

Exploring Symbioses by Single-Cell Genomics

JANINE KAMKE¹, KRISTINA BAYER¹, TANJA WOYKE², AND UTE HENTSCHEL^{1,*}

¹*Julius-von-Sachs Institute for Biological Sciences, University of Würzburg, Julius-von-Sachs Platz 3, 97082 Würzburg, Germany; and* ²*Department of Energy Joint Genome Institute, Walnut Creek, California*

Abstract. Single-cell genomics has advanced the field of microbiology from the analysis of microbial metagenomes where information is “drowning in a sea of sequences,” to recognizing each microbial cell as a separate and unique entity. Single-cell genomics employs Phi29 polymerase-mediated whole-genome amplification to yield microgram-range genomic DNA from single microbial cells. This method has now been applied to a handful of symbiotic systems, including bacterial symbionts of marine sponges, insects (grasshoppers, termites), and vertebrates (mouse, human). In each case, novel insights were obtained into the functional genomic repertoire of the bacterial partner, which, in turn, led to an improved understanding of the corresponding host. Single-cell genomics is particularly valuable when dealing with uncultivated microorganisms, as is still the case for many bacterial symbionts. In this review, we explore the power of single-cell genomics for symbiosis research and highlight recent insights into the symbiotic systems that were obtained by this approach.

Background

The fields of metagenomics, metatranscriptomics, and metaproteomics, among others collectively referred to as “omics,” have made a tremendous impact on symbiosis research (see other articles in this special issue). For many decades, symbiosis research was possible solely by descriptive approaches because neither could the symbionts be cultured (largely unchanged to this day) nor was there experimental access to many symbiotic systems (also largely unchanged). The implementation of cultivation-independent approaches based on 16S rRNA gene se-

quences thus initiated a major revolution by making it possible to place organisms that were frequently known only by electron microscopy into a phylogenetic context. 16S rRNA gene phylogenies further helped to delineate co-evolution and co-speciation events by comparing host and symbiont phylogenies.

The implementation of omics methods spurred a second wave of information in symbiosis research as it became possible to predict the genomic underpinnings of symbioses. For example, the discovery that many insects had genomically encoded nutritional interdependencies on their symbiotic bacteria represented a milestone discovery (*i.e.*, Gil *et al.*, 2003; Wu *et al.*, 2006; Moran *et al.*, 2008; Wilson *et al.*, 2010). Similarly, metagenomics provided novel insights into chemoautotrophic symbioses, in that sulfur-oxidizing and sulfate-reducing symbionts provide a gutless marine worm host with multiple sources of nutrition (Woyke *et al.*, 2006). Moreover, omics approaches have provided the first glimpse into the functional gene repertoire of marine sponges and their beneficial microbial consortia (Hallam *et al.*, 2006, Thomas *et al.*, 2010; Liu, M., *et al.*, 2011; Liu *et al.*, 2012; Fan *et al.*, 2012).

One recent addition to the omics repertoire is single-cell genomics. It relies on genomic sequence information from individual microbial cells and is entirely cultivation-independent. By use of Phi29 polymerase it is possible to obtain comprehensive genomic information from individual microbial cells—something that to our knowledge is not possible with any other technique to date (Hutchison and Venter, 2006; Binga *et al.*, 2008; Ishoey *et al.*, 2008). Single-cell genomics is especially well suited for symbiosis research in which the vast majority of symbionts have not been cultivated and are thus not accessible by conventional techniques. Here we present a brief overview of the methodology and its current limitations and challenges. We then review the current state of single-cell genomics techniques in symbiosis research using five recently published exam-

Received 11 April 2012; accepted 19 June 2012.

* Author to whom correspondence should be addressed: E-mail: ute.hentschel@uni-wuerzburg.de

Abbreviations: MDA, multiple displacement amplification; SAG, single amplified genome.

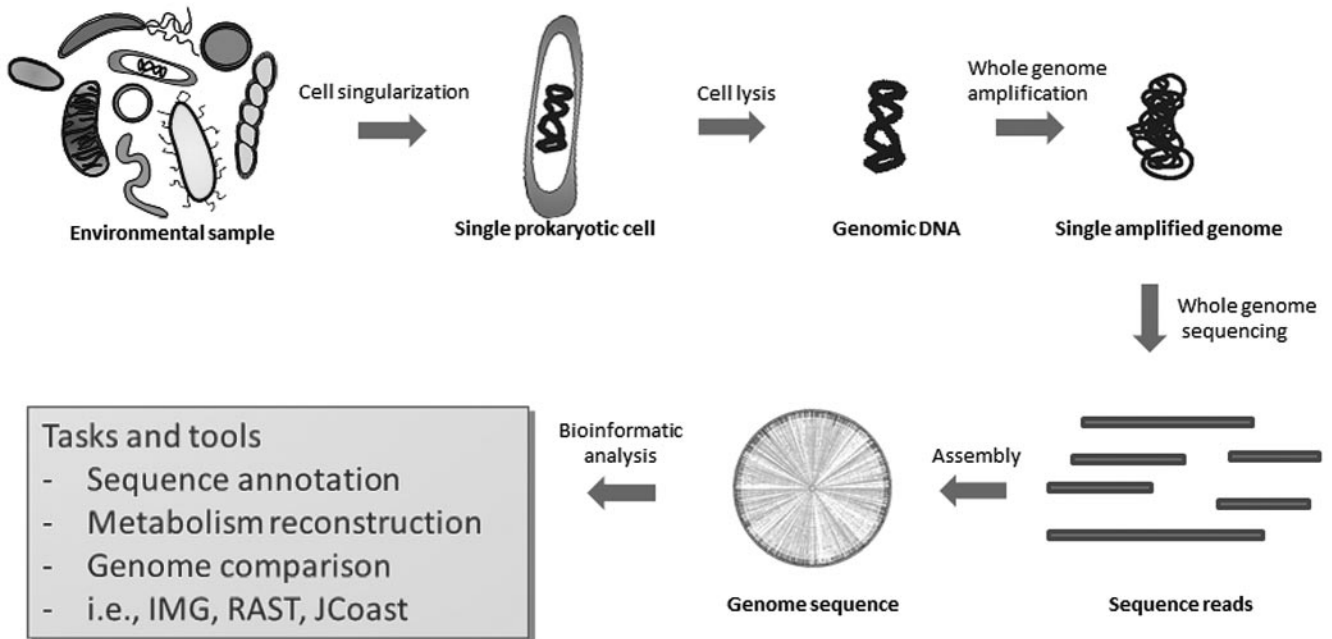


Figure 1. Diagram of the experimental procedures involved in single-cell genomic analysis.

ples. We further discuss how single-cell genomics has been applied to provide genomic insights into secondary metabolism, and we present future prospects of how this technique may lead to further advancements in the field. The application of single-cell genomics to other areas of microbiology is beyond the scope of this paper, and we refer the reader to other recent reviews (Hutchison and Venter, 2006; de Jager and Siezen, 2011; Kalisky and Quake, 2011; Yilmaz and Singh, 2011).

A Laboratory Primer on Single-Cell Genomics

The principle of single-cell genomics is to singularize microbial cells from environmental samples, to access the complete genomic material of a single cell, and to generate sufficient amounts of DNA by amplification for whole-genome sequencing (Fig. 1). The first step is the efficient singularization of the cells, which depends largely on the characteristics of the sample. If enrichments or even pure cultures are available, serial dilution is a possible method (Zhang *et al.*, 2006). One of the most commonly used methods for obtaining single cells from uncultivated microbiota is fluorescence-activated cell sorting (FACS). Microbial cells can be labeled with a fluorescent dye or subjected to fluorescence *in situ* hybridization (FISH) to target intact cells or cells of a distinct phylogenetic affiliation (Podar *et al.*, 2007; Yilmaz *et al.*, 2010). The Single Cell Genomics Center at the Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, offers services for single-

cell sorting and, optionally, also whole-genome amplification, thus making this technique available to any laboratory.

Microfluidic chambers have also proven successful for obtaining single amplified genomes (SAGs) (Marcy *et al.*, 2007a, b; Blainey *et al.*, 2011). This method uses reaction volumes of only 60 nl, which reduces the likelihood of contaminating the sample. Microfluidic devices have also been recently developed for FISH and tyramide-signal-amplification FISH (tsa-FISH) followed by cell sorting *via* flow cytometry directly on the device (Chen *et al.*, 2011; Liu, P., *et al.*, 2011). This approach holds great promise for 16S rRNA gene-based identification of single cells, while bearing low risks of contamination.

Micromanipulation techniques are particularly useful when the target microorganism is morphologically distinct (Ishoev *et al.*, 2008; Woyke *et al.*, 2010). Micromanipulation using microcapillaries has been used successfully in combination with FISH to target phylogenetically distinct cells of interest (Kvist *et al.*, 2007). Other options are micromanipulation by use of optical tweezers or laser capture microdissection. The first method has already been used in combination with microfluidic devices (Pamp *et al.*, 2012), while the latter was applied only to fixed bacterial samples (Klitgaard *et al.*, 2005), thus not permitting subsequent whole-genome amplification procedures.

Following single-cell separation, the next step is cell lysis to provide access to the genomic material, and subsequent multiple displacement amplification (MDA) (Lasken, 2007). MDA relies on the Phi29 polymerase enzyme that

amplifies randomly primed template DNA in an isothermal reaction with very high efficiencies. Because of the strand displacement activity of Phi29 polymerase, newly synthesized DNA becomes directly accessible for the next polymerase molecule, thus resulting in continuous DNA amplification. For a more technical description of the MDA process, laboratory protocols, and a list of commercially available kits, the interested reader is referred to several reviews (Lasken *et al.*, 2005; Silander and Saarela, 2008).

The next step, the whole-genome sequencing of MDA products, has been approached by a variety of sequencing methods. Most studies thus far have relied on 454 pyrosequencing using shotgun or paired end libraries (*e.g.*, Marcy *et al.*, 2007b; Musmann *et al.*, 2007; Blainey *et al.*, 2011; Siegl *et al.*, 2011). However, hybrid sequencing of 454 and Illumina techniques have been shown to produce better genome coverage than a single sequencing technique (Rodrigue *et al.*, 2009). Combinations of 454 and Sanger sequencing have also been used (Woyke *et al.*, 2009), and the first complete single-cell-derived genome was constructed using Sanger, 454, and Illumina techniques (Woyke *et al.*, 2010). As sequencing techniques are constantly evolving, novel approaches such as the PacBio (Pacific Biosciences of California, Inc.) or IonTorrent (Torrent Systems, Inc.) systems might enable even better draft genome recovery from single cells.

Technical Challenges

While the single-cell technology opens a major window of opportunities into symbioses research, the methodology is still subject to various technical challenges, which are detailed below.

Contamination

Contamination with non-target cells or DNA is one of the major challenges of the single-cell approach. Because single-cell whole-genome amplification *via* MDA is random hexamer primed, any piece of DNA in the reaction mix will co-amplify and compete for amplification with the low fg-range target DNA, if of sufficient length. Contaminating sequences not only reduce sequencing efficiency, but also may significantly confound the analysis of novel single-cell genomes. Contamination can be process-introduced or sample-introduced.

To prevent process contamination, the most stringent decontamination procedures are needed (Table 1). It is best practice to bleach-sterilize work areas and laboratory equipment and UV-irradiate all disposables, as well as buffers and water utilized within the single-cell pre-MDA work flow (Stepanuskas and Sieracki, 2007; Rodrigue *et al.*, 2009). Even with such stringent preventive measures, process-introduced contamination is a rather common phenomenon, largely due to the presence of contaminants such as

Table 1

Current technical challenges and potential solutions in single-cell genomics

Challenge	Potential Solution
Contamination	<ul style="list-style-type: none"> ● Bleach-sterilize work area and equipment ● UV-irradiate all disposables and reagents ● Perform two cycles of cell sorting (if possible) ● Extensively rinse the microbial cell (<i>i.e.</i>, after micromanipulation) ● Remove known contaminants (<i>i.e.</i>, <i>Delftia</i>, human) by binning methods post-MDA
Limited lysis	<ul style="list-style-type: none"> ● Chemically lyse <i>via</i> alkaline solution (KOH) ● Enzymatically lyse using lysozyme and/or proteases or custom-made enzyme cocktails ● Freeze/thaw, heat, <i>etc.</i>
Amplification bias	<ul style="list-style-type: none"> ● Reduce reaction volume, <i>i.e.</i>, by addition of crowding agents ● Normalize sequencing libraries ● Digitally normalize post-sequencing
Chimerism	<ul style="list-style-type: none"> ● Avoid long-mate pair libraries
Fragmented and partial nature of genomes	<ul style="list-style-type: none"> ● Pool individual single amplified genomes representing the same operational taxonomic unit (when applicable)

Delftia acidovorans in MDA reagents. As none of the commercial MDA reagents available to date are designed for single-cell applications but rather for the amplification of ng-range DNA or for many hundreds of cells, there has been no incentive or need to provide completely DNA-free reaction components.

To circumvent this issue, UV-irradiation of MDA reagents including Phi29 has been used successfully to minimize the co-amplification of free bacterial DNA found in commercial reagents during single-cell MDA (Woyke *et al.*, 2011). Moreover, it is possible to prepare ultra-pure Phi29 in house by using affinity-purification of recombinant Phi29 DNA (Blainey and Quake, 2011).

Sample-introduced contamination represents a slightly different challenge. If single cells are to be isolated using micromanipulation, where individual cells are transferred in rather large volumes, extensive rinsing of single cells by repetitive transfer to clean buffer drops on a slide may be helpful in shedding free DNA or additional small cells. However, this may not suffice, as seen in the *Sulcia* single-cell sequencing project, where more than 40% of the sequence reads were probably host-derived, despite extensive rinsing (Woyke *et al.*, 2010). Isolating single cells using fluorescence-activated cell sorting (FACS) has been shown to eliminate the carry-over of exogenous DNA by minimizing the transfer volume (droplet volume of 10 pl). Rodrigue and colleagues (2009) nicely demonstrated the successful removal of unwanted free DNA by performing two cycles of cell sorting.

Even when maximum preventive measures are taken, a thorough post-MDA quality control is advisable. Direct

Sanger sequencing of 16S rRNA gene PCR products generated for each MDA product has been used to identify the phylogeny of the single-cell genome. In addition, these data can be used to detect a potential contaminant by a thorough review of the Sanger sequencing chromatograms. As the rRNA genes generally represent only a minute fraction of the whole genome (<1%), simply relying on this method has proved rather inefficient. For example, even though no 16S rRNA gene was detected by PCR in the *Sulcia* single-cell project, more than 40% of the sequence data was derived from likely *Delftia* contamination (Woyke *et al.*, 2010). As was expected, the assembled *Delftia* contigs did not encode the 16S rRNA gene.

Post-sequencing, the bioinformatic detection and removal of common contaminants such as *Delftia* and human may be simple, but it can also be challenging if there is no clear discrimination between the target organism and the contaminant. Even when a combination of binning methods such as nucleotide signatures and phylogenetic assignments based on Blast analysis are applied, the distinction between true biological data associated with the single cell (*i.e.*, horizontal gene transfer, plasmids, phage infecting the single cell) and a potential contaminant may be blurred, particularly in highly fragmented assemblies or genomes that lack any near neighbor within the sequence database. For symbiont single-cell genomes inhabiting hosts that do not have a closely related sequenced representative, the confident identification of potential host-derived sequence can be a major challenge.

Limited lysis

Lysis is a key step in single-cell whole-genomes amplification, because it exposes the genomic DNA to make it accessible for amplification. As some cells may harbor only a single copy of their genome, lysis should be gentle so as to fully maintain the integrity of the DNA. Extensive nicking of the genomic DNA and, even more so, introduction of dsDNA breaks leads to complete loss of the linkage information at these genomic sites. Thus the lysis method of choice must be mild, yet harsh enough to enable access to the genomes for the majority of single cells. In an ideal experiment, lysis should be accomplished for each of the single cells isolated, providing access to the genetic make-up for every cell. To date, however, no universal lysis method for all taxa exists.

Cell lysis methods can generally be categorized as chemical, enzymatic, and physical (including acoustic such as sonication, and optical and mechanical), each having their weaknesses and strengths and suitability for prokaryotic single-cell genomics. The currently most common lysis method for a single bacterial cell is chemical lysis *via* alkaline solution (KOH). While easily applicable, this method for opening a cell has a success rate of only about

20% for various environments. Enzymatic lysis methods make use of cell-wall-cleaving enzymes such as lysozyme in combination with proteases (Tamminen and Virta, 2010; Fleming *et al.*, 2011). Although this is a gentle lysis method, vast variations in cell wall properties among different organisms render it unlikely to be universally applicable. Cocktails combining cell-wall-active enzymes with differing specificities such as lysozyme, achromopeptidase, mutanolysin, and lysostaphin may be a viable solution, albeit thorough decontamination may be a necessity. Alternative methods that have been applied or suggested for single-cell genomics include physical methods such as freeze/thawing, heat, and combinations of the various methods discussed here (Kvist *et al.*, 2007; Mussmann *et al.*, 2007; Siegl *et al.*, 2011). For a comprehensive review of single-cell lysis techniques, please refer to Brown and Audet (2008).

Amplification bias

An array of different approaches to reduce amplification bias have emerged over the years, with the most promising result shown by Marcy *et al.* (2007b). The authors demonstrated that a reduction in reaction volumes (60 nl as compared to 50 μ l) greatly reduces amplification bias. While sub-microliter-scale amplifications as achieved by microfluidic chips is not practical for many laboratories, crowding agents mimicking smaller reaction volumes, such as PEG and trehalose (Ballantyne *et al.*, 2006; Pan *et al.*, 2008) may be easier to implement into a laboratory's single-cell workflow (Table 1). Additional methods to deal with uneven genome coverage are applied post-MDA. Normalized libraries can be generated using protocols based on subtractive hybridization with the usage of duplex-specific nuclease (Rodrigue *et al.*, 2009; Swan *et al.*, 2011). The major drawback of normalized libraries is that these are labor-intensive and, unless they are adapted to a plate format, not suitable for high-throughput single-cell sequencing projects. Lastly, digital normalization, which informatically reduces the sequence information for over-represented regions of the genome prior to assembly, has proved to be of tremendous value for single-cell sequence data (Rodrigue *et al.*, 2009; A. Sczyrba, University of Bielefeld, pers. comm.)

Chimerism

The nature of MDA introduces chimeric rearrangements, which become apparent after sequencing and are found on the order of one chimeric junction every 20 kbp (Lasken and Stockwell, 2007; Marcy *et al.*, 2007b; Woyke *et al.*, 2009). Although some reports have shown the reduction of chimerism due to S1 nuclease treatment (Zhang *et al.*, 2006), other laboratories have failed to evidence this effect (Woyke *et al.*, 2009). The Lasken laboratory extensively characterized the types of chimera formed during MDA, determining that the majority (~85%) of these rearrange-

ments are inversion/deletion events (Lasken and Stockwell, 2007). Such artifacts can challenge assembly algorithms, and manual assembly curation may be required to break chimeric contigs. It is advisable to avoid long-mate pair libraries, which will provide a high percentage of incorrect pairing information (Table 1).

Fragmented and partial nature of single-cell genomes

Single-cell genomes sequenced to date range from partial genomes of a few hundred kilobases in assembly size (Youssef *et al.*, 2011) to a finished genome (Woyke *et al.*, 2010). The majority of single-cell genomes recovered with today's methodologies will likely be fragmented and partial in nature, rather than resembling a truly complete genome (Table 1). We suspect that the completely finished *Sulcia* single-cell genome will, at least for now, remain an exception rather than the rule (Woyke *et al.*, 2010). This "drafty" nature of most single-cell-derived genomes, in addition to the bias and chimera issues discussed above, makes the bioinformatic analysis of the data less straightforward than sequence data from isolate genomes. As single-cell genomics becomes increasingly popular, various tools specifically designed for genome data from single cells have become available. Over the last few years, several software packages for single-cell assembly have been released that address the problem of highly variable coverage rate in MDA-derived data. SmashCell (Simple Metagenomics Analysis SHell-for sequences from single Cells) is a software framework that combines assembly, gene prediction, and annotation of single-cell data (Harrington *et al.*, 2010). Assemblers that followed were IDBA-UD (Peng *et al.*, 2012) and Velvet-sc. (Chitsaz *et al.*, 2011), and most recently, the novel single-cell-specific assembler called SPAdes, developed by the Pevzner group (Bankevich *et al.*, 2012).

Automatic annotation and its manual refinement can be very challenging when dealing with fragmented single amplified genomes (SAGs), especially for cells with no closely related reference genomes available. Drafty single-cell genomes might not provide the necessary genomic context to securely annotate a gene, or the gene of interest itself is fragmented and thus cannot be annotated with high confidence. This in turn affects the ability to predict the existence of metabolic pathways, as certain key enzymes might be missing from the single-cell genome or are not clearly annotated. Thus tools generally used for comparative genomics are to be used with caution for single-cell data, and the analysis of single cells may be limited. This is a challenge that at present cannot readily be addressed with mere bioinformatics but might be overcome through the steadily increasing amount of available genomic data as well as further improvements in the recovery of single-cell genomes from the environment.

A current strategy to improve assembly and genome

recovery for single cells is the pooling of individual SAGs representing the same operational taxonomic unit (OTU) (Podar *et al.*, 2007; Warnecke and Hugenholtz, 2007; Blainey *et al.*, 2011). This strategy will either await the availability of replicate SAGs within the pool of randomly isolated single cells or require enrichment based, for example, on 16S rRNA probes. Blainey *et al.* (2011) demonstrated elegantly that with the pooling of five SAGs the recovery of more than 95% of a single-cell genome was approached, here that of "*Candidatus* Nitrosoarchaeum limnia SFB1," an ammonia-oxidizing archaeon. Although this is a fine strategy if the environmental population is clonal, it becomes more challenging in a heterogeneous population. Pre-binning of the SAG data using average nucleotide identity (ANI) (Konstantinidis *et al.*, 2006) prior to co-assembly of these datasets may be beneficial in this regard.

Although many laboratories have been working on improving some of the key aspects of single-cell genome sequencing in an attempt to improve the quality of the recovered genomes, there is still major room for progress. A goal to aim for would be the complete recovery of each single cell isolated; small steps toward this goal will aid in moving the field of single-cell genomics to the next level. In the meantime, even a few hundred kilobases from a single cell can be of tremendous value by giving insight into the coding potential of microbial dark matter and providing a long-needed link between phylogeny and function for this uncultured majority.

Application of Single-Cell Genomics to Symbioses

In the following section we introduce representative single-cell genomics studies in a host-associated context. This method has now been applied to a handful of symbiotic systems, including bacterial symbionts of marine sponges, insects (grasshoppers, termites), and vertebrates (mouse, human) (Fig. 2). We give an overview of each experimental system and show what contributions single-cell genomics has made to the corresponding field.

Poribacterial symbiont of marine sponges

Many marine sponges contain massive amounts of microorganisms within their mesohyl matrix, which can contribute up to 35% of the animal's biomass (Hentschel *et al.*, 2006, 2012; Taylor *et al.*, 2007; Webster and Taylor, 2012). Members of at least 30 bacterial phyla and both archaeal lineages were found by high-throughput sequencing technologies within sponge hosts (Webster *et al.*, 2010; Schmitt *et al.*, 2012). The candidate phylum Poribacteria is among the predominant microorganisms in these microbial consortia (Lafi *et al.*, 2009; Schmitt *et al.*, 2011). As with most sponge-associated microorganisms, little is known about the function that *Poribacteria* might play in

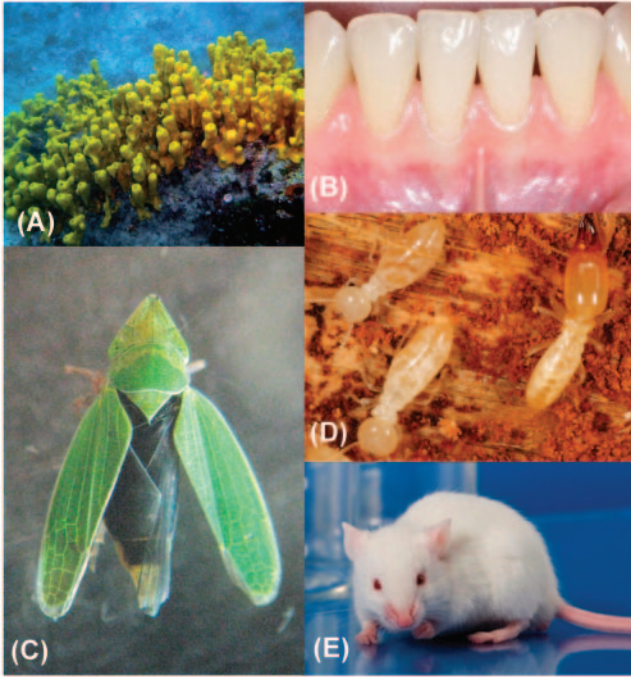


Figure 2. Host animals of symbionts, for which single-cell genomics has been employed: the marine sponge *Aplysina aerophoba* (A), human mouth (B), the green sharpshooter *Draeculacephala minerva* (C), the termite *Reticulitermes speratus* (D), and the mouse *Mus musculus* (E). Photo credits: Janine Kamke, University of Würzburg (A); Martin Linke, Prophylaxepraxis Ahaus (B); Damon Tighe, DOE Joint Genome Institute (C); Yuichi Hongoh, Tokyo Institute of Technology (D); photographer: Valdek Dmytrowski, photo courtesy of Taconic (E).

this symbiosis. This lack is largely attributed to the fact that none of the sponge symbionts have yet been cultured.

Siegl *et al.* (2011) were able to obtain a single poribacterial cell (Poribacteria WGA A3) from the marine sponge *Aplysina aerophoba* by a customized cell-separation protocol followed by FACS sorting (Fig. 2a, Table 2). Through alkaline lysis, MDA, and 454 pyrosequencing, 105 Mbp of raw sequence were recovered, which assembled to a genome size of 1.88 Mbp. While being somewhat flawed by the fragmented nature of many operons, more than 500 contigs contained at least one complete open reading frame that, in combination with the full dataset, led to a comprehensive analysis of the genomic repertoire of a single poribacterium. This study shows that it is possible to arrive at comprehensive genomic information from single cells, even when collected from exceedingly diverse samples, such as marine sponges.

The poribacterial single-cell genome sequence encoded genes involved in glycolysis, TCA cycle, and oxidative phosphorylation (Fig. 3). Furthermore, a purine and pyrimidine metabolism, pathways for canonical amino acids, were identified, rendering auxotrophy an unlikely event (Siegl *et al.*, 2011). Additionally, nitrite assimilation is highly likely to occur in *Poribacteria*, as indicated by the presence of two assimilatory nitrite reductases.

The single-cell genomics study on *Poribacteria* shows potential for degradative metabolism through the presence of several sulfatases, peptidases, and proteins related to

Table 2

A compilation of published reports in which single-cell genomics have been employed in a symbiosis context

Host system	Marine sponge <i>Aplysina aerophoba</i>	Sharpshooter <i>Draeculacephala minerva</i>	Termite gut protist <i>Trichonympha agilis</i>	Human subgingival crevice	Mouse intestine
Authors	Siegl <i>et al.</i> , 2011	Woyke <i>et al.</i> , 2010	Hongoh <i>et al.</i> , 2008	Marcy <i>et al.</i> , 2007b	Pamp <i>et al.</i> , 2012
Bacterial symbiont	Candidate phylum Poribacteria, WGA A3	Candidate <i>Sulcia muelleri</i> DMIN	Candidate phylum Termite group 1, Rs D-17	Candidate phylum TM7	Candidate clade <i>Arthromitus</i> , Clostridiaceae
Microbial diversity	Very high	Low	High	High	SFB monocolonized
Cell isolation method	FACS	Micromanipulation	Micromanipulation	Microfluidic chip	Laser tweezers and microfluidic chip
Coverage	~75% of a single genome	Closed single-cell genome	Closed composite genome	Not estimated	~99% of each of five genomes
Genome data	~1.88 Mbp assembled	0.24 Mbp	1.13 Mbp	2.86 Mbp	1.28-1.50 Mbp
Sequencing method	454 pyrosequencing	Sanger sequencing, 454 pyrosequencing, Illumina sequencing	Sanger sequencing, 454 pyrosequencing	454 pyrosequencing	Multiplex 454 pyrosequencing

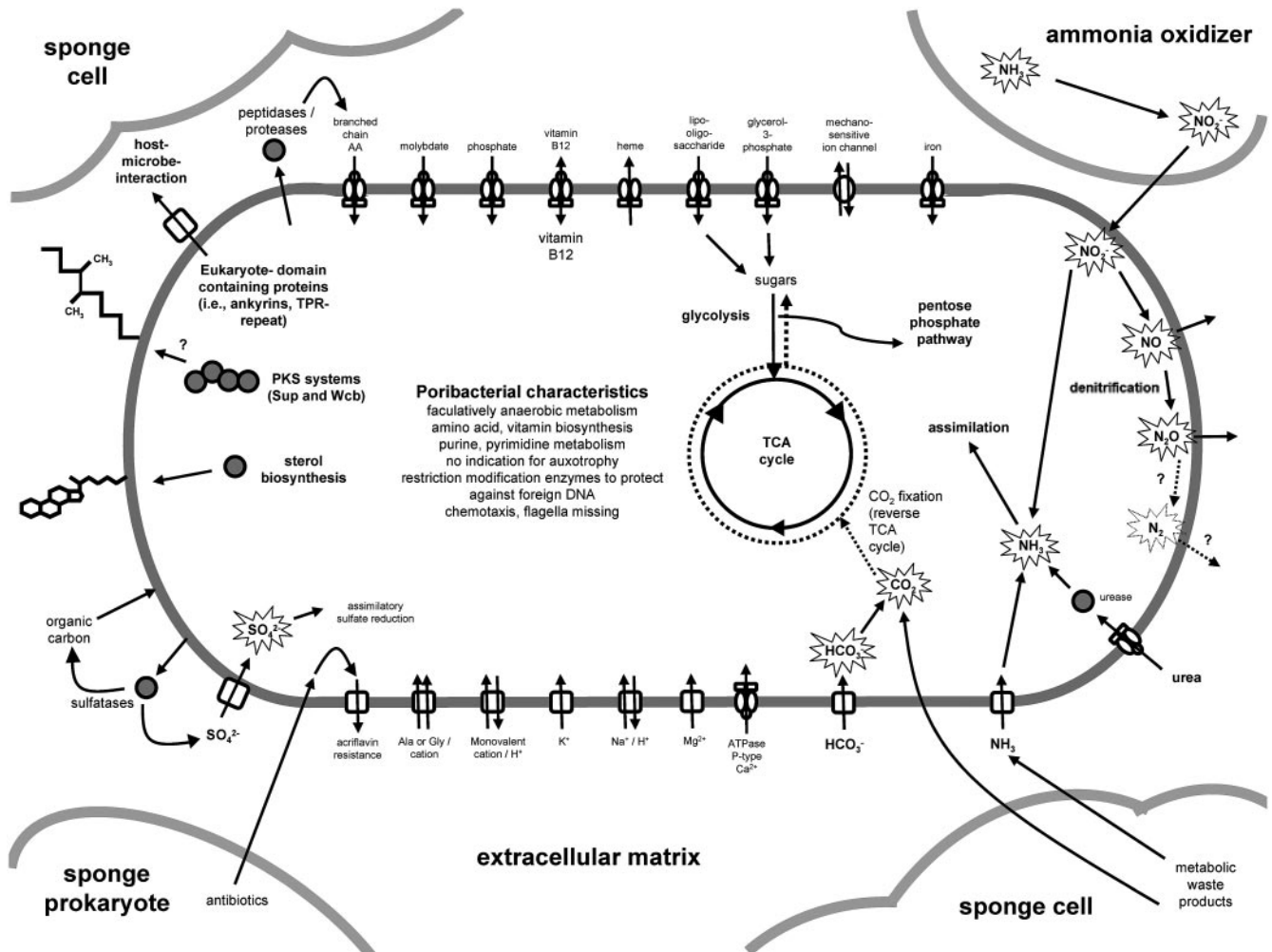


Figure 3. Predicted genomic features of poribacterial symbionts from marine sponges. Modified after Siegl *et al.* (2010), ISME J 5: 61–70, with kind permission by Nature Publishing Group.

N-glycan degradation (Siegl *et al.*, 2011). Degradation of complex substances was also suggested by Liu, M., *et al.* (2011) for another bacterial symbiont of *Cymbastella concentrica*. Furthermore, Hallam *et al.* (2006) reported on several proteases that might degrade extracellular matrix proteins as a defense mechanism on the genome of *C. symbiosum* in *Axinella mexicana*. Whether these degradative enzymes are used in defense or serve a nutritional purpose for the sponge symbionts would be an interesting question for future studies.

There are indications that mechanisms underpinning host-microbe interactions are encoded on the genome of “*Candidatus* Poribacteria WGA A3.” These are proteins with eukaryote-like domains, such as ankyrin repeats and tetratricopeptide repeats (TPR), which are likely to mediate protein-protein interactions (Siegl *et al.*, 2011). The latter have been found in all known genomic datasets from prokaryotic sponge symbionts (Hallam *et al.*, 2006; Thomas *et al.*, 2010; Liu, M., *et al.*, 2011), and the expression of

proteins with these domains has been confirmed (Liu *et al.*, 2012). Ankyrins are especially likely to be of relevance as they might be involved in the recognition of and protection from host phagocytosis (Liu *et al.*, 2012). This theory is further strengthened by the fact that proteins with ankyrin domains are found in other obligate intracellular pathogenic and symbiotic systems, where they interfere with host cell function (*e.g.*, Mavromatis *et al.*, 2006; Walker *et al.*, 2007; Habyarimana *et al.*, 2008; Voth *et al.*, 2009; Al-Khodori *et al.*, 2010; Murray *et al.*, 2011).

Intracellular bacterial symbiont of a sharpshooter

“*Candidatus* S. muelleri” and “*Candidatus* Baumannia cicadellincola” are the two obligate symbionts of the green sharpshooter *Draeculacephala minerva* (Fig. 2c). They are localized in the bacteriome, a specialized organelle of the insect that harbors obligate symbionts, and are vertically transmitted *via* the eggs to the next generation (Moran and

Baumann, 2000). *S. muelleri* was found as an intracellular symbiont in several insect species together with *B. cicadellinicola* or other symbiotic bacteria (Takiya *et al.*, 2006; McCutcheon *et al.*, 2009). These insects are feeding on plant sap, which lacks many essential nutrients that are instead supplied by their bacterial symbionts (Redak *et al.*, 2004; McCutcheon *et al.*, 2009). Symbioses with *S. muelleri* are believed to have been established at least 260 million years ago in an ancestor species of today's insects (Moran *et al.*, 2005).

Woyke and coworkers (2010) were able to sequence, nearly base-perfect, the complete genome of *S. muelleri* from *D. minerva* (*S. muelleri* DMIN) (Table 2). They achieved cell separation by dissecting the bacteriome and using light microscopy and micromanipulation to select single cells with the previously described morphology of *S. muelleri*. Through hybrid sequencing of 454 and Sanger techniques it was possible to generate a draft genome, which could be closed with additional Sanger finishing and Illumina polishing. The resulting genome has a size of 243,933 bp. Although the authors faced contamination problems through both MDA reagents and probable exogenous host DNA, they were able to extract *S. muelleri* reads by mapping their data against an existing *S. muelleri* genome from the glassy-winged sharpshooter (*S. muelleri* GWSS) (Wu *et al.*, 2006; McCutcheon and Moran, 2007). The study by Woyke and coworkers (2010) is an example of nearly ideal conditions for a single-cell genomics study because (i) only two bacterial symbionts are housed in the bacteriome, (ii) the target microorganism was morphologically distinct, (iii) the genomes of insect symbionts are extremely reduced and polyploid, thus making genome closure easier to accomplish, and (iv), a reference genome for *Candidatus Sulcia muelleri* was available.

By functional analysis of the genomic data, Woyke *et al.* (2010) found no difference in the metabolic capability encoded in their single-cell-derived genome and available reference sequences (Wu *et al.*, 2006; McCutcheon and Moran, 2007). The *Sulcia* genome encodes first and foremost biosynthesis pathways for eight essential amino acids. The metabolic capacities of *Sulcia* have shown to be complementary with those of the second sharpshooter symbiont "*Candidatus Baumannia cicadellinicola*," which has the genomic potential to produce additional essential amino acids and several vitamins (McCutcheon and Moran, 2007). Comparison of genes with lower similarity between two *S. muelleri* genomes (DMIN and GWSS) identified potential bacterial surface antigens (Woyke *et al.*, 2010). These proteins might be connected to host specificity, and further investigations could reveal interactions between bacteria and host cells.

To evaluate the quality of the newly obtained genome from the single-cell source, the authors compared it to a metagenomically derived *S. muelleri* genome, which they

constructed out of total DNA from 25 bacteriomes from *D. minerva*. The low number of sequence polymorphisms might represent population variations in otherwise conserved data (Woyke *et al.*, 2010). The heterogeneity within the population was further analyzed by detection of single-nucleotide polymorphisms (SNPs) in two independent *S. muelleri* metagenome datasets (a metagenome library of a single bacteriome and the before-mentioned metagenome library of 25 pooled bacteriomes). A low number of SNPs were detected in the latter only. These results suggest a low genetic variety, which is consistent with the stable genome contents and arrangements that have been shown for other primary symbionts of insects (Tamas *et al.*, 2002). *S. muelleri* sequences appear to evolve extremely slowly even when compared to other intracellular symbionts of insects (Takiya *et al.*, 2006).

The study by Woyke *et al.* (2010) proves clearly that obtaining a complete, high-quality genome from a single bacterial cell is possible. Previous data regarding symbiont metabolism were confirmed, and only minor differences between previously sequenced genomes of the same bacterial species were detected, thus supporting the current theory that genetic variety among microbial symbionts is rather low. This analysis reiterates the power of single-cell genomics for the investigation of insect symbiosis.

TG1 symbionts in termite gut protists

Termites are social insects that live mostly on dead plant and wood material. Their ability to digest lignocellulose renders termites and their associated microbial consortia of interest for biofuel production (Weng *et al.*, 2008; Scharf *et al.*, 2011). The eukaryotic and prokaryotic symbionts of termites are phylogenetically diverse, with often several hundred bacterial species being present (reviewed by Hongoh, 2010; Husseneder, 2010). The microbial gut protists are mostly found in lower termites and also harbor bacterial symbionts that aid in the digestion of wood particles (Stingl *et al.*, 2005; Ohkuma, 2008).

Hongoh *et al.* (2008) isolated a single *Trichonympha agilis* protist cell from the termite *Reticulitermes speratus*, via micromanipulation (Fig. 2d). They were able to retrieve approximately 10^3 cells of the bacterial phylotype Rs D-17, members of the candidate phylum termite group 1 (TG1). These bacteria are predominant and exclusive to the posterior of the host flagellate *T. agilis*. Multiple displacement amplification enabled the recovery of sufficient DNA for genome sequencing. Through the combination of genomic material from many clonal cells and subsequent whole-genome amplification, it was possible to retrieve a complete, composite genome sequence of Rs D-17 cells, with a size of approximately 1.13 Mbp.

The obtained genome sequence showed evidence for the metabolic adaptation of the bacterial symbiont to an intra-

cellular lifestyle. The Rs D-17 genome still contains several intact pathways for amino acid and cofactor biosynthesis. Several genes belonging to these pathways are duplicates, indicating the importance of those substances for the symbiosis. The eukaryotic partners in this interaction are dependent on the retrieval of amino acids and vitamins from bacterial symbionts, since their lignocellulose-based nutrition does not supply these essential compounds (Husseneder, 2010). In return, the TG1 bacteria are supplied with glucose-6-phosphate (a dominant carbon source) and glutamate (as a nitrogenous compound). Glutamate biosynthesis is disrupted in the bacterial genome but most likely supplied by the protist. The protist might also supply phosphorylated glucose to the symbiont, which helps preserve the symbiont's ATP reservoirs, which are predicted to be produced by fermentation of sugar to acetate. Thus, a complementary metabolism between the symbiotic partners was proposed (Hongoh *et al.*, 2008).

Hongoh *et al.* (2008) further proposed streamlining adaptation as a result of reduced genome size, the presence of several pseudogenes, and the presence of duplicated regions of metabolic relevance. This system is thus an excellent subject for further studies of evolutionary pressure on genomes of intracellular symbionts. It would be highly interesting to compare the TG1 symbiont genomes from different protist species, which are known to harbor specific and phylogenetically divergent TG1 bacteria (Stingl *et al.*, 2005; Ohkuma *et al.*, 2007). If genome adaptation is still ongoing, it might be possible to investigate both co-evolution with their respective protist hosts and convergent evolution of phylogenetically different TG1 clades.

Candidate phylum TM7 from human gingival crevice

The first study that applied single-cell genomics to host-associated bacteria was conducted on the candidate phylum TM7 from human mouth biofilms (Marcy *et al.*, 2007b). Representatives of the candidate TM7 phylum have been found in biofilms of the subgingival crevice in healthy humans and also in conjunction with periodontitis (Fig. 2b) (Colombo *et al.*, 2009; Crielaard *et al.*, 2011). The fact that these microorganisms are not highly abundant in the microbiome of the human mouth represents a special challenge when obtaining single cells. Cells were singularized from a biofilm sample using a microfluidic device in combination with light microscopy (Table 2) (Marcy *et al.*, 2007b). This permitted a more targeted selection of cells based on morphological properties. Marcy *et al.* (2007b) sequenced the genomic data from three single TM7 cells (Tm7 a–c), which resulted in datasets of various sizes (TM7 a: 2.86 Mbp assembled data, TM7 b: 10 Mbp unassembled data, TM7 c: 474 kbp assembled data). To ensure analysis with exclusively high-quality data, exclude contamination, and minimize the influence of possible MDA bias, the authors

used a very strict quality filtering and binning approach to construct the 963-kbp “TM7 metagenome” out of the combined data of all three cells.

On the basis of this information, the authors predicted the presence of several major pathways in the metagenome, such as glycolysis, TCA cycle, nucleotide biosynthesis, and biosynthesis and salvage pathways for several amino acids (Marcy *et al.*, 2007b). Furthermore, growth of TM7 microorganisms on oligosaccharides and amino acids was indicated, which is consistent with the high nutrient environment of the human mouth. Proteins involved in type IV pilus biosynthesis were also identified, possibly representing a virulence factor (Marcy *et al.*, 2007b). The authors further predicted UDP-*N*-acetylmuramyl tripeptide synthetase to be involved in virulence. In bifidobacteria, this enzyme is involved in peptidoglycan formation, which plays a role in chronic granulomatous inflammation (Simelyte *et al.*, 2003).

This study (Marcy *et al.*, 2007b) is an elegant example of how the single-cell genomics approach can result in novel information about candidate phyla where cultured representatives do not exist. Although the strict binning approach excluded a substantial number of potential TM7 reads, it still provided previously unavailable genomic information about this candidate phylum TM7.

Segmented filamentous bacteria from mouse intestine

Segmented filamentous bacteria (SFB), a specific clade of *Clostridia*, are host-specific symbionts that are present in the lower intestine of many vertebrates. SFB form segmented filaments, which are firmly attached to the epithelia of the host intestine. They are of special interest because they were shown to directly influence the host's immune system (reviewed by Reading and Kasper, 2011). Pamp *et al.* (2012) used a combination of laser tweezers and microfluidic chips to isolate five individual SFB filaments from feces of a SFB-monocolonized mouse (Fig. 2e, Table 2). Data were assembled individually and resulted in greater than 98% genome coverage for each genome. Additionally, two versions of composite genomes including data from all five cells were assembled, one *de novo* and one with an existing reference SFB genome (Prakash *et al.*, 2011).

The presence of nearly complete glycolysis and pentose phosphate pathways indicates metabolic heterotrophy in SFB. Genes involved in the electron transport chain were lacking, leading the authors to postulate substrate-level phosphorylation through phosphoglycerate, pyruvate, and acetate kinases, as well as the production of molecular hydrogen by pyruvate ferredoxin oxidoreductase (Pamp *et al.*, 2012). The genomic potential of SFB also indicates fermentation through several dehydrogenases coupled with substrate oxidation. Extracellular proteases and several

transport systems are also encoded, which might ensure breakdown of larger peptides and uptake of the resulting amino acids. The bacteria seem indeed to be dependent on the uptake of amino acids, vitamins, and co-factors (Pamp *et al.*, 2012). The authors further suggest an anaerobic metabolism for SFB, in which amino acids are taken up and fermented to sustain energy supply, and together with vitamins and co-factors, are used for maintenance of cellular processes.

Four specific protein sequence clusters were identified that distinguish SFBs from other members of the family *Clostridiaceae* 1. Because of their extracellular location and limited catalytic domains, these proteins might be components of surface structures that are involved in niche adaptation (Pamp *et al.*, 2012). The polymorphisms found in the surface proteins suggest the differentiation of SFB among their host animal strains, which may indicate the species (or strain)-specific adaptation of SFB to each host (Kuwahara *et al.*, 2011; Prakash *et al.*, 2011; Sczesnak *et al.*, 2011). Although sequence similarity was quite high between genomes from the same host species (98%–99%), the authors could detect distinct differences between different mice strains, thus supporting the hypothesis of host speciation.

With respect to factors mediating host-microbe interactions, the SFB genomes encode flagella that allow movement within the host mucous layer, or that might be involved in cell adhesion or even cell recognition by the host, which, in turn, can cause an immune response (Pamp *et al.*, 2012). Furthermore, homologs of myosin-cross-reactive antigen (MCRA) were identified, which can trigger an autoimmune disease during *Streptococcus* spp. infections (Wu *et al.*, 2010). MCRAs are also involved in adherence and survival in the host context in other bacteria (O’Flaherty and Klaenhammer, 2010; Malachowa *et al.*, 2011). Additional proteins with potential to mediate microbe-host interactions are ADP-ribosyltransferases (ADPRT). ADPRT are proteins that modify enzymes and are occasionally known as toxins secreted by some pathogenic bacteria. Four types of novel ADPRT that differ in sequence, structure, and location of domains were identified on the SFB genomes (Pamp *et al.*, 2012). Some of these proteins are found close to the phage-related genes and show similarities to toxins from other bacteria.

The SFB genomes also encode for proteins that protect the bacteria from host defenses and ensure their survival. For example, multidrug and toxic compound extrusion (MATE) proteins were found that might protect from antimicrobial substances (Pamp *et al.*, 2012). The authors further identified potential choloylglycine hydrolases that might aid in the protection from bile acid and O-acetyltransferases and polysaccharide deacetylase with the potential to weaken host lysozyme activity. Furthermore, the genome data suggest protection from oxidative stress by the pres-

ence of rubrerythrin and catalases, as well as immune evasion of SFB through catabolism of arginine (Pamp *et al.*, 2012). Finally, the SFB genomes also have the genomic potential to produce a protective extracellular polysaccharide capsule.

This study (Pamp *et al.*, 2012) provided unprecedented insights into the putative mechanisms of interaction between a specific clade of intestinal symbionts and epithelial cells of the mouse intestine. These insights were made possible by having access to an experimentally tractable host model (mouse) in which only the target symbionts were present. The study also benefitted from the availability of SFB reference genomes. As a result, several unique groups of proteins were identified with possible involvement in host-specification, interaction, and symbiont survival. These results have implications for bacterial genome evolution and speciation in the mammalian intestine.

Single-Cell Genomics: Linking Phylogeny With Function

In addition to providing genomic insights into single microbial cells, the method of single-cell genomics can be employed to link phylogenetic identity of a bacterium to a specific function (Swan *et al.*, 2011). Metagenomics has also been successfully used to describe novel microbial gene clusters for secondary metabolites (Kennedy *et al.*, 2010; Piel, 2011). However, in ecosystems with high-microbial diversity, it becomes nearly impossible (the “needle in the haystack” problem) to determine the phylogenetic origin of the sought-after biosynthetic pathway by metagenomics alone (but see exception in Beja *et al.*, 2000). In this context, a combination of single-cell genomics with metagenomics is the best solution (Siegl and Hentschel, 2010; Bayer *et al.*, 2012). The assignment of novel genes and functions to their biological origin is important for heterologous expression studies needed for sustainable metabolite production.

The study by Siegl and Hentschel (2010) aimed to clone secondary metabolite gene clusters from WGA products and to identify the corresponding microbial producers. For this purpose, the microbial consortia of a marine sponge were sorted by FACS and then subjected to WGA. A cosmid library was constructed from the WGA product of a sample containing two bacterial cells, one a member of the candidate phylum Poribacteria and one of a sponge-specific clade of *Chloroflexi*. Library screening led to the genomic characterization of two cosmid clones encoding a polyketide synthase (PKS) and a nonribosomal peptide synthetase (NRPS). PCR screening of WGA products from several additional, FACS-sorted cells supports the assignment of the Sup-PKS gene to *Poribacteria* and the novel NRPS gene to *Chloroflexi*. Here, the single-cell genomics approach has permitted the cloning of entire gene clusters from single

microbial cells of known phylogenetic origin, thus providing a sought-after link between phylogeny and function. One important drawback of this cosmid library approach was, however, the chimeric nature of the cloned WGA product, thus supporting the general impracticability of Phi29-amplified products for heterologous expression studies.

In a recent study, Bayer *et al.* (2012) explored FADH₂-dependent halogenase genes in microbial metagenomes and WGA products of FACS-sorted single cells of a marine sponge. Screening of a metagenomic library resulted in four halogenase-bearing clones that could not be taxonomically assigned. In the screened WGA products, 12 reactions were halogenase-positive, representing three distinct clades of these enzymes. For six of these products, a corresponding 16S rRNA gene could be identified for which purity was established by cloning of the respective PCR product and RFLP analysis of at least 32 clones. The WGA was considered to be derived from a single phylotype only if the restriction pattern was uniform. In this determination, deltaproteobacterial, actinobacterial, and poribacterial sponge symbionts were identified as possible producers of the three halogenase clades. The single-cell genomic analysis was the essential technique to allow for the assignment of a given function to specific microbial phylotypes.

Conclusions and Future Perspectives

The advantage of single-cell genomics for obtaining comprehensive information on individual microbial cells is undisputed. These analyses open up the opportunity to investigate sequence heterogeneity within symbionts of a single symbiotic host or host organ (*i.e.*, bacteriome), to delineate genomic differences, and to arrive at hypotheses on genome evolution and microbial speciation using individual microbial cells. Single-cell genomics is furthermore suitable for analysis of symbiosis systems with high microbial diversity in which even the most comprehensive metagenomic sequencing projects cannot address the question of which function belongs to which phylotype, and for investigations of candidate phyla where representatives have not been cultivated. Single-cell genomics has been fully embraced by the scientific community, and the road is clear ahead for it to become a mainstream technique in modern microbiology.

Acknowledgments

Financial support to U. H. was provided by the SFB630-grant TPA5 and by the Bavaria California Technology Center (BaCaTeC). T. W. was funded by the U.S. Department of Energy Joint Genome Institute, Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Literature Cited

- Al-Khodor, S., C. T. Price, A. Kalia, and Y. Abu Kwaik. 2010. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* **18**: 132–139.
- Ballantyne, K. N., R. A. van Oorschot, R. J. Mitchell, and I. Koukoulas. 2006. Molecular crowding increases the amplification success of multiple displacement amplification and short tandem repeat genotyping. *Anal. Biochem.* **355**: 298–303.
- Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski *et al.* 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.
- Bayer, K., M. Scheuermayer, L. Fieseler, and U. Hentschel. 2012. Genomic mining for novel FADH₂-dependent halogenases in marine sponge-associated microbial consortia. *Mar. Biotechnol.* DOI:10.1007/s10126-012-9455-2.
- Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Binga, E. K., R. S. Lasken, and J. D. Neufeld. 2008. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J.* **2**: 233–241.
- Blainey, P. C., and S. R. Quake. 2011. Digital mda for enumeration of total nucleic acid contamination. *Nucleic Acids Res.* **39**: e19.
- Blainey, P. C., A. C. Mosier, A. Potanina, C. A. Francis, and S. R. Quake. 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6**: e16626.
- Brown, R. B., and J. Audet. 2008. Current techniques for single-cell lysis. *J. R. Soc. Interface* **5** (Suppl. 2): S131–138.
- Chen, C. H., S. H. Cho, H. I. Chiang, F. Tsai, K. Zhang, and Y. H. Lo. 2011. Specific sorting of single bacterial cells with microfabricated fluorescence-activated cell sorting and tyramide signal amplification fluorescence in situ hybridization. *Anal. Chem.* **83**: 7269–7275.
- Chitsaz, H., J. L. Yee-Greenbaum, G. Tesler, M. J. Lombardo, C. L. Dupont, J. H. Badger, M. Novotny, D. B. Rusch, L. J. Fraser, N. A. Gormley *et al.* 2011. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* **29**: 915–921.
- Colombo, A. P., S. K. Boches, S. L. Cotton, J. M. Goodson, R. Kent, A. D. Haffajee, S. S. Socransky, H. Hasturk, T. E. Van Dyke, F. Dewhirst, and B. J. Paster. 2009. Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J. Periodontol.* **80**: 1421–1432.
- Crielaard, W., E. Zaura, A. A. Schuller, S. M. Huse, R. C. Montijn, and B. J. Keijser. 2011. Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med. Genomics* **4**: 22.
- de Jager, V., and R. J. Siezen. 2011. Single-cell genomics: unravelling the genomes of unculturable microorganisms. *Microb. Biotechnol.* **4**: 431–437.
- Fan, L., D. Reynolds, M. Liu, M. Stark, S. Kjelleberg, N. S. Webster, and T. Thomas. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. USA* www.pnas.org/cgi/doi/10.1073/pnas.1203287109.
- Fleming, E. J., A. E. Langdon, M. Martinez-Garcia, R. Stepanauskas, N. J. Poulton, E. D. Masland, and D. Emerson. 2011. What's new is old: resolving the identity of *Leptothrix ochracea* using single cell genomics, pyrosequencing and fish. *PLoS ONE* **6**: e17769.
- Gil, R., F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A.

- Latorre, C. Rausell, J. Kamerbeek, J. Gadau, B. Holdobler, R. C. van Ham, R. Gross, and A. Moya. 2003. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. USA* **100**: 9388–9393.
- Habyarimana, F., S. Al-Khodori, A. Kalia, J. E. Graham, C. T. Price, M. T. Garcia, and Y. A. Kwaik. 2008. Role for the ankyrin eukaryotic-like genes of *legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ. Microbiol.* **10**: 1460–1474.
- Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson, and E. F. DeLong. 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. USA* **103**: 18296–18301.
- Harrington, E. D., M. Arumugam, J. Raes, P. Bork, and D. A. Relman. 2010. SmashCell: a software framework for the analysis of single-cell amplified genome sequences. *Bioinformatics* **26**: 2979–2980.
- Hentschel, U., K. M. Usher, and M. W. Taylor. 2006. Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.* **55**: 167–177.
- Hentschel, U., J. Piel, S. M. Degnan, and M. W. Taylor. 2012. Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* (In press).
- Hongoh, Y. 2010. Diversity and genomes of uncultured microbial symbionts in the termite gut. *Biosci. Biotechnol. Biochem.* **74**: 1145–1151.
- Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma. 2008. Complete genome of the uncultured termite group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. USA* **105**: 5555–5560.
- Husseneder, C. 2010. Symbiosis in subterranean termites: a review of insights from molecular studies. *Environ. Entomol.* **39**: 378–388.
- Hutchison, C. A., and J. C. Venter. 2006. Single-cell genomics. *Nat. Biotech.* **24**: 657–658.
- Ishoey, T., T. Woyke, R. Stepanauskas, M. Novotny, and R. S. Lasken. 2008. Genomic sequencing of single microbial cells from environmental samples. *Curr. Opin. Microbiol.* **11**: 198–204.
- Kalisky, T., and S. R. Quake. 2011. Single-cell genomics. *Nat. Methods* **8**: 311–314.
- Kennedy, J., B. Flemer, S. A. Jackson, D. P. H. Lejon, J. P. Morrissey, F. O’Gara, and A. D. Dobson. 2010. Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Mar. Drugs* **8**: 608–628.
- Klitgaard, K., L. Molbak, T. K. Jensen, C. F. Lindboe, and M. Boye. 2005. Laser capture microdissection of bacterial cells targeted by fluorescence in situ hybridization. *BioTechniques* **39**: 864–868.
- Konstantinidis, K. T., A. Ramette, and J. M. Tiedje. 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl. Environ. Microbiol.* **72**: 7286–7293.
- Kuwahara, T., Y. Ogura, K. Oshima, K. Kurokawa, T. Ooka, H. Hirakawa, T. Itoh, H. Nakayama-Imaohji, M. Ichimura, K. Itoh, C. Ishifune, Y. Maekawa, K. Yasutomo, M. Hattori, and T. Hayashi. 2011. The lifestyle of the segmented filamentous bacterium: a non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing. *DNA Res.* **18**: 291–303.
- Kvist, T., B. K. Ahring, R. S. Lasken, and P. Westermann. 2007. Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* **74**: 926–935.
- Lafi, F. F., J. A. Fuerst, L. Fieseler, C. Engels, W. W. L. Goh, and U. Hentschel. 2009. Widespread distribution of *Poribacteria* in demospongiae. *Appl. Environ. Microbiol.* **75**: 5695–5699.
- Lasken, R., A. Raghunathan, T. Kvist, T. Ishoey, P. Westermann, B. K. Ahring, and R. Boissy. 2005. Multiple displacement amplification from single bacterial cells. Pp. 119–147 in *Whole Genome Amplification*, S. Hughes and R. Lasken, eds. Scion Publishing, Oxfordshire, United Kingdom.
- Lasken, R. S. 2007. Single-cell genomic sequencing using multiple displacement amplification. *Curr. Opin. Microbiol.* **10**: 510–516.
- Lasken, R. S., and T. B. Stockwell. 2007. Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol.* **7**: 19.
- Liu, M., L. Fan, L. Zhong, S. Kjelleberg, and T. Thomas. 2012. Metaproteogenomic analysis of a community of sponge symbionts. *ISME J.* **6**: 1515–1525.
- Liu, M. Y., S. Kjelleberg, and T. Thomas. 2011. Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. *ISME J.* **5**: 427–435.
- Liu, P., R. J. Meagher, Y. K. Light, S. Yilmaz, R. Chakraborty, A. P. Arkin, T. C. Hazen, and A. K. Singh. 2011. Microfluidic fluorescence in situ hybridization and flow cytometry (mufflowfish). *Lab Chip* **11**: 2673–2679.
- Malachowa, N., P. L. Kohler, P. M. Schlievert, O. N. Chuang, G. M. Dunny, S. D. Kobayashi, J. Miedzobrodzki, G. A. Bohach, and K. S. Seo. 2011. Characterization of a *Staphylococcus aureus* surface virulence factor that promotes resistance to oxidative killing and infectious endocarditis. *Infect. Immun.* **79**: 342–352.
- Marcy, Y., T. Ishoey, R. S. Lasken, T. B. Stockwell, B. P. Walenz, A. L. Halpern, K. Y. Beeson, S. M. Goldberg, and S. R. Quake. 2007a. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3**: 1702–1708.
- Marcy, Y., C. Ouverney, E. M. Bik, T. Losekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman, and S. R. Quake. 2007b. Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated tm7 microbes from the human mouth. *Proc. Natl. Acad. Sci. USA* **104**: 11889–11894.
- Mavromatis, K., C. K. Doyle, A. Lykidis, N. Ivanova, M. P. Francino, P. Chain, M. Shin, S. Malfatti, F. Larimer, A. Copeland *et al.* 2006. The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *J. Bacteriol.* **188**: 4015–4023.
- McCutcheon, J. P., and N. A. Moran. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. USA* **104**: 19392–19397.
- McCutcheon, J. P., B. R. McDonald, and N. A. Moran. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci. USA* **106**: 15394–15399.
- Moran, N. A., and P. Baumann. 2000. Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* **3**: 270–275.
- Moran, N. A., P. Tran, and N. M. Gerardo. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. *Appl. Environ. Microbiol.* **71**: 8802–8810.
- Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**: 165–190.
- Murray, J. D., R. R. Muni, I. Torres-Jerez, Y. Tang, S. Allen, M. Andriankaja, G. Li, A. Laxmi, X. Cheng, J. Wen *et al.* 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant J.* **65**: 244–252.
- Mussmann, M., F. Z. Hu, M. Richter, D. de Beer, A. Preisler, B. B. Jorgensen, M. Huntemann, F. O. Glockner, R. Amann, W. J. Koopman *et al.* 2007. Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol.* **5**: e230.
- O’Flaherty, S. J., and T. R. Klaenhammer. 2010. Functional and phenotypic characterization of a protein from *Lactobacillus acidophilus* involved in cell morphology, stress tolerance and adherence to intestinal cells. *Microbiology* **156**: 3360–3367.

- Ohkuma, M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol.* **16**: 345–352.
- Ohkuma, M., T. Sato, S. Noda, S. Ui, T. Kudo, and Y. Hongoh. 2007. The candidate phylum 'termite group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* **60**: 467–476.
- Pamp, S. J., E. D. Harrington, S. R. Quake, D. A. Relman, and P. C. Blainey. 2012. Single cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). *Genome Res.* DOI: 10.1101/gr.131482.111.
- Pan, X., A. E. Urban, D. Palejev, V. Schulz, F. Grubert, Y. Hu, M. Snyder, and S. M. Weissman. 2008. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proc. Natl. Acad. Sci. USA* **105**: 15499–15504.
- Peng, Y., H. C. Leung, S. M. Yiu, and F. Y. Chin. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**:1420–1428.
- Piel, J. 2011. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* **65**: 431–453.
- Podar, M., C. B. Abulencia, M. Walcher, D. Hutchison, K. Zengler, J. A. Garcia, T. Holland, D. Cotton, L. Hauser, and M. Keller. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* **73**: 3205–3214.
- Prakash, T., K. Oshima, H. Morita, S. Fukuda, A. Imaoka, N. Kumar, V. K. Sharma, S. W. Kim, M. Takahashi, N. Saitou et al. 2011. Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of th17 cell differentiation. *Cell Host Microbe* **10**: 273–284.
- Reading, N. C., and D. L. Kasper. 2011. The starting lineup: key microbial players in intestinal immunity and homeostasis. *Front. Microbiol.* **2**: 148.
- Redak, R. A., A. H. Purcell, J. R. Lopes, M. J. Blua, R. F. Mizell 3rd, and P. C. Andersen. 2004. The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. *Annu. Rev. Entomol.* **49**: 243–270.
- Rodrigue, S., R. R. Malmstrom, A. M. Berlin, B. W. Birren, M. R. Henn, and S. W. Chisholm. 2009. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS ONE* **4**: e6864.
- Scharf, M. E., Z. J. Karl, A. Sethi, and D. G. Boucias. 2011. Multiple levels of synergistic collaboration in termite lignocellulose digestion. *PLoS ONE* **6**: e21709.
- Schmitt, S., U. Hentschel, and M. W. Taylor. 2011. Deep sequencing reveals diversity and community structure of complex microbiota in five mediterranean sponges. *Hydrobiologia* **687**: 341–351.
- Schmitt, S., P. Tsai, J. Bell, J. Fromont, M. Ilan, N. Lindquist, T. Perez, A. Rodrigo, P. J. Schupp, J. Vacelet, N. Webster, U. Hentschel, and M. W. Taylor. 2012. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J.* **6**: 564–576.
- Szesnak, A., N. Segata, X. Qin, D. Gevers, J. F. Petrosino, C. Huttenhower, D. R. Littman, and I. I. Ivanov. 2011. The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10**: 260–272.
- Siegl, A., and U. Hentschel. 2010. Pks and nrps gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ. Microbiol. Rep.* **2**: 507–513.
- Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel. 2011. Single-cell genomics reveals the lifestyle of poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* **5**: 61–70.
- Silander, K., and J. Saarela. 2008. Whole genome amplification with phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol. Biol.* **439**: 1–18.
- Simelyte, E., M. Rimpilainen, X. Zhang, and P. Toivanen. 2003. Role of peptidoglycan subtypes in the pathogenesis of bacterial cell wall arthritis. *Ann. Rheum. Dis.* **62**: 976–982.
- Stepanauskas, R., and M. E. Sieracki. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl. Acad. Sci. USA* **104**: 9052–9057.
- Stingl, U., R. Radek, H. Yang, and A. Brune. 2005. "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71**: 1473–1479.
- Swan, B. K., M. Martinez-Garcia, C. M. Preston, A. Sczyrba, T. Woyke, D. Lamy, T. Reinthaler, N. J. Poulton, E. D. Masland, M. L. Gomez et al. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.
- Takiya, D. M., P. L. Tran, C. H. Dietrich, and N. A. Moran. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. *Mol. Ecol.* **15**: 4175–4191.
- Tamas, I., L. Klasson, B. Canbäck, A. K. Näslund, A.-S. Eriksson, J. J. Wernegreen, J. P. Sandström, N. A. Moran, and S. G. E. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**: 2376–2379.
- Tamminen, M., and M. Virta. 2010. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Nature Precedings* DOI:http://hdl.handle.net/10101/npre.2010.4761.1
- Taylor, M. W., R. Radax, D. Steger, and M. Wagner. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* **71**: 295–347.
- Thomas, T., D. Rusch, M. Z. Demaree, P. Y. Yung, M. Lewis, A. Halpern, K. B. Heidelberg, S. Egan, P. D. Steinberg, and S. Kjelleberg. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J.* **4**: 1557–1567.
- Voth, D. E., D. Howe, P. A. Beare, J. P. Vogel, N. Unsworth, J. E. Samuel, and R. A. Heinzen. 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J. Bacteriol.* **191**: 4232–4242.
- Walker, T., L. Klasson, M. Sebaihia, M. J. Sanders, N. R. Thomson, J. Parkhill, and S. P. Sinkins. 2007. Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol.* **5**: 39.
- Warnecke, F., and P. Hugenholz. 2007. Building on basic metagenomics with complementary technologies. *Genome Biol.* **8**: 231.
- Webster, N. S., and M. W. Taylor. 2012. Marine sponges and their microbial symbionts: love and other relationships. *Environ. Microbiol.* **14**: 335–346.
- Webster, N. S., M. W. Taylor, F. Behnam, S. Lucker, T. Rattei, S. Whalan, M. Horn, and M. Wagner. 2010. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ. Microbiol.* **12**: 2070–2082.
- Weng, J. K., X. Li, N. D. Bonawitz, and C. Chapple. 2008. Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. *Curr. Opin. Biotechnol.* **19**: 166–172.
- Wilson, A. C. C., P. D. Ashton, F. Calevro, H. Charles, S. Colella, G. Febvay, G. Jander, P. F. Kushlan, S. J. Macdonald, J. F. Schwartz, G. H. Thomas, and A. E. Douglas. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.* **19**: 249–258.
- Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J.

- Shapiro *et al.* 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955.
- Woyke, T., G. Xie, A. Copeland, J. M. González, C. Han, H. Kiss, J. H. Saw, P. Senin, C. Yang, S. Chatterji *et al.* 2009. Assembling the marine metagenome, one cell at a time. *PLoS ONE* **4**: e5299
- Woyke, T., D. Tighe, K. Mavromatis, A. Clum, A. Copeland, W. Schackwitz, A. Lapidus, D. Wu, J. P. McCutcheon, B. R. McDonald, N. A. Moran, J. Bristow, and J. F. Cheng. 2010. One bacterial cell, one complete genome. *PLoS ONE* **5**: e10314.
- Woyke, T., A. Sczyrba, J. Lee, C. Rinke, D. Tighe, S. Clingenpeel, R. Malmstrom, R. Stepanauskas, and J.-F. Cheng. 2011. Decontamination of mda reagents for single cell whole genome amplification. *PLoS ONE* **6**: e26161.
- Wu, D., S. C. Daugherty, S. E. Van Aken, G. H. Pai, K. L. Watkins, H. Khouri, L. J. Tallon, J. M. Zaborsky, H. E. Dunbar, P. L. Tran, N. A. Moran, and J. A. Eisen. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**: e188.
- Wu, H. J., Ivanov, II, J. Darce, K. Hattori, T. Shima, Y. Umesaki, D. R. Littman, C. Benoist, and D. Mathis. 2010. Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32**: 815–827.
- Yilmaz, S., and A. K. Singh. 2011. Single cell genome sequencing. *Curr. Opin. Biotechnol.* DOI 10.1016/j.copbio.2011.11.018.
- Yilmaz, S., M. F. Haroon, B. A. Rabkin, G. W. Tyson, and P. Hugenholtz. 2010. Fixation-free fluorescence in situ hybridization for targeted enrichment of microbial populations. *ISME J.* **4**: 1352–1356.
- Youssef, N. H., P. C. Blainey, S. R. Quake, and M. S. Elshahed. 2011. Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone spring, OK, USA). *Appl. Environ. Microbiol.* **77**: 7804–7814.
- Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church. 2006. Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* **24**: 680–686.