Exploring Symbioses by Single-Cell Genomics

JANINE KAMKE1, KRISTINA BAYER1, TANJA WOYKE2, AND UTE HENTSCHEL1,*

1Julius-von-Sachs Institute for Biological Sciences, University of Würzburg, Julius-von-Sachs Platz 3, 97082 Würzburg, Germany; and 2Department 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 sequences 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-
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 (Ishoey 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; Mussmann 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 workflow (Stepanauskas 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

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Potential Solution</th>
</tr>
</thead>
</table>
| Contamination            | ● 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.e., after micromanipulation) |
| Limited lysis            | ● Remove known contaminants (i.e., Delftia, human) by binning methods post-MDA |
| Amplification bias       | ● Chemically lyse via alkaline solution (KOH)  
|                          | ● Enzymatically lyze using lysozyme and/or proteases or custom-made enzyme cocktails |
| Chimerism                | ● Freeze/thaw, heat, etc. |
| Fragmented and partial nature of genomes | ● 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.
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, a-chromopeptidase, 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 µl) 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 (Woyke 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
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

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

Intracellular bacterial symbiont of a sharpshooter

“Candidatus S. muelleri” and “Candidatus Baumannia cicadellinicola” 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 adaption 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 (Simelyle 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 pentode 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. 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). MCRA 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 chloymphycin 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 presence of ruberythrin 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 cosm id 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 TP A5 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


Gil, R., F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A.


