BIOSYNTHESIS

Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins

Timothy R. Fallon[1]*†, Vikram V. Shende[1]†, Igor H. Wierzbicki[2], Amanda L. Pendleton[3,4], Nathan F. Watervoort[3,4], Robert P. Auber[3,4], David J. Gonzalez[2,5], Jennifer H. Wisecaver[3,4], Bradley S. Moore[1,5]*

Prymnesium parvum are harmful haptophyte algae that cause massive environmental fish kills. Their polyketide polyether toxins, the prymnesins, are among the largest nonpolymeric compounds in nature and have biosynthetic origins that have remained enigmatic for more than 40 years. In this work, we report the "PKZILLAs," massive *P. parvum* polyketide synthase (PKS) genes that have evaded previous detection. PKZILLA-1 and -2 encode giant protein products of 4.7 and 3.2 megadaltons that have 140 and 99 enzyme domains. Their predicted polyene product matches the proposed pre-prymnesin precursor of the 90-carbon–backbone A-type prymnesins. We further characterize the variant PKZILLA-B1, which is responsible for the shorter B-type analog prymnesin-B1, from *P. parvum* RCC3426 and thus establish a general model of haptophyte polyether biosynthetic logic. This work expands expectations of genetic and enzymatic size limits in biology.

Large-scale fish deaths caused by harmful algal blooms are global health, environmental, and food security problems (*1*). Anthropogenic causes appear to be hastening the severity and frequency of toxic eukaryotic microalgal blooms in freshwater and marine ecosystems, including the massive fish kill along the Oder River in 2022 by the golden alga *Prymnesium parvum* (Haptophyta) that decimated half of the river's fish population through Poland and Germany (*2*, *3*). Its hemolytic poisons, including prymnesin-1 and prymnesin-B1, are regularly implicated in worldwide mass fish die-offs (*4*, *5*) (Fig. 1) and are notable for their giant, fused polycyclic polyether structures. Prymnesin-1 (*6*), at 90 carbon atoms in length, joins other massive microalgal polyketide biotoxins such as palytoxin (*7*) and maitotoxin (*8*) that are among the largest nonpolymeric carbon chain molecules in nature (*9*). Collectively, this group of microalgal polyketides pose serious human and environmental health risks as neurotoxins. Fundamental knowledge of microalgal polyketide biology is still poorly understood because no causal biosynthesis gene has yet been identified in a haptophyte, chlorophyte, or dinoflagellate.

The chemical structure of giant microalgal polyethers such as prymnesin-1 imply a biosynthetic assembly-line construction of two-carbon chain length iterations to a polyene intermediate that undergoes epoxidation followed by a nucleophilic reaction cascade to construct their distinctive trans-fused ("ladder-frame") polyether frameworks (*10*, *11*). However, the biosynthesis of these massive microalgal toxins has remained an enigma despite a wealth of intimate knowledge of polyketide biochemistry from decades of research in bacteria and fungi (*12*). Despite numerous transcriptomic studies that identified orphan biosynthetic gene candidates in multimodular type I polyketide synthases (PKSs) from toxic microalgae (*13*, *14*), including numerous from diverse *P. parvum* strains (*15*), no PKS gene or fragment has yet been connected to a microalgal polyketide structure. The sheer size of microalgal polyether biotoxins presents experimental challenges. There is also a lack of methods to study their genetic origin. The model green alga *Chlamydomonas reinhardtii*, for instance, hosts a single large ~80–kilo–base pair (kbp) PKS known as PKS1 (Fig. 1D). Genetic knockout experiments established that PKS1 participates in formation of the zygospore cell wall, but its polyketide product remains unknown (*16*). Furthermore, unlike bacteria and fungi that organize their PKS encoding genes into polycistrons and biosynthetic gene clusters, other eukaryotes typically use monocistronic mRNAs and infrequently functionally colocalize most genes, which thus greatly obfuscates gene discovery efforts (*17*).

Here, we report the application of a customized gene annotation strategy that enabled the discovery of two massive PKS genes, PKZILLA-1 and -2, from *P. parvum* strain 12B1 that we propose are responsible for the complete backbone assembly of its notorious ladder-frame polyether toxin, prymnesin-1. Not only are the two giant PKZILLA "gigasynthases" organized consistently with the long-anticipated polyene intermediate structure of a microalgal lad… frame polyether, but PKZILLA-1 is larger t… the exceptionally large human protein titin (*18*) (Fig. 1D).

Genomic and transcriptomic evidence for the PKZILLA gigasynthases

We selected the A-type (*19*) prymnesin-1–producing *P. parvum* strain 12B1 as a model system to resolve microalgal polyether biosynthesis because its 116-Mbp genome and our recently published near–chromosome-level genomic assembly (*20*) make strain 12B1 relatively tractable among microalgae and other toxic *P. parvum* strains. We first cataloged PKS genes potentially involved in prymnesin-1 biosynthesis within our automated gene annotation (*20*) and identified 44 PKS genes that encoded relatively small proteins with one to three *trans*-acyltransferases (*trans*-AT) PKS modules. However, confirmatory tblastn queries that used PKS domains unveiled three seemingly contiguous and notably large PKS "hotspot" loci that stood apart at 137, 93, and 74 kbp on pseudochromosomes 17, 7, and 10, respectively. These hotspots showed a high concentration of visible coding regions that were only partially captured by 25 fragmented PKS gene models; thus, we hypothesized that they represented massive and misannotated single genes. Upon manual revision, we successfully constructed single-gene models from each hotspot, as described in the next paragraph, and dubbed the resulting genes PKZILLA-1, -2, and -3 (Figs. 1D and 2 and table S1). At final count, we annotated 22 PKS genes distributed across 16 of 34 pseudochromosomes that ranged in size from 3 to a notable 137 kbp (table S1).

Constructing the PKZILLA gene models from their candidate hotspots required several manual gene annotation interventions. The initial realignments of our Oxford Nanopore Technologies (ONT) long genomic DNA reads (*20*) to the PKZILLA hotspots revealed an assembly collapse of a tandem repetitive region within the coding N terminus of PKZILLA-1 that was fixed by targeted reassembly. After that revision, we found no further PKZILLA assembly concerns (fig. S1). To test for and classify transcriptional activity at these hotspots, we next analyzed coverage from a *P. parvum* 12B1 oligo-dT mRNA enrichment, i.e., poly-A tail pull-down Illumina RNA sequencing (RNA-seq) dataset (*21*) to localize the putative PKZILLA mRNA 3′ ends. This dataset indicated one transcriptional termination site per hotspot (fig. S2); however, as established for poly-A pull-down RNA-seq datasets, the coverage was negligible beyond ~10 kbp to the 5′ end and thus uninformative to 5′ transcriptional activity (fig. S2). To evaluate whether the full PKZILLA hotspots were transcriptionally active and consistent with single genes, we applied mRNA length-unbiased ribosomal RNA (rRNA)–depletion

[1]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography and University of California, San Diego, La Jolla, CA 92093, USA. [2]Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA. [3]Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA. [4]Purdue Center for Plant Biology, Purdue University, West Lafayette, IN 47907, USA. [5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.
*Corresponding author. Email: tfallon@ucsd.edu (T.R.F.); bsmoore@ucsd.edu (B.S.M.)
†These authors contributed equally to this work.

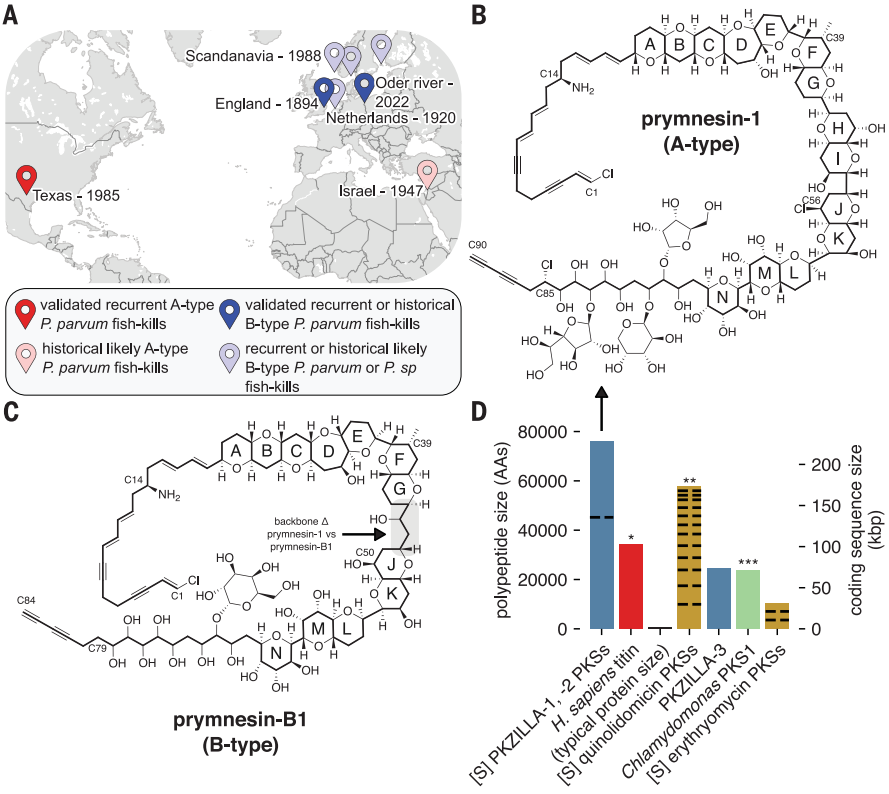

Fig. 1. Prymnesin, its source PKZILLA PKSs, and other large proteins and PKS systems. (**A**) Map of historical or ongoing major *P. parvum* or *Prymnesium* spp. fish-killing blooms: England (*49*, *50*), Netherlands (*51*), Israel (*6*, *52*, *53*), Texas (*5*, *54*, *55*), Scandinavia (*4*), and Oder River (*3*, *37*). (**B**) Molecular structure of prymnesin-1 (*31*). (**C**) Molecular structure of prymnesin-B1 (*56*). (**D**) Comparison of polypeptide and coding nucleotide sizes from representative PKSs or computationally summed PKS systems. Blue indicates PKZILLAs from *P. parvum* 12B1 (this work); [S] indicates computationally summed lengths for independent PKS proteins that participate in the same biosynthetic system; black dashed lines indicate divisions of PKS systems into independent proteins; red and * indicate the largest known protein (non-PKS) (*18*); gold indicates representative bacterial PKS systems, including the quinolidomicin (** indicates the previous largest known PKS system) (*42*) and erythromycin (*57*) PKSs; and green and *** indicate the previous largest genetically studied microalgal PKS (*16*). AA, amino acid.

Illumina RNA-seq by generating and sequencing two deoxyuridine triphosphate (dUTP)–stranded libraries from exponentially growing *P. parvum* cultures from the day and night phases. The low relative expression of the PKZILLAs combined with abundant rRNA reads from unoptimized rRNA depletion required four independent sequencing runs to accumulate sufficient coverage. Ultimately, we calculated that the PKZILLA transcript expression levels were uniformly low with transcripts per million (TPM) values of 1, 2, and 0.5 for PKZILLA-1, -2, and -3, respectively, in both day and night phases (table S1). These rRNA datasets further showed contiguous and sense-stranded transcriptional activity across the three PKZILLA gene models, which indicates one transcriptional start site per PKZILLA hotspot (Fig. 2 and fig. S2). Critically, the rRNA-depletion data identified the presence and location of the 34 PKZILLA introns, all of which showed canonical eukaryotic GT-AG splice sites (table S2 and figs. S3, S4, and S5). The gene model–derived PKZILLA polypeptide sequences show near-contiguous sequence similarity to known PKS domains, with limited evidence for internal breaks (figs. S6 and S7). Thus, we concluded that PKZILLA-1, -2, and -3 are single genes that each encode a single major transcriptional and translational product (Fig. 2). Notably, the calculated size of the PKZILLA-1 transcript at 136,071 nucleotides and the associated protein at 45,212 amino acids is about 25% larger than the mammalian muscle protein titin (*18*) (Fig. 1D and table S5).

Proteomic evidence for the PKZILLAs

To validate the PKZILLA proteins predicted by our gene models, we analyzed lyophilized *P. parvum* 12B1 biomass using an optimized bottom-up proteomics method. We identified and confidently validated 43 and 38 proteomic peptides from PKZILLA-1 and -2, respectively, yet none for PKZILLA-3 (fig. S8 and table S6). Only nine and six peptides from PKZILLA-1 and -2, respectively, were single copy (single-match) within a single predicted PKZILLA polypeptide ("protein-unique"). Instead, most of the detected peptides were multimatch peptides present in multiple copies, either protein-unique to a given PKZILLA or present in both PKZILLA-1 and -2 polypeptides (protein-multimatch) (figs. S9 and S10). This high proportion of multimatch peptides highlights the internally repetitive nature of the "gigamodular" PKZILLAs, both within and across proteins. These peptides were only present in the PKZILLA gene models and were not found anywhere else in six-frame translations of the 12B1 genome.

We next established which regions of the PKZILLA polypeptides were supported by proteomics. A complication is that most of the *P. parvum* proteomic data were multimatch peptides, which are rare in proteomic analyses of typical nonlarge, nonrepetitive proteins. Because they cannot be unambiguously assigned to a single polypeptide region, multimatch peptides are often ignored in downstream analyses in favor of simpler protein-unique single-match peptides (*22*). We judged that overlooking the multimatch peptides, although a simple solution, needlessly limited our analysis and discarded valuable data. We adapted to this challenge by subclassifying each protein-unique yet multimatch peptide that also only arose from the translation product of a single exon as "exon-unique" (fig. S10), which thus localized proteomic support to the exon rather than the residue level. Of the 43 PKZILLA-1 proteomic peptides, 14 met both the protein-unique and exon-unique criteria (fig. S10) and thus established unambiguous proteomic support for translation of 7 out of 17 PKZILLA-1 exons (41%), bounded upstream and downstream by exons 2 and 15, respectively (Fig. 2A). When considering the remaining 29 tryptic peptides despite their exon-multimatch or protein-multimatch ambiguity (fig. S10), we established increased proteomic support for 76% of the PKZILLA-1 exons (Fig. 2A). Applying the same criteria to PKZILLA-2, we measured proteomic support of 16 exon-unique peptides from 3 of the 12 exons (25%), bounded by exons 3 and 6, which increased to 75% of PKZILLA-2 exons after considering all PKZILLA matching peptides (Fig. 2B). Overall, these results confidently validate the translation of the PKZILLA-1 and -2 transcripts into proteins and are consistent with a single translational product per gene.

Annotation of the PKZILLA domain and modular structures and their compatibility with prymnesin

With proteomics-validated PKZILLA gene models in hand, we next tested their possible role in prymnesin-1 biosynthesis by annotating their PKS domains and evaluating their modular arrangement against the chemical structure of a retrobiosynthetically proposed pre-prymnesin

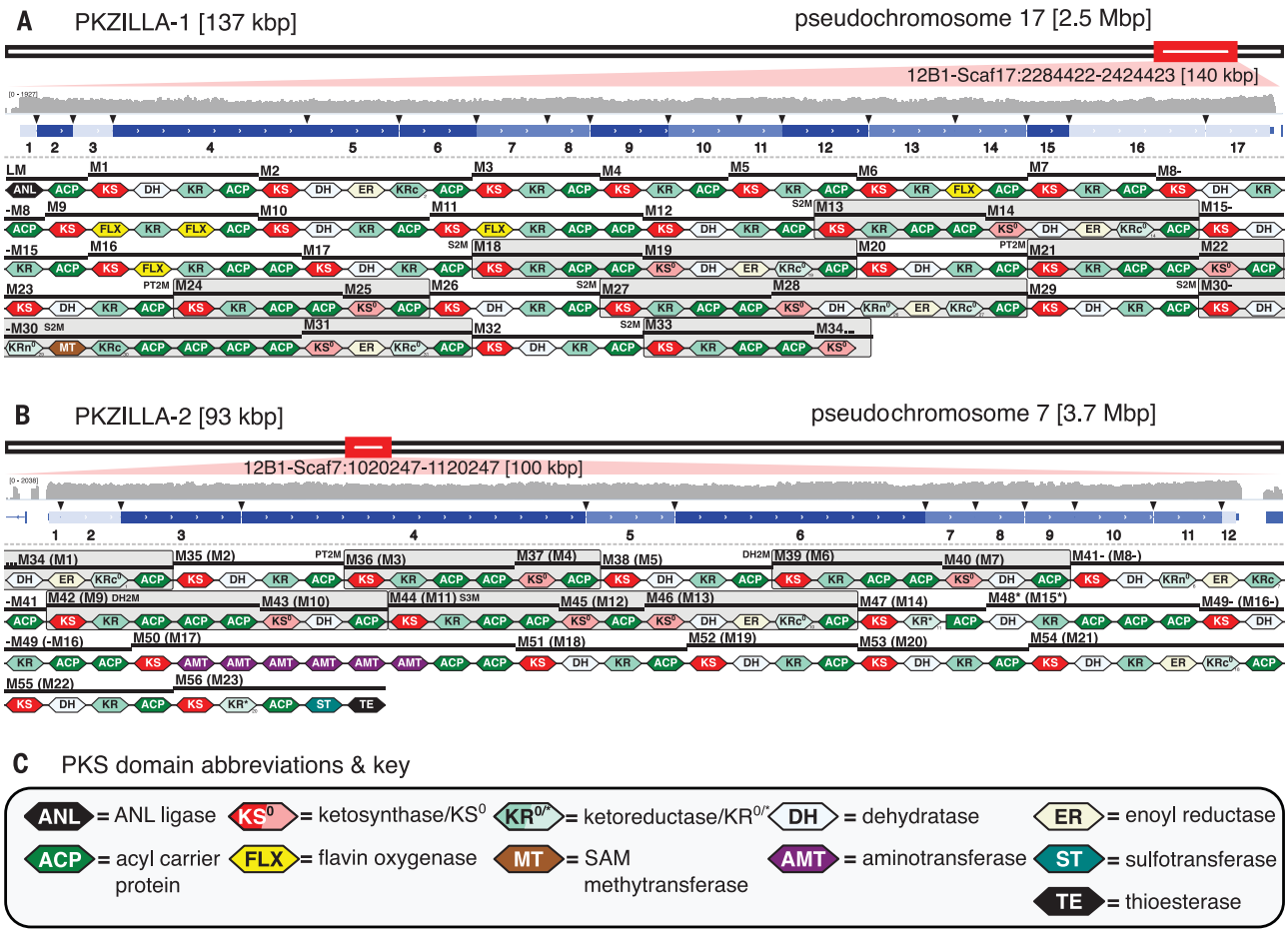

Fig. 2. Genomic, transcriptomic, and proteomic evidence for the PKZILLAs. (**A** to **C**) Genomic PKS hotspot loci with gene models and PKS domain and module annotations for (A) PKZILLA-1 and (B) PKZILLA-2. Red boxes denote chromosomal locations and relative sizes of the PKZILLA genes. The contiguous log-scale forward-stranded read coverage from the stranded rRNA-depletion RNA-seq (in gray) is shown across the PKZILLA gene models (in blue). Introns are highlighted with black arrows, and exons are numbered 1 to 17 for PKZILLA-1 and 1 to 12 for PKZILLA-2. See fig. S2 for an alternative view and figs. S3 and S4 for a detailed view of each intron. The numbered protein-coding exons are colored light blue, medium blue, or dark blue on the basis of whether supporting proteomic peptides from that exon were not detected, detected by protein-multimatch peptide matches alone, or detected by protein-unique plus exon-unique peptide matches, respectively (see section "Proteomic evidence for the PKZILLAs" in the text and fig. S10). Domain and module annotations (starting with the LM and M1 of PKZILLA-1 and ending with M56 of PKZILLA-2) are shown below the gene models; see key in (C). The bi- and trimodules are boxed in gray and categorized as S2/3M (saturating bi- and trimodule), PT2M (pass-through bimodule), and DH2M (dehydrating bimodule). See figs. S6 and S7 for non–length-normalized domains. KRn, N-terminal KR subdomain; KRc, C-terminal KR subdomain; KR*, catalytically novel or inactive KR.

biosynthetic precursor (PPBP; Fig. 3C). We identified 140 and 99 protein domains for PKZILLA-1 and -2, respectively (tables S3 to S5), using InterProScan (*23*). We also cataloged 30 candidate domains of unknown function (cDUFs); however, none of these cDUFs showed strong evidence of being unannotated enzyme domains (tables S7 and S8). The first two domains of PKZILLA-1, an acyl-CoA synthetase/NRPS adenylation domain/luciferase (ANL) superfamily ligase adjacent to an acyl carrier protein (ACP) domain (Fig. 2A), make up an unconventional, yet previously described (*24*), loading module (LM) to initiate polyketide chain elongation. PKZILLA-2 lacked any recognizable N-terminal loading domains; however, it does have a C-terminal thioesterase (TE) domain, consistent with polyketide chain termination. In the end, we organized the combined 239 domains into 56 *trans*-AT PKS modules, including module-34 (M34), split across the C terminus of PKZILLA-1 (Fig. 2A) and the N terminus of PKZILLA-2 (Fig. 2B and tables S4, S9, and S10).

Prymnesin-1 is devoid of standard polyketide termini and had an unknown direction of chain elongation. We resolved its biosynthetic directionality by first correlating the diagnostic polyol segment at C84 to C76 with the five adjacent modules M3 to M7 that contain a ketoreductase (KR) as the terminal reductive domain. On the basis of this observation, we could infer the directionality of biosynthesis, with the PKZILLA-1 LM initiating the polyketide biosynthetic pathway with a three-carbon carboxylate of yet unknown origin followed by chain extension with seven malonate molecules via M1 to M7 with interceding KR reduction to produce a hexol intermediate (fig. S21). Although five of the β-hydroxyls originate from malonate extender units, the out-of-sequence α-hydroxy group at C77 is instead likely installed by the α-hydroxylase flavoprotein (FLX) domain (*25*, *26*) contained in M6. PKZILLA-1 harbors two additional FLX domain–containing modules, M9 and M11, both of which align with the installation of the further α-hydroxylations at C71 and C67, respectively. The absence of domains that correspond to additional C–H oxygenations (C81 and C83) and halogenation (C85) suggests that these functionalities are installed by intermodular- or trans-acting enzymes during chain elongation or after polyketide assembly by means of oxidative enzymes (*27*).

The precyclized region that spans prymnesin-1's polyether rings (C74 to C20) coincides with a

break from canonical *trans*-AT modular architecture, with 11 nonelongating ketosynthase (KS^0)–containing modules interspersed among 16 canonical *trans*-AT PKS modules (Figs. 2 and 3A and tables S9 and S10). Two of the KS^0s are found as part of dehydrating bimodules (M39/M40, M42/M43; Fig. 2) first characterized in bacterial *trans*-AT pathways (*28*), wherein the first module [KS-KR(-ACP_n)] catalyzes chain elongation and ketone reduction and the second module, which consists of a KS^0-DH-ACP (where DH stands for dehydratase), performs the corresponding dehydration to yield an α,β-alkene thioester intermediate (fig. S21). Three other KS^0s are integrated into simple "pass-through bimodules" (M21/M22, M24/M25, M36/M37; Fig. 2), similar to those found in bacterial *trans*-AT biosynthetic gene clusters (*29*) whose minimal KS^0-ACP architecture preserves the oxidation state generated by the upstream module. The remaining KS^0s reside in previously unknown "saturating bimodules" and "saturating trimodules" (M13/M14, M18/M19, M27/M28, M30/M31, M33/M34, and M44/M45/M46; Fig. 2). In these systems, the second and third modules contain the full complement of reductive domains, which convert the β-hydroxy group to a saturated methylene at positions C64, C56, C44, C40, C36, and C22 (Fig. 3C and fig. S21). Notably, the saturating bimodule M30/M31 also contains the only methyltransferase (MT) domain in the entire assembly line and is positioned to install prymnesin's lone methyl group at C39 (Fig. 3 and fig. S21). We further confirmed the module-to-precursor alignment throughout this C74 to C20 region

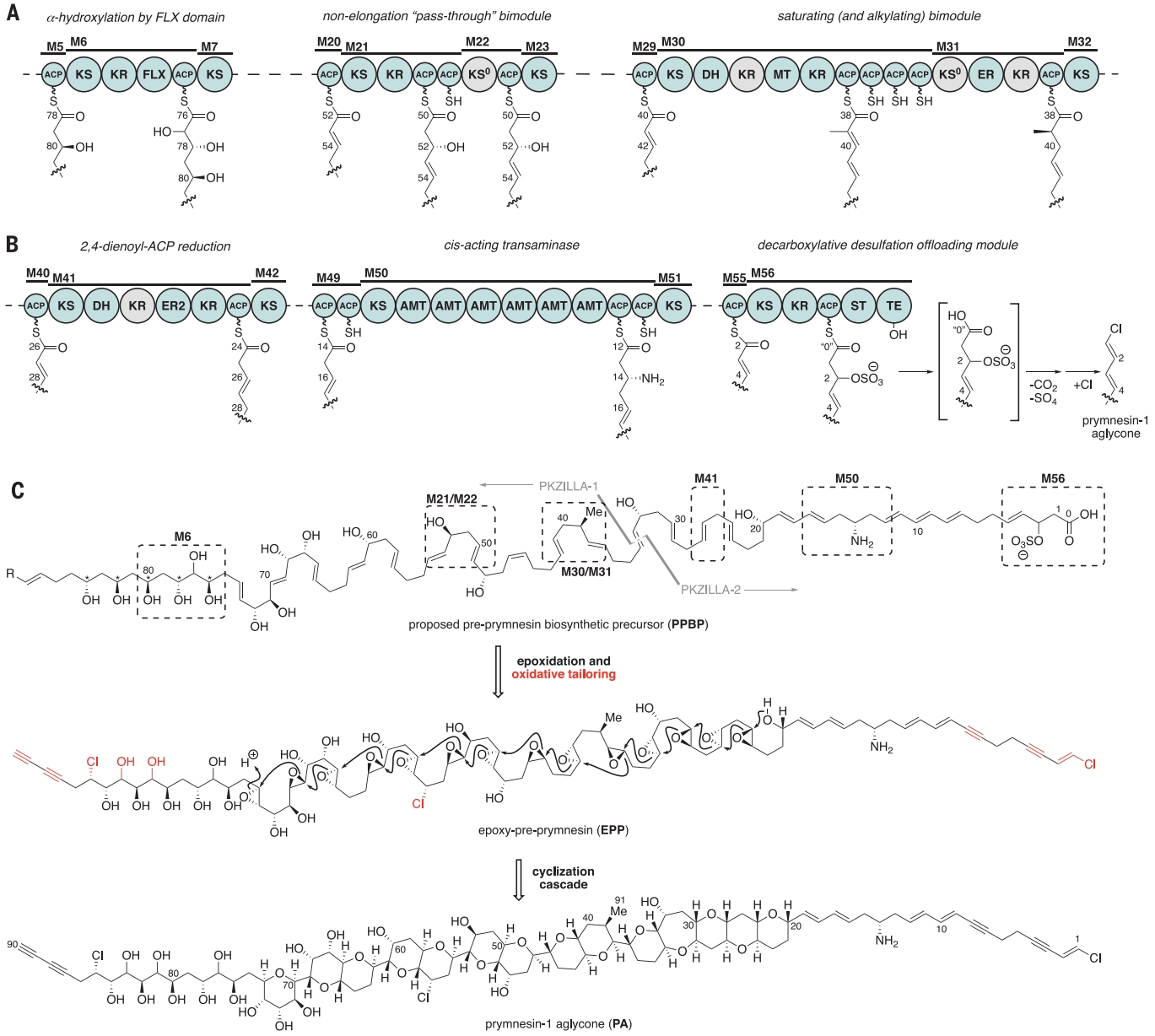

Fig. 3. Alignment of PKZILLA PKS modules with the proposed prymnesin biosynthetic precursor. (**A**) PKZILLA-1 snapshot assembly line reactions at (bi)modules M6, M21/M22, and M30/M31 that depict FLX-catalyzed α-hydroxylation, reducing nonelongation pass-through, and dual methylation-saturation reactions. Dotted lines indicate omitted contiguous PKZILLA protein sequence. (**B**) PKZILLA-2 snapshot assembly line reactions at modules M41, M50, and M56 that depict 2,4-dienoyl-ACP reduction, transamination, and decarboxylative desulfation reactions. (**C**) Proposed structures of the PKZILLA-compatible pre–prymnesin-1 biosynthetic precursor, its corresponding epoxy-pre–prymnesin-1 intermediate, and the prymnesin-1 aglycone (*31*). The unassigned starter R group in PPBP is presently unknown yet responsible for providing the terminal acetylene group. For simplicity, all oxidative modifications are depicted after assembly line. The snapshot reactions in (A) and (B) are correspondingly boxed in (C). See fig. S21 for further detail. Me, methyl.

by applying *trans*-AT KR heuristics (*30*) to bioinformatically predict the stereochemical outcome of reduction for each KR domain (table S12). These bioinformatic predictions matched with six of the seven previously assigned configurations from the most recent structure revision of prymnesin-1 (*31*), with the C32 hydroxyl as the exception (table S13 and fig. S21). By extrapolating these predictions, we propose β-hydroxy stereochemical assignments for the yet unassigned C84 to C76 region of prymnesin-1 (Fig. 3C and table S13).

In addition, the polyene segment (C19 to C1) contains several distinguishing structural features, all of which align with the final 10 modules of PKZILLA-2. The M49 dehydratase is positioned to catalyze a previously described (*32*) vinylogous dehydration to reconfigure the C19 to C16 diene out of conjugation relative to the ACP-tethered thioester, and the unconventional configuration of six consecutively arranged pyridoxal 5′-phosphate (PLP)–dependent aminotransferase (AMT) domains (*33*) in M50 is located at the precise position to incorporate the sole primary amine at C14 (Fig. 3B). Much like the initial modules in PKZILLA-1, the final modules in PKZILLA-2, M51 to M55, have

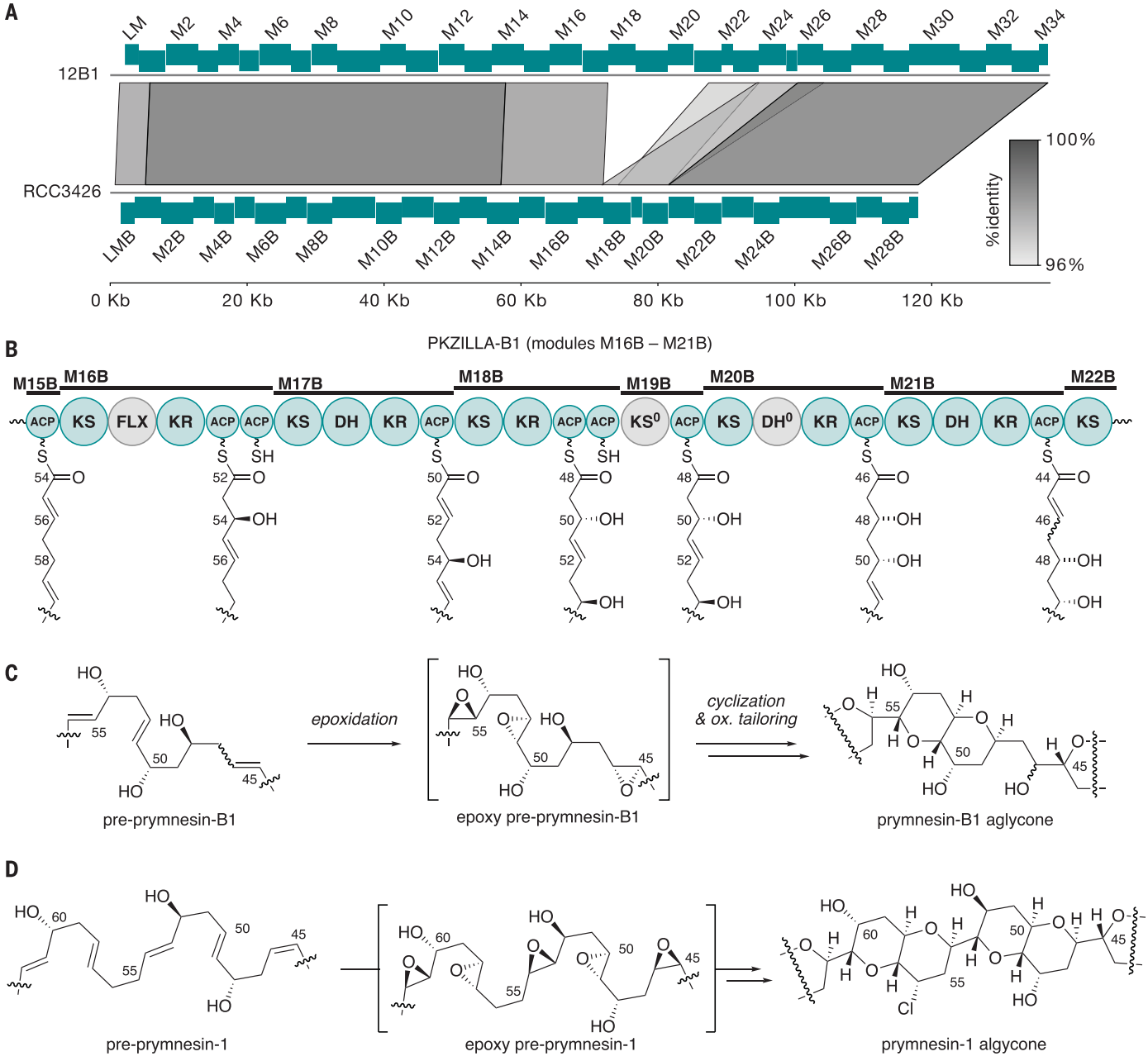

Fig. 4. Comparative genomics, biosynthesis, and chemistry of the A- and B-type prymnesins. (**A**) Synteny plot of PKZILLA-1 from *P. parvum* 12B1 (top) and PKZILLA-B1 from *P. parvum* RCC3426 (bottom) showing a 15-kbp deletion and other modifications relative to the 12B1 sequence. (**B**) Domain organization of PKZILLA-B1 modules M16B to M21B and its assembly line biosynthesis of the differentiating C45-C55 fragment. Inactive domains are colored gray. (**C**) Partial structure and proposed conversion of pre–prymnesin-B1 via epoxy-pre–prymnesin-B1 to prymnesin-B1. Because the C46 stereocenter in prymnesin-B1 has not yet been established (*56*), the configuration of the C45-C46 alkene is not drawn. Full structures are shown in fig. S23. (**D**) Corresponding partial structures and biosynthesis of prymnesin-1 and intermediates highlighting the structure region that is distinct from the B1 series shown in (C). The omitted structure regions are identical between the A- and B-type prymnesins, as supported by their respective PKZILLA gigasynthases. See Fig. 1 for full chemical structures.

standard *trans*-AT domain architecture and generate the C12 to C7 triene and a transient C4 to C3 alkene that must undergo further desaturation to yield prymnesin's observed alkyne. The final module, M56, is precedented by the terminal PKS module from curacin biosynthesis (*34*), wherein an unusual sulfotransferase domain sulfates the β-hydroxy and TE-mediated offloading initiates simultaneous decarboxylation and sulfate elimination to produce a terminal alkene (Fig. 3B). Because prymnesin terminates in a vinyl chloride, an additional halogenase must act before or after chain offloading to install the third and final chloride at C1 (*35*).

PKZILLA gene models establish broad prymnesin polyketide biosynthetic logic

We next predicted that A-type– and B-type–producing strains of *P. parvum* would differ at the PKZILLA-1 locus on the basis of structural differences between their prymnesins (*36*) (Fig. 1, B and C). Illumina genomic DNA reads from nine strains of diverse A- and B-type *P. parvum* strains (*20*) were aligned to the 12B1 genome, and the read coverage at the three PKZILLA loci was compared. Precise structural differences could not be determined from read mapping alone because of the abundance of multimapped reads across the highly repetitive PKZILLA loci (fig. S22). In general, regions of uniquely mapped reads in all three PKZILLA genes were conserved across A- and B-type strains; however, the four B-type strains showed reduced uniquely mapped read coverage at modules 18 and 21 of PKZILLA-1, which indicates a deletion relative to the 12B1 sequence (fig. S22A). To better characterize this structural difference, we used ONT long reads to create a de novo assembly of B-type strain RCC3426 and characterized its homologous PKZILLA-1 locus. Synteny analysis identified a single 15-kbp region in 12B1 PKZILLA-1 that is absent in RCC3426. The absent region was composed of a small portion of M17, all of M18 to M20, and more than half of M21 for PKZILLA-1, among other changes (Fig. 4A). Recently, a similar observation was reported in the B-type *P. parvum* ODER1 strain (*37*).

These changes in the B-type PKZILLA-1 gigasynthase "PKZILLA-B1" resulted in a smaller 29-modular synthase that corresponded to the six-carbon-shorter prymnesin-B1 molecule, in which three elongating and two nonelongating modules were absent in relation to PKZILLA-1 (Fig. 4A). Domain predictions based on the A-type PKZILLA model mapped seamlessly with the pre–prymnesin-B1 polyene intermediate (Fig. 4B and fig. S23). Two key differences included the replacement of the M18/M19 saturating bimodule with a reducing M18B/M19B bimodule and the tetradomain reducing module M20B with a nonfunctional DH domain (table S15). These changes in module organization support the C50 and C48 hydroxyls in the pre–prymnesin-B1 intermediate (Fig. 4C) that, in addition to the loss of three extension modules, distinguish prymnesin-B1 from prymnesin-1 (Fig. 4D).

Although much of the prymnesin assembly line conforms to *trans*-AT PKS biochemistry, there are a few module-to-precursor alignments that may signal distinctive enzyme-catalyzed steps that are consistent across the A- and B-type PKZILLAs. In the case of M40 and its dienoyl intermediate, we propose that the adjacent M41 module elongates the growing polyketide chain to a 2,4-dienoyl-ACP intermediate before reduction (Fig. 3B and fig. S21) by a phylogenetically distinct enoyl reductase "ER2" (table S10 and figs. S15 and S20) to generate a β,γ-alkene out of conjugation with the thioester carbonyl. This type of reduction has precedence in fatty acid biosynthesis on acyl–coenzyme A (acyl-CoA) intermediates (*38*). Similarly, module M48 is missing an explicit KS domain (Fig. 2, fig. S7, and table S10) and instead harbors the 186-residue sequence-unique cDUF9 in the analogous upstream position (tables S8 and S10) that may help recruit a restorative cis- or trans-acting KS enzyme (fig. S21). In addition, on the basis of the presently revised structure of prymnesin-1 (*31*), our biosynthetic model requires a (*Z*)-ene at C45-C46 to accommodate the stereochemical outcome of the prymnesin aglycone, although analysis at M26 predicted an (*E*)-ene product (table S12). Taken together, the sequence of the assembly-line operations in the A- and B-type prymnesin systems strongly supports a causal role for PKZILLAs as the gigasynthases responsible for synthesis of the prymnesin backbone and suggests future focus areas to identify the remaining biosynthetic enzymes for the polyether cascade.

Discussion

The structural elucidation of brevetoxin B as the causative agent of the toxic red tide dinoflagellate *Karenia brevis* more than four decades ago (*39*) established microalgae as exquisite producers of polycyclic polyether toxins. More than 150 members with five- to nine-membered cyclic ethers, including maitotoxin as the largest with 32 fused rings, have since been discovered (*40*) and have helped shape the field of marine natural products given their toxicity to both marine life and humans. One such marine polyether, halichondrin B, was even the inspiration for the synthetic derivative eribulin, which has since been approved as an anticancer drug (*41*). The discovery and initial characterization of the prymnesin PKZILLA gigasynthases now elucidates the long-standing question about how microalgae biosynthesize their giant polyketide polyether molecules.

The domain composition of the PKZILLA-1 and -2 gigasynthases are impeccably aligned with the cooperative assembly of the prymnesin carbon scaffold and support decades-old hypotheses that diverse microalgal ladder-frame polyethers from haptophytes and dinoflagellates are constructed from all (*E*)-linear polyene intermediates (*10*, *11*). Our discovery identified several unexpected features of the prymnesin assembly line. The sizes of the PKZILLA enzymes are stunning and surprisingly require just two modular *trans*-AT PKS proteins at 239 total domains for the construction of the 90-carbon-long prymnesin molecule. By contrast, the longest known bacterial polyketide, quinolidomicin at 68 carbons (*42*), is assembled by 13 PKS gene products (Fig. 1D). The notable size of the PKZILLAs expands our imagination on the capabilities of enzymes in the construction of complex molecules. In addition, many nonelongating KS^0s are featured in modules associated with the construction of the polycyclic interior of prymnesin, which may contribute to the timing and mechanism of polyether assembly.

No comparable PKS system has yet been identified from a toxic dinoflagellate (*14*); however, numerous studies have established that dinoflagellates encode large numbers of multimodular and single-domain type I PKSs (*43*), including a promising, yet reportedly incomplete, 35-kilonucleotide, seven-module PKS transcript candidate from the ciguatoxin-producing *Gambierdiscus polynesiensis* (*13*). If dinoflagellates similarly encode giant PKSs reminiscent of the PKZILLAs, then common transcriptomic practices involving poly-A pull-down RNA-seq may bias against giant transcripts and instead require length-unbiased rRNA-depletion RNA-seq alongside customized assembly and annotation, as performed in this study. Notably, the PKZILLAs went unrecognized from recent *P. parvum* transcriptomic (*15*) and genomic (*44*) analyses, which highlights the challenges of assembling and annotating giant PKS genes with their highly repetitive sequences. Retrospective analysis of the recent transcriptomic study (*15*), with hindsight from knowledge of the full PKZILLA gene models, revealed that the 3′ ends of the PKZILLA-1 and -2 transcripts were indeed captured, covering modules M28 to M34 (contig 7) and M54 to M56 (contig 6), respectively. These observations are similar to ours when viewed only by poly-A RNA-seq (fig. S2) and highlight the limitations of this common RNA-seq method when conducted on giant transcripts.

Dinoflagellate polyketides also share a distinctive biosynthetic feature that involves the irregular incorporation of intact and C1-deleted acetate building blocks as illuminated by isotope labeling studies (*45*). The prymnesin biosynthetic model, however, supports the intact incorporation of 43 contiguous malonate units, which is standard in most bacterial systems. The α-hydroxylating FLX domains and pass-through modules found within the PKZILLA

modules provide a tantalizing hypothesis for this yet-to-be-described dinoflagellate PKS biochemistry: Assembly-line oxidation to the α-ketone followed by transacylation by a KS^0 may lead to the excision of single-carbon atoms by decarbonylation, as seen previously in the biosynthesis of marine polyketides enterocin (*46*) and barbamide (*47*).

Although PKZILLA-1 and -2 are responsible for the construction of most of the prymnesin molecule, additional enzymes (acyltransferases, desaturases, hydroxylases, chlorinases, epoxidases, and glycosyltransferases) are needed to complete the full biosynthetic pathway and install prymnesin's remaining functional groups and sugar moieties. In contrast to the small alkaloid domoic acid with its clustered causal genes (*48*), the distribution of the PKZILLA-1 and -2 genes across separate pseudochromosomes indicates that prymnesin biosynthesis is not encoded within a single biosynthetic gene cluster and that its tailoring enzymes may also not be clustered. The discovery of the PKZILLAs and their role in prymnesin biosynthesis lays the foundation for the development and implementation of alternative linked omics approaches to fully uncover the complete suite of prymnesin biosynthetic enzymes. Moreover, PKZILLAs offer the opportunity to dissect the enzymology of ladder-frame polyether biosynthesis and will serve as a model to capture and dissect giant genes, transcripts, and proteins in specialized metabolism.

REFERENCES AND NOTES

1. G. M. Hallegraeff, D. M. Anderson, K. Davidson, F. Gianella, P. Hansen, *Fish-Killing Marine Algal Blooms: Causative Organisms, Ichthyotoxic Mechanisms, Impacts and Mitigation* (IOC Manuals and Guides 93, UNESCO, 2023).
2. A. W. Griffith, C. J. Gobler, *Harmful Algae* **91**, 101590 (2020).
3. J. Sobieraj, D. Metelski, *Toxins* **15**, 403 (2023).
4. B. Karlson *et al*., *Harmful Algae* **102**, 101989 (2021).
5. R. Patiño, V. G. Christensen, J. L. Graham, J. S. Rogosch, B. H. Rosen, *Water* **15**, 2808 (2023).
6. T. Igarashi, M. Satake, T. Yasumoto, *J. Am. Chem. Soc.* **121**, 8499–8511 (1999).
7. R. E. Moore, G. Bartolini, *J. Am. Chem. Soc.* **103**, 2491–2494 (1981).
8. M. Murata *et al*., *J. Am. Chem. Soc.* **115**, 2060–2062 (1993).
9. K. C. Nicolaou, M. O. Frederick, R. J. Aversa, *Angew. Chem. Int. Ed.* **47**, 7182–7225 (2008).
10. K. Nakanishi, *Toxicon* **23**, 473–479 (1985).
11. I. Vilotijevic, T. F. Jamison, *Science* **317**, 1189–1192 (2007).
12. A. Nivina, K. P. Yuet, J. Hsu, C. Khosla, *Chem. Rev.* **119**, 12524–12547 (2019).
13. F. M. Van Dolah *et al*., *PLOS ONE* **15**, e0231400 (2020).
14. A. Verma *et al*., *Microorganisms* **7**, 222 (2019).
15. K. Anestis *et al*., *Sci. Total Environ.* **795**, 148878 (2021).
16. N. Heimerl *et al*., *Plant J.* **95**, 268–281 (2018).
17. M. H. Medema, T. de Rond, B. S. Moore, *Nat. Rev. Genet.* **22**, 553–571 (2021).
18. M.-L. Bang *et al*., *Circ. Res.* **89**, 1065–1072 (2001).
19. S. B. Binzer *et al*., *Harmful Algae* **81**, 10–17 (2019).
20. J. H. Wisecaver *et al*., *Curr. Biol.* **33**, 2246–2259.e8 (2023).
21. National Center for Biotechnology Information (NCBI), Sequence Read Archive (SRA), *Prymnesium parvum* strain 12B Illumina reads - SRA PRJNA201451, SRR1685644; https://www.ncbi.nlm.nih.gov/sra/?term=SRR1685644.
22. Y. Dou *et al*., *Nat. Commun.* **14**, 5809 (2023).
23. P. Jones *et al*., *Bioinformatics* **30**, 1236–1240 (2014).
24. F. Hemmerling, K. E. Lebe, J. Wunderlich, F. Hahn, *ChemBioChem* **19**, 1006–1011 (2018).
25. F. Hemmerling *et al*., *Angew. Chem. Int. Ed.* **61**, e202116614 (2022).
26. A. J. Winter *et al*., *Angew. Chem. Int. Ed.* **62**, e202312514 (2023).
27. E. J. N. Helfrich, J. Piel, *Nat. Prod. Rep.* **33**, 231–316 (2016).
28. D. T. Wagner *et al*., *Structure* **25**, 1045–1055.e2 (2017).
29. J. Masschelein *et al*., *Nat. Chem.* **11**, 906–912 (2019).
30. E. J. N. Helfrich *et al*., *Nat. Chem. Biol.* **15**, 813–821 (2019).
31. M. Sasaki *et al*., *Tetrahedron Lett.* **47**, 5687–5691 (2006).
32. F. Taft, M. Brünjes, T. Knobloch, H. G. Floss, A. Kirschning, *J. Am. Chem. Soc.* **131**, 3812–3813 (2009).
33. Z. D. Aron, P. C. Dorrestein, J. R. Blackhall, N. L. Kelleher, C. T. Walsh, *J. Am. Chem. Soc.* **127**, 14986–14987 (2005).
34. L. Gu *et al*., *J. Am. Chem. Soc.* **131**, 16033–16035 (2009).
35. Y. Jiang, A. Kim, C. Olive, J. C. Lewis, *Angew. Chem. Int. Ed.* **63**, e202317860 (2024).
36. S. A. Rasmussen *et al*., *J. Nat. Prod.* **79**, 662–673 (2016).
37. H. Kuhl *et al*., *Curr. Biol.* 10.1016/j.cub.2024.06.033 (2024).
38. T. Hua *et al*., *J. Biol. Chem.* **287**, 28956–28965 (2012).
39. Y.-Y. Lin *et al*., *J. Am. Chem. Soc.* **103**, 6773–6775 (1981).
40. Z.-P. Jiang *et al*., *Chem. Sci.* **12**, 10197–10206 (2021).
41. J. Cortes, P. Schöffski, B. A. Littlefield, *Cancer Treat. Rev.* **70**, 190–198 (2018).
42. T. Hashimoto *et al*., *Org. Lett.* **20**, 7996–7999 (2018).
43. F. M. Van Dolah, G. S. Kohli, J. S. Morey, S. A. Murray, *J. Phycol.* **53**, 1325–1339 (2017).
44. J. Jian *et al*., *Sci. Total Environ.* **908**, 168042 (2024).
45. R. M. Van Wagoner, M. Satake, J. L. C. Wright, *Nat. Prod. Rep.* **31**, 1101–1137 (2014).
46. R. Teufel *et al*., *Nature* **503**, 552–556 (2013).
47. S. Guo *et al*., *J. Am. Chem. Soc.* **145**, 5017–5028 (2023).
48. J. K. Brunson *et al*., *Science* **361**, 1356–1358 (2018).
49. P. A. Holdway, R. A. Watson, B. Moss, *Freshw. Biol.* **8**, 295–311 (1978).
50. E. S. Hems, thesis, University of East Anglia, Norwich (2017).
51. F. Liebert, W. M. Deerns, Nederland Nationaal Archief - Inventaris van het archief van de Hoofinspecteur der Visserijen, (1907) 1910-1922 (1924) - Item 195: Verslag van een onderzoek naar de oorzaken van de vissterfte in de polder Workumer-Nieuwland nabij Workum, door N.W. Deerns van het Rijksinstituut voor Hydrografisch Visserijonderzoek, houdende een beschrijving van visdodende flagellaten (planktonorganismen). 1920 (1920); https://www.nationaalarchief.nl/onderzoeken/archief/2.11.12.
52. K. Reich, M. Aschner, *Palest. J. Bot.* **4**, 14–23 (1947).
53. M. Shilo, R. F. Rosenberger, *Ann. N. Y. Acad. Sci.* **90**, 866–876 (1960).
54. T. L. James, A. De La Cruz, *Tex. J. Sci.* **41**, 429–430 (1989).
55. D. L. Roelke *et al*., *J. Plankton Res.* **33**, 243–253 (2011).
56. S. A. Rasmussen *et al*., *J. Nat. Prod.* **79**, 2250–2256 (2016).
57. S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson, L. Katz, *Science* **252**, 675–679 (1991).
58. Y. Perez-Riverol *et al*., *Nucleic Acids Res.* **50**, D543–D552 (2022).
59. T. Fallon, Label-free bottom-up proteomics of Prymnesium pavum 12B1. PRIDE (2024); https://dx.doi.org/10.6019/PXD044632.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions and feedback from members of the Moore laboratory and from A. Lukowski [University of California, San Diego (UCSD)]. This publication includes data generated at the USCD IGM Genomics Center by using an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant (no. S10 OD026929). Large language models (LLMs)—including code generation by GitHub Copilot (Microsoft) within Visual Studio Code (Microsoft) and code generation and limited text rephrasing by the Claude 3 Opus application programming interface (Anthropic PBC)—were used in this work. The authors reviewed all LLM-generated text and rejected any inaccuracies. **Funding:** This work was funded by National Institutes of Health grants F32-ES032276 (to T.R.F.), F32-GM145146 (to V.V.S.), and R01-GM085770 (to B.S.M.) and National Science Foundation grant DEB-1831493 (to J.H.W.). **Author contributions:** Conceptualization: T.R.F., B.S.M.; Data curation: T.R.F., V.V.S., I.H.W., A.L.P., R.P.A.; Formal analysis: T.R.F., V.V.S., I.H.W.; Funding acquisition T.R.F., V.V.S., D.J.G., J.H.W., B.S.M.; Investigation: T.R.F., V.V.S., I.H.W., A.L.P., N.F.W., R.P.A.; Methodology: T.R.F., V.V.S., I.H.W., A.L.P., N.F.W., R.P.A.; Project administration: T.R.F., D.J.G., J.H.W., B.S.M.; Resources: T.R.F., B.S.M.; Software: T.R.F.; Supervision: T.R.F., B.S.M.; Validation: T.R.F., V.V.S., B.S.M.; Visualization: T.R.F., V.V.S., B.S.M.; Writing – original draft: T.R.F., V.V.S., B.S.M.; Writing – review and editing: T.R.F., V.V.S., I.H.W., A.L.P., N.F.W., R.P.A., D.J.G., J.H.W., B.S.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The *P. parvum* 12B1 v1.1 genome assembly, PKZILLA gene models, and updated gene annotation have been deposited to the National Center for Biotechnology Information (NCBI) Whole Genome Shotgun (WGS) database (accession number JBGBPQ010000000). Raw rRNA-depletion RNA-seq data have been deposited to the NCBI Sequence Read Archive (SRA) (BioProject PRJNA936443). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE (*58*) partner repository under the dataset identifier PXD044632 (*59*). Other extended datasets and analysis code (scripts) are available on Zenodo and/or Github and are both cited in-line throughout the manuscript and listed in table S16.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ado3290
Materials and Methods
Figs. S1 to S23
Tables S1 to S16
References (*60*–*137*)
MDAR Reproducibility Checklist

Submitted 30 January 2024; accepted 10 July 2024
10.1126/science.ado3290