INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE

Sharma *et al.*, *Int. J. Syst. Evol. Microbiol.* 2024;74:006198
DOI 10.1099/ijsem.0.006198

MICROBIOLOGY SOCIETY

OPEN ACCESS

Spirochaete genome identified in red abalone sample represents a novel genus *Candidatus* Haliotispira gen. nov. within the order *Spirochaetales*

Sharu Paul Sharma[1], Catherine M. Purcell[2], John R. Hyde[2] and Andrew J. Severin[1,*]

Abstract

A fully assembled spirochaete genome was identified as a contaminating scaffold in our red abalone (*Haliotis rufescens*) genome assembly. In this paper, we describe the analysis of this bacterial genome. The assembled spirochaete genome is 3.25 Mb in size with 48.5mol% G+C content. The proteomes of 38 species were compared with the spirochaete genome and it was discovered to form an independent branch within the family *Spirochaetaceae* on the phylogenetic tree. The comparison of 16S rRNA sequences and average nucleotide identity scores between the spirochaete genome with known species of different families in *Spirochaetia* indicate that it is an unknown species. Further, the percentage of conserved proteins compared to neighbouring taxa confirm that it does not belong to a known genus within *Spirochaetaceae*. We propose the name *Candidatus* Haliotispira prima gen. nov., sp. nov. based on its taxonomic placement and origin. We also tested for the presence of this species in different species of abalone and found that it is also present in white abalone (*Haliotis sorenseni*). In addition, we highlight the need for better classification of taxa within the class *Spirochaetia*.

INTRODUCTION

Bacteria belonging to the phylum *Spirochaetota* consist of a diverse group of motile bacteria that share the distinguishing physical feature of one or more endoflagella [1]. Bacteria in this phylum occupy a broad range of environments, ranging from microbial mats near deep-sea hydrothermal vents [2] to gingivitis-inducing oral biofilms [3] and to the gut of host organisms such as termites [4]. Notable pathogenic spirochaete species include *Borrelia recurrentis* (which causes louse-borne relapsing fever) [5], *Brachyspira aalborgi* (which causes human colonic spirochaetosis) [6], *Treponema pallidum* (whose subspecies cause syphilis and yaws) [7] and *Candidatus* Maribrachyspira akoyae (which is thought to be the causative agent of the Akoya oyster disease responsible for mass mortality events for cultured pearl oysters) [8].

There is a well-documented association between spirochaetes and molluscan hosts, particularly in bivalves and gastropods. These spirochaetes have been identified in temperate, Antarctic and deep-sea bivalve species, including the common cockle (*Cerastoderma edule*), Pacific oyster (*Crassostrea gigas*), saltwater clam (*Laternula elliptica*) [9] and in 14 species of North American and Eurasian gastropods representing six different families of freshwater snails (*Amnicolidae*, *Baicaliidae*, *Bithyniidae*, *Pyrgulidae*, *Lithoglyphidae* and *Benedictiidae*), which were collected from a variety of water bodies including deep water hydrothermal and gas hydrate zones [10]. In the above species, the spirochaetes were associated with the crystalline styles in their hosts; these rod-shaped gelatinous styles are found in the stomach lining of many molluscs and aid in releasing digestive enzymes ([11] and references therein). The function of the spirochaetes in the styles (if any) is uncertain but has been proposed as providing some digestive benefit or additional nutrient absorption for their hosts; however, in these studies, the spirochaetes were not present in every sampled specimen, thereby casting doubt on their role as an essential symbiont in these animals [8–10].

Author affiliations: [1]Genome Informatics Facility, Iowa State University, Ames, Iowa, USA; [2]NOAA Fisheries Southwest Fisheries Science Center, La Jolla, California, USA.
***Correspondence:** Andrew J. Severin, severin@iastate.edu
Keywords: abalone; Candidatus; genome; spirochaete.
Abbreviations: ANI, average nucleotide identity; LRR, leucine-rich repeat; POCP, percentage of conserved proteins.
The sequences of novel spirochaete are available. The whole genome sequence data are available under GenBank accession number CP123443 and 16S rRNA gene sequences at GenBank accession numbers OQ628450 and OQ628451. The complete dataset is available under Bioproject PRJNA941885.
One supplementary figures, four supplementary tables and one supplementary file are available with the online version of this article.
006198

Spirochaetes have also been found in the gills of bivalves and gastropods, including marine clam species (*Scrobucilaria* species, *Laternula elliptica*), soft shelled clams (*Mya arenaria*) and a family of cold-seep clams (*Lucinidae*) [9, 12]. More recently, spirochaetes have been identified in several species of abalone including the giant (*Haliotis gigantea*), disc (*Haliotis discus*) and varicoloured (*Haliotis diversicolor*) abalone [13]. Again, inconsistent presence and numbers of spirochaetes in the gills of sampled specimens (including those of the same species) leaves similar questions as to any possible role the spirochaetes may serve in the gill tissues of these hosts [13].

Very little is known about the diversity and function of these potential commensal bacteria, particularly in marine species such as abalone. There are seven species of abalone that are native to the western coast of North America and, unfortunately, all have undergone significant population declines due to factors including overexploitation, disease and climate change [14]. Two of these abalone, the white (*Haliotis sorenseni*) and black (*Haliotis cracherodii*), are now endangered. Most of these North American species are cultured either to help restore natural populations or for commercial markets to supply demand once met by wild-caught animals. Understanding the function of various microbiota within an organism's microbiome, whether beneficial or detrimental, may alter husbandry approaches and could significantly improve the health and growth of these abalone in culture and when placed back into the wild. Towards that aim, in this paper, we describe the genome of a novel spirochaete that was co-assembled with the red abalone (*Haliotis rufescens*) genome reported by Masonbrink *et al.* in 2019 [15].

METHODS

Spirochaete genome identification

The spirochaete genome was identified during the genome assembly of the red abalone (*Haliotis rufescens*) described in Masonbrink *et al.* [15]. Briefly, following a quality check with FastQC, the reads were assembled using MaSuRCA (version 3.2.2) then improved using HiRise2 version 2.1.2-ad17ecf8bf57 [16, 17]. The assembled scaffolds were compared to the National Center for Biotechnology Information (NCBI) nucleotide (nt) database (downloaded on 1 August 2022) using the NCBI Basic Local Alignment Search Tool (BLAST)−2.13.0+. A single scaffold was identified as having sequence similarity to a spirochaete genome and was removed from the abalone genome to be reported separately here.

Genome assembly statistics and annotation

The AssemblyStats nextflow pipeline was used to assess assembly quality (https://github.com/isugifnf/assemblystats). This workflow runs the assemblathon stats perl script and BUSCO (version 5.1.2) using the bacterial lineage dataset (created on: 6 March 2020) [18]. To annotate the spirochaete genome, the NCBI Prokaryotic Genome Annotation Pipeline (PGAP; 2022-08-11. build6275) was used [19] following the PGAP tutorial (https://bioinformaticsworkbook.org/dataAnalysis/GenomeAnnotation/PGAP_tutorial.html).

Phylogenetic analysis

Publicly available genome assembly data for 38 species were downloaded from NCBI; these sequences represented four orders in the class *Spirochaetia*. A few non spirochaete genomes were added to the dataset based on sequence similarity found using BLAST and PGAP taxcheck (Table S1, available in the online version of this article). Protein datasets from 38 species were used as inputs to infer orthologous genes in the novel genome assembly. Orthofinder (version 2.5.2) ran with the Diamond option and default parameters (dendroblast, fasttree, MCL inflation parameter=1.5) [20, 21], and generated phylogenetic trees with the Species Tree from All Genes algorithm [21]. The 16S rRNA genes from 42 genomes were compared to the new genome and a tree generated using fasttree (Fig. S1). The resulting species trees were visualized using iTOL [22].

Taxonomic classification

For species-level classification, average nucleotide identity (ANI) was calculated with BLAST+ using the online tool, JspeciesWS [23]. All 16S rRNA gene sequences were extracted from the species being compared using command line tools. MUSCLE (version 3.8.1551) was used for pairwise comparison of all available 16S rRNA sequences and a percent identity matrix was generated using ClustalO (version 1.2.4) [24, 25]. To ascertain the genus-level classification, the percentage of conserved proteins (POCP) was calculated using the formula ([(C1 +C2)/(T1 +T2)]*100), where C1 and C2 represent the conserved number of proteins, and T1 and T2 represent the total number of proteins in the two genomes being compared [26]. BLASTP was used to identify homologous proteins in pairs of genomes using an E-value of less than 1×10^{-5}, greater than 40% sequence identity and greater than 50% alignment length (https://github.com/hoelzer/pocp).

Data visualization

Heatmap matrices reflecting the ANI and POCP values across *Spirochaetia* species comparisons were generated using ComplexHeatmap package (version 2.15.1) in R [27] as outlined in the tutorial (https://datascience.101workbook.org/08-DataVisualization/02-GRAPHS/03 R/05-rstudio-tutorial-ComplexHeatmap).

Presence/absence analysis in other *Haliotis* species

In addition to the red abalone, raw sequencing reads from other abalone species were mapped to the spirochaete genome. This included data generated from epipodial tissues of a male and female specimen from green (*H. fulgens*), pink (*H. corrugata*), white (*H. sorenseni*) and black (*H. cracherodii*) abalone; and epipodial and mantle tissue from a male and female pinto (*H. kamtschatkana*) abalone from the Masonbrink *et al.* [15] study [15]. In addition, reads from gill tissue of disc (*H. discus hannai*; SRR9859071, SRR10127524) and greenlip (*H. laevigata*; SRR6678008), gut of pink (*H. corrugata*; SRR20073179), and foot muscle tissue of red abalone (*H. rufescens*; SRR19251568) were also mapped to the new spirochaete genome. Short reads were aligned using Hisat2 (version 2.2.0) with default options (zero threshold, so it reports all matches); long reads were aligned using Minimap2 (version 2.2) [28, 29]. Coverage statistics were calculated with Samtools (version 1.10), and the programs Jbrowse2 and Dotplotly (not shown) were used to visualize alignments [30, 31].

RESULTS AND DISCUSSION

Genome assembly and annotation

The assembled spirochaete genome was contained in a single scaffold within the red abalone genome assembly [15] and had a total nucleotide size of 3.25 Mb with a G+C content of 48.5 mol%. The alignment of the Illumina paired-end reads to this assembly indicated coverage of 99.99% of the assembled spirochaete genome with a 29.5 mean coverage depth across the genome. After using the bacteria_odb10 dataset to assess completeness of the assembly, out of the 124 BUSCO genes, 99 complete (79.8%) and seven (5.6%) fragmented BUSCO genes were identified, although 18(14.6%) were missing from the assembled genome. After using the annotated protein sequence, 100(80.6%) complete BUSCO genes were identified. These BUSCO scores are similar to other related published genomes (Table S2).

The Prokaryotic Genome Annotation Pipeline (PGAP) resulted in the annotation of 2430 coding genes in the new genome. The number is more than all the three species in the most closely related genus in the species tree, *Entomospira,* which have close to 1600 coding genes (Table S3). The number is smaller compared to neighbouring species in the 16S rRNA gene tree (Fig. S1), *Spirochaeta cellobiosiphila* and *Spirochaeta isovalerica*. Although *S. cellobiosiphila* is similar in genome size, the number of coding genes is higher at 3618 (Table S3). The difference in size and number of genes suggests biological differences between these genera and the new genome.

Of the coding genes in the spirochaete genome, Orthofinder placed 66% (1607) of the genes into orthogroups and 34% (819) into singletons. Notably, 56 genes contained across eight orthogroups were unique to this genome. An in-depth analysis of the orthogroups across the 39 spirochaete species indicated that there were 111 putative gene duplication events unique to this spirochaete genome. Just two of the orthogroups accounted for 45 of these duplication events, indicating a massive expansion of two gene types. The first orthogroup underwent 25 duplications of genes coding for BspA proteins. The second orthogroup, which underwent 20 duplications, contains genes for B-repeat-containing internalin protein family, including InlB and some similar hypothetical proteins. Both BspA and internalin proteins belong to the leucine-rich repeat (LRR) containing surface protein family, which are known to mediate host–pathogen interactions [32–34]. Another notable expansion was eight duplications in the glycosyltransferase family (glycosyltransferase or glycosyltransferase family two proteins), the enzymes that transfer sugar moieties from donor to an acceptor substrate. More research is warranted to understand how these gene expansions may or may not be related to its presence on the gills of abalone species.

Phylogenetic analysis

Proteomes from 38 species across 21 genera, including a proteobacteria as an outgroup, were compared with the proteome of the candidate spirochaete using Orthofinder [21] to generate a phylogenetic tree (Fig. 1). The 37 species of spirochaetes in the tree chosen for this comparison represent the class *Spirochaetia*, and span across four orders: *Brachyspirales*, *Brevinematales*, *Leptospirales* and *Spirochaetales*. The family *Spirochaetaceae*, which is part of the order *Spirochaetales*, is in turn divided into 12 genera: *Alkalispirochaeta*, *Bullifex*, *Clevelandina*, *Diplocalyx*, *Hollandina*, *Marispirochaeta*, *Oceanispirochaeta*, *Pillotina*, *Salinispira*, *Sediminispirochaeta*, *Spirochaeta* and *Thiospirochaeta*, as determined by the International Code of Nomenclature of Prokaryotes and the list of Prokaryotic names with Standing in Nomenclature (accessed on 17 November 2022) [35]. Orthofinder placed the candidate spirochaete on a separate branch next to the relatively newly discovered genus *Entomospira* within the family *Spirochaetacae* of *Spirochaetales* [36]. *Entomospira* were identified in the aquatic larval stages of mosquitos [37]. Although the closest species on this bacterial phylogenetic tree live in very different environments and host species, this is not unexpected for the diversity seen in spirochaetes. Even the same environment has been found to have diverse spirochaetes; for example, Husmann *et al*. [9] identified spirochaetes in four species of marine bivalves that spanned across spirochaete orders (i.e., *Brachyspirales* and *Spirochaetales*), and even across families within a single species of oyster (*Crassostrea gigas*, with spirochaetes in the genera of *Cristispira* and *Spirochaeta*) [9].

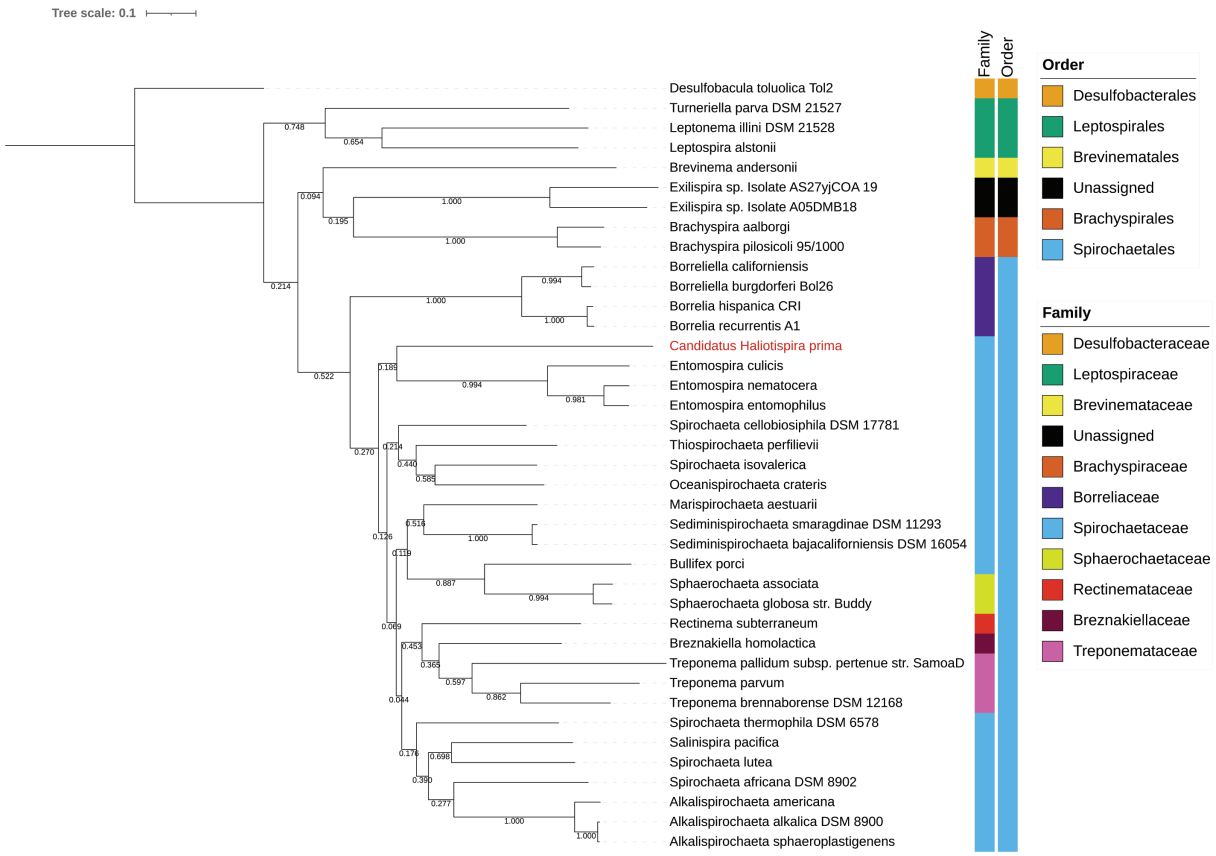

Fig. 1. STAG phylogenetic tree showing the relationship of *Candidatus* Haliotispira prima (in red) to representative spirochaetes. The tree is rooted to a proteobacteria outgroup, *Desulfobacula toluolica*. The numbers on the branches are boostrapped probabilities.

To be classified as the same species, two samples must have an average nucleotide identity (ANI) score above a 95% threshold, which corresponds to 70% DNA–DNA hybridization value [38]. The ANI values fell below this threshold for all species that were compared with this new genome (Fig. 2, Supplemental File ANIb). Organisms sharing an ANI value above the 95% threshold (Fig. 2, shown in dark green) indicates these organisms should be considered as the same species. *Candidatus* Haliotispira prima has ANI scores with the compared species far below the threshold (Supplemental File ANIb), strongly indicating that this spirochaete represents a distinct species. Interestingly, the ANI score between species *A. alkalica* and *A. sphaeroplastigenens* is higher than the 95% species threshold, revealing that these organisms should perhaps not be classified as distinct species (raw data in Supplemental File ANIb). Despite attempting to analyse all 16S rRNA gene sequences of the compared species, no matching species was identified. The percent identity values for the candidate spirochaete ranged from 14.5 to 92.4% (Supplemental File PIM), below the 97–99% threshold used for differentiating species [39]. The whole genome sequence was also aligned with those of neighbouring species with minimap2. Maps generated with Dotplotly showed no synteny and no significant alignments between the candidate genome and other species. The 16S rRNA gene tree also shows that the genes from *Candidatus* Haliotispira prima form an independent branch (Fig. S1). It also shows that the sequences from the new genome are close to *Spirochaeta* as the neighbouring genes belong to *Spirochaeta cellobiosiphila* and *Spirochaeta isovalerica*. In addition, the genome was also compared to 10 species with *Candidatus* status (Table S4) to check if it matches any of them. None of these species had significant matches to the candidate genome when comparing whole genomes or proteomes.

Both ANI and 16S rRNA comparisons have indicated that this represents a new species, yet genus-level demarcation requires the comparison of conserved proteins between pairs of different species. All the POCP values for the candidate species were below 20% (Fig. 3, Supplemental File POCP), which is well below the cut-off of at least 50% to be considered for genus-level classification [26]. This analysis indicates that the candidate spirochaete does not belong to the closest genus on the phylogenetic tree, *Entomospira*, or any other genera of *Spirochaetia* tested, indicating the discovery of an entirely novel genus. We propose the name *Candidatus* Haliotispira prima gen. nov., sp. nov. based on its taxonomic placement and origin in red abalone (*H. rufescens*).

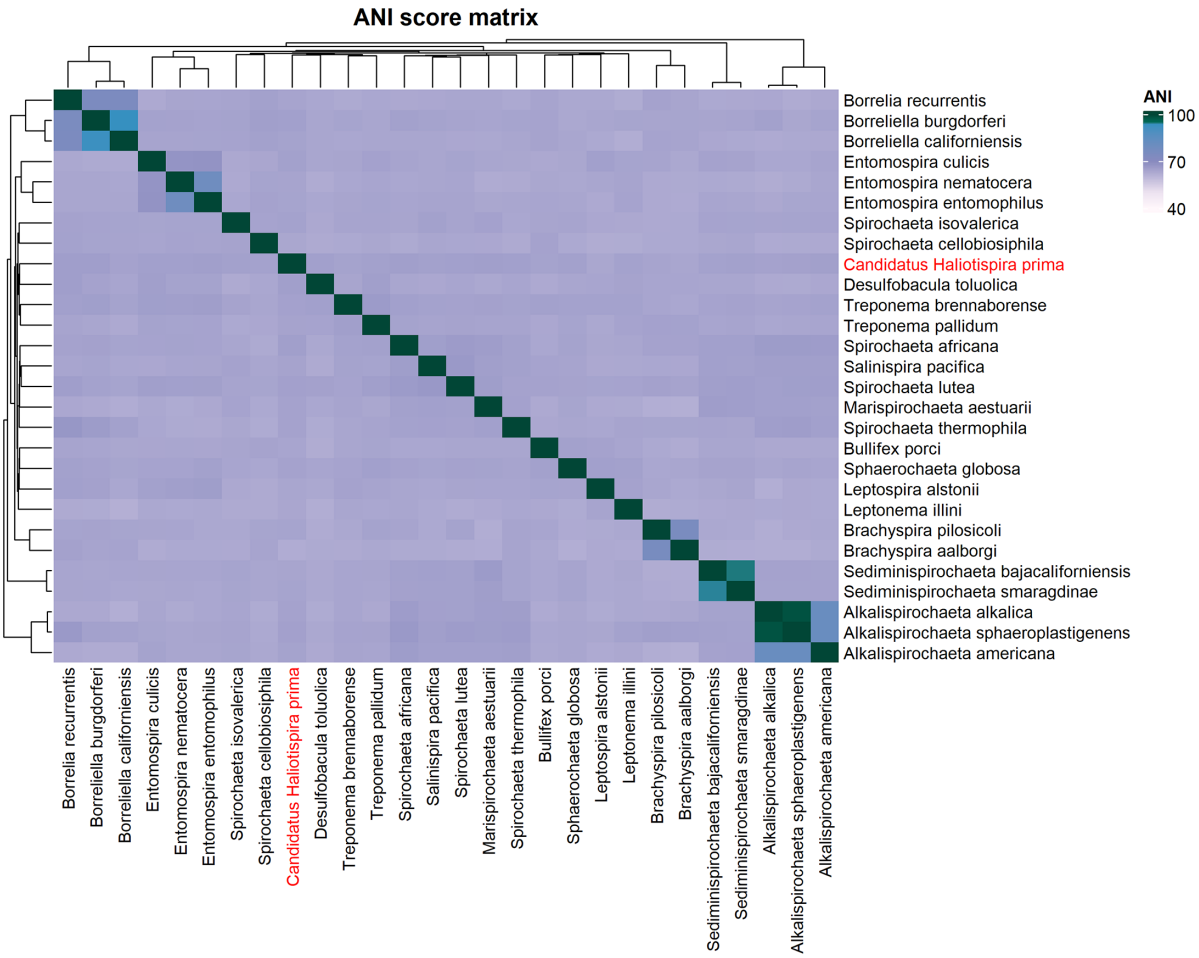

Fig. 2. ANI (average nucleotide identity) matrix showing *Candidatus* Haliotispira prima does not share ANI score above 95% threshold with any compared species. The dark green boxes along the diagonal indicate the above-threshold scores.

Need for improved classification

Most genera form monophyletic clades on the phylogenetic tree presented here (Fig. 1), yet species within the genus *Spirochaeta* lack this distinguishing feature and appear polyphyletic. For example, *Spirochaeta isovalerica* groups with *Oceanispirochaeta crateris* whereas *Spirochaeta lutea* and *Salinispira pacifica* group together. *S. cellobiosiphila*, *S. africana* and *S. thermophila* each seem to be independent lineages. While this adds evidence to the hypothesis that *Spirochaeta* is polyphyletic, it certainly needs a finer phylogenetic classification, as suggested previously [36, 40]. POCP values further support this conclusion, as the *Spirochaeta* species surveyed here were below the 50% threshold. In contrast, the species in the genera *Borrelia* and *Borreliella* have POCP values well above this threshold and their differentiation has been contested by various authors [41, 42]. Likewise, ANI values between *Alkalispirochaeta alkalica* and *Alkalispirochaeta sphaeroplastigenens* indicate a subspecies classification may be more appropriate, as ANI values were greater than 98%, higher than the current threshold of 95% for different species. Note that although *Rectinema subterraneum* now falls under *Rectinemataceae* it stills shows up under *Treponemataceae* in LPSN [35, 43].

Presence in other species

As described above, the *Candidatus* Haliotispira prima genome was found in DNA extracted from gill tissue of red abalone (*H. rufescens*). We wanted to know if this microbe was present in other abalone species. Illumina short-read data from Masonbrink *et al.* [15] and from publicly available raw sequencing data from other studies (see Methods) were mapped to the *Candidatus* Haliotispira prima genome to determine if this spirochaete was also found in these species. These data included short read data from black, white, green, pink, pinto, disc and greenlip abalone, and an additional red abalone specimen. Illumina reads from five white abalone samples mapped to the spirochaete genome. Four of the samples were extracted from epipodial tissues collected from F2 generation captive-bred white abalone that are part of a restoration culture effort for this endangered species, and all individuals were living in the same facility. The fifth sample was collected from the gill tissue of a wild white abalone from Punta

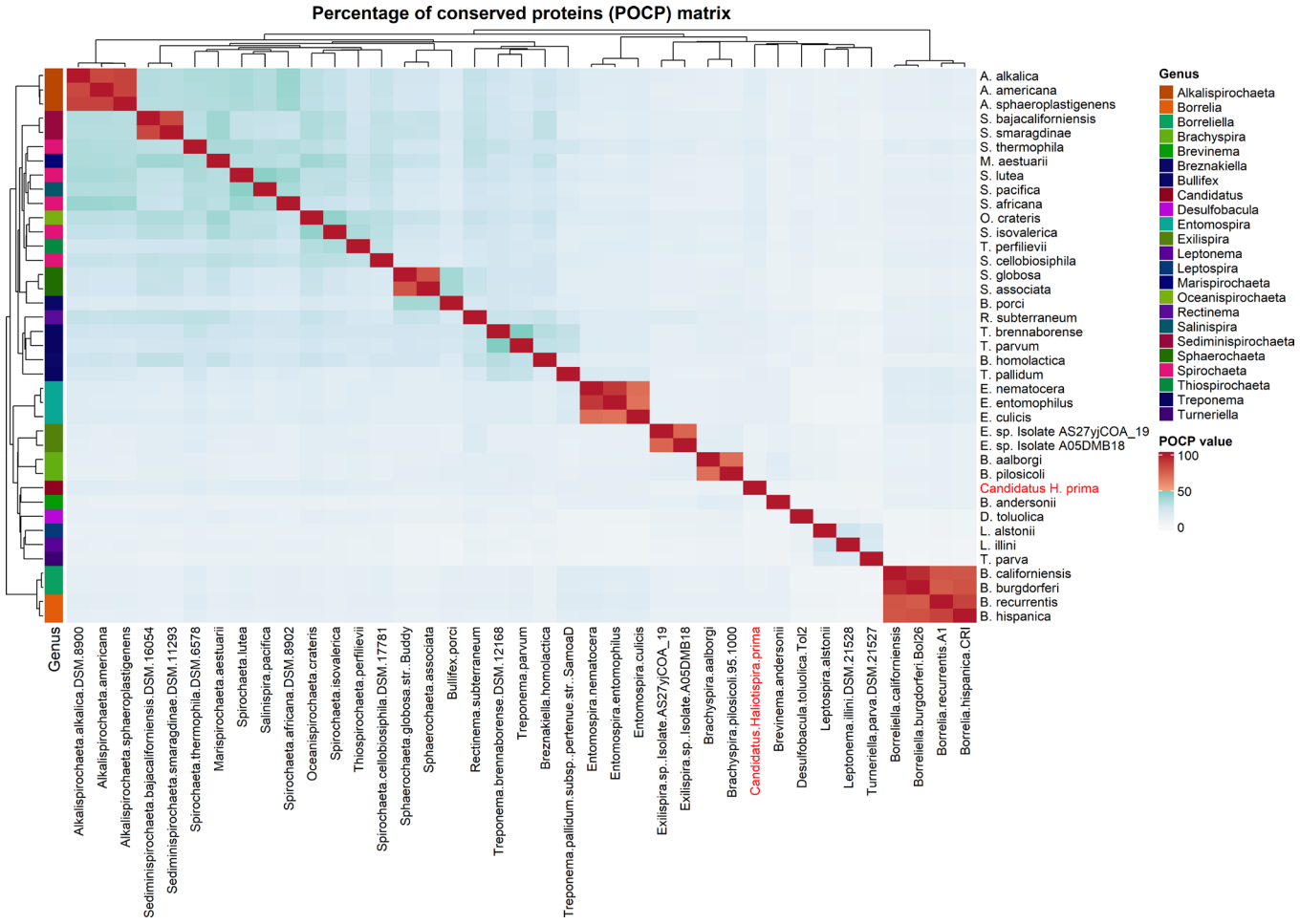

Fig. 3. A POCP matrix where all values corresponding to *Candidatus* Haliotispira prima are well below the 50% threshold for genera. Most species belonging to the same genus form clusters, as evidenced by the high POCP values on the matrix. The *Spirochaeta* species tested here have values lower than the 50% threshold, whereas the division between *Borrelia* and *Borreliella* genera is above the 50% threshold.

Canoas, Baja California, Mexico. Of all the samples, only short reads from the wild white specimens had any alignment coverage. Of those samples, only two had alignment coverage above 3%, the sample collected from Mexico at 52% and the sample collected from La Jolla, CA at 22% coverage of the spirochaete genome. This suggests that this spirochaete is also present in white abalone. The lack of alignment coverage in other species does not necessarily mean that *Candidatus* Haliotispira prima is absent in those species, but may be a result of not being present in the tissue (or type of tissue) sampled from those individuals. Additional sequence data collected from different tissue types and from abalone sampled in different environments (both within and across species) may reveal a pattern of association for this spirochaete species and possibly identify other spirochaetes belonging to this genus. Further exploration may also help reveal the nature of the spirochaete–abalone symbiosis (e.g., commensal, mutual, or parasitic).

CONCLUSION

The goal of this study was to taxonomically classify a whole bacterial genome sequence co-assembled with the genome of red abalone (*H. rufescens*) [15]. The *Candidatus* Haliotispira prima has a circular genome of 3.25 Mb. This size is typical for a spirochaete but is larger than the genomes of the *Entomospira* species (closest on the phylogenetic tree) which average about 1.8 Mb in size (Table S3). The annotation resulted in identification of 2430 coding genes, which is dissimilar to neighbouring species on phylogenetic trees. This new species was found to be present in *H. rufescens* and *H. sorenseni*. Lineage-specific expansion of gene families is an important means of adaptation [44–46]. The massive expansion of the LRR-containing surface protein family in *Candidatus* Haliotispira prima could indicate the expansion of host range [45]. Since the LRR-containing surface protein family is associated with host–pathogen interactions, the potential pathogenicity of this species and potentially other species that might

belong to the *Candidatus* Haliotispira genus needs to be investigated. Although no closely related spirochaetes have not been reported as molluscan pathogens, it is not clear whether *Candidatus* Haliotispira prima is a pathogen or non-pathogenic symbiont of *H. rufescens* and *H. sorenseni*. The availability of this whole genome sequence will enable other researchers to classify new species belonging to *Candidatus* Haliotispira gen. nov.

DESCRIPTION OF *CANDIDATUS* HALIOTISPIRA GEN. NOV.

Candidatus Haliotispira (Ha.li.o.ti.spi'ra. N.L. fem. n. *Haliotis*, the abalone genus; L. fem. n. spira, a spiral; N.L. fem. n. Haliotispira, a spiral from Haliotis).

Candidatus Haliotispira gen. nov. is a spirochaete first discovered in *Haliotis rufescens* (hence the name Haliotispira) samples. This genus is placed within the family *Spirochaetacae* of the order *Spirochaetales*, class *Spirochaetia*, phylum *Spirochaetota*. This new genus is proposed based on a lack of sequence similarity and low POCP scores. The POCP values are below the commonly accepted thresholds for shared genera designation.

DESCRIPTION OF *CANDIDATUS* HALIOTISPIRA PRIMA SP. NOV

Candidatus Haliotispira prima (pri'ma. L. fem. ord. num. prima, the first)

Candidatus Haliotispira prima gen. nov., sp. nov. is the first described bacterium in this novel genus (hence prima). This species was differentiated from other members of *Spirochaetacae* based on phylogenetic analysis and absence of sequence similarity. The assembled genome forms an independent branch on the spirochaete phylogenetic tree and analyses using ANI and 16S rRNA reveal low shared similarity scores. ANI values are below the commonly accepted thresholds for shared species designation. This species also exhibits significant gene expansion of the LRR-containing surface protein family, which is involved in host–pathogen interactions. The whole genome sequence data are available under GenBank accession number CP123443 and 16S rRNA gene sequences with GenBank accession numbers OQ628450 and OQ628451.

Funding information
This research was supported in part by an appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by ORAU under DOE contract number DE-SC0014664. This project received funding under award 15WCR012 from NOAA Fisheries Service, in cooperation with the Saltonstall–Kennedy Grant Program, funding from the UC Davis Office of Research Signature Research in Genomics Program, and from NOAA Sea Grant (project number R/HCME-10). All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of NOAA Fisheries, USDA, DOE, or ORAU/ORISE. Support for this research was also provided by XSEDE [47] (grant number MCB140217) and from California Sea Grant (project number R/HCME-10). XSEDE is supported by the National Science Foundation (grant number ACI-1548562).

Acknowledgements
This work used the Extreme Science and Engineering Discovery Environment (XSEDE). The majority of the computational work was performed on HPC clusters at Iowa State University.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Paster BJ. Spirochaetes. In: Trujillo ME, DeVos P, Hedlund B, Kämpfer P and Rainey FA (eds). *Bergey's Manual of Systematics of Archaea and Bacteria*. Wiley; 2015.
2. Harwood CS, Jannasch HW, Canale-Parola E. Anaerobic spirochete from a deep-sea hydrothermal vent. *Appl Environ Microbiol* 1982;44:234–237.
3. Seshadri R, Myers GSA, Tettelin H, Eisen JA, Heidelberg JF, *et al*. Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A* 2004;101:5646–5651.
4. Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. Acetogenesis from H2 plus CO2 by spirochetes from termite guts. *Science* 1999;283:686–689.
5. Cutler SJ, Moss J, Fukunaga M, Wright DJ, Fekade D, *et al*. *Borrelia recurrentis* characterization and comparison with relapsing-fever, Lyme-associated, and other *Borrelia* spp. *Int J Syst Bacteriol* 1997;47:958–968.
6. Mikosza AS, La T, Brooke CJ, Lindboe CF, Ward PB, *et al*. PCR amplification from fixed tissue indicates frequent involvement of *Brachyspira aalborgi* in human intestinal spirochetosis. *J Clin Microbiol* 1999;37:2093–2098.
7. Smajs D, Norris SJ, Weinstock GM. Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. *Infect Genet Evol* 2012;12:191–202.
8. Matsuyama T, Yasuike M, Fujiwara A, Nakamura Y, Takano T, *et al*. A Spirochaete is suggested as the causative agent of Akoya oyster disease by metagenomic analysis. *PLoS One* 2017;12:e0182280.
9. Husmann G, Gerdts G, Wichels A. Spirochetes in crystalline styles of marine bivalves: group-specific PCR detection and 16S rRNA sequence analysis. *J Shellfish Res* 2010;29:1069–1075.
10. Sitnikova T, Michel E, Tulupova Y, Khanaev I, Parfenova V, *et al*. Spirochetes in gastropods from Lake Baikal and North American freshwaters: new multi-family, multi-habitat host records. *Symbiosis* 2012;56:103–110.
11. Michl SC, Windisch W, Geist J. Function of the crystalline style and first detection of laminarinase activity in freshwater mussels of the genus *Anodonta*. *J Molluscan Stud* 2014;80:198–200.
12. Duperron S, Fiala-Médioni A, Caprais J-C, Olu K, Sibuet M. Evidence for chemoautotrophic symbiosis in a Mediterranean cold

seep clam (Bivalvia: Lucinidae): comparative sequence analysis of bacterial 16S rRNA, APS reductase and RubisCO genes. *FEMS Microbiol Ecol* 2007;59:64–70.

13. Mizutani Y, Mori T, Miyazaki T, Fukuzaki S, Tanaka R. Microbial community analysis in the gills of abalones suggested possible dominance of epsilonproteobacterium in *Haliotis gigantea*. *PeerJ* 2020;8:e9326.
14. Karpov K, Haaker P, Taniguchi I, Rogers-Bennett L. Serial depletion and the collapse of the California Abalone (*Haliotis* spp.) fishery. *Can Spec Publ Fisheries Aquatic Sci* 2000:11–24.
15. Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, *et al*. An annotated genome for *Haliotis rufescens* (red abalone) and resequenced green, pink, pinto, black, and white abalone species. *Genome Biol Evol* 2019;11:431–438.
16. Andrews S. Fastqc - a quality control tool for high throughput sequence data. *Babraham Bioinforma* 2010.
17. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, *et al*. The MaSuRCA genome assembler. *Bioinformatics* 2013;29:2669–2677.
18. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;31:3210–3212.
19. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, *et al*. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
20. Emms DM, Kelly S. STRIDE: Species Tree Root Inference from Gene Duplication Events. *Mol Biol Evol* 2017;34:3267–3278.
21. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;16:1–4.
22. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
23. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 2016;32:929–931.
24. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 2004;5:1–9.
25. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* 2011;7:539.
26. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, *et al*. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.
27. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;32:2847–2849.
28. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019;37:907–915.
29. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
30. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008.
31. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, *et al*. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* 2023;24:1–21.
32. Inagaki S, Onishi S, Kuramitsu HK, Sharma A. *Porphyromonas gingivalis* vesicles enhance attachment, and the leucine-rich repeat BspA protein is required for invasion of epithelial cells by "*Tannerella forsythia*" *Infect Immun* 2006;74:5023–5028.
33. Ebbes M, Bleymüller WM, Cernescu M, Nölker R, Brutschy B, *et al*. Fold and function of the InlB B-repeat. *J Biol Chem* 2011;286:15496–15506.
34. Handrich MR, Garg SG, Sommerville EW, Hirt RP, Gould SB. Characterization of the BspA and Pmp protein family of trichomonads. *Parasit Vectors* 2019;12:1–5.
35. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.
36. Graña-Miraglia L, Sikutova S, Vancová M, Bílý T, Fingerle V, *et al*. Spirochetes isolated from arthropods constitute a novel genus *Entomospira* genus novum within the order Spirochaetales. *Sci Rep* 2020;10:17053.
37. Sikutová S, Halouzka J, Mendel J, Knoz J, Rudolf I. Novel spirochetes isolated from mosquitoes and black flies in the Czech republic. *J Vector Ecol* 2010;35:50–55.
38. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, *et al*. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.
39. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 2014;64:346–351.
40. Ben Hania W, Joseph M, Schumann P, Bunk B, Fiebig A, *et al*. Complete genome sequence and description of *Salinispira pacifica* gen. nov., sp. nov., a novel spirochaete isolated from a hypersaline microbial mat. *Stand Genomic Sci* 2015;10:1–4.
41. Margos G, Gofton A, Wibberg D, Dangel A, Marosevic D, *et al*. The genus Borrelia reloaded. *PLoS One* 2018;13:e0208432.
42. Margos G, Castillo-Ramirez S, Cutler S, Dessau RB, Eikeland R, *et al*. Rejection of the name *Borreliella* and all proposed species comb. nov. placed therein. *Int J Syst Evol Microbiol* 2020;70:3577–3581.
43. Brune A, Song Y, Oren A, Paster BJ. A new family for 'Termite gut Treponemes': description of *Breznakiellaceae* fam. nov., *Gracilinema caldarium* gen. nov., comb. nov., *Leadbettera azotonutricia* gen. nov., comb. nov., *Helmutkoenigia isoptericolens* gen. nov., comb. nov., and *Zuelzera stenostrepta* gen. nov., comb. nov., and proposal of *Rectinemataceae* fam. nov. *Int J Syst Evol Microbiol* 2022;72:005439.
44. Jordan IK, Makarova KS, Spouge JL, Wolf YI, Koonin EV. Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res* 2001;11:555–565.
45. Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, *et al*. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics* 2016;17:1–7.
46. Lespinet O, Wolf YI, Koonin EV, Aravind L. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res* 2002;12:1048–1059.
47. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, *et al*. XSEDE: accelerating scientific discovery. *Comput Sci Eng* 2014;16:62–74.