Polyketide synthase genes and molecular trade-offs in the ichthyotoxic species Prymnesium parvum

Anestis, Konstantinos; Kohli, Gurjeet Singh; Wohlrab, Silke; Varga, Elisabeth; Larsen, Thomas Ostenfeld; Hansen, Per Juel; John, Uwe

Published in:
Science of the Total Environment

DOI:
10.1016/j.scitotenv.2021.148878

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Polyketide synthase genes and molecular trade-offs in the ichthyotoxic species *Prymnesium parvum*

Konstantinos Anestis a, Gurjeet Singh Kohli a, Sylke Wohlrab a,b, Elisabeth Varga c, Thomas Ostenfeld Larsen d, Per Juel Hansen e, Uwe John a,b,⁎

a Ecological Chemistry, Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570, Bremerhaven, Germany
b Helmholtz Institute for Functional Marine Biodiversity, Ammerländer Heerstraße 231, 26129 Oldenburg, Germany
c Department of Food Chemistry and Toxicology, Faculty of Chemistry, University of Vienna, Währinger Straße 40, 1090 Vienna, Austria
d Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 221, 2800 Kongens Lyngby, Denmark
e Marine Biology Section, University of Copenhagen, Strandpromenaden 5, 3000 Helsingør, Denmark

HIGHLIGHTS

- The ichthyotoxic species *Prymnesium parvum* produces the lytic toxins, prymnesins.
- Nine *P. parvum* strains with various toxin contents were studied for the presence of polyketide synthase genes.
- Numerous polyketide synthase genes and their functional domain organization are reported for all nine strains.
- Gene expression analysis highlights the occurrence of molecular trade-offs related to cellular toxin content.

GRAPHICAL ABSTRACT

ABSTRACT

*Prymnesium parvum* is a bloom forming haptophyte that has been responsible for numerous fish kill events across the world. The toxicity of *P. parvum* has been attributed to the production of large polyketide compounds, collectively called prymnesins, which based on their structure can be divided into A-, B- and C-type. The polyketide chemical nature of prymnesins indicates the potential involvement of polyketide synthases (PKSs) in their biosynthesis. However, little is known about the presence of PKSs in *P. parvum* as well as the potential molecular trade-offs of toxin biosynthesis. In the current study, we generated and analyzed the transcriptomes of nine *P. parvum* strains that produce different toxin types and have various cellular toxin contents. Numerous type I PKSs, ranging from 37 to 109, were found among the strains. Larger modular type I PKSs were mainly retrieved from strains with high cellular toxin levels and eight consensus transcripts were present in all nine strains. Gene expression variance analysis revealed potential molecular trade-offs associated with cellular toxin quantity, showing that basic metabolic processes seem to correlate negatively with cellular toxin content. These findings point towards the presence of metabolic costs for maintaining high cellular toxin quantity. The detailed analysis of PKSs in *P. parvum* is the first step towards better understanding the molecular basis of the biosynthesis of prymnesins and contributes to the development of molecular tools for efficient monitoring of future blooms.

© 2021 Published by Elsevier B.V.
1. Introduction

Worldwide blooms of the haptophyte Prymnesium parvum occur frequently and negatively impact fishing and aquaculture industries (Roelke et al., 2011, 2016). P. parvum is a cosmopolitan species, that owing to its high plasticity, can be found in freshwater, brackish and marine ecosystems (Granéli et al., 2012). Historically, much of the knowledge on the toxic effects of P. parvum on other organisms have been derived from simple bioassay studies, and not via quantitative measurements of the causative toxins. Nutrient availability and the element composition in the water have been shown to influence the cellular and extracellular levels of the toxins. In general, nutrient deplete conditions induce higher toxicity in the bioassays. For example, toxicity towards fish erythrocytes was highly enhanced with phosphorous limitation (Beszteri et al., 2012), whereas high light availability and high temperature seem to reduce the toxic effects of prymnesins (Qin et al., 2020; Taylor et al., 2021). P. parvum belongs to the constitutive mixoplankton (Flynn et al., 2019), and can feed on bacteria and other protists. The release of toxins by P. parvum mediates feeding through the immobilization and lysis of its prey (Skovgaard and Hansen, 2003; Tillmann, 2003). However, attempts to demonstrate enhanced feeding under nitrogen and phosphorous limitation have not been successful (Skovgaard et al., 2003; Lundgren et al., 2016).

The toxicity of P. parvum has been attributed to its ability to produce large ladder-frame polyether compounds known as prymnesins (Igarashi et al., 1999). Prymnesins belong to the group of polyketides, an extremely diverse group of compounds produced by a wide range of organisms, from prokaryotes, unicellular algae to higher eukaryotes (Wright and Cembella, 1998; Staunton and Weissman, 2001). Prymnesins were first characterized back in 1999 and reported as prymnesin 1 and 2 (Igarashi et al., 1999). Recently, a new B-type of prymnesin was isolated and structurally characterized by NMR (Rasmussen et al., 2016), and high-resolution mass spectrometric analyses demonstrated the existence of an additional third type of prymnesin, altogether leading to the classification of three prymnesin types, the A-, B- and C-types (Rasmussen et al., 2016). The three types differ in the size of the backbone: the A-type being the longest with 91 carbons, followed by B-type and C-type with 85 and 83 carbons, respectively. Within each type there is a huge structural diversity due to variances in chlorination, saturation and attached sugar moieties, and altogether 51 prymnesin congeners have been reported so far (Binzer et al., 2019). The chemical diversity of polyketides is reflected by a plethora of associated biological functions including anticancer and immunosuppressive properties (Martinez Andrade et al., 2018).

The carbon chain of polyketide compounds is synthesized by polyketide synthases (PKSs) (Rein and Borrone, 1999). PKS related genes share a common evolutionary history with fatty acid synthases (FASs) (Kohli et al., 2016). Both PKSs and FASs have a standard core structure, where the ketosynthase domain (KS) catalyzes the condensation of the acyl units in synergy with an acyltransferase (AT) and an acyl carrier protein (ACP) (Cane et al., 1998; Jenke-Kodama et al., 2005). After the condensation of an acyl unit, the product can be subject to further modifications by the presence of other domains such as dehydratase (DH), enoyl reductase (ER) and ketoreductase (KR). These domains can respectively produce a double bond, a fully-reduced methylene or a hydroxy group (Weissman, 2015). As a result, the organization of these domains has a different effect on the product’s final structure. The biosynthesis of a polyketide is terminated by the presence of a thioesterase domain (TE), which hydrolyzes the polyketide compound from the ACP.

Type I PKSs can be iterative, possessing all the catalytic domains in a single protein, and can elongate a chain in a repeated way, functioning like fatty acid synthases in animals and fungi. Modular type I PKSs occur as distinct modules, where each module contains all the prerequisite domains needed for catalyzing the condensation reaction, leading to the elongation of the polyketide chain by two carbon units. In type II PKSs, each catalytic domain is a distinct protein and they function independently and iteratively to type II FASs in bacteria and plants (Cane et al., 1998; Jenke-Kodama et al., 2005). Type III PKSs also function in an iterative manner and consist of self-contained homodimeric enzymes with each monomer catalyzing a specific function. Type III PKSs differ from the other PKS types due to their ability to perform condensation without using acyl carrier proteins (Ferrer et al., 1999).

The role of polyketide synthases in the biosynthesis of marine secondary metabolites has driven extensive studies to elucidate their presence in marine protists and their evolutionary history (e.g. John et al., 2008; Monroe and Dolah, 2008; Kohli et al., 2016). Genome- and transcriptome-based studies on dinoflagellates have shed light on the diversity and presence of PKSs in this potentially toxic and phylogenetically diverse group. The majority of PKS transcripts from dinoflagellates belong to type I PKSs, alongside sequences from other protists. Both modular and single mono-functional domains are present, with the latter showing similar structure to type II PKSs (Monroe and Dolah, 2008). Large modular type I PKSs have been found in many dinoflagellate species including the ciguatoxin-producing Gambierdiscus polysynesiensis, palytoxin-like producing Ostreopsis species and the brevetoxin producing Karenia brevis (Kohli et al., 2017; Van Dolah et al., 2017; Verma et al., 2019). Diatoms are another diverse group of marine protists with the potential to be toxic. The Pseudo-nitzschia genus is known for the production of the neurotoxin domoic acid and the combination of transcriptomic and biochemical approaches has led to the identification of the complete biosynthetic pathway of domoic acid (Brunson et al., 2018; Haroardóttir et al., 2019). The lack of tools to genetically manipulate the majority of toxins producing organisms has also been an obstacle in incorporating functional genomics in the study of the molecular mechanisms involved in polyketides’ biosynthesis (Kohli et al., 2016; Lauritano et al., 2019).

In haptophytes, transcriptomic surveys of 12 strains indicated the presence of type I PKSs (Kohli et al., 2016). However, not all strain screened in this study are toxic producing and given the involvement of PKSs in a wide range of biosynthetic pathways, no direct connection to toxin production was feasible. The Chrysochromulina genus contains species that produce ichthyotoxic compounds of unknown structure, but presumably similar to prymnesins. The only available transcriptomic study comes for the species Chrysochromulina polyepis, where thirteen putative PKSs were found based on expressed sequence tags (John et al., 2010). In comparison to dinoflagellates, few detailed studies have been conducted concerning the potential involvement of PKSs in toxin biosynthesis in haptophytes (Freitag et al., 2011; Beszteri et al., 2012).

In the current study, we compared the prymnesin profile, the transcriptomes with a focus on the presence of PKSs, and the corresponding expressed metabolic and cellular functions of nine P. parvum strains. Given the existence of three different prymnesin types, we chose three strains for each of the prymnesin types, A, B and C. The PKS transcripts derived from the transcriptomes were analyzed through phylogenetics of the KS domain to assess the relationship between prymnesin type and PKS genes to get insights into PKSs evolution within haptophytes and P. parvum in particular. Furthermore, we measured the cellular toxin content of all nine strains, relating patterns between toxin types and cellular toxin contents with the presence of PKS genes. Moreover, we investigated gene expression pattern of cellular, metabolic and regulative processes in relation to the cellular toxin and composition to understand the potential molecular and metabolic trade-offs of toxin production in P. parvum.

2. Materials and methods

2.1. Cultures

Nine strains of P. parvum were obtained from various culture collections, and further information on the year and site of isolation is provided in the Supplementary information. All strains were grown at...
17 °C using standard K-medium prepared with sterile filtered North Sea seawater of a salinity of 30. All strains were grown with a light:dark cycle of 16:8 h and a light intensity of 80 μmol photons·m⁻²·s⁻¹. Prior to establishing the experimental cultures, the cells were rendered axenic using a cocktail of antibiotics (165 μg·mL⁻¹ ampicillin, 33.3 μg·mL⁻¹ gentamicin, 100 μg·mL⁻¹ streptomycin, 1 μg·mL⁻¹ chloramphenicol, 10 μg·mL⁻¹ ciprofloxacin). The antibiotic treatment lasted 4 days and was performed a second time after one week. The axenicity of the cultures was validated by fluorescence microscopy after staining with 4′,6-diamidino-2-phenylindole (DAPI).

2.2. Toxin extraction, quantification and correlation analysis

Toxin extraction and quantification were performed according to the protocol described in Svenssen et al., 2019, with small modifications. In general, the biomass on each filter was extracted twice with two 20 mL MeOH each for 30 min using an ultrasonic batch. The samples were centrifuged in between at 4300 × g for 15 min at 4 °C. The combined extract (40 mL) was evaporated to dryness using a CentriVap Benchtop Vacuum Concentrator (Labconco Corporation, Kansas City, MO, USA) at 35 °C. Reconstitution was performed with 1 mL methanol:H₂O (90:10, v:v) and short-time ultrasonic bath treatment. The solution was centrifuged to separate residues from the glass fiber filters, and the supernatant was transferred to an HPLC glass vial. HPLC-FLD measurements were performed after derivatization with the AccQ-Fluor Reagent Kit (Waters Cooperation, Milford/MA, USA) with a 1200 HPLC system (Agilent Technologies, Waldbronn, Germany) using fumonisins B₁ and B₂ as external calibrants due to the lack of standards. Prymnesins and fumonisins share a primary amine group as structural feature important for the derivatization process using AccQ-Fluor. The obtained results are an approximation of the Prymnesin content in the samples. HPLC-HRMS-measurements were performed to confirm the presence of Prymnesins and identify the specific Prymnesin analogues using a 1290 UHPLC system coupled to a 6550 iFunnel QTOF LC/MS (both from Agilent Technologies). Chromatographic separation was achieved with a Kinetex F5 (2.1 × 100 mm, 2.6 μm, Phenomenex, Aschaffenburg, Germany) column using a water-acetonitrile gradient (eluent A: H₂O, eluent B: acetonitrile: H₂O (90:10, v:v)) and short-time ultrasonic bath treatment. The solution was centrifuged to separate residues from the glass fiber filters, and the supernatant was transferred to an HPLC glass vial. HPLC-FLD measurements were performed after derivatization with the AccQ-Fluor Reagent Kit (Waters Cooperation, Milford/MA, USA) with a 1200 HPLC system (Agilent Technologies, Waldbronn, Germany) using fumonisins B₁ and B₂ as external calibrants due to the lack of standards. Prymnesins and fumonisins share a primary amine group as structural feature important for the derivatization process using AccQ-Fluor. The obtained results are an approximation of the Prymnesin content in the samples. HPLC-HRMS-measurements were performed to confirm the presence of Prymnesins and identify the specific Prymnesin analogues using a 1290 UHPLC system coupled to a 6550 iFunnel QTOF LC/MS (both from Agilent Technologies). Chromatographic separation was achieved with a Kinetex F5 (2.1 × 100 mm, 2.6 μm, Phenomenex, Aschaffenburg, Germany) column using a water-acetonitrile gradient (eluent A: H₂O, eluent B: acetonitrile: H₂O (90:10, v:v)). The mass spectrometer was operated in the positive ionization mode in a scanning range of m/z 250 to 1700 with 3 scans per second. The total amount of extracted Prymnesins was divided by the total number of harvested cells in order to obtain the concentration of Prymnesins per cell.

A linear model was used in order to test for a potential effect of growth rates on toxin contents and the toxin data were therefore power transformed (x^0.25) to achieve a normal distribution of the residuals. All analysis were performed with the R base packages.

2.3. RNA extraction, library generation and assembly analysis

Cells for RNA extraction were harvested at the same time point as for toxin extraction (initial exponential phase and cell concentration of ~100,000 cells mL⁻¹) by centrifugation at 1500 × g for 10 min and the obtained pellet was transferred to 1 mL of Triragent mixed with glass beads. RNA isolation was performed as described in Wohlrab et al. (2017). Libraries for sequencing were prepared using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, USA). The reads were trimmed using the CLC Genomics Workbench with the default settings. Trinity (v2.8.4) was used for assembling, with a set-up minimum contig length of 300 bp (Haas et al., 2013). The total contigs were clustered together with a similarity threshold of 98% using CD-hit (Fu et al., 2012) while contigs containing homopolymer stretches were removed using an in-house script. The quality of the assemblies was assessed using BUSCO and reference datasets for both Stramenopiles and Alveolates (Waterhouse et al., 2017).

The assemblies were screened for the presence of NRPS, PKS and FAS transcripts using HMMER (Finn et al., 2011) and an in-house HMM database with an applied E-value cutoff of ≤10⁻¹⁰ as described in Kohli et al., 2016. The sequences coding for PKS domains were afterward further analyzed using Pfam (Punta et al., 2012) and conserved domain searches (Marchler-bauer et al., 2017) to get the structure of PKS and NRPS/PKS contigs. The annotation of the contig domains, protein alignments and phylogenetic analyses were all performed with the Geneious software (Kearse et al., 2012). The sequences were aligned using MAFFT (Katoh et al., 2002), while phylogenetic analysis was carried out using RaxML (Stamatakis, 2006) and selected the LG model of rate heterogeneity with 1000 bootstraps (Le and Gascuel, 2008). The reference assembly was annotated using Trinotate v3.0.2 and the top blastx hit with e-value ≤10⁻¹⁵ was selected in order to increase the accuracy of protein assignment. The resulting Pfam domains were assigned to KEGG Orthologies which were further used for the gene expression analysis.

The gene expression variance analysis was done in the R environment using the ‘variancePartition’ package (Hoffman and Schadt, 2016). The read counts were normalized based on the total number of reads per sample and further variance stabilizing transformation with the Deseq2 package (Love et al., 2014), as recommended by Hoffman and Schadt (2016).

3. Results and discussion

3.1. Strain phylogeny and toxin content

The transcriptome derived 18S rDNA sequences were used to construct a phylogenetic tree. In this way, we validated that no cross-contamination occurred during the experiments and subsequent extractions. The phylogenetic placement of the strains (Fig. 1) is in accordance with the findings from which have used the internal transcribed spacer (ITS) of the rRNA operon and supports the concept of the monophyletic origin of the Prymnesin-type based on the producing strains of a given phylogenetic clade. The strains have been isolated from different areas of the world and no clear pattern was observed between their geographic distribution and phylogenetic placement.

In the current study, the cellular toxin contents varied among the nine strains of P. parvum. We confirmed the presence of described Prymnesin analogues (Supplementary information), which have been previously detected. It is yet unknown, if and at which level the different analogues contribute to the potential toxic effects. Comparison of toxic effects of P. parvum strains on a fish species (Onchorhynchus mykiss) and a microalga (Teleaulax acuta) gave no correlation with strains low in toxicity towards the microalga being highly toxic towards fish (Blossom et al., 2014); however these strains produced different types of Prymnesins. In the current study, toxin contents were at the same level for A-type producing strains, which contained 0.045 to 0.076 fmol of toxin per cell. In B-type producing strains, two of them contained the highest amounts of toxin in this study, with an average of 0.63 fmol per cell for K-0081 and 0.533 fmol per cell for K-0374. Among the C-type producing strains, strains K-0252 and NIES-1017 contained very low, almost undetectable amounts of Prymnesins, while RCC-1436 contained the highest amount of Prymnesins. In a previous study, Svenssen et al. (2019) developed an indirect method for estimating the Prymnesin content of B-type producing strains, and showed that the majority of the toxin is found intracellular rather than being released into the medium. In the same study, strains K-0081 and K-0374 were selected for Prymnesin B-type quantification. The amount of cellular toxin content varied from our finding, with K-0081 containing ~5 times more Prymnesin than K-0374. These findings highlight the importance of the applied and well controlled culturing methods as probably one of the essential drivers for toxin production and allelopathy in P. parvum. The cellular toxin content could be indicative for the potential allelopathy of P. parvum; therefore, sensitive but also ecological relevant bioassays should be used in order to elucidate the relation between these two parameters: cellular toxin content and expressed extra cellular lytic toxicity.
3.2. De novo transcriptomic assembly

On average 149 ± 34 million reads were generated for every strain (Table 1). The raw read sequences have been deposited at NCBI (National Center for Biotechnology Information) under the BioProject PRJNA718746. Strain specific de novo transcriptomes were assembled. After clustering all contigs with a similarity of 98% and a minimum length of 300 bp, the number of contigs varied among the strains, from 56,645 to 80,915 (Table 1). According to BUSCO (Benchmarking Universal Single-Copy Orthologs), the completeness of the assemblies was evaluated using as reference a total of 171 highly conserved proteins belonging to Stramenopiles and Alveolates (Waterhouse et al., 2017). The percentage of total proteins with a match in either complete single or duplicated copy varied with an average of 78.06 ± 2.56. Previous studies of haptophytes transcriptomic PKS surveys have not analyzed their completeness according to recent tools and thus a direct comparison is not possible (e.g. John et al., 2010; Beszteri et al., 2012). The reference assembly used for the gene expression variance analysis had a BUSCO completeness of 81.5%. The total number of transcripts was 186,144 and KEGG orthologies were assigned for 44,408.

3.3. Type I modular polyketide synthase encoding transcripts

Modular type I PKSs consist of consecutive modules, each of them catalyzing the addition of an extender unit (acyl group) to the growing polyketide product. The sequence of the domains on each module and their combination decides the nature of the final chemical structure of the produced compound. As a result, the functional annotation and domain organization of modular Type I PKSs can be an important step towards correlating which gene clusters that encode the enzymes making a given carbon backbone. Our analyses revealed the presence of a plethora of PKS-related contigs in all strains of *P. parvum*. The

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Type A</th>
<th></th>
<th></th>
<th>Type B</th>
<th></th>
<th></th>
<th>Type C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NIES-1812</td>
<td>PPDW-02</td>
<td>UTEX-2797</td>
<td>K-0081</td>
<td>UIO-223</td>
<td>K-0374</td>
<td>K-0252</td>
<td>NIES-1017</td>
</tr>
<tr>
<td>Assembled reads</td>
<td>134,628,014</td>
<td>173,930,042</td>
<td>108,806,618</td>
<td>174,294,139</td>
<td>177,225,633</td>
<td>116,245,080</td>
<td>124,143,170</td>
<td>129,669,753</td>
</tr>
<tr>
<td>Total contigs ≥ 300 bp</td>
<td>59,570</td>
<td>64,714</td>
<td>61,656</td>
<td>60,561</td>
<td>54,645</td>
<td>80,915</td>
<td>60,981</td>
<td>58,249</td>
</tr>
<tr>
<td>BUSCO % completeness</td>
<td>70.85</td>
<td>74.91</td>
<td>83.76</td>
<td>80.81</td>
<td>77.12</td>
<td>73.80</td>
<td>81.92</td>
<td>84.13</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>56.3</td>
<td>55.7</td>
<td>56.9</td>
<td>56.8</td>
<td>56.3</td>
<td>55.8</td>
<td>56.9</td>
<td>57</td>
</tr>
<tr>
<td>Total PKS related transcripts</td>
<td>109</td>
<td>96</td>
<td>91</td>
<td>37</td>
<td>67</td>
<td>68</td>
<td>64</td>
<td>83</td>
</tr>
<tr>
<td>Total multi-domain transcripts</td>
<td>49</td>
<td>47</td>
<td>35</td>
<td>25</td>
<td>46</td>
<td>41</td>
<td>33</td>
<td>42</td>
</tr>
<tr>
<td>#KS domains</td>
<td>112</td>
<td>100</td>
<td>84</td>
<td>68</td>
<td>77</td>
<td>83</td>
<td>67</td>
<td>89</td>
</tr>
<tr>
<td>#KR domains</td>
<td>25</td>
<td>35</td>
<td>19</td>
<td>50</td>
<td>29</td>
<td>28</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>#DH domains</td>
<td>23</td>
<td>25</td>
<td>11</td>
<td>32</td>
<td>31</td>
<td>21</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>#ER domains</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>#ACP domains</td>
<td>62</td>
<td>69</td>
<td>42</td>
<td>70</td>
<td>51</td>
<td>52</td>
<td>31</td>
<td>49</td>
</tr>
<tr>
<td>Complete modules at longest transcript</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 1. Phylogeny of 18S rRNA extracted from the transcriptomes of all nine *P. parvum* strains (A). Phylogenetic analysis of type I polyketide synthases (PKSs) and non-ribosomal polyketide synthetases/synthase (NRPSs) ketoacyl synthase (KS) domains. The alignment included 412 KS sequences, out of which 208 belong to the nine *P. parvum* strains. KS domains with an identity of ≥98% were excluded in order to avoid redundancy and to improve the alignment. RaxML with 1000 bootstraps was used for the construction of the phylogenetic tree.
The total number of PKS-related contigs varied from 37 to 109. The KS domain catalyzes the condensation of carbon units and is highly conserved, thus making it the most suitable one for phylogenetic purposes (John et al., 2008). Therefore, the resulting KS domains of *P. parvum* were phylogenetically analyzed in order to identify type I PKS transcripts. Additional KS domains from other haptophytes, or corresponding outgroup species (Kohli et al., 2016), were used to check for common clades found within haptophytes (Fig. 1). The KS domains were divided based on whether they coded the full or partial domain and their presence on multi- or single-domain contigs (Supplementary information). In general, full KS domains in multi-modular transcripts accounted for 24.6–72.6% of the total KS domains. The partial KS domains in multi-modular transcripts had a relative contribution ranging from 12.3–33.3%, while KS domains in single-domain transcripts accounted for 15.1–55.3% of the total KS domains.

The resulting phylogeny is in its general topology in accordance with previous studies that involved KS domains from a wide range of protists (John et al., 2008; Kohli et al., 2016). The *P. parvum* ketoacyl synthase (KS) domains fell into a well-supported haptophyte specific group (Fig. 2). Within this group, the *P. parvum* KS domains dominated in three clades. Two of those clades, A and C, were well supported with both of them having a bootstrap value of 100. In clade A, all KS domain sequences belonged to *P. parvum*, while in clade C there was a KS domain belonging to *Chrysochromulina brevisillum*, another prymnesiophyte. All KS domains we examined for the amino acid sequence of their active sites and the level of conservation. In principle, active sites are conserved among the sequences with some exceptions, which presented clade specific patterns. In type I PKSs, conserved active site residues seem to be important for the functionality of the KS domains (Kwon et al., 1998). In clade A, the amino acid sequence QGLGS was found in all KS domains and is divergent to the expected HGTGT motif. Moreover, in clade C, some KSs possessed NTACS instead of the expected DTACS. However, the active site C, which is responsible for the decarboxylative condensation, was maintained in all sequences. The divergent active site residues could attribute different function to enzymes, something indicated by the formation of distinct clades by the corresponding KS sequences (Eichholz et al., 2012).

Besides having *P. parvum* specific KS clades, no clear relation to the prymnesin type produced by the strains was found. Moreover, KS domains belonging to the same modular PKS contigs were not only found in the *P. parvum* dominated ones, but also fell into mixed haptophyte clades. The high number of KS domains and the corresponding other functional domains found in *P. parvum* and other species indicate a relaxed selection pressure leading to highly diverse domains driven by evolutionary processes such as gene duplications and domain shuffling (Beedessee et al., 2019). Moreover, no clear pattern was observed in terms of the total number of retrieved KS domains and the level of cellular toxin content. Nonetheless, a high number of modular PKS genes was found in the highly toxic strain K-0081 and most KS domains (84%) made part of modular PKS transcripts.

Modular type I PKSs have been detected in many marine protists and can have considerable gene length as indicated by their products (John et al., 2008). In the current study, larger transcripts were found in strain K-0081, with a 41,957 nt long transcript (K0081_21_c0_g1_i7_frame3) encoding for a 10-module type I PKS. When further assembled, the total

---

**Figure 2.** Detailed phylogeny of type I polyketide synthases (PKSs) and non-ribosomal polyketide synthases (NRPS) ketoacyl synthase (KS) domains. Outgroups correspond to those indicated in Fig. 1. The outer ring represents the prymnesin type that is produced by the corresponding strain. In clades A, B and C, *P. parvum* ketoacyl synthase (KS) domains dominated.
length of the contig was 49,007 nt and 11 modules as it assembled with K0081_21_c0_g3_l1_frame1r and had an overlap of 1473 nt and two mismatches. In total, we identified the presence of eight consensus modular type I PKS contigs in all nine strains (Fig. 3B). The identification of the consensus contigs was based upon the identity of the KS domains of the corresponding transcripts and the overall domain organization of the contigs (Fig. 3B). However, not the same contig length was found in all strains. In general, long modular type I PKS contigs were retrieved from the highly toxic B strains and smaller contigs from the other strains aligned to them. All PKSs retrieved by each strain and their functional domain organizations can be found at the Supplementary information.

Genome wide studies to mine PKSs from microalgae has led to the description of large modular type I PKSs, up to 14 modules long (Sasso et al., 2012). In Ostreococcus sp., PKSs account for 1.5% of the genome, illustrating the increased fitness the products of these genes attribute to their holders (John et al., 2008). The diversity of polyketides is not only restricted in their biosynthetic mechanisms, but also in their function. For example, in *Chlamydomonas reinhardtii* a type I PKS gene with an estimated mass of 2.3 MDa is involved in zygospore maturation (Heimerl et al., 2018). High number of KS encoding transcripts, 115 and 121 for ichthyotoxic *Ostreopsis* sp. and *K. brevis* respectively, have been discovered (Van Dolah et al., 2017; Verma et al., 2019). Larger contigs and a higher number of modular PKS transcripts were retrieved from a highly toxic strain of *Gambierdiscus* sp. compared to a non-toxic one (van Dolah et al., 2020). In this case, the longest PKS transcript consisted of seven modules and could predict a part of the polyether backbone of the produced ciguatoxin. In another study, PKS transcripts were higher expressed in a toxic strain of *G. balechii* as compared to a non-toxin one (Wu et al., 2020). Other studies have suggested no differences in terms of total number of PKS transcripts when toxic and non-toxic species or strains of the same species were used, suggesting their constitutive presence in the genome of the organisms (Verma et al., 2019; Vingiani et al., 2020).

The functional annotation of the modular type I PKSs in *P. parvum* resulted in the expected PKS domains, but also revealed the presence of some additional ones (Fig. 3B). Such ‘unusual’ PKS domains have previously been found in toxigenic dinoflagellates such as *Gambierdiscus* sp. and *K. brevis* (Kohli et al., 2015, Van Dolah et al., 2017). The presence

Fig. 3. Heatmap of log2TPM gene expression of 8 consensus modular type I PKS contigs (A) and their domain organization (B). KS = ketoacyl synthase, KR = ketoreductase, ER = enoyl reductase, ACP = acyl carrier protein, DH = dehydratase, TE = thioesterase, AT = acyltransferase, HMG = hydroxymethylglutaryl-coenzyme A synthase, CRO = enoyl-CoA hydratase/isomerase, MT = methyltransferase. The *symbol indicates the contigs for which the cellular toxin content explained >70% of their gene expression variance.
of such ‘unusual’ PKS domains in distinct genera could be a result of either convergent evolution or horizontal gene transfer (Orr et al., 2013). Independently of the events that led to this result, secondary metabolism (which PKSs make part of) is subject to increased evolution which is driven by environmental pressures and selection based on the fitness the product attributes to an organism (Pichersky and Gang, 2000). A methyltransferase domain (MT) belonging to the methyltrasf_12 family was detected in three modular contigs. This domain belongs to the CL0063 clan and contains member proteins that are S-adenosylmethionine-dependent methyltransferases (SAM). S-adenosyl-L-methionine is acting as a donor of methyl groups that are added to a wide variety of substrates that include lipids and nucleic acids (Martin and McMillan, 2003). In PKSs, this domain catalyzes the addition of methyl branches (Piel, 2010). An enoyl-CoA hydratase/isomerase domain was also found and showed high similarity to the crotonase superfamily proteins and the CL0127 clan. This superfamily contains multiple enzymes with the ability to perform processes such as dehalogenation, hydration or isomerization. This domain was usually followed by a hydroxymethylglutaryl-coenzyme A synthase. This enzyme is member of the thiolase-like superfamily and the CL0046 clan. Members of this protein family are involved in both the degradation and biosynthesis of fatty acid synthases. Finally, a sulfotransferase domain was found in a modular contig and was followed by a thioesterase (TE) domain. This sulfotransferase domain belongs to the CL0023 clan and catalyzes the addition of a sulfo group to the substrate.

The identification of ‘unusual’ PKS domains highlights the complexity of haptophyte derived polyketides and the involvement of domains with variable effect on the final product. The study of these additional domains is limited and further investigations would contribute substantially to our understanding of chemical compounds biosynthesized by protists. This is especially important, as marine natural products can have a wide range of medical applications which would include antibiotics and anticancer medicines (Martínez Andrade et al., 2018; Lauritano et al., 2020).

Non-ribosomal peptide synthases (NRPS) catalyse the incorporation of amino acids to the final product. Hybrid PKS/NRPS products involve the addition of both acetate groups and amino acids. Three different PKS/NRPS hybrids were found in the current study, but they were not present in the transcriptomes of all strains. The phylogenetic analysis of the KS domains from PKS/NRPS hybrids supports their bacterial origin.
The expression pattern of the found modular PKS transcripts as transcripts per million (TPM) was further analyzed (Fig. 3A). The expression variance of two consensus contigs was found to be explained by the cellular toxin content, rendering them alongside their phylogeny as potential candidate genes in the biosynthesis of prymnesins. Moreover, the contigs clustered together in the case of B-type prymnesin producing strains, while there was a mixed clustering for the A- and C-types producing ones. Single-domain contigs from low toxin content strains showed high KS similarity and clustered with modular contigs, which indicates lower expression in low toxin content strains and the requirement of deeper sequencing in order to obtain full length transcripts.

3.4. Gene expression variance and potential ecological implications

The gradient of toxin content across the nine strains studied here was used as a parameter in order to search for molecular trade-offs involved in toxin production. The production of complex and large toxic compounds such as prymnesins may come at a considerable cost via the investment of carbon and energy resources, which, could instead be allocated to cell proliferation. We decomposed the total variance in the gene expression matrix into partial variances explained by the factors ‘toxin content’ and ’prymnesin type’. The number of transcripts with a KEGG annotation for which the toxin content explained the majority of the expression variance was 6335. The further exploration of these transcripts was done only for the ones that are involved in metabolic pathways. The general pattern revealed a downregulation of the cell’s metabolism as a trade-off for high cellular toxin contents (Fig. 4C). For all KEGG metabolic pathways, 1892 transcripts showed a negative correlation to toxin content, while 631 transcripts showed a positive correlation.

Trade-offs of cellular toxin content and growth rates have been studied in the paralytic shellfish toxin (PST) producing dinoflagellates of the genus Alexandrium (Blossom et al., 2019). In this study, strains with higher cellular PST contents had lower growth rates compared to strains with low cellular PST content. Our study indicated similar patterns in P. parvum: strains with high prymnesin content had lower growth rates compared to strains with low cellular prymnesin content (Fig. 4A; B). Although differences in growth rates could only explain 15% of the observed variability in toxin contents, the negative correlation was statistically significant (p = 0.043). The fact that growth rate explained 15% of the toxin variability highlights that multiple factors influence toxin production.

Prymnesin production seems to be constitutive in all P. parvum strains studied up to date, indicating the importance of toxin production in this species (Binzer et al., 2019). Within natural populations, the presence of strains of various toxicity attributes different advantages in individual and collective level (Tillmann and Hansen, 2009). Highly toxic strains play an important role to the success of the given species as they often have a broad impact on competitors and grazers (Donk and Ianora, 2011). The advantage of toxigenic strains with regard to grazing and competition has been shown for several species (John et al., 2002; Tillmann and Hansen, 2009) and a mutual facilitation of toxic strains for non-toxic ones has been postulated (John et al., 2015). However, if both toxic and non-toxic strains of a given species co-occur in populations, it may not always be to the benefit of the toxic strains, if the non-toxic strains are not affected by the released toxins. This has, in fact, been shown in experiments with a mixture of toxic and non-toxic P. parvum strains in the presence of a competitor; the non-toxic strains increased in relative abundance during the incubation (Driscoll et al., 2013). Initially, the highly toxic strains produced and released toxic compounds, which eliminated competitors and thus attributed a competitive advantage of the species. Within the context of the theory of the “public goods”, non-toxic or low toxic strains (also described as “cheaters”) are benefited by the elimination of competition as they have easier access to nutrients (Driscoll and Pepper, 2010; John et al., 2015). In the same study, the increase in frequency of low toxic strains highlights the fine balance and trade-offs between high and low toxicity. The growth advantage of low toxic strains would imply their later dominance in a community and the elimination of slowly growing strains. However, an important parameter to take into account is grazing, as low toxic strains will be preferred for consumption (Selander et al., 2006; Tillmann et al., 2007).

4. Conclusion

The study of the molecular basis of toxin biosynthesis is challenging due to the diversity of both chemical structures and involved genes. A plethora of polyketide synthase genes were found in P. parvum and this is the first step for future studies to elucidate the mechanisms involved in the study of prymnesins. Eight modular type I PKSs were present in the transcriptomes in all nine strains, while the expression of two contigs was explained by the cellular prymnesin content. The expression variance analysis suggests the downregulation of basic metabolic processes as a potential trade-off for high cellular toxin content. This work contributes to the developments of molecular tools for monitoring harmful blooms of the ichthyotoxic species P. parvum by providing a detailed analysis of a group of genes potentially involved in the production of prymnesins.

CRediT authorship contribution statement

Konstantinos Anestis: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization Gurjeet Singh Kohli: Conceptualization, Methodology, Writing – Review & Editing Sylke Wohlrab: Methodology, Writing – Review & Editing, Visualization Elisabeth Varga: Methodology, Investigation, Writing – Review & Editing, Funding acquisition Thomas Ostenfeld Larsen: Writing – Review & Editing, Per Juel Hansen: Writing – Review & Editing, Funding acquisition Uwe John: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgements

E.V. thanks Franz Berthiller of the University of Natural Resources and Life Sciences, Vienna (BOKU) for the access to the Agilent 6550 iFunnel QTOF LC/MS system.

Funding

This work has been funded by European Union’s Horizon 2020 research and innovation programme under grant agreement No. 766327. E.V. received funding of the Austrian Science Fund (FWF) through an Erwin-Schrödinger fellowship [J3895-N28].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.148878.

References


