Packaging of *Dinoroseobacter shibae* DNA into Gene Transfer Agent Particles Is Not Random

Jürgen Tomasch[1,3,*], Hui Wang[1], April T.K. Hall[2], Diana Patzelt[1,6], Matthias Preusse[3], Jörn Petersen[4], Henner Brinkmann[4], Boyke Bunk[4], Sabin Bhuju[5], Michael Jarek[5], Robert Geffers[5], Andrew S. Lang[2], and Irene Wagner-Döbler[1]

[1]Group Microbial Communication, Helmholtz-Centre for Infection Research, Braunschweig, Germany

[2]Department of Biology, Memorial University of Newfoundland, St John's, Newfoundland and Labrador, Canada

[3]Department of Molecular Bacteriology, Helmholtz-Centre for Infection Research, Braunschweig, Germany

[4]Department of Microbial Ecology and Diversity Research, Leibniz Institute DSMZ—German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

[5]Group Genome Analytics, Helmholtz-Center for Infection Research, Braunschweig, Germany

[6]Present address: Sartorius Stedim Biotech GmbH, Göttingen, Germany

***Corresponding author:** E-mail: juergen.tomasch@helmholtz-hzi.de.

Accepted: January 8, 2018

Abstract

Gene transfer agents (GTAs) are phage-like particles which contain a fragment of genomic DNA of the bacterial or archaeal producer and deliver this to a recipient cell. GTA gene clusters are present in the genomes of almost all marine *Rhodobacteraceae* (Roseobacters) and might be important contributors to horizontal gene transfer in the world's oceans. For all organisms studied so far, no obvious evidence of sequence specificity or other nonrandom process responsible for packaging genomic DNA into GTAs has been found. Here, we show that knock-out of an autoinducer synthase gene of *Dinoroseobacter shibae* resulted in overproduction and release of functional GTA particles (DsGTA). Next-generation sequencing of the 4.2-kb DNA fragments isolated from DsGTAs revealed that packaging was not random. DNA from low-GC conjugative plasmids but not from high-GC chromids was excluded from packaging. Seven chromosomal regions were strongly overrepresented in DNA isolated from DsGTA. These packaging peaks lacked identifiable conserved sequence motifs that might represent recognition sites for the GTA terminase complex. Low-GC regions of the chromosome, including the origin and terminus of replication, were underrepresented in DNA isolated from DsGTAs. DNA methylation reduced packaging frequency while the level of gene expression had no influence. Chromosomal regions found to be over- and underrepresented in DsGTA-DNA were regularly spaced. We propose that a "headful" type of packaging is initiated at the sites of coverage peaks and, after linearization of the chromosomal DNA, proceeds in both directions from the initiation site. GC-content, DNA-modifications, and chromatin structure might influence at which sides GTA packaging can be initiated.

Key words: gene transfer agent, GTA, Roseobacter, horizontal gene transfer.

Introduction

Prokaryotes, even different strains of one species, often differ remarkably in their genomic inventories. This variability in genome content cannot be explained by gene loss alone (Dagan and Martin 2007) but is the result of a considerable amount of horizontal gene transfer (HGT), which has been shown to be a main driver of prokaryote evolution (McInerney et al. 2017). Mechanisms for HGT in prokaryotes include transformation, that is, direct uptake of naked DNA from the environment, conjugational transfer of plasmids that requires direct contact of donor and recipient cells (Thomas and Nielsen 2005), and transduction by phages (Canchaya et al. 2003). In 1974 an additional, unique way for HGT of genomic DNA, independent of cell–cell contact or the presence of naked DNA in the environment, was discovered in *Rhodobacter capsulatus* (Marrs 1974). Later, the responsible *R. capsulatus* gene transfer agent (RcGTA) was found to resemble a small tailed phage that contains fragments of cellular genomic DNA instead of viral DNA

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

(Yen et al. 1979). Since then, Gene transfer agent (GTA) particles of different evolutionary origins have been identified in various other bacteria (Humphrey et al. 1997; Biers et al. 2008) and the archaeon *Methanococcus voltae* (Bertani 1999).

In *Alphaproteobacteria*, two phylogenetically unrelated types of GTAs have been identified, namely the *Rhodobacterales* GTA, with the best studied RcGTA, and the *Bartonella* GTA (BaGTA). Those related to RcGTA have been identified in many *Alphaproteobacteria* and are particularly widespread in members of the order *Rhodobacterales* where they are organized in a conserved gene cluster of approximately 15 kb (Lang and Beatty 2007; Newton et al. 2010; Huang et al. 2011). As RcGTA packages only approximately 4 kb of DNA it has to be considered non-self-transmissible (Lang et al. 2017). GTAs isolated from *Rhodobacteraceae* of the Roseobacter group transferred resistance markers with high frequencies to both cultures and natural communities of marine bacteria, suggesting an important role of GTAs for the gene flow in the world's oceans (McDaniel et al. 2010). Furthermore, phylogenomic analysis identified GTAs as the main source of genes acquired through HGT in members of the genus *Phaeobacter*, consisting mainly of marine surface colonizers (Freese et al. 2017).

The *Rhodobacterales* GTAs probably evolved from a prophage of the *Siphoviridae* family that lost genes for integration into and excision from the host DNA as well as for replication, whereas genes for DNA packaging and structural components were retained (Huang et al. 2011), and these genes were subsequently brought under the control of the host gene regulatory systems (Mercer et al. 2012). In *R. capsulatus*, genes coding for the competence system, needed for uptake of DNA from GTAs by the receiving cell, are under control of the same regulators as the GTA genes (Brimacombe et al. 2015). Recently it has been found that the RcGTA is induced by nutrient depletion, possibly as a consequence of increased levels of guanosine-tetraphosphate (ppGpp) (Westbye et al. 2017). Notably, only a small subset of a *R. capsulatus* culture (<3%) expresses RcGTA genes, leading to lysis of the producing cells (Fogg et al. 2012; Hynes et al. 2012).

A second phylogenetically distinct alphaproteobacterial GTA has been found in the intracellular bacteria from the genus *Bartonella* (Berglund et al. 2009; Tamarit et al. 2017). As in *R. capsulatus*, BaGTA production is limited to a small fraction of the population and these producers lyse to release GTA particles (Quebatte et al. 2017). The authors suggested that the fittest subpopulation of *B. henselae* cells, indicated by low levels of ppGpp, sacrifices itself to spread beneficial mutations via GTAs throughout the population of actively dividing cells.

The mechanism of packaging DNA into GTA particles is currently not known. Based upon homology of key components like terminase and portal proteins it is assumed that it resembles the "headful" type of packaging that is used by many double-stranded DNA phages (Lang et al. 2017). The parts of the genomes that are mobilized by packaging into GTAs have been determined for *R. capsulatus* as well as *B. grahamii* and *B. henselae*. These studies suggest random packaging of host genomic DNA into GTAs.

Marker transfer and DNA digestion analysis in the pregenomic era suggested random packaging of genomic DNA into RcGTAs (Marrs 1974; Humphrey et al. 1997; Bertani 1999). These initial findings later gained support by a study employing whole-genome microarray hybridization of RcGTA-DNA, which showed that only the GTA gene cluster itself was underrepresented, by 25%, in packaged DNA, whereas the remaining genomic DNA was randomly packaged (Hynes et al. 2012). Since the GTA cluster is highly transcribed in the subset of the population that is responsible for producing the GTA particles, it was hypothesized that access to the GTA encoding DNA was blocked by the RNA polymerase complex (Hynes et al. 2012).

For BaGTA, a virulence region important for host-adaptability of the mouse pathogen *B. grahamii* was found to be overrepresented in GTA-DNA (Berglund et al. 2009). Genomic DNA of this region was present in higher copy number in the cells due to run-off replication from a phage-derived origin of replication, thus packaging in this case can be considered random but dependent on the copy number of the DNA (Berglund et al. 2009). A recent transposon sequencing study in *B. henselae* confirmed the relationship between DNA copy number and frequency of packaging of chromosomal DNA but additionally showed that DNA from plasmids is not transferred by BaGTAs (Quebatte et al. 2017).

The genome of the dinoflagellate-associated bacterium *Dinoroseobacter shibae* consists of a 3.79-Mb chromosome and five extrachromosomal replicons (Wagner-Döbler et al. 2010). Two of them (153 and 72 kb) have a similar high GC-content (65–68%) and codon-usage as the chromosome and are therefore classified as "chromids"' according to Harrison et al. (2010). The other three (191, 126, and 86 kb) have a lower GC-content (60–61%) and deviating codon-usage and represent typical plasmids (Petersen et al. 2013). The 191- and 126-kb plasmids harbor type IV secretion systems that are employed for their conjugational transfer (Patzelt et al. 2016). A complete GTA gene cluster is present on the *D. shibae* chromosome (Wang et al. 2014). Similar to *R. capsulatus* (Mercer et al. 2012), the *D. shibae* GTA (DsGTA) gene cluster is under the control of the CtrA phosphorelay (Wang et al. 2014). This signal-transduction system is integrated into the *D. shibae* quorum sensing (QS) system as it is induced by the product of the autoinducer synthase $LuxI_1$ (Patzelt et al. 2013) and then induces expression of two additional QS response regulator/autoinducer synthase operons, $luxR/I_2$, and $luxB/I_3$ (Wang et al. 2014). Here, we show that knockout of the autoinducer synthase gene $luxI_2$ results in overproduction of DsGTAs. Using next-generation sequencing, we analyze packaging of *D. shibae* DNA into GTA particles.

Materials and Methods

Bacterial Strains and Plasmids

Dinoroseobacter shibae strains used in this study, *E. coli* strains, plasmids and primers used for construction of the *D. shibae* Δ*luxI*$_2$ (Dshi_2852) deletion strain are listed in supplementary table S1, Supplementary Material online. *D. shibae* Δ*luxI*$_2$ was generated by replacing the gene with a gentamycin resistance cassette via double homologous recombination as previously described (Patzelt et al. 2013). The knock-out vector pJB5603*luxI*$_2$:: Gm^r was subcloned in *E. coli* ST18 and introduced into *D. shibae* DFL12 via conjugation. Resequencing of the knock-out strain confirmed deletion of the wild-type gene, insertion of the resistance cassette into the correct locus and identified one secondary mutation. Details on knock-out construction and confirmation can be found in supplementary methods, note S1 and figure S1, Supplementary Material online.

Cultivation

Dinoroseobacter shibae strains were grown in 1/2 Difco Marine Broth 2216 (BD, Sparks, MD) or artificial seawater medium (ASM, supplementary methods) at 30 °C and 160 rpm. *Ruegeria pomeroyi* was grown in 1/2 yeast extract-tryptone-sea salt (YTSS) medium (4 g l^{-1} tryptone, 2.5 g l^{-1} yeast extract, and 15 g l^{-1} sea salts) at 20 °C and 160 rpm.

Detection of the GTA Major Capsid Protein by Western Blot

Dinoroseobacter shibae strains were grown for 19 h in ASM. Samples from *Ruegeria pomeroyi*, grown in 1/2 YTSS for 19 h, were included as positive controls as this organism is known to produce GTA particles (Biers et al. 2008) and has detectable GTA capsid protein (Fu et al. 2010). Cells from 0.4 ml of each culture were pelleted and resuspended in an equal volume of TE buffer. Culture supernatants were prepared from cultures by centrifugation for 2 min at 17,000 × g, with removal of a fraction of the resulting supernatant to a new tube. The samples (5 and 10 μl for cells and supernatants, respectively) were mixed with SDS-PAGE sample buffer (NEB) and incubated 5 min at 98 °C before electrophoresis. SDS-PAGE and immunoblotting were performed as described previously (Fu et al. 2010) using the primary antibody targeting a region of the GTA major capsid protein (Agrisera AB, Vännäs, Sweden) that is conserved among many *Rhodobacterales*. The images were captured using the ImageQuant LAS 4000 instrument (GE Life Sciences, Uppsala, Sweden) and were each uniformly adjusted for brightness and contrast.

Gene Transfer Bioassays

Gene transfer agent donor strains were grown for 24 h in ASM at 30 °C and 160 rpm. The recipient strain, *D. shibae* DFL12, was grown under the same conditions in 1/2 Mb. The donor cultures were passed through 0.22-μm filters and 0.5 ml of this filtrate was added to 0.1 ml of the recipient cells. The mixture was then incubated at 30 °C for 1 h without shaking, for 4 h and 160 rpm, followed by addition of 0.9 ml ASM and a further incubation at 30 °C and 160 rpm for 18 h. The cells were then plated on 1/2 Mb with gentamycin sulfate (150 μg ml^{-1}) and incubated at 30 °C for 2–4 days until colonies were visible for counting. The negative controls for these assays were filtrates of the gentamycin-sensitive strain, DFL12, and the gentamycin-resistant strain, Δ*luxI*$_1$, added to the recipient cells. Filtrate-only controls that contained no recipient cells were also included to confirm no cells passed through the 0.22-μm filters.

Purification of DsGTA Particles for Electron Microscopy

Dinoroseobacter shibae Δ*luxI*$_2$ was grown 1/2 Mb for 29 h at 30 °C with shaking at 160 RPM. The cells were pelleted by centrifugation and NaCl was added to the supernatant to a final concentration of 1 M. Polyethylene glycol (PEG) 6000 was added to make a final concentration of 10% w/v and dissolved by stirring at room temperature. The solution was then incubated at 4 °C for 18 h and the precipitated material pelleted by centrifugation at 10,000 × g for 20 min at 20 °C. The supernatant was poured off and pelleted material was resuspended in 0.1 M ammonium acetate (pH 7) with shaking at room temperature for 2 h followed by gentle pipetting. This material was centrifuged at 2,400 × g for 2 min at room temperature and the supernatant collected into a new tube. Triton X-100 was added to a final concentration of 10% (v/v) with gentle mixing by inversion. This was centrifuged at 1,10,000 × g for 2 h at 20 °C, and most of the supernatant then removed. The pellet was resuspended in the remaining solution (~0.5 ml) by gentle pipetting and the tube refilled with 0.1 M ammonium acetate. The tube was centrifuged again for 1 h and the previous steps repeated. After this third centrifugation, almost all of the supernatant was removed and the pellet left to resuspend in the small amount of remaining liquid with shaking at 4 °C for 18 h. The material was homogenized by gentle pipetting, transferred to a microfuge tube, and centrifuged at 17,000 × g for 1 min at room temperature. The supernatant was collected for electron microscopy imaging at the University of Guelph (Canada) Molecular and Cellular Imaging Facility. A 5-μl sample was placed onto a 200 mesh copper grid with formvar/carbon coating, adsorbed for 1 min, and blotted off with filter paper. The grid was then floated on a drop of 2% uranyl acetate. The sample was viewed in a FEI Tecnai F20 electron microscope at 200 kV. The images were collected with a Gatan 4 K bottom-mount camera using the Gatan DigitalMicrograph software.

Isolation of DsGTA-DNA

DsGTA-DNA was isolated from three independent cultures. The cultures were treated as for the Electron microscopy up to

the point of resuspension of the PEG precipitates. In this case, the pellets were resuspended in TE buffer with gentle shaking at room temperature for 2 h. Free nucleic acids were removed by treatment with DNase I (0.01 U μl^{-1}) and RNase A (0.08 U μl^{-1}) at 37 °C in 1× DNase buffer (New England Biolabs) for 1 h. The sample was heated at 75 °C for 15 min after addition of EDTA to a final concentration of 5 mM. Nucleic acids were purified by phenol: chloroform: isoamyl alcohol (25:24:1) extraction and ethanol precipitation. The purified DNA was subjected to restriction enzyme digestion using the enzymes EcoRI and BfaI (New England Biolabs), according to the manufacturer's protocol. The untreated DNA sample was incubated under the same conditions (i.e., in 1× reaction buffer at 37 °C) with the addition of dH_2O in place of enzyme. The approximately 4.2-kb band of DsGTA-DNA was excised from the gel and purified with the Wizard SV Gel and PCR Clean-Up System (Promega, Fitchburg, WI), followed by ethanol precipitation.

Isolation of Genomic DNA

Genomic DNA was isolated from 5 ml samples of two of the *D. shibae* cultures used for isolation of DsGTA-DNA using the NucleoSpin tissue kit (Machery-Nagel, Düren, Germany) according to manufacturer's protocols.

Isolation of RNA

RNA for transcriptome analysis of the three different *D. shibae* strains was isolated from two independent cultures. RNA for analysis of the relationship of DsGTA packaging and transcription was isolated from two cultures from which DsGTA-DNA was isolated. Cells were collected by centrifugation at 12,000 × g for 1 min at 4 °C, covered with 1 ml Trizol (Ambion, Germany), immediately frozen in liquid N_2 and stored at −70 °C until processing. For RNA extraction, cells were homogenized with ~0.3 g of glass beads in the FastPrep-24 instrument (MP Biomedicals, CA) at 6.0 m/s for 3 min and then incubated for 5 min at room temperature. Samples were centrifuged at 12,000 × g for 10 min at 4 °C and the supernatants were transferred to fresh tubes, followed by addition of 100 µl of 1-bromo-3-chloropropane (Sigma-Aldrich, St. Louis, MO) and incubation for 10 min at room temperature. Samples were centrifuged at 12,000 × g for 10 min at 4 °C, after which the aqueous phase was transferred to new tubes and mixed with 500 µl of absolute ethanol. Extracts were purified using RNeasy Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. In addition, samples were treated with DNAse I (Qiagen, Hilden, Germany).

Sequencing of RNA

Total RNA was treated with RiboZero (Bacteria) kit (Illumina, San Diego, CA) following the manufacturer's protocol in order to remove ribosomal RNA from the samples. Single end, strand specific cDNA libraries were prepared from rRNA depleted total RNA using Scriptseq v2 RNA-Seq Library Preparation Kit (Illumina) following the manufacturers protocol. For sequencing equal volume of libraries (12 PM) was multiplexed on a single lane. Sequencing was done on the HiSeq 2500 (Illumina) using TruSeq SBS Kit v3—HS (Illumina) for 50 cycles resulting in 50-bp reads. Image analysis and base calling were performed using the Illumina pipeline v 1.8 (Illumina).

Sequencing of Genomic and DsGTA-DNA

Paired end cDNA libraries were prepared using the NEBNext Ultra DNA Library Preparation Kit for Illumina (NEB, Ipswich, MA). Sequencing was done on the MiSeq 2000 (Illumina) using MiSeq Reagent Kits v2 (Illumina) for 250 cycles resulting in 2×250 bp paired end reads. Image analysis and base calling were performed using the Illumina pipeline v 1.8 (Illumina).

PacBio-Sequencing and DNA Modification Calling

PacBio-sequencing was performed as previously described (Bartling et al. 2017). Methylome analysis was performed using the "RS_Modification_and_Motif_Analysis.1" protocol included in SMRT Portal version 2.3.0. 99.5 and 96.6% of the identified m6A and m4C motifs were methylated, respectively.

Processing of Sequence Data

The demultiplexed raw fastq-files were quality-controlled using the FASTQ-mcf suite (https://github.com/ExpressionAnalysis/ea-utils; last accessed January 15, 2018). Low quality bases (Phred-score < 30) and identified Illumina adaptors were clipped from the sequences. Reads were mapped to reference genomes using bowtie2 (Langmead and Salzberg 2012) with default parameters for single or paired end reads. Ambiguously mapping reads were randomly distributed between all regions to which they could be assigned. The resulting sam-files were converted to indexed binary format and pile-up format using samtools (Li et al. 2009). Accession numbers of reference *D. shibae* DNA sequences: chromosome, 3.79 Mb [NC_009952.1]; pDSHI01, 191 kb [NC_009955.1]; pDSHI02, 153 kb [NC_009956.1]; pDSHI03, 126 kb [NC_009957.1]; pDSHI04, 86 kb [NC_009958.1]; pDSHI05, 72 kb [NC_009959.1].

Statistical Analysis

Pile-up files were loaded into the R statistical environment (R version 3.4.0). Reproducibility of biological replicates was assessed by visual inspection of scatterplots and boxplots and by calculating Spearman rank correlation (supplementary figs. S7–S9, Supplementary Material online). Distribution of DsGTA and genomic coverage was compared with theoretical data following a normal distribution using the qqnorm()

function in R. The mean coverage by DsGTA-DNA, genomic DNA, and RNA as well as the mean GC content were calculated for sliding windows of 4 kb using the rollapply function of the zoo package (Zeileis and Grothendieck 2005). GC-content for sliding windows was calculated using the letterFrequency function of the Biostrings package (Pagès et al. 2017). Only methylated sites with an identification Phred score > 30 (99.9% probability of correct identification) were used in this study and summed up for the sliding window range. Spearman rank correlation was calculated for DsGTA coverage with GC-content and $\log_2$-transformed RNA coverage, respectively. The autocorrelation function, available from the R stats package, calculates the Pearson correlation from all pairs of observations that have the same distance from each other. This is done for increasing distances and makes it possible to find distances at which the correlation of two observations (in this case DsGTA coverage) peaks, indicating periodicity of the data. Cumulative GC/AT-skew for identification of the origin and terminus of replication has been performed using the oriloc function of the seqinr package (Charif and Lobry 2007).

Code Availability

Linux shell and R scripts for analysis are available at https://github.com/Juergent79/GTA; last accessed January 15, 2018.

Identification of Phage Genes

The tool PHAST (Zhou et al. 2011) was used to identify phage derived genes in the genome of *D. shibae*.

DNA Motif Discovery

DNA sequence of the seven DsGTA coverage peaks was used as primary input for the motif discovery algorithm of the MEME suite 4.12.0 (Bailey et al. 2009). Motif discovery was run in the discriminative mode, with the seven peak regions indicated in figure 3*A* as query input and either the whole chromosomal DNA sequence or DNA sequences of chromosomal regions with coverage between 200 and 500 reads/nt as controls (secondary) input. The second control avoided inclusion of low GC regions that showed an extremely low coverage.

Data Availability

RNA, genomic, and DsGTA Illumina sequencing data have been made publicly available at the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena; last accessed January 15, 2018) under the project accession number PRJEB20656. PacBio data used for evaluation of *D. shibae* $\Delta luxI_2$ (supplementary fig. S1 and note S1, Supplementary Material online) has not been deposited in a public database yet but will be available from the authors upon request.

Results

Only weak expression of the DsGTA gene cluster could be observed in wild type *D. shibae* and virtually no expression was found in the $luxI_1$ knockout. However, when we knocked out the CtrA-controlled $luxI_2$ gene (supplementary fig. S1 and note S1, Supplementary Material online), strong expression of the DsGTA gene cluster was observed (fig. 1*A*). The DsGTA major capsid protein could accordingly be detected in cell extracts and supernatants of *D. shibae* $\Delta luxI_2$, indicating that this strain produces and releases GTA particles. A weak signal was also observed in cell-extracts, but not supernatants of the wild type, indicating a low GTA production rate in the native strain under the conditions studied (fig. 1*B*). The gentamicin resistance cassette used as the marker for knock-out construction could be successfully transferred by cell-free supernatants of *D. shibae* $\Delta luxI_2$, but not $\Delta luxI_1$, to the wild type demonstrating that DsGTAs are functional and that *D. shibae* cells are able to receive and incorporate DNA from the GTAs (fig. 1*C*). A gentamycin-resistant strain obtained via GTA-mediated marker transfer also produced and released increased numbers of GTAs relative to the wild type strain, validating the role of the $luxI_2$ gene in repressing DsGTA production (fig. 1*D*). DsGTA resembles a small tailed phage with a head diameter of approximately 33 nm and a tail length of approximately 48 nm (fig. 1*E*). It packages DNA of 4.2 kb in size, similar to RcGTA-DNA. Treatment with restriction enzymes produced smearing below this band, indicating that the DNA was double-stranded and that it was not a specific sequence of DNA (fig. 1*F*).

We used next-generation sequencing to analyze whether packaging of DNA into DsGTA particles is random, or shows preferences for distinct genomic loci. To this end, DsGTA-DNA, genomic DNA, and RNA from the same samples were sequenced and mapped on the *D. shibae* genome. This allowed us to determine the relative copy number of replicons, to address the possibility of run-off replication found in in *B. grahamii* and *B. henselae* (Berglund et al. 2009; Quebatte et al. 2017), and to study the relationship between transcriptional activity and packaging frequency similar to *R. capsulatus* (Hynes et al. 2012). Furthermore, we used PacBio sequencing data to determine methylation of the *D. shibae* DNA. We calculated the mean coverage for sliding windows of 4 kb, for the chromosome and extrachromosomal replicons. Coverage of the multipartite genome of *D. shibae* by DNA isolated from DsGTA particles showed a clear deviation from normal distribution and a long tail for lower coverage (fig. 2*A*). Coverage by genome sequencing of the according strains mostly followed a normal distribution, with only few regions being over or underrepresented (supplementary fig. S2, Supplementary Material online).

Coverage distribution differed for the six *D. shibae* replicons (fig. 2*B* and supplementary table S1, Supplementary Material online). The chromosome showed the highest

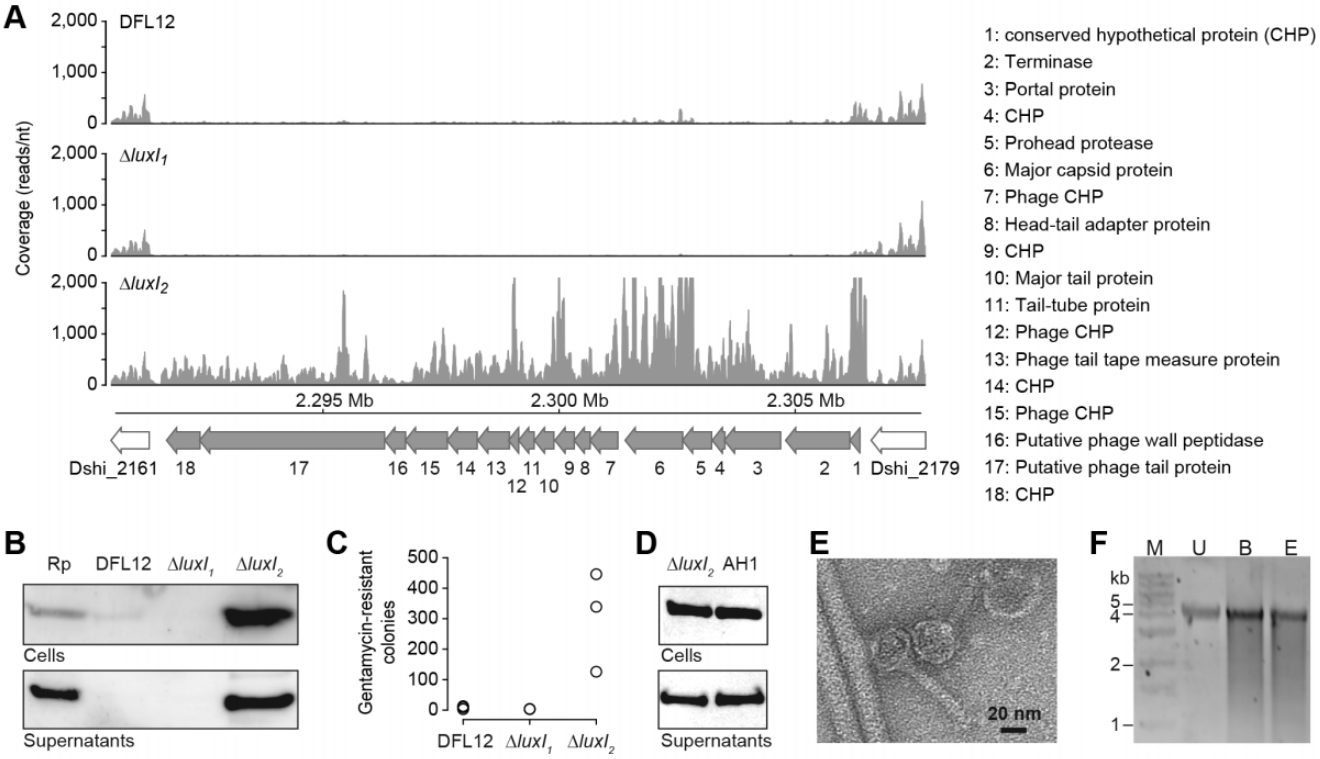

Fig. 1.—Biosynthesis of functional GTAs in *D. shibae* Δ*luxI*$_2$. (*A*) Expression of the DsGTA gene cluster in *D. shibae* DFL12, Δ*luxI*$_1$, and Δ*luxI*$_2$ as determined by strand-specific RNA-seq. Coverage of the negative strand is shown. DsGTA genes conserved in *Rhodobacterales* are displayed as grey arrows. Genes are numbered according to the reference *R. capsulatus*; predicted functions are indicated on the right. Reproducibility of RNA sequencing is shown in supplementary fig. S7, Supplementary Material online. (*B*) Detection of the GTA major capsid protein in cell extracts and culture supernatants of *Ruegeria pomeroyi* (Rp, positive control) and the aforementioned *D. shibae* strains using western blotting. (*C*) Transfer of a gentamycin-resistance cassette to *D. shibae* DFL12 by supernatants from different *D. shibae* strains. Individual data from three (DFL12, Δ*luxI*$_2$) and two (Δ*luxI*$_1$) independent experiments is shown. (*D*) Transfer of the gentamycin-resistance cassette disruption of *luxI*$_2$ via GTA results in overproduction of GTAs in the resulting recipient strain (AH1). (*E*) Electron micrograph showing two DsGTA particles. The filamentous structures are most likely flagella. (*F*) Visualization of the DNA isolated from GTAs: untreated DNA (U) and DNA digested with BamHI (*B*) and EcoRI (*E*) are shown.

variability (between 10 and 763 reads/nt, median: 181 reads/nt); the 72-kb chromid had a coverage similar to the chromosome (median: 192 reads/nt) while the 153-kb chromid showed the highest abundance within DsGTA-DNA (median: 1082 reads/nt). Coverage of the plasmids was low overall (median between 4 and 6 reads/nt), but indicated that at least parts of the 191- and 86-kb plasmids were present in DsGTA-DNA (maximum: 237 and 258 reads/nt, respectively) whereas the 126-kb plasmid was almost absent (maximum: 8 reads/nt). In the case of the 153-kb chromid, the high coverage could partly be explained by the 50% higher copy number of this replicon, whereas the copy number of all other replicons was close to one chromosome equivalent (fig. 2*C*). Next, we analyzed the coverage of each replicon in detail, starting with the extrachromosomal replicons. Differences in DNA content or a sequencing bias could be excluded as a source explaining coverage differences of chromids and plasmids (fig. 2*D*). Both chromids were completely, albeit not evenly, covered by GTA-DNA. In particular, coverage of the 153-kb chromid showed a conspicuous gap between 95 and 102 kb. The 126-kb plasmid showed virtually no mapping reads and thus can be considered not to be packaged into DsGTA particles. The 191 and 86 kb showed a similar low coverage except for one 7.5-kb region with very high coverage and sharp boundaries. We found that the DNA sequences of these regions were identical to the low coverage region on the 153-kb chromid. It contains a transposable element with the gene *thiC* in the center that has been spread to the different replicons (*thiC* locus, supplementary fig. S3*A*, Supplementary Material online). Thus, random mapping should lead to an approximately equal distribution of the ambiguous reads between the three paralogous target loci. Assigning all the reads to the locus on the 153-kb chromid would fill the gap on this chromid and resolve observed discrepancies of read mapping (supplementary fig. S3 and note S2, Supplementary Material online). Therefore, we conclude that the large majority of reads assigned to the *thiC* locus originated from the 153-kb chromid, and that plasmids are virtually excluded from being packaged into GTA particles. Plasmids showed a 60% and 40% higher frequency of

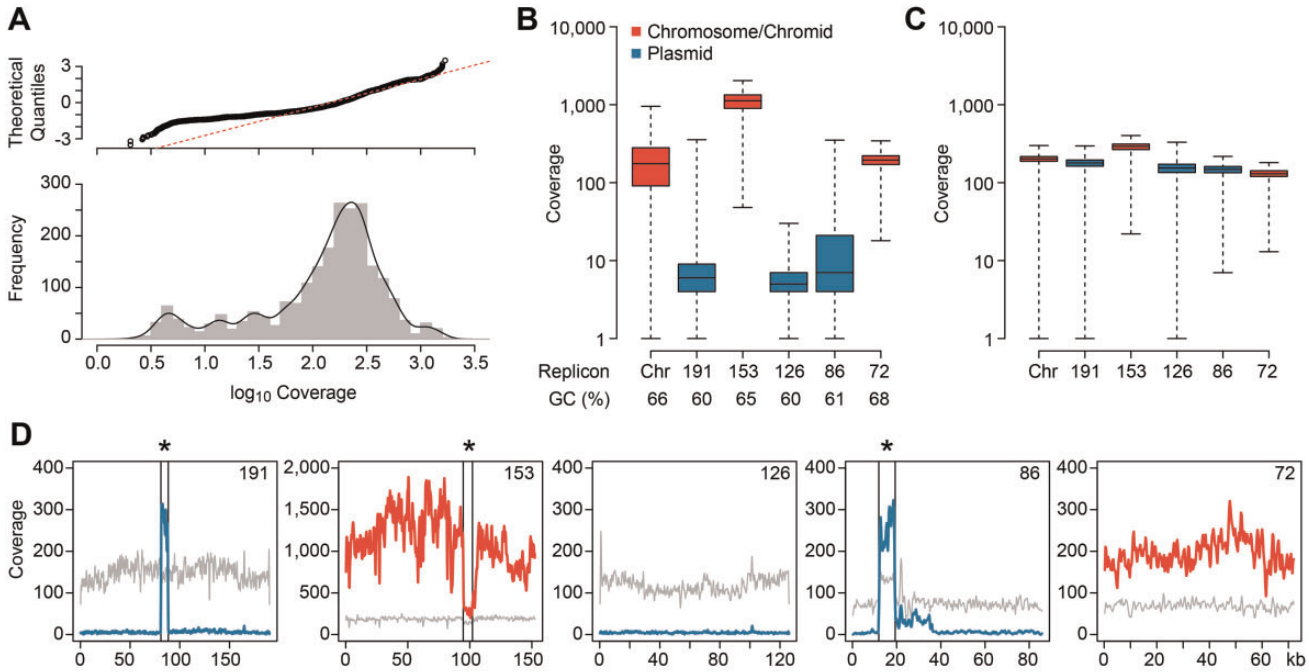

Fig. 2.—Packaging of DNA originating from different replicons. (*A*) Quantile-quantile plot comparing $\log_{10}$-transformed sequencing read-coverage of DsGTA-DNA to a theoretical normal distribution (upper panel) and histogram showing distribution of $\log_{10}$-coverage (lower panel). Mean values for 4-kb bins overlapping by 2-kb are shown. Red line in the upper panel indicates normal distributed data. (*B*) Distribution of $\log_{10}$-coverage of DsGTA-DNA by replicon. Average GC-content of each replicon is indicated. (*C*) Distribution of $\log_{10}$-coverage of *D. shibae* genomic DNA by replicon. The boxplots presented in *B* and *C* show the 25th and 75th quartiles as whiskers, the interquartile range as box and the median as line. These data have been reproduced (supplementary fig. S8, Supplementary Material online). (*D*) Coverage of extrachromosomal replicons by sequencing reads from DsGTA-DNA (blue and red, according to type of replicon) and genomic DNA (grey). Position of a paralogous region to which reads mapped ambiguously is indicated by an asterisk (supplementary fig. S3 and note S2, Supplementary Material online).

methylated adenosines within 4-kb sliding windows than the chromids and the chromosome, respectively. The frequency of methylated cytosines was similar for all replicons (supplementary fig. S4 and tables S2 and S3, Supplementary Material online).

The chromosome showed a clearly nonrandom coverage within the DsGTA-DNA (fig. 3*A*). In contrast, genome sequencing demonstrated a regular pattern of coverage and thus run-off replication, or some other cause of differences in the amounts of different regions of the chromosome being present in cells, could be excluded as a potential explanation for the observed pattern in the DsGTA-DNA coverage (fig. 3*A*). Four almost regularly spaced regions were clearly underrepresented, including the origin and terminus of replication (fig. 3*A*, supplementary fig. S5, Supplementary Material online). In addition, the GTA gene cluster itself showed very low packaging. In contrast, seven regions were preferentially packaged into DsGTA particles. Between those clearly over- and underrepresented regions, several local maxima and minima exist. Regions with low coverage tended to have a low GC-content (fig. 3*B*), consistent with the low coverage of the low-GC plasmids. The cooccurrence of local minima in GC-content and GTA packaging was most obvious for the origin and terminus of replication. A local minimum in GTA coverage within peak 4 could be observed coinciding with a local minimum of GC-content (fig. 3*A*). On the other hand, less frequent packaging of the GTA cluster could not be explained by GC-content. Overall, the strong positive correlation between packaging and GC-content was highly significant (fig. 3*D*). We could not identify DNA motifs such as possible recognition sites for restriction enzymes, phage-like recognition sites, or repeats that were enriched in regions with high packaging rate using the MEME suite for sequence motif discovery. Besides the GTA gene cluster, one defective phage region, lacking a phage replicase gene, was identified in the *D. shibae* genome at position 98.7–102.8 kb, just outside peak 1 (supplementary table S4, Supplementary Material online). Genes encoding transposable elements or tRNAs were also not found within the DsGTA coverage peaks (supplementary table S5, Supplementary Material online). The origin and terminus of replication showed a higher frequency of methylated adenosines (m6A) (fig. 3*C*), whereas the frequency of methylated cytosines (m4C) lacked such a pattern (supplementary fig. S6, Supplementary Material online). The negative correlation between packaging into DsGTAs and the number of m6A-sites was overall lower than for the

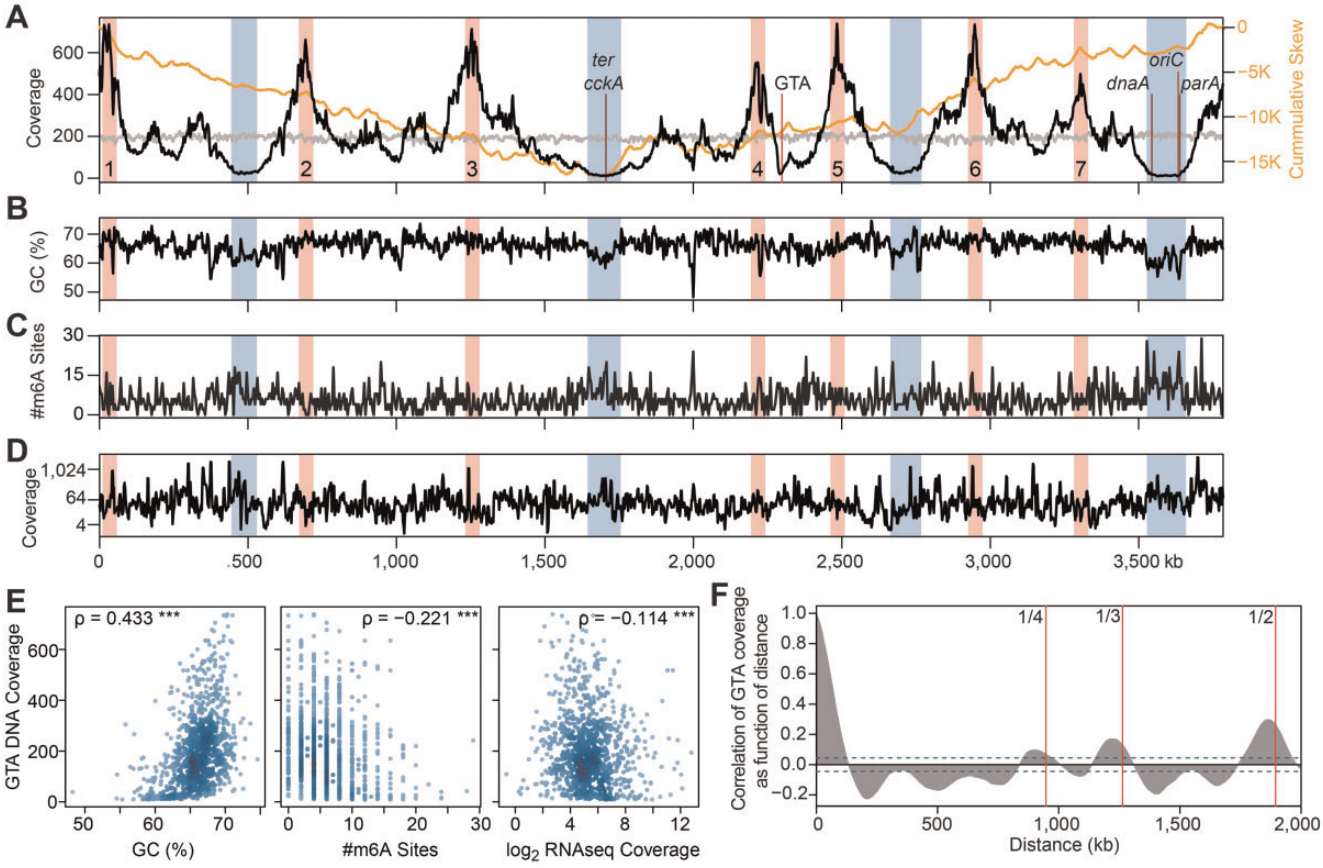

Fig. 3.—Packaging of chromosomal DNA into GTAs in comparison to chromosomal position, GC-content, and transcription. (*A*) Coverage of the chromosome by DsGTA (black) and genomic DNA (grey). Over and underrepresented regions are highlighted in red and blue, respectively. Minimum of cumulative GC/AT-skew (yellow) indicates the terminus of replication (*ter*), which is identical with the position of the gene *cckA*. The origin of replication (*oriC*) has been identified between the genes *dnaA* and *parA* (supplementary fig. S5, Supplementary Material online). These data have been reproduced (supplementary fig. S9, Supplementary Material online). (*B*) GC-content of the chromosome. (*C*) Number of methylated Adenosines (m6A) within the GANTC motif. (*D*) Log_2 coverage of the chromosome by RNA-seq reads for both strands. Data in *A* to *D* have been calculated for sliding windows of 4 kb. (*E*) Scatterplots of DsGTA coverage versus GC-content, m6A methylation and log_2 RNA-seq coverage. Spearman's ρ is indicated ($P < 0.001$). (*F*) Autocorrelation function showing Pearson correlation calculated from DsGTA coverage for all pairs of sequences with the same distance from each other sorted by increasing distance. One quarter, one third, and half the chromosomal length away from *oriC* are indicted by red lines. The dashed blue line indicates the 95% confidence interval.

correlation with the GC-content (fig. 3*E*). Gene expression did not have a large influence on packaging frequency, neither viewed globally for both strands (fig. 3*D*) nor when expression of genes by strand was investigated (supplementary fig. S6, Supplementary Material online). Only a weak, but still significant, negative correlation was found (fig. 3*D*). Within peaks 3, 4, and 6, local minima corresponded to highly expressed genes (fig. 3*A* and supplementary table S5, Supplementary Material online). The first impression of a regular spacing of the different packaging frequency regions gained support from an autocorrelation analysis. For this method, adapted from time-series analysis, Pearson correlation is calculated from all pairs of coverage values that have the same distance from each other and this is done for increasing distances. The resulting plot, showing the correlation ordered by increasing distance, indicated a significant periodicity of the DsGTA coverage data with maximal correlation at distances slightly less than one quarter, one third, and half the chromosome size (fig. 3*F*).

Discussion

Here, we demonstrated that knockout of the autoinducer synthase-encoding gene *luxI*$_2$ leads to overproduction of GTA particles by *D. shibae*. Thus, the product of $LuxI_2$ likely acts as a repressor of GTA gene expression. Transcription of both the GTA gene cluster as well as the operon coding for $LuxI_2$ and the AHL-binding transcription factor $LuxR_2$ is induced by the same regulatory system, the CtrA phosphorelay (Wang et al. 2014). Coactivation of the GTA gene cluster and its repressor might offer an explanation for bistability of GTA gene expression with only a subset of cells producing GTA particles (Fogg et al. 2012; Hynes et al. 2016; Quebatte et al.

2017) although we do not have direct evidence from single cell experiments with *D. shibae*. Gene expression is a stochastic or noisy process. Bistability can arise from noise in the expression of transcriptional regulators embedded in positive or negative feedback loops (Dubnau and Losick 2006). It is tempting to speculate that noisy expression of *luxI*/R_2 induces bifurcation of the population into cells expressing GTA genes and those that do not. However, further experiments will be needed in order to understand regulation of GTA production in *D. shibae*.

Packaging of *D. shibae* genomic DNA into GTA particles is strikingly different from that reported for the other organisms for which whole genome studies exist (Berglund et al. 2009; Hynes et al. 2012). In these cases microarrays had been used to determine packaging frequency of genomic DNA, a less sensitive method compared with DNA sequencing due to the limited dynamic range of the obtained fluorescence signal. However, run-off replication and the resulting higher packaging frequency of the amplified region could be reliably detected by this method (Berglund et al. 2009). Furthermore, transposon-sequencing supported the microarray data for BaGTA (Quebatte et al. 2017). Thus, it is unlikely that the differences in packaging frequency between this and the previous studies represent an artifact of the different methods used.

The question of exactly how the packaging of bacterial DNA into GTAs is performed has not been addressed yet. However, the available data would suggest a "headful" type of packaging, which is common in double-stranded DNA phages (Casjens 2011; Oliveira et al. 2013; Black 2015). In this packaging system, the often highly concatemeric and branched phage DNA is channeled into the empty capsid by an enzyme complex located at the capsid entry (portal) and it is condensed 10,000-fold in the process (Berndsen et al. 2015). This biomotor, which is the fastest and most powerful molecular motor known to date (Smith 2011), consists of three essential components, each of them forming oligomeric ring structures: the portal protein, and the large and small subunits of the terminase complex, TerL and TerS, respectively (Casjens 2011). Packaging is initiated by TerS binding to recognition sites within the concatemeric phage DNA. A recent study suggests that the DNA is wrapped around the oligomeric TerS ring (Gao et al. 2016) and TerS-bound DNA is then recognized and initially cut by TerL at a defined site. As the nuclease domain of TerL cuts DNA unspecifically, the site of DNA linearization is determined by the structure of the TerS-DNA nucleoprotein complex (Djacem et al. 2017). The ATP-dependent motor domain of TerL then transports the linear DNA through the portal ring into the capsid. When the capsid is full, another cut releases linear DNA that can be packaged into the next capsid (Black 2015). TerS presentation of DNA towards TerL is indispensable for production and packaging of linear DNA in vivo. However, Phage T4 DNA packaging is up to 100% efficient in vitro with three components only: small linear DNA of any sequence, the large terminase TerL, and purified prohead with portal proteins (Black and Rao 2012). Thus, if linear DNA is present in the cell, packaging could be possible even without recognition sites for TerS.

How then can random packaging of chromosomal DNA into GTAs be accomplished? Genes coding for terminase and portal proteins are present in alphaproteobacterial GTA gene clusters (Lang and Beatty 2007; Tamarit et al. 2017). On the other hand, we were not able to identify sequence motifs that could act as recognition sites for TerS within the peaks of high packaging. Two hypotheses might help to explain the observed packaging frequency. First, the *D. shibae* genome might encode recognition sites for TerS at the coverage maxima that are so weakly conserved that we cannot detect them, or the physical properties of the DNA, for example, how easy it is to wrap the DNA around TerS, are more important than the sequence itself. Second, perhaps there are more recognition sites throughout the genome, making it difficult to identify those enriched in the packaging peaks by discriminative analysis. Orientation, three-dimensional structure, DNA modification and the formation of nucleoprotein complexes of the *D. shibae* chromosome and plasmids might then determine if a region can be accessed for packaging. Many *Alphaproteobacteria*, including *D. shibae* (Patzelt et al. 2013), utilize a polar mode of cell division. In the best-studied model organism *Caulobacter crescentus*, the origin of replication is located at one cell pole, and the arms of the chromosome extend the length of the cell towards the terminus of replication which is bound to the opposite cell pole (Bowman et al. 2008). Physical attachment to the cell poles might sterically hinder DNA to be bound by TerS, explaining the virtually absent packaging of DNA from the chromosomal origin and terminus of replication. Using chromosome conformation capture and deep sequencing, the three-dimensional structure of the *C. crescentus* chromosome was mapped (Le et al. 2013). It showed multiple spatial domains which were stable throughout the cell cycle; the chromosome contained back-bone regions as well as so-called plectonemes (super-coiled structures). A similar chromatin-like structure was also found in *Bacillus subtilis* and *Escherichia coli* (Dame and Tark-Dame 2016). The nature of chromosomal organization in *D. shibae* is currently unknown, but if it is comparable to that of these well-studied model organisms, such organization into domains that are dependent on localization and interaction frequency might result in some regions in which the DNA is more exposed and therefore more accessible for packaging into GTAs. This might explain the regular distribution of packaging peaks throughout the chromosome.

Regardless of the details of DNA recognition by TerS, once DNA is bound by the protein it would then be cut by TerL, resulting in linear DNA that can be further packaged into a GTA head. As successive packaging might randomly abort after this initial linearization, the GTA coverage would

decrease with increasing distance from the site of initial recognition and cutting, explaining the observed coverage pattern. The probability of TerS binding to DNA might be high for the seven major peaks and lower for the minor peaks on the chromosome. However, novel methodological approaches will be required to test these hypotheses. It will also be interesting to determine if nonrandom packaging of GTAs is a unique feature of *D. shibae*, or also found in other *Rhodobacterales*.

From an evolutionary point of view it is remarkable that the two different mechanisms for HGT mobilize different sets of genes in *D. shibae*. However, as exemplified by the *thiC* locus, duplication allows mobile genetic elements to be transferred to other cells by both the GTA and conjugation mechanisms.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) within the Transregional Collaborative Research Centre "Roseobacter" (TRR 51/2 TB04). A.S.L.'s lab was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada (341561) and A.T.K.H. was supported by an NSERC Undergraduate Student Research Award. A.S.L. thanks Fraser Davidson, Purvikalyan Pallegar and Ekaterina Steffen for assistance with the western blotting, and Robert Harris and Cezar Khursigara at the University of Guelph Molecular & Cellular Imaging Facility for assistance with the electron microscopy. The authors are grateful to the two anonymous referees whose comments and suggestions helped to improve the manuscript.

Author Contribution

J.T., D.P., A.S.L., and I.W.D. designed the study. H.W., A.T.K.H., D.P., and A.S.L. performed experiments. S.B., M.J., and R.G. sequenced the samples. J.T. analyzed the data with major contributions from D.P., M.P., J.P., H.B., and A.S.L. J.T. wrote the manuscript with the help of M.P., J.P., H.B., A.S.L., and I.W.D. All authors read and approved the final manuscript.

Literature Cited

Bailey TL, et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37(Web Server issue):W202–W208.

Bartling P, et al. 2017. The composite 259-Kb plasmid of martelella mediterranea DSM 17316(T)-A natural replicon with functional RepABC modules from Rhodobacteraceae and Rhizobiaceae. Front Microbiol. 8:1787.

Berglund EC, et al. 2009. Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS Genet. 5(7):e1000546.

Berndsen ZT, Keller N, Smith DE. 2015. Continuous allosteric regulation of a viral packaging motor by a sensor that detects the density and conformation of packaged DNA. Biophys J. 108(2):315–324.

Bertani G. 1999. Transduction-like gene transfer in the methanogen *Methanococcus voltae*. J Bacteriol. 181(10):2992–3002.

Biers EJ, et al. 2008. Occurrence and expression of gene transfer agent genes in marine Bacterioplankton. Appl Environ Microbiol. 74(10):2933–2939.

Black LW. 2015. Old, new, and widely true: the bacteriophage T4 DNA packaging mechanism. Virology 479–480:650–656.

Black LW, Rao VB. 2012. Structure, assembly, and DNA packaging of the bacteriophage T4 head. Adv Virus Res. 82:119–153.

Bowman GR, et al. 2008. A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. Cell 134(6):945–955.

Brimacombe CA, Ding H, Johnson JA, Beatty JT, de Boer P. 2015. Homologues of genetic transformation DNA import genes are required for *Rhodobacter capsulatus* gene transfer agent recipient capability regulated by the response regulator CtrA. J Bacteriol. 197(16):2653–2663.

Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H. 2003. Phage as agents of lateral gene transfer. Curr Opin Microbiol. 6(4):417–424.

Casjens SR. 2011. The DNA-packaging nanomotor of tailed bacteriophages. Nat Rev Microbiol. 9(9):647–657.

Charif D, Lobry JR. 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman H.E, Vendruscolo M, editors. Structural approaches to sequence evolution. p. 207–232. New York: Springer.

Dagan T, Martin W. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc Natl Acad Sci U S A. 104(3):870–875.

Dame RT, Tark-Dame M. 2016. Bacterial chromatin: converging views at different scales. Curr Opin Cell Biol. 40:60–65.

Djacem K, Tavares P, Oliveira L. 2017. Bacteriophage SPP1 pac cleavage: a precise cut without sequence specificity requirement. J Mol Biol. 429(9):1381–1395.

Dubnau D, Losick R. 2006. Bistability in bacteria. Mol Microbiol. 61(3):564–572.

Fogg PC, Westbye AB, Beatty JT. 2012. One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. PLoS One 7(8):e43772.

Freese HM, et al. 2017. Trajectories and drivers of genome evolution in surface-associated marine Phaeobacter. Genome Biol Evol. 9(12):3297–3311.

Fu Y, et al. 2010. High diversity of rhodobacterales in the subarctic north atlantic ocean and gene transfer agent protein expression in isolated strains. Aquat Microb Ecol. 59(3):283–293.

Gao S, Zhang L, Rao VB. 2016. Exclusion of small terminase mediated DNA threading models for genome packaging in bacteriophage T4. Nucleic Acids Res. 44(9):4425–4439.

Harrison PW, Lower RP, Kim NK, Young JP. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol. 18(4):141–148.

Huang S, Zhang Y, Chen F, Jiao N. 2011. Complete genome sequence of a marine roseophage provides evidence into the evolution of gene transfer agents in alphaproteobacteria. Virol J. 8(1):124.

Humphrey SB, Stanton TB, Jensen NS, Zuerner RL. 1997. Purification and characterization of VSH-1, a generalized transducing bacteriophage of serpulina hyodysenteriae. J Bacteriol. 179(2):323–329.

Hynes AP, et al. 2012. DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA. Mol Microbiol. 85(2):314–325.

Hynes AP, et al. 2016. Functional and evolutionary characterization of a gene transfer agent's multilocus "genome." Mol Biol Evol. 33(10):2530–2543.

Lang AS, Beatty JT. 2007. Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends Microbiol. 15(2):54–62.

Lang AS, Westbye AB, Beatty JT. 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. Annu Rev Virol. 4(1):87–104.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. Nat Methods. 9(4):357–359.

Le TB, Imakaev MV, Mirny LA, Laub MT. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. Science 342(6159):731–734.

Li H, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079.

Marrs B. 1974. Genetic recombination in *Rhodopseudomonas capsulata*. Proc Natl Acad Sci U S A. 71(3):971–973.

McDaniel LD, et al. 2010. High frequency of horizontal gene transfer in the oceans. Science 330(6000):50.

McInerney JO, McNally A, O'Connell MJ. 2017. Why prokaryotes have pangenomes. Nat Microbiol. 2(4):17040.

Mercer RG, et al. 2012. Regulatory systems controlling motility and gene transfer agent production and release in *Rhodobacter capsulatus*. FEMS Microbiol Lett. 331(1):53–62.

Newton RJ, et al. 2010. Genome characteristics of a generalist marine bacterial lineage. ISME J. 4(6):784–798.

Oliveira L, Tavares P, Alonso JC. 2013. Headful DNA packaging: bacteriophage SPP1 as a model system. Virus Res. 173(2):247–259.

Pagès H, Aboyoun P, Gentleman R, DebRoy S. 2017. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.44.2. Available at: http://bioconductor.org/packages/3.5/bioc/html/Biostrings.html, last accessed January 15, 2018.

Patzelt D, et al. 2016. Gene flow across genus barriers—conjugation of *Dinoroseobacter shibae*'s 191-Kb killer plasmid into *Phaeobacter inhibens* and AHL-mediated expression of type IV secretion systems. Front Microbiol. 7:742.

Patzelt D, et al. 2013. You are what you talk: quorum sensing induces individual morphologies and cell division modes in *Dinoroseobacter shibae*. ISME J. 7(12):2274–2286.

Petersen J, Frank O, Goker M, Pradella S. 2013. Extrachromosomal, extraordinary and essential—the plasmids of the Roseobacter clade. Appl Microbiol Biotechnol. 97(7):2805–2815.

Quebatte M, et al. 2017. Gene transfer agent promotes evolvability within the fittest subpopulation of a bacterial pathogen. Cell Syst. 4(6):611–621.

Smith DE. 2011. Single-molecule studies of viral DNA packaging. Curr Opin Virol. 1(2):134–141.

Tamarit D, et al. 2017. Origin and evolution of the Bartonella gene transfer agent. Mol Biol Evol. doi: 10.1093/molbev/msx299.

Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol. 3(9):711–721.

Wagner-Döbler I, et al. 2010. The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a Hitchhiker's guide to life in the sea. ISME J. 4(1):61–77.

Wang H, et al. 2014. The CtrA phosphorelay integrates differentiation and communication in the marine alphaproteobacterium *Dinoroseobacter shibae*. BMC Genomics 15(1):130.

Westbye AB, O'Neill Z, Schellenberg-Beaver T, Beatty JT. 2017. The *Rhodobacter capsulatus* gene transfer agent is induced by nutrient depletion and the RNAP omega subunit. Microbiology 163(9): 1355–1363.

Yen HC, Hu NT, Marrs BL. 1979. Characterization of the gene transfer agent made by an overproducer mutant of *Rhodopseudomonas capsulata*. J Mol Biol. 131(2):157–168.

Zeileis A, Grothendieck G. 2005. Zoo: S3 infrastructure for regular and irregular time series. J Stat Softw. 14(6): 1–27.

Zhou Y, et al. 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39(Web Server issue):W347–W352.

Associate editor: Bill Martin