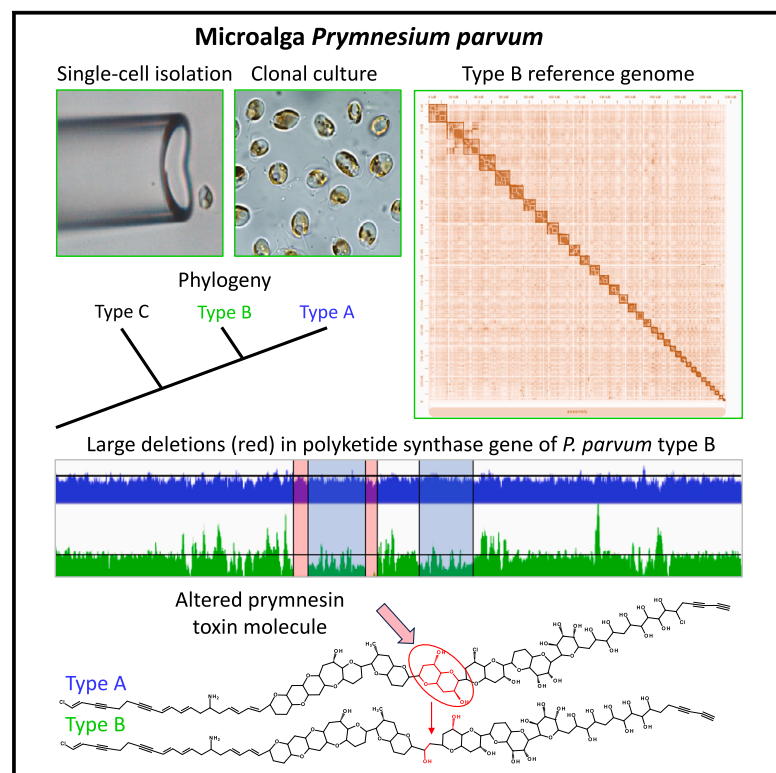


Current Biology

The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins

Graphical abstract



Authors

Heiner Kuhl, Jürgen F.H. Strassert, Dora Čertnerová, ..., Jan Köhler, Michael T. Monaghan, Matthias Stöck

Correspondence

michael.monaghan@igb-berlin.de (M.T.M.),
matthias.stoeck@igb-berlin.de (M.S.)

In brief

Kuhl and Strassert et al. present a reference-quality genome of the microalga *Prymnesium parvum* s.l. and reveal a large, recently evolved deletion in a PKS gene that is linked to the production of prymnesin, a toxin responsible for catastrophic fish kills and loss of aquatic life worldwide.

Highlights

- A reference-quality genome of *Prymnesium parvum* type B was sequenced and assembled
- The relatively large type B genome shows transposon expansion in all 34 chromosomes
- A 20-kbp deletion in a PKS gene is linked to the unique structure of the type B toxin



Article

The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins

Heiner Kuhl,^{1,7,8,10} Jürgen F.H. Strassert,^{1,7,9,10} Dora Čertnerová,^{2,3} Elisabeth Varga,^{4,5} Eva Kreuz,^{1,7} Dunja K. Lamatsch,³ Sven Wuertz,^{1,7} Jan Köhler,^{1,7} Michael T. Monaghan,^{1,6,7,*} and Matthias Stöck^{1,7,11,*}

¹Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany

²Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

³Research Department for Limnology, Mondsee, University of Innsbruck, Mondsee, Austria

⁴Department of Food Chemistry and Toxicology, University of Vienna, Vienna, Austria

⁵Unit Food Hygiene and Technology, Centre for Food Science and Veterinary Public Health, Clinical Department for Farm Animals and Food System Science, University of Veterinary Medicine, Vienna, Austria

⁶Institut für Biologie, Freie Universität Berlin, Berlin, Germany

⁷X (formerly Twitter): @LeibnizIGB

⁸X (formerly Twitter): @KuhlGenomics

⁹X (formerly Twitter): @jfh_strassert

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: michael.monaghan@igb-berlin.de (M.T.M.), matthias.stoeck@igb-berlin.de (M.S.)

<https://doi.org/10.1016/j.cub.2024.06.033>

SUMMARY

The catastrophic loss of aquatic life in the Central European Oder River in 2022, caused by a toxic bloom of the haptophyte microalga *Prymnesium parvum* (in a wide sense, s.l.), underscores the need to improve our understanding of the genomic basis of the toxin. Previous morphological, phylogenetic, and genomic studies have revealed cryptic diversity within *P. parvum* s.l. and uncovered three clade-specific (types A, B, and C) prymnesin toxins. Here, we used state-of-the-art long-read sequencing and assembled the first haplotype-resolved diploid genome of a *P. parvum* type B from the strain responsible for the Oder disaster. Comparative analyses with type A genomes uncovered a genome-size expansion driven by repetitive elements in type B. We also found conserved synteny but divergent evolution in several polyketide synthase (PKS) genes, which are known to underlie toxin production in combination with environmental cues. We identified an approximately 20-kbp deletion in the largest PKS gene of type B that we link to differences in the chemical structure of types A and B prymnesins. Flow cytometry and electron microscopy analyses confirmed diploidy in the Oder River strain and revealed differences to closely related strains in both ploidy and morphology. Our results provide unprecedented resolution of strain diversity in *P. parvum* s.l. and a better understanding of the genomic basis of toxin variability in haptophytes. The reference-quality genome will enable us to better understand changes in microbial diversity in the face of increasing environmental pressures and provides a basis for strain-level monitoring of invasive *Prymnesium* in the future.

INTRODUCTION

In the summer of 2022, an anthropogenic environmental disaster struck the Central European Oder River, resulting in a significant loss of aquatic life due to the proliferation of a strain of the toxin-producing microalga, *Prymnesium parvum* s.l. (“sensu lato”—in a wide sense) (Haptophyta, Prymnesiophyceae, Prymnesiales, Prymnesiaceae; sometimes incorrectly classified as a “golden alga”).¹ The toxins released by this brackish-water mixotroph, which measures only 5–10 μm and carries two flagella for active movement and a specialized organelle (haptoneura) for attaching to prey, led to the death of a thousand metric tons of fish, mussels, and

snails along the entire Oder River in Poland and Germany.^{1–3} The invasive *Prymnesium* has been identified as the cause of massive fish kills since the mid-20th century,^{4,5} but only in the last three decades has much of its cryptic diversity been recognized. Electron microscopy revealed variation in the organic scales among strains or between stages of what may be a haplo-diplontic life cycle.^{6–8} Phylogenetic analyses revealed two (internal transcribed spacer 1 [ITS1])⁷ or three (ITS1 or ITS1 + 2)^{1,9–11} different clades (types A, B, and C)¹⁰ by these evolutionarily relatively conserved DNA markers. Toxicological analyses later revealed clade-specific allelopathic toxins (prymnesins) that correspond to these three clades.¹⁰ These compounds consist of a ladder-frame



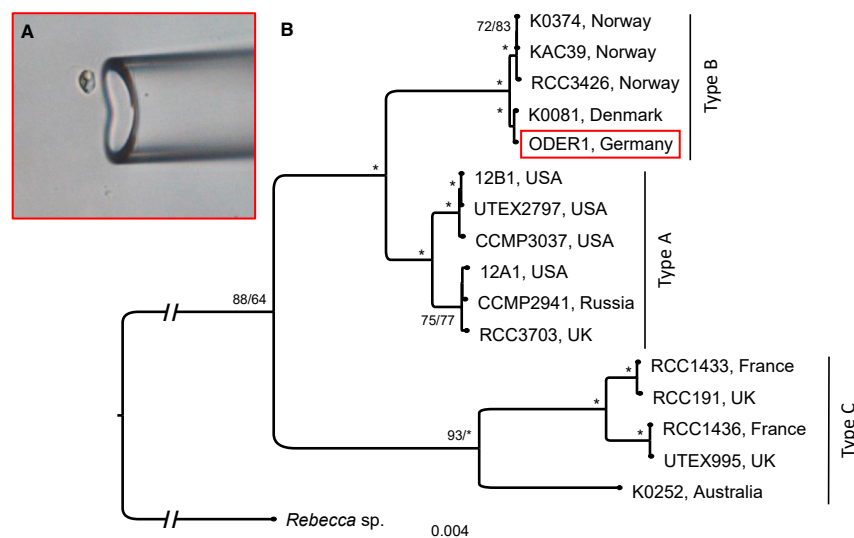


Figure 1. Strain isolation and phylogenetic relationship of *Prymnesium parvum* ODER1

(A) Photomicrograph showing the isolation of a living cell of the ODER1 strain (obtained from the Oder River; [Figure S5](#)) with a micromanipulator; all material (cells serving for DNA and RNA extraction) used for whole-genome sequencing was generated by cultivation from a single cell.

(B) Maximum likelihood phylogeny of 16 *P. parvum* s.l. strains based on chloroplast genomes, rooted with the haptophyte *Rebecca* sp. Node labels indicate ultra-fast-bootstrap (UFBS)/Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT) support values. Values >95% are indicated by asterisks. Reference-quality genomes exist only for the type B clade that we newly report here (red box) and for the type A clade (12B1, UTEX2797, CCMP3037).^{15,16} The closest relative of the diploid *P. parvum* ODER1 strain is K-0081, which has a triploid genome, according to our SNP allele frequency analysis and flow cytometry data ([Figures 3A and 3B](#)).

polyether backbone, whose length defines the type, as well as various pentose and/or hexose units.^{10,12} Flow cytometry, transcriptomics,^{13,14} and whole-genome sequencing^{15,16} studies indicate that *P. parvum* s.l. is a complex of at least 40 genetically distinct strains (i.e., forms within each of the types A, B, and C; [Figure 1](#)) that differ in genome size, ploidy, or both, and produce type-specific prymnesins as well as strain-specific mixtures of differently glycosylated/halogenated prymnesin variants.^{15,16} The polyketide-like chemical nature of these prymnesins suggests that polyketide synthases (PKSs) contribute to their biosynthesis,^{14,17} while glycosyltransferases are essential for polysaccharide modifications.

A better understanding of hidden diversity within *P. parvum* s.l. and the unambiguous identification of these harmful bloom-forming algae require genomic data.¹⁶ Accurate identification will also facilitate the identification of prymnesins, which is of particular interest because their structure and toxicity are type dependent.¹⁸ To understand how the genes of the major protein families evolve and enable strain-specific toxin production, comparative genomics using chromosome-scale high-quality reference genomes of *P. parvum* is essential. The knowledge gained could contribute to mitigating future impacts of this globally relevant threat, which is of increasing relevance due to freshwater salinization and climate change.¹

Here, we present the haplotype-resolved diploid genome of the *P. parvum* type B strain (hereafter “*ODER1*”) that caused the Oder disaster. Compared with a diploid type A genome, we find enormous, evenly distributed size expansion of all *ODER1* strain chromosomes (“pseudo-chromosomes”; [STAR Methods](#)). We also observe a difference in ploidy compared with the most closely related type B genome from Denmark, which is triploid in our analysis of genome size. Analysis of PKS genes shows evolutionary changes between types A and B and—in the case of glycosyltransferases—even between closely related type B genomes, which may underlie their different toxin structures and, potentially, the ecologically relevant toxicity.

RESULTS

Chloroplast phylogenetic tree assigns *P. parvum* ODER1 to the type B clade

Using long-read data (ONT Minlon; [STAR Methods](#)), we assembled a complete chloroplast genome of *P. parvum* strain ODER1 ([Figure 1A](#)). Screening short-read assemblies¹⁶ of other *P. parvum* s.l. strains for chloroplast sequences enabled us to reconstruct a well-supported phylogenetic tree of type A, B, and C strains ([Figure 1B](#)). *P. parvum* ODER1 is most closely related to a type B strain isolated in 1985 from brackish water in northwestern Denmark (K-0081) and forms a well-supported clade with other type B strains from Norway (RCC3426, KAC-39, and K-0374). The results suggest that similarity can be explained by geographic proximity and corroborates prior results from ITS sequencing of ODER1.¹

P. parvum ODER1 diploid genome assembly and comparison with type A assemblies

Using state-of-the-art HIFI sequencing and Hi-C data ([Figure S1; STAR Methods](#)), we obtained a diploid (2n) genome assembly of *P. parvum* ODER1, making it the first reference-quality assembly of a type B strain. Compared with publicly available type A reference assemblies,^{15,16} statistics and BUSCO scores ([Tables 1 and S1](#)) were consistently improved in our assembly. The most striking difference is the genome size. The size of each of our two haploid *P. parvum* ODER1 assemblies (236/237 Mbp) exceeds those of the type A strains 12B1 (94 Mbp) and CCMP3037 (107 Mbp) by a factor of more than two. We did not compare our ODER1 assembly to type A strain UTEX2797 because two publications present contradictory assembly results for this strain,^{15,16} which may be caused by high heterozygosity, hybridization, or issues of cultivation.

Dot plots to compare assemblies of strain 12B1 (type A) and ODER1 (type B) show a conserved collinearity/synteny between the 1n = 34 Hi-C, supported chromosomes in both types A and B. This excludes the possibility of size differences being the result of large-scale duplications or polyploidization ([Figure 2A](#)).

Table 1. Diploid long-read genome assembly results of *Prymnesium parvum* ODER1

	<i>P. parvum</i> ODER1 (this study)		<i>P. parvum</i> 12B1 (Wisecaver et al. ¹⁶)	<i>P. parvum</i> CCMP3037 (Jian et al. ¹⁵)
Type	B	B	A	A
Haplotype	hap1	hap2	collapsed	collapsed
Total length (bp)	236,176,151	237,246,831	93,538,114	107,321,770
Chromosomal scaffolds	34	34	34	34
Contig number	314	288	225	362
Scaffold n50 (bp)	8,642,580	8,552,854	3,203,049	3,786,890
Contig n50 (bp)	1,791,285	2,319,971	852,115	968,388
Largest contig (bp)	8,233,435	6,580,969	3,281,684	5,352,942
Repeat content	63.98%	64.23%	29.4%	29.67%

The diploid genome was resolved into two chromosome-level haploid genome assemblies of similar quality (hap1, hap2) using HIFI, ONT, and Hi-C sequencing data. Comparison with *P. parvum* type A genomes reveals an increase in genome size and repeat content in type B *P. parvum* ODER1, despite both strains being diploid. For an assessment of genome completeness using BUSCO scores, see Table S1.

Regarding structural variation, only a few inter-chromosomal rearrangements are visible, while intra-chromosomal rearrangements occur more often. This is especially true for chromosome 1 of ODER1, which matches Scaf2 of strain 12B1 (Figure 2A).

De novo repeat annotation shows that the genome-size augmentation of ODER1 is due to expansion of repetitive elements and not due to differences in gene content. Repeat content in *P. parvum* ODER1 is 64% compared with 29%–30% in both type A reference genomes. The non-repetitive fraction of the two genomes that were assembled from PacBio HIFI reads, ODER1 type B (85 Mbp) and CCMP3037 type A (75 Mbp), is similar in size, which suggests that the genome-size expansion is mainly driven by repetitive elements. Gypsy- and Copia-like retrotransposons in particular have expanded in *P. parvum* ODER1 compared with *P. parvum* 12B1 (Figure 2B). Complete open reading frames (ORFs) of retroviral genes can be predicted on many annotated repeat elements, suggesting their recent activity. Recent evolution is also supported by the detection of haplotype-specific retroelement integrations.

For comparisons between type B strains, only short reads and assemblies from these were available.¹⁶ Due to the high repeat content, we found these assemblies highly fragmented, which hindered comparison on the chromosomal level. Interestingly, K-0081 has been described as a tetraploid¹⁶ and is the closest relative of ODER1 in our chloroplast-genome phylogeny. We mapped the available short reads of K-0081 to the ODER1 assembly and found low divergence between the genomes and an allele frequency spectrum (Figure 3A) with a peak at around 33% of heterozygous variant read coverage in K-0081, which suggests triploidy rather than tetraploidy. The allele frequency plot for ODER1 clearly supported a 50% peak of heterozygous variant reads, as expected for a diploid organism. To further investigate the DNA content difference and to assess ploidy level, we performed propidium iodide (PI) flow cytometry measurements of strains ODER1 and K-0081. The DNA content of strain ODER1 is 0.55 ± 0.01 pg (534 Mbp), supporting diploidy, while strain K-0081 has a DNA content of 0.77 ± 0.02 pg (756 Mbp), supporting triploidy (Table S2). The DNA content difference between strains was further verified by simultaneous analysis (Figure 3B).

The microfibrils of the cell-covering scales, which consist of proteins and carbohydrates, showed a radial arrangement on

both faces (proximal and distal) in the ODER1 strain (Figure 3C). In rare cases, whether a pattern was radial or spiral could not be determined. A lack of spiral patterns fits the hypothesis that diploid *Prymnesium* cells lack scales with spirally wound microfibrils on the distal face, while both spiral and radial patterns occur frequently in haploid cells.^{6,8,19,20} The scales of the triploid K-0081 strain showed both types of microfibril arrangements (Figure 3D), allowing us to distinguish ODER1 from K-0081 based on morphological features. The thickness of the inflexed rim of the scales, which cover the cell in two layers, varied in both strains from narrow (outer layer) to wide (inner layer).

Analyses of PKSs in *P. parvum* type A and B long-read reference genomes

PKSs are involved in the synthesis of prymnesin toxins and can be extremely large proteins because of their repeated domain structure. Some PKSs identified from *P. parvum* type A genomes hold the “world record” in protein size (>40,000 amino acid residues) and have been named “PKZILLA-1” and “PKZILLA-2.”¹⁷ The domain structure of both enzymes has been related to the prymnesin structure. Due to the presence of large exons, small introns, and repetitiveness, the corresponding genes pose a challenge to most annotation tools and to short-read transcriptomics. Interestingly, we found that an *ab initio* annotation tool, GENSCAN,²¹ performed best on the prediction of PKS. This is likely the result of the tool targeting maximum ORF lengths rather than typical exon and intron sizes. Using the two haploid ODER1 assemblies and the available *P. parvum* type A long-read assemblies (12B1, CCMP3037, and UTEX2797),^{15,16} we were able to predict 26 phylogenetic clades of PKS gene sequences (Figure S2A). Although most sequence clades ($n = 18$) have the corresponding PKS gene present in all genomes examined, we found three clades in which the PKS gene is deleted and one clade in which the PKS gene is tandemly repeated in ODER1 (Figures S2A and S2B). This pattern indicates divergent evolution of the PKS gene family among type A and B strains, which likely contributes to the production of structurally different toxins, prymnesins A and B. Interestingly, we also found differences in PKS tandem duplications between the two haplotype assemblies of ODER1, suggesting very recent evolutionary changes (Figure S2C).

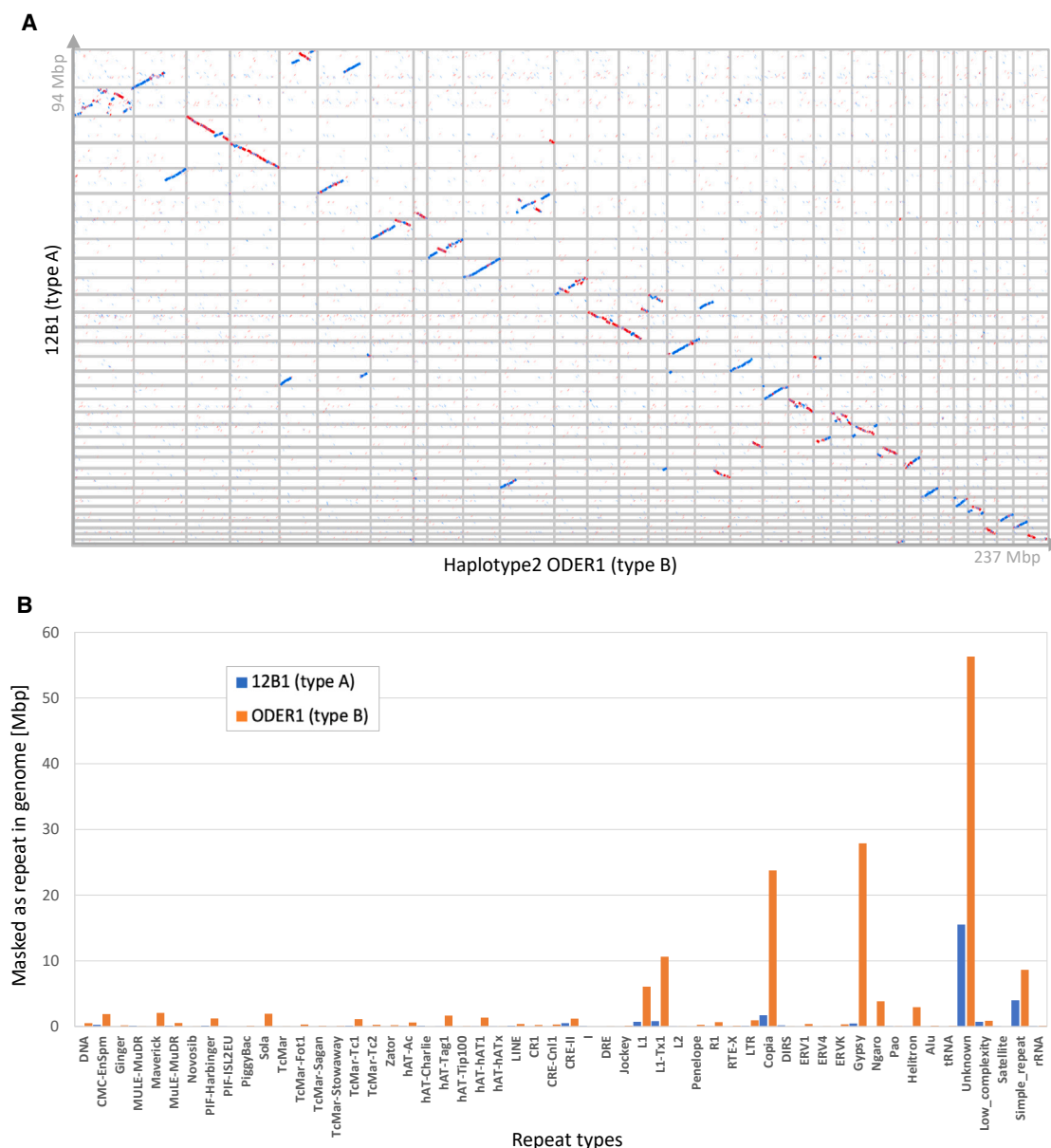


Figure 2. Haploid genome-size expansion between *P. parvum* types A and B

(A) Dot plot comparison between assembled chromosomes of diploid *P. parvum* ODER1 (type B, x axis; Figure S1) and diploid *P. parvum* 12B1 (type A, y axis) shows enormous though relatively evenly distributed size expansion of all chromosomes in the ODER1 strain and some structural rearrangements (blue, forward strand; red, reverse complement strand).

(B) Expansion of several repeat classes (L1, Copia, Gypsy, unknown) explains major genome-size differences between the type A and type B strains.

A deletion in the PKS gene of ODER1 explains structural differences of type A and B prymnesins

PKS gene family analysis made evident a size difference in the largest-predicted PKS protein (PKZILLA-1¹⁷) between ODER1 and the type A strains, while another large PKS (PKZILLA-2) is conserved (Figure 4). Comparison of the corresponding genomic regions between ODER1 and the CCMP3037 assemblies, using HIFI long-read data that were assembled without gaps in the regions, revealed a large deletion (~20 kbp) and a smaller duplication (~4 kbp) that were both specific to ODER1. No signatures of viral contributions (e.g., LTRs,

LINES) are found in the ODER1 PKZILLA-1 gene. Interproscan (<https://www.ebi.ac.uk/interpro/>) of the corresponding sequences show that the deletion removes six KS3_2 domains from ODER1 PKZILLA-1, while the duplication adds one KS3_2 domain (Figure S3). We also found the larger deletion in other type B strains by mapping published short reads against the CCMP3037 assembly and inspecting read coverage (Figure 5). In sum, the five missing KS3_2 domains in type B strain PKZILLA-1 could explain the missing 1,6-dioxadecalin core unit (Figure 4, highlighted in red)¹⁰ in the structure of the prymnesin B toxins.

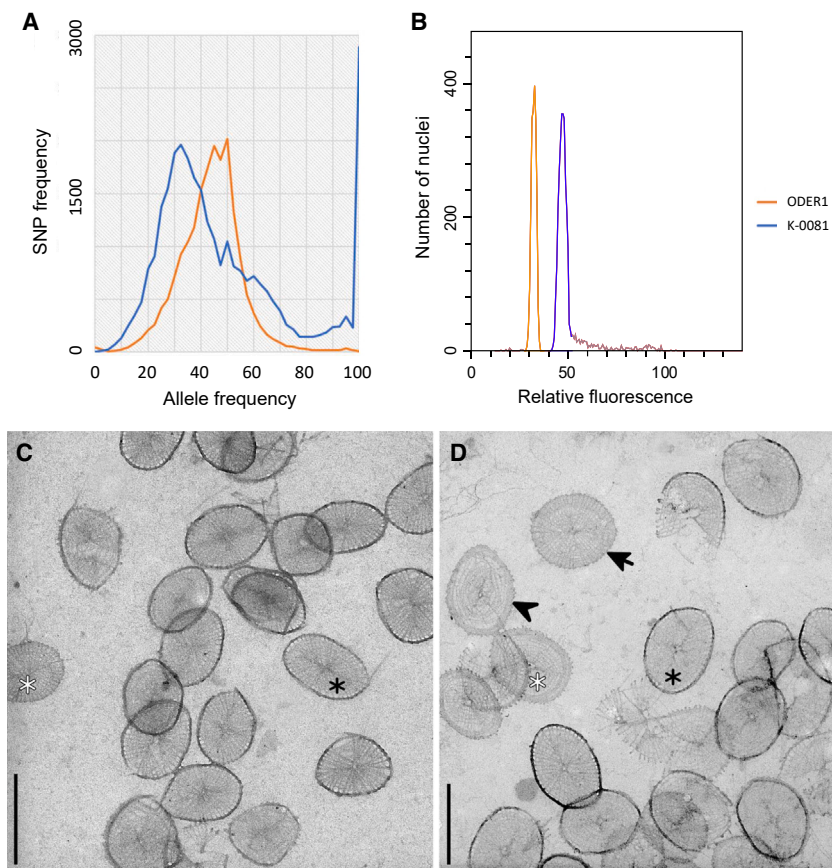


Figure 3. Ploidy inference of closely related *Prymnesium parvum* strains ODER1 and K-0081

(A) As expected for a diploid organism, the ODER1 strain showed a peak at 50% allele frequency, while the K-0081 strain had a peak at 33% and a shoulder at around 60%, suggesting a triploid rather than tetraploid genome. The number of homozygous variants in K-0081 was low (28,884 variants with 100% allele frequency and coverage $\geq 20\times$), underlining the very close relationship of both strains.

(B) Simultaneous flow cytometric analysis of the type B strains ODER1 and K-0081. The relative fluorescence of nuclei stained with propidium iodide shows ploidy-level difference between the strains. The resulting DNA content of strain ODER1 (0.55 pg) corresponds to diploidy, while the DNA content of strain K-0081 (0.77 pg) corresponds to triploidy (see also Table S2).

(C and D) Transmission electron microscopy (TEM) showing the body scales of the outer and inner layers of *P. parvum* (C) strain ODER1 and (D) strain K-0081. Scales of the outer layer are characterized by a narrow rim (black asterisk) and those of the inner layer by a wide rim (white asterisk). Scales of ODER1 exhibit a radial arrangement of microfibrils on both faces (typical for diploid cells) whereas scales of K-0081 exhibit a radial pattern on the proximal face (arrow) and a spirally wound pattern on the distal face (arrowhead; for more details, please see Figure S6). Scale bars, 400 nm.

DISCUSSION

Our work provides unprecedented detail regarding the biological agent that, by its toxin production, was the mechanism causing the Oder River disaster in 2022. To our knowledge, this study presents the first reference-quality assembly of a type B *P. parvum* genome and the first haplotype-resolved genome of a haptophyte microalga. Our ODER1 reference genome provides insights into the genetic basis and variability of the toxin production. The chemical structure of the type B toxin backbone is distinguished from that of type A by the absence of a 1,6-dioxadecalin-core unit,¹ and the underlying PKS genes show type-specific differences on the level of gene families and gene structures (Figures 4, 5, S2, and S3). The structural differences within the largest PKS gene in types A and B result in the gain or loss of five ketide-synthase domains in the respective proteins. This change in domain number is close to the theoretical counts of 3–4 that have been predicted based on the number of C atoms in the prymnesin backbone.¹⁰ We thus hypothesize that an evolutionary change in the PKS genes provides the basis for a typical type B toxin. The large 20-kbp deletion in type B *P. parvum* s.l. (Figures 4 and 5) does not appear to result in a deleterious shift in the PKS ORF because a compromised gene function would imply a loss of toxicity, which has not been observed in any known type B. Prymnesin structural modifications specific to the Oder River are also characterized by different expressions of

certain prymnesins. This might be connected to differences of K-0081 and ODER1 regarding a glycosyltransferase gene (Figure S4).¹ Recently, it was reported that the prymnesin type influences its toxicity,¹⁸ with cytotoxic potencies ranked as type A > C > B. Our hypothesis that changes to the PKS gene that lead to the difference between types A and B prymnesin backbones may have an influence on the folding capacities of the large prymnesin molecule and may result in different toxicity. Whether or not the lack of additional sugar units has an impact on the degree of toxicity is not yet known but is currently a subject of study within the ODER-SO project (<https://www.igb-berlin.de/en/oder-so>).

The new ODER1 reference genome further adds to the evidence by Wisecaver et al.¹⁶ that the hidden diversity within *P. parvum* s.l. indicates the occurrence of cryptic species and can only be comprehensively understood using state-of-the-art genomics. The ODER1 assembly provides direct evidence that the expansion of retroelements was a driver of haploid genome-size evolution, as earlier inferred from indirect methods (i.e., estimated from highly fragmented short-read assemblies).¹⁶ Their high repeat content renders genome assembly of type B more complex than type A strains and underlines the need for high-accuracy long-read approaches. In the future, repetitive (retro)viral sequences in the genome may help us to discover potentially infectious viruses in environmental samples that might serve as agents of biological control to fight *P. parvum* blooms.²³

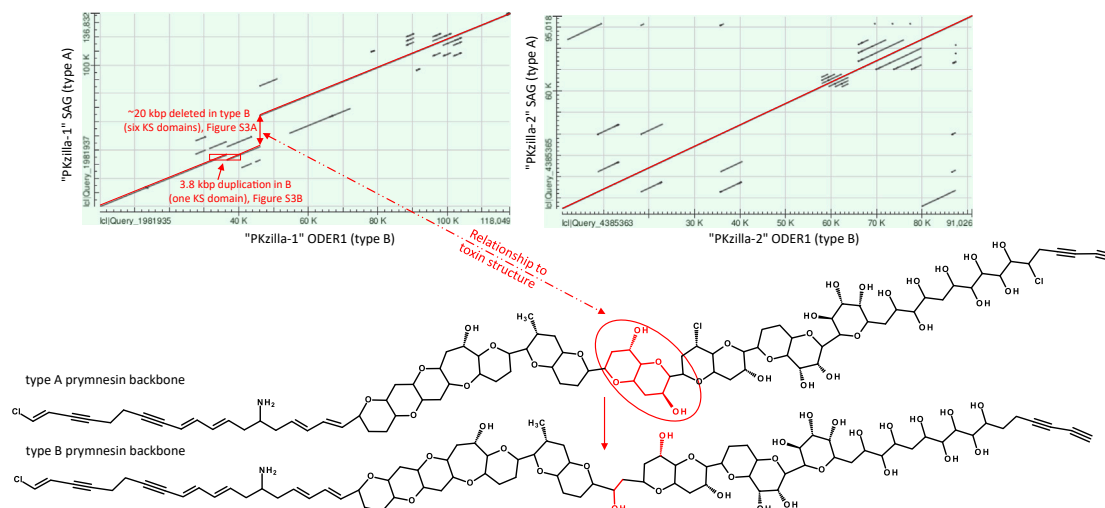


Figure 4. Identification of structural changes in the largest polyketide synthases in *Prymnesium parvum* type B versus type A

Structural differences between PKZILLA-1 genes in type A and type B strains may correspond to structural changes of the prymnesin toxins produced by these genes. In contrast, PKZILLA-2 does not show structural variation between types. The backbone structure of type A prymnesins according to Igarashi et al.,²² with three incorporated chlorine atoms (C-1, C-56, and C-85), and those of type B prymnesins according to Rasmussen et al.,¹² with one chlorine at position C-1 are displayed. Although the type A prymnesin backbone consists of 91 carbons, the type B has only 85 carbon atoms due to the replacement of one 1,6-dioxadecal core unit (marked in red) with a short acyclic C2-linkage. For further information, see Figures S2–S4.

In addition to haploid genome-size evolution, differences in ploidy—which may involve life cycle stages and evolutionary change—contribute to the diversity within *P. parvum* s.l. and complicate our understanding of their taxonomy and systematics. Comparative flow cytometric analyses as well as our genomic and morphological data clearly show that the ODER1 strain is a diploid form. During their life cycle, *P. parvum* s.l. microalgae alternate between haploid and diploid life stages, each of which can reproduce by asexual mitotic division⁸ and may be able to generate blooms. Thus, genotypically highly similar if not identical life forms may

occur. According to our analyses, the closest known relative of the diploid ODER1 strain is the triploid K-0081 from Denmark. It remains unclear whether this difference is indicative of previously undescribed genetic and genomic plasticity in these algae, or is an artifact of nearly four decades of K-0081 cultivation. Of note, K-0081 was previously described as tetraploid, possibly due to an underestimation of its haploid genome size based on short-read sequencing in combination with flow cytometric data that often result in 10%–20% larger genome size estimates than sequencing data.²⁴

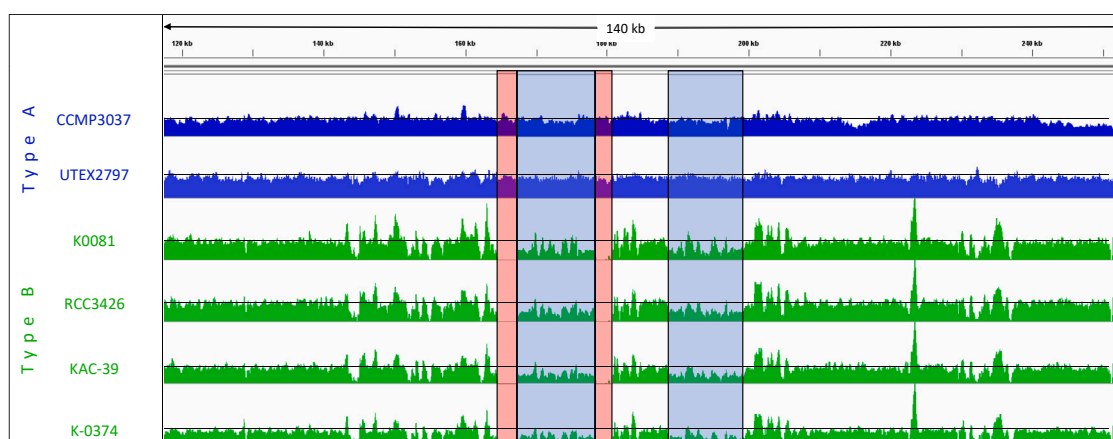


Figure 5. Structural differences of the largest polyketide synthases between several type A and B *Prymnesium parvum*

PKZILLA-1 deletion in four additional type B strains (green: K-0081, K-0374, KAC-39, and RCC3426) compared with two type A strains (blue: CMP3037 and UTEX2797), as revealed by short-read mapping to the genomic region of type A PKZILLA-1 (from the genome assembly of CCMP3037). Due to the repetitive nature of the PKZILLA-1 gene, large parts cannot be covered by uniquely mapped reads. Non-uniquely mapped reads are randomly distributed between two larger regions (light blue) that exhibit reduced (~50%) read coverage in type B, indicating a deletion in type B or a duplication in type A. Borders of the left region do not have any read coverage (red), further supporting the deletion of this part of the gene in type B strains. For further information, see Figures S2–S4.

Although the lack of spiral arrangement in microfibrils in our analysis of scale morphology supports diploidy of ODER1, scales had not been examined in triploid *P. parvum* s.l. before. Our morphological data neither corroborate nor contradict the K-0081 triploidy inferred from flow cytometric and SNP allele frequency analyses. The triploid K-0081 exhibits radial and spirally wound microfibril arrangement, so far known only in haploid cells.^{6,8,19,20}

Together with the available type A reference genomes, our high-quality genome of type B presents an important basic research contribution for comparative genomic analyses of this globally relevant group of microalgae. Using short-read sequencing, the reference genomes (type C is pending) now allow, when detected in a specific region, taxonomically determining *P. parvum* s.l. with little effort in whole-genome detail. This will form the basis for the development of taxon-specific control methods and surveillance of its potential evolutionary adaptation and change. The link between PKS and toxins enables a better understanding of the mechanistic relationships between gene expression, toxin production, and ecological/environmental conditions. To understand the interaction of these factors in natural water bodies, growth and toxicity experiments are required to predict *P. parvum* strain-specific blooms and the causal ecological conditions. *P. parvum* remained present in the entire Oder River after the summer 2022 bloom, as documented using a molecular-quantification (qPCR) assay. It shows the presence of this microalga (03/2023–02/2024), including major shifts in quantities (Figure S5), suggesting a potential for future blooms when the compound environmental conditions that trigger such a disaster are again fulfilled.¹

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Culturing methods
- **METHOD DETAILS**
 - High Molecular Weight (HMW) DNA extraction and sequencing
 - Hi-C library construction and sequencing
 - Diploid genome assembly
 - RNA extraction, sequencing, and transcriptome assembly
 - Repeat analysis and comparison
 - Annotation
 - Transmission electron microscopy
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - *Prymnesium parvum* phylogeny based on chloroplast DNA
 - Estimation of DNA content by flow cytometry and ploidy level assignment using genomics
 - qPCR-assay to quantify *Prymnesium parvum* in water samples

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.06.033>.

ACKNOWLEDGMENTS

We thank colleagues at the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) for help with field and laboratory work, especially Katrin Preuss, Miriam Diemer, Shambhavi Parmar, and Karla Münzer, who raised the *Prymnesium* cultures and prepared samples; Vanessa Bremerich for cartographic support; and Martin Pusch for supporting our research proposal. Hi-C sequencing was performed at the Sequencing Core Facility of the Max Planck Institute for Molecular Genetics (Berlin, Germany). We thank Susan Mbedi (Berlin Center for Genomics in Biodiversity Research) for help with sequencing and Renate Radek (Freie Universität Berlin) for kind assistance with electron microscopy. *P. parvum* strain K-0081 was obtained from the Norwegian Culture Collection of Algae (NORCCA, Oslo); for strains UTEX2797 and RCC1436, we thank Per Juel Hansen (University of Copenhagen). D.Č. is grateful for the financial support of the Martina Roeselová Memorial Fellowship granted by the IOCB Tech Foundation. Our study was conducted as part of the project ODER~SO, financed by the German Federal Agency for Nature Conservation (BfN) with funds from the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV).

AUTHOR CONTRIBUTIONS

Conceptualization, M.S., M.T.M., H.K., J.F.H.S., and D.K.L.; resources, J.K., J.F.H.S., and M.S.; data collection and curation, H.K., J.F.H.S., D.Č., and E.K.; software, H.K.; formal analysis, H.K., J.F.H.S., D.Č., and S.W.; funding acquisition, M.S. and M.T.M.; writing – original draft, H.K., M.S., J.F.H.S., E.V., and M.T.M.; writing – review and editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 29, 2024

Revised: May 13, 2024

Accepted: June 11, 2024

Published: July 9, 2024

REFERENCES

1. Köhler, J., Varga, E., Spahr, S., Gessner, J., Stelzer, K., Brandt, G., Mahecha, M.D., Kraemer, G., Pusch, M., Wolter, C., et al. (2024). Unpredicted ecosystem response to compound human impacts in a European river. *Sci. Rep.* 14, 16445. <https://doi.org/10.1038/s41598-024-66943-9>.
2. Szlauer-Łukaszewska, A., Ławicki, Ł., Engel, J., Drewniak, E., Cieżak, K., and Marchowski, D. (2024). Quantifying a mass mortality event in freshwater wildlife within the Lower Odra River: insights from a large European river. *Sci. Total Environ.* 907, 167898. <https://doi.org/10.1016/j.scitotenv.2023.167898>.
3. Sobieraj, J., and Metelski, D. (2023). Insights into toxic *Prymnesium parvum* blooms as a cause of the ecological disaster on the Odra River. *Toxins (Basel)* 15, 403. <https://doi.org/10.3390/toxins15060403>.
4. Sproston, N.G. (1946). Fish mortality due to a brown flagellate. *Nature* 158, 70. <https://doi.org/10.1038/158070b0>.
5. Our Molecular Biology Correspondent (1967). Polymerase activity and dissociation. *Nature* 216, 852–853.
6. Green, J.C., Hibberd, D.J., and Pienaar, R.N. (1982). The taxonomy of *Prymnesium* (Prymnesiophyceae) including a description of a new cosmopolitan species, *P. patellifera* sp. nov., and further observations on *P. parvum* N. carter. *Br. Phycol. J.* 17, 363–382. <https://doi.org/10.1080/00071618200650381>.
7. Larsen, A., and Medlin, L.K. (1997). Inter- and intraspecific genetic variation in twelve *Prymaesium* (haptophyceae) clones. *J. Phycol.* 33, 1007–1015. <https://doi.org/10.1111/j.0022-3646.1997.01007.x>.
8. Gastineau, R., Davidovich, N.A., Hallegraef, G., Probert, I., and Mouget, J.-L. (2014). Reproduction in microalgae. In *Reproductive Biology of*

- Plants, K.G. Ramawa, J.-M. Mérillon, and K.R. Shivanna, eds. (CRC Press), pp. 1–28.
9. Lutz-Carrillo, D.J., Southard, G.M., and Fries, L.T. (2010). Global genetic relationships among isolates of golden alga (*Prymnesium parvum*) 1. J. American Water Resour. Assoc. 46, 24–32. <https://doi.org/10.1111/j.1752-1688.2009.00388.x>.
10. Binzer, S.B., Svenssen, D.K., Daugbjerg, N., Alves-de-Souza, C., Pinto, E., Hansen, P.J., Larsen, T.O., and Varga, E. (2019). A-, B- and C-type prymnesins are clade specific compounds and chemotaxonomic markers in *Prymnesium parvum*. Harmful Algae 81, 10–17. <https://doi.org/10.1016/j.hal.2018.11.010>.
11. Driscoll, W.W., Wisecaver, J.H., Hackett, J.D., Espinosa, N.J., Padway, J., Engers, J.E., and Bower, J.A. (2023). Behavioural differences underlie toxicity and predation variation in blooms of *Prymnesium parvum*. Ecol. Lett. 26, 677–691. <https://doi.org/10.1111/ele.14172>.
12. Rasmussen, S.A., Meier, S., Andersen, N.G., Blossom, H.E., Duus, J.O., Nielsen, K.F., Hansen, P.J., and Larsen, T.O. (2016). Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. J. Nat. Prod. 79, 2250–2256. <https://doi.org/10.1021/acs.jnatprod.6b00345>.
13. Koid, A.E., Liu, Z., Terrado, R., Jones, A.C., Caron, D.A., and Heidelberg, K.B. (2014). Comparative transcriptome analysis of four prymnesiophyte algae. PLoS One 9, e97801. <https://doi.org/10.1371/journal.pone.0097801>.
14. Anestis, K., Kohli, G.S., Wohrlab, S., Varga, E., Larsen, T.O., Hansen, P.J., and John, U. (2021). Polyketide synthase genes and molecular trade-offs in the ichthyotoxic species *Prymnesium parvum*. Sci. Total Environ. 795, 148878. <https://doi.org/10.1016/j.scitotenv.2021.148878>.
15. Jian, J., Wu, Z., Silva-Núñez, A., Li, X., Zheng, X., Luo, B., Liu, Y., Fang, X., Workman, C.T., Larsen, T.O., et al. (2024). Long-read genome sequencing provides novel insights into the harmful algal bloom species *Prymnesium parvum*. Sci. Total Environ. 908, 168042. <https://doi.org/10.1016/j.scitotenv.2023.168042>.
16. Wisecaver, J.H., Auber, R.P., Pendleton, A.L., Watervoor, N.F., Fallon, T.R., Riedling, O.L., Manning, S.R., Moore, B.S., and Driscoll, W.W. (2023). Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. Curr. Biol. 33, 2246–2259.e8. <https://doi.org/10.1016/j.cub.2023.05.003>.
17. Fallon, T.R., Shende, V.V., Wierzbicki, I.H., Auber, R.P., Gonzalez, D.J., Wisecaver, J.H., and Moore, B.S. (2024). Giant polyketide synthase enzymes biosynthesize a giant marine polyether biotoxin. Preprint at bioRxiv. <https://doi.org/10.1101/2024.01.29.577497>.
18. Varga, E., Prause, H.-C., Riepl, M., Hochmayr, N., Berk, D., Attakpah, E., Kiss, E., Medić, N., Del Favero, G., Larsen, T.O., et al. (2024). Cytotoxicity of *Prymnesium parvum* extracts and prymnesin analogs on epithelial fish gill cells RTgill-W1 and the human colon cell line HCEC-1CT. Arch. Toxicol. 98, 999–1014. <https://doi.org/10.1007/s00204-023-03663-5>.
19. Larsen, A., and Edvardsen, B. (1998). Relative ploidy levels in *Prymnesium parvum* and *P. patelliferum* (Haptophyta) analyzed by flow cytometry. Phycologia 37, 412–424. <https://doi.org/10.2216/i0031-8884-37-6-412.1>.
20. Eikrem, W., Medlin, L.K., Henderiks, J., Rokitta, S., Rost, B., Probert, I., Throndsen, J., and Edvardsen, B. (2017). Haptophyta BT - Handbook of the Protists, J.M. Archibald, A.G.B. Simpson, and C.H. Slamovits, eds. (Springer International Publishing), pp. 893–953.
21. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94. <https://doi.org/10.1006/jmbi.1997.0951>.
22. Igarashi, T., Satake, M., and Yasumoto, T. (1996). Prymnesin-2: a potent ichthyotoxic and hemolytic glycoside isolated from the red tide alga *Prymnesium parvum*. J. Am. Chem. Soc. 118, 479–480. <https://doi.org/10.1021/ja9534112>.
23. Wagstaff, B.A., Vladu, I.C., Barclay, J.E., Schroeder, D.C., Malin, G., and Field, R.A. (2017). Isolation and characterization of a double stranded DNA megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. Viruses 9, 40. <https://doi.org/10.3390/v9030040>.
24. Kasai, F., O'Brien, P.C.M., and Ferguson-Smith, M.A. (2013). Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping. Genomics 102, 468–471. <https://doi.org/10.1016/j.ygeno.2013.09.002>.
25. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
26. Zhang, H., Song, L., Wang, X., Cheng, H., Wang, C., Meyer, C.A., Liu, T., Tang, M., Aluru, S., Yue, F., et al. (2021). Fast alignment and preprocessing of chromatin profiles with Chromap. Nat. Commun. 12, 6566. <https://doi.org/10.1038/s41467-021-26865-w>.
27. Zheng, Z., Li, S., Su, J., Leung, A.W.-S., Lam, T.-W., and Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. Nat. Comput. Sci. 2, 797–803. <https://doi.org/10.1038/s43588-022-00387-x>.
28. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. 38, 5825–5829. <https://doi.org/10.1093/molbev/msab293>.
29. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
30. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18, 170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
31. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
32. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101. <https://doi.org/10.1016/j.cels.2015.07.012>.
33. Frith, M.C., and Kawaguchi, R. (2015). Split-alignment of genomes finds orthologies more accurately. Genome Biol. 16, 106. <https://doi.org/10.1186/s13059-015-0670-9>.
34. Kiebas, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011). Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487–493. <https://doi.org/10.1101/gr.113985.110>.
35. Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32, 2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>.
36. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
37. Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics 39, btad014. <https://doi.org/10.1093/bioinformatics/btad014>.
38. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14, 708–715. <https://doi.org/10.1101/gr.1933104>.
39. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA 117, 9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
40. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
41. Perte, M., Perte, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of

- a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. <https://doi.org/10.1038/nbt.3122>.
42. Niknafs, Y.S., Pandian, B., Iyer, H.K., Chinnaiyan, A.M., and Iyer, M.K. (2017). TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat. Methods* 14, 68–70. <https://doi.org/10.1038/nmeth.4078>.
43. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
44. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
45. Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. <https://doi.org/10.1038/s41592-019-0669-3>.
46. Zhou, C., McCarthy, S.A., and Durbin, R. (2023). YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* 39, btac808. <https://doi.org/10.1093/bioinformatics/btac808>.
47. Kuhl, H. (2024). HANNO: efficient high-throughput annotation of protein coding genes in eukaryote genomes (v0.4). Zenodo. <https://doi.org/10.5281/zenodo.11532370>.
48. Auber, R., and Wisecaver, J.H. (2019). Algal nuclei isolation for nanopore sequencing of HMW DNA V.3 protocols.io. <https://doi.org/10.17504/protocols.io.5i2g4ge>.
49. Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37, 4572–4574. <https://doi.org/10.1093/bioinformatics/btab705>.
50. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>.
51. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>.
52. Dpooležel, J., Binarová, P., and Lcretti, S. (1989). Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* 31, 113–120. <https://doi.org/10.1007/BF02907241>.
53. Temsch, E.M., Greilhuber, J., and Krisai, R. (2010). Genome size in liverworts. *Preslia* 82, 63–80.
54. Veselý, P., Bureš, P., Šmarda, P., and Pavlíček, T. (2012). Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Ann. Bot.* 109, 65–75. <https://doi.org/10.1093/aob/mcr267>.
55. Doležel, J., and Bartoš, J. (2005). Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* 95, 99–110. <https://doi.org/10.1093/aob/mci005>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw read data	This paper	BioProject: PRJNA1101072
Assembled genomes and transcriptomes	This paper	BioProject: PRJNA1101072
Experimental models: Organisms/strains		
<i>Prymnesium parvum</i> ODER1	Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)	ODER1
<i>Prymnesium parvum</i> K-0081	Norwegian Culture Collection of Algae (NORCCA)	K-0081
<i>Prymnesium parvum</i> RCC1436	Prof. Per Juel Hansen (University of Copenhagen)	RCC1436
<i>Prymnesium parvum</i> UTEX2797	Prof. Per Juel Hansen (University of Copenhagen)	UTEX2797
Oligonucleotides		
5'-CACATCCGATCGTGTCTGC-3'	This paper	PrymF2239
5'-GGCACAAACGACTTGGTAGG-3'	This paper	PrymR2384
Software and algorithms		
BUSCO v. 3.0.2	Simão et al. ²⁵	https://busco.ezlab.org/
Chromap v. 0.2.4-r467	Zhang et al. ²⁶	https://github.com/haowenz/chromap
CLAIR3 v. 1.0.4	Zheng et al. ²⁷	https://github.com/HKU-BAL/Clair3
eggNOG v. 5.0.2	Cantalapiedra et al. ²⁸	http://eggno-mapper.embl.de/
Flye v. 2.9.2-b1786	Kolmogorov et al. ²⁹	https://github.com/fenderglass/Flye
GENSCAN v. 1998	Burge and Karlin ²¹	http://hollywood.mit.edu/GENSCANinfo.html
Hifiasm v. 0.19.6-r595	Cheng et al. ³⁰	https://github.com/chhylp123/hifiasm
IQ-TREE 2 v. 1.6.12	Minh et al. ³¹	http://www.iqtree.org/
Juicebox v. 1.11.08	Durand et al. ³²	https://aidenlab.org/juicebox/
Last-split v. 1454	Frith and Kawaguchi ³³	https://gitlab.com/mcfrith/last/-/blob/main/doc/last-split.rst
LAST v. 1454	Kielbasa et al. ³⁴	https://gitlab.com/mcfrith/last
Minidot v. 0.3-r179	Li ³⁵	https://github.com/thackl/minidot
Minimap2 v. 2.26-r1175	Li ³⁶	https://github.com/lh3/minimap2
Miniprot v. 0.12-r237	Li ³⁷	https://github.com/lh3/miniprot
MULTIZ v. 11.2	Blanchette et al. ³⁸	https://github.com/multiz/multiz
RepeatModeler v. open-1.0.8	Flynn et al. ³⁹	https://www.repeatmasker.org/RepeatModeler/
SAMtools v. 1.17	Danecek et al. ⁴⁰	https://github.com/samtools/
StringTie v. 2.2.1	Pertea et al. ⁴¹	https://ccb.jhu.edu/software/stringtie/
TACO v. 0.7.3	Niknafs et al. ⁴²	https://tacorna.github.io/
TransDecoder v. 5.5.0	Haas et al. ⁴³	https://github.com/sghignone/TransDecoder
Trimmomatic v. 0.39	Bolger et al. ⁴⁴	https://github.com/usadellab/Trimmomatic
Trinity v. 2.10.0	Haas et al. ⁴³	https://github.com/trinityrnaseq/trinityrnaseq/releases
WTDBG2 v. 2.2	Ruan and Li ⁴⁵	https://github.com/ruanjue/wtdbg2
YaHS v. 1.2a.1	Zhou et al. ⁴⁶	https://github.com/c-zhou/yahs

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Matthias Stöck (matthias.stoeck@igb-berlin.de).

Materials availability

Prymnesium parvum ODER1 can be obtained from Jan Köhler, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany.

Data and code availability

- Genomic and transcriptomic read data newly generated for this study are available from NCBI as BioProject PRJNA1101072. A genome browser for the ODER1 assembly has been established at <http://genomes.igb-berlin.de:8081>
- Code developed for the paper is available at <https://github.com/HMPNK/HANNO47>
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Culturing methods

Sampling and cultivation of *Prymnesium parvum* ODER1: A water sample from the Oder River was taken in Küstrin (river-km 617) on the 19th of August 2022. The sample was then filtered (Nuclepore Track-Etched membrane filters, 5- μ m pore size, 25-mm diameter, Whatman) and the retained cells transferred to water originating from Lake Müggelsee in Berlin that had been sterile-filtered using 0.2- μ m pore size (Chromafil CA-20/25 S, Macherey-Nagel, Düren, Germany; cellulose acetate, 25-mm diameter), autoclaved, and supplemented with NaCl to reach a salinity of 0.2%. To avoid potential mixing of different strains that may have been present, and to minimize contamination, we started cultures from single *P. parvum* cells that were isolated using a micromanipulator (MMO-202ND; Narishige, Tokyo, Japan) equipped with a microinjector (CellTramm Oil; Eppendorf, Hamburg, Germany) that held a capillary (Figure 1A). One randomly picked cell was used to create a cell line named ‘ODER1’. Isolated cells were propagated in f/2 Medium (Culture Collection of Algae, University of Göttingen, Germany; https://sagdb.uni-goettingen.de/culture_media/) at a salinity of 0.5%, 20 °C, and a 12 h:12 h light-dark cycle (80 μ mol photons m⁻² s⁻¹). To quickly obtain high cell densities needed for RNA sequencing (see below), some cells were separated and propagated in 1/2 SWES Brackish Water (Culture Collection of Algae, University of Göttingen, Germany) at a salinity of 3%.

METHOD DETAILS

High Molecular Weight (HMW) DNA extraction and sequencing

Once *P. parvum* ODER1 cell concentrations reached approximately 1 million cells/mL⁻¹, ca. 400 mL of culture were isolated following the nuclei isolation protocol published at www.protocols.io by Auber and Wisecaver.⁴⁸ High molecular weight (HMW) DNA was extracted from isolated nuclei using the Nanobind plant nuclei kit (Pacific Biosciences, Menlo Park, USA) according to the manufacturer’s protocol. DNA was quantified using UV-spectroscopy (NanoDrop, Thermo Fisher Scientific, Dreieich, Germany) and fluorometry (Qbit, Thermo Fisher Scientific, Dreieich, Germany). To check DNA fragment size quality, a sequencing run was performed on a Minlon sequencer (Oxford nanopore Technologies, Oxford, UK) using 1 μ g HMW DNA (sheared five times by a G23 needle), the LSK-110 library preparation kit, and a R9.4.1 flow cell. ONT sequencing reads, base-called using Guppy 6 and the dna_r9.4.1_450bps_plant_sup.cfg model, passed N50 read length >15 kbp. The final sequencing of the HMW DNA was done on a Pacbio Revio sequencer at Novogene (UK) using circular consensus read mode (CCS/HIFI).

Hi-C library construction and sequencing

The Arima High Coverage HiC Kit (Arima Genomics, Carlsbad, CA, USA) was used to construct a Hi-C sequencing library. We used about 7.5×10^7 *P. parvum* cells and followed the manufacturer’s protocols for nucleated blood. The sequencing library was constructed using the ARIMA protocols for the Accel NGS 2S Plus DNA Library Kit (Swift Biosciences, Ann Arbor, USA). This library was amplified by nine cycles of PCR. As a quality check, the library was sequenced on our in-house Minlon device as described above. The production-scale sequencing of the Hi-C library was performed on a NextSeq2000 sequencer (Illumina) using 150 bp paired-end read mode at the Berlin Center for Genomics in Biodiversity Research.

Diploid genome assembly

The longest 1.5 million CCS/HIFI reads (N50 read length 19,300 bp; 29.4 Gbp in total) were used for *de novo* assembly with Hifiasm³⁰ (0.19.6-r595). Hi-C Illumina data were included to improve haplotype phasing (options: -h1 -h2). ONT reads were included to allow for gap closure (option: -ul). The Hi-C Illumina reads were independently mapped to the resulting Hifiasm haplotype 1 or haplotype 2 contigs using Chromap,²⁶ and these were scaffolded to chromosome-scale by YaHS.⁴⁶ Technically, these scaffolds are pseudo-chromosomes, i.e., chromosomal-sized scaffolds constructed from Hi-C data without karyotype or genetic linkage map anchoring. The Hi-C scaffolds were manually curated using Juicebox,³² and bacterial contamination could be removed as these contigs had clearly reduced Hi-C signals. A few gaps in the assemblies could be closed because neighboring contigs had long overlaps. The two haploid assemblies were aligned to each other with Minimap2 and stringent mapping parameters “-x asm5” for genomes with divergence of less than 5%.^{36,49} Results were plotted in a dotplot fashion using Minidot.³⁵ The haploid assemblies were also

compared to the *P. parvum* 12B1 type A reference genome¹⁶ (Minimap2, Minidot but using less stringent alignment parameters: -x map-ont). A genome browser for the ODER1 assembly has been established at <http://genomes.igb-berlin.de:8081/>. Genome assembly, whole genome sequencing (WGS) data and RNA-seq reads are available from NCBI as Bioproject PRJNA1101072.

RNA extraction, sequencing, and transcriptome assembly

Cultured *P. parvum* strains UTEX2797 and RCC1436 were obtained from the University of Copenhagen and raised in the same growth conditions (3% salinity) used for ODER1 as described above. After attaining cell densities of about 1 million cells/mL, 130 mL of each culture (ODER1, UTEX2797, RCC1436) was filtered using GF/F glass fiber filters (pore size 0.7 μ m), and total RNA was extracted using the Transcriptome RNeasy PowerWater Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Libraries were prepared with the TruSeq stranded mRNA library protocol (poly A selection). Transcriptomes were then sequenced (PE 150 bp) at MacroGen Europe on the Illumina NovaSeq platform. Reads were trimmed with Trimmomatic⁴⁴ (using the options LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 4:15, and MINLEN: 36), assembled with Trinity, and proteins were predicted with TransDecoder⁴³ using default settings.

Repeat analysis and comparison

De novo repeat analysis was performed on *P. parvum* ODER1 as well as on *P. parvum* 12B1 using RepeatModeler/RepeatMasker.³⁹ Repeat annotations of different repeat classes were summarized by the script "buildSummary.pl" to allow for comparison between the different genomes.

Annotation

A set of TransDecoder proteins from our *P. parvum* types A, B, and C transcriptome assemblies (UTEX2797, ODER1, RCC1436, respectively) and Wisecaver et al.¹⁶ (strain 12B1 annotated proteins) was compiled and then splice-aligned with the genome assembly using Miniprot³⁷ with gtf output. Transcript sequences were splice-aligned with the genome assembly using Minimap2 (-x splice), supported by a splice junction file generated from the prior protein alignments. Minimap2 sam output was converted to gtf format. The strand of the mRNA alignments in the gtf file was corrected using the information in the sam "ts:" fields, if necessary. The resulting gtf files of genomic exon coordinates from protein and transcript alignments were combined using StringTie⁴¹ and TACO,⁴² and the genomic coordinates of CDS exons were calculated by TransDecoder.⁴³ All resulting gene-models were functionally annotated by eggNOG,²⁸ best protein matches (LAST aligner),³⁴ and BUSCO.²⁵ A single best gene model was chosen from a cluster of gene models according to scoring of its functional annotation or its CDS length (if no functional annotation was assigned).

Some genes coding for polyketide synthases (PKSs) were difficult to annotate, because they are extremely large, are only weakly expressed, and no reference proteins were available. We found that these genes could be reasonably well annotated by GENSCAN²¹ *ab initio* gene prediction, while other gene prediction tools like Augustus failed. Thus, we performed GENSCAN prediction on several *P. parvum* genomes to be able to perform a PKS gene family analysis.

Transmission electron microscopy

Whole-mount preparations of *P. parvum* cells (ODER1 and K-0081) were used for transmission electron microscopy (TEM). The cells were fixed for 1 min in 2% OsO₄ vapor, rinsed in distilled water, and subsequently stained with 2% aqueous uranyl acetate (1–3 min). The cells were then rinsed one more time and air-dried prior to examination using a Philips CM 120 BioTwin electron microscope. Micrographs were edited with Photopea (www.photopea.com) to mask tiny holes in the formvar film coating the TEM grids.

QUANTIFICATION AND STATISTICAL ANALYSIS

Prymnesium parvum phylogeny based on chloroplast DNA

ONT Minlon data obtained from test sequencing *P. parvum* ODER1 DNA extractions was assembled using WTDBG2⁴⁵ and polished using the ONT-long-reads by Flye.²⁹ This yielded a single contig chloroplast genome. Similarly, a chloroplast genome was produced from a sample of *Rebecca* sp. (strain ID 3408, CCAC, University Duisburg-Essen), a haptophyte from the Pavlovaceae family, to serve as an outgroup in the phylogeny. Short-read genome assemblies of other *P. parvum* s.l. genomes from Wisecaver et al.¹⁶ were downloaded from the corresponding FigShare repository and aligned to the ODER1 chloroplast genome by Minimap2.³⁶ Alignments were converted to maf format (paftools.js view -f maf) and screened for ortholog matches using Last-split.³³ Pair-wise maf files were combined into a multiple alignment using MULTIZ.³⁸ The multiple alignment maf file of 17 taxa was converted to aligned fasta format, nucleotide residues with fewer than 15 aligned taxa were removed. A maximum likelihood phylogenetic tree was calculated with IQ-TREE 2.³¹ The best-fit model (K3Pu+F+I+G4) was chosen using the Bayesian inference criterion (BIC) in IQ-TREE 2. Ultra-fast-bootstrap (UFBS)⁵⁰ and Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT)⁵¹ methods were applied to calculate branch support values.

Estimation of DNA content by flow cytometry and ploidy level assignment using genomics

Prymnesium parvum strain K-0081 was obtained from the Norwegian Culture Collection of Algae (NORCCA, Oslo) and propagated in f/2 medium at a salinity of 1%. The DNA content of *P. parvum* ODER1 and its closest relative *P. parvum* K-0081 was estimated using propidium iodide flow cytometry (PI FCM). To do so, 1 mL of well-grown culture was centrifuged (5 min, 2040 g; Eppendorf) and the

superfluous medium was removed by pipetting. The cell pellet was flash-frozen in liquid nitrogen, causing a rupture of cells and release of the nuclei. Next, 350 μL of ice-cold nuclei isolation buffer LB01 (15 mM Tris, 2 mM Na_2EDTA , 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, 0.1% (v/v) Triton X-100; pH = 8.0)⁵² was added. The resulting suspension was thoroughly shaken and kept on ice. Nuclei of two flowering plant species, selected to closely match the DNA content of the investigated sample without overlapping, *Solanum pseudocapsicum* (2C = 2.59 pg)⁵³ and *Carex acutiformis* (2C = 0.82 pg)⁵⁴ were used as internal standards for strain K-0081 and ODER1, respectively. To release the nuclei of the standard, a ca. 20 mg piece of fresh leaf tissue was chopped with a razor blade in a plastic Petri dish with 250 μL of ice-cold LB01 buffer. Both suspensions (algal and standard nuclei) were mixed thoroughly and filtered through a 42 μm nylon mesh into a 3.5 mL cuvette fitting the flow cytometer. Following 20 min incubation at room temperature, a staining solution consisting of 550 μL of LB01 lysis buffer, 50 $\mu\text{g mL}^{-1}$ propidium iodide, 50 $\mu\text{g mL}^{-1}$ RNase IIA and 2 $\mu\text{L mL}^{-1}$ β -mercaptoethanol was added. After 5 min incubation at room temperature, the relative fluorescence of at least 10,000 particles was recorded using a CytoFLEX S cytometer (Beckman Coulter, Indianapolis, IN, USA), equipped with a yellow-green laser (561 nm, 30 mW). Histograms were analyzed using CytExpert 2.4.0.28 software (Beckman Coulter). The DNA content was calculated as sample G_1 peak mean fluorescence/standard G_1 peak mean fluorescence $\times$ standard 2C DNA content.⁵⁵ To minimize random instrumental shift, both strains were analyzed at least three times on separate days and the measurements averaged. To corroborate DNA content differences, simultaneous analysis of strains ODER1 and K-0081 was performed.

We analyzed the allele frequency distribution of variants in K-0081 and ODER1 genomes. The HIFI data of *P. parvum* ODER1 were mapped to the ODER1 haplotype 2 assembly using Minimap2 (-a -x map-hifi), Illumina data for *P. parvum* K-0081 were mapped using parameters for short reads (-a -x sr). SAMtools⁴⁰ served to create sorted bam files of the data. CLAIR3²⁷ was applied to call variants on both datasets. From the resulting vcf files, we calculated the percentage of alternate allelic reads for each variant with total read coverage ≥ 20 and plotted this as a frequency distribution.

QPCR-assay to quantify *Prymnesium parvum* in water samples

Samples were taken monthly from March 2023 until February 2024 and fixed with Lugol solution. As a standard, samples from the Oder catastrophe ($n = 80$) in summer 2022 were counted microscopically (10×100) to calculate the correlation factor between qPCR-amplicon copy number of a DNA marker (Internal Transcribed Spacer 2: ITS2) and *Prymnesium* cell counts. Each 10-mL sample was centrifuged at 4,600 g at 4 °C for 1 h 45 min and then resuspended in 360 μL of lysis buffer ATL (Qiagen). DNA was extracted in duplicate with the QIAamp DNA Mini Kit (Qiagen) followed by cleanup with the OneStep PCR Inhibitor Removal Kit (Zymo Research). QPCR quantification was carried out in duplicate for each replicate with newly designed primers PrymF2239 (5'-CACATCC GATCGTGTCTGC-3') and PrymR2384 (5'-GGCACAACGACTTGGTAGG-3') at 96 °C (3 min), followed by 40 cycles of denaturation at 96 °C (30 s), annealing at 67 °C (30 s), extension at 72 °C (30 sec) in 25- μL reaction volumes (2 U Platinum Taq DNA-polymerase, 4 μL extracted DNA, 1.5 mM MgCl_2 , 9.6 pg μL^{-1} BSA, 400 nM each primer, 0.2 mM dNTPs, 1x SYBR Green I) and calculated from a standard dilution series of a PCR product (quantified with QuantiFluor dsDNA System, Promega with a DeNovix DS-11 spectrometer, Biozym, Hessisch Oldendorf, Germany). Cell equivalents were calculated from the ITS2 copy number ($R^2 = 0.87$).