Dataset of genome sequence, de novo assembly, and functional annotation of Ruegeria sp. (PBVC088), a marine bacterium associated with the toxin-producing harmful dinoflagellate, Pyrodinium bahamense var. compressum

ABSTRACT

The dataset comprises a whole-genome sequence of Ruegeria sp. PBVC088, a symbiotic (Gram-negative) bacterium associated with Pyrodinium bahamense var. compressum, which has been associated with harmful algal blooms in the coastal waters of west Sabah, Malaysia. Harmful algal blooms contribute to economic losses for the aquaculture industry, as well as human illnesses and fatalities due to paralytic shellfish poisoning. Bacteria-algae dynamics have posited that the interaction is potentially responsible for the toxin production during a toxic harmful algal bloom event. Despite the expanding body of literature on the capabilities of these bacteria to metabolize, produce, and modify toxins autonomously, it has yet to be confirmed that these toxin-producing bacteria are capable of autonomous toxin synthesis. Saxitoxin, a paralytic shellfish poisoning toxin, is produced by a unique biosynthetic pathway, where the genetic basis for the saxitoxin production was first reported in the saxitoxin-producing cyanobacteria strain Cylindrospermopsis raciborskii T3 (NCBI accession no. DQ787200). The genes responsible for saxitoxin biosynthesis in dinoflagellates, have yet to be fully elucidated. The identification of cyanobacteria saxitoxin biosynthesis genes (sxt) may eventually lead to the identification of homologous genes within the dinoflagellates. Previous studies on the diversity of the bacterial communities associated with the same toxic P. bahamense harmful alga has been carried out by using both the culture-dependent 16S ribosomal RNA gene sequence analysis and culture-independent 16S metagenomic sequence analysis. This study extends the knowledge pertaining to the genomic aspect of an associated bacterium isolated from P. bahamense alga by adopting a whole genome sequencing approach. Here, we report the genome sequencing, de novo assembly, and annotation data of a bacterium, Ruegeria sp. PBVC088, associated with harmful alga P. bahamense, which can be referenced by researchers to identify the genes and pathways related to toxin biosynthesis from a much larger data set. The genome of Ruegeria sp. PBVC088 was sequenced using the Illumina MiSeq platform with 250 bp paired-end reads. The number of reads generated from the MiSeq sequencer was 1,135,484, with an estimated coverage of 100X. The estimated genome size for the marine bacterium was computed to be 5.78 Mb. Annotation of the genome predicted 5,689 gene sequences, which were assigned putative functions based on homology to existing protein sequences in public databases. In addition, annotation of genes related to saxitoxin biosynthesis pathway was also performed. Raw fastq reads and the final version of

the genome assembly have been deposited in the National Center for Biotechnology Information (NCBI) (BioProject: PRJNA324753, WGS: LZNT00000000, SRA: SRR3646181). The genome data provided here are expected to better understand the genetic processes involved in saxitoxin biosynthesis in marine bacteria associated with dinoflagellates.