.; Taylor, J.; DiRuggiero, J. Response of extremophile microbiome to a rare rainfall reveals a two-step adaptation mechanism. *bioRxiv* **2018**, *bioRxiv*, 442525.
 43. Becker, E.A.; Seitzer, P.M.; Tritt, A.; Larsen, D.; Krusor, M.; Yao, A.I.; Wu, D.; Madern, D.; Eisen, J.A.; Darling, A.E.; et al. Phylogenetically Driven Sequencing of Extremely Halophilic Archaea Reveals Strategies for Static and Dynamic Osmo-response. *PLoS Genet.* **2014**, *10*, e1004784.
 44. DeMaere, M.Z.; Williams, T.J.; Allen, M.A.; Brown, M.V.; Gibson, J.A.E.; Rich, J.; Lauro, F.M.; Dyall-Smith, M.; Davenport, K.W.; Woyke, T.; et al. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl Acad Sci USA* **2013**, *110*, 16939–16944.
 45. Tschitschko, B.; Erdmann, S.; Roux, S.; Panwar, P.; Brazendale, S.; Cavicchioli, R.; DeMaere, M.Z.; A Allen, M.; Williams, T.J.; Hancock, A.M.; et al. Genomic variation and biogeography of Antarctic haloarchaea. *Microbiome* **2018**, *6*, 113.
 46. Mobberley, J.M.; Lindemann, S.R.; Bernstein, H.C.; Moran, J.J.; Renslow, R.S.; Babauta, J.; Hu, D.; Beyenal, H.; Nelson, W.C. Organismal and spatial partitioning of energy and macronutrient transformations within a hypersaline mat. *FEMS Microbiology Ecology* **2017**, *93*.
 47. Loukas, A.; Kappas, I.; Abatzopoulos, T.J. HaloDom: A new database of halophiles across all life domains. *J. of Biol Res.-Thessaloniki* **2018**, *25*, 2.
 48. Solden, L.; Lloyd, K.; Wrighton, K. The bright side of microbial dark matter: Lessons learned from the uncultivated majority. *Current Opinion in Microbiology* **2016**, *31*, 217–226.
 49. Ventosa, A.; De La Haba, R.R.; Sánchez-Porro, C.; Papke, R.T. Microbial diversity of hypersaline environments: A metagenomic approach. *Current Opinion in Microbiology* **2015**, *25*, 80–87.
 50. Narasingarao, P.; Podell, S.; A Ugalde, J.; Brochier-Armanet, C.; Emerson, J.B.; Brocks, J.J.; Heidelberg, K.B.; Banfield, J.F.; E Allen, E. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **2011**, *6*, 81–93.
 51. Pašić, L.; Fernández, A.B.; Martín-Cuadrado, A.-B.; Papke, R.T.; Rodríguez-Brito, B.; Ghai, R.; Mizuno, C.M.; McMahon, K.D.; Stepanauskas, R.; Rohwer, F.; et al. New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Sci Rep.* **2011**, *1*, 135.

52. Andrade, K.; Logemann, J.; Heidelberg, K.B.; Emerson, J.B.; Comolli, L.R.; A Hug, L.; Probst, A.J.; Keillar, A.; Thomas, B.C.; Miller, C.S.; et al. Metagenomic and lipid analyses reveal a diel cycle in a hypersaline microbial ecosystem. *ISME J.* **2015**, *9*, 2697–2711.
53. Podell, S.; Ugalde, J.A.; Narasingarao, P.; Banfield, J.F.; Heidelberg, K.B.; Allen, E.E. Assembly-Driven Community Genomics of a Hypersaline Microbial Ecosystem. *PLoS ONE* **2013**, *8*, e61692.
54. Vavourakis, C.D.; Andrei, A.-S.; Mehrshad, M.; Ghai, R.; Sorokin, D.Y.; Muyzer, G. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* **2018**, *6*, 168.
55. Emerson, J.B.; Andrade, K.; Thomas, B.C.; Norman, A.; Allen, E.E.; Heidelberg, K.B.; Banfield, J.F. Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013**, 370871.
56. Ramos-Barbero, M.D.; Martínez, J.M.; Almansa, C.; Rodríguez, N.; Villamor, J.; Gomariz, M.; Escudero, C.; Rubin, S.; Antón, J.; Martínez-García, M.; et al. Prokaryotic and viral community structure in the singular chaotropic salt lake salar de uyuni. *Environmental Microbiology* **2019**, doi:10.1111/1462-2920.14549.
57. Ramos-Barbero, M.D.; Martin-Cuadrado, A.-B.; Viver, T.; Santos, F.; Martinez-Garcia, M.; Antón, J. Recovering microbial genomes from metagenomes in hypersaline environments: The Good, the Bad and the Ugly. *Systematic and Applied Microbiology* **2019**, *42*, 30–40.
58. Di Meglio, L.; Santos, F.; Gomariz, M.; Almansa, C.; López, C.; Antón, J.; Nercessian, D. Seasonal dynamics of extremely halophilic microbial communities in three Argentinian salterns. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw184.
59. Berlanga, M.; Palau, M.; Guerrero, R. Functional Stability and Community Dynamics during Spring and Autumn Seasons Over 3 Years in Camargue Microbial Mats. *Front. Microbiol.* **2017**, *8*, 2619.
60. Ruvindy, R.; White, 3rd, R.A.; Neilan, B.A.; Burns, B.P. Unravelling core microbial metabolisms in the hypersaline microbial mats of Shark Bay using high-throughput metagenomics. *ISME J.* **2016**, *10*, 183–196.
61. Wong, H.L.; White, R.A.; Visscher, P.T.; Charlesworth, J.C.; Vázquez-Campos, X.; Burns, B.P. Disentangling the drivers of functional complexity at the metagenomic level in Shark Bay microbial mat microbiomes. *ISME J.* **2018**, *12*, 2619–2639.
62. White, R.A.; Wong, H.L.; Ruvindy, R.; Neilan, B.A.; Burns, B.P.; Iii, R.A.W. Viral Communities of Shark Bay Modern Stromatolites. *Front. Microbiol.* **2018**, *9*, 1223.
63. Speth, D.R.; Lagkouravdos, I.; Wang, Y.; Qian, P.-Y.; Dutilh, B.E.; Jetten, M.S.M. Draft Genome of *Scalindua rubra*, Obtained from the Interface Above the Discovery Deep Brine in the Red Sea, Sheds Light on Potential Salt Adaptation Strategies in Anammox Bacteria. *Microb Ecol* **2017**, *74*, 1–5.
64. Guan, Y.; Hikmawan, T.; Antunes, A.; Ngugi, D.; Stingl, U. Diversity of methanogens and sulfate-reducing bacteria in the interfaces of five deep-sea anoxic brines of the Red Sea. *Research in Microbiology* **2015**, *166*, 688–699.
65. Pachiadaki, M.G.; Yakimov, M.M.; Lacono, V.; Leadbetter, E.; Edgcomb, V. Unveiling microbial activities along the halocline of Thetis, a deep-sea hypersaline anoxic basin. *ISME J.* **2014**, *8*, 2478–2489.
66. Crits-Christoph, A.; Robinson, C.K.; Ma, B.; Ravel, J.; Wierzbos, J.; Ascaso, C.; Artieda, O.; Souza-Egipsy, V.; Casero, M.C.; DiRuggiero, J. Phylogenetic and Functional Substrate Specificity for Endolithic Microbial Communities in Hyper-Arid Environments. *Front. Microbiol.* **2016**, *7*, 301.
67. Narayan, A.; Patel, V.; Singh, P.; Patel, A.; Jain, K.; Karthikeyan, K.; Shah, A.; Madamwar, D. Response of microbial community structure to seasonal fluctuation on soils of Rann of Kachchh, Gujarat, India: Representing microbial dynamics and functional potential. *Ecological Genetics and Genomics* **2018**, *6*, 22–32.
68. Pandit, A.S.; Joshi, M.N.; Bhargava, P.; Shaikh, I.; Ayachit, G.N.; Raj, S.R.; Saxena, A.K.; Bagatharia, S.B. A snapshot of microbial communities from the Kutch: One of the largest salt deserts in the World. *Extremophiles* **2015**, *19*, 973–987.
69. Cuadros-Orellana, S.; Martin-Cuadrado, A.-B.; Legault, B.; D'Auria, G.; Zhaxybayeva, O.; Papke, R.T.; Rodriguez-Valera, F. Genomic plasticity in prokaryotes: The case of the square haloarchaeon. *ISME J.* **2007**, *1*, 235–245.
70. Papke, R.T.; Koenig, J.E.; Rodriguez-Valera, F.; Doolittle, W.F. Frequent Recombination in a Saltern Population of *Halorubrum*. *Science* **2004**, *306*, 1928–1929.
71. Pašić, L.; Rodriguez-Mueller, B.; Martin-Cuadrado, A.-B.; Rodriguez-Valera, F.; Mira, A.; Rohwer, F. Metagenomic islands of hyperhalophiles: The case of *Salinibacter ruber*. *BMC Genomics* **2009**, *10*, 570.
72. Martin-Cuadrado, A.-B.; Pašić, L.; Rodriguez-Valera, F. Diversity of the cell-wall associated genomic island of the archaeon *Haloquadratum walsbyi*. *BMC Genomics* **2015**, *16*, 13950.

73. Chen, Y.-C.; Liu, T.; Yu, C.-H.; Chiang, T.-Y.; Hwang, C.-C. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS ONE* **2013**, *8*, e62856.
74. Haft, D.H.; DiCuccio, M.; Badretdin, A.; Brover, V.; Chetvernin, V.; O'Neill, K.; Li, W.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; et al. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **2017**, *46*, D851–D860.
75. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834.
76. Kang, D.D.; Froula, J.; Egan, R.; Wang, Z.; Rahmann, S. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, *3*, 1165.
77. Haro-Moreno, J.M.; López-Pérez, M.; De La Torre, J.R.; Picazo, A.; Camacho, A.; Rodríguez-Valera, F. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* **2018**, *6*, 128.
78. Olm, M.R.; Brown, C.T.; Brooks, B.; Banfield, J.F. dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **2017**, *11*, 2864–2868.
79. Goodrich, J.K.; Di Rienzi, S.C.; Poole, A.C.; Koren, O.; Walters, W.A.; Caporaso, J.G.; Knight, R.; Ley, R.E. Conducting a microbiome study. *Cell* **2014**, *158*, 250–262.
80. Vavourakis, C.D.; Ghai, R.; Rodríguez-Valera, F.; Sorokin, D.Y.; Tringe, S.G.; Hugenholtz, P.; Muyzer, G. Metagenomic Insights into the Uncultured Diversity and Physiology of Microbes in Four Hypersaline Soda Lake Brines. *Front. Microbiol.* **2016**, *7*, 533.
81. Stewart, R.D.; Auffret, M.D.; Warr, A.; Wiser, A.H.; Press, M.O.; Langford, K.W.; Liachko, I.; Snelling, T.J.; Dewhurst, R.J.; Walker, A.W.; et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun* **2018**, *9*, 870.
82. Gibbons, S.M.; Duvallet, C.; Alm, E.J. Correcting for batch effects in case-control microbiome studies. *PLoS Comput Biol* **2018**, *14*, e1006102.
83. Paul, S.; Bag, S.K.; Das, S.; Harvill, E.T.; Dutta, C. Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* **2008**, *9*.
84. Jones, M.B.; Highlander, S.K.; Anderson, E.L.; Li, W.; Dayrit, M.; Klitgord, N.; Fabani, M.M.; Seguritan, V.; Green, J.; Pride, D.T.; et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc. Natl Acad Sci USA* **2015**, *112*, 14024–14029.
85. Tamames, J.; Puente-Sanchez, F. SqueezeM, a highly portable, fully automatic metagenomic analysis pipeline. *Bioinformatics* **2018**, *bioRxiv*, 347559.
86. Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **2017**, *33*, 3137–3139.
87. Mikheenko, A.; Saveliev, V.; Gurevich, A. MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics* **2015**, *32*, 1088–1090.
88. Sczyrba, A.; Hofmann, P.; Belmann, P.; Koslicki, D.; Janssen, S.; Dröge, J.; Gregor, I.; Majda, S.; Fiedler, J.; Dahms, E.; et al. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nat. Methods* **2017**, *14*, 1063–1071.
89. Li, D.; Luo, R.; Liu, C.-M.; Leung, C.-M.; Ting, H.-F.; Sadakane, K.; Yamashita, H.; Lam, T.-W. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **2016**, *102*, 3–11.
90. Vollmers, J.; Wiegand, S.; Kaster, A.-K. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters!. *PLoS ONE* **2017**, *12*, e0169662.
91. Wheeler, D.L. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2001**, *29*, 11–16.
92. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **2015**, *44*, D733–D745.
93. Chen, Y.; Ye, W.; Zhang, Y.; Xu, Y. High speed BLASTN: An accelerated MegaBLAST search tool. *Nucleic Acids Research* **2015**, *43*, 7762–7768.
94. Miller, C.S.; Baker, B.J.; Thomas, B.C.; Singer, S.W.; Banfield, J.F. EMIRGE: Reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol* **2011**, *12*, R44.

95. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glockner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41*, D590–6.
96. Chen, I.A.; Markowitz, V.M.; Chu, K.; Palaniappan, K.; Szeto, E.; Pillay, M.; Ratner, A.; Huang, J.; Andersen, E.; Huntemann, M.; et al. IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **2017**, *45*, D507–D516.
97. Abubucker, S.; Segata, N.; Goll, J.; Schubert, A.M.; Izard, J.; Cantarel, B.L.; Rodriguez-Mueller, B.; Zucker, J.; Thiagarajan, M.; Henrissat, B.; et al. Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Comput Biol* **2012**, *8*, e1002358.
98. Alneberg, J.; Bjarnason, B.S.; De Bruijn, I.; Schirmer, M.; Quick, J.; Ijaz, U.Z.; Lahti, L.; Loman, N.J.; Andersson, A.F.; Quince, C. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **2014**, *11*, 1144–1146.
99. Sieber, C.M.K.; Probst, A.J.; Sharrar, A.; Thomas, B.C.; Hess, M.; Tringe, S.G.; Banfield, J.F. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol* **2018**, *3*, 836–843.
100. Wommack, K.E.; Bhavsar, J.; Ravel, J. Metagenomics: Read length matters. *Appl. Environ. Microbiol.* **2008**, *74*, 1453–1463.
101. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **2015**, *13*, 278–289.
102. Frank, J.A.; Pan, Y.; Tooming-Klunderud, A.; Eijsink, V.G.H.; McHardy, A.C.; Nederbragt, A.J.; Pope, P.B.; Nederbragt, A.; Pope, P. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. *Sci Rep.* **2016**, *6*, 25373.
103. Brown, B.L.; Watson, M.; Minot, S.S.; Rivera, M.C.; Franklin, R.B. MinION™ nanopore sequencing of environmental metagenomes: A synthetic approach. *GigaScience* **2017**, *6*, 1–10.
104. Rang, F.J.; Kloosterman, W.P.; De Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol* **2018**, *19*, 90.
105. Driscoll, C.B.; Otten, T.G.; Brown, N.M.; Dreher, T.W. Towards long-read metagenomics: Complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. *Stand. in Genomic Sci* **2017**, *12*, 9.
106. Moss, E.; Bishara, A.; Tkachenko, E.; Kang, J.B.; Andermann, T.M.; Wood, C.; Handy, C.; Ji, H.; Batzoglou, S.; Bhatt, A.S. De novo assembly of microbial genomes from human gut metagenomes using barcoded short read sequences. *bioRxiv* **2017**, doi:10.1101/125211.
107. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; A Sasani, T.; Tyson, J.R.; Beggs, A.D.; Dillthey, A.T.; Fiddes, I.T.; et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol* **2018**, *36*, 338–345.
108. Press, M.O.; Wiser, A.H.; Kronenberg, Z.N.; Langford, K.W.; Shakya, M.; Lo, C.-C.; Mueller, K.A.; Sullivan, S.T.; Chain, P.S.G.; Liachko, I. Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *Genomics* **2017**, *bioRxiv*, 198713.
109. Burton, J.N.; Liachko, I.; Dunham, M.J.; Shendure, J. Species-Level Deconvolution of Metagenome Assemblies with Hi-C–Based Contact Probability Maps. *G3* **2014**, *4*, 1339–1346.
110. Lavelle, A.; Sokol, H. Gut microbiota: Beyond metagenomics, metatranscriptomics illuminates microbiome functionality in IBD. *Nat. Rev. Gastroenterol Hepatol* **2018**, *15*, 193–194.
111. Ren, M.; Zhang, Z.; Wang, X.; Zhou, Z.; Chen, D.; Zeng, H.; Zhao, S.; Chen, L.; Hu, Y.; Zhang, C.; et al. Diversity and Contributions to Nitrogen Cycling and Carbon Fixation of Soil Salinity Shaped Microbial Communities in Tarim Basin. *Front. Microbiol.* **2018**, *9*, 431.
112. Garoutte, A.; Cardenas, E.; Tiedje, J.; Howe, A. Methodologies for probing the metatranscriptome of grassland soil. *Journal of Microbiological Methods* **2016**, *131*, 122–129.
113. Jiang, Y.; Xiong, X.; Danska, J.; Parkinson, J. Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* **2016**, *4*, 13780.

