Phycological Research

Phycological Research 2019 doi: 10.1111/pre.12392

Vampirovibrio chlorellavorus draft genome sequence, annotation, and preliminary characterization of pathogenicity determinants

Blake T. Hovde,[1]* **Seth A. Steichen**,[2] **Shawn R. Starkenburg**[1] **and Judith K. Brown**[2]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA and [2]School of Plant Sciences, The University of Arizona, Tucson, Arizona, USA

SUMMARY

Vampirovibrio chlorellavorus is recognized as a pathogen of commercially-relevant *Chlorella* species. Algal infection and total loss of productivity (biomass) often occurs when susceptible algal hosts are cultivated in outdoor open pond systems. The pathogenic life cycle of this bacterium has been inferred from laboratory and field observations, and corroborated in part by the genomic analyses for two Arizona isolates recovered from an open algal reactor. *V. chlorellavorus* predation has been reported to occur in geographically- and environmentally-diverse conditions. Genomic analyses of these and additional field isolates is expected to reveal new information about the extent of ecological diversity and genes involved in host-pathogen interactions. The draft genome sequences for two isolates of the predatory *V. chlorellavorus* (*Cyanobacteria*; Ca. Melainabacteria*)* from an outdoor cultivation system located in the Arizona Sonoran Desert were assembled and annotated. The genomes were sequenced and analyzed to identify genes (proteins) with predicted involvement in predation, infection, and cell death of *Chlorella* host species prioritized for biofuel production at sites identified as highly suitable for algal production in the southwestern USA. Genomic analyses identified several predicted genes encoding secreted proteins that are potentially involved in pathogenicity, and at least three apparently complete sets of virulence (Vir) genes, characteristic of the VirB-VirD type system encoding the canonical VirB1-11 and VirD4 proteins, respectively. Additional protein functions were predicted suggesting their involvement in quorum sensing and motility. The genomes of two previously uncharacterized *V. chlorellavorus* isolates reveal nucleotide and protein level divergence between each other, and a previously sequenced *V. chlorellavorus* genome. This new knowledge will enhance the fundamental understanding of *trans*-kingdom interactions between a unique cosmopolitan cyanobacterial pathogen and its green microalgal host, of broad interest as a source of harvestable biomass for biofuels or bioproducts.

Key words: Algae pathogens, algal pathogen, *chlorella*, melainabacteria, VirB.

INTRODUCTION

Several predatory bacteria are known to infect algae, including species that are of interest for cultivation as biofuel feedstocks (Gromov & Mamkaeva 1980; Lee *et al.* 2018; Steichen & Brown 2018). Sequencing, annotation, and *in silico* functional analyses of the algal pathogen genomes are essential to deciphering bacterial-host interactions within the infection cycle. One such pathosystem is that involving the bacterial pathogen *V. chlorellavorus* and its microalgal host, *Chlorella sorokiniana* (Shihira & Krauss 1965) strain DOE1412 (UTEX B 3016) (Steichen & Brown 2018). This predatory bacterium is reported to have a narrow host range encompassing three *Chlorella* species, *C. kessleri, C. sorokiniana,* and *C. vulgaris* (Coder & Starr 1978; Coder & Goff 1986) and *Micractinium inermum* (Ganuza *et al.* 2016), which also falls within the family Chlorellaceae. *V. chlorellavorus* forms a structure appearing to be involved in the attachment of the bacterium to its algal host cell wall. This has previously been observed by microscopy (Gromov & Mamkaeva 1972), and so is expected to be involved in the attachment to other susceptible host alga. In a recent study (Park *et al.* 2018), a sample of the bacterium recovered from an experimental outdoor reactor in Arizona and maintained in a co-culture with its *Chlorella* host was observed to attach to the outer cell surface of *C. sorokiniana* by scanning electron microscopy (SEM) (Fig. 1).

A genome sequence for *V. chlorellavorus* (GenBank: GCA_001858525.1) (Soo *et al.* 2015) was previously determined from a lyophilized sample collected in the Ukraine during 1978 (herein 'Vc_UKR') (Soo *et al.* 2015). The annotation of the Vc_UKR genome supported previous hypotheses that this predatory bacterium is manifest as multiple morphological and biochemical stages during the infection cycle (Soo *et al.* 2015). With respect to predicted virulence factors appearing to be essential during early stages of *Chlorella* host infection, the identification of genes predicted to encode proteins recognizable as a Type IV secretion system (T4SS) components, was suggestive of a translocator mechanism for secreting/transporting and uptake of macromolecules across the bacterial cell envelope. The T4SS of *V. chlorellavorus* was identified within three complete sets of virulence (Vir) genes

*To whom correspondence should be addressed.
Email: hovdebt@lanl.gov
Communicating editor: Rei Narikawa
Received 6 January 2019; accepted 29 May 2019.

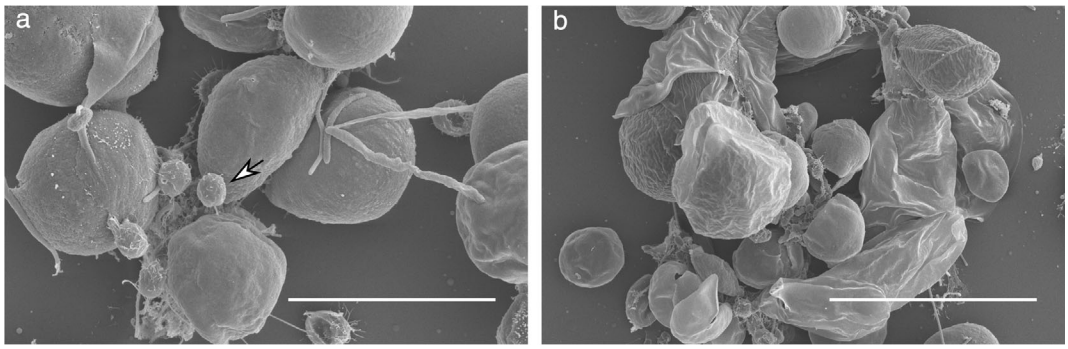

Fig. 1. Scanning Electron Micrographs of *Chlorella sorokiniana* and attached *Vampirovibrio chlorellavorus* cells. Images of samples collected from Arizona test site at (a) 10 000 × magnification with *V. chlorellavorus* indicated by the white arrow and (b) 5000 × magnification. Scale bars are displayed in white representing (a) 5.0 μm and (b) 10.0 μm.

that are characteristic of the VirB-VirD type system encoding canonical VirB1-11 and VirD4 proteins, respectively (Soo *et al.* 2015).

The objective of this study was to sequence and annotate two previously uncharacterized isolates of *V. chlorellavorus* infecting *C. sorokiniana* from a pilot-scale microalgae production facility in Tucson, Arizona, USA; Isolate Vc_AZ_1 (identified in 2016) and Vc_AZ_2 (identified in 2017) for comparative analyses. First, identification of proteins with predicted involvement in one or more types of secretion systems was carried out to probe the possible basis for pathogenicity. Secondly, the means by which *V. chlorellavorus* may deliver effector proteins into the algal host cell to initiate and culminate in infection were predicted. The identification of genes encoded in the complete bacterial genome sequence that are involved in processes leading up to and culminating in the infection of *Chlorella* spp. cells by *V. chlorellavorus,* including host finding, attachment, reproduction/binary fission, and plasmid replication is crucial to characterize this infection cycle. Further genomic based analyses may aid in revealing the basis for the susceptibility of certain *Chlorella* host-species, over those that are non-hosts (Shihira & Krauss 1965) of this predatory bacterial pathogen.

MATERIALS AND METHODS

Isolate information and *in vivo* co-culturing of the algal host and *V. chlorellavorus*

Arizona isolates of *V. chlorellavorus* (Vc_AZ) were collected from outdoor experimental algal cultivation ponds, containing *C. sorokiniana* strain DOE1412 (UTEX B3016), on the University of Arizona campus in Tucson, AZ USA. The samples from March 2016 and May 2017 were designated Vc_AZ_1 and Vc_AZ_2, respectively. The co-culture was maintained live in the laboratory by serial transfer of bacterial cells with *C. sorokiniana* DOE1412 host cells into BG-11 media at 2–3 week intervals to a culture of pathogen-free DOE1412 cells at 1.0 optical density (750 nm). Cultures used as the source of genomic DNA for sequencing were grown in 250 mL flasks in modified BG-11 medium (Huesemann *et al.* 2016) using continuous agitation (120 RPM, Mini Shaker 15 Digital Orbital Shaker, VWR, USA) and a 12 h light:dark cycle. Samples of Vc_AZ_2 algal cultures exhibiting characteristic symptoms of infection by *V. chlorellavorus*, consisting of yellowing and reduced live cell counts, were collected. The stage of infection was additionally characterized by light microscopic observation of aggregated algal cells with cleared cytoplasmic content and closely associated coccoid bacteria. The cells were collected by centrifugation for 5 min at 5930 × g 7000 rpm (J2-21 model, Beckman, USA). The resulting supernatant was filtered using a 2 μm Whatman filter membrane to remove large cells and debris. The filtrate containing the 'free' *V. chlorellavorus* cells of approximately 1.0 μm diameter was centrifuged at 27 200 × g for 10 min to enrich for bacterial cells. The pellet was resuspended in 1 mL sterile BG-11 media. The concentration of the *V. chlorellavorus* cells was estimated to be approximately $1.02E^8$ 16S copies mL^{-1}, based on quantitative PCR (qPCR) analysis (Steichen & Brown 2018).

Genome sequencing and assembly

Isolate Vc_AZ_1; collected May 2016: Shotgun sequencing was performed on a combination of Carbenicillin treated samples and untreated samples of *V. chlorellavorus* co-cultured with *C. sorokiniana* DOE1412, as the algal host. Carbenicillin was used to reduce the bacterial load of the metagenomic sample, but is ineffective as a treatment for *V. chlorellavorus* (Personal communication – Dr. Satish Rajamani/unpublished results). The DNA was purified with the ZR Fungal/Bacterial DNA Miniprep Kit (Zymo Research). Three samples were prepared for Illumina MiSeq 2 × 300 bp sequencing using the Nextera DNA Library Preparation Kit (Illumina). Sequencing of two samples from Carbenicillin-treated cultures yielded 17.7 million and 4.6 million reads, respectively, and the untreated sample yielded 5.7 million reads. Raw reads for Vc_AZ_1 were deposited in the NCBI SRA database under accession numbers SRR7132426-SRR7132428. To maximize genome depth of coverage, sequences from these three samples were combined and trimmed for read quality (q score > 29) prior to assembly. Assembly of the metagenome reads was performed using IDBA_UD version 1.1.1 (Peng *et al.* 2012) on the EDGE platform (Li *et al.* 2017). Minimum and maximum k-mer length were set to

31 and 121, respectively along with step size of 20 and a minimum assembled contig size of 200 bp. The resulting assembly was 139.8 million bp contained in 71 830 contigs corresponding to the genomes of the *Chlorella* host as well as >10 bacterial species. To identify contigs in the assembled metagenome *V. chlorellavorus* origin, the published *V. chlorellavorus* genome sequence (Soo *et al.* 2015) was used as a subject sequence, querying all assembled contigs against using BLASTN (version 2.2.28) with default settings (Zhang *et al.* 2000). Twenty contigs were identified to be of *V. chlorellavorus* origin based on the BLASTN similarity searches. Contigs that contained overlapping sequence were combined into single larger contigs, reducing the final genome assembly to 14 contigs ranging in size from 5358 bp to 492 360 bp in length. A *de-novo* isolation of the Vc_AZ_1 contigs was performed, utilizing read coverage versus GC content, independent of utilizing the Vc_URK as a reference. This analysis resulted in the same cohort of Vc_AZ_1 contigs being identified with no additional contigs added to the draft genome.

Isolate Vc_AZ_2; collected March 2017: *V. chlorellavorus* was prepared and enriched by filtration as described above. Sample preparation, sequencing and assembly was performed by Phase Genomics (Seattle, WA) using the ProxiMeta Hi-C metagenomic deconvolution assembly service. Raw reads for Vc_AZ_2 were deposited in the NCBI SRA database under accession number SRR7985616.

Genome annotation

PROKKA (v1.11) (Seemann 2014) was used to annotate the final genome assemblies of *V. chlorellavorus.* Vc_AZ_1 had a total of 2506 putative genes annotated while Vc_AZ_2 had a total of 2718 putative genes annotated. Genome assemblies and annotations for these isolates have been deposited respectively, under the accession number QFWH00000000 (NCBI BioSample SAMN09087149) and QFWI00000000 (NCBI BioSample SAMN09087150) in NCBI GenBank and are available under the NCBI BioProject number PRJNA464406. To query the completeness of both genome assemblies, BUSCO (Simão *et al.* 2015) was used to identify highly conserved single copy bacterial gene orthologs. Selected Vc_AZ gene sequences were manually annotated by alignment with the Vc_UKR genome sequence (Soo *et al.* 2015) using BLASTN.

Gene sequence analysis

The predicted VirB Vc_AZ sequences were identified by alignment with protein sequences available in the Swissprot database (release 2018_01, http://www.uniprot.org/) (Bateman *et al.* 2017) using BLASTX (States & Gish 1994), using the default search parameters. The translated amino acid sequences of the 2718 predicted Vc_AZ genes were analyzed to identify the conserved domains of documented function by alignment with both the unrestricted, full and the more reliable curated NCBI Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd/) using RPS-BLAST with an expected minimum threshold of 0.01 (Marchler-Bauer *et al.* 2017). Each Vc_AZ protein sequence was assigned a functional annotation by searching for clusters of orthologous genes available in the eggNOG 4.5 database (http://eggnogdb.embl.de/#/app/home), using DIAMOND and the default parameters (Huerta-Cepas *et al.* 2017). The results obtained using the two approaches were aggregated and manually curated. In addition, the Vc_AZ predicted VirB-encoding genes were compared to the results reported by Soo *et al.* (2015) with respect to the predicted VirB operons, by mapping to the VirB11 locus using the 'gggenes' package, available in R studio (R Core Team 2013). Genome synteny plots between Vc_AZ_1 and two assemblies were computed using the progressive Mauve algorithm with match seed weight and minimum LCB score options set to 'automatic' calculation (Darling *et al.* 2004).

The amino acid sequences for predicted bacterial secretion system (SS) proteins were deduced based on the bacterial codon-use table (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi; Table 11), and annotated based on comparisons with the most closely related orthologs in NCBI Conserved Domain Database (CDD). Candidate ABC transporter gene sequences were extracted by manual curation of Conserved Domains Database and eggNOG annotation results by filtering for domains with ABC and ATPase related functional predictions, and then classified according to the class and family designations as previously described (Dassa & Bouige 2001).

RESULTS

Genome sequencing and annotation

The Vc_AZ_1 and Vc_AZ_2 draft genomes contain 2.78 and 2.98 Mbp, respectively. (Table 1). Vc_AZ_1 genome assembly is supported by an average fold read coverage of 62.5X with a total of 3.39% of all metagenome reads mapping back to the *V. chlorellavorus* genome. To determine the degree of completeness, the annotated genes were analyzed against a bacterial BUSCO (Benchmarking Universal Single-Copy Orthologs) database of conserved universal single copy genes. Among the two genomes assembled, 136 and 137 of 148 bacterial BUSCO genes were identified for Vc_AZ_1 and Vc_AZ_2, respectively (Appendix S1). When Vc_AZ_1 and Vc_AZ_2 BUSCO results were compared to Vc_UKR (Soo *et al.* 2015), the results were nearly identical, with 138 of 148 BUSCO genes identified. Collectively, these results indicated that the genome of each of the newly sequenced *V. chlorellavorus* isolates are nearly complete based on the high level of conserved

Table 1. Genome assembly and annotation statistics

	Vc_AZ_1 (Metagenome assembly)	Vc_AZ_2 (Hi-C assembly)
Genome size (Mb)	2.78	2.98
Assembled contigs	11	38
GC (%)	54.8	53.0
Putative genes	2506	2718
tRNA genes	43	47
Bacterial BUSCO genes identified	136/148 (91.9%)	137/148 (92.6%)

BUSCOs between all assemblies. The lack of duplicated or fragmented BUSCOs also provided support that the genome assemblies were of high quality (Appendix S1).

Isolate comparisons based on pairwise nucleotide identity

The three isolates (Vc_AZ_1, Vc_AZ_2 and Vc_UKR) were found to be significantly unique from each other with respect to overall shared nucleotide identity. Using average nucleotide identity (ANI) of conserved homologous regions, the three genomes shared only 83–84.5% identity (Appendix S2), based on OrthoANI analysis (Yoon *et al.* 2017). Further, less than 61% of the total available genome sequence was conserved among all pairwise comparisons.

In silico identification of selected secretion system-like genes with predicted involvement in pathogenicity

A comparative analysis of proteins with predicted involvement in *V. chlorellavorus* secretion systems and the mechanisms by which this pathogen may deliver effector proteins into algal host cells was performed. Putative ATP binding cassette (ABC) transporters, previously shown by the functional validation in well-characterized pathosystems to be involved in coupling ATP hydrolysis to solute uptake and efflux across the cell membrane in bacteria and eukaryotes (Davidson & Chen 2004), were identified based on predicted homologies with functional ABC transporters and ATPase activities. These proteins are of interest because in pathogenic bacteria, they act as virulence factors directly involved in host attachment or intrusion, nutrient uptake, and/or the secretion of toxins or anti-microbial factors (peptides, other). Among these, several have been implicated in the attachment of *V. chlorellavorus* to its algal host cell wall, a step that is predicted to precede the enzymatic dissolution of cellular contents released concomitant with the collapse of infected algal cells (Soo *et al.* 2015).

Most bacterial T4 secretion system (T4SS) operons have been shown to be arranged canonically within a 12 component VirB system, like that which was first described for the plant pathogen *Agrobacterium tumefaciens* (Smith & Townsend) (Christie 2016). Manual curation of the functionally annotated regions showed that the AZ_Vc_2 genome encoded a nearly complete set of VirB-like genes arranged and oriented similarly to the three operons encoded by the type isolate, Vc_UKR, of *V. chlorellavorus* (Fig. 2). The Vc_AZ_2 VirB4-encoding pathogenicity island was most homologous to the VirB4 operon on contig 21.19 (NCBI: LAPX01000019.1) in Vc_UKR. The VirB4 orthologs on these pathogenicity operons share an average nucleotide identity of 84.5% and average alignment length of 952 bp (n = 15), while plasmid 1.27 and plasmid 2.28 had 79.7% ID/240 bp^{-1} alignment length averages (n = 11) and 78.5%/230 bp^{-1} alignment length averages (n = 16), respectively (Appendix S3). Among these genes, 11 were identified to have a predicted involvement in T4SS-related processes (Table 2), based on sequence homology with the VirB genes of *A. tumefaciens*, and components of the homologous SS system genes represented by *tra* and *trb*, encoded by *Escherichia coli* and other bacteria (Christie 2016).

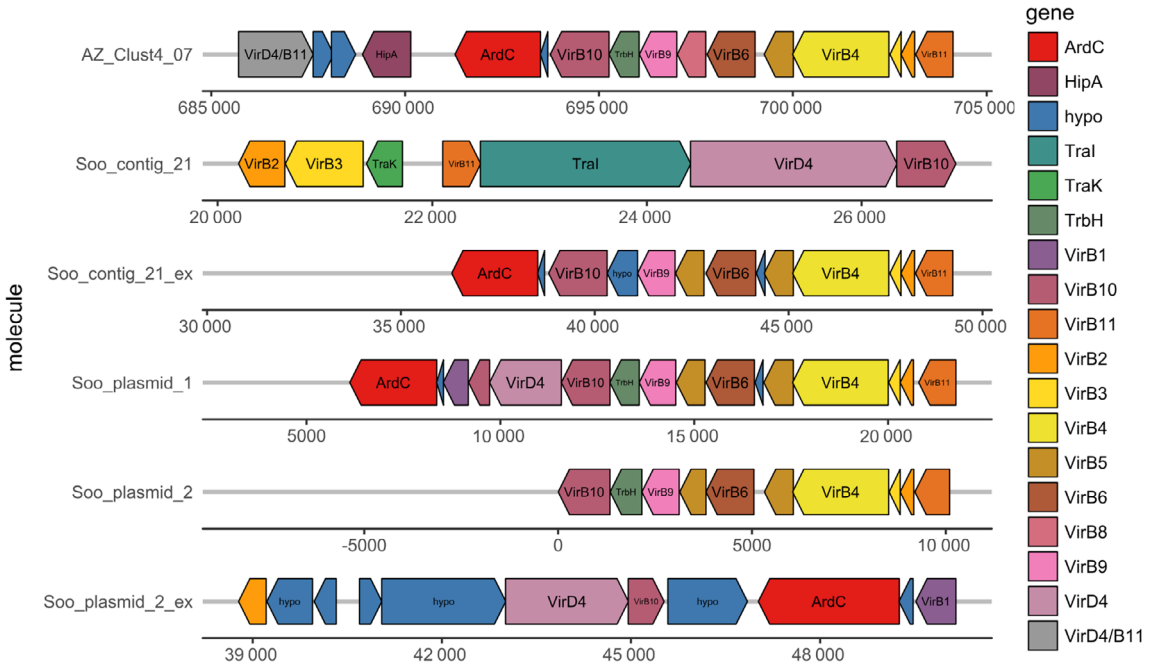

Fig. 2. Gene Organization of the Type IV Secretion System Coding Regions. The operon in the AZ_Vc_2 genome assembly of *V. chlorellavorus* (contig: AZ_Clust4_07), compared to operons reported by Soo *et al.* 2015 located on chromosomal contig 21.19, plasmid 1.27, and plasmid 2.28. Genes were mapped by first aligning the VirB11 gene from each of the four compared operons. [Color figure can be viewed at wileyonlinelibrary.com]

Table 2. Selected annotation results of the putative type IV secretion system genes identified in the genome assembly of the Arizona isolate, Vc_AZ_2, of *Vampirovibrio chlorellavorus*. Annotation by PROKKA of the Vc_AZ_2 genome was compared to the annotations of the UKR_Vc T4SS operon reported by Soo *et al*. (2015), compared against the UniProt database (Bateman *et al*. 2017), the EMBL eggNOG database (Huerta-Cepas *et al*. 2017) and the NCBI Conserved Domains Database (CDD) (Marchler-Bauer *et al*. 2017)

Putative gene	Prokka annotation	Soo *et al*. (2015)		BLASTX vs. UniProt		eggNOG mapper		NCBI CDD	
		Contig	Annotation	Description	%ID	Predicted name	Description	Full database	Curated database
Clust4 00660	TraG	21.19	VirD4	TraG	48%	TRAG	Conjugal transfer coupling protein TraG	TrwB AAD bind superfamily	RecA-like NTPases superfamily
Clust4 00666	PtlG	21.19	VirB10	TrbI	45%	TRBI	Conjugation TrbI family protein	VirB10 like superfamily	VirB10
Clust4 00667	Hypothetical	21.19	TrbH	NA	NA	NA	NA	TrbH superfamily	NA
Clust4 00668	Hypothetical	21.19	VirB9	TrbG	36%	TRBG	Transfer protein TrbG	VirB9 CagX TrbG superfamily	VirB9 CagX TrbG
Clust4 00669	Hypothetical	21.19	VirB5	TrbF	28%	TRBF	Conjugal transfer protein	VirB8 like superfamily	TrbF
Clust4 00670	Hypothetical	21.19	TrbL (VirB6)	TrbL	32%	TRBL	Transfer protein trbL	TrbL superfamily	NA
Clust4 00671	Hypothetical	21.19	TrbJ (VirB5)	NA	NA	NA	P-type conjugative transfer protein TrbJ	VirB5 like superfamily	NA
Clust4 00672	VirB4	21.19	VirB4	TrbE	46%	TRBE	Conjugal transfer ATPase TrbE	CagE TrbE VirB superfamily	ABC ATPase
Clust4 00673	Hypothetical	21.19	VirB3	TrbD	42%	NA	Conjugal transfer protein trbD	VirB3 superfamily	NA
Clust4 00674	Hypothetical	NA	VirB2	NA	NA	NA	Conjugal transfer protein Trbc	TrbC superfamily	Peptidase M48 M56 superfamily
Clust4 00675	PtlH	21.19	VirB11	TrbB	45%	TRBB	P-type conjugative transfer ATPase TrbB	P-loop NTPase superfamily	RecA-like NTPases superfamily

The VirB gene cluster found in the Vc_AZ_2 and Vc_UKR genomes was not recovered otherwise detectable in the Vc_AZ_1 assembly. Syntenic analysis between the genomes of the two AZ isolates indicated that the genomic regions flanking the VirB gene cluster in Vc_AZ_2 were conserved on contig 5 of the Vc_AZ_1 assembly (Fig. 3). Because this 'missing region' occurs within an assembled contig, it seems unlikely that it could have been excluded during genome assembly. Considering the observed pathogenic nature of this isolate, it seems further unlikely that the bacterium would be suitably fit or even pathogenic without the predicted VirB-like functions. Of note, the VirB gene clusters in Vc_AZ_2 and Vc_UKR do not contain a gene encoding VirB1, suggested to be essential for T-pilus assembly (Zupan *et al*. 2007). Further,

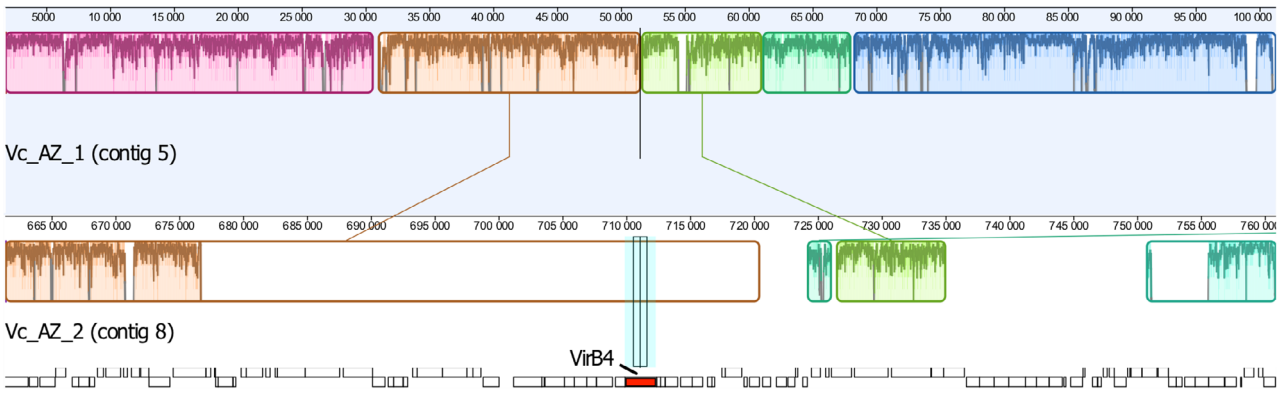

Fig. 3. Positional analysis Vc_AZ_2 Type IV secretion system gene location in Vc_AZ_1 genome assembly. Comparison of locally-collinear blocks of sequences for the two genomes, determined using a progressive MAUVE alignment, are displayed as the same color on top and bottom tracks with the connecting lines and white space indicating no detected synteny. Gene annotations are shown as open squares, and the Vc_AZ_2 VirB4 gene highlighted in red. [Color figure can be viewed at wileyonlinelibrary.com]

Vc_AZ_2 also lacks a VirB7 locus, a canonical lipoprotein component of the outer membrane complex.

The Vc_AZ_2 genome was also found to encode 34 genes with predicted ABC transporter- or ATPase-activity (Appendix S4). The substrate specificity of ATPase units and ABC transporter classifications were inferred from conserved domain annotations, revealing a wide range of transport capabilities, including import and export of apparently essential sugars, peptides, and iron-siderophore complexes, predicted to enhance or be required for pathogenicity (Köster 2001). All the protein sequences are conserved in Vc_AZ_1, except for three loci; Clust4_00026, Clust4_00672, Clust4_01801, and Clust4_01946, having predicted functional involvement in DNA repair, T4SS, chromosome segregation, and a toxin/antitoxin system, respectively (Appendix S4).

DISCUSSION

Genome sequencing of additional isolates of *V. chlorellavorus* from a *Chlorella* dominated outdoor cultivation pond suggests that the Genus *Vampirovibrio* was unexpectedly more genomically and functionally diverse than could have been anticipated. Given the relatively high levels of divergence between these three representatives of the same (putative) species, additional sequencing and characterization of these pathogenic bacterial isolates, and others from diverse habitats are required to enhance the understanding of the range of diversity and extent of virulence within this formidable pathogen of *Chlorella* spp., including the vulnerability to targeted, knowledge-based mitigation strategies designed to prevent losses of entire reactor runs, in otherwise optimal microalgal arid production sites.

The genome assembly of the *V. chlorellavorus* isolate, Vc_AZ_1, did not harbor recognizable VirB gene homologs of the T4SS components that are expected to be essential for virulence and pathogenicity. In contrast, the other Arizona isolate for which the genome sequence was determined, Vc_AZ_2, was found to encode homologs of T4SS. However, neither of the plasmids reported from the type isolate from Ukraine, Vc_UKR were recovered from either Arizona isolate. Despite these results, both of the Arizona isolates were associated with algal cultures that experienced cell collapse. Thus, there is no obvious explanation for the ability to recover the expected VirB homologs (T4SS components) from one but not the other AZ isolate. Similarly, the basis for the apparent absence of bacterial plasmids, like those associated with the type isolate, is not known.

The three *V. chlorellavorus* genome assemblies each contained a single copy of the 16S ribosomal RNA-encoding gene that shared >99% nucleotide identity with one other, despite the observation that the overall gene content and shared nucleotide identities were notably different. These results are consistent with those from other studies that have recognized the limitations of using a single gene as a species marker for bacterial identification. The availability of several divergent genome sequences of *V. chlorellavorus* will enable a comparison of multiple genes to explore bacterial diversity and carry out the first detailed ecological studies of this host-pathogen complex.

Additional studies are needed to verify the functional predictions reported herein, and to elucidate temporal gene expression of the T4SS components and ABC transporters, in relation to the infection cycle, which are expected to be among the core suite of bacterial effector-host protein interactors, that together with specific environmental factors favor *V. chlorellavorus* infection of their algal hosts.

In summary, this research provides new perspectives toward understanding the diversity of the genus, *Vampirovibrio*, for which until now has been represented by a single exemplar species, and to be characterized at the genome level. The results of this research have direct relevance to the success of large-scale commercial algal production projects underway to advance US energy security (biofuels), and the production of aquaculture feedstocks and algal-based nutraceuticals. Additionally, this study has helped to identify selected virulence-related gene groups employed by this obligate bacterial pathogen for the successful predation of susceptible *Chlorella* species.

ACKNOWLEDGMENTS

This work was supported by the Bioenergy Technology Office within the US Department of Energy Office of Energy Efficiency and Renewable Energy, contract NL0029949 (WBS 1.3.1.600) and by the US Department of Energy, Grant DE-EE0006269. The authors declare no conflicts of interest that could be perceived to influence the outcomes of the research.

REFERENCES

Bateman, A., Martin, M. J., O'Donovan, C. *et al.* 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**: D158–69.

Christie, P. J. 2016. The mosaic type IV secretion systems. *EcoSal Plus.* **7**: 1-2.

Coder, D. M. and Goff, L. J. 1986. The host range of the chlorellavorous bacterium ('*Vampirovibrio chlorellavorus*'). *J. Phycol.* **22**: 543–6.

Coder, D. M. and Starr, M. P. 1978. Antagonistic association of the chlorellavorus bacterium ('Bdellovibrio' chlorellavorus) with *Chlorella vulgaris*. *Curr. Microbiol.* **1**: 59–64.

Darling, A. C. E., Mau, B., Blattner, F. R. and Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**: 1394–403.

Dassa, E. and Bouige, P. 2001. The ABCs of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Res. Microbiol.* **152**: 211–29.

Davidson, A. L. and Chen, J. 2004. ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem.* **73**: 241–68.

Ganuza, E., Sellers, C. E., Bennett, B. W., Lyons, E. M. and Carney, L. T. 2016. A novel treatment protects *Chlorella* at commercial scale from the predatory bacterium *Vampirovibrio chlorellavorus*. *Aquat. Microbiol.* **7**: 848.

Gromov, B. V. and Mamkaeva, K. A. 1972. Electron microscopic study of parasitism by *Bdellovibrio chlorellavorus* bacteria on cells of the green alga *Chlorella vulgaris*. *Tsitologiia* **14**: 256–60.

Gromov, B. V. and Mamkaeva, K. A. 1980. New genus of bacteria, *Vampirovibrio*, parasitizing *chlorella*, and previously assigned to the genus Bdellovibrio. *Mikrobiologiia* **49**: 165–7.

Huerta-Cepas, J., Forslund, K., Coelho, L. P. *et al.* 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**: 2115–22.

Huesemann, M., Crowe, B., Waller, P. *et al.* 2016. A validated model to predict microalgae growth in outdoor pond cultures subjected to fluctuating light intensities and water temperatures. *Algal Res.* **13**: 195–206.

Köster, W. 2001. ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Res. Microbiol.* **152**: 291–301.

Lee, P. A., Martinez, K. J. L., Letcher, P. M., Corcoran, A. A. and Ryan, R. A. 2018. A novel predatory bacterium infecting the eukaryotic alga *Nannochloropsis*. *Algal Res.* **32**: 314–20.

Li, P.-E., Lo, C.-C., Anderson, J. J. *et al.* 2017. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Res.* **45**: 67–80.

Marchler-Bauer, A., Bo, Y., Han, L. *et al.* 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**: D200–3.

Park, S. H., Steichen, S. A., Li, X., Ogden, K. and Brown, J. K. 2018. Association of *Vampirovibrio chlorellavorus* with decline and death of *Chlorella sorokiniana* in outdoor reactors. *J. Appl. Phycol.* **31**: 1131–42.

Peng, Y., Leung, H. C. M., Yiu, S. M. and Chin, F. Y. L. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–8.

R Core Team, 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. R Foundation for Statistical Computing. [cited on January 2018]. Available from: http://www.R-project.org/

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–9.

Shihira, I. and Krauss, R. W. 1965. *Chlorella. Physiology and taxonomy of forty-one isolates*. pp. 1–97. Maryland: University of Maryland, College Park.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–2.

Soo, R. M., Woodcroft, B. J., Parks, D. H., Tyson, G. W. and Hugenholtz, P. 2015. Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*. *PeerJ* **3**: e968.

States, D. J. and Gish, W. 1994. Combined use of sequence similarity and codon bias for coding region identification. *J. Comput. Biol.* **1**: 39–50.

Steichen, S. A. and Brown, J. K. 2018. Real-time quantitative detection of *Vampirovibrio chlorellavorus,* an obligate bacterial pathogen of *Chlorella sorokiniana*. *J. Appl. Phycol.* **31**: 1117–29.

Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S. J. and Chun, J. 2017, 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**: 1281–6.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**: 203–14.

Zupan, J., Hackworth, C. A., Aguilar, J., Ward, D. and Zambryski, P. 2007. VirB1* promotes T-pilus formation in the vir-type IV secretion system of *Agrobacterium tumefaciens*. *J. Bacteriol.* **189**: 6551–63.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. BUSCO Gene Inventory in all three sequenced *Vampirovibrio chlorellavorus* genomes

Appendix S2. Nucleotide Identity Analysis of three *V. chlorellavorus* genomes

Appendix S3. Alignment of the annotated genes of the T4SS VirB operon from Vc_AZ_2 to genes within the three VirB operons found in the UKR_Vc genome assembly

Appendix S4. Predicted ATPase and ABC transporter genes in the Vc_AZ_2 genome