Dozens of Toxin-Related Genes Are Expressed in a Nontoxic Strain of the Dinoflagellate Heterocapsa circularisquama

Tovah Salcedo,1 Ravi J. Upadhyay,1 Keizo Nagasaki,2 and Debashish Bhattacharya*,1

1Department of Ecology, Evolution and Natural Resources and Institute of Marine and Coastal Sciences, Rutgers University
2National Research Institute of Fisheries and Environment of Inland Sea, Fisheries Research Agency, Hiroshima, Japan

*Corresponding author: E-mail: bhattacharya@aesop.rutgers.edu.
Associate editor: Douglas Crawford

Abstract

The dinoflagellate Heterocapsa circularisquama is lethal to a variety of marine organisms, in particular, commercially important farmed bivalves. Unlike most dinoflagellate toxins, which are polyketides, the only described toxin from H. circularisquama (H2-a) is a porphyrin derivative that functions in light. It is unknown whether H2-a is produced specifically for its lytic properties. We searched for toxin-related genes in the transcriptome of a nontoxic strain of H. circularisquama, and surprisingly found the richest set of toxin-related genes yet described in dinoflagellates. There are 87 distinct expressed sequence tag contigs that encode polyketide synthases and nonribosomal peptide synthases, as well as 8 contigs that are involved in porphyrin biosynthesis. Phylogenomic analysis shows that many toxin-related genes are widely distributed among dinoflagellates. Our data likely indicate a variety of unknown metabolic functions for the toxin-related genes in H. circularisquama because they were identified in a nontoxic strain raised in unialgal culture.

Key words: polyketide synthase, nonribosomal peptide synthase, porphyrin, harmful algal bloom.

Dinoflagellates are globally important aquatic primary producers, best known as the source of harmful algal blooms, or “red tides,” which often pose danger to humans and diverse marine organisms (Taylor 1990). The evolution and ecology of toxicity remains poorly understood. For example, Karlodinium veneficum uses karlotoxins during predation (Sheng et al. 2010), but it has been proposed that the closely related Karenia brevis produces brevetoxins for osmoregulation, and its harmful effects may be incidental (Errera and Campbell 2011). Both of these and the majority of known dinoflagellate toxins are produced by polyketide synthase (PKS) and nonribosomal peptide synthase (NRPS) genes (Kellmann et al. 2010). Given the absence of complete genome data from dinoflagellates, the number of toxin-related genes in these taxa, the evolutionary origins of these sequences, and their genomic organization (e.g., linkage, sequential organization) are unknown.

Recent advances in sequencing technology have made it possible to generate genome-wide data from nonmodel ecologically important taxa (Ekblom and Galindo 2011), including dinoflagellates, which have immense nuclear genomes ranging in size from 1.5 to 220 Gbp (LaJeunesse et al. 2005; Wiseceaver and Hackett 2011). Here, we studied Heterocapsa circularisquama, the only dinoflagellate for which a naturally occurring (single stranded) ssRNA virus has been implicated in bloom demise (Nagasaki 2008). Heterocapsa circularisquama is polymorphic for killing diverse organisms (e.g., Uchida et al. 1995; Kamiyama and Arima 1997; Kim et al. 2000), although the nature of toxicity remains unclear. Only one substance capable of causing in vitro toxicity has been isolated thus far from H. circularisquama: a porphyrin derivative that requires light to cause cell lysis (H2-a; Miyazaki et al. 2005). We asked two questions in our work: 1) which toxin genes are present in the transcriptome of H. circularisquama, relating to H2-a or polyketides, and 2) is the sequenced strain of H. circularisquama lethal using an in vitro rotifer model?

We generated a single lane of Illumina high-throughput sequence data (150 × 150 bp paired-end reads) to assemble the transcriptome of H. circularisquama strain HCLG-1. This resulted in 1.3 Gbp of sequence data that were assembled into 46,344 expressed sequence tag (EST) contigs of length >200 bp with ≥4× average contig coverage. These unique EST contigs likely represent >80% of the genes in H. circularisquama based on in-depth transcriptome analysis of the saxitoxin producing dinoflagellate Alexandrium tamarense (Moustafa et al. 2010). Blast2GO was used to assign annotations to 15,523 EST contigs, 87 of which are likely toxin related, including 61 PKS- and 13 NRPS-encoding contigs (Table 1). This is significantly more than the eight PKS genes previously identified in K. brevis (Monroe and Van Dolah 2008) or the 23 putative PKS/NRPS genes present in A. tamarense (Salcedo T, unpublished data). From our phylogenomic analysis, 29 of 75 PKS/NRPS-encoding EST contigs could be aligned with published sequences (which includes the GenBank RefSeq and dbEST databases, TBoostDB, and data from the Joint Genome Institute), and 15 of those alignments included sufficient taxa and amino acid positions to build trees (supplementary fig. 1, Supplementary Material online). Of these maximum likelihood phylogenies, seven include only dinoflagellates and six include dinoflagellates and other lineages, but support a shared gene history amongst at least two dinoflagellate genera. These results are consistent with a BLASTx analysis.
using the 15 EST contigs as queries against the GenBank "nr" database (results not shown). Of the remaining trees, one (contig 30097) groups *H. circularisquama* and *H. triquetra* sequences but lacks other dinoflagellates, whereas the second (contig 51475) groups *H. circularisquama* with bacteria. This latter EST contig is polyadenylated but the genomic copy, found within a partial assembly of the *H. circularisquama* genome (Salcedo T, Bhattacharya D, unpublished data), did not contain identifiable intron splice sites. Whereas, it is unclear whether introns are frequent in dinoflagellate nuclear genes (reviewed in Lin 2011), we cannot currently rule out the possibility that this sequence is a bacterial contamination. Proteins encoded by at least two *H. circularisquama* EST contigs were present in 10 of the 15 trees, consistent with the existence of PKS/NRPS gene families, and in some cases suggestive of *H. circularisquama*-specific gene family expansions (e.g., contig 4065, contig 2305). Taken together, the phylogenomic data suggest that many toxin-related genes have a shared evolutionary history with distantly related dinoflagellates, despite the relatively small number of published toxin-related gene sequences from this group and the likely failure to detect or align highly diverged orthologs within what may be lineage-specific multigene families.

An additional eight EST contigs were found to encode proteins present in the porphyrin pathway (Table 1), confirming the expression of necessary precursors for H2-a; however, the enzymes for finalizing the toxin structure are not known (Miyazaki et al. 2005). By comparing EST contigs with the preliminary draft genome assembly, we identified at least one intron splice site in most toxin-related EST contigs (Table 1), and found that none of the EST contigs mapped to genomic DNA derived from naturally occurring intracellular bacteria (Maki and Imai 2001) in our unialgal culture. These data demonstrate a dinoflagellate origin of the majority of toxin-related EST contigs. We also identified possible homologs for 12 of the 26 genes thought to be involved in biosynthesis of the polyketide saxitoxin (e.g., Moustafa et al. 2009; Stüken et al. 2011; supplementary table 1, Supplementary Material online).

In *K. brevis*, seven of eight PKS genes show sequence similarity to, and thus potentially share a common evolutionary history with, Type I PKSs, but are single-domain genes, like Type II PKSs (Monroe and Van Dolah 2008). We found one EST contig that encodes two Type I PKS domains (fig.1), but this sequence did not have a significant BLAST hit to any other sequence in our local database. Because our data do not include the dinoflagellate splice leader sequence (Lidie and Van Dolah 2007) for every EST contig, there may be additional multidomain PKS genes present in *H. circularisquama* that were missed due to incomplete 5′-termini of cDNAs.

All EST contigs in our analysis were derived from strain HCLG-1, initially selected for its rapid growth rate. Because

### Table 1. Categories of Toxin-Related Genes by Type of Toxin or Pathway.

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Number of ESTs</th>
<th>Mean Amino Acid Lengtha</th>
<th>Number of ESTs with Intron Supportb</th>
<th>Contig ID Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acid synthesisc</td>
<td>12</td>
<td>1142 (599-3365)</td>
<td>9</td>
<td>30162, 32575, 33540, 39817, 29948, 40879, 42157, 86385, 43817, 64052, 6473, 66200</td>
</tr>
<tr>
<td>Hybrid PKS/NRPS</td>
<td>1</td>
<td>1210 (n/a)</td>
<td>1</td>
<td>114413</td>
</tr>
<tr>
<td>PKS</td>
<td>61</td>
<td>542 (214-1160)</td>
<td>44</td>
<td>104138, 111988, 112446, 113133, 113953, 113953, 1398, 2078, 2305, 2599, 29762, 3030, 30829, 31316, 32871, 33328, 33604, 34123, 34980, 35211, 4065, 40970, 43779, 43836, 45090, 51475, 60224, 62363, 63067, 6353, 87808, 89408, 89953, 91665, 104400, 112378, 120309, 15094, 16922, 1819, 30108, 33248, 34496, 63025, 8159, 99209, 10137, 112102, 112293, 12613, 30178, 31982, 32837, 36924, 37397, 4077, 40822, 46850, 65336, 68654, 51603, 7276, 10743, 111965, 1641, 30097, 35797, 41902, 44662, 59419, 59527, 63162, 64374, 9741, 112741</td>
</tr>
<tr>
<td>NRPS</td>
<td>13</td>
<td>471 (330-1084)</td>
<td>9</td>
<td>111969, 112302, 112706, 37431, 52339, 62001, 86724, 46238</td>
</tr>
<tr>
<td>Precursors for porphyrin</td>
<td>8</td>
<td>879 (219-1678)</td>
<td>7</td>
<td>111969, 112302, 112706, 37431, 52339, 62001, 86724, 46238</td>
</tr>
</tbody>
</table>

a Values in parentheses indicate minimum and maximum lengths.
b Intron splice junctions predicted using Augustus (see Methods).
c Due to cross talk between fatty acid biosynthesis and PKS/NRPS biosynthesis, we cannot differentiate toxic from nontoxic genes.

![Fig. 1](image_url). Schematic image of an EST contig that encodes two PKS domains. Bioinformatic analyses that support each domain are listed under the domains (see methods in supplemental document 1, Supplementary Material online).
not all isolates of *H. circularisquama* cause harm in in vitro assays (Kim et al. 2002), we performed a mortality assay using a marine rotifer to describe one toxin producing phenotype of HCLG-1. Surprisingly, HCLG-1 did not kill rotifers (fig. 2). What, then, explains the expression of dozens of potentially toxin-related genes by these nontoxic cells? Furthermore, what role might PKS genes play in a species in which the only known toxin pathway (in toxic strains) presumably does not involve these proteins? With regard to the first question, it is possible that toxin genes in dinoflagellates are constitutively expressed and regulated posttranscriptionally, as was proposed for nontoxic *K. brevis* strains that express PKS genes (Monroe et al. 2010). In fact, Moustafa et al. (2010) showed that 73% of the transcriptome (including the putative PKS and NRPS genes) in *A. tamarense* shows no significant change in gene expression across different culture conditions (i.e., comprises a “core” transcriptome). Regarding the second question, given that there is a known overlap between PKS biology and fatty acid biosynthesis (Rein and Borrone 1999), dinoflagellates may use products of PKS genes for other functions. Without a robust genome sequence and comprehensive description of metabolites, it remains unclear how the dozens of unique PKS EST contigs identified here map to PKS pathways, and how those pathways relate to toxin phenotypes, in particular given phylogenic support for dinoflagellate-specific genes. Additional transcriptomic and metabolic data from strains capable of killing rotifers in vitro are required to help clarify what role these PKS genes may play in the biology of *H. circularisquama*. Regardless, *H. circularisquama* clearly harbors a large genomic arsenal to support production of both polyketide toxins known from other dinoflagellates and the nonpolyketide H2-a toxin already implicated in shellfish death. These data provide testable hypotheses for future research into the ecology and biochemistry of dinoflagellate toxicity.

**Methods**

Methods are described in supplementary document 1, Supplementary Material online. Toxin-related EST contigs found in *H. circularisquama* are available in supplementary document 1, Supplementary Material online.

**Supplementary Material**

The supplementary document 1, figure 1, and table 1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This work was supported by generous start-up funds from Rutgers University and grants from the National Science Foundation to D.B. We thank Cheong Xin Chan for help with phylogenomic analysis. We thank Divino Rajah for technical assistance with Illumina sequencing, Dana Price for bioinformatic support, Emily Monroe for helpful discussion, and Paul Falkowski for use of his lab facilities. We thank Kay Bidle (Rutgers) for critical project input and Yuji Tomaru (Fisheries Research Agency) for supplying algal strains. Two anonymous reviewers provided valuable comments on this manuscript.

**References**


LaJeunesse TC, Lamber G, Andersen RA, Coffroth MA, Galbraith DW. 2005. *Symbiodinium* (Pyrrhophyta) genome sizes...
(DNA content) are smallest among dinoflagellates. J Phycol. 41:880–886.