

**CRYPTIC SPECIATION AND HYBRIDIZATION IN THE TOXIC ALGAL
BLOOM-FORMING SPECIES COMPLEX PRYMNESIUM PARVUM**

by

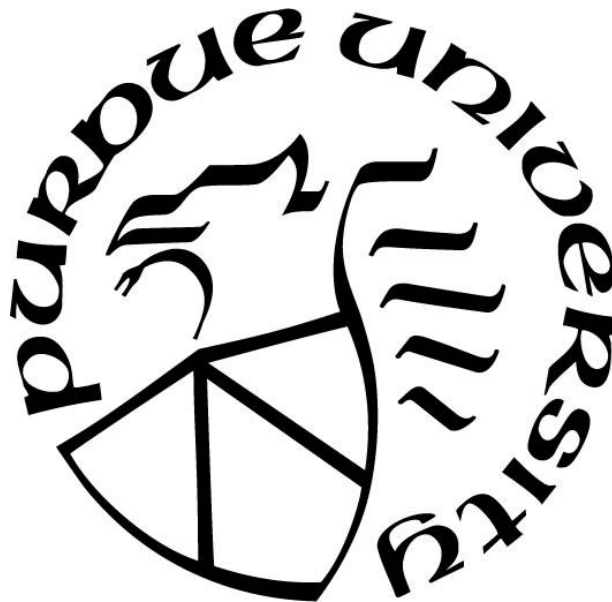
Nathan Watervoort

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Biochemistry

West Lafayette, Indiana

December 2026

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Prof. Jennifer Wisecaver, Chair

Department of Biochemistry

Prof. Brian Dilkes, Member

Department of Biochemistry

Prof. Joshua Widhalm, Member

Department of Horticulture & Landscape Architecture

Prof. Mary Catherine Aime, Member

Department of Botany and Plant Pathology

Approved by:

Dr. Joseph Ogas and Dr. Robert Stahelin

Dedicated to Mamma, Pappa, Tim, Ben, and Elise. I could not have done this without you.

ACKNOWLEDGMENTS

I am very fortunate to have worked with many talented and kind people throughout this thesis. I want to thank my thesis committee: Dr. Brian Dilkes, Dr. Josh Widhalm, and Dr. Cathy Aime. Their insightful comments and unique points of view greatly improved the strength of the science within this work. I also want to thank Dr. Jody Banks for her mentorship and insights throughout this work.

The Wisecaver Lab was also an incredibly supportive environment which I have greatly appreciated. I want to thank all previous Wisecaver Lab members, and in particular I want to thank Timilehin Jeje, Dr. Katelin Burow, Dr. Kate Eastman, Dr. Amanda Pendleton, Dr. Joshua Trujillo, and Dr. Robert Auber for their positive presence and support. Along similar lines, I also want to thank the Center for Plant Biology Trainee Association for fostering a positive and supportive environment.

I have also benefitted from many sources of financial and technical support that supported me throughout my degree. I want to thank Purdue University, Washington State University, and the National Science Foundation for their generous support. I also want to thank the Department of Biochemistry for their support through the Bird Stair Fellowship. This thesis was also supported technically by the Bindley Biosciences Flow Cytometry and Metabolomics cores. I specifically want to thank Dr. Jill Hutchcroft and Dr. Bruce Cooper for their technical support. I was to thank the Purdue Rosen Center for Advanced Computing, the Washington State University Center for Institutional Research Computing, and NSF ACCESS for computing support.

I finally want to thank my amazing research advisor, Dr. Jennifer Wisecaver. Through two institutions and two states, she has been an incredible mentor without which nothing here would be possible. Her laboratory is a fantastic learning environment which pushed me to grow into a better scientist. I have been incredibly fortunate to benefit from her patient guidance throughout my doctoral research, and I am much better off for it.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	4
TABLE OF CONTENTS.....	5
LIST OF TABLES	8
LIST OF FIGURES	9
LIST OF ABBREVIATIONS.....	10
ABSTRACT.....	11
1. INTRODUCTION	12
1.1 The <i>Prymnesium parvum</i> species complex.....	17
1.2 References.....	20
2. COMPARISON OF TWO GENOME ASSEMBLIES FOR <i>PRYMNESIUM PARVUM</i> STRAIN UTEX2797.....	32
2.1 Introduction.....	32
2.2 Methods.....	34
2.2.1 Data sources.....	34
2.2.2 Illumina Sequencing	35
2.2.3 Read processing and k-mer counting.....	35
2.2.4 SNP analysis	35
2.3 Results and Discussion	36
2.4 Conclusion	39
2.5 References.....	40
3. HYBRIDIZATION BETWEEN TOXIC BLOOM-FORMING ALGAE OF THE <i>PRYMNESIUM PARVUM SENSU LATO</i> SPECIES COMPLEX.....	44
3.1 Abstract.....	44
3.2 Introduction.....	44
3.3 Results.....	47
3.3.1 Variation in heterozygosity and ploidy in <i>P. parvum</i> clade A	48
3.3.2 Phylogenetic analyses identify a new A3 lineage and additional A1xA2 hybrids	51
3.3.3 A1xA2 hybrids contain either A1- or A2-like chloroplasts and display complex evolutionary relationships.....	53

3.3.4 Chromosomal and segmental anomalies accumulate in laboratory hybrid strains	55
3.3.5 A1x A1 hybrids contain more genome anomalies than non-hybrid strains.....	58
3.4 Discussion	60
3.5 Materials and Methods.....	64
3.5.1 Taxon Sampling.....	64
3.5.2 Algal Culturing	65
3.5.3 Single-cell reisolation	66
3.5.4 DNA sequencing, read filtering, and read mapping	66
3.5.5 Preliminary SNP phylogenetic network	67
3.5.6 Genome size estimation via flow cytometry	68
3.5.7 Estimation of Heterozygosity	68
3.5.8 <i>P. parvum</i> clade A phylogenetic analysis.....	69
3.5.9 Identification of Ancestry Informative Markeres	70
3.5.10 Chloroplast SNP analyses.....	70
3.5.11 Structural variant identification	71
3.6 References.....	72
4. HARMFUL ALGAL BLOOMS OF <i>PRYMNESIUM PARVUM SENSU LATO</i> IN TEXAS CONSIST OF MULTIPLE CRYPTIC SPECIES	79
4.1 Abstract	79
4.2 Introduction.....	79
4.3 Methods.....	81
4.3.1 Sampling of recent Texas blooms.....	81
4.3.2 Isolation and maintenance of algal cultures.....	83
4.3.3 Flow cytometry for DNA content estimation	83
4.3.4 DNA sequencing.....	84
4.3.5 Read filtering	85
4.3.6 SNP calling and network creation	86
4.3.7 Chromosomal anomalies	87
4.3.8 Chemotyping.....	87
4.4 Results.....	88
4.4.1 Texas <i>Prymnesium parvum</i> s.l. blooms vary in cell density and toxicity.....	88

4.4.2 A majority of strains from 2018-2020 Texas blooms are B-types	91
4.4.3 Multiple cryptic species can be found in the same bloom.....	94
4.4.4 Recent Texas isolates vary in prymnesin production	95
4.4.5 American B-types have chromosomal anomalies.....	96
4.5 Discussion	97
4.5.1 Two cryptic species of <i>P. parvum</i> s.l. are present in Texas blooms.....	97
4.5.2 Multiple cryptic species are found in the same blooms.....	98
4.5.3 B-type <i>P. parvum</i> s.l. accumulate chromosomal anomalies.....	99
4.6 References.....	99
5. CONCLUSION.....	105
5.1 References.....	111
APPENDIX A. SUPPLEMENTAL FIGURE	115
APPENDIX B. SUPPLEMENTAL TABLES	132
PUBLICATIONS.....	150
VITA.....	151

LIST OF TABLES

Table 2.1 Statistics from all A-type <i>P. parvum</i> s.l. long-read assemblies	34
Table 3.1 Description of <i>P. parvum</i> clade A strains.	50
Table 3.2 List of structural anomalies in single-cell reisolates compared to progenitor strains...	57
Table 4.1 Water sample metadata.	82

LIST OF FIGURES

Figure 1.1 Recognized microbial cryptic species complexes across the tree of life. Tree topology is from Burki et al., (2020), branch lengths are not informative.	16
Figure 1.2 A) Prymnesin structure of prymnesin-1, an A-type prymnesin. B) Light micrograph of an A1 strain of <i>Prymnesium parvum</i> s.l.: UTEX3227.	18
Figure 1.3 Model phylogeny of <i>Prymnesium parvum</i> s.l. based on works published before this thesis.	19
Figure 2.1. Model of evolutionary relationships between <i>P. parvum</i> s.l. strains present in this analysis based on phylogenetic and genomic analyses performed by Wisecaver et al. (2023).	33
Figure 2.2. K-mer frequency plots for CCMP 3037 and UTEX 2797. Data from Jian et al. (A,B); from Wisecaver et al. (C,D); and resequenced for this study (E,F).	37
Figure 2.3. SNP analysis of <i>P. parvum</i> s.l. type A strains. A) PCA analysis B) Phylogenetic network. LR, long read; SR, short read.	38
Figure 3.1 Heterozygosity and genome size variation among <i>P. parvum</i> clade A strains.	48
Figure 3.2 Phylogenetic analysis of <i>P. parvum</i> clade A strains.	52
Figure 3.3 Genetic variation in A1xA2 hybrids.	54
Figure 3.4 Scaffold 8 structural anomalies in UTEX3226 and single-cell reisolates.	56
Figure 3.5 Structural anomalies detected in A1, A2, and A1xA2 strains.	58
Figure 3.6 Proposed model of <i>P. parvum</i> clade A lifecycle and hybridization.	61
Figure 4.1 Location of the eight sampled bodies of water.	89
Figure 4.2 Bifurcating phylogenetic tree of <i>P. parvum</i> s.l. strains.	90
Figure 4.3 Phylogenetic network of <i>P. parvum</i> s.l. strains.	92
Figure 4.4 Flow cytometry estimates of genome sizes.	94
Figure 4.5 Chemical analysis of recent Texas strains.	95
Figure 4.6 Chromosomal anomalies in Texas B-type strains.	96
Figure 5.1 Model phylogeny of the <i>P. parvum</i> s.l. species complex incorporating results from this thesis (*). CCMP1926/1927 have not been chemotyped and may be a new lineage of <i>P. parvum</i> s.l. entirely.	109

LIST OF ABBREVIATIONS

AIM	Ancestry Informative Marker
DNBseq	DNA Nanoball Sequencing
ONT	Oxford Nanopore
s.l.	<i>Sensu lato</i> /in a broad sense

ABSTRACT

Historically, the dominant way to determine species identity was with morphology. However, recent advancements in DNA sequencing have made it possible to genetically identify species. This frequently leads to clashes between morphological species concepts and genetic species concepts called cryptic species complexes, wherein one morphospecies may contain several genetically defined species. These cryptic species complexes are becoming more widely recognized and are particularly problematic in microbial eukaryotes. However, recognition of microbial cryptic species complexes in microbial eukaryotes lags behind their discovery. One microbial eukaryote cryptic species complex is the toxic haptophyte alga *Prymnesium parvum sensu lato* (s.l.). *P. parvum* s.l. is composed of at least three genetically distinct but morphologically identical species. Each species of *P. parvum* s.l. forms toxic algal blooms, produces different forms of the potent ichthyotoxin prymnesin, have different nucleotide content, and have different genome sizes. The *P. parvum* species complex was recently discovered, but failing to acknowledge its existence hinders experimental efforts and leads to significant confusion in the field. Here, I show that failing to account for different *P. parvum* s.l. cryptic species limits experimental success. I demonstrate different cryptic species of *P. parvum* s.l. can hybridize and generate unstable genomes. I also found that the genetic diversity of the *P. parvum* species complex is reflected in local populations and can be found in the same bloom. Taken together, *P. parvum* s.l. is a useful case study to demonstrate the importance of recognizing cryptic species diversity for experimental success and to fully delineate a species complex in microbial eukaryotes.

1. INTRODUCTION

Species are one of the fundamental units of evolution between which comparisons are made (De Queiroz 2005). Although species are incredibly important for making biological comparisons, delineating a species is difficult and can be done with several frameworks, called species concepts. One of the most common species concepts is the biological species concept, which defines species as groups of organisms that are capable of reproducing with one another and that do not exchange genetic information with other species.

However, the clear line between species, as defined by the biological species concept, frequently does not survive contact with biological reality. Gene flow between different described species is not uncommon. Hybridization between species is a known phenomenon and an important speciation mechanism in many eukaryotes, including plants and animals (Hermansen et al. 2011; Prysacz et al. 2015). Recurrent hybridization can lead to the integration of genes from one species into the genome of another through a process called introgression, which has been very important in the evolution of many eukaryotes such as *Triticum aestivum* (bread wheat) (Levy and Feldman 2022). Horizontal gene transfer of individual genes between species has also conferred advantageous traits to many species, such as the ability to scavenge host ATP in the parasites *Chlamydia* and *Rickettsia* (Winkler and Neuhaus 1999; Koonin et al. 2001). Nevertheless, the biological species concept remains the most widely used definition of species in part because it is broadly applicable and based on a biological process (Bobay and Ochman 2017; Stankowski and Ravinet 2021).

In addition to these leaky mating barriers leading to gene flow across species, delimiting biological species can be difficult due to myriad other processes. Complete life-cycle descriptions, including sexual stages and what triggers them, can be rare (Rizos et al. 2024). Although many species with poorly described life cycles may have unknown sexual stages to their life cycles, all prokaryotes and a minority (~5%) of eukaryotes are asexual (Schön et al. 2008). Asexual species may be classifiable under the biological species concept if the barriers to horizontal gene transfer can be considered a mating barrier, but this is an area of ongoing debate (Bobay and Ochman 2017; Hörandl 2018).

Given the limitations of the biological species concept, researchers studying different lineages of life have attempted to delineate species using different means. Large multicellular eukaryotes that are visible with the naked eye are called “macrobial eukaryotes” and are frequently defined based on their morphology. Morphological classifications are quick and cheap, and thus very accessible (Dunn 2003). Modern applications like iNaturalist have also made morphological classifications even more accessible (Unger et al. 2021; Mason et al. 2025). Historically, most available data for phylogenetic analyses were also morphological (Simpson 1951). However, morphological classifications frequently encompass several genetically distinct populations (M. Zhao et al. 2016; Yan et al. 2018; Zhong et al. 2020). Morphological characteristics can also converge between otherwise divergent groups, making evolutionarily divergent species appear similar (Leander 2008; Lindgren et al. 2012; Huang et al. 2016).

In the genomic era, many modern studies on macrobial eukaryote speciation incorporate genetic information to delineate species. Approximately 90% of species delimitation studies from 2013 to 2023 used some form of conserved genetic marker to separate species (Singhal et al. 2025). Although the exact marker varies between major lineages, popular options include portions of the ribosomal RNA operon, the rubisco large subunit for photosynthetic organisms, and the cytochrome c oxidase subunit 1 in metazoans (Antil et al. 2023). However, marker-based delineations use limited amounts of sequence data, which may not provide enough phylogenetic signal. Individual barcodes recapitulate known species relationships, as defined by the author, only 70 to 77% of the time (Meier et al. 2006; Elias et al. 2007; Dupuis et al. 2012). Part of the reason for this is that markers can evolve at different, often slower, rates than the rest of the genome and result in little-to-no phylogenetic signal between otherwise distinct species (Dupuis et al. 2012). Incorporating multiple markers significantly increases the chance of successfully identifying species, but can be significantly more resource and labor intensive (Dupuis et al. 2012). Whole-genome sequencing is a promising alternative to marker-based species-delineation because it incorporates every nuclear marker and can delineate lineages that individual markers do not have the resolution to separate (Galen, Nunes, et al. 2018; Church et al. 2025). Similarly, organellar sequencing can be used to resolve species in some groups, but due to difference in inheritance patterns, may not have the same evolutionary history as the nuclear genome (Galen, Nunes, et al. 2018). Despite the benefit of whole-genome sequencing compared to marker-based approaches,

the cost of collecting genomic data and maintaining the computational infrastructure to process it has limited its implementation (Singhal et al. 2025).

Researchers have employed similar techniques to delineate species in prokaryotes using phenotypic and genetic differences (Varghese et al. 2015). Originally, prokaryotic species were defined using phenotypic (e.g., morphology, pathogenicity, ecological roles, in certain cases disease symptoms) and chemotaxonomic (e.g., lipid composition, GC%, isozymes) information (Stackebrandt 2002). But since the 1960s, genetic approaches have been integrated alongside phenotypic and chemotaxonomic information to create an integrated polyphasic approach (Vandamme et al. 1996; Mende et al. 2013). An example of a prokaryotic species delineated using a polyphasic approach is the bacterium *Anaerococcus senegalensis*. Lagier et al. (2012) first suggested that *A. senegalensis* may have been a novel species due to it having a low 16S sequence similarity ($\leq 97.8\%$) to other characterized *Anaerococcus* species. Lagier et al. (2012) further described *A. senegalensis* using genomic (GC%, gene content), phenotypic (enzymatic activity), and chemotaxonomic (total protein composition) characteristics, which supported characterizing *A. senegalensis* as a unique species. However, there is no gold-standard method to integrate these different lines of evidence for all prokaryotes, and exact tests or metrics differ between prokaryotic species (Ramasamy et al. 2014).

Eukaryotes that are not visible with the naked eye are called microbial eukaryotes. Microbial eukaryotes are a polyphyletic group and encompass a massive diversity of lineages (Burki et al. 2020). Like prokaryotes, microbial eukaryotes are microbial, and can have massive population sizes and cosmopolitan distributions (Finlay 2002). At the same time, microbial eukaryotes frequently (but not always) reproduce sexually and have larger genome sizes often comparable to those of macrobial eukaryotes (Ruvindy et al. 2023; Yadav et al. 2023). Microbial eukaryotes are also ecologically important and require rapid species identification as much as any other lineage of life (Brooks et al. 2016; Anderson et al. 2021; Sobieraj and Metelski 2023).

Traditionally, microbial eukaryotes were described via morphology, like macrobial eukaryotes (Caron et al. 2009). This is especially true for the majority of microbial eukaryotes that cannot be cultured (Caron et al. 2009). However, genetic approaches have also become important in defining microbial eukaryote species. DNA barcode-based analyses are popular, but difficult in microbial

eukaryotes because any universal barcode would have to be conserved across all eukaryotes. Although efforts are still underway to find such a barcode, there is not currently an accepted universal microbial eukaryote barcode (Pawlowski et al. 2012). Instead, the Consortium for the Barcoding of Life recommends a two-step approach for microbial eukaryotes using the V4 region of the 18S rDNA gene followed by group-specific barcodes (Pawlowski et al. 2012; Antil et al. 2023). However, this two-step approach does not extend equally to all microbial eukaryotes, as some groups such as ciliates have no secondary barcode that encompasses the entire group (Y. Zhao et al. 2016). Recent advancements in whole-genome sequencing are promising for microbial eukaryote species identification and have revealed a wealth of new taxa (Caron and Hu 2019). However, whole-genome sequencing is more challenging in microbial eukaryotes than in prokaryotes due to microbial eukaryotes having larger and more repetitive genomes (McGrath and Katz 2004).

As researchers define species using different species concepts, species definitions made using different species concepts can frequently disagree. Morphology does not necessarily change according to genetic distance and may not faithfully recapture evolutionary relationships. Cases where a species defined under the morphological species concept obscures species that could be delineated under alternative species concept (such as genetic or biological species concepts) are called cryptic species complexes. Prominent examples of cryptic species complexes are found in a wide range of organisms. These examples include *Anopheles* mosquitoes (Zhong et al. 2020), *Andrias* salamanders (Yan et al. 2018), and *Laspinema* cyanobacteria (Dvořák et al. 2024).

Cryptic species complexes are particularly problematic in microbial species, including microbial eukaryotes. Compared to macrobial eukaryotes, morphology in microbial eukaryotes is more difficult to see, has fewer diagnostic characteristics, and changes less rapidly (Taylor et al. 2006). Many studies also incorporate ultrastructural morphology, which requires electron microscopy to view (Green et al. 1982; Amato et al. 2007; Amato and Montresor 2008). Although incorporating ultrastructural features into morphological classifications has yielded more diagnostic characteristics, it also exacerbates the difficulty in observing morphological features. Combined, these factors mean that fewer species of microbial eukaryotes are morphologically classifiable.

However, less than 6% of currently described cryptic species complexes are microbial eukaryotes (Struck et al. 2018).

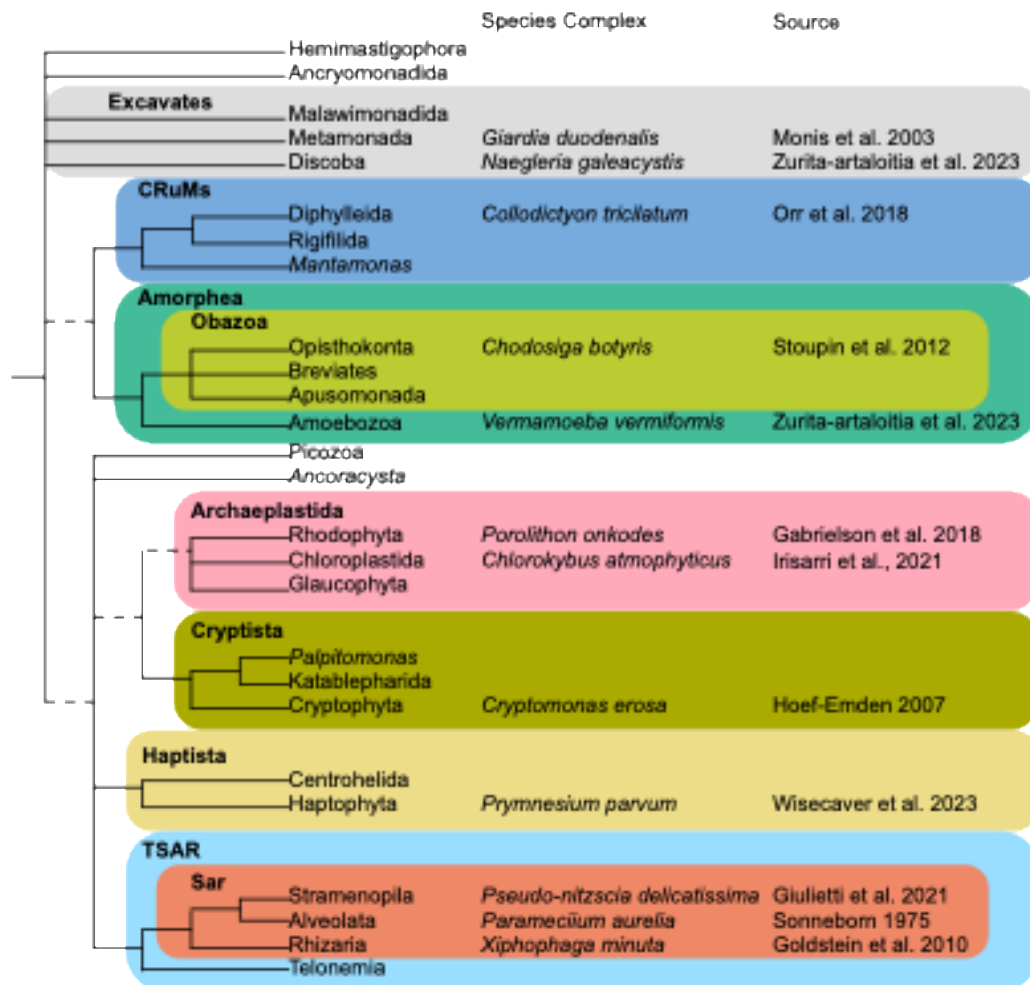


Figure 1.1 Recognized microbial cryptic species complexes across the tree of life. Tree topology is from Burki et al., (2020), branch lengths are not informative.

Despite the sparsity of described microbial cryptic species complexes, recognized microbial eukaryote cryptic species complexes are found across the eukaryotic tree of life (Table B.11, Figure 1.1) (Caron and Hu 2019). Every toxic eukaryotic clade of the UNESCO Taxonomic Reference List of Harmful Microalgae has cryptic species representatives (Lilly et al. 2007; Riisberg and Edvardsen 2008; Lundholm et al. 2009; Quijano-Scheggia et al. 2009; Klöpper et al. 2013; John et al. 2014; Galen, Borner, et al. 2018; Wisecaver et al. 2023). Microbial plant pathogens so frequently form cryptic species complexes that Cros and Groenewald (2005) stated,

“show me a plant pathogen and I will show you a species complex.” Microbial fungi are also frequently cryptic species complexes, and several prominent examples can be found across diverse lineages of fungi (Lockhart et al. 2008; Rouxel et al. 2013; Alastruey-Izquierdo et al. 2014; Fernandez-Pittol et al. 2022; Fernandez-Pittol et al. 2022; O’Donnell et al. 2022).

The rate of microbial eukaryote cryptic species discovery is also increasing as the cost to sequence DNA plummets (Caron and Hu 2019). However, the traditional belief that microbial eukaryotes are less diverse remains (Finlay and Clarke 1999; Fenchel 2005; Caron and Hu 2019). This gap is concerning as working with cryptic species both requires careful laboratory handling but also allows for investigations into unique biological phenomena. There is a need to develop a framework to emphasize cryptic species diversity before formal species descriptions to preserve the integrity of the literature surrounding individual complexes and allow more complete descriptions of the complex.

1.1 The *Prymnesium parvum* species complex

A prime example of a recently discovered microbial eukaryotic cryptic species complex is *Prymnesium parvum sensu lato* (s.l., in the broad sense) (Figure 1.2B). *P. parvum* s.l. are haptophyte algae known for forming highly damaging toxic blooms (Southard et al. 2010; Roelke et al. 2016; Binzer et al. 2019; Wisecaver et al. 2023; Watervoort et al. 2026). In some cases, *P. parvum* s.l. blooms will kill a majority of fish and mollusks in a body of water, devastating local ecosystems and causing millions of dollars of damage (Free et al. 2023; Sobieraj and Metelski 2023). *P. parvum* s.l. blooms have gotten worse over time and are difficult to predict (Grover et al. 2010; Tábora-Sarmiento et al. 2022). The exact reasons why *P. parvum* s.l. blooms are difficult to predict are unclear, but may include interspecific competition and strain-level temperature responses (Grover et al. 2010). Generalized *P. parvum* s.l. growth models predict that *P. parvum* s.l. will expand in range as many bodies of water appear suitable for *P. parvum* s.l. growth but have not yet reported any blooms. These vulnerable but currently unaffected areas include reservoirs in western Europe and southeastern Australia (Macêdo et al. 2023).

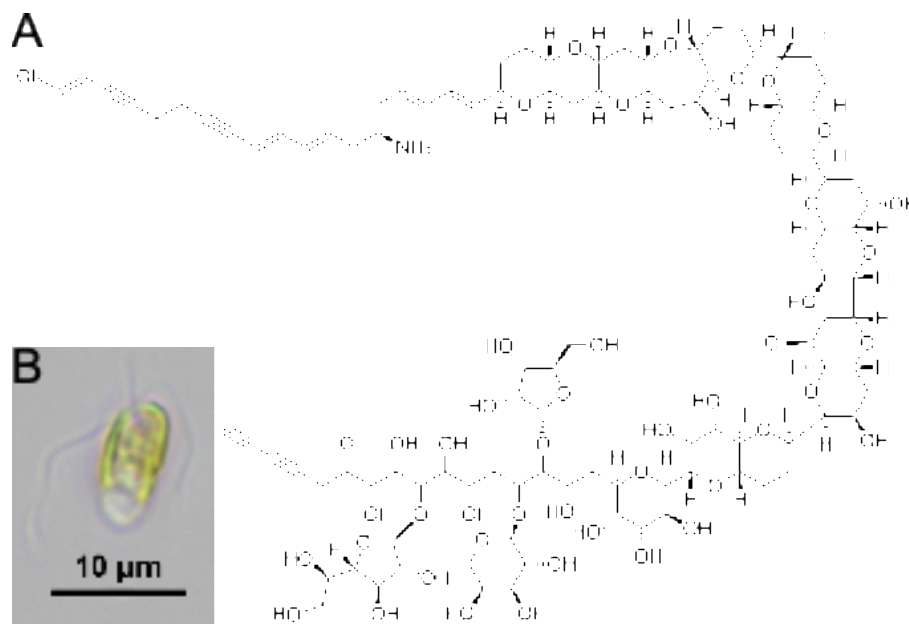


Figure 1.2 A) Prymnesin structure of prymnesin-1, an A-type prymnesin. B) Light micrograph of an A1 strain of *Prymnesium parvum* s.l.: UTEX3227.

P. parvum s.l. was first described on the Isle of Wight in the United Kingdom by Carter in (1937). Earlier fish kills in the Netherlands (Liebert and Deerns 1920) and Germany (Lenz 1933) were also attributed to *P. parvum* s.l. after the fact. Since then, *P. parvum* has been isolated from every continent except Antarctica (Farrow 1969; Larsen and Bryant 1998; Southard et al. 2010). In the United States, James and De La Cruz were the first to document a *P. parvum* s.l. bloom in the Pecos River in 1985 that killed over 100,000 fish and millions of mollusks (Linam et al. 1991; Southard et al. 2010). *P. parvum* s.l. has now been detected in 25 different US states, including Hawaii, with blooms recorded in 21 states (Patiño et al. 2023).

The toxicity of *P. parvum* s.l. largely comes from its massive polyketide toxin prymnesin (Figure 1.2A; Igarashi et al. 1996; Igarashi et al. 1999; Manning and La Claire 2010). Prymnesins are a group of lytic compounds that likely function by forming pores in cell membranes, increasing cell permeability and (in the case of fish) leading to respiratory failure (Ulitzur and Shilo 1966; Svendsen et al. 2018; Varga et al. 2024). There are three classes of prymnesins, which are designated A-, B-, and C-types (Rasmussen et al. 2016; Binzer et al. 2019). Extremely large polyketide synthases (PKZILLAs) are involved in the production of prymnesins, and are among the largest genes currently known (Fallon et al. 2024). So far, the PKZILLAs involved in producing

A- and B-type prymnesins, and homologous gene candidates responsible for producing C-type prymnesins, have been identified (Fallon et al. 2024; Kuhl et al. 2024).

Binzer et al. (2019), using an internal transcribed spacer phylogeny of 26 strains, showed that prymnesins are produced in a clade-specific manner, with A-, B-, and C-type prymnesins produced by strains in A-, B-, and C-type clades. Wisecaver et al. (2023) expanded upon this analysis with genome level data and found that A-, B-, and C-type clades have very different genome sizes and nucleotide compositions. The combined results from Binzer et al. (2019) and Wisecaver et al. (2023) indicate that A-, B-, and C-types are separate cryptic species (Figure 1.3). Moreover, A- and C-type *P. parvum* can be separated into multiple subclades, which may also represent separate species (Figure 3). C-types can be separated into three clades (C1, C2, and C3) which have elevated rates of synonymous substitutions compared to within-clade comparisons of A- and B- types. The C-type subclades have different haploid genome sizes ranging from 151 to 235 Mbp, further indicating that these subclades likely represent separate cryptic species (Wisecaver et al. 2023).

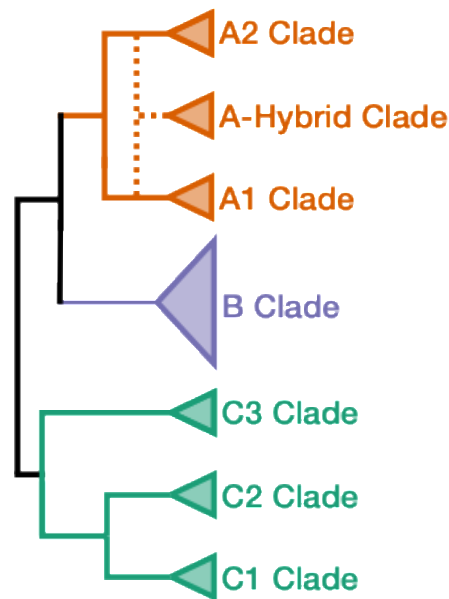


Figure 1.3 Model phylogeny of *Prymnesium parvum* s.l. based on works published before this thesis.

Similar to the C-type subclades, A-type strains can be separated into discrete subclades designated A1 and A2, although the nucleotide divergence between A-type subclades is not as great as

between C-type subclades (Wisecaver et al. 2023). Wisecaver et al. (2023) found that A1 and A2 strains could also form A-hybrid strains. The strain UTEX2797 appeared to be a hybrid because roughly half of its single-copy genes grouped within the A1 clade and the other half grouped within the A2 clade. The scaffolds of UTEX2797 also formed syntenic pairs between one A1-like scaffold and one A2-like scaffold. Another strain (UTEX3226/UTEX3226) had high levels of heterozygosity, suggesting that it may also be a hybrid strain (Wisecaver et al. 2023). However, a second study of UTEX2797 found no evidence of hybridization and asserted that UTEX2797 belonged to the A1 clade (Jian et al. 2024).

Cryptic speciation in *P. parvum* s.l. requires careful handling to prevent strain mix-ups and attribute results to the proper species. In Chapter 2, I present evidence that Jian et al. (2024) inadvertently misidentified UTEX2797 as an A1 strain. Strain mix ups are easy in *P. parvum* s.l. due to the cryptic nature of the species complex. However, cryptic speciation also allows for unique biological phenomena. In Chapter 3, I show that hybridization is not unique to UTEX2797, and that A-hybrid *P. parvum* s.l. are frequently isolated from North American waters. A-hybrid *P. parvum* s.l. rapidly accumulate chromosomal anomalies. In Chapter 4, I find that A-type and B-type *P. parvum* s.l. can be found in the same blooms. I also find that B-type *P. parvum* accumulate chromosomal anomalies similar to those observed in A-types, which points to the importance of structural variation in *P. parvum* s.l. genome evolution.

1.2 References

- Alastruey-Izquierdo A, Alcazar-Fuoli L, Cuenca-Estrella M. 2014. Antifungal Susceptibility Profile of Cryptic Species of *Aspergillus*. *Mycopathologia*. 178(5–6):427–433. <https://doi.org/10.1007/s11046-014-9775-z>
- Amato A et al. 2007. Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms. *Protist*. 158(2):193–207. <https://doi.org/10.1016/j.protis.2006.10.001>
- Amato A, Montresor M. 2008. Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. *Phycologia*. 47(5):487–497. <https://doi.org/10.2216/07-92.1>
- Anderson DM et al. 2021. Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae*. 102:101975. <https://doi.org/10.1016/j.hal.2021.101975>

- Antil S et al. 2023. DNA barcoding, an effective tool for species identification: a review. *Mol Biol Rep.* 50(1):761–775. <https://doi.org/10.1007/s11033-022-08015-7>
- Bendif EM et al. 2014. Genetic delineation between and within the widespread coccolithophore morpho-species *Emiliania huxleyi* and *Gephyrocapsa oceanica* (Haptophyta). *Journal of Phycology.* 50(1):140–148. <https://doi.org/10.1111/jpy.12147>
- Binzer SB et al. 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>
- Blake DP et al. 2021. Genetic and biological characterisation of three cryptic *Eimeria* operational taxonomic units that infect chickens (*Gallus gallus domesticus*). *International Journal for Parasitology.* 51(8):621–634. <https://doi.org/10.1016/j.ijpara.2020.12.004>
- Bobay L-M, Ochman H. 2017. Biological Species Are Universal across Life’s Domains. *Genome Biology and Evolution.* 9(3):491–501. <https://doi.org/10.1093/gbe/evx026>
- Brooks BW et al. 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Enviro Toxic and Chemistry.* 35(1):6–13. <https://doi.org/10.1002/etc.3220>
- Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The New Tree of Eukaryotes. *Trends in Ecology & Evolution.* 35(1):43–55. <https://doi.org/10.1016/j.tree.2019.08.008>
- Cai R et al. 2020. Cryptic species in the parasitic *Amoebophrya* species complex revealed by a polyphasic approach. *Sci Rep.* 10(1):2531. <https://doi.org/10.1038/s41598-020-59524-z>
- Caron DA et al. 2009. Protists are microbes too: a perspective. *The ISME Journal.* 3(1):4–12. <https://doi.org/10.1038/ismej.2008.101>
- Caron DA, Hu SK. 2019. Are We Overestimating Protistan Diversity in Nature? *Trends in Microbiology.* 27(3):197–205. <https://doi.org/10.1016/j.tim.2018.10.009>
- Carter N. 1937. New or interesting algae from brackish water. *Archiv für Protistenkunde.* 90:1–68. <https://cir.nii.ac.jp/crid/1573668926043051648>
- Church SH et al. 2025. Population genomics of a sailing siphonophore reveals genetic structure in the open ocean. *Current Biology.* 35(15):3556–3569.e6. <https://doi.org/10.1016/j.cub.2025.05.066>
- De Luca D, Piredda R, Sarno D, Kooistra WHCF. 2021. Resolving cryptic species complexes in marine protists: phylogenetic haplotype networks meet global DNA metabarcoding datasets. *The ISME Journal.* 15(7):1931–1942. <https://doi.org/10.1038/s41396-021-00895-0>
- De Queiroz K. 2005. Ernst Mayr and the modern concept of species. *Proc Natl Acad Sci USA.* 102(suppl_1):6600–6607. <https://doi.org/10.1073/pnas.0502030102>

- Degerlund M et al. 2012. Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *Journal of Plankton Research*. 34(5):416–431. <https://doi.org/10.1093/plankt/fbs004>
- Díaz-Tapia P, Ly M, Verbruggen H. 2020. Extensive cryptic diversity in the widely distributed *Polysiphonia scopulorum* (Rhodomelaceae, Rhodophyta): Molecular species delimitation and morphometric analyses. *Molecular Phylogenetics and Evolution*. 152:106909. <https://doi.org/10.1016/j.ympev.2020.106909>
- Dudiuk C, Theill L, Gamarra S, Garcia-Effron G. 2017. Detection of Cryptic *Candida* Species Recognized as Human Pathogens Through Molecular Biology Techniques. *Curr Fungal Infect Rep*. 11(4):176–183. <https://doi.org/10.1007/s12281-017-0294-5>
- Dunn CP. 2003. Keeping taxonomy based in morphology. *Trends in Ecology & Evolution*. 18(6):270–271. [https://doi.org/10.1016/S0169-5347\(03\)00094-6](https://doi.org/10.1016/S0169-5347(03)00094-6)
- Dupuis JR, Roe AD, Sperling FAH. 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*. 21(18):4422–4436. <https://doi.org/10.1111/j.1365-294X.2012.05642.x>
- Dvořák P et al. 2024. Population genomics resolves cryptic species of the ecologically flexible genus *Laspinema* (Cyanobacteria). *Journal of Phycology*. 60(4):871–885. <https://doi.org/10.1111/jpy.13475>
- Elias M et al. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proc R Soc B*. 274(1627):2881–2889. <https://doi.org/10.1098/rspb.2007.1035>
- Fallon TR et al. 2024. Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins. *Science*. 385(6709):671–678. <https://doi.org/10.1126/science.ado3290>
- Farrow GA. 1969. Note on the association of *Prymnesium* with fish mortalities. *Water Research*. 3(5):375–379. [https://doi.org/10.1016/0043-1354\(69\)90090-6](https://doi.org/10.1016/0043-1354(69)90090-6)
- Fenchel T. 2005. Cosmopolitan microbes and their cryptic species. *Aquat Microb Ecol*. 41:49–54. <https://doi.org/10.3354/ame041049>
- Fernandez-Pittol M et al. 2022. Aspergillosis by cryptic *Aspergillus* species: A case series and review of the literature. *Revista Iberoamericana de Micología*. 39(2):44–49. <https://doi.org/10.1016/j.riam.2022.04.002>
- Finlay BJ. 2002. Global Dispersal of Free-Living Microbial Eukaryote Species. *Science*. 296(5570):1061–1063. <https://doi.org/10.1126/science.1070710>
- Finlay BJ, Clarke KJ. 1999. Ubiquitous dispersal of microbial species. *Nature*. 400(6747):828–828. <https://doi.org/10.1038/23616>

- Free G et al. 2023. An EU analysis of the ecological disaster in the Oder River of 2022: lessons learned and research based recommendations to avoid future ecological damage in EU rivers, a joint analysis from DG ENV, JRC and the EEA. Publications Office of the European Union. [accessed 2024 Oct 21]. <https://data.europa.eu/doi/10.2760/067386>
- Gabrielson PW, Hughey JR, Diaz-Pulido G. 2018. Genomics reveals abundant speciation in the coral reef building alga *Porolithon onkodes* (Corallinales, Rhodophyta). *Journal of Phycology*. 54(4):429–434. <https://doi.org/10.1111/jpy.12761>
- Galen SC, Borner J, et al. 2018. The polyphyly of *Plasmodium* : comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R Soc open sci*. 5(5):171780. <https://doi.org/10.1098/rsos.171780>
- Galen SC, Nunes R, Sweet PR, Perkins SL. 2018. Integrating coalescent species delimitation with analysis of host specificity reveals extensive cryptic diversity despite minimal mitochondrial divergence in the malaria parasite genus *Leucocytozoon*. *BMC Evol Biol*. 18(1):128. <https://doi.org/10.1186/s12862-018-1242-x>
- Gilbert D, Mitchell E, Bourdier G, Francez A-J. 2003. Population dynamics and food preferences of the testate amoeba *Nebela tinctoria* major-bohemica-collaris complex (Protozoa) in a Sphagnum peatland. *Acta protozoologica*, Nencki Institute of Experimental Biology, 2003/42/2/99-104. <https://hal.inrae.fr/hal-02677116v1>
- Giulietti S et al. 2021. Integrative taxonomy of the *Pseudo-nitzschia* (Bacillariophyceae) populations in the NW Adriatic Sea, with a focus on a novel cryptic species in the *P. delicatissima* species complex. *Phycologia*. 60(3):247–264. <https://doi.org/10.1080/00318884.2021.1899733>
- Goldstein ST, Habura A, Richardson EA, Bowser SS. 2010. XIPHOPHAGA MINUTA, AND X. ALLOMINUTA, NOV. GEN., NOV. SPP., NEW MONOTHALAMID FORAMINIFERA FROM COASTAL GEORGIA (USA): CRYPTIC SPECIES, GAMETOGENESIS, AND AN UNUSUAL FORM OF CHLOROPLAST SEQUESTRATION. *The Journal of Foraminiferal Research*. 40(1):3–15. <https://doi.org/10.2113/gsjfr.40.1.3>
- Gómez A, Serra M, Carvalho GR, Lunt DH. 2002. SPECIATION IN ANCIENT CRYPTIC SPECIES COMPLEXES: EVIDENCE FROM THE MOLECULAR PHYLOGENY OF *BRACHIONUS Plicatilis* (ROTIFERA). *Evolution*. 56(7):1431–1444. <https://doi.org/10.1111/j.0014-3820.2002.tb01455.x>
- Green JC, Hibberd DJ, Pienaar RN. 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. *British Phycological Journal*. 17(4):363–382. <https://doi.org/10.1080/00071618200650381>
- Grover JP, Baker JW, Roelke DL, Brooks BW. 2010. Current Status of Mathematical Models for Population Dynamics of *Prymnesium parvum* in a Texas Reservoir. *JAWRA Journal of the American Water Resources Association*. 46(1):92–107. <https://doi.org/10.1111/j.1752-1688.2009.00393.x>

- Grzebyk D et al. 2022. Insight into the morphology and genetic diversity of the *Chaetoceros tenuissimus* (Bacillariophyta) species complex. *European Journal of Phycology*. 57(4):507–525. <https://doi.org/10.1080/09670262.2022.2029949>
- Hausmann K et al. 2006. Cryptic species in a morphospecies complex of heterotrophic flagellates : the case study of *Caecitellus* spp. *Acta Protozoologica*. 45:415–431. <https://agro.icm.edu.pl/agro/element/bwmeta1.element.agro-article-30c44cb2-39ce-47d8-bbe3-75a4b314d0f0>
- Hermansen JS et al. 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*. 20(18):3812–3822. <https://doi.org/10.1111/j.1365-294X.2011.05183.x>
- Hoef-Emden K. 2007. Revision of the Genus *Cryptomonas* (Cryptophyceae) II: Incongruences between the Classical Morphospecies Concept and Molecular Phylogeny in Smaller Pyrenoid-less Cells. *Phycologia*. 46(4):402–428. <https://doi.org/10.2216/06-83.1>
- Hörandl E. 2018. The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. *TAXON*. 67(6):1066–1081. <https://doi.org/10.12705/676.5>
- Huang C-H et al. 2016. Resolution of Brassicaceae Phylogeny Using Nuclear Genes Uncovers Nested Radiations and Supports Convergent Morphological Evolution. *Mol Biol Evol*. 33(2):394–412. <https://doi.org/10.1093/molbev/msv226>
- Ichimura T. 1981. Mating types and reproductive isolation in *Closterium ehrenbergii meneghini*. *Bot Mag Tokyo*. 94(4):325–334. <https://doi.org/10.1007/BF02493393>
- Igai K et al. 2022. Fine-scale genetic diversity and putative ecotypes of oxymonad protists coinhabiting the hindgut of *Reticulitermes speratus*. *Molecular Ecology*. 31(4):1317–1331. <https://doi.org/10.1111/mec.16309>
- Igarashi T, Satake M, Yasumoto T. 1996. Prymnesin-2: A Potent Ichthyotoxic and Hemolytic Glycoside Isolated from the Red Tide Alga *Prymnesium parvum*. *J Am Chem Soc*. 118(2):479–480. <https://doi.org/10.1021/ja9534112>
- Igarashi T, Satake M, Yasumoto T. 1999. Structures and Partial Stereochemical Assignments for Prymnesin-1 and Prymnesin-2: Potent Hemolytic and Ichthyotoxic Glycosides Isolated from the Red Tide Alga *Prymnesium parvum*. *Journal American Chemical Society*. 121(37):8499–8511. <https://doi.org/10.1021/ja991740e>
- Irisarri I et al. 2021. Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc R Soc B*. 288(1963):20212168. <https://doi.org/10.1098/rspb.2021.2168>
- Jian J 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>

- John U et al. 2014. Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: The introduction of five species with emphasis on molecular-based (rDNA) classification. *Protist*. 165(6):779–804. <https://doi.org/10.1016/j.protis.2014.10.001>
- Klöpfer S et al. 2013. Phylogeny and morphology of a *Chattonella* (Raphidophyceae) species from the Mediterranean Sea: what is *C. subsalsa* ? *European Journal of Phycology*. 48(1):79–92. <https://doi.org/10.1080/09670262.2013.771412>
- Kollár J et al. 2026. A polyphasic taxonomy of the pseudocryptic diatom species complex *Pinnularia acidicola* , including description of new species. *European Journal of Phycology*. 1–25. <https://doi.org/10.1080/09670262.2026.2626682>
- Koonin EV, Makarova KS, Aravind L. 2001. Horizontal Gene Transfer in Prokaryotes: Quantification and Classification. *Annu Rev Microbiol*. 55(1):709–742. <https://doi.org/10.1146/annurev.micro.55.1.709>
- Kordbacheh A, Garbalena G, Walsh EJ. 2017. Population structure and cryptic species in the cosmopolitan rotifer *Euchlanis dilatata*. *Zoological Journal of the Linnean Society*. 181(4):757–777. <https://doi.org/10.1093/zoolinnean/zlx027>
- Kuhl H et al. 2024. The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins. *Current Biology*. 34(16):3698–3706. <https://doi.org/10.1016/j.cub.2024.06.033>
- Lagier J-C et al. 2012. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci*. 6(1):116–125. <https://doi.org/10.4056/sigs.2415480>
- Larsen A, Bryant S. 1998. Growth rate and toxicity of *Prymnesium parvum* and *Prymnesium patelliferum* (haptophyta) in response to changes in salinity, light and temperature. *Sarsia*. 83(5):409–418. <https://doi.org/10.1080/00364827.1998.10413700>
- Leander BS. 2008. Different modes of convergent evolution reflect phylogenetic distances: a reply to Arendt and Reznick. *Trends in Ecology & Evolution*. 23(9):481–482. <https://doi.org/10.1016/j.tree.2008.04.012>
- Lenz F. 1933. Untersuchungen zur Limnologie von Strandseen. *SIL Proceedings*, 1922-2010. 6(1):166–177. <https://doi.org/10.1080/03680770.1933.11898529>
- Levy AA, Feldman M. 2022. Evolution and origin of bread wheat. *The Plant Cell*. 34(7):2549–2567. <https://doi.org/10.1093/plcell/koac130>
- Liebert F, Deerns W. 1920. Onderzoek naar de oorzak van een Vischsterfte in den Polder Workumer Nieuwland, nabij Workum. *Verhandlungen en Rapporten uitgegeven door Rijk instituten voor Visscherijonderzoek*. 1:81–93. https://books.google.com/books/about/Onderzoek_naar_de_oorzaak_van_een_vischs.htm?id=PN6x0AEACAAJ

- Lilly EL, Halanych KM, Anderson DM. 2007. Species boundaries and global biogeography of the *Alexandrium tamarense* complex (Dinophyceae)1. *Journal of Phycology*. 43(6):1329–1338. <https://doi.org/10.1111/j.1529-8817.2007.00420.x>
- Linam G, Ralph J, Glass J. 1991. Toxic blooms, an unusual algae threatens aquatic resources. Chihuahuan Desert Discovery. 28(Winter). <https://tpwd.texas.gov/landwater/water/environconcerns/hab/ga/news/toxicblooms.phtml>
- Lindgren AR, Pankey MS, Hochberg FG, Oakley TH. 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. *BMC Evol Biol*. 12(1):129. <https://doi.org/10.1186/1471-2148-12-129>
- Lockhart SR, Messer SA, Pfaller MA, Diekema DJ. 2008. Geographic Distribution and Antifungal Susceptibility of the Newly Described Species *Candida orthopsilosis* and *Candida metapsilosis* in Comparison to the Closely Related Species *Candida parapsilosis*. *J Clin Microbiol*. 46(8):2659–2664. <https://doi.org/10.1128/JCM.00803-08>
- Lundholm N et al. 2009. IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae. Accessed at <https://www.marinespecies.org/hab> on yyyy-mm-dd. [accessed 2026 Jan 29]. <https://www.marinespecies.org/imis.php?dasid=3124&doiid=362>. <https://doi.org/10.14284/362>
- Luo Z et al. 2017. Cryptic diversity within the harmful dinoflagellate *Akashiwo sanguinea* in coastal Chinese waters is related to differentiated ecological niches. *Harmful Algae*. 66:88–96. <https://doi.org/10.1016/j.hal.2017.05.008>
- Macêdo RL, Haubrock P, Rocha O. 2023. Towards effective management of the marine-origin *Prymnesium parvum* (Haptophyta): a growing concern in freshwater reservoirs? *Harmful Algae*. 102513. <https://doi.org/10.1016/j.hal.2023.102513>
- Manning SR, La Claire JW. 2010. Prymnesins: toxic metabolites of the golden alga, *Prymnesium parvum* Carter (Haptophyta). *Mar Drugs*. 8(3):678–704. <https://doi.org/10.3390/md8030678>
- Mason BM et al. 2025. iNaturalist accelerates biodiversity research. *BioScience*. 75(11):953–965. <https://doi.org/10.1093/biosci/biaf104>
- McGrath CL, Katz LA. 2004. Genome diversity in microbial eukaryotes. *Trends in Ecology & Evolution*. 19(1):32–38. <https://doi.org/10.1016/j.tree.2003.10.007>
- Meier R, Shiyang K, Vaidya G, Ng PKL. 2006. DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success Hedin M, editor. *Systematic Biology*. 55(5):715–728. <https://doi.org/10.1080/10635150600969864>
- Mende DR, Sunagawa S, Zeller G, Bork P. 2013. Accurate and universal delineation of prokaryotic species. *Nat Methods*. 10(9):881–884. <https://doi.org/10.1038/nmeth.2575>

- Monis PT, Andrews RH, Mayrhofer G, Ey PL. 2003. Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. *Infection, Genetics and Evolution*. 3(1):29–38. [https://doi.org/10.1016/S1567-1348\(02\)00149-1](https://doi.org/10.1016/S1567-1348(02)00149-1)
- Morard R et al. 2024. The global genetic diversity of planktonic foraminifera reveals the structure of cryptic speciation in plankton. *Biological Reviews*. 99(4):1218–1241. <https://doi.org/10.1111/brv.13065>
- O'Donnell K et al. 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. *Fungal Genetics and Biology*. 41(6):600–623. <https://doi.org/10.1016/j.fgb.2004.03.003>
- O'Donnell K et al. 2022. DNA Sequence-Based Identification of *Fusarium* : A Work in Progress. *Plant Disease*. 106(6):1597–1609. <https://doi.org/10.1094/PDIS-09-21-2035-SR>
- Oliverio AM, Lahr DJG, Nguyen T, Katz LA. 2014. Cryptic Diversity within Morphospecies of Testate Amoebozoa (Amoebozoa: Arcellinida) in New England Bogs and Fens. *Protist*. 165(2):196–207. <https://doi.org/10.1016/j.protis.2014.02.001>
- Orr RJS et al. 2018. Enigmatic Diphyllatea eukaryotes: culturing and targeted PacBio RS amplicon sequencing reveals a higher order taxonomic diversity and global distribution. *BMC Evol Biol*. 18(1):115. <https://doi.org/10.1186/s12862-018-1224-z>
- Patiño R et al. 2023. Toxic algae in inland waters of the conterminous united states—a review and synthesis. *Water*. 15(15):2808. <https://doi.org/10.3390/w15152808>
- Pawlowski J et al. 2012. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS Biol*. 10(11):e1001419. <https://doi.org/10.1371/journal.pbio.1001419>
- Proctor RH et al. 2004. Discontinuous distribution of fumonisin biosynthetic genes in the *Gibberella fujikuroi* species complex. *Mycological Research*. 108(7):815–822. <https://doi.org/10.1017/S0953756204000577>
- Pryszcz LP et al. 2015. The Genomic Aftermath of Hybridization in the Opportunistic Pathogen *Candida metapsilosis*. *PLOS Genetics*. 11(10):e1005626. <https://doi.org/10.1371/journal.pgen.1005626>
- Quijano-Scheggia SI et al. 2009. Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia*. 48(6):492–509. <https://doi.org/10.2216/08-21.1>
- Ramasamy D et al. 2014. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *International Journal of Systematic and Evolutionary Microbiology*. 64(Pt_2):384–391. <https://doi.org/10.1099/ijs.0.057091-0>

- Rasmussen SA et al. 2016. Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J Nat Prod*. 79(9):2250–2256. <https://doi.org/10.1021/acs.jnatprod.6b00345>
- Read BA et al. 2013. Pan genome of the phytoplankton *Emiliana underpins* its global distribution. *Nature*. 499(7457):209–213. <https://doi.org/10.1038/nature12221>
- Riisberg I, Edvardsen B. 2008. Genetic variation in bloom-forming ichthyotoxic *Pseudochattonella* species (Dictyochophyceae, Heterokonta) using nuclear, mitochondrial and plastid DNA sequence data. *European Journal of Phycology*. 43(4):413–422. <https://doi.org/10.1080/09670260802299602>
- Rizos I, Frada MJ, Bittner L, Not F. 2024. Life cycle strategies in free-living unicellular eukaryotes: Diversity, evolution, and current molecular tools to unravel the private life of microorganisms. *J Eukaryotic Microbiology*. 71(6):e13052. <https://doi.org/10.1111/jeu.13052>
- Roelke DL et al. 2016. A chronicle of a killer alga in the west: ecology, assessment, and management of *Prymnesium parvum* blooms. *Hydrobiologia*. 764(1):29–50. <https://doi.org/10.1007/s10750-015-2273-6>
- Rouxel M et al. 2013. Phylogenetic and experimental evidence for host-specialized cryptic species in a biotrophic oomycete. *New Phytologist*. 197(1):251–263. <https://doi.org/10.1111/nph.12016>
- Ruggiero MV et al. 2015. Diversity and temporal pattern of *Pseudo-nitzschia* species (Bacillariophyceae) through the molecular lens. *Harmful Algae*. 42:15–24. <https://doi.org/10.1016/j.hal.2014.12.001>
- Ruvindy R et al. 2023. Genomic copy number variability at the genus, species and population levels impacts in situ ecological analyses of dinoflagellates and harmful algal blooms. *ISME Communications*. 3(1):70. <https://doi.org/10.1038/s43705-023-00274-0>
- Schön I, Lamatsch DK, Martens K. 2008. Lessons to Learn from Ancient Asexuals. In: Egel R, Lankenau D-H, editors. *Recombination and Meiosis*. Vol. 3. Springer Berlin Heidelberg; p 341–376 (Genome Dynamics and Stability). [accessed 2026 Sept 14]. http://link.springer.com/10.1007/7050_2007_032. https://doi.org/10.1007/7050_2007_032
- Shazib SUA et al. 2025. Phylogeny and species delimitation of ciliates in the genus *Spirostomum* (class Heterotrichea) using single-cell transcriptomes. *BMC Ecol Evo*. 25(1):17. <https://doi.org/10.1186/s12862-025-02353-3>
- Simon EM, Nanney DL, Doerder FP. 2008. The “*Tetrahymena pyriformis*” complex of cryptic species. *Biodivers Conserv*. 17(2):365–380. <https://doi.org/10.1007/s10531-007-9255-6>

- Simon N et al. 2017. Revision of the Genus *Micromonas* Manton et Parke (Chlorophyta, Mamiellophyceae), of the Type Species *M. pusilla* (Butcher) Manton & Parke and of the Species *M. commoda* van Baren, Bachy and Worden and Description of Two New Species Based on the Genetic and Phenotypic Characterization of Cultured Isolates. *Protist*. 168(5):612–635. <https://doi.org/10.1016/j.protis.2017.09.002>
- Simpson GG. 1951. The Species Concept. *Evolution*. 5(4):285. <https://doi.org/10.2307/2405675>
- Singhal S et al. 2025. A Genomic Perspective on Species Delimitation. *Annual Review of Ecology, Evolution, and Systematics*. 56(1):467–489. <https://doi.org/10.1146/annurev-ecolsys-102723-055311>
- Škaloud P, Rindi F. 2013. Ecological Differentiation of Cryptic Species within an Asexual Protist Morphospecies: A Case Study of Filamentous Green Alga *Klebsormidium* (Streptophyta). *J Eukaryotic Microbiology*. 60(4):350–362. <https://doi.org/10.1111/jeu.12040>
- Slapeta J, López-García P, Moreira D. 2006. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol*. 23(1):23–29. <https://doi.org/10.1093/molbev/msj001>
- Sobieraj J, Metelski D. 2023. Insights into toxic *Prymnesium parvum* blooms as a cause of the ecological disaster on the Odra River. *Toxins*. 15(6):403. <https://doi.org/10.3390/toxins15060403>
- Sonneborn TM. 1975. The *Paramecium aurelia* complex of fourteen sibling species. *Transactions of the American Microscopical Society*. 94(2):155–178. <https://doi.org/10.2307/3224977>
- Southard GM, Fries LT, Barkoh A. 2010. *Prymnesium parvum*: The Texas experience. *JAWRA Journal of the American Water Resources Association*. 46(1):14–23. <https://doi.org/10.1111/j.1752-1688.2009.00387.x>
- Stackebrandt E. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*. 52(3):1043–1047. <https://doi.org/10.1099/ijs.0.02360-0>
- Stankowski S, Ravinet M. 2021. Quantifying the use of species concepts. *Current Biology*. 31(9):R428–R429. <https://doi.org/10.1016/j.cub.2021.03.060>
- Stewart IA et al. 2001. Genotypic variability in subarctic Atlantic planktic foraminifera. *Marine Micropaleontology*. 43(1–2):143–153. [https://doi.org/10.1016/S0377-8398\(01\)00024-X](https://doi.org/10.1016/S0377-8398(01)00024-X)
- Stoupin D et al. 2012. Cryptic diversity within the choanoflagellate morphospecies complex *Codosiga botrytis* – Phylogeny and morphology of ancient and modern isolates. *European Journal of Protistology*. 48(4):263–273. <https://doi.org/10.1016/j.ejop.2012.01.004>
- Struck TH et al. 2018. Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution*. 33(3):153–163. <https://doi.org/10.1016/j.tree.2017.11.007>

- Svendsen MBS, Andersen NR, Hansen PJ, Steffensen JF. 2018. Effects of Harmful Algal Blooms on Fish: Insights from *Prymnesium parvum*. *Fishes*. 3(1):11. <https://doi.org/10.3390/fishes3010011>
- Tábor-Sarmiento S, Patiño R, Portillo-Quintero C, Coldren C. 2022. Air, land, and water variables associated with the first appearance and current spatial distribution of toxic *Prymnesium parvum* blooms in reservoirs of the southern great plains, USA. *Science of The Total Environment*. 836:155567. <https://doi.org/10.1016/j.scitotenv.2022.155567>
- Taylor JW et al. 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Phil Trans R Soc B*. 361(1475):1947–1963. <https://doi.org/10.1098/rstb.2006.1923>
- Ulitzur S, Shilo M. 1966. Mode of Action of *Prymnesium parvum* Ichthyotoxin. *The Journal of Protozoology*. 13(2):332–336. <https://doi.org/10.1111/j.1550-7408.1966.tb01915.x>
- Unger S, Rollins M, Tietz A, Dumais H. 2021. iNaturalist as an engaging tool for identifying organisms in outdoor activities. *Journal of Biological Education*. 55(5):537–547. <https://doi.org/10.1080/00219266.2020.1739114>
- Vandamme P et al. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev*. 60(2):407–438. <https://doi.org/10.1128/mr.60.2.407-438.1996>
- Varga E 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>
- Varghese NJ et al. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res*. 43(14):6761–6771. <https://doi.org/10.1093/nar/gkv657>
- Watervoort NF, Jeje T, Dilkes BP, Wisecaver JH. 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum* sensu lato species complex. *bioRxiv*. 2026.06.03.730016. <https://doi.org/10.64898/2026.06.03.730016>
- Winkler HH, Neuhaus HE. 1999. Non-mitochondrial ATP transport. *Trends in Biochemical Sciences*. 24(2):64–68. [https://doi.org/10.1016/S0968-0004\(98\)01334-6](https://doi.org/10.1016/S0968-0004(98)01334-6)
- Wisecaver JH et al. 2023. Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. *Current Biology*. 33(11):2246-2259.e8. <https://doi.org/10.1016/j.cub.2023.05.003>
- Wu Y et al. 2022. Cryptic speciation of benthic *Prorocentrum* (Dinophyceae) species and their potential as ecological indicators. *Journal of Sea Research*. 190:102304. <https://doi.org/10.1016/j.seares.2022.102304>
- Yadav V, Sun S, Heitman J. 2023. On the evolution of variation in sexual reproduction through the prism of eukaryotic microbes. *Proceedings of the National Academy of Sciences*. 120(10):e2219120120. <https://doi.org/10.1073/pnas.2219120120>

- Yan F et al. 2018. The Chinese giant salamander exemplifies the hidden extinction of cryptic species. *Current Biology*. 28(10):R590–R592. <https://doi.org/10.1016/j.cub.2018.04.004>
- Zhao M et al. 2016. The famous cultivated mushroom Bailinggu is a separate species of the *Pleurotus eryngii* species complex. *Sci Rep*. 6(1):33066. <https://doi.org/10.1038/srep33066>
- Zhao Y et al. 2016. Utility of combining morphological characters, nuclear and mitochondrial genes: An attempt to resolve the conflicts of species identification for ciliated protists. *Molecular Phylogenetics and Evolution*. 94:718–729. <https://doi.org/10.1016/j.ympev.2015.10.017>
- Zhong D et al. 2020. Extensive new *Anopheles* cryptic species involved in human malaria transmission in western Kenya. *Sci Rep*. 10(1):16139. <https://doi.org/10.1038/s41598-020-73073-5>
- Zlatogursky VV, Klimov VI. 2016. Barcoding Heliozoa: Perspectives of 18S rDNA for Distinguishing Between *Acanthocystis* Species. *Protist*. 167(6):555–567. <https://doi.org/10.1016/j.protis.2016.09.004>
- Zou S et al. 2016. How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). *Sci Rep*. 6(1):36822. <https://doi.org/10.1038/srep36822>
- Zurita-Artaloitia JM, Rivera J, Vinuesa P. 2023. Extensive Cryptic Diversity and Ecological Associations Uncovered among Mexican and Global Collections of *Naegleria* and *Vermamoeba* Species by 18S Ribosomal DNA, Internal Transcribed Spacer, and Cytochrome Oxidase Subunit I Sequence Analysis Santos ALD, editor. *Microbiol Spectr*. 11(2):e03795-22. <https://doi.org/10.1128/spectrum.03795-22>

2. COMPARISON OF TWO GENOME ASSEMBLIES FOR *PRYMNESIUM PARVUM* STRAIN UTEX2797

2.1 Introduction

Cryptic species complexes, consisting of two or more genetically distinct species with indistinguishable morphology, are major contributors to genetic diversity in many protist lineages (John et al., 2014; Simon et al., 2008; e.g., Sonneborn, 1975). *Prymnesium parvum* (*sensu lato*, s.l.) is a toxic, bloom-forming haptophyte now recognized as such a complex (Binzer et al., 2019; Wisecaver et al., 2023). Although the exact number of constituent species is unclear, recent comparative studies have revealed substantial genetic divergence between strains (Anestis et al., 2021; Kuhl et al., 2024; Wisecaver et al., 2023). Blooms of *P. parvum* s.l. have become more frequent in recent decades, exemplified by a particularly severe 500-kilometer-long bloom in the Oder River between Germany and Poland in 2022 (Free et al., 2023; Köhler et al., 2024). As more laboratories obtain *P. parvum* s.l. strains, the absence of reliable morphological characters heightens the risk of strain misidentification.

The *P. parvum* s.l. species complex is comprised of three main clades (A, B, and C), each producing a distinct form of the toxin prymnesin (Binzer et al., 2019). A comparative genomic analysis of 15 *P. parvum* s.l. strains by Wisecaver et al. (2023) further divided clade A into three subclades. Subclade A1 included two U.S. strains, UTEX3227 and CCMP 3037; subclade A2 contained two European strains, RCC 3703 and CCMP 2941; and the final type A subclade comprised two strains, UTEX 2797 and UTEX3226, both isolated from Texas waterbodies in 2002 and 2010, respectively. Both UTEX 2797 and UTEX3226 showed evidence of hybridization between the A1 and A2 subclades. Specifically, both strains had elevated levels of heterozygosity compared to the other 13 strains in the analysis. Both strains also contained roughly twice as much DNA content per nucleus (~280 Mbp as estimated via flow cytometry) than their k-mer-based estimate of genome size (~120 Mbp). Wisecaver et al (2023) assembled the genome of UTEX 2797 (hereafter UTEX 2797-W) using a combination of Oxford Nanopore Technologies (ONT) long reads, Illumina short reads, and Hi-C reads for scaffolding. They performed a phylogenomic analysis of the genes on the resulting 66 scaffolds and showed that most scaffolds formed syntenic pairs with genes on one member of the scaffold-pair grouping with the A1 subclade and genes on

the other member of the scaffold-pair grouping with the A2 subclade. Taken together, the data suggested that UTEX 2797 and UTEX3226 were formed via hybridization between A1-like and A2-like parents with the apparent retention of the full, or near full, genomic compliment of each parent (Figure 2.1).



Figure 2.1. Model of evolutionary relationships between *P. parvum* s.l. strains present in this analysis based on phylogenetic and genomic analyses performed by Wisecaver et al. (2023).

Recently, Jian et al. (2024) described two new *de novo* genome assemblies for CCMP 3037 and UTEX 2797 (hereafter CCMP 3037-J and UTEX 2797-J), which were created using a combination of PacBio HiFi long reads and DNBSseq short reads (Table 2.1). The authors reported genome assembly sizes of 107.32 Mbp for strain CCMP 3037-J and 97.56 Mbp for strain UTEX 2797-J. Notably, the UTEX 2797-J assembly was half the size of the 197.59 Mbp UTEX 2797-W genome assembly from Wisecaver et al. (2023). Jian et al. (2024) found no evidence of high-heterozygosity or hybridization in their UTEX 2797-J assembly.

Table 2.1 Statistics from all A-type *P. parvum* s.l. long-read assemblies

Assembly	UTEX3227	UTEX 2797-W (Wisecaver)	CCMP 3037-J (Jian)	UTEX 2797-J (Jian)
Source	Wisecaver et al. 2023	Wisecaver et al. 2023	Jian et al. 2024	Jian et al. 2024
Read Inputs	Illumina, ONT, Hi-C	Illumina, ONT, Hi-C	HiFi, DNBSeg	HiFi, DNBSeg
Assembly Length (Mbp)	93.538	197.593	107.322	97.559
Scaffold Number	34	66	na	na
Scaffold N50 (Mbp)	3.203	3.431	na	na
Contig Number	225	585	362	463
Contig N50 (Mbp)	0.852	0.548	0.968	0.597
GC%	57.74	57.34	57.38	57.59

Beyond the problems caused by contradicting assemblies in the field, UTEX 2797 is a widely used laboratory strain featured in numerous published studies (e.g., Binzer et al., 2019; Blossom et al., 2014; Chávez Montes et al., 2024; Lysgaard et al., 2018; Rashel and Patiño, 2017; Richardson and Patiño, 2021). Confusion surrounding the two UTEX 2797 assemblies has led to the strain's exclusion from a recent analysis of the *P. parvum* s.l. complex (see Kuhl et al., 2024). To identify the cause of the discrepancy between the UTEX 2797-W and UTEX 2797-J assemblies, we conducted a series of tests on nucleic acid sequences from UTEX 2797 and CCMP 3037. Our analysis incorporated newly generated sequences from freshly obtained cultures of both strains, as well as sequences of UTEX 2797 acquired by a third party.

2.2 Methods

2.2.1 Data sources

Assemblies were downloaded from CNGB and Figshare for assemblies from Jian et al. (2024) and Wisecaver et al. (2023), respectively. Statistics were calculated using stats.sh from the BBTools v38.91 bioinformatics suite (Bushnell, 2017).

2.2.2 Illumina Sequencing

Fresh cultures of UTEX 2797 and CCMP 3037 were acquired directly from their respective culture collections. Genomic DNA was isolated from each strain and Illumina short-read sequencing was performed (Novogene Corporation Inc., Sacramento, CA) as described in Wisecaver et al. (2023). Briefly, cultures were grown to late-exponential phase before pelleting and snap-freezing in liquid nitrogen. DNA was extracted from cell-pellets using a CTAB-based protocol (Auber, 2019).

2.2.3 Read processing and k-mer counting

UTEX 2797 and CCMP 3037 short-read libraries from both publications were compared alongside the resequenced libraries generated in this study (Jian et al., 2024; Wisecaver et al., 2023). All reads were processed according to the pipeline in Wisecaver et al. (2023). Briefly, bacterial contamination was flagged and removed using BlobTools v1.1.1 (Laetsch and Blaxter, 2017), and the filtered Illumina reads were randomly subsampled to equal read counts ($n = 40$ million pairs) with BBMap reformat.sh v38.87 to facilitate comparison across plots. K-mers were counted with KMC v3.1.1 (Kokot et al., 2017).

2.2.4 SNP analysis

Two reference genomes are currently available for *P. parvum* s.l.: the prymnesin type A producing strain UTEX3227 sequenced by Wisecaver et al. (2023) and the type B producing strain ODER1 sequenced by Kuhl et al. (2024). To avoid reference bias that could occur when aligning to the type A genome, reads were aligned to haplotype 1 of the ODER1 type B genome (Kuhl et al., 2024) using BWA-MEM v0.7.17 (Li, 2013) for Illumina short reads and Minimap2 v2.28 (Li, 2018) for ONT and PacBio long reads. In addition, we downloaded Illumina RNAseq reads for UTEX 2797 sequenced by a third party (Kuhl et al., 2024), which were aligned to the ODER1 genome using STAR v2.7.11b (Dobin et al., 2013). All properly paired read alignments were filtered to remove read duplicates and secondary/chimeric alignments using Samtools v1.9 and GATK v4.2.6.1 MarkDuplicates (Brouard et al., 2019; Danecek et al., 2021; Li et al., 2009; Van der Auwera and O'Connor, 2020).

SNPs were called using BCFtools v1.17 mpileup (Danecek et al., 2021) and filtered to only those present in all samples with a minimum quality of 30 using VCFtools v0.1.16x (Danecek et al., 2011), which yielded 77,168 total SNPs. A principle components analysis (PCA) plot was created in PLINK v1.90b6.21 (Purcell et al., 2007) based on 6,783 independent variants that were extracted using a window size of 50 kb, step size of 10 kb, and r^2 of 0.1. All 77,168 SNPs were used to generate a distance matrix in PLINK which was visualized using NeighborNet implemented in SplitsTree4 v4.19.2 with the lambda fraction set to 1 and variance set to Ordinary Least Squares (Bryant, 2003; Huson, 1998).

2.3 Results and Discussion

The k-mer plots for the CCMP 3037 and UTEX 2797 libraries generated by Wisecaver et al. (2023) and Jian et al. (2024) were compared to resequenced libraries for each strain. In the Jian et al. (2024) data, strain CCMP 3037 is highly heterozygous, as indicated by the two distinct peaks in the k-mer frequency plot (Figure 2.2A). Conversely, strain UTEX 2797 shows only one dominant k-mer peak, indicative of far less heterozygosity (Figure 2.2B). The pattern is the opposite in the Wisecaver et al. (2023) data: CCMP 3037 has low heterozygosity (Figure 2.2C), while UTEX 2797 is high (Figure 2.2D). The resequenced reads demonstrate the same pattern of heterozygosity as previously published by Wisecaver et al. (2023) and the opposite pattern of what was published by Jian et al. (2024) (Figure 2.2E,F). This potentially indicated that the strain IDs were switched for the Jian et al. (2024) investigation.

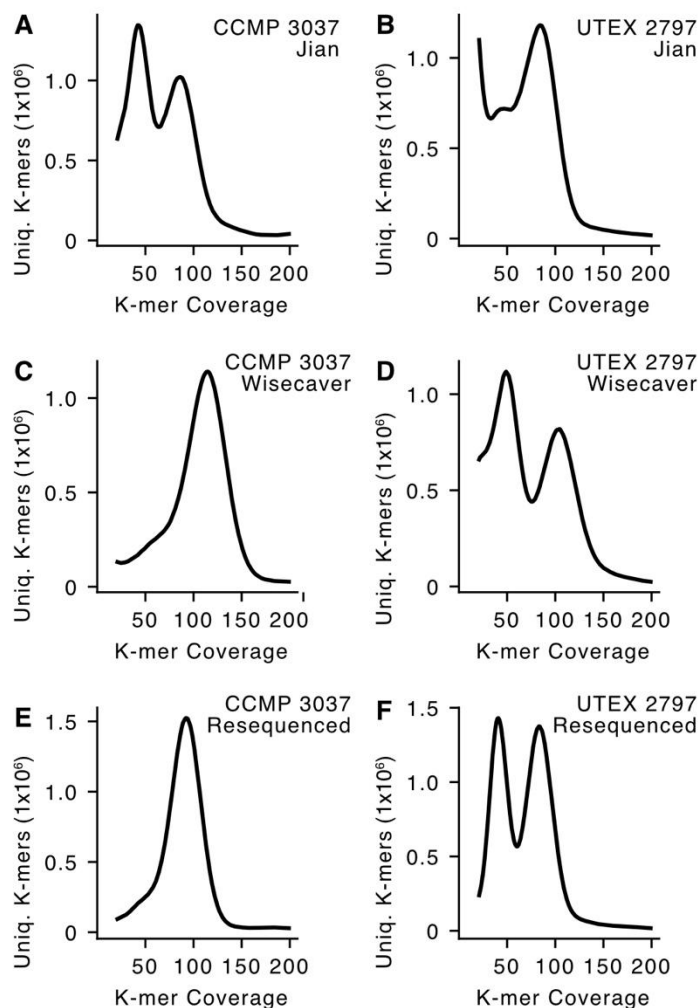


Figure 2.2. K-mer frequency plots for CCMP 3037 and UTEX 2797. Data from Jian et al. (A,B); from Wisecaver et al. (C,D); and resequenced for this study (E,F).

In the PCA of the resulting SNPs, the strains clustered into three distinct groups along the first principal component axis (Figure 2.3A). These three clades correspond to the phylogenetic clades proposed by Wisecaver et al. (2023). The first group contains strains from the A2 subclade: CCMP 2941 and RCC 3703. The second group contains strains of the A1 subclade: UTEX3227 and CCMP 3037 sequenced by Wisecaver et al. (2023), hereafter CCMP 3037-W. UTEX 2797-J and the freshly reacquired and resequenced CCMP 3037 also cluster in this first group. The Illumina reads from CCMP 3037-W, UTEX 2797-J, and the resequenced CCMP 3037 almost entirely overlap one another in PC space, which illustrates the strong similarity between the three samples and indicates that they correspond to the same strain. The third group contains the two hybrid strains proposed by Wisecaver et al. (2023): UTEX3226 and UTEX 2797-W. CCMP 3037-J and

the freshly reacquired and resequenced UTEX 2797 also cluster within this third group. Similar to the first cluster, the Illumina reads from UTEX 2797-W, CCMP 3037-J, and resequenced UTEX 2797 overlap one another. Critically, the UTEX 2797 RNAseq reads independently sequenced by Kuhl et al. (2024) also cluster in this third group demonstrating that UTEX 2797-W (Wisecaver et al. 2023), the freshly reacquired and resequenced UTEX 2797 (this paper), the UTEX 2797 used for RNAseq (Kuhl et al., 2024), and CCMP 3037-J (Jian et al. 2023) are all UTEX 2797.

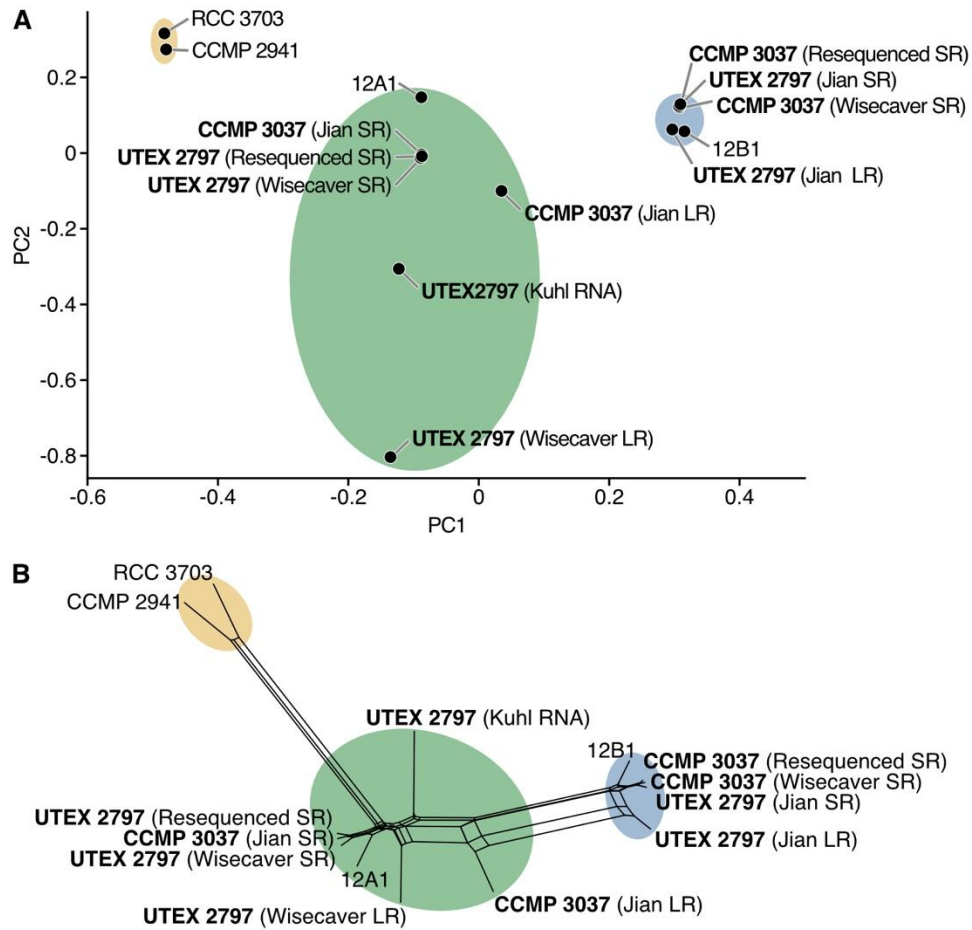


Figure 2.3. SNP analysis of *P. parvum* s.l. type A strains. A) PCA analysis B) Phylogenetic network. LR, long read; SR, short read.

All filtered variants were used to create a NeighborNet (Bryant, 2003; Huson, 1998) phylogenetic network, which recovered the same trends observed in the PCA (Figure 2.3B). UTEX 2797-J grouped with the A1 strains from Wisecaver et al (2023), and CCMP 3037-J grouped with the hybrid strains from Wisecaver et al (2023). UTEX 2797 from Kuhl et al (2024) also grouped with

the hybrid strains from Wisecaver et al. (Figure 2.3B). Taken together, these analyses show that the CCMP 3037 and UTEX 2797 strain labels were inadvertently swapped in Jian et al. (2024).

If CCMP 3037-J is derived from UTEX 2797 sequence, why then why does it lack the high heterozygosity and gene duplication seen in UTEX 2797-W? To explore this, we examined the HiFi long-read coverage for the CCMP 3037-J assembly. However, the libraries available in the China National GeneBank DataBase (CNGDB, accession CNP0004284) contain far fewer reads, and consequently much lower coverage, than reported by Jian et al. (2024). The deposited data provide only ~6.5X coverage for each strain, whereas Hifiasm (Cheng et al. 2021), the assembler used by Jian et al., recommends at least four times that amount (~13X per haplotype, i.e., 26X total for a diploid/hybrid strain). Such low coverage could have led to haplotype collapse, producing a single consensus haplotype in CCMP 3037-J and thereby masking any gene duplication present in UTEX 2797. However, without access to the complete read set used in the original CCMP 3037-J assembly, this remains speculative.

2.4 Conclusion

Misidentification of samples used in genomic analyses is an ongoing challenge for the field, resulting in corrections (e.g., Dixon et al., 2024) or retractions (e.g., Christensen et al., 2021; Mu et al., 2017; Smith and Editors, 2018) depending on the context. The discrepancy between the UTEX 2797-W and UTEX 2797-J assemblies has propagated confusion and resulted in the strain's exclusion from a subsequent genomic comparison of *P. parvum* s.l. (see Kuhl et al., 2024). This is unfortunate as UTEX 2797 has been used in numerous prior studies of the species complex (e.g., Binzer et al., 2019; Blossom et al., 2014; Chávez Montes et al., 2024; Lysgaard et al., 2018; Rashed and Patiño, 2017; Richardson and Patiño, 2021) and therefore should be included in genome-level investigations when appropriate. The elevated heterozygosity present in UTEX 2797 and the strain's proposed hybrid ancestry (Wisecaver et al. 2023) is an intriguing aspect of its biology worthy of further study. Strain misidentification is a particular concern when working with microbial and cryptic species such as *P. parvum* s.l. because the lack of morphological identifiers necessitates that mix ups be detected via sequencing. The issue highlighted here shows the dire need for clade-specific markers to efficiently confirm *P. parvum* s.l. strain identity and emphasize the importance for markers in developing species complexes.

2.5 References

- Anestis, K., Kohli, G.S., Wohlrab, S., Varga, E., Larsen, T.O., Hansen, P.J., 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>
- Auber, R., 2019. Total DNA extraction from plant tissue using CTAB method [WWW Document]. [protocols.io](https://dx.doi.org/10.17504/protocols.io.bamnic5e). URL <https://dx.doi.org/10.17504/protocols.io.bamnic5e> (accessed 2.11.22).
- Binzer, S.B., Svenssen, D.K., Daugbjerg, N., Alves-de-Souza, C., Pinto, E., Hansen, P.J., Larsen, T.O., 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>
- Blossom, H.E., Rasmussen, S.A., Andersen, N.G., Larsen, T.O., Nielsen, K.F., Hansen, P.J., 2014. *Prymnesium parvum* revisited: relationship between allelopathy, ichthyotoxicity, and chemical profiles in 5 strains. *Aquat. Toxicol. Amst. Neth.* 157, 159–166. <https://doi.org/10.1016/j.aquatox.2014.10.006>
- Brouard, J.-S., Schenkel, F., Marete, A., Bissonnette, N., 2019. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J. Anim. Sci. Biotechnol.* 10, 44. <https://doi.org/10.1186/s40104-019-0359-0>
- Bryant, D., 2003. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Mol. Biol. Evol.* 21, 255–265. <https://doi.org/10.1093/molbev/msh018>
- Bushnell, B., 2017. BBTools Software Package. <https://sourceforge.net/projects/bbmap/>.
- Chávez Montes, R.A., Mary, M.A., Rashel, R.H., Fokar, M., Herrera-Estrella, L., Lopez-Arredondo, D., Patiño, R., 2024. Hormetic and transcriptomic responses of the toxic alga *Prymnesium parvum* to glyphosate. *Sci. Total Environ.* 954, 176451. <https://doi.org/10.1016/j.scitotenv.2024.176451>
- Christensen, K.A., Rondeau, E.B., Minkley, D.R., Leong, J.S., Nugent, C.M., Danzmann, R.G., Ferguson, M.M., Stadnik, A., Devlin, R.H., Muzzerall, R., Edwards, M., Davidson, W.S., Koop, B.F., 2021. Retraction: The Arctic charr (*Salvelinus alpinus*) genome and transcriptome assembly. *PLOS ONE* 16, e0247083. <https://doi.org/10.1371/journal.pone.0247083>
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>

- Dixon, O.F.L., De Silva, C., Phillips, B.T., Gallagher, A.J., 2024. Correction to: Novel deep-sea observations reveal twilight zone occurrence for two species of pelagic sharks: the bignose shark *Carcharhinus altimus* and the silky shark *Carcharhinus falciformis*. *Environ. Biol. Fishes* 107, 1171–1174. <https://doi.org/10.1007/s10641-024-01610-6>
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Free, G., Van de Bund, W., Gawlik, B., Van Wijk, L., Wood, M., Guagnini, E., Koutelos, K., Annunziato, A., Grizzetti, B., Vigiak, O., Gnecci, M., Poikane, S., Christiansen, T., Whalley, C., Antognazza, F., Zerger, B., Hoeve, R.J., Stielstra, H., 2023. An EU analysis of the ecological disaster in the Oder River of 2022: lessons learned and research based recommendations to avoid future ecological damage in EU rivers, a joint analysis from DG ENV, JRC and the EEA. Publications Office of the European Union. <https://doi.org/10.2760/067386>
- Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68–73. <https://doi.org/10.1093/bioinformatics/14.1.68>
- Jian, J., Wu, Z., Silva-Núñez, A., Li, X., Zheng, X., Luo, B., Liu, Y., Fang, X., Workman, C.T., Larsen, T.O., Hansen, P.J., Sonnenschein, E.C., 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>
- John, U., Litaker, R.W., Montresor, M., Murray, S., Brosnahan, M.L., Anderson, D.M., 2014. Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: The introduction of five species with emphasis on molecular-based (rDNA) classification. *Protist* 165, 779–804. <https://doi.org/10.1016/j.protis.2014.10.001>
- Köhler, J., Varga, E., Spahr, S., Gessner, J., Stelzer, K., Brandt, G., Mahecha, M.D., Kraemer, G., Pusch, M., Wolter, C., Monaghan, M.T., Stöck, M., Goldhammer, T., 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>
- Kokot, M., Długosz, M., Deorowicz, S., 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33, 2759–2761. <https://doi.org/10.1093/bioinformatics/btx304>
- Kuhl, H., Strassert, J.F.H., Čertnerová, D., Varga, E., Kreuz, E., Lamatsch, D.K., Wuertz, S., Köhler, J., Monaghan, M.T., Stöck, M., 2024. The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins. *Curr. Biol.* S0960982224008170. <https://doi.org/10.1016/j.cub.2024.06.033>
- Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. *F1000Research* 6, 1287. <https://doi.org/10.12688/f1000research.12232.1>
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>

- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [accessed 2023 Oct 11]. <http://arxiv.org/abs/1303.3997>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lysgaard, M.L., Eckford-Soper, L., Daugbjerg, N., 2018. Growth rates of three geographically separated strains of the ichthyotoxic *Prymnesium parvum* (Prymnesiophyceae) in response to six different pH levels. *Estuar. Coast. Shelf Sci.* 204, 98–102. <https://doi.org/10.1016/j.ecss.2018.02.030>
- Mu, X., Yang, Y., Liu, Y., Luo, D., Xu, M., Wei, H., Gu, D., Song, H., Hu, Y., 2017. Retraction Note to: The complete mitochondrial genomes of two freshwater snails provide new protein-coding gene rearrangement models and phylogenetic implications. *Parasit. Vectors* 10, 350. <https://doi.org/10.1186/s13071-017-2287-1>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>
- Rashel, R.H., Patiño, R., 2017. Influence of genetic background, salinity, and inoculum size on growth of the ichthyotoxic golden alga (*Prymnesium parvum*). *Harmful Algae* 66, 97–104. <https://doi.org/10.1016/j.hal.2017.05.010>
- Richardson, E.T., Patiño, R., 2021. Growth of the harmful alga, *Prymnesium parvum* (Prymnesiophyceae), after gradual and abrupt increases in salinity. *J. Phycol.* 57, 1335–1344. <https://doi.org/10.1111/jpy.13172>
- Simon, E.M., Nanney, D.L., Doerder, F.P., 2008. The “*Tetrahymena pyriformis*” complex of cryptic species. *Biodivers. Conserv.* 17, 365–380. <https://doi.org/10.1007/s10531-007-9255-6>
- Smith, C.M., Editors, T.P.O., 2018. Retraction: Virulent *Diuraphis noxia* Aphids Over-Express Calcium Signaling Proteins to Overcome Defenses of Aphid-Resistant Wheat Plants. *PLOS ONE* 13, e0191678. <https://doi.org/10.1371/journal.pone.0191678>
- Sonneborn, T.M., 1975. The *Paramecium aurelia* complex of fourteen sibling species. *Trans. Am. Microsc. Soc.* 94, 155–178. <https://doi.org/10.2307/3224977>
- Van der Auwera, G.A., O'Connor, B.D., 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media. https://catalog.nlm.nih.gov/permalink/01NLM_INST/1fua1rm/alma9917773213406676

Wisecaver, J.H., Auber, R.P., Watervoort, N.F., Fallon, T.R., Riedling, O.L., Manning, S.R., Moore, B.S., 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>

3. HYBRIDIZATION BETWEEN TOXIC BLOOM-FORMING ALGAE OF THE *PRYMNESIUM PARVUM SENSU LATO* SPECIES COMPLEX

Content in this chapter has been submitted for review and is pending publication with Genome Biology and Evolution:

Watervoort NF, Jeje T, Dilkes BP, Wisecaver JH. 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum sensu lato* species complex. *In Review*.

3.1 Abstract

The potential for hybridization to act as a driver of genetic diversity and adaptation in harmful algal bloom-forming species has received scarce attention, despite growing recognition of its occurrence in diverse protist and algal lineages. *Prymnesium parvum* s.l. (Haptophyta), is a cryptic species complex whose members form ecosystem- disruptive toxic algal blooms around the world. A prior genome analysis showed that UTEX2797, a widely used laboratory strain, originated via hybridization between clade A1 and clade A2 of this species complex. To assess the extent of A1×A2 hybridization in *P. parvum* s.l., we screened the genomes of 28 strains and identified 16 additional A1×A2 hybrid strains isolated from inland Texas or the eastern United States between 2001 and 2020. Chloroplast haplotypes indicated that hybridization between A1 and A2 may have occurred multiple times, and hybrids with different chloroplast haplotypes have been co-isolated from blooms in Texas in 2013 and 2020. Additionally, strain NIES1812 from Okinawa, Japan was sufficiently divergent from A1 and A2 to warrant designation as a separate clade, which we name A3. These results provide evidence for a facultative sexual life cycle in *P. parvum* s.l. and expand our understanding of the extensive cryptic genetic diversity present in the species complex. The frequent isolation of hybrid strains from North American blooms suggests that hybridization is common and represents a significant source of adaptive potential in these economically and ecologically damaging organisms.

3.2 Introduction

Hybridization is the result of sexual reproduction between two genetically isolated populations or species. In plants, hybridization coupled with whole genome duplication (allopolyploidy) is a

powerful evolutionary force, driving rapid diversification through the formation of new hybrid species and promoting increased rates of gene innovation and acquisition of novel traits (Soltis and Soltis 2009). Recently, whole-genome sequencing has revealed numerous examples of hybridization in fungi, including emergent pathogens (Steensels et al. 2021). In these cases, hybridization events are inferred from comparative analyses that rely on genome-level sequence data. Despite limited genomic data in most lineages of microbial eukaryotes, hybridization has been documented in ciliates (Obert et al. 2022), diatoms (Nomaguchi et al. 2018; Kim et al. 2020), dinoflagellates (Brosnahan et al. 2010), green algae (Craig et al. 2025), haptophytes (Wisecaver et al. 2023), trypanosomes (Matos et al. 2022) and other parasitic protozoa (Detwiler and Criscione 2010; Inbar et al. 2019). The growing number of hybrids identified across diverse eukaryotic lineages beyond plants and fungi suggests that hybridization may play an important evolutionary role in these groups as well.

Prymnesium parvum s.l. (*sensu lato*, in the broad sense) is a cryptic species complex of unicellular algae in the haptophyte lineage of eukaryotes. Species within the complex form toxic, ecosystem-disruptive algal blooms around the world (Roelke and Manning 2018; Patiño et al. 2023). *P. parvum* s.l. produce hemolytic toxins, prymnesins, that are among the largest metabolites known. The proteins responsible for prymnesin biosynthesis are comprised of ~100 separate enzymatic domains and are encoded by massive (~100 kbp) modular polyketide synthase genes called PKZILLAs (Fallon et al. 2024). To date, three distinct prymnesins (A-, B-, and C-type) have been identified that differ in the number of carbons in their backbone molecule, with A-type prymnesins being the largest (90 carbons) and C-types being the smallest (83 carbons) (Rasmussen et al. 2016). Relative toxicity also varies between prymnesin types with A-type prymnesins having the highest level of cytotoxicity to epithelial fish gill and human colon cell lines and B-type prymnesins having the lowest level of cytotoxicity (Varga et al. 2024). Production of the different prymnesins segregates phylogenetically across the *P. parvum* s.l. species tree forming three large A-, B-, and C-type prymnesin-producing clades (Binzer et al. 2019).

Extremely high levels of genetic diversity exist both within and among the A-, B-, and C-type clades. Haploid genome size estimates across the species complex range from 115 Mbp to 359 Mbp, and vegetatively haploid, diploid, triploid, and tetraploid strains can be maintained in stable

culture (Wisecaver et al. 2023; Kuhl et al. 2024). Although cryptic speciation is recognized within *P. parvum* s.l., the boundaries between species remain unclear. Increased divergence between some A- and C-type strains suggests the presence of distinct populations or potentially multiple species within these two clades (Wisecaver et al. 2023). The habitat range of *P. parvum* s.l. appears to be expanding, and blooms are predicted to increase both in frequency and intensity in the future (Tábora-Sarmiento et al. 2022; Macêdo et al. 2023; Sobieraj and Metelski 2023). Consequently, increasing attention is being paid toward the genetic diversity present in *P. parvum* s.l. to identify genes and traits responsible for their ecological success.

Recently, Wisecaver et al. (2023) provided genome-scale evidence of hybridization within *P. parvum* s.l. clade A. A chromosome-scale de novo assembly of *P. parvum* strain UTEX2797 revealed two co-linear genomes with divergent haplotypes. One UTEX2797 haplotype grouped phylogenetically with two strains from the United States: strain UTEX3227 and CCMP3037, hereafter referred to as clade A1. The second UTEX2797 haplotype grouped with two strains from Europe: RCC3703 and CCMP2941, hereafter referred to as clade A2. UTEX2797 contained roughly twice as much DNA content per nucleus (~280 Mbp, as estimated via flow cytometry), than its k-mer-based estimate of genome size (~120 Mbp) indicating the strain was an allodiploid hybrid (Wisecaver et al. 2023). UTEX2797 was first isolated in 2002 from a Colorado River sample from Texas (United States) and has since been used in numerous growth and sequencing experiments of *P. parvum* s.l. (e.g., Blossom et al. 2014; Talarski et al. 2016; Rashel and Patiño 2017; Lysgaard et al. 2018; Anestis et al. 2021; Richardson and Patiño 2021; Bannon et al. 2024; Chávez Montes et al. 2024; Varga et al. 2024). In their analysis, Wisecaver et al. (2023) suspected that an additional strain, UTEX3226, also from Texas (Table 1), was likely a hybrid due to elevated rates of heterozygosity in genome sequence data comparable to that of UTEX2797. To date, the hybrid nature of strain UTEX3226 has not been confirmed and the extent of hybridization in *P. parvum* s.l. is unknown.

To investigate whether additional hybrid strains exist beyond UTEX2797, we screened a collection of 38 *P. parvum* s.l. strains for signatures of A1×A2 hybridization (Table B.1). This included 25 *P. parvum* s.l. strains from culture collections and 10 strains isolated from recent Texas blooms. We identified numerous hybrid strains, all from North American blooms in the 21st century,

suggesting that hybrids are common in, and well adapted to, these environments. We present evidence that A1×A2 hybridization has likely occurred multiple times. The genomes of hybrid strains contain numerous structural variants relative to parental strains. Together, these results demonstrate that the hybrid genome found in UTEX2797 is not unusual for *P. parvum s.l.* and suggest that hybridization is an ongoing process in the evolution of *P. parvum* clade A.

3.3 Results

Heterozygosity was qualitatively assessed in K-mer coverage plots by the presence and relative height of a second peak at half the coverage of the homozygous peak (Figure 3.1, Figure A.2). High heterozygosity was previously observed in strain UTEX2797 and is indicative of its hybrid origin (Wisecaver et al. 2023). Fourteen additional strains showed a high level of heterozygosity, comparable to that of UTEX2797. All 15 high-heterozygosity strains were diploid, which we classified as putative allodiploid hybrids (Figure 3.1).

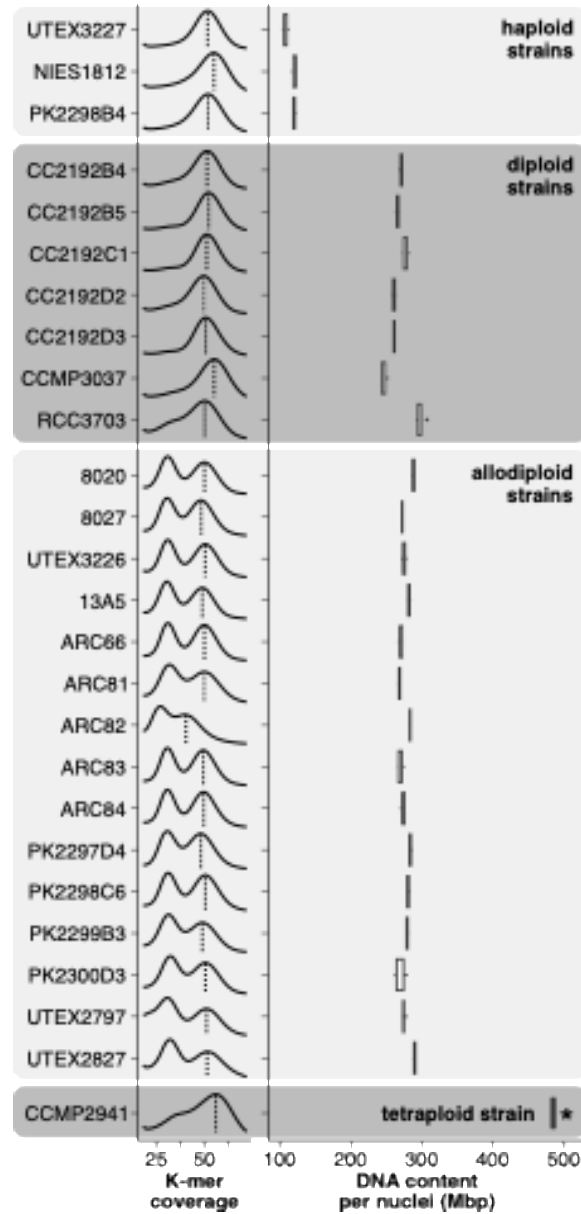


Figure 3.1 Heterozygosity and genome size variation among *P. parvum* clade A strains.

Filtered Illumina reads were randomly subsampled to 10 million pairs per strain prior calculation of K-mer coverage to control for differences in sequencing depth between strains. Homozygous k-mer peaks are indicated by the vertical dashed lines. Boxplots depict total DNA content per nuclei (Mbp) based on flow cytometry. (*) Flow cytometry measurements for CCMP2941 were taken from Wisecaver et al. (2023).

3.3.1 Variation in heterozygosity and ploidy in *P. parvum* clade A

The ploidy of all *P. parvum* clade A strains was inferred using their total nuclear DNA content compared to that of the haploid reference strain, UTEX3227, which has a nuclear DNA content of

107 Mbp (Table 3.1, Table B.2). In addition to UTEX3227, two other strains appear haploid: NIES1812 isolated in 2005 from Okinawa, Japan (DNA content = 120 Mbp) and PK2298B4, a newly isolated strain from a 2020 bloom in Possum Kingdom Lake, Texas (DNA content = 120 Mbp) (Table 3.1, Figure 3.1). Twenty-two strains were diploid with nuclear DNA content roughly twice that of the haploid strains (average DNA content = 298 Mbp) (Table 3.1, Figure 3.1). Diploids include four strains (CCMP3037, RCC3707, UTEX2797, and UTEX3226) previously identified by Wisecaver et al. (2023) as well as 18 newly characterized diploid strains all isolated from the United States between 2002 and 2020, including twelve strains from Texas, four from North Carolina, and two from South Carolina (Table 3.1). We did not have a sample of CCMP2941 in the lab for flow cytometry at this time, but previously reported a nuclear DNA content of 485 Mbp for this strain (Wisecaver et al. 2023), indicating it is a tetraploid (Table 3.1).

Heterozygosity was qualitatively assessed in K-mer coverage plots by the presence and relative height of a second peak at half the coverage of the homozygous peak (Figure 3.1, Figure S2). High heterozygosity was previously observed in strain UTEX2797 and is indicative of its hybrid origin (Wisecaver et al. 2023). Fourteen additional strains showed a high level of heterozygosity, comparable to that of UTEX2797. All 15 high-heterozygosity strains were diploid, which we classified as putative allodiploid hybrids (Figure 3.1).

Table 3.1 Description of *P. parvum* clade A strains.

Subclade	Strain ID	Date isolated	Isolation location	DNA content per nuclei (Mbp)	Ploidy	Dominant AIM ancestry
A1	UTEX3227	2010	Moss Creek Lake, TX, USA	107	haploid	A1
A1	CC2192B4	2019	Lake Colorado City, TX, USA	271	diploid	A1
A1	CC2192B5	2019	Lake Colorado City, TX, USA	265	diploid	A1
A1	CC2192C1	2019	Lake Colorado City, TX, USA	277	diploid	A1
A1	CC2192D2	2019	Lake Colorado City, TX, USA	261	diploid	A1
A1	CC2192D3	2019	Lake Colorado City, TX, USA	261	diploid	A1
A1	CCMP3037	2008	Twin Buttes Lake, Laramie, WY, USA	246	diploid	A1
A1	PK2298B4	2020	Possum Kingdom Lake, TX, USA	120	haploid	A1
A2	CCMP2941	2007	Repnoe, Tambov Oblast, Russia	485*	tetraploid*	A2
A2	RCC3703	1953	Millport, Great Cumbrae, UK	297	diploid	A2
A3	NIES1812	2005	Yufu Isl., Okinawa, Japan	120	haploid	--†
A hybrid	8020	2013	Possum Kingdom Lake, TX, USA	288	allodiploid	A1×A2
A hybrid	8027	2013	Possum Kingdom Lake, TX, USA	272	allodiploid	A1×A2
A hybrid	UTEX3226	2010	Moss Creek Lake, TX, USA	275	allodiploid	A1×A2
A hybrid	13A5	2010	E.V. Spence Reservoir, TX, USA	282	allodiploid	A1×A2
A hybrid	ARC66	2001	Hilton Head, SC, USA	270	allodiploid	A1×A2
A hybrid	ARC81	2002	Artesian Aquafer, Elizabeth City, NC, USA	268	allodiploid	A1×A2
A hybrid	ARC82	2002	Artesian Aquafer, Elizabeth City, NC, USA	283	allodiploid	A1×A2
A hybrid	ARC83	2002	Artesian Aquafer, Elizabeth City, NC, USA	270	allodiploid	A1×A2
A hybrid	ARC84	2002	Artesian Aquafer, Elizabeth City, NC, USA	273	allodiploid	A1×A2
A hybrid	PK2297D4	2020	Possum Kingdom Lake, TX, USA	284	allodiploid	A1×A2
A hybrid	PK2298C6	2020	Possum Kingdom Lake, TX, USA	280	allodiploid	A1×A2
A hybrid	PK2299B3	2020	Possum Kingdom Lake, TX, USA	279	allodiploid	A1×A2
A hybrid	PK2300D3	2020	Possum Kingdom Lake, TX, USA	269	allodiploid	A1×A2
A hybrid	UTEX2797	2002	US Hwy 183, Texas Colorado River, TX, USA	274	allodiploid	A1×A2
A hybrid	UTEX2827	2002	Charleston, SC, USA	290	allodiploid	A1×A2

* Genome size and C-value estimates for CCMP2941 are from (Wisecaver et al. 2023). † As an outgroup, A3 strain NIES1812 showed equal proportions of homozygous A1 and A2 AIMs (see Figure 3.2C).

3.3.2 Phylogenetic analyses identify a new A3 lineage and additional A1x A2 hybrids

Short reads from all A-type strains were aligned to the UTEX3227 v1.1 nuclear genome assembly (Fallon et al. 2024) and used to call 453,986 single nucleotide polymorphisms (SNPs) shared by the genomes of at least two strains. Strains formed four clusters in a SNP-based phylogenetic network (Figure 3.2A). Three clusters correspond to the A1, A2, and A-hybrid subclades identified in Wisecaver et al. (2023). The A1 subclade was comprised of both haploid (UTEX3227 and PK2298B4) and diploid strains (CCMP3037 from Wyoming and five new diploid strains isolated from a bloom in Lake Colorado City, Texas in 2019) (Figure 3.2A, Table 3.1). No additional A2 strains were identified in this analysis beyond the two already described in Wisecaver et al. (2023): RCC3703, a diploid strain from Great Cumbrae, United Kingdom, and CCMP2941, a tetraploid strain from Repnøe, Russia (Figure 3.2A, Table 3.1). All putative allodiploid hybrids identified based on K-mer frequencies and flow cytometry grouped with UTEX2797 in the A-hybrid clade (Figure 3.2A, Table 3.1). To investigate whether the hybridization signal could be due to culturing issues (e.g., two divergent diploid strains present in a single culture), we created seven single-cell reisolates of UTEX2797 (n=4) and UTEX3226 (n=3). These new isolates all showed the same level of heterozygosity as their progenitor cultures and grouped within the A-hybrid clade (Figure 3.2A, Figure A.2).

Strain NIES1812 from Okinawa, Japan grouped separately from all other strains in our analysis (Figure 3.2A). An earlier rDNA phylogeny placed NIES1812 as an outgroup to the A1 and A2 subclades (Binzer et al. 2019). Our SNP network was not informative with respect to the relative branching pattern of the A-type subclades as it lacked an outgroup. Therefore, to evaluate the placement of NIES1812 relative to A1 and A2, we used 9,769 single-copy genes for concatenation- and coalescent-based species tree analyses. The same species tree topology was recovered in both analyses (Figure A.3) and placed NIES1812 as the outgroup to all other A-type clades in our analysis (Figure 3.2B). Site concordance factors (sCF) and gene concordance factors (gCF) indicated strong support for the A1 and A2 subclades as well as the B- and C-type clades (Figure 3.2B). However, the analysis indicated a moderate amount of discordance regarding the placement of NIES1812 (sCF=31, gCF=42, Figure 3.2B) relative to the A1 and A2 subclades.

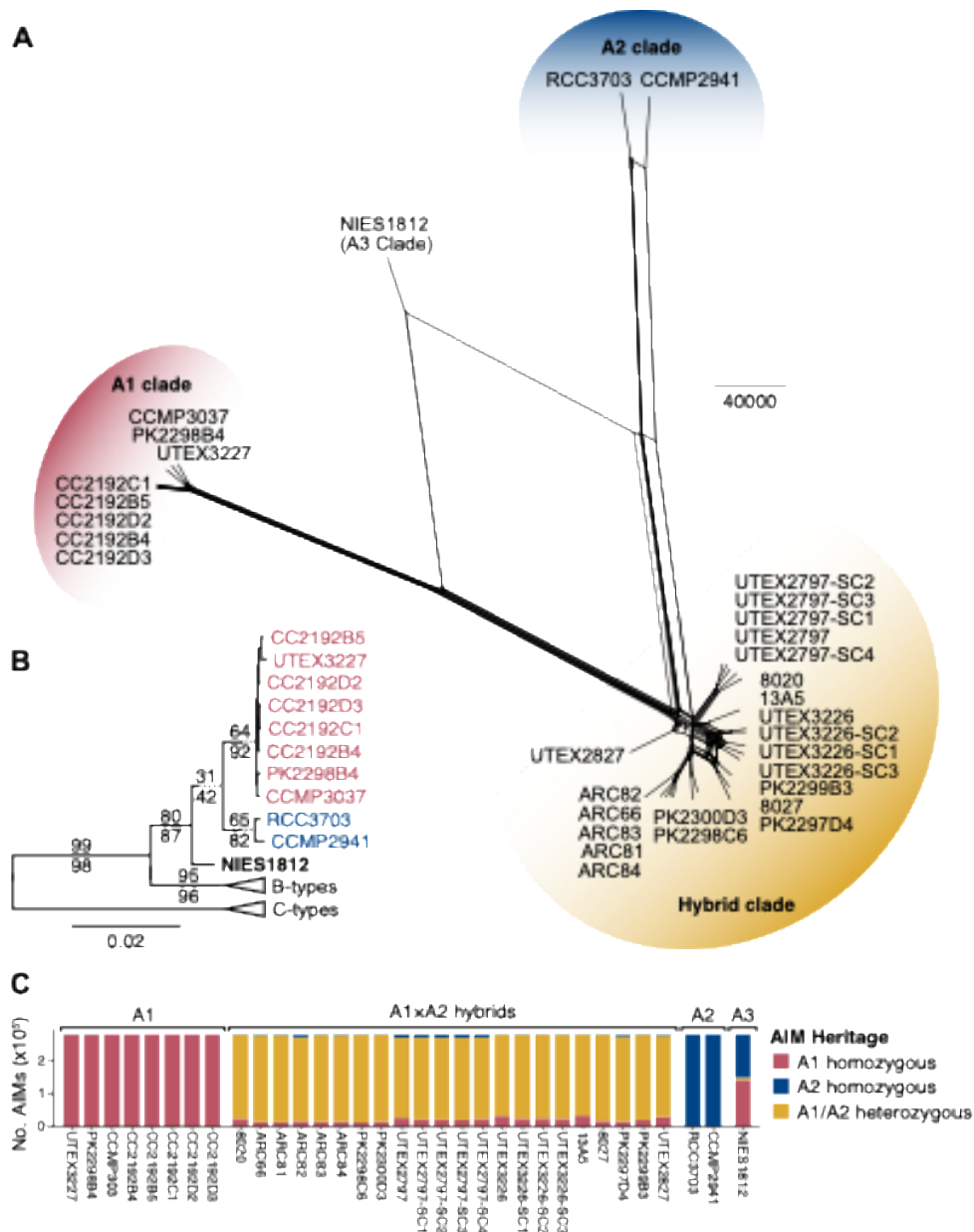


Figure 3.2 Phylogenetic analysis of *P. parvum* clade A strains.

A) Network of clade A strains based on distance matrix of 453,986 SNPs. Edge lengths represent absolute SNP differences with the scale bar corresponding to 40,000 SNPs. B) Concatenation-based Maximum Likelihood species phylogeny (excluding hybrid strains) based on 9,769 single copy genes showing the relative branching pattern of clade A1 (pink), A2 (blue), and the new A3 clade comprised of NIES1812 (bolded). Support values for nodes are reported on the preceding branch as site concordance factors (above) and gene concordance factors (below). The B and C clades have been collapsed for clarity; see Figure S3 for full tree. C) Proportion of AIMs that are A1 homozygous, A2 homozygous, or A1/A2 heterozygous per strain.

Of the 453,986 SNPs in our analysis, 61.22% (n=277,926) were completely fixed for different alleles in diploid A1 strains compared to A2 strains. At these fixed sites, hereafter referred to as Ancestrally Informative Markers (AIMs), allodiploid hybrids are expected to be fully heterozygous, provided several assumptions. These assumptions are 1) that the A1 and A2 subclades represent the true parental lineages of hybrids; 2) that the SNPs are truly fixed between the A1 and A2 lineages; and 3) that minimal subsequent backcrossing or genome evolution has taken place in hybrid strains. Matching this expectation, the majority of AIMs in all allodiploid hybrids strains were heterozygous (Figure 3.2C, Table B.4), lending further support to these strains being A1×A2 hybrids. On the high end, 94.68% of AIMs were heterozygous in strain PK2298C6, 4.72% were homozygous for the A1 allele, and 0.60% were homozygous for the A2 allele (Figure 3.2C, Table B.4). The hybrid strain with the lowest level of heterozygosity was UTEX2797; 87.23% of AIMs were heterozygous, 9.55% were homozygous for the A1 allele, and 3.22% were homozygous for the A2 allele (Figure 3.2C, Table B.4). In contrast to the pattern observed in hybrid strains, AIMs in NIES1812 were predominantly homozygous (as expected for a haploid strain) for either the A1 allele (50.77%) or the A2 allele (47.14%) (Figure 3.2C, Table B.4). These SNPs did not show evidence of A3 arising as a homoploid hybrid of A1 and A2 as AIMs were evenly distributed across the genome (Figure A.4). This pattern is consistent with its phylogenetic placement sister to A1/A2 and the AIM alleles present in NIES1812 likely representing the ancestral state present in their common ancestor.

3.3.3 A1×A2 hybrids contain either A1- or A2-like chloroplasts and display complex evolutionary relationships

To investigate chloroplast inheritance patterns in A1×A2 hybrids, short reads from A1, A2, and A1×A2 hybrid strains were aligned to the UTEX3227 v1 chloroplast genome assembly (Wisecaver et al. 2023) and used to identify 188 chloroplast-encoded SNPs. Chloroplast SNPs were strongly linked and formed two distinct chloroplast haplotypes (cpHaps) corresponding to the A1 and A2 subclades (Figure 3.3A). Consistent with single-strain cultures of hybrids, no evidence of chloroplast heteroplasmy was detected in hybrid strains. Instead, nine hybrid strains carried an A1-like chloroplast (hereafter referred to as cpHap-A1), and six hybrids carried an A2-like chloroplast (hereafter referred to as cpHap-A2; Figure 3.3A). All single-cell reisolates of UTEX2797 (cpHap-A1) and UTEX3226 (cpHap-A2) had the same cpHap as their progenitor cultures (Figure 3.3A).

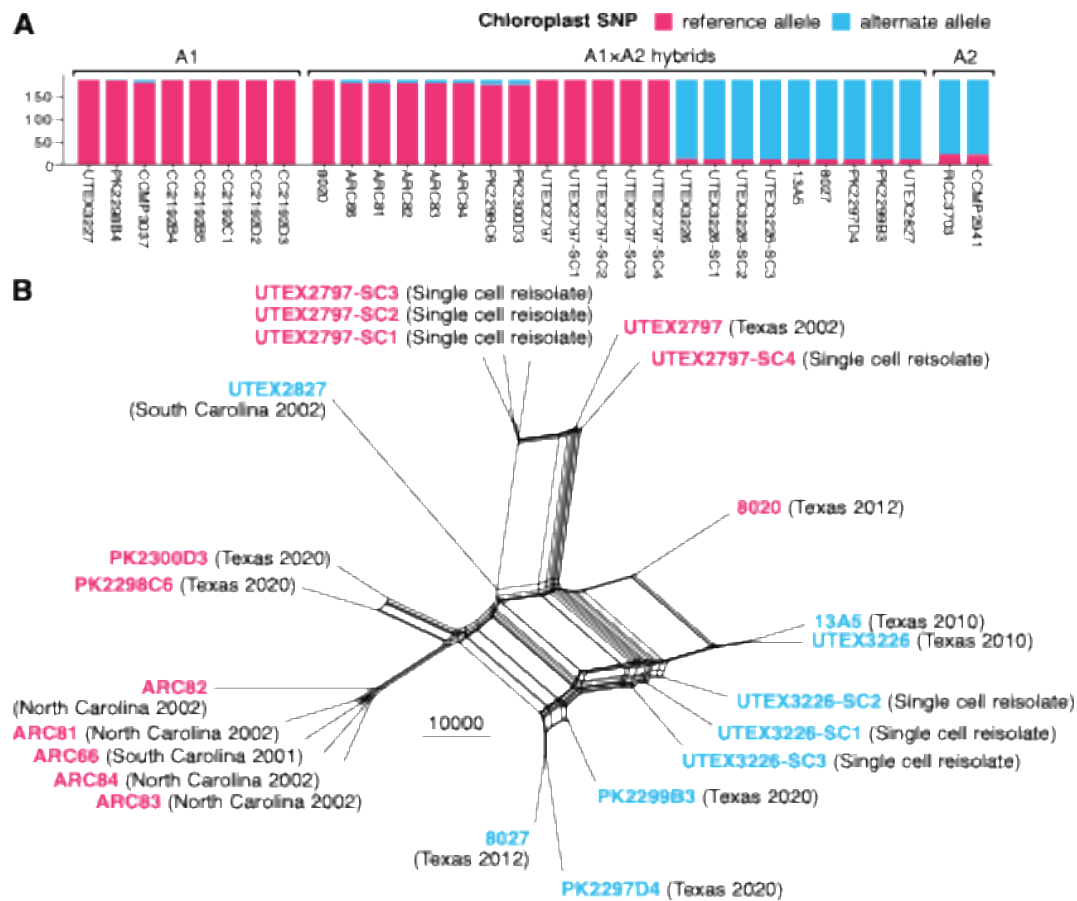


Figure 3.3 Genetic variation in A1x A2 hybrids.

A) Chloroplast SNPs. B) Neighbor net analysis of A1x A2 hybrids: cpHap-A1 (pink), cpHap-A2 (blue). Edge lengths represent absolute SNP differences with the scale bar corresponding to 10,000 SNPs.

Strains with different cpHaps were co-isolated from the same blooms. In Lake Possum Kingdom, Texas, strains 8020 (cpHap-A1) and 8027 (cpHap-A2) were co-isolated from the same bloom in 2012. Hybrids with both cpHap-A1 (PK2298C6 and PK2300D3) and cpHap-A2 (PK2297D4 and PK2299B3) were also co-isolated in 2020. The three Lake Possum Kingdom strains belonging to cpHap-A2 cluster together in the nuclear SNP phylogenetic network, but the three cpHap-A1 strains do not (Figure 3.3B). Occurrence of cpHap-A1 and cpHap-A2 hybrid strains extends beyond Lake Possum Kingdom. Strain ARC66 (cpHap-A1) and UTEX2827 (cpHap-A2) were both isolated from water samples from South Carolina in 2001 and 2002, respectively.

The clear segregation of hybrid chloroplast variation is contrasted by a more complex pattern in the nuclear genome. An analysis of molecular variance (AMOVA) showed that cpHap explained

only 22% of the variance present in the hybrid nuclear SNP dataset compared to 98% of the variance present in chloroplast SNPs ($p < 0.001$ for both). We also classified hybrid strains based on region of isolation, inland Texas (n=16) or coastal Carolinas (n=6), and found geography explained a similar amount of nuclear variance (26%, $p < 0.001$). The interaction between region and cpHap was explored further using hierarchical AMOVA (Table B.4). When region was nested within cpHap, cpHap lost significance ($\Phi_{CT} = -0.055$, $p = 0.344$) while region within cpHap remained significant ($\Phi_{SC} = 0.419$, $p < 0.001$). The pattern was reversed when cpHap was nested within region ($\Phi_{CT} = 0.069$, $p = 0.333$ for region; $\Phi_{SC} = 0.370$, $p < 0.001$ for cpHap). The AMOVA results indicate that hybrid genome variation reflects both geography and chloroplast type, but that the two factors are partially overlapping rather than hierarchical.

3.3.4 Chromosomal and segmental anomalies accumulate in laboratory hybrid strains

Chromosomal anomalies including aneuploidy (gain or loss of whole chromosomes) and large-scale structural mutations resulting in gain or loss of chromosomal segments have been reported in hybrid genomes of other microbial eukaryotes (van den Broek et al. 2015; Inbar et al. 2019; Maeda et al. 2022; Craig et al. 2025). To investigate whether similar chromosomal anomalies were present in A1×A2 hybrids, we plotted normalized coverage (NC; i.e., normalized read depth) alongside the relative abundance of A1- vs A2-derived reads. Numerous large-scale structural anomalies suggestive of duplications and losses of either A1 or A2 haplotypes were clearly visible in genome plots of hybrid strains (e.g., Figure 3.4). We chose to first compare strains UTEX2797 and UTEX3226 to their single-cell reisolates. This approach circumvents potential reference bias by treating each progenitor as its own baseline. Moreover, anomalies observed in reisolates have arisen *de novo* after the progenitor strains were first cultured in 2002 and 2012, respectively. We classified anomalies into three categories: chromosomal anomalies spanned whole scaffolds, telomeric anomalies involved the terminal ends of scaffolds, and interstitial anomalies were found in scaffold interiors.

In total, four chromosomal anomalies, four telomeric anomalies, and five interstitial anomalies were identified in single-cell reisolates (Table 3.2, Figures A.4-A.9). Critically, genetic information absent in the progenitor culture remained absent in all single-cell reisolates. For example, Scaffold 22 in UTEX2797 has three large regions (two telomeric and one interstitial)

with no heterozygosity; these regions showed a decrease in NC and loss of A2 alleles consistent with loss of the A2 haplotype. All UTEX2797 single-cell reisolates show the same loss; the A2 haplotype never reappears in these regions (Figure A.8).

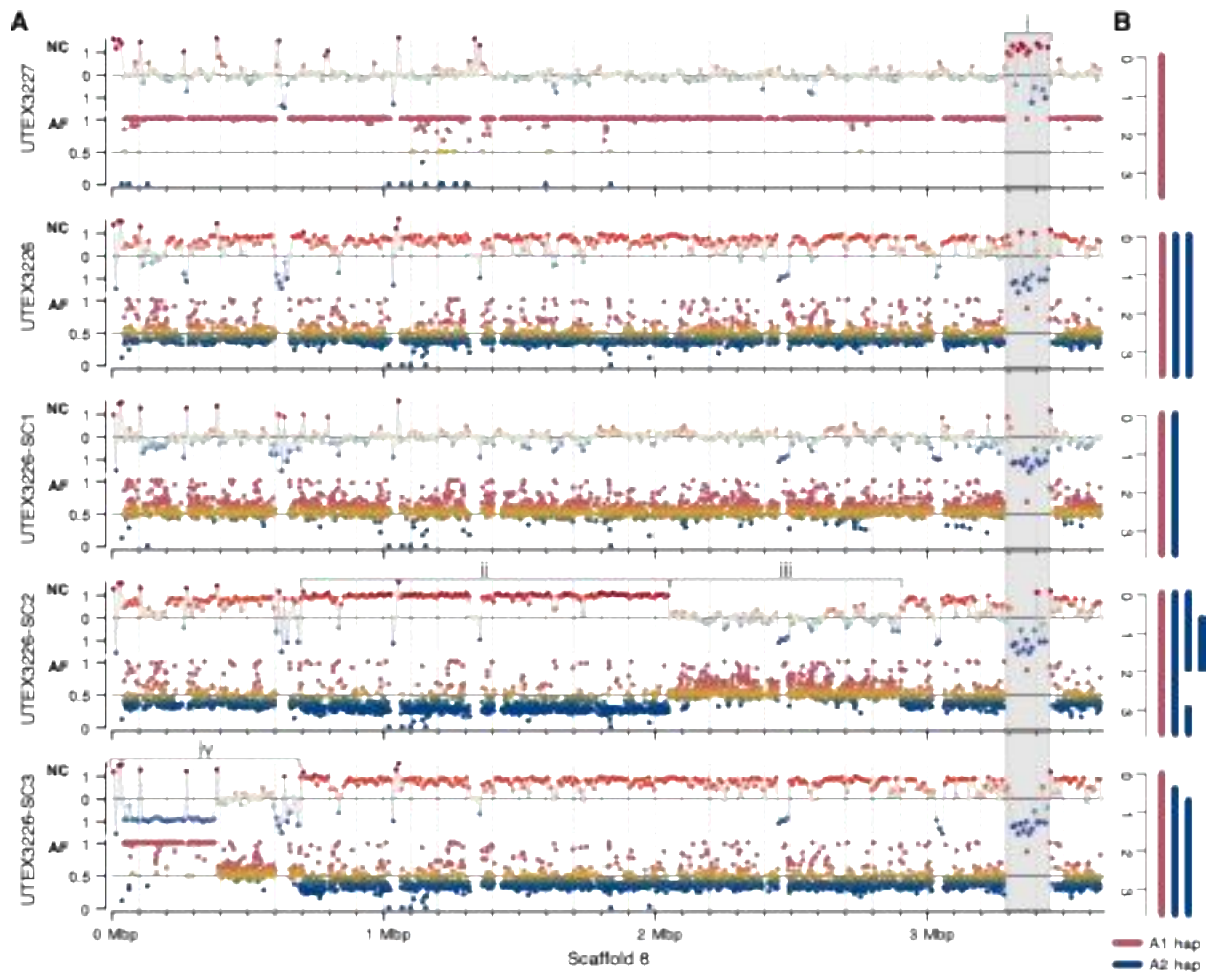


Figure 3.4 Scaffold 8 structural anomalies in UTEX3226 and single-cell reisolates.

A) For each strain, top track indicates log2 ratio of normalized genome coverage (NC); bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). Region i highlights a misassembly in the reference genome that was skipped in our analysis. Regions ii and iii correspond to interstitial anomalies in UTEX3226-SC2. Region iv indicates two overlapping telomeric anomalies in UTEX3226-SC3. B) Models of scaffold 8 organization (y-axis = Mbp) indicating regions duplicated or lost in single-cell reisolates relative to UTEX3226.

Table 3.2 List of structural anomalies in single-cell reisolates compared to progenitor strains.

Anomaly Type	Re-isolate(s) with variant	Scaffold	Start (Mbp)	End (Mbp)	Variant subtype
chromosomal	UTEX3226-SC1	8	scaffold start	scaffold end	reversion to disomy
chromosomal	UTEX2797-SC1	9	scaffold start	scaffold end	reversion to disomy
chromosomal	UTEX2797-SC2, UTEX2797-SC3	22	scaffold start	scaffold end	A1-derived trisomy
chromosomal	UTEX3226-SC1, UTEX3226-SC3	25	scaffold start	scaffold end	A1-derived trisomy
telomeric	UTEX3226-SC3	8	scaffold start	0.7	A2 telomeric segmental loss
telomeric	UTEX3226-SC3	8	scaffold start	0.39	A2 telomeric segmental loss
telomeric	UTEX2797-SC2, UTEX2797-SC3	8	1.25	scaffold end	A2 telomeric segmental duplication
telomeric	UTEX2797-SC2, UTEX2797-SC3	9	scaffold start	0.3	A2 telomeric segmental loss
interstitial	UTEX3226-SC1	2	3.19	3.39	A1 interstitial segmental loss
interstitial	UTEX3226-SC2	8	0.65	2.05	A2 interstitial segmental duplication
interstitial	UTEX3226-SC2	8	2.05	2.9	A2 interstitial segmental loss
interstitial	UTEX2797-SC1, UTEX2797-SC2, UTEX2797-SC3, UTEX2797-SC4	21	1.49	1.55	A2 interstitial segmental loss
interstitial	UTEX2797-SC1, UTEX2797-SC4	21	1.58	1.7	A2 interstitial segmental loss

To illustrate the range and dynamics of anomalies observed, we describe Scaffold 8 in detail below, as it contains examples of all three categories across the progenitor strain, UTEX3226, and its reisolates (Figure 3.4). In UTEX3226, Scaffold 8 shows elevated A2 allele frequency (A2_AF) across its full length alongside an increase in NC, which is consistent with a duplication of the A2-derived chromosome. Single-cell re isolate UTEX3226-SC1 showed a decrease in both NC and A2_AF indicating a reversion to disomy. UTEX3226-SC2 contains two examples of interstitial anomalies: Scaffold8:650000-2050000 shows increased NC and A2_AF indicative of an A2

interstitial duplication; Scaffold8:2050000-2900000 shows decreased NC and A2_AF indicative of A2 interstitial loss (Figure 3.4). Lastly, UTEX3226-SC3 contains two examples of telomeric losses, both involving the beginning of its two A2-derived chromosomes (Figure 3.4).

3.3.5 A1xA1 hybrids contain more genome anomalies than non-hybrid strains

We extended our analysis of genome structural anomalies to all *P. parvum* A1, A2, and A1×A2 hybrid strains, excluding single-cell reisolates, which were analyzed separately above. Across 24 strains, the median count of genome anomalies was 2.5 with seven strains (six non-hybrid and one hybrid) having no anomalies (Figure 3.5A). None of the diploid A1-type strains isolated in 2020 contained any anomalies. Anomaly counts were significantly higher in hybrid strains compared to non-hybrid strains (Wilcoxon-Mann-Whitney test, $Z = 2.55$, $p = 0.0096$; Figure 3.5B).

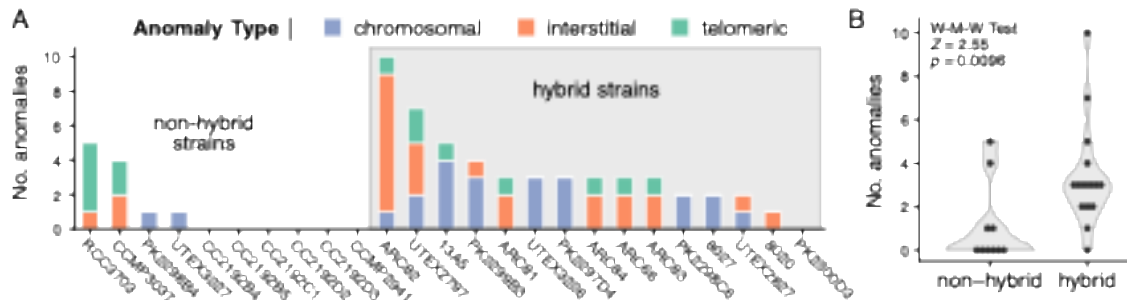


Figure 3.5 Structural anomalies detected in A1, A2, and A1x2 strains.

A) Number of chromosomal, interstitial, and telomeric anomalies detected per strain. B) Violin plots comparing hybrid vs non-hybrids. Statistical analysis was performed using a Wilcoxon-Mann-Whitney test.

The strain with the most anomalies was hybrid strain ARC82 ($n = 10$). Yet, this count represents an underestimate as Scaffold 28 in ARC82 displayed extreme variability in both NC and AF, which was indicative of significant chromosomal rearrangement relative to the UTEX3227 reference assembly, and prevented accurate inference of anomaly count for this scaffold in ARC82 (Figure A.10). We struggled to accurately estimate the number of structural anomalies in a second strain, UTEX2827, which had two clear anomalies: a Scaffold 9 polysomy and an interstitial loss on Scaffold26 (Figure A.6). However, several other regions in UTEX2827 showed high variability

NC and AF that prevented an accurate count of structural anomalies for this strain across several scaffolds (Figure A.11).

Unlike anomalies in single-cell reisolates, whose timing of origin is constrained by the known isolation dates of the progenitor strains, the evolutionary timing of anomalies detected in the remaining strains is often uncertain. Some anomalies may have evolved in lab cultures, whereas others may represent genomic variation present in natural populations. Four cpHap-A1 strains from North and South Carolina (ARC66, ARC81, ARC83, and ARC84) share the same three structural anomalies: an A2 interstitial loss on Scaffold 4, an A1 telomeric loss on Scaffold 13, and an A1 interstitial duplication on Scaffold 26 (Table B.6). Similarly, all five cpHap-A2 strains from Texas (8027, UTEX3226, 13A5, PK2297D4, and PK2299B3) are A1-disomic for Scaffold 27, having lost the A2-derived copy (Table B.6). The presence of the same structural anomalies in closely related but independently isolated strains suggests these genetic variants were present in the natural population from which they were sampled. In contrast, ARC82 was sampled from the same location and in the same year as ARC81, ARC83, and ARC84 but shares only one anomaly (the A1 interstitial duplication on Scaffold 26) with these strains (Table B.6). Other closely related strains (e.g., PK2298C6 and PK2300D3; UTEX3226 and 13A5; and PK2297D4 and PK2299B3) also differ in some anomalous regions (Table B.6). Whether these differences represent standing variation present in natural populations or culturing-induced differences requires further study.

Lastly, two scaffolds show evidence for repeated polysomies (Table B.6). Scaffold 8 was polysomic in 20.8% of strains in our analysis: four cpHap-A2 strains (UTEX3226, 13A5, PK2297D4, and PK2299B3) and one cpHap-A1 strain (PK2298C6). This polysomy is notably absent in PK2300D3 (closely related to PK2298C6) and 8027 (closely related to the four cpHap-A2 strains), suggesting that polysomy of Scaffold 8 is either recurrent or frequently lost. This is consistent with the pattern observed in UTEX3226-SC1, which reverted to disomy from the polysomic state present in UTEX3226 (Figure 3.4). Similarly, Scaffold 19 was polysomic in 37.5% of strains in our analysis: five cpHap-A2 strains (8027, UTEX3226, 13A5, PK2297D4, PK2299B3), two cpHap-A1 strains (PK2298C6, ARC82), and two A1 strains (PK2298B4, UTEX3227). The presence of Scaffold 19 polysomies in both hybrid and non-hybrid strains

suggests that some chromosomes may be prone to polysomy regardless of hybrid status, though whether this reflects a shared ancestral variant or recurrent independent events remains unclear.

3.4 Discussion

Strains of *P. parvum* s.l. have been studied in laboratory cultures for over 70 years and have never been observed to reproduce sexually. Nevertheless, the existence of A1×A2 hybrid strains provides strong evidence for a facultative sexual lifecycle in these algae. The lifecycles of other haptophyte algae are poorly understood outside of the coccolithophores, which have complex sexual haplodiplontic life cycles (Bousquet et al. 2025). In coccolithophores, lifecycle phases are heteromorphic; cells alternate between a flagellated haploid phase and a calcified diploid phase, each capable of clonal reproduction. *P. parvum* s.l. likely undergoes a similar haplodiplontic lifecycle, but the haploid and diploid phases are morphologically indistinguishable (Figure 3.6). Different A1×A2 hybrid strains carry chloroplast haplotypes from either the A1 or A2 parent, indicating that while chloroplast inheritance is uniparental during syngamy, both parental lineages can serve as the chloroplast donor (Figure 3.6). Additional work is needed to determine whether distinct mating types exist in *P. parvum* s.l. as in other algae, such as *Chlamydomonas* (Umen and Coelho 2019). In the past, research on the *P. parvum* s.l. lifecycle has been complicated by the cryptic nature of the species complex, both in terms of cryptic divergence between clades and cryptic lifecycle phases. Identification of multiple haploid and diploid strains within the A1 clade (Table 3.1) will facilitate new research into the *P. parvum* s.l. lifecycle.

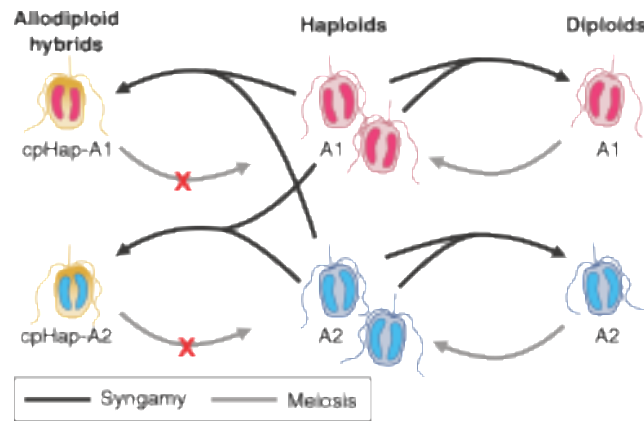


Figure 3.6 Proposed model of *P. parvum* clade A lifecycle and hybridization

Allodiploid hybrids are formed via syngamy between A1 and A2 haploid cells; genetic incompatibilities prevent meiosis (red X). During hybridization, chloroplasts can be inherited from either the A1 parent (cpHap-A1, pink) or the A2 parent (cpHap-A2, blue). All cell types are capable of asexual reproduction via mitosis (not shown).

The lack of intermediate strains that group in between the hybrid clade and either the A1 or A2 clades (Figure 3.2A) suggests a lack of backcrossing between hybrids and their parent populations. One possible explanation for this pattern could be that A1×A2 hybrids cannot complete meiosis (Figure 3.6) as is the case in some other hybrid systems (Bizzarri et al. 2016; Montecinos et al. 2017). This could be due to divergent parental karyotypes such that homologous chromosomes are unable to pair up correctly during meiosis (Hunter et al. 1996). Alternatively, genetic incompatibilities between the A1 and A2 alleles could lead to low viability of cells post meiosis (Xu and He 2011). The large number of chromosomal abnormalities present in our existing hybrid lab strains would certainly reduce the number of viable gametes produced following meiosis, but the exact timing of when most abnormalities evolved is unclear. In our phylogenomic analysis of protein-coding genes, the A1 and A2 clades clearly segregate with strong branch support (Figure 3.2B), consistent with the accumulation of significant genetic differences since their divergence. Additional chromosome-scale genome assemblies and comparative karyotyping of *P. parvum* clade A are needed to investigate whether the A1 and A2 strains possess chromosomal number differences and/or structural rearrangements that would inhibit meiosis in hybrids. While A1×A2 hybrids appear reproductively isolated from their parental populations, examples from yeasts suggest that hybrid fertility can sometimes be evolutionarily rescued through mutation of mating

type loci (Ortiz-Merino et al. 2017; Braun-Galleani et al. 2018), raising questions about the long-term evolutionary fate of *P. parvum* clade A hybrids.

The hybrid strains in our analysis are likely the products of multiple A1×A2 hybridization events. The existence of two chloroplast haplotypes is a strong indication that hybridization has occurred at least twice. However, if hybrid strains were the result of only two events (one event giving rise to cpHap-A1 strains and the other event producing cpHap-A2), that should be reflected in the nuclear data as well. Nuclear variation does cluster based on cpHap, but only after accounting for geography (Table 3.2). This pattern is more consistent with independent hybridization events producing cpHap-A1 and cpHap-A2 strains in Texas separate from those that gave rise to the same cpHap in the Carolinas.

Despite most strains in our analysis being from North America, *P. parvum* type A has a broad global distribution with cataloged strains from Europe, Israel, Japan, and Australia (Binzer et al. 2019). In our analysis, strain NIES1812 from Okinawa, Japan grouped separately from the A1 and A2 clades, consistent with earlier work (Binzer et al. 2019), indicating that the strain is sufficiently divergent to warrant designation as a separate A3 subclade. Overall, the *P. parvum* type A subclades show some geographic clustering, with A1 strains from North America, A2 strains from Europe, and the single A3 clade from Japan. However, this pattern should be interpreted cautiously, as *P. parvum* type A distribution and diversity remain poorly understood even in North America, where sampling has been most extensive.

The possibility of multiple hybridization events in *P. parvum* clade A raises the intriguing prospect of one or more hybrid zones where A1 and A2 populations overlap. Hybrid zones are well documented in plants and animals (Harrison 1993), but less common in microbial eukaryotes. This may be due to difficulty in detection rather than biological barriers to hybridization. Here, *P. parvum* A1×A2 hybrids are morphologically indistinguishable from nonhybrid strains and must be identified based on genetic characteristics. Moreover, A1- and A2-type strains are also morphologically identical to one another, and the true range of either parental population is unknown. Despite limited sampling, we know that different lineages of *P. parvum* type A can co-occur in nature. A1 and A1×A2 hybrid strains were co-isolated from the same water sample twice in Texas: once in 2010 from a bloom in Moss Creek Lake (Driscoll et al. 2013) and again in 2020

from a bloom in Possum Kingdom Lake (Table 3.1). Characterizing the genetic diversity of *P. parvum* s.l. present at a location based on a limited number of isolated strains may give a biased representation of true diversity. Culture-independent methods, such as single-cell sequencing, could be employed in future studies of *P. parvum* s.l. to determine where or if hybrid zones exist.

Rapid evolution of extensive segmental and chromosome-scale anomalies appears to be a common feature of hybrid genomes in microbial eukaryotes, with precedents in fungi, diatoms, trypanosomatids, and green algae (van den Broek et al. 2015; Inbar et al. 2019; Maeda et al. 2022; Craig et al. 2025). These anomalies can have functional consequences. For example, in lager brewing yeast strains, different aneuploid variants among hybrids were associated with distinct phenotypes in flocculation and diacetyl production (van den Broek et al. 2015). Heterosis, also known as hybrid vigor, where hybrid strains show increased growth compared to parental strains has been documented in natural hybrid fungal endophytes (Li et al. 2024) and demonstrated experimentally in *Saccharomyces* spp. (Bernardes et al. 2017). Conversely, some variants may be maladaptive or only tolerated under controlled laboratory conditions. Similar to observations of the green algal *Auxenochlorella* hybrid (Craig et al. 2025), we observed high levels of structural anomalies between single-cell isolates and their progenitor cultures, suggesting that genome instability may accumulate more readily in stable culture conditions than in natural environments. However, our analysis also identified anomalies shared between independently isolated strains, indicating some aneuploid and segmental variation is present in natural populations of *P. parvum*. Future studies are needed to determine how many of the chromosomal abnormalities observed in lab strains are also present in natural populations.

The number of structural anomalies identified in *P. parvum* clade A strains was striking and indicates that substantial genomic diversity exists within this group. Moreover, our estimates represent a lower bound on the true extent of structural variation present in laboratory strains. Reference bias is one important limitation as any sequence absent in UTEX3227, including A2-specific regions, will go undetected in our analysis. Additionally, we could only characterize anomalies as losses or duplications relative to the reference genome and could not characterize other types of structural variation such as translocations or inversions. Our results underscore the

need for additional chromosome-scale genome assemblies, including from A2 strains and A1xA2 hybrids, to fully reconstruct genome structural differences present in *P. parvum* clade A.

The functional and ecological consequences of hybridization in *P. parvum* s.l. remain poorly understood yet may be directly relevant to predicting and managing *P. parvum* s.l. harmful algal blooms given the prevalence of hybrid strains in North American bloom samples. Phenotypic variation is known to exist between clade A strains, including differences in prymnesin profiles (Binzer et al. 2019) as well as growth rates and feeding behavior among A1 and hybrid strains (Driscoll et al. 2013; Driscoll et al. 2023). No A-type prymnesin producing strain has been more extensively studied than hybrid strain UTEX2797, which has served as a model strain for the clade in numerous experiments. Notably, A-type prymnesins from UTEX2797 were found to be the most potent among tested strains (Varga et al. 2024), but two extracts from UTEX2797 sampled three years apart showed significant differences in prymnesin profiles and potency. Our findings suggest that genome instability during laboratory culture could contribute to such variation, raising important questions about the reproducibility of studies conducted across time with the same strain. More broadly, it remains unclear how representative observations made with UTEX2797 are of non-hybrid strains, or whether hybrid and non-hybrid populations differ in their propensity to form toxic blooms. Given the ongoing threat of *P. parvum* s.l. blooms to freshwater systems worldwide (Roelke et al. 2016; Macêdo et al. 2023) and the substantial genomic variation documented here among laboratory strains, understanding the relationship between genetic diversity, hybridization, and toxicity is an important topic for future investigation.

3.5 Materials and Methods

3.5.1 Taxon Sampling

Illumina sequence data was acquired for 38 *P. parvum* strains from cultures in the Wisecaver Lab (see DNA sequencing, read filtering, and read mapping). Of these strains, 28 were from culture collections, including the Algal Resource Collection (6 strains courtesy of Dr. Schonna Manning); the National Center for Marine Algae and Microbiota (2 strains); the Culture Collection of Algae at the University of Texas at Austin (5 strains, 2 courtesy of Dr. William Driscoll); the Norwegian Culture Collection of algae (3 strains courtesy of Dr. Tim Fallon); the Roscoff Culture Collection

(5 strains); the Kalmar Culture Collection (1 strain courtesy of Dr. Tim Fallon), and the Microbial Culture Collection at the National Institute of Environmental Studies, Japan (3 strains). Three additional strains were acquired from the personal collection of Dr. William Driscoll. See Supplemental Table B.1 for additional strain information.

The remaining 10 strains were isolated from 2019 and 2020 *P. parvum* s.l. blooms in Texas. Water samples from a 2019 bloom in Lake Colorado City and a 2020 bloom in Possum Kingdom Lake were collected by the Texas Parks and Wildlife Department (TPWD) as part of the agency's Golden Alga monitoring program and mailed courtesy of Greg Southard. To generate clonal cultures, water samples were passed through a MoFlo Astrios High Speed Cell Sorter (Beckman Coulter, Brea, CA, USA). A gated population was identified based on forward scatter and chlorophyll fluorescence of existing *P. parvum* s.l. pure culture, and positive events were sorted into 96-well plates, one event per well. To create the initial plate media, TPWD water samples were passed through a 0.22 μm polyethersulfone filter (CELLTREAT, Pepperell, MA, USA). Plates were grown at 20°C using a 12:12 light:dark cycle and irradiance of approximately 100 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$. After three weeks, any well consisting of *P. parvum* was retained as an isolate and transferred to 20 mL of L1-Si media in sterile 25 x 150 mm borosilicate glass test tubes. This resulted in 5 new isolates from the Lake Colorado City bloom and 5 new isolates from the Possum Kingdom Lake bloom.

3.5.2 Algal Culturing

L1-Si media was prepared using Instant Ocean Sea Salt (Spectrum Brands, Blacksburg, VA, USA) at a salinity of 6 and a L1-Si Media Kit obtained from the National Center for Marine Algae and Microbiota (East Boothbay, ME, USA). All cultures were maintained in a 12:12 light:dark cycle at 20°C. Media and salinities used for each strain can be found in Table B.1.

Cultures were periodically treated with antibiotics to reduce bacterial contamination (Guillard 2005). Two milliliters of dense ($>5 \times 10^5$ cells/mL) late-exponential phase cultures were transferred sequentially into media containing different antibiotics: first ampicillin (100 mg/L) and neomycin (100 mg/L) for 48 hours, followed by kanamycin (100 mg/L) and penicillin (100 mg/L) for another

48 hours, and chloramphenicol (100 mg/L) for a final 24 hours. Subsequently, 2 mL of culture was transferred to fresh L1-Si media and allowed to fully recover.

3.5.3 Single-cell reisolation

UTEX2797 and UTEX3226 cultures were grown to late exponential phase before single-cell reisolation via serial dilution, following established protocols for single-cell isolation of microalgae (Andersen and Kawachi 2005; Lu et al. 2025). Cells were serially diluted to 1 cell/mL in L1-Si media (salinity 6) and transferred to 96-well plates with 100 μ L of diluted culture per well. Cultures were allowed to recover, and wells were considered derived from a single cell if ≤ 9 wells per plate had growth.

3.5.4 DNA sequencing, read filtering, and read mapping

Cultures for DNA sequencing were grown in 100 mL L1-Si media in 250 mL glass Erlenmeyer flasks. For each sample, 50 mL of late-exponential phase *P. parvum* s.l. culture was spun down at 4000 x g for 10 minutes. Genomic DNA was extracted from cell pellets using Plant DNAzol (Thermo Fisher #10978021, Waltham, MA, USA) according to lab protocol (Wisecaver 2026). Sequencing libraries were constructed and sequenced to produce 2x150 bp paired-end reads using an Illumina NovaSeq 6000 by Novogene Corporation Inc. (Sacramento, CA, USA).

Illumina read quality was assessed with FastQC v0.11.9 (Babraham Bioinformatics 2011) and reads were trimmed as needed using FastP v0.20.1 with quality, low complexity, polyG, and polyX filtering enabled (Chen et al. 2018). Although antibiotic treatments were performed as necessary, all cultures were considered xenic. Therefore, reads derived from contamination were identified and removed via the following pipeline. For each sample, reads were first assembled with ABySS v2.3.1 (Simpson et al. 2009) using a k-mer size of 96. Assembly scaffolds were queried against the NCBI nucleotide (nt) database (accessed September 11th, 2021) using blastn v2.11.0+ (Camacho et al. 2009) and against a custom RefSeq database (O’Leary et al. 2016) using DIAMOND v2.0.8 (Buchfink et al. 2015). The custom RefSeq database was composed of RefSeq release 207 supplemented with predicted protein sequences from marine microbial eukaryotic transcriptome sequencing project (MMETSP) (Keeling et al. 2014) and the 1000 plants

transcriptome sequencing project (1KP) (Matasci et al. 2014). Reads were aligned against the ABySS assembly using minimap2 v2.24-r1122 (Li 2018). Assembly scaffolds were filtered using BlobTools v1.1.1 (Laetsch and Blaxter 2017). The BlobTools taxrule ‘bestsumorder’ determined the taxonomic assignment of each scaffold, prioritizing information from protein hits first. Scaffolds denoted as non-eukaryotic in origin were removed to produce a filtered assembly. BBSplit v38.87 from the BBTools software suite (Bushnell 2017) was used to retain reads which mapped to the filtered assembly (hereafter referred to as filtered reads).

Filtered short reads for all strains were aligned to the UTEX3227 v1.1 genome (GenBank assembly GCA_041296205.1; Wisecaver et al. 2023; Fallon et al. 2024) using BWA-MEM v0.7.17 (Li 2013). Read alignments were filtered to include only reads mapped in proper pairs with a minimum quality of 30 via SAMtools v1.9 (Li et al. 2009; Danecek et al. 2021). Secondary and chimeric alignments flagged by BWA-MEM v0.7.17 (Li 2013) were also removed. PCR duplicates were removed with GATK MarkDuplicates v4.2.6.1 (McKenna et al. 2010). Median read alignment rate was 90.4% and ranged from 78.2% (ARC82) to 96.6% (CCMP3037) (Table B.6).

3.5.5 Preliminary SNP phylogenetic network

A preliminary SNP network was constructed to first identify all *P. parvum* clade A strains for downstream analyses (see analysis flowchart Figure A.12). Variants were called against the UTEX3227 v1.1 genome using BCFtools v1.17 mpileup (Danecek et al. 2021) using BAM alignment files for all 38 strains (excluding single-cell reisolates) as input (Table B.1). Variants were first quality filtered using VCFtools 0.1.16 (Danecek et al. 2011) to retain only biallelic SNPs with a minimum quality of 100 and a minimum p-value of 1×10^{-6} . For constructing the phylogenetic network, biallelic SNPs were further filtered using VCFtools so that the final dataset included those present in all individuals, a minimum mean depth of 20, and a minimum minor allele frequency of 0.04 (i.e. at least two samples). The resulting 816,587 SNPs were used to generate a square distance matrix with PLINK v1.90b6.21 (Purcell et al. 2007). A network depicting this matrix was created using NeighborNet with default options in SplitsTree4 v4.19.2 (Huson 1998; Bryant 2003). Strains that grouped with strains known to produce A-type prymnesin (Binzer et al. 2019; Wisecaver et al. 2023) were included in all downstream analyses, with the

exception of ARC140. ARC140 was originally chemotyped as producing B-type prymnesin (Binzer et al. 2019), but clustered with type A strains in our analysis, so it was excluded.

3.5.6 Genome size estimation via flow cytometry

Exponentially growing *P. parvum* s.l. culture (500 µL) was centrifuged at 1000 x g for 10 minutes. The supernatant was discarded and the pellet flash-frozen in liquid nitrogen and resuspended in 500 µL ice-cold nuclear isolation buffer LB01 consisting of 15 mM tris, 2 mM Na₂EDTA, 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, and 0.1% (v/v) Triton X-100 at pH 8.0 (Dpooležel et al. 1989; Čertnerová and Galbraith 2021). Nuclei were kept at 4°C and analyzed within 24 hours.

All nuclei were treated with 0.4 mg/mL RNaseA (Invitrogen #12091021, Waltham, MA, USA) and 0.05 mg/mL propidium iodide (Invitrogen #P3566, Waltham, MA, USA) at 4°C for 3 hours in the dark. Following incubation, nuclei were passed through a 35 µm cell-strainer cap (Falcon #352235, Corning, NY, USA) and analyzed on a Cytex Northern Lights flow cytometer equipped with a 488-nm laser and 14 fluorescent detection channels (Cytex Biosciences, Fremont, CA, USA). Nuclei were identified via forward scatter and fluorescence in the B7 channel corresponding to peak propidium iodide fluorescence. Fluorescence values corresponding to at least 1000 nuclei per sample were gathered using the batch analysis feature in SpectroFlo (Cytex Biosciences, Fremont, CA, USA). Biological replicates were performed on separate days at approximately the same time. DNA content of UTEX3227, RCC3703, CCMP3037, UTEX2797, UTEX3226, and K0081 from Wisecaver et al., (2023) were used to create a standard curve to convert fluorescence into total DNA content in picograms of DNA for all other strains (Table B.2). Total DNA content was converted to base pairs using the formula of 1 pg = 0.98 x 10⁹ base pairs (Cavaller-Smith 1985).

3.5.7 Estimation of Heterozygosity

K-mers were counted from filtered reads using KMC v3.1.1 with a k-mer length (-k) of 21, minimal k-mer occurrence (-ci) of 1, and maximal k-mer occurrence (-cs) of 10,000 (Kokot et al. 2017). K-mer frequencies were modeled using GenomeScope2 v2.0 (Ranallo-Benavidez et al.

2020) using a k-mer length of 21 and a ploidy value of 1,2, or 4 depending on the strain (Table 3.1). Strains with heterozygosity above 1% and a pronounced heterozygous peak were flagged as high heterozygosity.

3.5.8 *P. parvum* clade A phylogenetic analysis

Variants were called against the UTEX3227 v1.1 genome using BCFtools v1.17 mpileup (Danecek et al. 2021), as described above using BAM alignment files for all clade A strains and single-cell reisolates (Table B.1). Variants were filtered using VCFtools 0.1.16 (Danecek et al. 2011) and used to create a SNP phylogenetic network as described above (see analysis flowchart Figure A.13). The resulting A-type only dataset was comprised on 453,986 SNPs.

To construct a gene-based species tree, cleaned short-read DNA libraries from non-hybrid A-types, B-types, and C-types were assembled into contigs via ABySS v2.3.1 (Simpson et al. 2009). Assemblies were masked with RepeatMasker v4.1.1 (Smit et al. 2017) using the modeled repetitive element library generated by Wisecaver et al., (2023). This masked genome was then used to create a hints file with BRAKER v2.1.5 (Hoff et al. 2019; Br  na et al. 2021) supplemented with evidence from Odb10-protzoa and all predicted haptophyte proteins from MMETSP (Keeling et al. 2014; Kriventseva et al. 2019). This hints file and an AUGUSTUS species-specific training file from Wisecaver et al., (2023) was used to predict coding sequences on all masked genomes with BRAKER v2.1.5 (Hoff et al. 2019; Br  na et al. 2021). The longest coding sequence for each locus was extracted using a custom python script and used input for OrthoFinder v2.5.5 (Emms and Kelly 2015; Emms and Kelly 2019). OrthoFinder identified 9836 single copy orthogroups present in all strains. Nucleotide coding sequences were aligned with MAFFT in the GUIDANCE v2.02 software suite (Kato   and Standley 2013; Sela et al. 2015). Alignments were trimmed using trimAl v1.4.1 (Capella-Gutierrez et al. 2009). All alignments smaller than 150 bp were discarded, leaving 9769 orthogroups for species tree construction. All alignments were concatenated and used to create a concatenated species tree via IQTree2 v2.2.2.9 (Nguyen et al. 2015; Minh et al. 2020). Size-filtered alignments were also used to create individual gene trees using IQTree2 v2.2.2.9 (Nguyen et al. 2015; Minh et al. 2020). The coalescence-based phylogeny estimation was conducted using ASTRAL v5.7.1 (Zhang et al. 2018). Gene and site concordance values were calculated for both the coalescent and concatenated species trees via IQTree2 v2.2.2.9 and

visualized via FigTree v1.4.4 (Nguyen et al. 2015; Rambaut 2018; Minh et al. 2020). See Figure A.14 for analysis flowchart.

3.5.9 Identification of Ancestry Informative Markers

Ancestry Informative Markers (AIMs) diagnostic of A1 or A2 ancestry were identified as previously described (Blom et al. 2024; Thörn et al. 2024). Sites that were fixed (i.e., $F_{ST} = 1$) between diploid A1 genomes (parental population 1) and A2 genomes (parental population 2) were identified with the Weir and Cockerham method using VCFtools v0.1.16 (Danecek et al. 2011). This identified 277,926 AIMs. The F_{ST} was then calculated between each parental population and all other clade A strains individually, and AIMs in these strains were classified based on the difference between the two F_{ST} values. An AIM was classified as A1/A1 homozygous if $F_{ST_vs_A1} = 0$ and $F_{ST_vs_A2} = 1$. An AIM was classified as A2/A2 homozygous if $F_{ST_vs_A1} = 1$ and $F_{ST_vs_A2} = 0$. Lastly, an AIM was classified as A1/A2 heterozygous if both values were non-zero and the absolute value (abv) of the difference was less than 0.2; no AIMs had intermediate values (abv > 0.2 and < 1).

3.5.10 Chloroplast SNP analyses

To identify reads that mapped preferentially to chloroplast genome, the UTEX3227_v1.1 nuclear genome was combined with the UTEX3227 chloroplast genome assembly from Wisecaver et al. (2023). Filtered short reads from A1, A2, and A1×A2 strains were aligned to the combined assembly using BWA-MEM v0.7.17 (Li 2013). Alignments were filtered to only include those mapped in proper pairs with a minimum quality of 30 via SAMtools (Li et al. 2009; Danecek et al. 2021). Supplementary and chimeric alignments as marked by BWA-MEM v0.7.17 (Li 2013) were also removed. PCR duplicates were removed with GATK MarkDuplicates v4.2.6.1. Variants were called using BCFtools v1.17 (Danecek et al. 2021) with minimum mapping and base qualities of 20. Variants were filtered using VCFtools 0.1.16 (Danecek et al. 2011) to only retain biallelic SNPs present in all individuals with a minimum quality of 30, a minimum minor allele frequency of 0.05, and a minimum mean depth of 20. See Figure A.15 for analysis flowchart. The resulting chloroplast dataset consisted of 188 SNPs.

To assess the partitioning of nuclear genetic variation among chloroplast haplotypes and geographic region of isolation, we conducted AMOVAs on the nuclear 453,986 SNP dataset and the chloroplast SNP dataset (Table B.5). AMOVAs were performed using the *pegas* function implemented in the *poppr* package (Kamvar et al., 2014) in R. Statistical significance of variance components was assessed by permutation tests ($n = 999$).

3.5.11 Structural variant identification

Normalized genome coverage (NC) was calculated for each A1, A2, and A1×A2 strain with *deepTools* v3.5.5 (Ramírez et al. 2016) using the filtered BAM alignment files as input, bin size of 10 kbp, RPGC normalization, and an effective genome size of 93,538,723 (corresponding to the length of the UTEX3227_v1.1 assembly). Uniformity of genome coverage was evaluated, and the five smallest scaffolds 30-34 were excluded from the coverage analysis due to increased variability in coverage both within and between strains. A1 diploid strains CC2192B5 and CC2192D3 were selected to as reference euploid strains, as these strains had the lowest coefficient of variation in coverage across chromosomes and no outlier chromosomes after scaffolds 30-34 were excluded. The $\log_2$ ratio of NC for each 10 kbp window was calculated as $\log_2\text{NC} = \log_2(x_i + 0.01) - \log_2(\mu_{\text{diploid}} + 0.01)$, where x_i was the NC for strain i in a bin and μ_{diploid} was the mean NC for the reference strains in the same bin.

A1 allele frequency (A1_AF) was calculated for all AIMs by extracting the A1 allele depth from the VCF file divided by the total read depth for that site. The mean $\log_2\text{NC}$ for 50 kbp windows and mean A1_AF for 20 kbp windows were plotted using *Matplotlib* (Hunter 2007). Diploid regions that have maintained both A1 and A2 parental alleles are expected to have an average $\log_2\text{NC} \approx 0$ and A1_AF ≈ 0.5 . Large regions that deviate from these expected values are indicative of structural variation. For example, a monosomic region resulting from a loss of the A1 haplotype would have half the coverage of diploid regions ($\log_2\text{NC} \approx -1$) and a corresponding drop in A1_AF to zero; a trisomic region resulting from a duplication of the A1 haplotype would have 1.5x coverage ($\log_2\text{NC} \approx 0.58$) and a corresponding 2-to-1 A1-to-A2 allele ratio (A1_AF ≈ 0.67). However, these numbers represent ideal values. Actual strain data displayed significantly more noise due to variation in read mapping quality across the genome and between haplotypes. Due to genome reference bias, reads derived from the A1 haplotype aligned better than A2 haplotype-

derived reads. This resulted in elevated average A1_AF (≈ 0.57) in background regions of the genome (Table B.4). Plots were manually inspected to identify regions of the genome (~ 50 kbp or larger) that deviated significantly from $\log_2\text{NC} = 0$ and $\text{A1_AF} = 0.57$.

3.6 References

- Andersen RA, Kawachi M. 2005. Traditional microalgae isolation techniques. In: *Algal Culturing Techniques*. Elsevier, Burlington, MA; p 83–100. <https://doi.org/10.1016/B978-012088426-1/50007-X>
- Anestis K et al. 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>
- Babraham Bioinformatics. 2011. FastQC A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bannon CC et al. 2024. Influence of biotic and abiotic factors on prymnesin profiles in three strains of *Prymnesium parvum*. *Algal Res.* 78:103390. <https://doi.org/10.1016/j.algal.2024.103390>
- Bernardes JP, Stelkens RB, Greig D. 2017. Heterosis in hybrids within and between yeast species. *J Evol Biol.* 30(3):538–548. <https://doi.org/10.1111/jeb.13023>
- Binzer SB et al. 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>
- Bizzarri M, Giudici P, Cassanelli S, Solieri L. 2016. Chimeric sex-determining chromosomal regions and dysregulation of cell-type identity in a sterile *Zygosaccharomyces allodiploid* yeast. *PLOS ONE.* 11(4):e0152558. <https://doi.org/10.1371/journal.pone.0152558>
- Blom MPK et al. 2024. Hybridization in birds-of-paradise: widespread ancestral gene flow despite strong sexual selection in a lek-mating system. *iScience.* 27(7):110300. <https://doi.org/10.1016/j.isci.2024.110300>
- Blossom HE et al. 2014. *Prymnesium parvum* revisited: relationship between allelopathy, ichthyotoxicity, and chemical profiles in 5 strains. *Aquat Toxicol.* 157:159–166. <https://doi.org/10.1016/j.aquatox.2014.10.006>
- Bousquet L et al. 2025. Life cycle and morphogenetic differentiation in heteromorphic cell types of a cosmopolitan marine microalga. *New Phytol.* 245(5):1969–1984. <https://doi.org/10.1111/nph.20360>

- Braun-Galleani S et al. 2018. *Zygosaccharomyces pseudobailii*, another yeast interspecies hybrid that regained fertility by damaging one of its MAT loci. *FEMS Yeast Res.* 18(7):foy079. <https://doi.org/10.1093/femsyr/foy079>
- Brosnahan ML et al. 2010. Outbreeding lethality between toxic Group I and nontoxic Group III *Alexandrium tamarense* spp. isolates: Predominance of heterotypic encystment and implications for mating interactions and biogeography. *Deep-Sea Res Part II Top Stud Oceanogr.* 57(3–4):175–189. <https://doi.org/10.1016/j.dsr2.2009.09.005>
- Brůna T et al. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma.* 3(1):lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Bryant D. 2003. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 21(2):255–265. <https://doi.org/10.1093/molbev/msh018>
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12(1):59–60. <https://doi.org/10.1038/nmeth.3176>
- Bushnell B. 2017. BBTools Software Package. <https://sourceforge.net/projects/bbmap/>
- Camacho C et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10(1):421. <https://doi.org/10.1186/1471-2105-10-421>
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Cavaller-Smith T. 1985. The evolution of genome size. John Wiley and Sons Inc., New York, NY. <https://www.osti.gov/biblio/6076917>
- Čertnerová D, Galbraith DW. 2021. Best practices in the flow cytometry of microalgae. *Cytometry A.* 99(4):359–364. <https://doi.org/10.1002/cyto.a.24328>
- Chávez Montes RA et al. 2024. Hormetic and transcriptomic responses of the toxic alga *Prymnesium parvum* to glyphosate. *Sci Total Environ.* 954:176451. <https://doi.org/10.1016/j.scitotenv.2024.176451>
- Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Craig RJ et al. 2025. Targeted genetic manipulation and yeast-like evolutionary genomics in the green alga *Auxenochlorella*. *Plant Cell.* 37(11):koaf259. <https://doi.org/10.1093/plcell/koaf259>
- Danecek P et al. 2011. The variant call format and VCFtools. *Bioinformatics.* 27(15):2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>

- Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*. 10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>
- Detwiler JT, Criscione CD. 2010. An infectious topic in reticulate evolution: introgression and hybridization in animal parasites. *Genes*. 1(1):102–123. <https://doi.org/10.3390/genes1010102>
- Dpooležel J, Binarová P, Lcretti S. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biol Plant*. 31(2):113–120. <https://doi.org/10.1007/BF02907241>
- Driscoll WW et al. 2023. Behavioural differences underlie toxicity and predation variation in blooms of *Prymnesium parvum*. *Ecol Lett*. 26(5):677–691. <https://doi.org/10.1111/ele.14172>
- Driscoll WW, Espinosa NJ, Eldakar OT, Hackett JD. 2013. Allelopathy as an emergent, exploitable public good in the bloom-forming microalga *Prymnesium parvum*. *Evolution*. 67(6):1582–1590. <https://doi.org/10.1111/evo.12030>
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 16(1):157. <https://doi.org/10.1186/s13059-015-0721-2>
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>
- Fallon TR et al. 2024. Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins. *Science*. 385(6709):671–678. <https://doi.org/10.1126/science.ado3290>
- Guillard R. 2005. Purification methods for microalgae. In: *Algal Culturing Techniques*. Elsevier, Burlington, MA; p 117–132. <https://doi.org/10.1016/B978-012088426-1/50009-3>
- Harrison RG. 1993. Hybrid zones and the evolutionary process. Oxford University Press, New York. <https://doi.org/10.1093/oso/9780195069174.001.0001>
- Hoff K, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol Clifton NJ*. 1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5
- Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 9(3):90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Hunter N, Chambers SR, Louis EJ, Borts RH. 1996. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J*. 15(7):1726–1733. <https://doi.org/https://doi.org/10.1002/j.1460-2075.1996.tb00518.x>
- Huson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*. 14(1):68–73. <https://doi.org/10.1093/bioinformatics/14.1.68>

- Inbar E et al. 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in *Leishmania*. *PLoS Genet.* 15(5):e1008042. <https://doi.org/10.1371/journal.pgen.1008042>
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. <https://doi.org/10.1093/molbev/mst010>
- Keeling PJ et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* 12(6):e1001889. <https://doi.org/10.1371/journal.pbio.1001889>
- Kim JH et al. 2020. Sexual reproduction and genetic polymorphism within the cosmopolitan marine diatom *Pseudo-nitzschia pungens*. *Sci Rep.* 10(1):10653. <https://doi.org/10.1038/s41598-020-67547-9>
- Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics.* 33(17):2759–2761. <https://doi.org/10.1093/bioinformatics/btx304>
- Kriventseva EV et al. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 47(D1):D807–D811. <https://doi.org/10.1093/nar/gky1053>
- Kuhl H et al. 2024. The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins. *Curr Biol.* 34(16):3698–3706. <https://doi.org/10.1016/j.cub.2024.06.033>
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Research.* 6:1287. <https://doi.org/10.12688/f1000research.12232.1>
- Li H et al. 2009. The sequence alignment/map format and SAMtools. *Bioinforma Oxf Engl.* 25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [accessed 2023 Oct 11]. <http://arxiv.org/abs/1303.3997>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences *Birol I*, editor. *Bioinformatics.* 34(18):3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li Z et al. 2024. Intraspecific diploidization of a halophyte root fungus drives heterosis. *Nat Commun.* 15(1):5872. <https://doi.org/10.1038/s41467-024-49468-7>
- Lu Z et al. 2025. Coevolution of marine phytoplankton and *Alteromonas* bacteria in response to pCO₂ and coculture. *ISME J.* 19(1):wrae259. <https://doi.org/10.1093/ismejo/wrae259>

- Lysgaard ML, Eckford-Soper L, Daugbjerg N. 2018. Growth rates of three geographically separated strains of the ichthyotoxic *Prymnesium parvum* (Prymnesiophyceae) in response to six different pH levels. *Estuar Coast Shelf Sci.* 204:98–102. <https://doi.org/10.1016/j.ecss.2018.02.030>
- Macêdo RL, Haubrock P, Rocha O. 2023. Towards effective management of the marine-origin *Prymnesium parvum* (Haptophyta): a growing concern in freshwater reservoirs? *Harmful Algae.* 102513. <https://doi.org/10.1016/j.hal.2023.102513>
- Maeda Y et al. 2022. Chromosome-scale genome assembly of the marine oleaginous diatom *Fistulifera solaris*. *Mar Biotechnol.* 24(4):788–800. <https://doi.org/10.1007/s10126-022-10147-7>
- Matasci N et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience.* 3(1):17. <https://doi.org/10.1186/2047-217X-3-17>
- Matos GM et al. 2022. Microevolution of *Trypanosoma cruzi* reveals hybridization and clonal mechanisms driving rapid genome diversification. *eLife.* 11:e75237. <https://doi.org/10.7554/eLife.75237>
- McKenna A et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Minh BQ et al. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Montecinos AE et al. 2017. Hybridization between two cryptic filamentous brown seaweeds along the shore: analysing pre- and postzygotic barriers in populations of individuals with varying ploidy levels. *Mol Ecol.* 26(13):3497–3512. <https://doi.org/10.1111/mec.14098>
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274. <https://doi.org/10.1093/molbev/msu300>
- Nomaguchi T et al. 2018. Homoeolog expression bias in allopolyploid oleaginous marine diatom *Fistulifera solaris*. *BMC Genomics.* 19(1):330. <https://doi.org/10.1186/s12864-018-4691-0>
- Obert T, Zhang T, Rurik I, Vďačný P. 2022. First molecular evidence of hybridization in endosymbiotic ciliates (Protista, Ciliophora). *Front Microbiol.* 13:1067315. <https://doi.org/10.3389/fmicb.2022.1067315>
- O’Leary NA et al. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44(D1):D733–D745. <https://doi.org/10.1093/nar/gkv1189>

- Ortiz-Merino RA et al. 2017. Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLOS Biol.* 15(5):e2002128. <https://doi.org/10.1371/journal.pbio.2002128>
- Patiño R et al. 2023. Toxic algae in inland waters of the conterminous united states—a review and synthesis. *Water.* 15(15):2808. <https://doi.org/10.3390/w15152808>
- Purcell S et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575. <https://doi.org/10.1086/519795>
- Rambaut A. 2018. FigTree. GitHub repository. <https://github.com/rambaut/figtree>. <https://github.com/rambaut/figtree/releases>
- Ramírez F et al. 2016. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44(W1):W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 11(1):1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rashel RH, Patiño R. 2017. Influence of genetic background, salinity, and inoculum size on growth of the ichthyotoxic golden alga (*Prymnesium parvum*). *Harmful Algae.* 66:97–104. <https://doi.org/10.1016/j.hal.2017.05.010>
- Rasmussen SA et al. 2016. Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J Nat Prod.* 79(9):2250–2256. <https://doi.org/10.1021/acs.jnatprod.6b00345>
- Richardson ET, Patiño R. 2021. Growth of the harmful alga, *Prymnesium parvum* (*Prymnesiophyceae*), after gradual and abrupt increases in salinity. *J Phycol.* 57(4):1335–1344. <https://doi.org/10.1111/jpy.13172>
- Roelke DL, Manning SR. 2018. Harmful algal species fact sheet: *Prymnesium parvum* (Carter) “golden algae.” In: *Harmful Algal Blooms*. John Wiley & Sons, Ltd; p 629–632. <https://doi.org/10.1002/9781118994672.ch16q>
- Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43(W1):W7–W14. <https://doi.org/10.1093/nar/gkv318>
- Simpson JT et al. 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19(6):1117–1123. <https://doi.org/10.1101/gr.089532.108>
- Smit AF, Hubley R, Green P. 2017. RepeatMasker Open-4.0. <http://repeatmasker.org>
- Sobieraj J, Metelski D. 2023. Insights into toxic *Prymnesium parvum* blooms as a cause of the ecological disaster on the Odra River. *Toxins.* 15(6):403. <https://doi.org/10.3390/toxins15060403>

- Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588. <https://doi.org/10.1146/annurev.arplant.043008.092039>
- Steensels J, Gallone B, Verstrepen KJ. 2021. Interspecific hybridization as a driver of fungal evolution and adaptation. *Nat Rev Microbiol.* 19(8):485–500. <https://doi.org/10.1038/s41579-021-00537-4>
- Tábora-Sarmiento S, Patiño R, Portillo-Quintero C, Coldren C. 2022. Air, land, and water variables associated with the first appearance and current spatial distribution of toxic *Prymnesium parvum* blooms in reservoirs of the southern great plains, USA. *Sci Total Environ.* 836:155567. <https://doi.org/10.1016/j.scitotenv.2022.155567>
- Talarski A, Manning SR, La Claire JW. 2016. Transcriptome analysis of the euryhaline alga, *Prymnesium parvum* (Prymnesiophyceae): Effects of salinity on differential gene expression. *Phycologia.* 55:33–44. <https://doi.org/10.2216/15-74.1>
- Thörn F et al. 2024. Contemporary intergeneric hybridization and backcrossing among birds-of-paradise. *Evol Lett.* 8(5):680–694. <https://doi.org/10.1093/evlett/qrae023>
- Umen J, Coelho S. 2019. Algal sex determination and the evolution of anisogamy. *Annu Rev Microbiol.* 73:267–291. <https://doi.org/10.1146/annurev-micro-020518-120011>
- van den Broek M et al. 2015. Chromosomal copy number variation in *Saccharomyces pastorianus* Is evidence for extensive genome dynamics in industrial lager brewing strains. *Appl Environ Microbiol.* 81(18):6253–6267. <https://doi.org/10.1128/AEM.01263-15>
- Varga E 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>
- Wisecaver J. 2026. Wisecaver Lab algal DNA extraction with plant DNAzol reagent. <https://dx.doi.org/10.17504/protocols.io.kxygxp4dol8j/v1>
- Wisecaver JH et al. 2023. Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. *Curr Biol.* 33(11):2246–2259.e8. <https://doi.org/10.1016/j.cub.2023.05.003>
- Xu M, He X. 2011. Genetic incompatibility dampens hybrid fertility more than hybrid viability: yeast as a case study. *PLOS ONE.* 6(4):e18341. <https://doi.org/10.1371/journal.pone.0018341>
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 19(6):153. <https://doi.org/10.1186/s12859-018-2129-y>

4. HARMFUL ALGAL BLOOMS OF *PRYMNESIUM PARVUM SENSU LATO* IN TEXAS CONSIST OF MULTIPLE CRYPTIC SPECIES

4.1 Abstract

Harmful algal blooms are a growing threat to global ecosystems. *Prymnesium parvum sensu lato* (s.l.) is a group of unicellular haptophyte algae which are known for forming highly ichthyotoxic blooms. *P. parvum* s.l. is a cryptic species complex with member species that differ in genome size, gene content, and toxin production. Cryptic species of *P. parvum* s.l. form three main clades called A, B, and C-types each of which produce a unique variant of the toxin prymnesin. To determine whether cryptic species of *P. parvum* can co-occur within a geographic region or an individual bloom, we surveyed nine water samples from 2018-2020 taken from seven reservoirs across central Texas, generating 54 new *P. parvum* s.l. isolates. We genetically and metabolically characterized isolates using Illumina whole-genome sequencing and HPLC-QTOF-MS untargeted metabolomics. A majority (n=41) of recent isolates were B-types, whilst the remainder (n=13) were A-types. Two Texas water samples contained both A- and B-type *P. parvum*, indicating that the two cryptic species can co-occur within a single bloom. Texas B-types also have segmental and aneuploid chromosomal anomalies corresponding to gains and losses of DNA. These findings indicate that the extensive genome diversity observed within the *P. parvum* species complex is also reflected in its blooms and local populations. Future investigations are needed to understand the eco-evolutionary mechanisms that create and maintain exceptionally high levels of diversity in the *P. parvum* species complex.

4.2 Introduction

Algal blooms threaten global aquatic ecosystems and water quality (Liu and Klaiber, 2023; Sobieraj and Metelski, 2023). Although algal blooms are not new, recently they have become increasingly damaging due to factors such as climate change and anthropogenic nutrient loading (Hallegraeff et al., 2021; Kazmi et al., 2022). One type of toxic algae, *Prymnesium parvum sensu lato* (s.l.), is notable due to its highly ichthyotoxic blooms (Sobieraj and Metelski, 2023; Southard et al., 2010). *P. parvum* s.l. blooms devastate local ecosystems, exemplified by a large bloom in the Oder River which killed 50% of local fish populations (Mora et al., 2024; Sobieraj and Metelski,

2023). *P. parvum* s.l. blooms are also expanding in range, frequency, and intensity (Roelke et al., 2016; Tábora-Sarmiento et al., 2022).

Binzer et al. (2019) found that *P. parvum* s.l. is a species complex composed of at least three cryptic member species. These species are called A-, B-, or C-types based on the form of the polyketide toxin prymnesin they produce (Binzer et al., 2019). Two of the three member species, A-types and C-types, are further divided into subclades that may also be cryptic member species (Watervoort et al., 2026; Wisecaver et al., 2023).

In the United States, blooms are the most frequent and severe in Texas (Southard et al., 2010). Previously, all publicly accessioned strains of *P. parvum* s.l. from Texas were chemically characterized as A-types (Binzer et al., 2019; Rasmussen et al., 2016; Wisecaver et al., 2023). However, ITS-based phylogenies suggest that B-type *P. parvum* are also present in Texas (Driscoll et al., 2023; Lutz-Carrillo et al., 2010). These studies examined two Texas B-type strains, which were both sister to European B-types (Driscoll et al., 2023; Lutz-Carrillo et al., 2010). However, neither of the two Texas B-type strains were accessioned, so the only genetic information known about Texas B-types are ITS sequences. The divergence between American and European B-types on a whole-genome level, and the overall genetic diversity of Texas B-types, is unknown.

The first documented *P. parvum* s.l. bloom occurred in the Netherlands in 1920 (Liebert and Deerns, 1920). The first documented *P. parvum* s.l. bloom in the United States occurred in 1985 along the Pecos River in Texas, which killed millions of fish and mollusks (Linam et al., 1991; Southard et al., 2010). Since then, *P. parvum* s.l. blooms in Texas have become more intense and have caused managers to stop stocking sport fish in certain lakes (Brooks et al., 2011). The direct and measurable economic damage of *P. parvum* s.l. blooms in Texas was estimated at 13 million USD in 2010, although total damages are likely much greater (Southard et al., 2010). *P. parvum* s.l. blooms have been detected in 23 total states, stretching from the western to eastern US seaboards and Hawaii (Patiño et al., 2023; Roelke et al., 2016). *P. parvum* s.l. blooms are associated with higher atmospheric carbon dioxide concentrations, water hardness, and salinity (Tábora-Sarmiento et al., 2022; VanLandeghem et al., 2015). However, *P. parvum* s.l. blooms are difficult to predict, and *P. parvum* s.l. bloom models fail to capture seasonal dynamics in both summer and winter (Grover et al., 2010). The reasons why *P. parvum* s.l. blooms are difficult to

predict are not clear, but may include variable strain-level or clade-level growth preferences, *P. parvum* s.l.'s poorly understood life cycle, and interspecific competition (Grover et al., 2010).

We surveyed local *P. parvum* s.l. populations in Texas and generated 54 new isolates. We examined these isolates using whole-genome sequencing, untargeted metabolomics, and flow cytometry to characterize the genetic and metabolic diversity of local populations of *P. parvum* s.l. in Texas.

4.3 Methods

4.3.1 Sampling of recent Texas blooms

The Texas Parks and Wildlife Department (TPWD) collected water samples as part of the agency's *P. parvum* s.l. (golden alga) monitoring program and shipped them courtesy of Greg Southard (Table 4.1). TPWD sampled two waterbodies (Lake Possum Kingdom and Lake Granbury) in multiple locations at the same time and two different waterbodies (Lake E.V. Spence and Lake Colorado City) twice across two years (2019 and 2020). TPWD determined community composition as "*P. parvum* s.l. dominant" if a majority of cells were *P. parvum* s.l., otherwise community compositions were "mixed algal communities." TPWD also reported the ichthyotoxicity of each sample via ichthyotoxicity units (ITUs) which were found using a *Pimephales promelas* (fathead minnow) assay. Four *P. promelas* fry were subjected to an environmental water sample for two hours in three conditions: A: undiluted water sample + cofactor (0.003 M DADPA/0.02 M Tris in dH₂O, pH=9), B: 20% water sample + cofactor, C: undiluted water sample. An ITU of 0 indicates no mortality, one ITU is complete mortality in condition A only, five ITUs is complete mortality in conditions A and B, and 25 ITUs is complete mortality in all three conditions. Incomplete mortality in any condition indicates an intermediate ITU value (e.g., intermediate mortality in condition B with complete mortality in condition A is an ITU of one to five).

Table 4.1 Water sample metadata.

Sample ID	Date	Waterbody	Site	Cell Count (cells/mL)	Toxicity (ITUs)	Algal Community
CT2072	11/27/18	ComancheTrailsPark	Odessa	9,000	≥ 25	Not Indicated
BP2119	2/4/19	BealPark	Midland	19,000	≥ 25	Mixed Algal Community
CR2175	4/9/19	ConchoRiver	FM1692	9,000	1	Mixed Algal Community
MC2191	12/3/19	MossCreek	Boatramp	10,000	1 – 5	Mixed Algal Community
CC2192	12/3/19	ColoradoCity	StatePark	27,000	≥ 25	<i>P. parvum</i> s.l. Dominant
ES2194	12/3/19	EVSpence	LakeviewPark	12,000	1 – 5	Mixed Algal Community
CC2255	2/17/20	ColoradoCity	StatePark	17,000	1	Mixed Algal Community
ES2259	2/17/20	EVSpence	LakeviewPark	18,000	≥ 25	Mixed Algal Community
PK2294	6/1/20	PossumKingdom	1–RockCreek	19,000	5	<i>P. parvum</i> s.l. Dominant
PK2296	6/1/20	PossumKingdom	3–WillowBeach	9,000	1	<i>P. parvum</i> s.l. Dominant
PK2297	6/1/20	PossumKingdom	4–MainLake	19,000	≥ 25	<i>P. parvum</i> s.l. Dominant
PK2298	6/1/20	PossumKingdom	5–CutMainLake	17,000	≥ 25	<i>P. parvum</i> s.l. Dominant
PK2299	6/1/20	PossumKingdom	6–DeepElm	11,000	1	<i>P. parvum</i> s.l. Dominant
PK2300	6/1/20	PossumKingdom	7–PP	7,000	1	Mixed Algal Community
GR2308	6/2/20	Granbury	HunterPark	5,000	5	Mixed Algal Community
GR2309	6/2/20	Granbury	HolidayEstates	12,000	≥ 25	Mixed Algal Community
GR2310	6/2/20	Granbury	Dam	1,000	1	Mixed Algal Community

4.3.2 Isolation and maintenance of algal cultures

Water samples were processed by a MoFlo Astrios High Speed Cell Sorter to generate clonal samples (Beckman Coulter, Brea, CA, USA). Events with similar forward scatter and chlorophyll fluorescence to *P. parvum* s.l. pure culture were sorted into 96-well plates, one event per well. Cells were sorted into media generated by passing TPWD water samples through a 0.22 μm polyethersulfone filter (CELLTREAT, Pepperell, MA, USA). Plates were incubated at 20 °C using a 12:12 light:dark cycle and irradiance of approximately 100 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$. After three weeks, any well with purely *P. parvum* s.l. cells were retained as an isolate and transferred into 20 mL of L1-Si media in sterile 25 x 150 mm borosilicate glass test tubes. Isolates were then maintained in L1-Si media made using an L1-Si Media Kit from the National Center for Marine Algae and Microbiota (East Boothbay, ME, USA). Isolate IDs (e.g. PK2298B4) are a prefix corresponding to the water sample they were collected from (PK2298; Table 4.1), and an isolate-unique letter and number (B4). All isolates with the same prefix were isolated from the same approximately 50 mL sample of water.

In addition to the new Texas isolates, 29 accessioned strains of *P. parvum* s.l. were acquired from culture collections. All cultures were maintained in a 12:12 light:dark cycle at 20 °C; media, salinities, and sequencing date for each strain can be found in Table B.1.

P. parvum s.l. cultures were periodically treated with antibiotics to reduce bacterial contamination (Guillard, 2005). Two milliliters of dense ($>5 \times 10^5$ cells/mL) late-exponential phase cultures were transferred sequentially into media containing three different antibiotic cocktails: ampicillin (100 mg/L) and neomycin (100 mg/L) for 48 hours, followed by kanamycin (100 mg/L) and penicillin (100 mg/L) for another 48 hours, and chloramphenicol (100 mg/L) for a final 24 hours. Subsequently, two mL of culture was transferred to fresh L1-Si media and allowed to fully recover.

4.3.3 Flow cytometry for DNA content estimation

Exponentially growing *P. parvum* s.l. culture (1 mL) was centrifuged at 1000 x g for ten minutes. The supernatant was discarded and the pellet flash-frozen in liquid nitrogen before resuspension in one mL ice-cold LB01 buffer (15 mM Tris, two mM Na_2EDTA , 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, 0.1% (v/v) Triton X-100, pH=8.0) (Čertnerová and

Galbraith, 2021; Dpooležel et al., 1989). Nuclei of *Chlorella vulgaris* and *Arabidopsis thaliana* served as known standards.

An exponentially growing *C. vulgaris* culture was vortexed briefly to resuspend cells before taking a one mL aliquot. Cells were pelleted at 1000 x g for ten minutes, resuspended in one mL ice-cold LB01 buffer, and transferred to two mL ZR BashingBead™ Lysis Tubes (Zymo Research Cat# S6003-50, Irvine, CA, USA). Cell suspensions were agitated using a Disruptor Genie (Scientific Industries Model # SI-D238, Bohemia, NY, USA) for five minutes and transferred to a new tube.

For *A. thaliana* samples, 1.5 mL of ice-cold LB01 was added to 100 mg of *A. thaliana* Col-0 rosette tissue and chopped on ice for three minutes. The chopped solution was passed through a 40 µm sterile cell strainer (Fisherbrand #22363547, Pittsburgh, PA, USA). All nuclei were kept at 4 °C until use and were always analyzed on the same day as they were harvested.

All nuclei were treated with 0.4 mg/mL RNaseA (Invitrogen #12091021, Waltham, MA, USA) and 0.05 mg/mL propidium iodide (Invitrogen #P3566, Waltham, MA, USA) at 4 °C for three hours in the dark. Following incubation, nuclei were passed through a 35 µm cell-strainer cap (Falcon #352235, Corning, NY, USA) and subsequently analyzed on a BD Fortessa Cell Analyzer (BD Biosciences, Franklin Lakes, NJ, USA) with a flow rate of “Low.” Fluorescence values were gathered using a 488 nm laser and bandpass filters of 586/15 and 710/50 nm. Data from at least 300 nuclei were collected per sample and were processed using the FlowCal v1.3 Python library (Castillo-Hair et al., 2016). Fluorescence values in the 586/15 region for all genome standards were used to generate a line of best fit via linear regression, which was used to convert fluorescence values of *P. parvum* s.l. nuclei in the same fluorescence region into picograms of DNA. Total DNA content in picograms was converted to base pairs using the formula of 1 pg = 0.98 x 10⁹ base pairs (Cavaller-Smith, 1985).

4.3.4 DNA sequencing

P. parvum s.l. cultures for DNA sequencing were grown in 100 mL L1-Si media in 250 mL glass Erlenmeyer flasks. For each sample, 50 mL of late-exponential phase *P. parvum* s.l. culture was centrifuged at 4000 x g for ten minutes. Genomic DNA was extracted using one of two methods

(Table S1): Plant DNAzol (Thermo Fisher #10978021, Waltham, MA, USA) according to Wisecaver and Auber (2022), or Quick DNA Plant/Seed Miniprep (Zymo Research #D6020, Irvine, CA, USA) following manufacturer recommendations. Sequencing libraries were constructed by Novogene Corporation Inc. (Sacramento, CA, USA), and 2x150 bp paired-end reads were sequenced using an Illumina NovaSeq 6000.

4.3.5 Read filtering

Reads were checked for any major quality issues using FastQC v0.11.9 (Babraham Bioinformatics, 2011) before being trimmed by FastP v0.20.1 (Chen et al., 2018). Although antibiotic treatments were performed as necessary, all cultures were considered xenic. Therefore, reads derived from contamination were identified and removed through a multi-pass assembly pipeline. For each sample, reads were first assembled with ABySS v2.3.1 (Simpson et al., 2009) using a k-mer size of 96. Assembly scaffolds were queried against the NCBI nucleotide (nt) database (accessed September 11th, 2021) using blastn v2.11.0+ (Camacho et al., 2009) and against a custom RefSeq database (O’Leary et al., 2016) using DIAMOND v2.0.8 (Buchfink et al., 2015). The custom RefSeq database was composed of RefSeq release 207 supplemented with predicted protein sequences from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (Keeling et al., 2014) and the 1000 Plants Transcriptome Sequencing Project (1KP) (Matasci et al., 2014). First-pass ABySS assembly scaffolds were filtered using BlobTools v1.1.1 (Laetsch and Blaxter, 2017). The BlobTools taxrule ‘bestsumorder’ determined the taxonomic assignment of each scaffold, prioritizing information from protein hits first. Scaffolds denoted as non-eukaryotic in origin were removed to produce a filtered assembly. Reads that mapped against the BlobTools-filtered assembly (hereafter referred to as filtered Illumina reads) were retained using BBSplit v38.87 from the BBTools software suite (Bushnell, 2017).

K-mers were counted from filtered reads using KMC v3.1.1 with a k-mer length (-k) of 21, minimal k-mer occurrence (-ci) of 1, and maximal k-mer occurrence (-cs) of 10,000 (Kokot et al., 2017). K-mer frequencies were modeled using GenomeScope2 v2.0 (Ranallo-Benavidez et al., 2020) using a k-mer length of 21 and a ploidy value of 1,2, or 4 depending on the strain (see Table B.13). Strains with heterozygosity above 1% and a pronounced heterozygous peak were flagged as high heterozygosity.

4.3.6 SNP calling and network creation

Filtered Illumina reads were aligned against the 12B1 (UTEX3227) v1.1 genome (Fallon et al., 2024; Wisecaver et al., 2023) using BWA-MEM (Li, 2013). Note that hereafter, we refer to strain 12B1 (Wisecaver et al., 2023) by its UTEX accession, UTEX3227, to avoid confusion with the *P. parvum* B-type clade. Alignments were filtered using Samtools view to retain those with a quality score over 30 (Danecek et al., 2021; Li et al., 2009). SAM flags assigned by BWA MEM were also used to filter alignments by only retaining reads mapped in a proper pair (0x2) and excluding unmapped reads (0x4). Filtered Illumina read alignments were deduplicated with GATK MarkDuplicates (McKenna et al., 2010; Van der Auwera and O'Connor, 2020).

SNPs were called from deduplicated and filtered Illumina alignments using BCFtools mpileup and call. BCFtools mpileup used reads with minimum read and base qualities of 20 to generate the vcf and report allelic depth (AD), genotype depth (DP), and strand bias (SP). BCFtools call used the multiallelic caller and reported genotype quality (GQ) and genotype probability (GP) (Danecek et al., 2021). The vcf was filtered with VCFtools to remove indels and retain biallelic SNPs present in 95% of individuals. SNPs were filtered to have a minimum quality of 30, a minor allele frequency above 0.02, a minimum depth of 10, and follow Hardy-Weinberg Equilibrium with a P-value of at least 1E-6 (Danecek et al., 2011). A square distance matrix was generated using filtered SNPs with PLINK v1.9 (Purcell et al., 2007). This distance matrix was then used to generate a phylogenetic network using the NeighborNet algorithm as implemented in SplitsTree4 v4.19.2.

The vcf called against the UTEX3227 v1.1 assembly was converted into a multiple sequence alignment using vcf2phylip v2.8 with default options (Ortiz, 2019). This alignment was used to build a bifurcating tree via IQTree v1.6.12 with default options (Nguyen et al., 2015). This tree was then visualized and annotated using iTOLv7 (Letunic and Bork, 2024).

Read alignment, read filtering, read deduplication, and network creation were repeated otherwise identically for all B-type strains against the ODER1 haplotype 1 reference genome (Kuhl et al., 2024).

4.3.7 Chromosomal anomalies

Normalized read coverage for every Texas B-type isolate was calculated using filtered Illumina alignments against the ODER1 haplotype 1 reference using deepTools v3.5.5 with a bin size of 25kbp, RPGC normalization, and an effective genome size of 236,176,151 bp (Kuhl et al., 2024; Ramírez et al., 2016). The smallest chromosome (CM083889.1) was excluded from downstream analysis as its coverage was consistently highly variable. The American B-type diploid isolates PK2299B6, CC2255C5, and CT2072C4 were selected to act as euploid reference strains, because they had few outlier chromosomes, low average coverage variability, and came from different water samples. The $\log_2$ ratio of normalized coverage (NC) for each ten kbp window was calculated as $\log_2\text{NC} = \log_2(x_i + 0.01) - \log_2(\mu_{\text{diploid}} + 0.01)$, where x_i was the NC for strain i in a bin and μ_{diploid} was the mean NC for the reference strains in the same bin. An area of variable coverage was considered a chromosomal anomaly if it fulfilled three criteria: 1) was present in a non-variable state in at least one strain, 2) had a $|\log_2\text{NC}| \geq 0.3$, and 3) had a consistent pattern for at least 100 kb.

4.3.8 Chemotyping

P. parvum s.l. cultures were grown in 200 mL L1 -Si +NH₄Cl supplemented with 1.5 mM NaHCO₃. Once cells grew to late exponential phase (205,000-1,940,000 cells/mL) they were counted via phase counting chamber (Hausser Scientific #3100, Horsham, PA, USA) and harvested by vacuum filtration onto 47 mm Whatman GF/B glass fiber filters (Sigma-Aldrich #WHA1821047, St. Louis, MO, USA). Filters were stored in polypropylene tubes at -80 C until ready for extraction.

Filters were shaken with five mL ethyl acetate at 37 °C and 220 RPM for 20 minutes, and the extract was discarded. This step was repeated until the ethyl acetate was colorless. The filter was then shaken with five mL methanol at 37 °C and 220 RPM for 30 minutes. The filter and methanol were centrifuged at 4000 x g for ten minutes. The filter was then discarded, and the methanol extract evaporated to dryness. The residue was then resuspended in 300 µL methanol and passed through a 0.2 µm PES filter.

Separations were performed on an Agilent 1290 system (Agilent Technologies, Palo Alto, CA, USA), with a mobile phase flow rate of 0.4 mL/min. The prymnesins were assayed with a five μ L injection using a Waters Acquity BEH C18 column (1.7 μ m, 2.1 $\times$ 100 mm) (Waters # 186002352, Milford, MA, USA), where mobile phase A and B were 0.1% formic acid in ddH₂O and acetonitrile, respectively. Initial conditions were 90:10 A:B, followed by a linear gradient to 0:100 at 12 min, with an isocratic hold until 14 min. Column re-equilibration was performed by returning to 90:10 A:B at 15 minutes and holding until 21 min. The mass analysis was obtained in positive ionization mode using an Agilent 6546 Q-TOF (Agilent Technologies #G6546AA, Palo Alto, CA, USA) mass spectrometer with ESI capillary voltage of +4.0 kV. MS/MS was performed in Data Dependent Analysis (DDA) mode, with a range of 100 – 1700 m/z and fixed collision energies of 10, 20, and 40 eV. Mass accuracy was improved by infusing Agilent Reference Mass Correction Solution (G1969-85001).

Peak deconvolution, integration, and statistical analysis were performed using MS-DIAL (v. 4.9) [<http://prime.psc.riken.jp/compms/msdial/main.html>]. Peak annotations were performed using the MassBank of North America MS/MS library, based on authentic standards (v. 17) [<http://prime.psc.riken.jp/compms/msdial/main.html#MSP>]. Mass tolerances for identification were 0.005 Da for MS1 and 0.01 Da for MS2.

4.4 Results

4.4.1 Texas *Prymnesium parvum* s.l. blooms vary in cell density and toxicity

Texas Parks and Wildlife (TPWD) collected 17 water samples from eight waterbodies (Table 4.1, Figure 4.2). Community composition could vary within the same body of water, as seen in Lake Possum Kingdom. Sample PK2300 from Lake Possum Kingdom had a mixed algal community, while the other Lake Possum Kingdom samples (PK2294-PK2299) were *P. parvum* s.l. dominant. *P. parvum* s.l. cell densities ranged from 1000 to 27,000 cells/mL, and the toxicity of the water samples varied from low (1 ichthyotoxicity unit/ITU) to high toxicity ($\geq$ 25 ITUs; Table 4.1). Cell densities and toxicities could vary within the same body of water. Lake Possum Kingdom cell densities ranged from 7,000 to 19,000 *P. parvum* s.l. cells/mL, and one to 25 ITUs. Although cell densities broadly correlated with toxicity, some water samples with comparatively low cell density

were highly toxic, and others with comparatively high cell density were less toxic. For example, sample CT2072 from Comanche Trails Park had a moderate cell count of 9,000 *P. parvum* s.l. cells/mL and ≥ 25 ITUs (total fish death), while sample CC2255 from Lake Colorado City had a higher cell concentration with 17,000 *P. parvum* s.l. cells/mL and one ITU (minimal fish death).

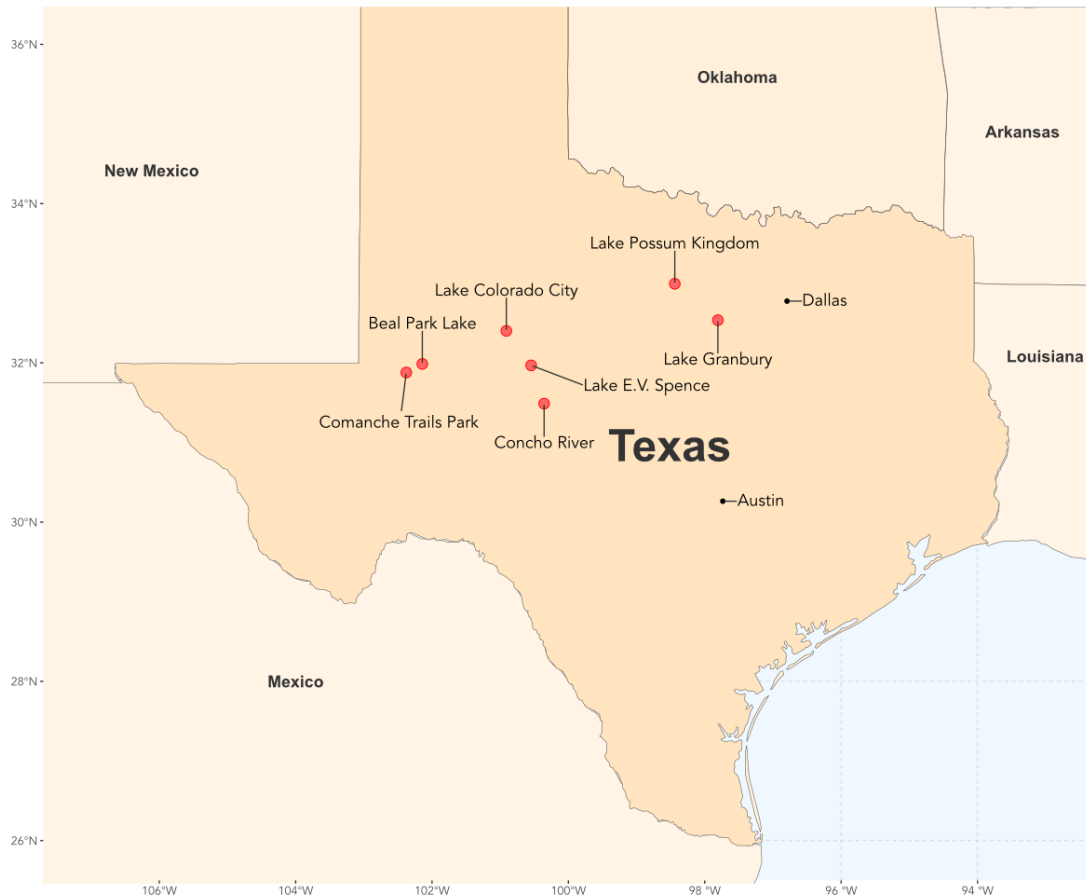


Figure 4.1 Location of the eight sampled bodies of water.

Lake Possum Kingdom, Lake Granbury, Lake Colorado City, and Lake E.V. Spence were sampled multiple times for this study.

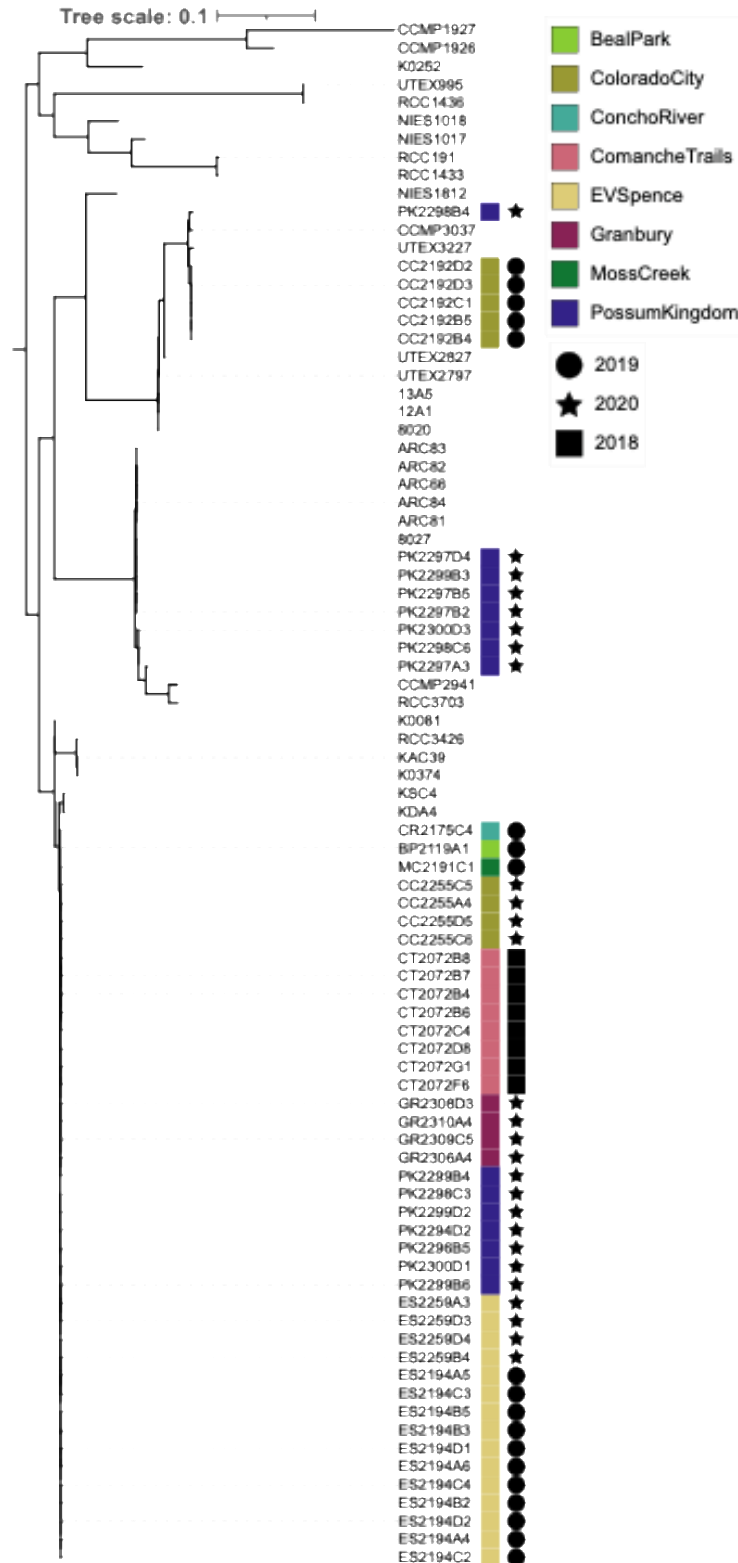


Figure 4.2 Bifurcating phylogenetic tree of *P. parvum* s.l. strains.

Phylogeny generated from 260,007 SNPs called against the UTEX3227 V1.1 reference assembly. The year and body of water each Texas strain was isolated in are indicated next to strain IDs.

4.4.2 A majority of strains from 2018-2020 Texas blooms are B-types

We aligned short reads from 85 *P. parvum* s.l. strains, including 54 new Texas isolates, against the UTEX3227 v1.1 (clade A1) reference genome to generate a phylogenetic network (Fallon et al., 2024; Wisecaver et al., 2023). We called single nucleotide polymorphisms (SNPs) from these alignments and generated a phylogenetic network from 260,007 SNPs. Previously genetically and chemically characterized A-, B-, and C-type strains grouped as expected, forming three main clades (Figure 4.3) (Binzer et al., 2019; Watervoort et al., 2026; Wisecaver et al., 2023). The A-, and C-type subclades (A1, A2, A3, A-hybrid, C1, C2, C3) were also regenerated, and contained previously genetically characterized strains (Watervoort et al., 2026; Wisecaver et al., 2023).

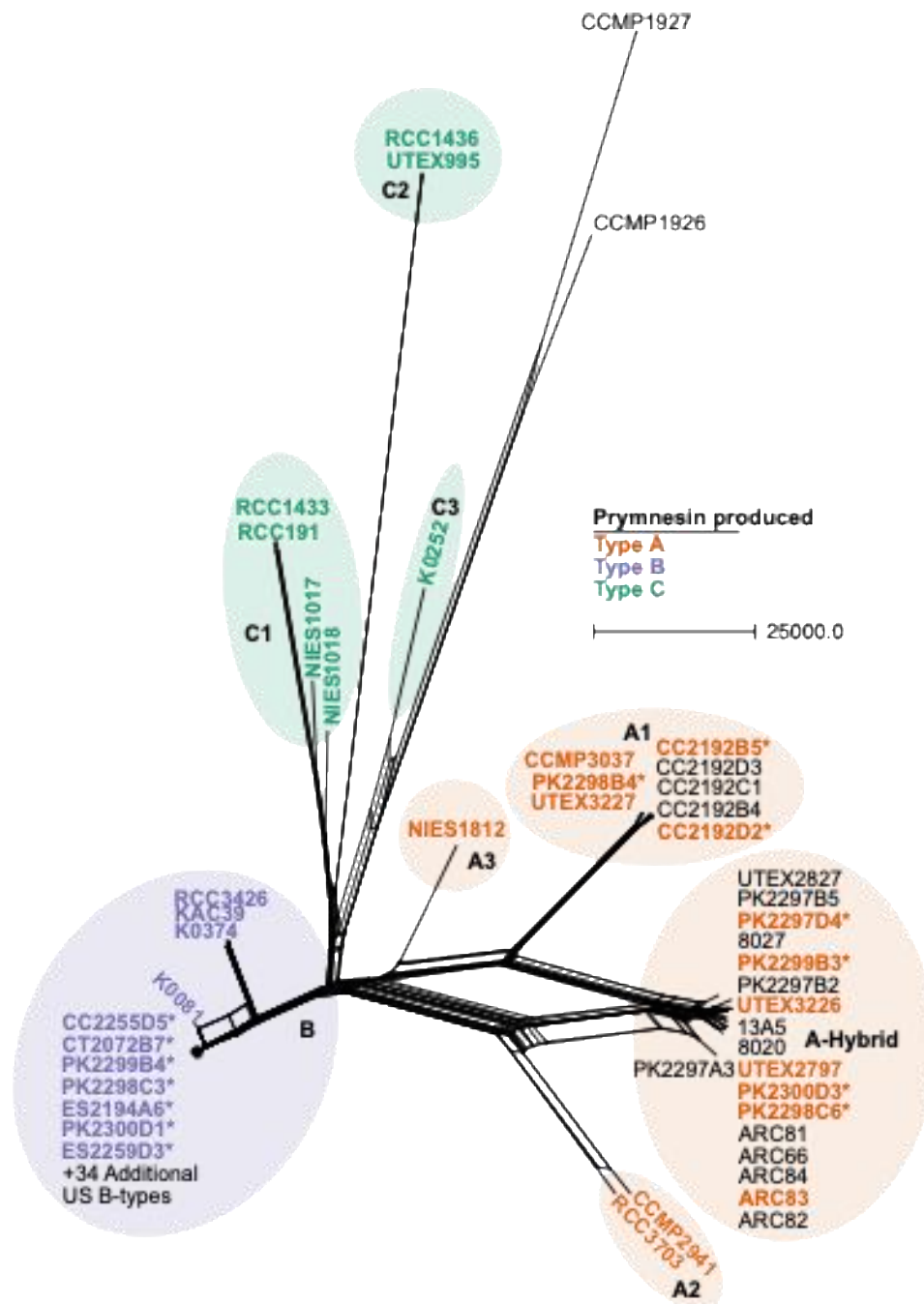


Figure 4.3 Phylogenetic network of *P. parvum* s.l. strains.

NeighborNet phylogenetic network built from 260,007 SNPs called against the UTEX3227 V1.1 reference assembly. Bolded strains have been chemically characterized before and are colored by prymnesin production. Clades are labeled based on identities from previous studies. *= characterized in this study.

Most (n=41) Texas isolates grouped within the B-type clade, with the remainder grouping in the A1 (n=6) or A-hybrid clades (n=7). No Texas isolates grouped in any C-type clades. To maximize phylogenetic resolution for the B-type clade, we re-analyzed B-type strains in a second phylogenetic network (Figure A.16) created from 136,987 SNPs called against the ODER1 B-type reference genome (Kuhl et al., 2024). Read filtering, alignment, and SNP filtering were otherwise identical to the network created with the UTEX3227 v1.1 genome. Texas B-types were more similar to each other than European B-types. Texas B-type isolates differ from other Texas B-types by 7,145 SNPs on average. In comparison, European B-type strains differ from other European B-types by 97,959 SNPs on average. European B-type strains grouped separately from American B-types.

All assayed Texas B-types had roughly diploid levels of DNA content with an average genome size of 658 Mbp (Figure 4.4). The genome sizes of Texas B-types also varied with a standard deviation of 24.92 Mbp. PK2299B4 had the largest genome size (725 Mbp), while PK2296B6 had the smallest (610 Mbp). Strain PK2299B4 also had high run-to-run variation in genome size, varying between 819 and 636 Mbp. Although there was significant strain-to-strain genome size variation (one-way ANOVA, $p=0.00982$), this was not driven by body of water (one-way ANOVA, $p=0.47$), year of isolation (one-way ANOVA, $p=0.26$), or gene content (linear regression, $p=0.37$; Figure A.17).

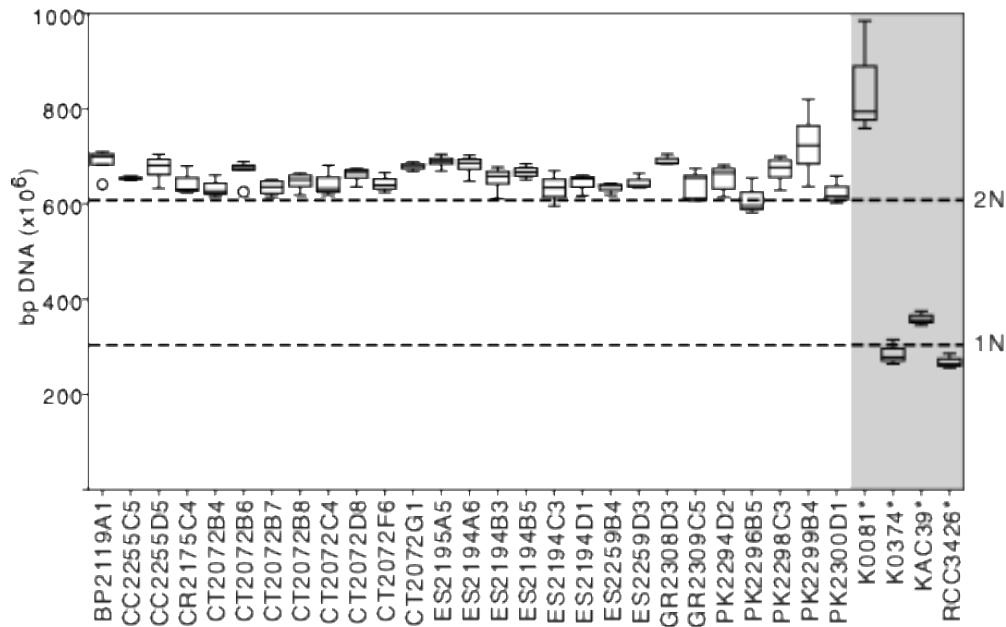


Figure 4.4 Flow cytometry estimates of genome sizes.

Estimates of B-type haploid (1N) and diploid (2N) genome sizes are from Wisecaver et al., (2023).

* flow cytometry data of B-types from Wisecaver et al., (2023).

Most of the A-type strains in this analysis were previously examined by Watervoort et al. (2026). However, three additional A-type isolates characterized here (PK2297A3, PK2297B2, PK2297B5) belonged to the A-hybrid clade (Watervoort et al., 2026). All Texas A-type strains were assigned to either the A1 or A-hybrid clades.

4.4.3 Multiple cryptic species can be found in the same bloom

Lake Possum Kingdom was sampled six times in 2020, yielding one A1 haploid (PK2298B4), seven A-Hybrid, and seven B-type isolates. Isolates from all three clades could be found in the same water sample (PK2298, Figure 4.2). PK2299 and PK2300 yielded both A-hybrid and B-type isolates, but no A1 isolates. The remaining Possum Kingdom 2020 samples yielded only B-type (PK2294, PK2296) or A-hybrid (PK2297) isolates. The other bloom that was sampled multiple times, Lake Granbury, only yielded B-type isolates (prefix GR2306-GR2310).

Two bodies of water (Lake Colorado City, Lake E.V. Spence) were sampled over several years. All isolates from Lake E.V. Spence (prefix ES2194/ES2259) belonged to the B-type clade. Isolates

from a Lake Colorado City sample taken in December of 2019 (prefix CC2192) were all A-types, while isolates from the same location taken in February of 2020 (prefix CC2192) were all B-types.

4.4.4 Recent Texas isolates vary in prymnesin production

We chemically characterized two A1, five A-hybrid, and seven B-type isolates with methods from Binzer et al. (2019). Total prymnesin abundances were normalized with cell counts obtained during cell harvesting for prymnesin extraction. All A- isolates produced A-type prymnesins, and all B-type isolates produced B-type prymnesins (Figure 4.5). At least one prymnesin was detected in every isolate (Table B.9, TableB.10, Figure 4.5).

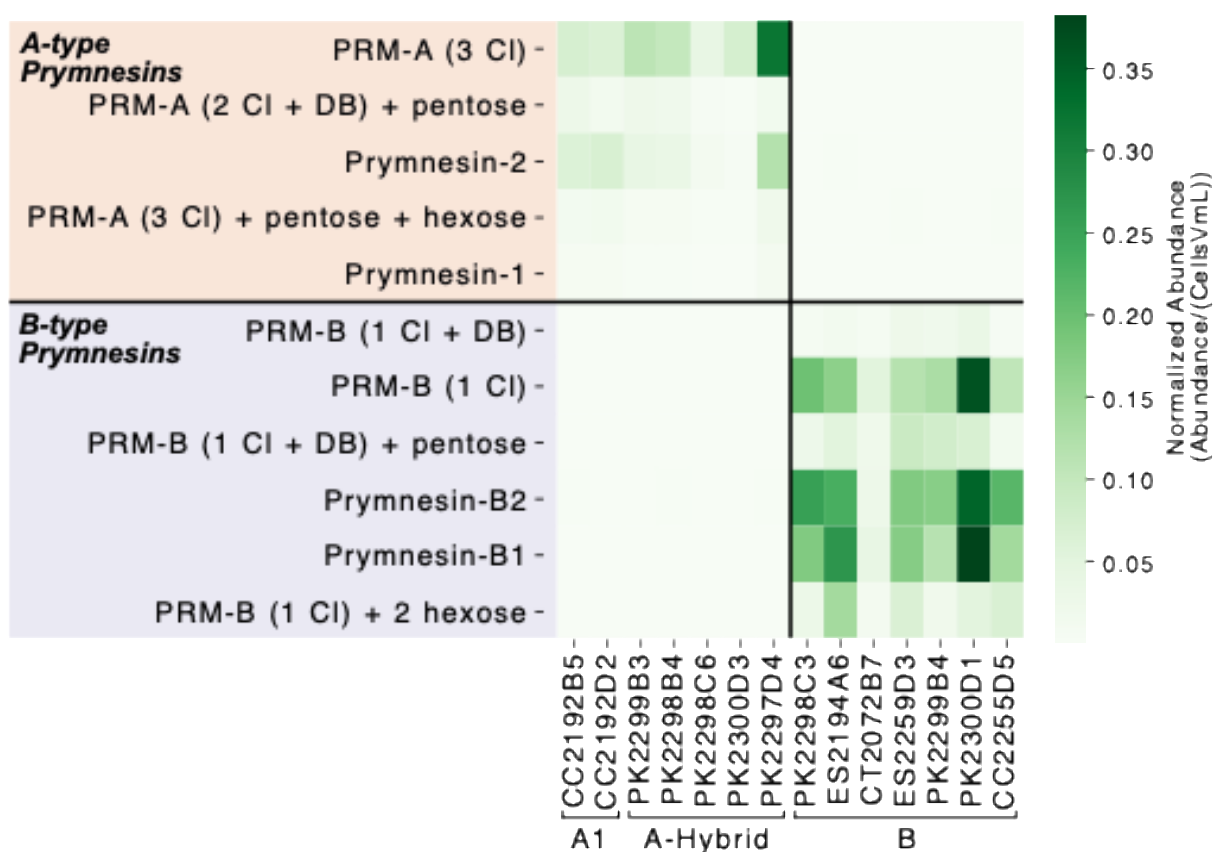


Figure 4.5 Chemical analysis of recent Texas strains.

Prymnesin abundances from 14 recent Texas strains identified via LC-MS following methods from Binzer et al., (2019). Abundances were normalized by dividing absolute abundances by cells/mL obtained immediately prior to prymnesin extraction.

We detected five different A-type prymnesins first identified by Binzer et al. (2019) in recent A-type Texas isolates (Figure 4.5). PRM-A (3 Cl) and Prymnesin-2 comprised 74% to 94% of the total A-type prymnesin abundances in each strain. We also detected six B-type prymnesins first identified by Binzer et al. (2019) in recent B-type Texas isolates (Figure 4.5). The three most abundant B-type prymnesins (PRM-B (1 Cl), Prymnesin-B2, Prymnesin-B1) comprised 72% to 91% of total B-type prymnesin abundances.

4.4.5 American B-types have chromosomal anomalies

Significant chromosomal anomalies, including aneuploidy (gain or loss of whole chromosomes) as well as large-scale structural mutations resulting in gain or loss of chromosomal segments, were identified in *P. parvum* A-type strains (Watervoort et al., 2026). To investigate whether similar anomalies are present in B-type strains, we compared normalized genome coverage between B-type strains using the ODER1 haplotype 1 genome assembly as a reference.

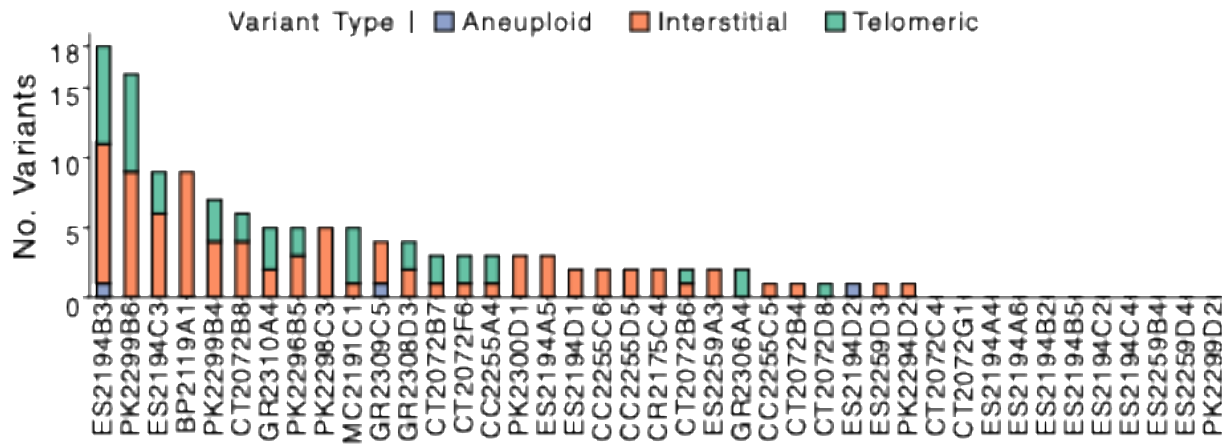


Figure 4.6 Chromosomal anomalies in Texas B-type strains.

Anomalies are color-coded by anomaly type, and information on all anomalies can be found in Table B.7.

We identified 128 total chromosomal anomalies, deduplicated to 104 unique anomalies in Texas B-type isolates (Figure 4.6). We categorized chromosomal anomalies as aneuploid (n=3, whole chromosome), interstitial (n=82, within a chromosome), or telomeric (n=43, at the end of a chromosome). Chromosomal anomalies were present in 29 of 34 total scaffolds in the ODER1 haplotype 1 assembly (Table B.7). ES2194B3, ES2194D2, and GR2309C5 are trisomic for

scaffolds 25, 1, and 20, respectively. No aneuploid variants were losses of DNA. Most (n=84) chromosomal anomalies were segmental regions of decreased coverage suggestive of DNA loss, 41 anomalies were segmental regions of increased coverage suggestive of DNA gain, and three were aneuploid gains of DNA.

The isolate with the most chromosomal anomalies was ES2194B3 (n=18), from a water sample collected from Lake E.V. Spence in December 2019 (Figure 4.6). Other isolates from the same water sample had fewer (≤ 9) anomalies. Isolates from Lake Possum Kingdom have a similar range in anomaly quantity (1-16). A total of 11 strains did not have any anomalies (Table B.14). American B-types had similar amounts of anomalies to A-type hybrids (Wilcox-Mann-Whitney $p=0.18$) but more anomalies than A-type diploids (Wilcox-Mann-Whitney $p=0.05$).

4.5 Discussion

4.5.1 Two cryptic species of *P. parvum* s.l. are present in Texas blooms

All prior genome-level investigations of Texas *P. parvum* s.l. found that all strains were A-types (Binzer et al., 2019; Rasmussen et al., 2016; Wisecaver et al., 2023). The isolates assayed in this study were mainly B-types (n=41). A minority of recent isolates (n=13) were A-types, and none were C-types. (Figure 4.2, Figure 4.3). This is the first study that has sequenced whole genomes of B-types from the North American continent.

All recent Texas B-type *P. parvum* s.l. isolates formed a single clade with limited genetic distance which was sister to European strains (Figure A.16). We were unable to procure cultures for the previously identified American B-type strains (8010, Lake Diversion). However, both 8010 and Lake Diversion also grouped sister to European B-types in previous ITS phylogenies (Driscoll et al., 2023; Lutz-Carrillo et al., 2010).

Texas B-types are low-heterozygosity diploids; the only other known B-type diploid is ODER1 from Europe (Kuhl et al., 2024). Total DNA content of Texas B-types can vary significantly but not to the extent that is explained by ploidy differences. This variation is also not explained by body of water, isolation date, or gene content (Figure A.17). This DNA variation may be due to

repetitive element expansion, which was found to drive the expansion of B-type genome sizes compared to A-type genomes (Kuhl et al., 2024).

We also chemically characterized seven B-type *P. parvum* s.l. isolates via untargeted metabolomics, all of which produced B-type prymnesins (Figure 4.5). Toxin production varied, but every strain produced at least one B-type prymnesin. *P. parvum* s.l. is known to have low- and high-toxicity strains; however, low-toxicity strains still produced toxins (Driscoll et al., 2023, 2013). Prymnesin production is also only one measure of toxicity, and future studies comparing allelopathy or hemolytic activity of Texas B-types are needed.

The A clade of *P. parvum* shows more within-clade genetic variation than the B clade (Figure 4.3). Three different A-type karyotypes (haploid, diploid, and hybrid) can be found, two of which (haploid and hybrid) have been isolated from the same bloom. Three new A1 x A2 hybrid isolates (PK2297A3, PK2297B2, PK2297B5) were identified in this study (Figure 4.3). The presence of A-type *P. parvum* s.l. was also verified via untargeted metabolomics of seven phylogenetically characterized A-type isolates (Figure 4.4).

4.5.2 Multiple cryptic species are found in the same blooms

A-type and B-type *P. parvum* were both isolated from the same bloom in Lake Possum Kingdom in 2020. This is not wholly unique, as Lake Colorado City had an A-type *P. parvum* bloom in December 2019 followed by a B-type *P. parvum* bloom in February 2020. Moreover, Driscoll et al (2023) examined another bloom in Lake Possum Kingdom in 2013 using an ITS-based tree and found both A-types and B-types. However, this analysis not only compares more individuals, but also constructs a much more robust phylogeny.

The sympatry of both A- and B-types supports that both A- and B-types are separate species from each other. If no mating barriers existed, then admixture between A- and B-types would create intermediate genotypes, which are not observed. The sympatry of both A- and B-types also suggests that there may be some form of niche differentiation between them, or that one species is in the process of outcompeting another. If A- and B-types are truly coexisting, then additional phenotyping studies of Texas *P. parvum* s.l. could help identify what these niches may be.

The lines between species in the *P. parvum* s.l. complex are difficult to place. The broad differences between A-, B-, and C-types and the sympatry of A- and B-types indicate that at least three species of *P. parvum* s.l. exist (Wisecaver et al., 2023). A- and C-types also have subclades which may be separate species as well. A-type subclades can form stable hybrids, which shows that mating barriers in the *P. parvum* complex can be leaky (Watervoort et al., 2026).

4.5.3 B-type *P. parvum* s.l. accumulate chromosomal anomalies

In comparison to the number of chromosomal abnormalities detected in A-type *P. parvum* Watervoort et al. (2026), the frequency of chromosomal anomalies in B-types was similar to A-hybrid strains but more than non-hybrid A1/A2 strains. This may be due to the B-types having a larger genome size than A-types, elevating the rates of chromosomal anomalies between non-hybrid strains. Fewer aneuploid anomalies were also observed in B-types than A-types. The presence of chromosomal anomalies in another lineage of *P. parvum* s.l. suggests that this is a major mechanism by which all *P. parvum* s.l. genomes evolve. This has not been investigated in C-type *P. parvum* s.l. due to the lack of a reference genome, but the presence of chromosomal anomalies in two out of three major groups of *P. parvum* s.l. suggests that C-type strains also accumulate chromosomal anomalies.

4.6 References

- Babraham Bioinformatics. 2011. FastQC A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Binzer, S.B., Svenssen, D.K., Daughjerg, N., Alves-de-Souza, C., Pinto, E., Hansen, P.J., Larsen, T.O., 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>
- Brooks, B.W., Grover, J.P., Roelke, D.L., 2011. *Prymnesium parvum*: An emerging threat to inland waters. *Environ. Toxicol. Chem.* 30, 1955–1964. <https://doi.org/10.1002/etc.613>
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. <https://doi.org/10.1038/nmeth.3176>
- Bushnell B. 2017. BBTools Software Package. <https://sourceforge.net/projects/bbmap/>

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Castillo-Hair, S.M., Sexton, J.T., Landry, B.P., Olson, E.J., Igoshin, O.A., Tabor, J.J., 2016. FlowCal: A user-friendly, open source software tool for automatically converting flow cytometry Data from arbitrary to calibrated units. *ACS Synth. Biol.* 5, 774–780. <https://doi.org/10.1021/acssynbio.5b00284>
- Cavaller-Smith T. 1985. The evolution of genome size. John Wiley and Sons Inc., New York, NY. <https://www.osti.gov/biblio/6076917>
- Čertnerová, D., Galbraith, D.W., 2021. Best practices in the flow cytometry of microalgae. *Cytometry A* 99, 359–364. <https://doi.org/10.1002/cyto.a.24328>
- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>
- Dpooležel, J., Binarová, P., 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>
- Driscoll, W.W., Espinosa, N.J., Eldakar, O.T., Hackett, J.D., 2013. Allelopathy as an emergent, exploitable public good in the bloom-forming microalga *Prymnesium parvum*. *Evolution* 67, 1582–1590. <https://doi.org/10.1111/evo.12030>
- Driscoll, W.W., Wisecaver, J.H., Hackett, J.D., Espinosa, N.J., Padway, J., Engers, J.E., 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>
- Fallon, T.R., Shende, V.V., Wierzbicki, I.H., Pendleton, A.L., Watervoot, N.F., Auber, R.P., Gonzalez, D.J., Wisecaver, J.H., Moore, B.S., 2024. Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins. *Science* 385, 671–678. <https://doi.org/10.1126/science.ado3290>
- Grover, J.P., Baker, J.W., Roelke, D.L., Brooks, B.W., 2010. Current Status of Mathematical Models for Population Dynamics of *Prymnesium parvum* in a Texas Reservoir. *JAWRA J. Am. Water Resour. Assoc.* 46, 92–107. <https://doi.org/10.1111/j.1752-1688.2009.00393.x>

- Guillard, R., 2005. Purification methods for microalgae, in: *Algal Culturing Techniques*. Elsevier, Burlington, MA, pp. 117–132. <https://doi.org/10.1016/B978-012088426-1/50009-3>
- Hallegraeff, G.M., Anderson, D.M., Belin, C., Bottein, M.-Y.D., Bresnan, E., Chinain, M., Enevoldsen, H., Iwataki, M., Karlson, B., McKenzie, C.H., Sunesen, I., Pitcher, G.C., Provoost, P., Richardson, A., Schweibold, L., Tester, P.A., Trainer, V.L., Yñiguez, A.T., Zingone, A., 2021. Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts. *Commun. Earth Environ.* 2, 1–10. <https://doi.org/10.1038/s43247-021-00178-8>
- Kazmi, S.S.U.H., Yapa, N., Karunarathna, S.C., Suwannarach, N., 2022. Perceived Intensification in Harmful Algal Blooms Is a Wave of Cumulative Threat to the Aquatic Ecosystems. *Biology* 11, 852. <https://doi.org/10.3390/biology11060852>
- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Ryneerson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaultot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* 12, e1001889. <https://doi.org/10.1371/journal.pbio.1001889>
- Kokot, M., Długosz, M., Deorowicz, S., 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* 33, 2759–2761. <https://doi.org/10.1093/bioinformatics/btx304>
- Kuhl, H., Strassert, J.F.H., Čertnerová, D., Varga, E., Kreuz, E., Lamatsch, D.K., Wuertz, S., Köhler, J., Monaghan, M.T., Stöck, M., 2024. The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins. *Curr. Biol.* 34, 3698–3706. <https://doi.org/10.1016/j.cub.2024.06.033>
- Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: interrogation of genome assemblies. *F1000Research* 6, 1287. <https://doi.org/10.12688/f1000research.12232.1>
- Letunic, I., Bork, P., 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. <https://doi.org/10.1093/nar/gkae268>
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://doi.org/10.48550/arXiv.1303.3997>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinforma. Oxf. Engl.* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liebert F, Deerns W. 1920. Onderzoek naar de oorzak van een Vischsterfte in den Polder Workumer Nieuwland, nabij Workum. *Verhandungen en Rapporten uitgegeven door Rijkinstututen voor Visscherijonderzoek.* 1:81–93. https://books.google.com/books/about/Onderzoek_naar_de_oorzaak_van_een_vischs.htm?id=PN6x0AEACAAJ
- Linam G, Ralph J, Glass J. 1991. Toxic blooms, an unusual algae threatens aquatic resources. *Chihuahuan Desert Discovery.* 28 (Winter). <https://tpwd.texas.gov/landwater/water/environconcerns/hab/ga/news/toxicblooms.phtml>
- Liu, Y., Klaiber, H.A., 2023. Don't Drink the Water! The Impact of Harmful Algal Blooms on Household Averting Expenditure. *Environ. Resour. Econ.* 86, 29–55. <https://doi.org/10.1007/s10640-023-00786-2>
- Lutz-Carrillo, D.J., Southard, G.M., Fries, L.T., 2010. Global genetic relationships among isolates of golden alga (*Prymnesium parvum*). *J. Am. Water Resour. Assoc.* 46, 24–32. <https://doi.org/10.1111/j.1752-1688.2009.00388.x>
- Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J.G., Gitzendanner, M.A., Wafula, E., Der, J.P., dePamphilis, C.W., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D.E., Soltis, P.S., Rothfels, C., Pokorny, L., Shaw, J.A., DeGironimo, L., Stevenson, D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., Wong, G.K.-S., 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3, 17. <https://doi.org/10.1186/2047-217X-3-17>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mora, D., Schlüsener, M., Fischer, H., Kleinteich, J., Schulz, M., Ternes, T., Thiel, J., Wick, A., Krenek, S., 2024. From genes to toxins: Profiling *Prymnesium parvum* during a riverine harmful algal bloom. *Harmful Algae* 136, 102644. <https://doi.org/10.1016/j.hal.2024.102644>
- Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>

- O’Leary, N.A., Wright, M.W., Brister, J.R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O’Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D., Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. <https://doi.org/10.1093/nar/gkv1189>
- Ortiz, E.M., 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. <https://doi.org/10.5281/ZENODO.2540861>
- Patiño, R., Christensen, V.G., Graham, J.L., Rogosch, J.S., Rosen, B.H., 2023. Toxic algae in inland waters of the conterminous united states—a review and synthesis. *Water* 15, 2808. <https://doi.org/10.3390/w15152808>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>
- Rasmussen, S.A., Meier, S., Andersen, N.G., Blossom, H.E., Duus, J.Ø., Nielsen, K.F., Hansen, P.J., 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>
- Roelke, D.L., Barkoh, A., Brooks, B.W., Grover, J.P., Hambright, K.D., LaClaire, J.W., Moeller, P.D.R., Patino, R., 2016. A chronicle of a killer alga in the west: ecology, assessment, and management of *Prymnesium parvum* blooms. *Hydrobiologia* 764, 29–50. <https://doi.org/10.1007/s10750-015-2273-6>
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 19, 1117–1123. <https://doi.org/10.1101/gr.089532.108>

- Sobieraj, J., Metelski, D., 2023. Insights into toxic *Prymnesium parvum* blooms as a cause of the ecological disaster on the Odra River. *Toxins* 15, 403. <https://doi.org/10.3390/toxins15060403>
- Southard, G.M., Fries, L.T., Barkoh, A., 2010. *Prymnesium parvum*: The Texas experience. *JAWRA J. Am. Water Resour. Assoc.* 46, 14–23. <https://doi.org/10.1111/j.1752-1688.2009.00387.x>
- Tábor-Sarmiento, S., Patiño, R., Portillo-Quintero, C., Coldren, C., 2022. Air, land, and water variables associated with the first appearance and current spatial distribution of toxic *Prymnesium parvum* blooms in reservoirs of the southern great plains, USA. *Sci. Total Environ.* 836, 155567. <https://doi.org/10.1016/j.scitotenv.2022.155567>
- Van der Auwera, G.A., O'Connor, B.D., 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media. https://catalog.nlm.nih.gov/permalink/01NLM_INST/1fua1rm/alma9917773213406676
- VanLandeghem, M.M., Farooqi, M., Southard, G.M., Patiño, R., 2015. Associations between Water Physicochemistry and *Prymnesium parvum* Presence, Abundance, and Toxicity in West Texas Reservoirs. *JAWRA J. Am. Water Resour. Assoc.* 51, 471–486. <https://doi.org/10.1111/jawr.12262>
- Watervoort, N.F., Jeje, T., Dilkes, B.P., Wisecaver, J.H., 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum* sensu lato species complex. *bioRxiv* 2026.06.03.730016. <https://doi.org/10.64898/2026.06.03.730016>
- Wisecaver, J., Auber, R., 2022. Wisecaver Lab algal DNA Extraction with Plant DNAzol Reagent [WWW Document]. *protocols.io*. URL [dx.doi.org/10.17504/protocols.io.kxygxp4dol8j/v1](https://doi.org/10.17504/protocols.io.kxygxp4dol8j/v1)
- Wisecaver, J.H., Auber, R.P., Pendleton, A.L., Watervoort, N.F., Fallon, T.R., Riedling, O.L., Manning, S.R., Moore, B.S., 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>

5. CONCLUSION

Cryptic species complexes are common in microbial eukaryotes and are found across the eukaryotic tree of life (Figure 1.1). Many algal bloom-forming species are cryptic species complexes. In fact, cryptic species complexes can be found in every toxic eukaryotic clade in the UNESCO Taxonomic Reference List of Harmful Microalgae (Lilly et al. 2007; Riisberg and Edvardsen 2008; Lundholm et al. 2009; Quijano-Scheggia et al. 2009; Klöpper et al. 2013; John et al. 2014; Galen et al. 2018; Wisecaver et al. 2023).

The toxic microalga *Prymnesium parvum* s.l. is an example of a cryptic species complex in microbial eukaryotes. Binzer et al. (2019) first identified the complex by using an rDNA (ITS1, ITS2, 5.8s) phylogeny to split *P. parvum* s.l. into three main clades: A-types, B-types, and C-types. Wisecaver et al. (2023) used a genome-level analysis to further split A-types into A1, A2, and A-hybrid subclades and split C-types into C1, C2, and C3 subclades. In Chapter 3, I used a genome-level analysis to identify one more potential subclade of *P. parvum* s.l.: A3. Clades within the species complex have distinct genome sizes, gene content, and nucleotide composition (Wisecaver et al. 2023).

P. parvum s.l. member species are not consistently acknowledged in modern studies. The most common reason cryptic species of *P. parvum* s.l. are not considered is that studies use strains isolated from nearby blooms but do not sequence or chemically characterize them (Briand et al. 2025; Abhijith et al. 2026 May). Alternatively, studies use a previously characterized and accessioned strain but draw complex-wide conclusions despite surveying only one species in the *P. parvum* species complex (Boucher et al. 2025; Mehdizadeh Allaf et al. 2026). Properly acknowledging the *P. parvum* species complex in experimental design requires more than sampling multiple species; it also requires proper handling of the morphologically identical *P. parvum* s.l. cryptic species. Without experimental validation of which cryptic species a *P. parvum* s.l. strain belongs to, strain mix-ups are easy. In Chapter 2, I described a strain mix-up in which two strains of *P. parvum* s.l.: CCMP3037 and UTEX2797 were inadvertently swapped prior to sequencing.

Strain swaps may also be more common than they initially appear. In Chapter 3, I identified ARC140 as an A-type in my analysis; however, the strain was labeled a B-type in Binzer et al. (2019). Strain incongruencies can be found outside of this thesis, as K0374 was an A-type in Beszteri et al. (Beszteri et al. 2026) but was a B-type in Binzer et al. (2019). K0252 was a B-type in Rasmussen et al. (2016) but a C-type in Binzer et al. (2019). The prevalence of potential strain swaps highlights the importance of depositing strains in publicly accessible culture collections to not only allow future investigations into those strains, but also to safeguard against strain swaps.

Cryptic speciation raises immediate methodological concerns for studying *P. parvum*. Proper handling of *P. parvum* strains is prudent in the short term, but it is likely necessary to develop markers or other rapid identification tests to differentiate cryptic species. In the case of the strain swap identified in Chapter 2, checking the phylogenetic placement or heterozygosity of both sequenced strains would have indicated the swap. Future studies of *P. parvum* s.l. should also report strain IDs and select strains from appropriate cryptic species when designing experiments. Researchers should also carefully consider if the results from a study that did not include their cryptic species are truly applicable to their research. In cases where strains have not been characterized before, researchers should categorize strains with as much phylogenetic resolution as possible, be it whole-genome sequencing, rDNA/marker gene phylogenies, or prymnesin chemotyping.

Although cryptic speciation in *P. parvum* s.l. poses methodological challenges, it also generates interesting biological phenomena. The A-type clade of *P. parvum* s.l. is separated into four clades: A1, A2, A3, and A-hybrid. The A-hybrid clade is formed by hybridization between A1 and A2 strains (Wisecaver et al. 2023; Chapter 3). A-hybrid strains are frequently recovered in North American *P. parvum* blooms and are likely formed by multiple hybridization events. A-hybrid genomes also frequently accumulate chromosomal anomalies, even between single-cell re-isolates of the same progenitor strains. In Chapter 4, I demonstrate that chromosomal anomalies are also present in B-types. Although chromosomal anomalies are also found in both A- and B-type homozygous diploids, far more are present in A-hybrid strains. The genomic instability observed in *P. parvum* may be the result of primarily asexual reproduction in diploid strains, as seen in the green alga *Auxenochlorella* (Craig et al. 2025). Strains of the same cryptic species from the same

water sample may not have identical structural anomalies. However, these anomalies may have arisen during laboratory cultivation, which should be investigated in the future. C-types may also have chromosomal anomalies like A- and B-types, but identifying C-type chromosomal anomalies requires a C-type reference genome.

Cryptic species diversity in the *P. parvum* species complex is also relevant to local ecosystems. In Chapter 4, I found that *P. parvum* s.l. blooms in Texas from 2018 to 2020 could be composed of A1 strains, A-hybrid strains, B-type strains, or a combination of all three. The existence of multiple cryptic species of *P. parvum* s.l. in sympatry suggests that some niche differentiation exists between them. When combined with the fact that A-type prymnesins are more toxic than B-type prymnesins, the genetic diversity of *P. parvum* s.l. blooms likely has implications for future monitoring and management programs (Varga et al. 2024).

Currently isolated strains are also heavily biased towards bloom-like conditions. Some strains have been isolated from non-bloom conditions, such as CCMP 1926, CCMP 1927, and select isolates from Texas in Chapter 4, but strains from bloom conditions are more common. Isolating strains from non-bloom conditions is more challenging than from bloom conditions because *P. parvum* s.l. takes up a smaller fraction of total biomass. Environmental metabarcoding or genetic markers could survey the genetic diversity of *P. parvum* s.l. without isolating strains, although this would not be at the same resolution as whole-genome sequencing. Environmental metagenomics is very promising and has yielded a fragmented *Prymnesium* sp. metagenome-assembled genome (Ettinger et al. 2025). Similar analyses could sample more genome-wide genetic diversity in ambient conditions than marker-based analyses.

Genetic separation within A-, B-, and C-type clades appears to be somewhat driven by geographic separation. Within A- and B-types, strains from separate continents form separate clades, although sampling is limited for some continents and some clades (Wisecaver et al. 2023; Chapter 4). C-types do not have as clear a signal, as Asian and European strains can be found in the C1 clade, although Asian C1 strains are sister to European C1 strains (Chapter 4). Additionally, the only known C3 strain (K0252) is the only sequenced C-type from Oceania, which supports the impacts of geographic distance driving differentiation (Wisecaver et al. 2023). All three major clades (A-, B-, and C-types) can be found around the world, although B-types are only found in North America

and Europe (Binzer et al. 2019; Wisecaver et al. 2023; Chapter 4). Co-occurrence of A- and B-types in individual blooms in Texas, and A-, B-, and C-types in Europe does confound any geographic trends beyond the subclade level. This co-occurrence may be due to more recent introductions of other clades to areas with existing *P. parvum* s.l. populations (i.e. B-types recently introduced to an area with endemic A-types), but there is no historical data to support this. Sampling of *P. parvum* s.l. is also not even, and continents such as Asia or Africa are greatly underrepresented. Geographic distance has driven cryptic speciation in other microbial eukaryotes which were previously thought to be purely cosmopolitan (Lilly et al. 2007; Kooistra et al. 2008; De Luca et al. 2021).

The genetic diversity of *Prymnesium parvum* s.l. is likely much greater than what is currently known (Figure 5.1). Increased sampling and sequencing of *P. parvum* s.l. frequently reveals clades with one to two strains instead of newly sampled strains fitting purely into established clades. Currently, there are five (A2, A3, C2, C3, and CCMP1926/CCMP1927) low-membership clades which are found in A- and C-types. More clades of *P. parvum* s.l. may be found by increasing sampling of under-sampled continents such as Africa or Asia. The CCMP1926/CCMP1927 clade is sister to the established A-, B-, and C-type lineages and has not been chemically characterized. These two strains may be highly divergent A-, B-, or C-type strains or potentially another chemotype entirely.

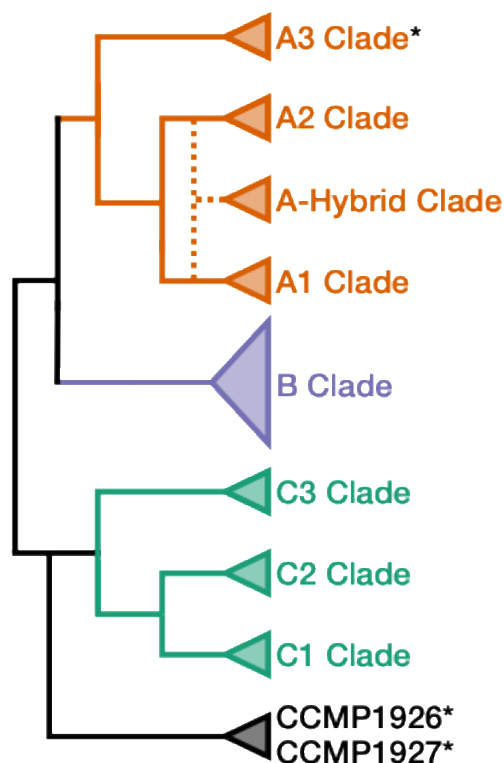


Figure 5.1 Model phylogeny of the *P. parvum* s.l. species complex incorporating results from this thesis (*). CCMP1926/1927 have not been chemotyped and may be a new lineage of *P. parvum* s.l. entirely.

To date, large-scale investigations into *P. parvum* have mostly focused on the A- and B-type clades (Wisecaver et al. 2023; Chapter 3; Chapter 4). Comparatively, far less is known about the C-type clade partly because it does not have a reference genome or a resolved prymnesin structure. The C-type clade, like the A-type clade, has three distinct subclades, which raises the possibility of hybridization in C-types similar to that seen in A-types. Once genomic resources like a C-type reference are available, the genetic diversity of C-types is fertile ground for future studies.

Similar lessons to those learned in *P. parvum* s.l. have been learned in other cryptic species complexes. One of the best described cryptic species complexes in microbial eukaryotes is the *Alexandrium tamarense* species complex. The morphospecies *Alexandrium catenella*, *A. fundyense*, and *A. tamarense* together form the *A. tamarense* species complex (Balech 1985). Lilly et al. (2007) found that *A. tamarense* s.l. strains formed five distinct clades on an LSU rDNA phylogeny, and designated these clades as groups 1-5. Every member of groups 1 and 4 was toxic,

whilst the other three were not. Mating barriers also separate groups, as inter-group mating experiments found that each group combination either formed no offspring or inviable offspring (Brosnahan et al. 2010). Each group of *A. tamarense* s.l. was formally accepted as a different species of *Alexandrium* based on both genetic distances and mating incompatibilities (John et al. 2014).

Upon discovering the five groups comprising the *A. tamerense* species complex, Lilly et al. (2007) advocated for using group names (1-5) instead of morphospecies. These group names were adopted and helped guide studies whilst taxonomic and nomenclature debates were ongoing (Collins et al. 2009; Brosnahan et al. 2010; Touzet et al. 2010). *P. parvum* s.l. is undergoing something similar, as A-, B-, and C-types are used to describe separate species of *P. parvum* s.l. before formal taxonomic revision (Bannon et al. 2024; Kuhl et al. 2024; Varga et al. 2024). The results from *P. parvum* s.l. and *A. tamarense* s.l. highlight the importance of using language to indicate broader cryptic diversity in the time between discovery and formal taxonomic revision.

P. parvum s.l. also used to be separated into two species: *P. parvum* and *P. patelliferum* (Green et al. 1982). *P. parvum* and *P. patelliferum* were differentiated based on ultrastructural differences in glycoprotein scales that coat their plasma membranes (Larsen et al. 1993). However, a later study found that cells could switch between *P. patelliferum* and *P. parvum* morphotypes (Larsen and Edvardsen 1998). Another study also found no difference between *P. parvum* and *P. patelliferum* on an ITS-based phylogeny (Larsen and Medlin 1997). *P. parvum* and *P. patelliferum* were then formally collapsed into one species: *P. parvum* s.l. (Larsen 1999). The cryptophyte *Teleaulax amphioxeia* was also erroneously separated into two species: *Teleaulax amphioxeia* and *Plagioselmis prolunga*. *T. amphioxeia* and *P. prolunga* are morphologically distinct, as *P. prolunga* is covered in periplast “plates” and is up to 8 µm long whilst *T. amphioxeia* is unclad and is up to 13 µm long (Altenburger et al. 2020). Altenburger et al. (2020) found that these two species were indistinguishable on an 18s phylogeny. Their flow cytometry analysis found that *P. prolunga* had half the amount of DNA as *T. amphioxeia*, which suggests that *P. prolunga* may be the haploid life stage of *T. amphioxeia*. *T. amphioxeia* cultures also changed into the *P. prolunga* morphology once cultures were undisturbed for ~1.5 months (Altenburger et al. 2020). Both *P.*

parvum s.l. and *T. amphioxia* s.l. are cautionary tales about oversplitting species on morphological characteristics when the life cycle of an organism is not fully known.

Taken together, the *P. parvum* s.l., *T. amphioxeia* s.l., and *A. tamarense* s.l. case studies demonstrate that morphology does not reliably distinguish species. The unreliability of morphology is evolutionarily conserved as cryptic species complexes are widespread across the eukaryotic tree of life (Figure 1.1). Morphology should not be entirely abandoned, as genetic approaches are more expensive and less accessible than morphological approaches (Dunn 2003). Nevertheless, the inability of morphology to consistently resolve species-level identification is a real concern that should be heeded in future studies.

5.1 References

- Abhijith M et al. 2026. Understanding the bacterial community associated with potentially toxic haptophyte *Prymnesium parvum* using culture-based and next- generation sequencing techniques. *Total Environment Microbiology*. 100092. <https://doi.org/10.1016/j.temicr.2026.100092>
- Altenburger A et al. 2020. Dimorphism in cryptophytes—The case of *Teleaulax amphioxeia* / *Plagioselmis prolonga* and its ecological implications. *Sci Adv*. 6(37):eabb1611. <https://doi.org/10.1126/sciadv.abb1611>
- Balech E. 1985. The genus *Alexandrium* or *Gonyaulax* of the *tamarensis* group. *Toxic dinoflagellates*. 32–38. <https://cir.nii.ac.jp/crid/1573950399076488192>
- Bannon CC et al. 2024. Influence of biotic and abiotic factors on prymnesin profiles in three strains of *Prymnesium parvum*. *Algal Research*. 78:103390. <https://doi.org/10.1016/j.algal.2024.103390>
- Beszteri S et al. 2026. Eco-physiological tolerances of *Prymnesium parvum* (Haptophyta, Prymnesiophyceae) strains – defined by clade assignment or local adaptation? *Phycologia*. 65(2):93–102. <https://doi.org/10.1080/00318884.2026.2617251>
- Binzer SB et al. 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>
- Boucher C et al. 2025. *Prymnesium parvum*: Correlation between lytic activity and prymnesin content, both increasing with phosphorus-depletion and lower cell density. *Algal Research*. 85:103830. <https://doi.org/10.1016/j.algal.2024.103830>

- Briand E et al. 2025. Species-Specific Exudates From the Harmful Algae *Alexandrium minutum* and *Prymnesium parvum* Drive the Structure of a Synthetic Microbial Community. *Environmental Microbiology*. 27(7):e70150. <https://doi.org/10.1111/1462-2920.70150>
- Brosnahan ML et al. 2010. Outbreeding lethality between toxic Group I and nontoxic Group III *Alexandrium tamarense* spp. isolates: Predominance of heterotypic encystment and implications for mating interactions and biogeography. *Deep Sea Res 2 Top Stud Oceanogr*. 57(3–4):175–189. <https://doi.org/10.1016/j.dsr2.2009.09.005>
- Collins C et al. 2009. IDENTIFICATION AND TOXICITY OF ALEXANDRIUM TAMARENSE (DINOPHYCEAE) IN SCOTTISH WATERS¹. *Journal of Phycology*. 45(3):692–703. <https://doi.org/10.1111/j.1529-8817.2009.00678.x>
- Craig RJ et al. 2025. Targeted genetic manipulation and yeast-like evolutionary genomics in the green alga *Auxenochlorella*. *Plant Cell*. 37(11):koaf259. <https://doi.org/10.1093/plcell/koaf259>
- De Luca D, Piredda R, Sarno D, Kooistra WHCF. 2021. Resolving cryptic species complexes in marine protists: phylogenetic haplotype networks meet global DNA metabarcoding datasets. *The ISME Journal*. 15(7):1931–1942. <https://doi.org/10.1038/s41396-021-00895-0>
- Dunn CP. 2003. Keeping taxonomy based in morphology. *Trends in Ecology & Evolution*. 18(6):270–271. [https://doi.org/10.1016/S0169-5347\(03\)00094-6](https://doi.org/10.1016/S0169-5347(03)00094-6)
- Ettinger CL, Eisen JA, Stajich JE. 2025. Eukaryotic MAGs recovered from deep metagenomic sequencing of the seagrass, *Zostera marina*, include a novel chytrid in the order Lobulomycetales. [accessed 2026 Sept 15]. <http://biorxiv.org/lookup/doi/10.1101/2025.02.11.637735>. <https://doi.org/10.1101/2025.02.11.637735>
- Galen SC et al. 2018. The polyphyly of *Plasmodium* : comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R Soc open sci*. 5(5):171780. <https://doi.org/10.1098/rsos.171780>
- Green JC, Hibberd DJ, Pienaar RN. 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. *British Phycological Journal*. 17(4):363–382. <https://doi.org/10.1080/00071618200650381>
- John U et al. 2014. Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: The introduction of five species with emphasis on molecular-based (rDNA) classification. *Protist*. 165(6):779–804. <https://doi.org/10.1016/j.protis.2014.10.001>
- Klöpper S et al. 2013. Phylogeny and morphology of a *Chattonella* (Raphidophyceae) species from the Mediterranean Sea: what is *C. subsalsa* ? *European Journal of Phycology*. 48(1):79–92. <https://doi.org/10.1080/09670262.2013.771412>

- Kooistra WHCF et al. 2008. Global Diversity and Biogeography of *Skeletonema* Species (Bacillariophyta). *Protist*. 159(2):177–193. <https://doi.org/10.1016/j.protis.2007.09.004>
- Kuhl H et al. 2024. The haplotype-resolved *Prymnesium parvum* (type B) microalga genome reveals the genetic basis of its fish-killing toxins. *Current Biology*. 34(16):3698–3706. <https://doi.org/10.1016/j.cub.2024.06.033>
- Larsen A. 1999. *Prymnesium parvum* and *P. patelliferum* (Haptophyta)—one species. *Phycologia*. 38(6):541–543. <https://doi.org/10.2216/i0031-8884-38-6-541.1>
- Larsen A, Edvardsen B. 1998. Relative ploidy levels in *Prymnesium parvum* and *P. patelliferum* (Haptophyta) analyzed by flow cytometry. *Phycologia*. 37(6):412–424. <https://doi.org/10.2216/i0031-8884-37-6-412.1>
- Larsen A, Eikrem W, Paasche E. 1993. Growth and toxicity in *Prymnesium patelliferum* (Prymnesiophyceae) isolated from Norwegian waters. *Can J Bot*. 71(10):1357–1362. <https://doi.org/10.1139/b93-161>
- Larsen A, Medlin LK. 1997. Inter- and intraspecific genetic variation in twelve *Prymnesium* (Haptophyceae) clones. *Journal of Phycology*. 33(6):1007–1015. <https://doi.org/10.1111/j.0022-3646.1997.01007.x>
- Lilly EL, Halanynch KM, Anderson DM. 2007. Species boundaries and global biogeography of the *Alexandrium tamarense* complex (Dinophyceae)1. *Journal of Phycology*. 43(6):1329–1338. <https://doi.org/10.1111/j.1529-8817.2007.00420.x>
- Lundholm N et al. 2009. IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae. Accessed at <https://www.marinespecies.org/hab> on yyyy-mm-dd. [accessed 2026 Jan 29]. <https://www.marinespecies.org/imis.php?dasid=3124&doiid=362>. <https://doi.org/10.14284/362>
- Mehdizadeh Allaf M et al. 2026. Chemical Control of Ichthyotoxic Algal Blooms in Aquaculture: Assessing Algicide Impacts on Cellular Motility and Bloom Suppression. *Microorganisms*. 14(5):1086. <https://doi.org/10.3390/microorganisms14051086>
- Quijano-Scheggia SI et al. 2009. Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia*. 48(6):492–509. <https://doi.org/10.2216/08-21.1>
- Rasmussen SA et al. 2016. Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J Nat Prod*. 79(9):2250–2256. <https://doi.org/10.1021/acs.jnatprod.6b00345>
- Riisberg I, Edvardsen B. 2008. Genetic variation in bloom-forming ichthyotoxic *Pseudochattonella* species (Dictyochophyceae, Heterokonta) using nuclear, mitochondrial and plastid DNA sequence data. *European Journal of Phycology*. 43(4):413–422. <https://doi.org/10.1080/09670260802299602>

- Touzet N et al. 2010. Co-Occurrence of the West European (Gr.III) and North American (Gr.I) Ribotypes of *Alexandrium tamarense* (Dinophyceae) in Shetland, Scotland. *Protist.* 161(3):370–384. <https://doi.org/10.1016/j.protis.2009.12.001>
- Varga E 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>
- Watervoort NF, Jeje T, Dilkes BP, Wisecaver JH. 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum* sensu lato species complex. *bioRxiv.* 2026.06.03.730016. <https://doi.org/10.64898/2026.06.03.730016>
- Wisecaver JH et al. 2023. Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. *Current Biology.* 33(11):2246-2259.e8. <https://doi.org/10.1016/j.cub.2023.05.003>

APPENDIX A. SUPPLEMENTAL FIGURE

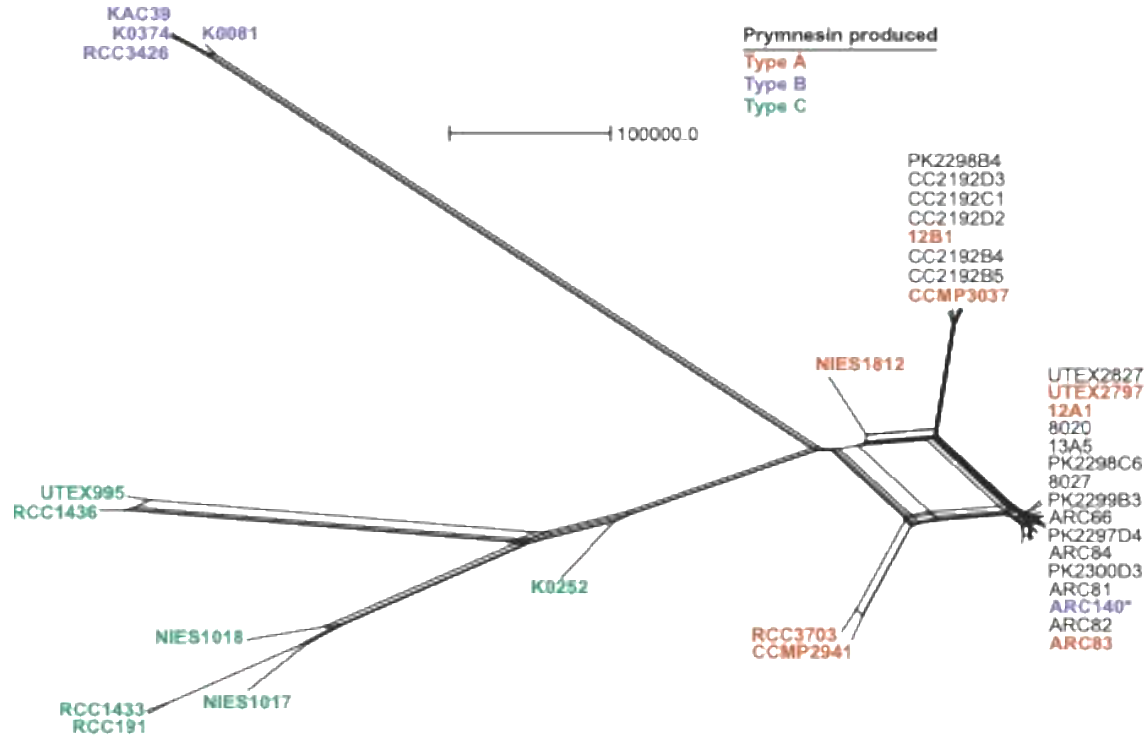


Figure A.1: *Pymnesium parvum sensu lato* network

Chemotyped strains are color coded according to Wisecaver et al. (2023) and Binzer et al. (2019). Strains that grouped within the type A clade were included in all downstream analyses, with the exception of ARC140(*). ARC140 was originally chemotyped as producing prymnesin type B (Binzer et al. 2019), but clustered with type A strains in our analysis. Edge length represents absolute SNP difference with the scale bar corresponding to 100,000 SNPs.

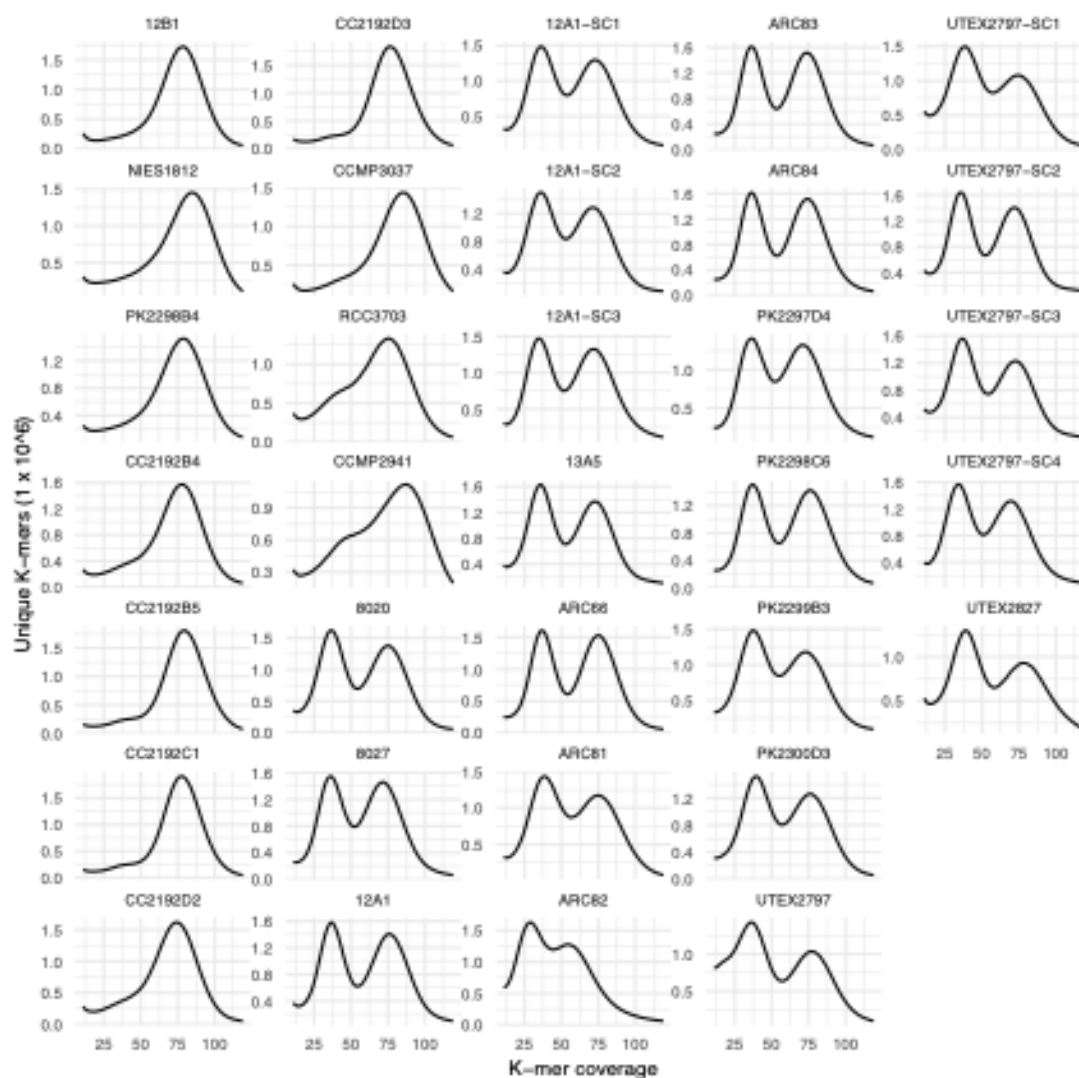


Figure A.2: Heterozygosity in *P. parvum* clade A strains

Filtered Illumina reads were randomly subsampled to 10 million pairs per strain prior calculation of K-mer coverage to control for differences in sequencing depth between strains.

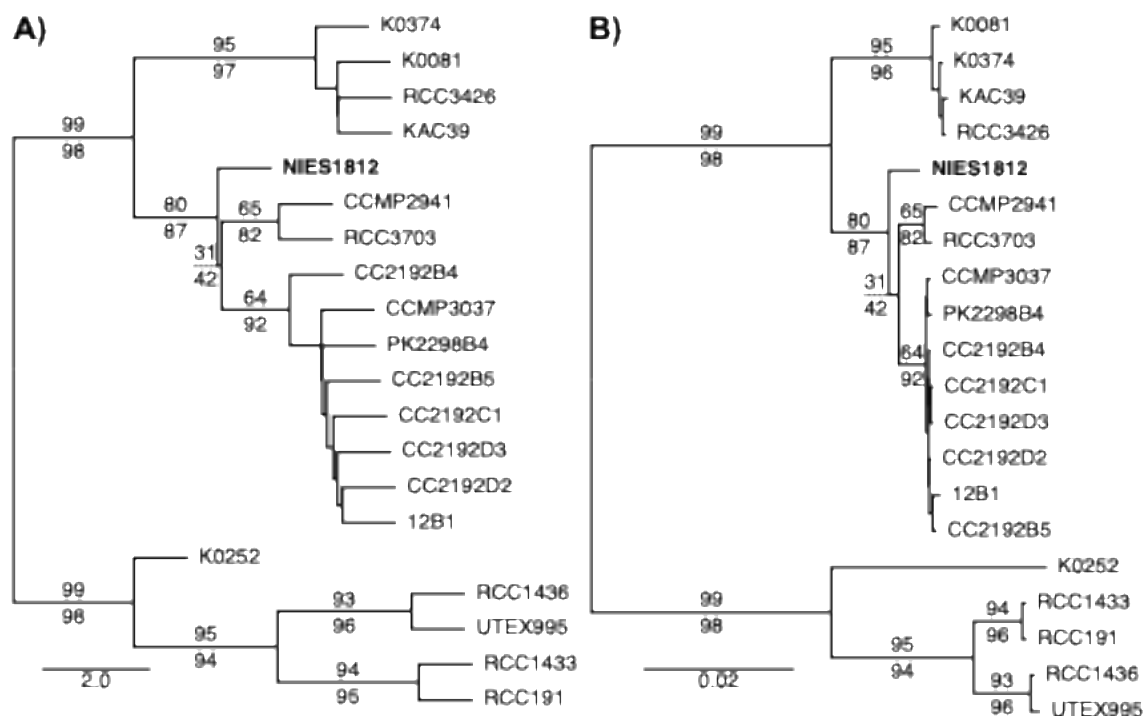


Figure A.3: Phylogenetic trees

A) Coalescent and B) Concatenated species trees for the *P. parvum* s.l. complex (excluding hybrids) using 9,769 single copy genes. Support values for nodes are reported as site/gene concordance factors above/below the preceding branch. All concordance factors for major clades of *P. parvum* are reported. Ultrafast bootstrap support for all branches was ≥ 99 . NIES1812 is bolded to highlight its position.

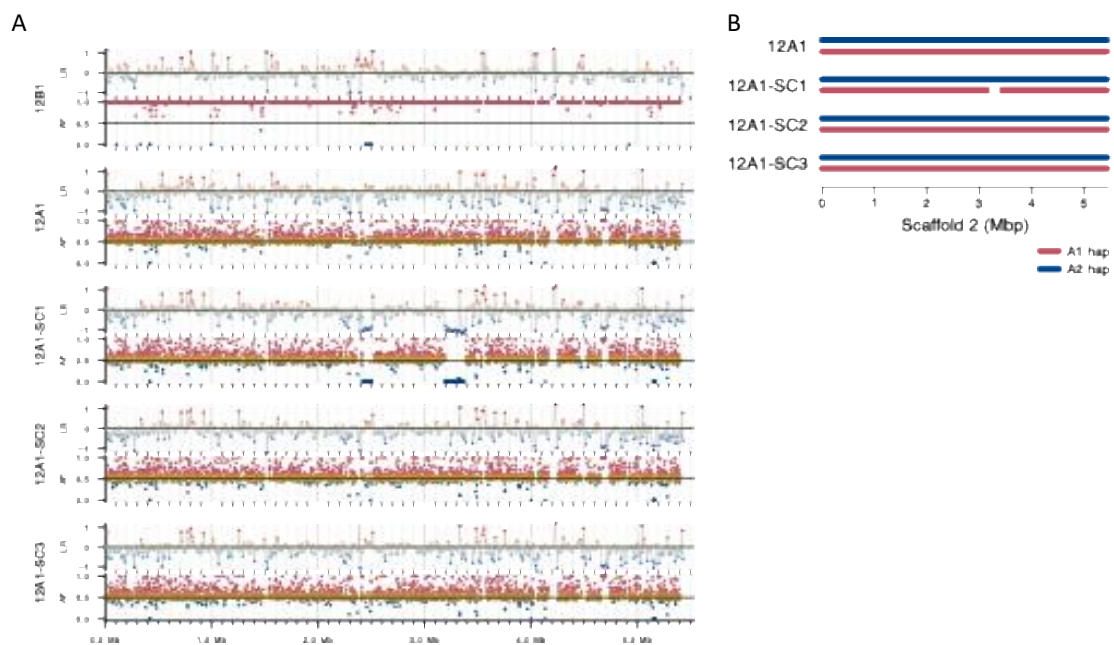


Figure A.4: Scaffold 2 structural anomalies in UTEX3226 single cell reisolates.

A) Top track indicates log2 ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

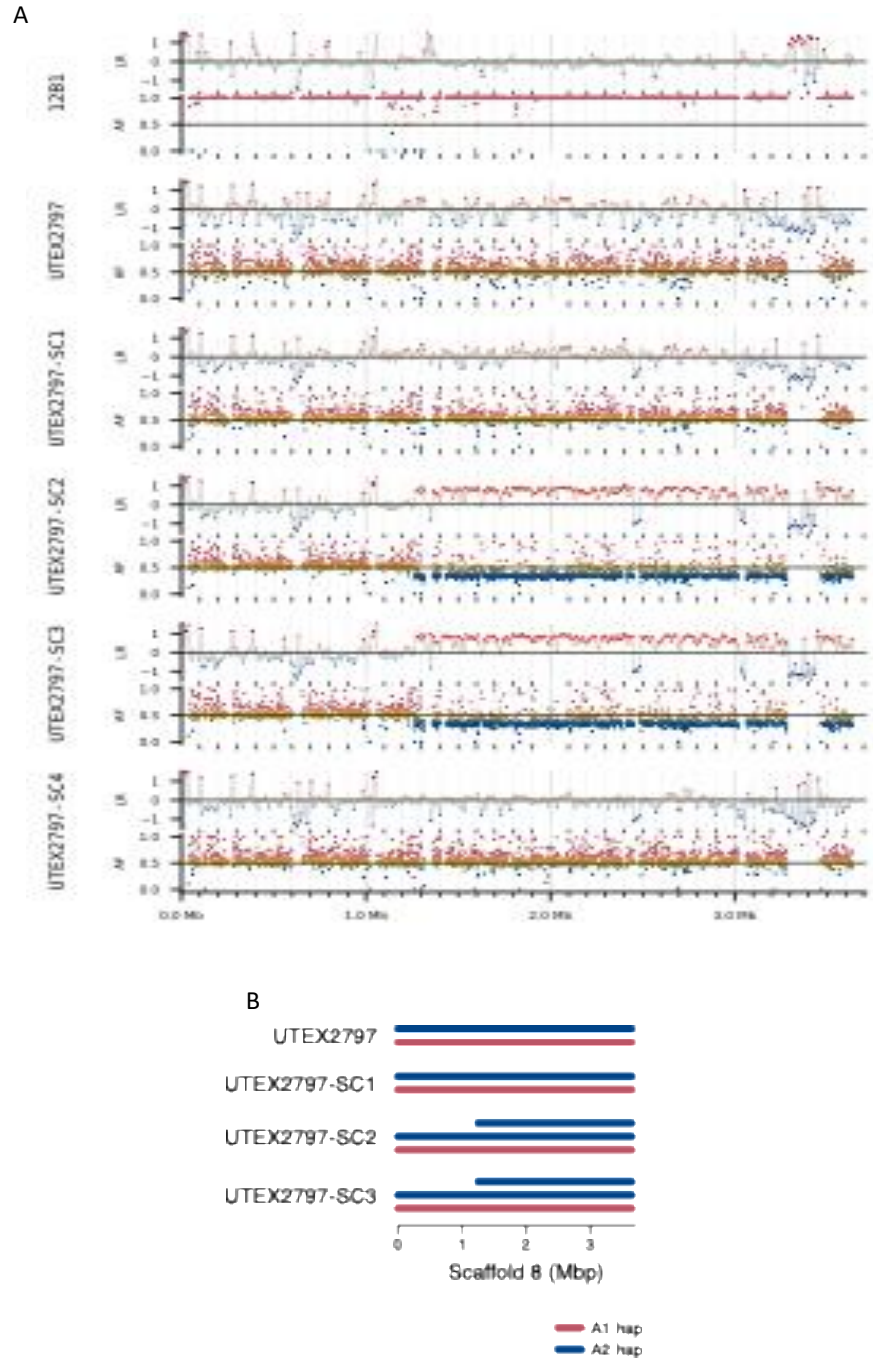


Figure A.5: Scaffold 8 structural anomalies in UTEX2797 single cell reisolates.

A) Top track indicates log2 ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

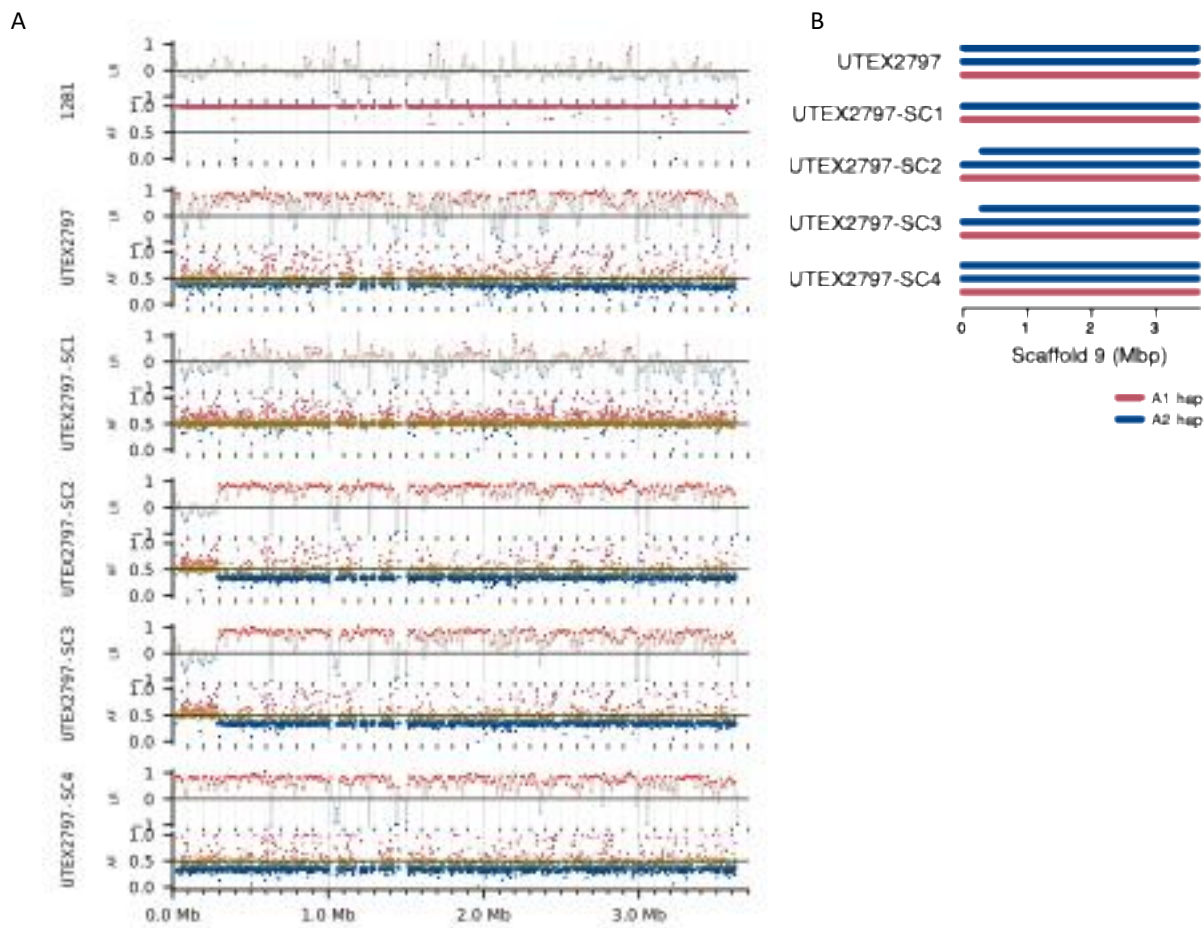


Figure A.6: Scaffold 9 structural anomalies in UTEX2797 single cell reisolates.

A) Top track indicates log₂ ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

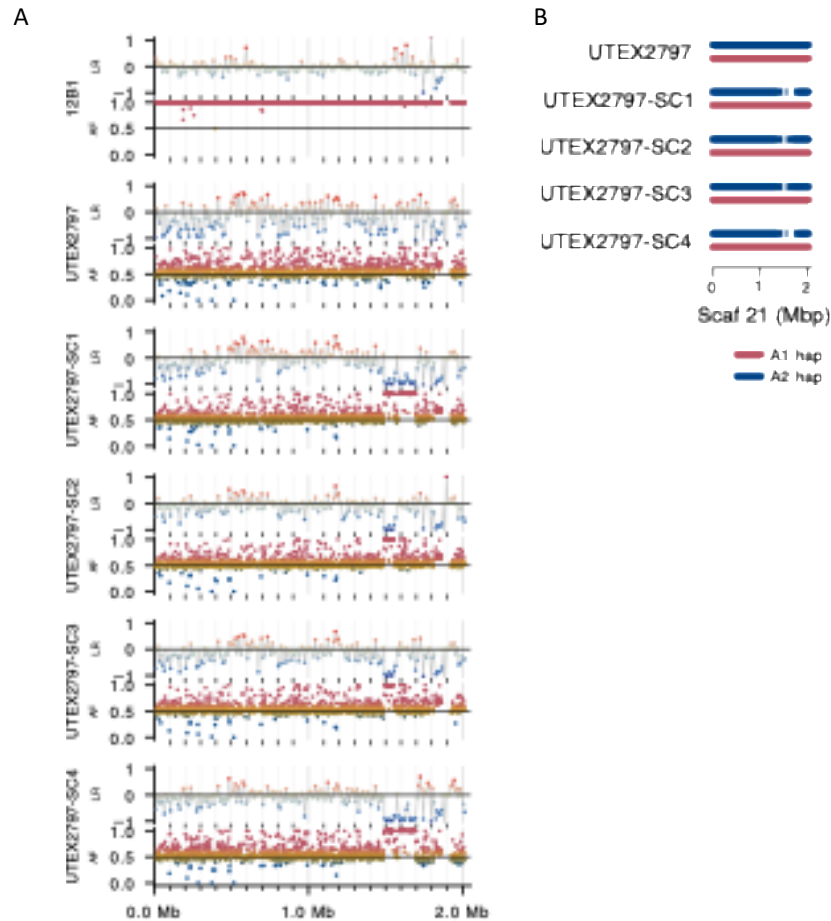


Figure A.7: Scaffold 21 structural anomalies in UTEX2797 single cell reisolates.

A) Top track indicates log2 ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

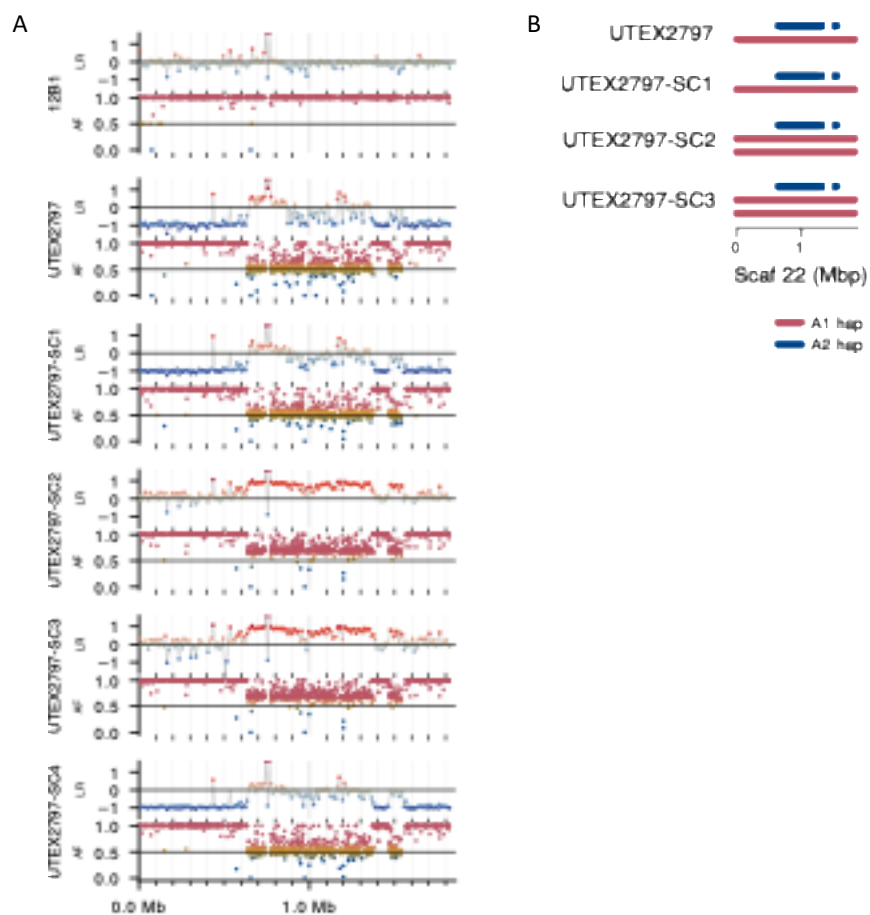


Figure A.8: Scaffold 22 structural anomalies in UTEX2797 single cell reisolates.

A) Top track indicates log2 ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

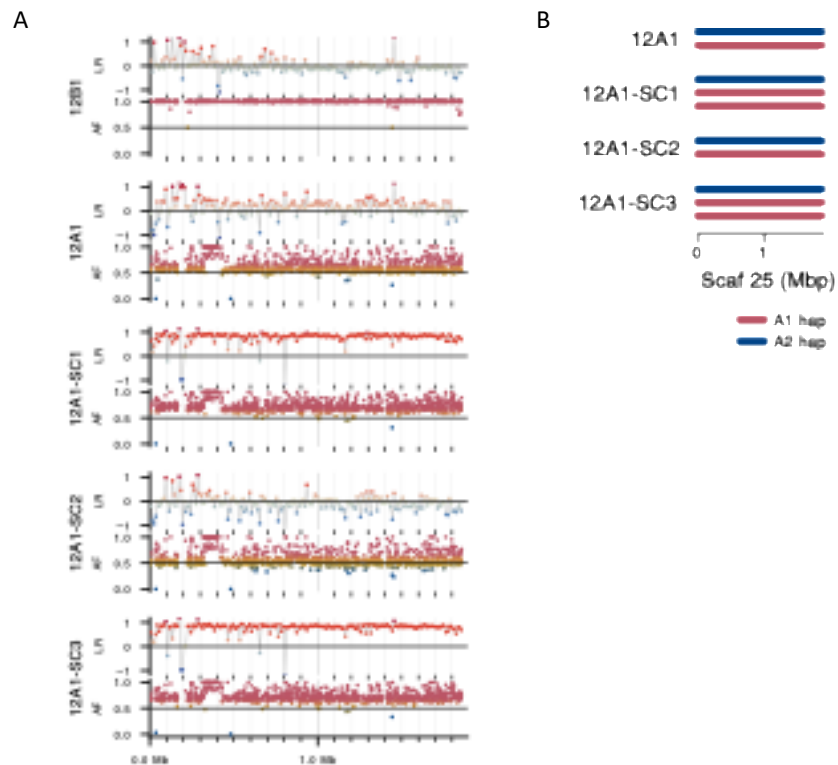


Figure A.9: Scaffold 25 structural anomalies in UTEX3226 single cell reisolates.

A) Top track indicates log2 ratio (LR) of normalized genome coverage; bottom track indicates A1 AIM allele frequency (AF; A1 read depth / total read depth). B) Model of scaffold organization.

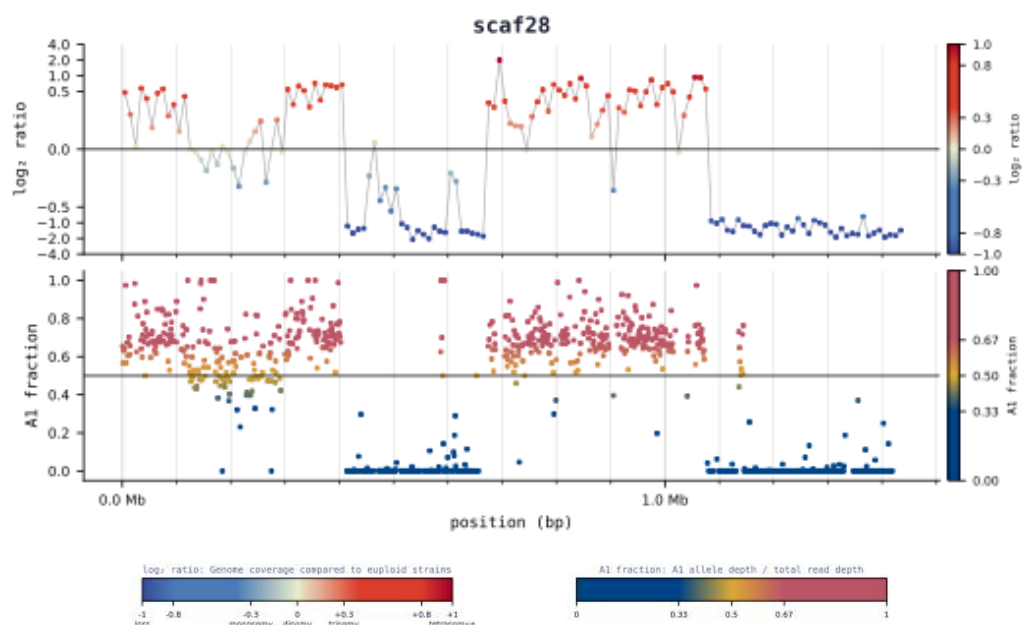


Figure A.10: Variation in coverage and allele frequency in Scaffold 28 in strain ARC82.

Top track indicates $\log_2$ ratio of normalized genome coverage; bottom track indicates A1 AIM allele frequency (A1 read depth / total read depth).

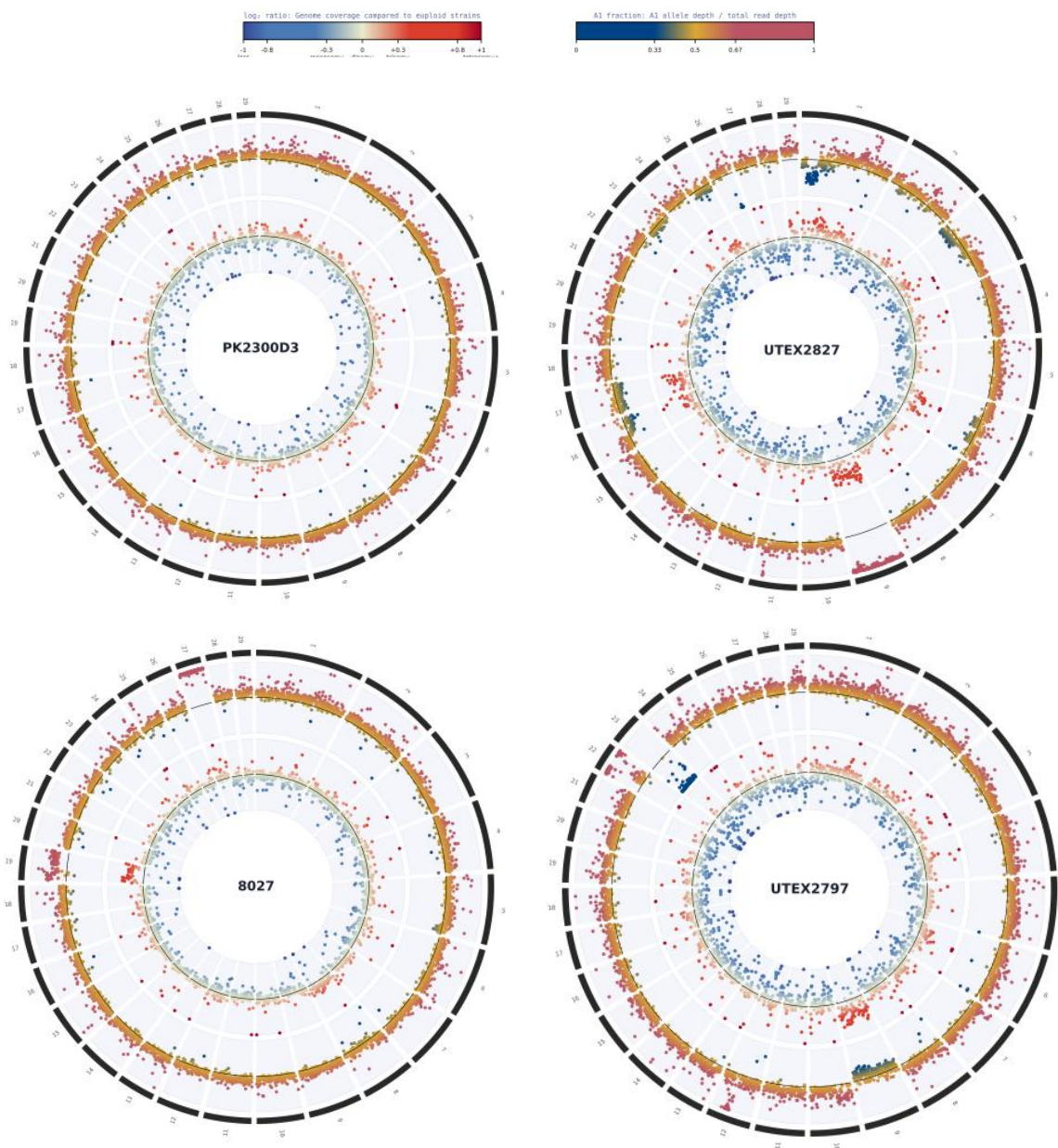


Figure A.11: Circular genome plots.

Inner tracks indicate log₂ ratio of genome coverage; outer tracks indicates A1 allele fraction (A1 read depth / total read depth). Strain PK2300D3 (top left) showed no structural anomalies. Strain UTEX2797 and 8027 (bottom plots) how examples of abnormalities on Scaffolds 9, 19, 22, 23, and 27. High variability in genome coverage and allele frequency in strain UTEX2827 (top right) prevented an accurate count of structural anomalies.

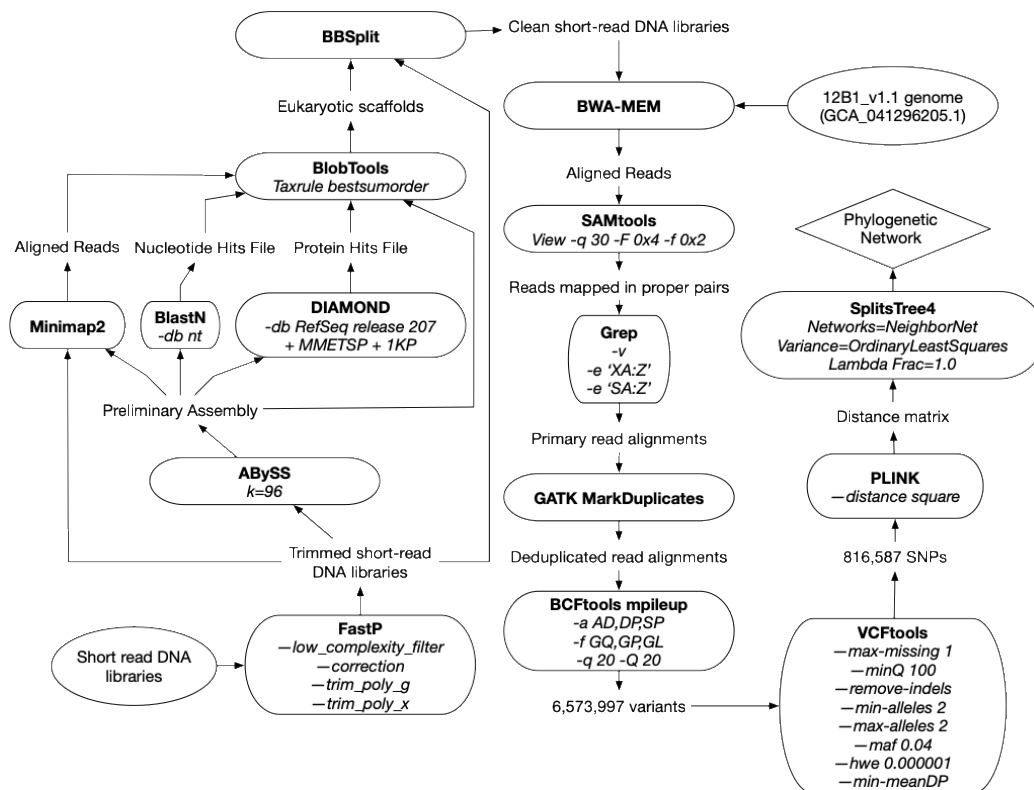
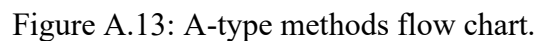


Figure A.12: Preliminary network flow chart.

Flow chart describing the process from short-read libraries to phylogenetic network used to identify A-type strains.



127

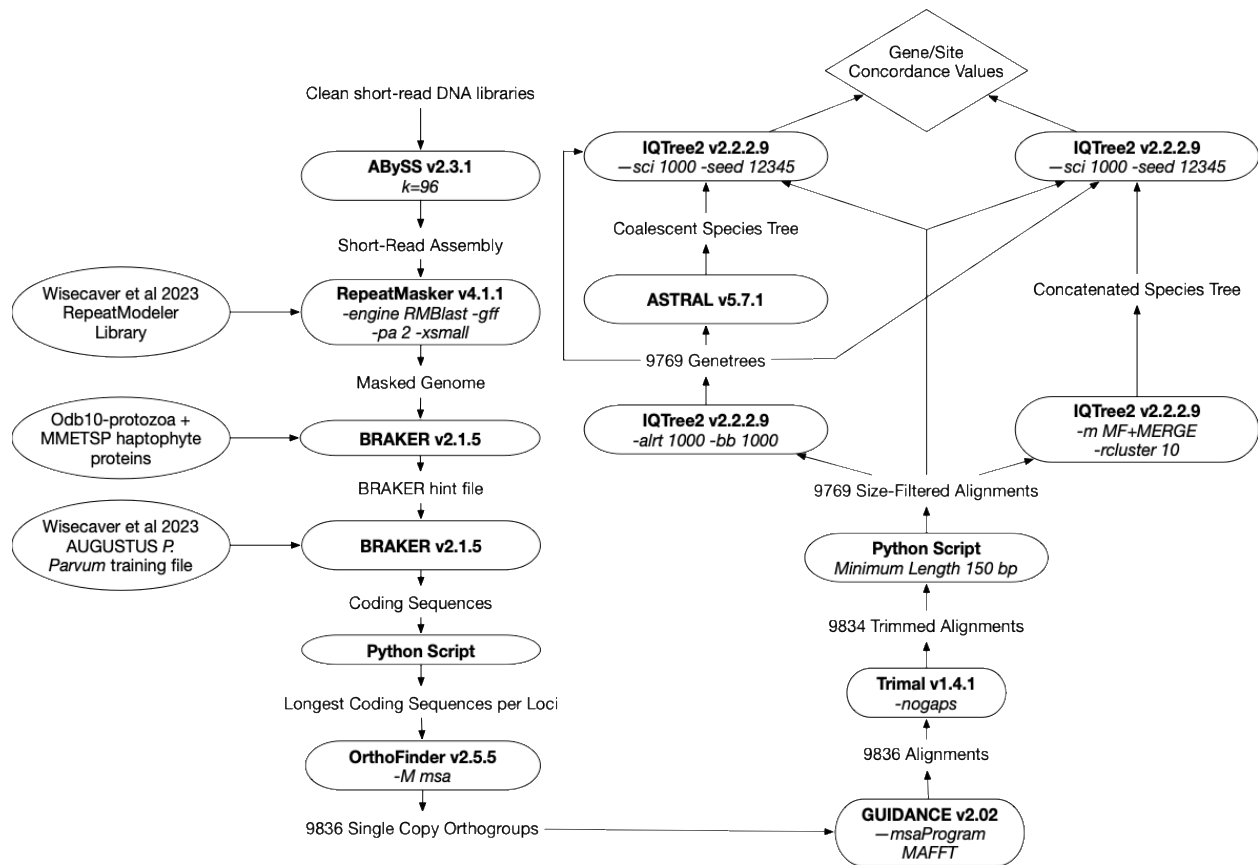


Figure A.14: Species tree creation methods flow chart.

Flow chart depicting Orthofinder analysis of all non-hybrid A, B, and C-strains, note that filtered short-read libraries are identical to those created at the start of Figure S12.

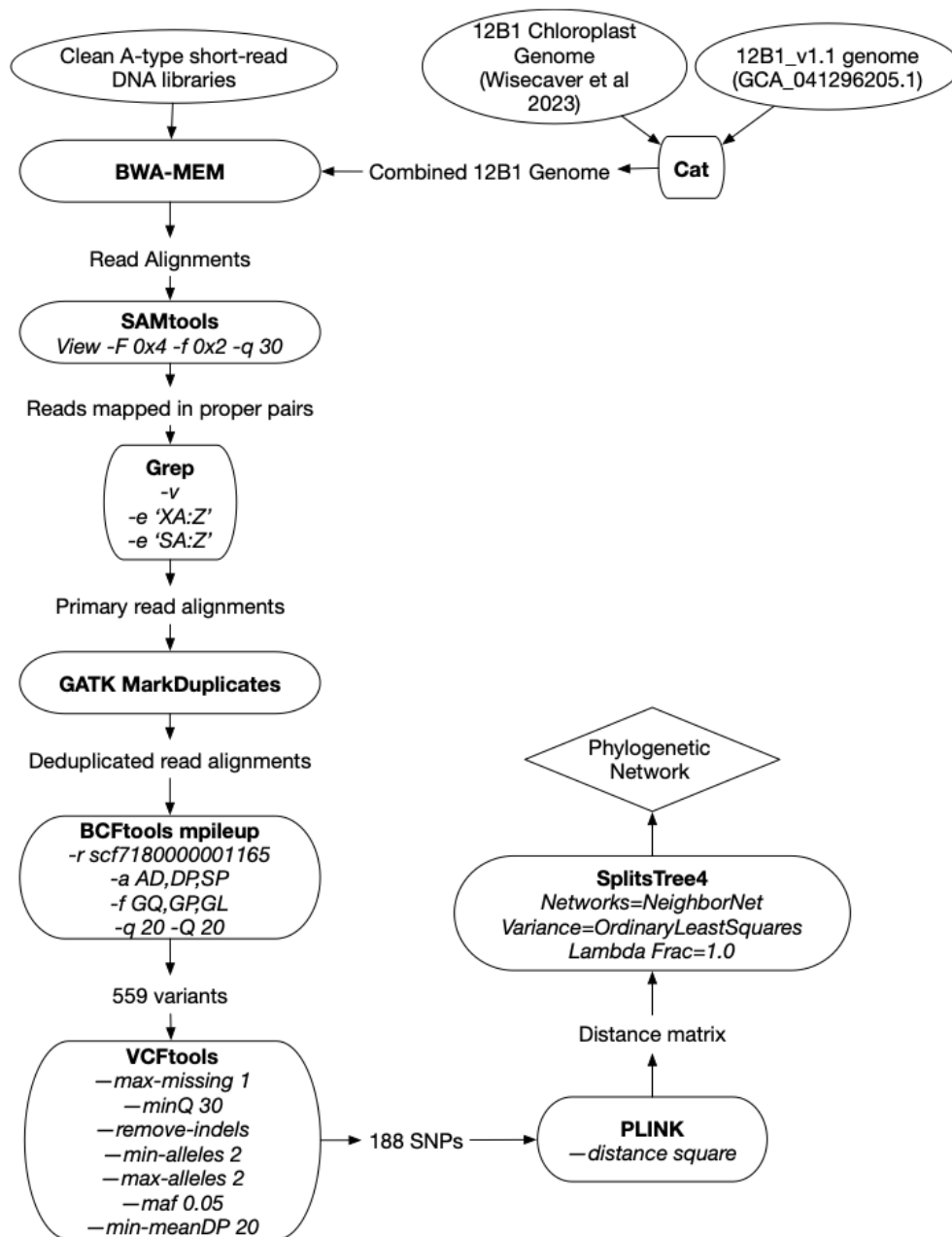


Figure A.15: Chloroplast methods flow chart.
Flow chart describing chloroplast specific analyses.

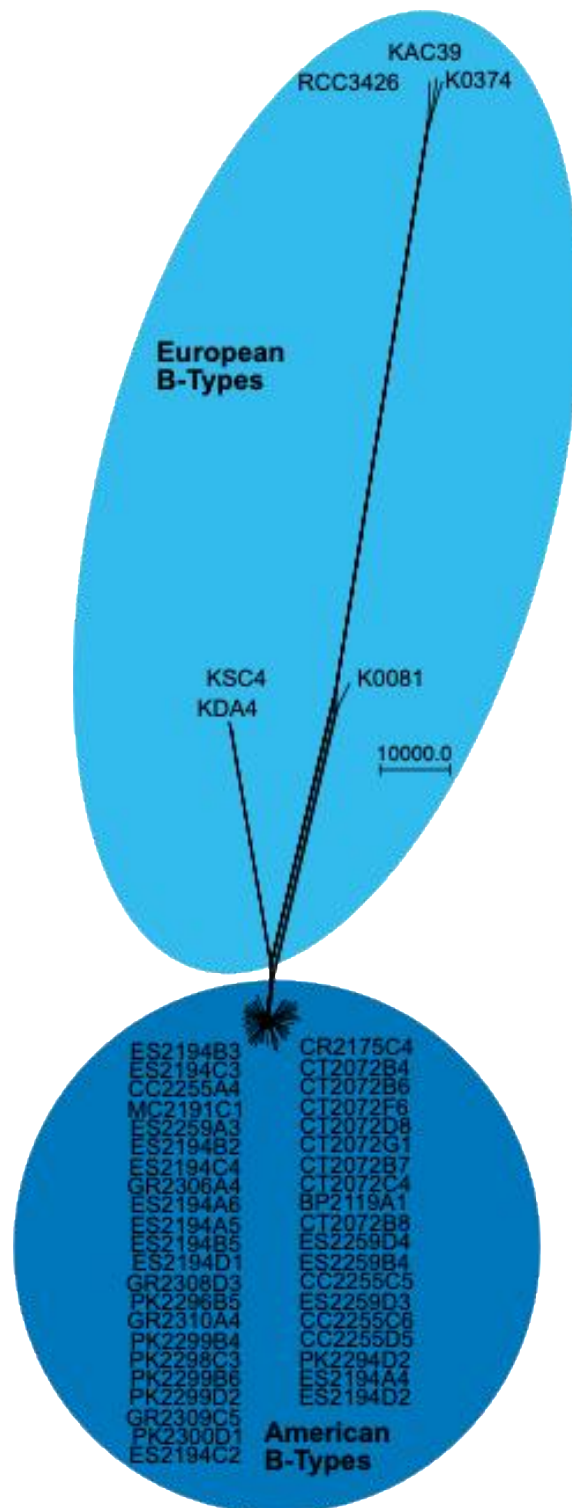


Figure A.16: B-type *P. parvum* s.l. network

Phylogenetic network created from 136,987 SNPs called against the ODER1 Haplotype 1 B-type reference genome.

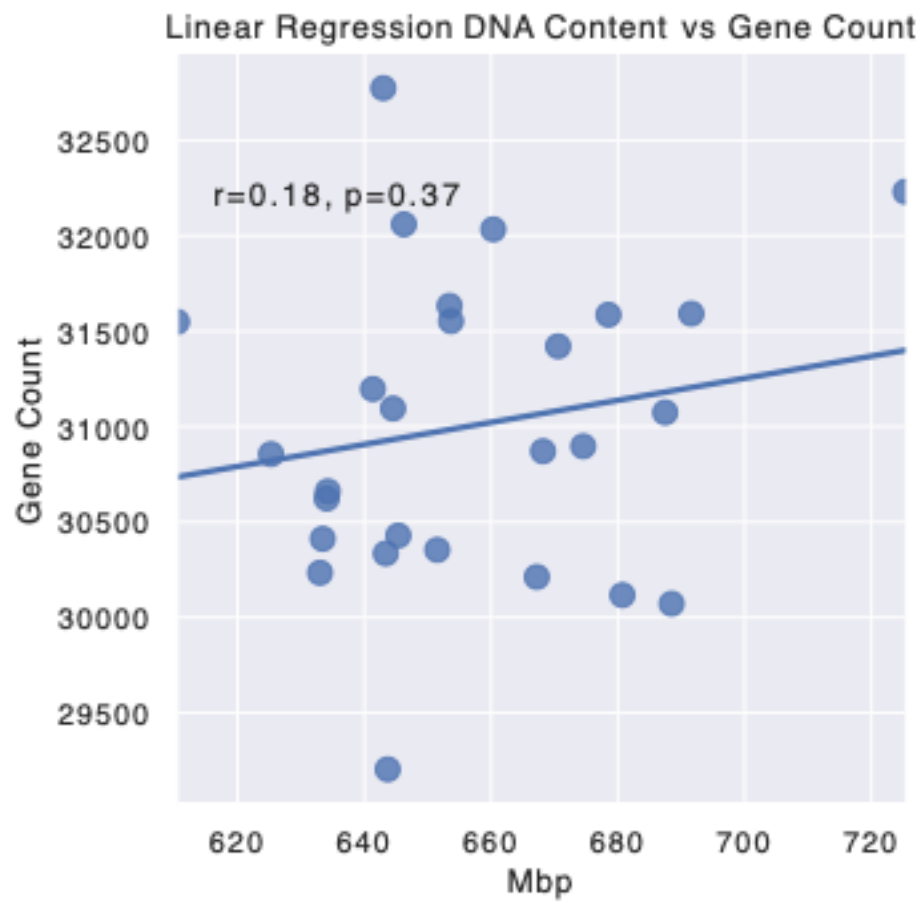


Figure A.17: B-type genome size vs gene content linear regression

APPENDIX B. SUPPLEMENTAL TABLES

Table B.1: Strain information

All cultures were maintained in L1-Si media prepared using Instant Ocean at the salinity levels indicated below. *CCMP2941 is not maintained in the lab currently.

ID	Synonyms	Salinity (PSU)	Location	Year Isolated
8020		6	Lake Possum Kingdom, TX, USA	2013
8027		6	Lake Possum Kingdom, TX, USA	2013
13A5		6	Lake EV Spence, TX, USA	2010
ARC66		6	Hilton Head, SC, USA	2001
ARC81		6	Artesian Aquafer, Elizabeth City, NC, USA	2002
ARC82		6	Artesian Aquafer, Elizabeth City, NC, USA	2002
ARC83		6	Artesian Aquafer, Elizabeth City, NC, USA	2002
ARC84		6	Artesian Aquafer, Elizabeth City, NC, USA	2002
BP2119A1		6	Beal Park, TX, USA	2019
CC2192B4		6	Lake Colorado City, TX, USA	2019
CC2192B5		6	Lake Colorado City, TX, USA	2019
CC2192C1		6	Lake Colorado City, TX, USA	2019
CC2192D2		6	Lake Colorado City, TX, USA	2019
CC2192D3		6	Lake Colorado City, TX, USA	2019
CC2255A4		6	Lake Colorado City, TX, USA	2020
CC2255C5		6	Lake Colorado City, TX, USA	2020
CC2255C6		6	Lake Colorado City, TX, USA	2020
CC2255D5		6	Lake Colorado City, TX, USA	2020
CCMP1926		6	Bigelow Laboratory dock, W Boothbay Harbor, ME USA	1997
CCMP1927		6	ME	1999
CCMP2941	A10,785c	N/A	Lake Repnoye, Russia	2007
CCMP3037		6	Twin Buttes Lake, Laramie, WY, USA	2008
CR2175C4		6	Concho River, TX, USA	2019
CT2072B4		6	Comanche Trails Park, TX, USA	2018
CT2072B6		6	Comanche Trails Park, TX, USA	2018
CT2072B7		6	Comanche Trails Park, TX, USA	2018
CT2072B8		6	Comanche Trails Park, TX, USA	2018
CT2072C4		6	Comanche Trails Park, TX, USA	2018
CT2072D8		6	Comanche Trails Park, TX, USA	2018
CT2072F6		6	Comanche Trails Park, TX, USA	2018
CT2072G1		6	Comanche Trails Park, TX, USA	2018
ES2194A4		6	Lake EV Spence, TX, USA	2019
ES2194A5		6	Lake EV Spence, TX, USA	2019
ES2194A6		6	Lake EV Spence, TX, USA	2019
ES2194B2		6	Lake EV Spence, TX, USA	2019
ES2194B3		6	Lake EV Spence, TX, USA	2019
ES2194B5		6	Lake EV Spence, TX, USA	2019
ES2194C2		6	Lake EV Spence, TX, USA	2019
ES2194C3		6	Lake EV Spence, TX, USA	2019
ES2194C4		6	Lake EV Spence, TX, USA	2019
ES2194D1		6	Lake EV Spence, TX, USA	2019
ES2194D2		6	Lake EV Spence, TX, USA	2019
ES2259A3		6	Lake EV Spence, TX, USA	2020
ES2259B4		6	Lake EV Spence, TX, USA	2020
ES2259D3		6	Lake EV Spence, TX, USA	2020
ES2259D4		6	Lake EV Spence, TX, USA	2020
GR2306A4		6	Lake Granbury, TX, USA	2020

Table B.1: continued

ID	Synonyms	Salinity (PSU)	Location	Year Isolated
GR2308D3		6	Lake Granbury, TX, USA	2020
GR2309C5		6	Lake Granbury, TX, USA	2020
GR2310A4		6	Lake Granbury, TX, USA	2020
K0081		6	Flade Sø, Thy. Denmark	1985
K0252	RCC3427,UIO55	6	Wisons Promontor, Norman Bay, Victoria, Australia	1987
K0374	UIO56,RCC3428	6	Norway	1989
KAC39		6	Oslo, Norway	Unknown
MC2191C1		6	Moss Creek Lake, TX, USA	2019
NIES1017	#22	32	Shimajiri, Hirara, Okinawa, Japan	1997
NIES1018	M-25	32	Shimajiri, Hirara, Okinawa, Japan	2002
NIES1812	TKB-251	32	Yufu Isl., Okinawa, Japan	2005
PK2294D2		6	Lake Possum Kingdom, TX, USA	2020
PK2296B5		6	Lake Possum Kingdom, TX, USA	2020
PK2297A3		6	Lake Possum Kingdom, TX, USA	2020
PK2297B2		6	Lake Possum Kingdom, TX, USA	2020
PK2297B5		6	Lake Possum Kingdom, TX, USA	2020
PK2297D4		6	Lake Possum Kingdom, TX, USA	2020
PK2298B4		6	Lake Possum Kingdom, TX, USA	2020
PK2298C3		6	Lake Possum Kingdom, TX, USA	2020
PK2298C6		6	Lake Possum Kingdom, TX, USA	2020
PK2299B3		6	Lake Possum Kingdom, TX, USA	2020
PK2299B4		6	Lake Possum Kingdom, TX, USA	2020
PK2299B6		6	Lake Possum Kingdom, TX, USA	2020
PK2299D2		6	Lake Possum Kingdom, TX, USA	2020
PK2300D1		6	Lake Possum Kingdom, TX, USA	2020
PK2300D3		6	Lake Possum Kingdom, TX, USA	2020
RCC1433	AC36,HAP36	32	Sallanelles, France	2007
RCC1436	AC45,HAP45,AR-69	32	Plestin Les Grèves, France	1977
RCC191	CCAP946/4,946/4,DH189,PLY5 27,RCC3705,CCMP709	32	West End of Fleet, Dorset, UK	Unknown
RCC3426	UIO54	32	Ryfylke, Norway	1989
RCC3703	PRYM,CCAP946/6,Millport 65, SMB A 65,CCMP708,CCAP946/6	6	Millport, Isle of Cumbrae, UK	1953
UTEX2797		6	US Hwy 183, Texas Colorado River, TX, USA	2002
UTEX2797- SC1		6	Single cell reisolate of UTEX2797	
UTEX2797- SC2		6	Single cell reisolate of UTEX2797	
UTEX2797- SC3		6	Single cell reisolate of UTEX2797	
UTEX2797- SC4		6	Single cell reisolate of UTEX2797	
UTEX2827		32	Charleston, SC, USA	2002
UTEX3226	12A1	6	Moss Creek Lake, TX, USA	2010
UTEX3226- SC1		6	Single cell reisolate of UTEX3226	
UTEX3226- SC2		6	Single cell reisolate of UTEX3226	
UTEX3226- SC3		6	Single cell reisolate of UTEX3226	
UTEX3227	12B1	6	Moss Creek Lake, TX, USA	2010
UTEX995	CCAP946/1B, NIVA3/88,SAG B 127.79,Plymouth 94,RCC3706	6	River Blackwater, Essex, UK	1952

Table B.2: A-type flow cytometry

Flow cytometry RFI was converted to pg DNA using a standard curve composed of UTEX3227 (0.11796 pg) CCMP3037 (0.24172 pg), UTEX2797 (0.28003 pg), UTEX3226 (0.29731 pg), RCC3703 (0.28609 pg), K0081 (0.86291 pg). Pg DNA was converted to mbp DNA using the formula: 1 pg = 0.98×10^9 base pairs

Strain ID	Ploidy	Flow Cytometry Replications	Estimated DNA content (pg, mean)	Estimated DNA content (pg, std)	Estimated DNA content (Mbp, mean)	Estimated DNA content (Mbp, std)
UTEX3227	haploid	3	0.109655	0.00421	107	4.125626
NIES1812	haploid	3	0.122746	0.004183	120	4.098953
PK2298B4	haploid	3	0.122192	0.002281	120	2.234957
8020	diploid	3	0.293523	0.002941	288	2.882497
8027	diploid	3	0.277652	0.001153	272	1.129454
UTEX3226	diploid	3	0.280284	0.003845	275	3.767983
13A5	diploid	3	0.287597	0.002381	282	2.33316
ARC140	diploid	3	0.283379	0.000945	278	0.926222
ARC66	diploid	3	0.275731	0.003934	270	3.855211
ARC81	diploid	3	0.273563	0.001766	268	1.731155
ARC82	diploid	3	0.28895	0.001041	283	1.020607
ARC83	diploid	3	0.275282	0.00516	270	5.056746
ARC84	diploid	3	0.278837	0.003863	273	3.785792
CC2192B4	diploid	3	0.27623	0.002697	271	2.643114
CC2192B5	diploid	3	0.27076	0.002797	265	2.740665
CC2192C1	diploid	3	0.282513	0.005158	277	5.054615
CC2192D2	diploid	3	0.265828	0.002856	261	2.798967
CC2192D3	diploid	3	0.266254	0.001598	261	1.566157
CCMP3037	diploid	3	0.251411	0.004771	246	4.675246
K0081	diploid	3	0.861559	0.001363	844	1.335784
PK2297D4	diploid	3	0.289401	0.002109	284	2.066395
PK2298C6	diploid	3	0.286181	0.002449	280	2.400049
PK2299B3	diploid	3	0.284832	0.001744	279	1.709153
PK2300D3	diploid	3	0.274963	0.010056	269	9.855301
RCC3703	diploid	3	0.303326	0.006932	297	6.793744
UTEX2797	diploid	3	0.279789	0.001691	274	1.656744
UTEX2827	diploid	3	0.295432	0.001949	290	1.909688

Table B.3: B-type flow cytometry.

Flow cytometry results for Texas B-type strains, results were converted from RFI to pg DNA using a standard curve of *A. thaliana* endoreduplicated nuclei and *C. vulgaris* nuclei. pg DNA was converted to Mbp using the formula $1 \text{ pg} = 0.98 \text{E}9 \text{ base pairs}$. All strains assayed in this table are diploid.

Strain ID	Flow Cytometry Replications	Estimated DNA content (pg, mean)	Estimated DNA content (pg, std)	Estimated DNA content (Mbp, mean)	Estimated DNA content (Mbp, std)
BP2119A1	4	0.701493	0.03215	687	31.50737
CC2255C5	3	0.667021	0.004977	653	4.877171
CC2255D5	4	0.688293	0.031762	674	31.1263
CR2175C4	3	0.657679	0.031803	644	31.16699
CT2072B4	3	0.647183	0.024409	634	23.92126
CT2072B6	5	0.681819	0.025407	668	24.89863
CT2072B7	4	0.647007	0.01832	634	17.95324
CT2072B8	4	0.65943	0.021191	646	20.76721
CT2072C4	3	0.656829	0.033268	643	32.60241
CT2072D8	4	0.673818	0.01805	660	17.68864
CT2072F6	3	0.656118	0.022167	642	21.72372
CT2072G1	3	0.692364	0.010343	678	10.13631
ES2194A5	4	0.702529	0.014997	688	14.69746
ES2194A6	4	0.694598	0.024164	680	23.68119
ES2194B3	4	0.664804	0.029262	651	28.67669
ES2194B5	3	0.680828	0.017731	667	17.37631
ES2194C3	3	0.645929	0.038273	633	37.50801
ES2194D1	3	0.656513	0.023957	643	23.47807
ES2259B4	3	0.646364	0.013654	633	13.38132
ES2259D3	3	0.658538	0.017053	645	16.71169
GR2308D3	3	0.7057	0.012075	691	11.83366
GR2309C5	5	0.654438	0.031514	641	30.88412
PK2294D2	5	0.666778	0.029845	653	29.24778
PK2296B5	3	0.622905	0.039425	610	38.63699
PK2298C3	4	0.684258	0.032616	670	31.96324
PK2299B4	4	0.740173	0.078949	725	77.37026
PK2300D1	3	0.638046	0.030253	625	29.64801

Table B.4: Ancestry informative markers

Strains CC2192B4, CC2192D2, CC2192B5, CC2192D3, CC2192C1, CCMP3037, CCMP2941, and RCC3703 are not shown as they are used as in the A1/A2 reference populations and cannot have AIMS apart from those reference populations

Strain ID	Subclade	A1 homozygous count	A1 homozygous %	A2 homozygous count	A2 homozygous %	A1-A2 heterozygous count	A1-A2 heterozygous %
NIES1812	A3	141,100	50.77%	131,017	47.14%	5,810	2.09%
UTEX3227	A1	276,873	99.62%	986	0.35%	68	0.02%
PK2298B4	A1	276,968	99.65%	808	0.29%	151	0.05%
8020	A hybrid	24,554	8.83%	1,476	0.53%	251,897	90.63%
8027	A hybrid	17,705	6.37%	1,458	0.52%	258,764	93.11%
UTEX3226	A hybrid	32,952	11.86%	1,257	0.45%	243,718	87.69%
UTEX3226-SC1	A hybrid	24,622	8.86%	1,792	0.64%	251,513	90.50%
UTEX3226-SC2	A hybrid	25,313	9.11%	1,285	0.46%	251,329	90.43%
UTEX3226-SC3	A hybrid	24,206	8.71%	1,482	0.53%	252,239	90.76%
13A5	A hybrid	33,680	12.12%	1,247	0.45%	243,000	87.43%
ARC66	A hybrid	14,620	5.26%	2,675	0.96%	260,632	93.78%
ARC81	A hybrid	14,109	5.08%	2,269	0.82%	261,549	94.11%
ARC82	A hybrid	17,073	6.14%	7,690	2.77%	253,164	91.09%
ARC83	A hybrid	15,625	5.62%	2,028	0.73%	260,274	93.65%
ARC84	A hybrid	15,864	5.71%	2,562	0.92%	259,501	93.37%
PK2297D4	A hybrid	16,392	5.90%	5,456	1.96%	256,079	92.14%
PK2298C6	A hybrid	13,111	4.72%	1,667	0.60%	263,149	94.68%
PK2299B3	A hybrid	20,136	7.25%	1,846	0.66%	255,945	92.09%
PK2300D3	A hybrid	15,555	5.60%	1,606	0.58%	260,766	93.83%
UTEX2797	A hybrid	26,541	9.55%	8,961	3.22%	242,425	87.23%
UTEX2797-SC1	A hybrid	21,453	7.72%	8,571	3.08%	247,903	89.20%
UTEX2797-SC2	A hybrid	21,532	7.75%	9,433	3.39%	246,962	88.86%
UTEX2797-SC3	A hybrid	20,824	7.49%	9,429	3.39%	247,674	89.11%
UTEX2797-SC4	A hybrid	23,956	8.62%	8,339	3.00%	245,632	88.38%
UTEX2827	A hybrid	29,773	10.71%	3,303	1.19%	244,851	88.10%

Table B.5: AMOVA

Hierarchical AMOVA investigating the effects of chloroplast haplotype and geographic region on molecular variation present in A1×A2 hybrids. * Estimated variance for chloroplast haplotype (cpHap) in Test 1 was negative and therefore excluded from summation of total variance and from proportion calculation.

Test	Source	Estimated variance	Proportion	Φ	P value
1	cpHap	-676	-- *	$\Phi_{CT} = -0.055$	0.344
1	Region (within cpHap)	5391	42%*	$\Phi_{SC} = 0.419$	<0.001
1	Within region	7477	58%*	--	
1	Total	12868	--	$\Phi_{ST} = 0.387$	
2	Region	882	7%	$\Phi_{CT} = 0.069$	0.333
2	cpHap (within region)	4399	34%	$\Phi_{SC} = 0.370$	<0.001
2	Within cpHap	7477	59%	--	
2	Total	12758	--	$\Phi_{ST} = 0.414$	

Table B.6: A-type strain anomalies

Anomaly type	Strain	Scaffold	Anomaly start	Anomaly end	Anomaly subtype
interstitial variant	ARC82	1	6.3	6.38	A2 interstitial segmental loss
interstitial variant	PK2299B3	2	3.2	3.32	A1 interstitial segmental loss
telomeric variant	CCMP3037	3	4.4	5.2	telomeric segmental duplication
telomeric variant	RCC3703	3	4.56	5.05	telomeric segmental duplication
interstitial variant	8020	4	3.45	3.7	A1 interstitial segmental loss
interstitial variant	ARC66	4	2.24	2.44	A2 interstitial segmental loss
interstitial variant	ARC81	4	2.24	2.44	A2 interstitial segmental loss
interstitial variant	ARC82	4	2.55	3.45	A1 interstitial segmental duplication
interstitial variant	ARC82	4	2.9	3.3	A2 interstitial segmental loss
interstitial variant	ARC83	4	2.24	2.44	A2 interstitial segmental loss
interstitial variant	ARC84	4	2.24	2.44	A2 interstitial segmental loss
interstitial variant	CCMP3037	4	0.75	0.9	interstitial segmental duplication
interstitial variant	ARC82	6	4.35	4.6	A2 interstitial segmental loss
chromosomal variant	UTEX3226	8	0	3.65	A2-derived polysomy
chromosomal variant	13A5	8	0.39	3.65	A2-derived polysomy
telomeric variant	13A5	8	0	0.39	A2 telomeric segmental loss
interstitial variant	ARC82	8	1.3	1.55	A1 interstitial segmental duplication
chromosomal variant	PK2297D4	8	0	3.65	A2-derived polysomy
chromosomal variant	PK2298C6	8	0	3.65	A2-derived polysomy
chromosomal variant	PK2299B3	8	0	3.65	A2-derived polysomy
interstitial variant	RCC3703	8	1.45	2.1	interstitial segmental loss
chromosomal variant	UTEX2797	9	0	3.65	A2-derived polysomy
chromosomal variant	UTEX2827	9	0	3.65	A1 disomy
interstitial variant	UTEX2797	12	1.6	1.78	A2 interstitial segmental loss
interstitial variant	UTEX2797	12	1.82	2	A2 interstitial segmental loss
telomeric variant	ARC66	13	2.45	2.6	A1 telomeric segmental loss
telomeric variant	ARC81	13	2.45	2.6	A1 telomeric segmental loss
telomeric variant	ARC83	13	2.45	2.6	A1 telomeric segmental loss
telomeric variant	ARC84	13	2.45	2.6	A1 telomeric segmental loss
interstitial variant	ARC82	16	1.15	2.15	A1 Interstitial segmental loss
interstitial variant	CCMP3037	18	1.17	1.8	interstitial segmental duplication
telomeric variant	CCMP3037	18	0	0.22	telomeric segmental loss
chromosomal variant	8027	19	0	2.4	A1-derived polysomy

Table B.6: continued

Anomaly type	Strain	Scaffold	Anomaly start	Anomaly end	Anomaly subtype
chromosomal variant	UTEX3226	19	0	2.4	A1-derived polysomy
chromosomal variant	UTEX3227	19	0	2.4	A1-derived polysomy
chromosomal variant	13A5	19	0	2.4	A1-derived polysomy
chromosomal variant	ARC82	19	0	2.4	A1-derived polysomy
chromosomal variant	PK2297D4	19	0	2.4	A1-derived polysomy
chromosomal variant	PK2298B4	19	0	2.4	A1-derived polysomy
chromosomal variant	PK2298C6	19	0	2.4	A1-derived polysomy
chromosomal variant	PK2298C6	19	0	2.4	A1-derived polysomy
telomeric variant	RCC3703	20	0	0.5	telomeric segmental loss
telomeric variant	RCC3703	22	0.5	1.85	telomeric segmental loss
interstitial variant	UTEX2797	22	1.37	1.47	A2 interstitial segmental loss
telomeric variant	UTEX2797	22	1.55	1.85	A2 telomeric segmental loss
telomeric variant	UTEX2797	22	0	0.64	A2 telomeric segmental loss
chromosomal variant	UTEX2797	23	0	2.05	A2 monosomy
telomeric variant	ARC82	24	0	0.1	A2 telomeric segmental loss
chromosomal variant	13A5	25	0	1.9	A1-derived polysomy
telomeric variant	RCC3703	25	0.2	0.9	telomeric segmental loss
interstitial variant	ARC66	26	1.48	1.54	A1 interstitial segmental duplication
interstitial variant	ARC81	26	1.48	1.54	A1 interstitial segmental duplication
interstitial variant	ARC82	26	1.48	1.54	A1 interstitial segmental duplication
interstitial variant	ARC82	26	0.44	0.82	A1 interstitial segmental loss
interstitial variant	ARC83	26	1.48	1.54	A1 interstitial segmental duplication
interstitial variant	ARC84	26	1.48	1.54	A1 interstitial segmental duplication
interstitial variant	UTEX2827	26	1.35	1.57	A1 interstitial segmental loss
chromosomal variant	8027	27	0	1.75	A1 disomy
chromosomal variant	UTEX3226	27	0	1.75	A1 disomy
chromosomal variant	13A5	27	0	1.75	A1 disomy
chromosomal variant	PK2297D4	27	0	1.75	A1 disomy
chromosomal variant	PK2299B3	27	0	1.75	A1 disomy

Table B.7 B-type strain anomalies

Anomaly Type	DNA Change	Strain	ODER1 Scaffold	Anomaly Start (mbp)	Anomaly Stop (mbp)	Log2 Coverage
Interstitial	Loss	BP2119A1	1	11150000	11275000	-0.6515587
Interstitial	Loss	CT2072B7	1	11150000	11275000	-0.6724
Interstitial	Loss	CT2072B8	1	11150000	11275000	-0.5896
Interstitial	Loss	CT2072B6	1	11450000	12075000	-1.0213
Aneuploid	Gain	ES2194D2	1	0	14654278	0.40905
Interstitial	Loss	ES2259A3	1	8850000	8975000	-0.707
Interstitial	Loss	ES2194B3	2	9850000	10025000	-0.9974
Interstitial	Loss	ES2194C3	2	9850000	10025000	-1.1055
Interstitial	Gain	CC2255C5	3	2450000	3100000	0.46648
Interstitial	Loss	ES2194A5	3	10450000	10650000	-0.5858
Interstitial	Loss	ES2194B3	3	1700000	1900000	-0.8004
Interstitial	Loss	ES2194C3	3	1700000	1900000	-0.8365
Interstitial	Loss	PK2298C3	3	10100000	10775000	-0.8575
Interstitial	Loss	BP2119A1	4	725000	1300000	-0.9864
Interstitial	Loss	BP2119A1	4	1375000	1525000	-0.8317
Interstitial	Loss	CT2072B4	4	9025000	9425000	-1.2816
Interstitial	Loss	GR2309C5	4	2125000	2275000	-0.8735
Interstitial	Gain	MC2191C1	4	11175000	12425000	0.68279
Telomeric	Loss	MC2191C1	4	12425000	12631994	-0.7521
Telomeric	Loss	PK2296B5	4	0	2950000	-0.9665
Interstitial	Loss	PK2296B5	4	3100000	3325000	-0.9228
Interstitial	Loss	ES2194B3	5	6575000	6750000	-0.9494
Interstitial	Loss	ES2194C3	5	6575000	6750000	-1.053
Interstitial	Loss	CR2175C4	7	9450000	9600000	-0.4912
Telomeric	Gain	ES2194B3	7	8425000	9609977	0.54317
Telomeric	Loss	PK2296B5	7	9425000	9609977	-0.7091
Telomeric	Loss	PK2299B6	7	0	250000	-1.4367
Telomeric	Gain	GR2306A4	8	0	4125000	0.31104
Telomeric	Gain	GR2306A4	8	4400000	9298367	0.35238
Telomeric	Gain	GR2310A4	8	0	4225000	0.34843
Interstitial	Gain	GR2310A4	8	4350000	4675000	0.42777
Interstitial	Loss	PK2298C3	8	2000000	2150000	-0.44
Interstitial	Loss	CR2175C4	9	550000	1075000	-1.7627
Interstitial	Loss	CT2072F6	10	3475000	3600000	-0.9981
Interstitial	Gain	ES2194D1	10	4725000	4900000	1.52869
Interstitial	Gain	ES2194D1	10	4950000	5175000	1.35659
Interstitial	Gain	BP2119A1	11	2525000	4125000	0.49235
Interstitial	Loss	CT2072B8	11	7625000	8125000	-1.2248
Telomeric	Loss	CT2072D8	11	6300000	8642580	-0.3302
Telomeric	Gain	MC2191C1	11	0	2525000	0.51644
Interstitial	Gain	BP2119A1	13	275000	1025000	0.61857
Interstitial	Gain	BP2119A1	13	1150000	2275000	0.47192
Interstitial	Loss	CC2255C6	13	1800000	2000000	-0.7039
Interstitial	Loss	CC2255D5	13	1800000	2000000	-0.7482
Interstitial	Loss	GR2308D3	13	500000	1000000	-0.5347
Interstitial	Loss	PK2299B6	13	250000	1050000	-1.4583
Telomeric	Gain	CT2072B7	14	0	4800000	0.42005
Telomeric	Gain	CT2072F6	14	0	4800000	0.3735
Interstitial	Loss	ES2194A5	14	5025000	5650000	-0.6761
Interstitial	Loss	ES2194A5	14	5775000	6750000	-0.7955
Interstitial	Loss	ES2194B3	14	5100000	5650000	-0.5464
Interstitial	Loss	ES2194B3	14	5775000	7650000	-0.7641
Interstitial	Loss	ES2194B3	14	7825000	8075000	-0.6762
Interstitial	Loss	GR2309C5	14	5750000	6700000	-0.8082

Table B.7: continued

Anomaly Type	DNA Change	Strain	ODER1 Scaffold	Anomaly Start (mbp)	Anomaly Stop (mbp)	Log2 Coverage
Interstitial	Loss	PK2300D1	14	6875000	7750000	-0.444
Interstitial	Loss	ES2194B3	15	5875000	6175000	-0.5193
Telomeric	Loss	ES2194B3	15	6300000	7581251	-0.7003
Interstitial	Gain	CT2072B8	16	550000	2450000	0.43745
Telomeric	Loss	PK2299B4	16	6100000	6861420	-1.0799
Telomeric	Loss	ES2194B3	18	5725000	5830881	-0.9145
Telomeric	Loss	ES2194C3	18	5675000	5830881	-0.885
Telomeric	Gain	GR2308D3	18	0	1125000	0.46963
Telomeric	Loss	GR2308D3	18	5575000	5830881	-1.0086
Interstitial	Loss	GR2309C5	18	5450000	5600000	-0.6486
Interstitial	Loss	GR2310A4	18	5450000	5625000	-0.7127
Interstitial	Loss	PK2294D2	18	5450000	5600000	-0.5955
Interstitial	Loss	PK2296B5	18	5450000	5625000	-0.6904
Interstitial	Loss	PK2298C3	18	5450000	5625000	-0.6553
Interstitial	Loss	PK2299B4	18	5450000	5625000	-0.8183
Interstitial	Loss	PK2299B6	18	5450000	5600000	-0.6635
Interstitial	Loss	PK2300D1	18	5450000	5625000	-0.6853
Telomeric	Loss	PK2299B6	18	0	725000	-0.6146
Interstitial	Gain	PK2296B5	19	4875000	5750000	0.54521
Interstitial	Gain	ES2259D3	20	3350000	3450000	0.5618
Aneuploid	Gain	GR2309C5	20	0	4952708	0.49135
Telomeric	Gain	PK2299B4	20	0	1075000	0.66127
Telomeric	Gain	PK2299B6	20	2750000	4952708	0.6445
Interstitial	Loss	BP2119A1	21	1550000	1700000	-0.3448
Telomeric	Loss	ES2194B3	21	0	125000	-0.7291
Telomeric	Loss	ES2194B3	21	4475000	4935105	-0.7141
Telomeric	Loss	MC2191C1	21	0	1550000	-0.5441
Telomeric	Loss	MC2191C1	21	4475000	4935105	-0.5365
Telomeric	Gain	PK2299B6	22	0	1025000	0.67778
Telomeric	Gain	PK2299B6	22	4300000	4942902	0.3683
Interstitial	Loss	CC2255C6	23	1825000	2150000	-1.2396
Interstitial	Loss	CC2255D5	23	1825000	2150000	-1.2895747
Telomeric	Gain	CT2072B7	23	625000	4425941	0.48947
Telomeric	Gain	CT2072F6	23	625000	4425941	0.44339
Interstitial	Loss	PK2300D1	24	3975000	4225000	-1.0237
Interstitial	Gain	BP2119A1	25	850000	1025000	0.44162
Aneuploid	Gain	ES2194B3	25	0	4464133	0.42413
Telomeric	Gain	PK2299B6	25	0	425000	0.44237
Interstitial	Gain	PK2299B6	25	550000	1000000	0.4643
Interstitial	Gain	PK2299B6	25	1525000	1925000	0.47095
Interstitial	Gain	PK2299B4	26	275000	1300000	0.42341
Telomeric	Gain	PK2299B4	26	1775000	4038700	0.36832
Telomeric	Loss	CC2255A4	27	0	325000	-0.9928
Interstitial	Loss	CC2255A4	27	425000	600000	-0.764
Telomeric	Loss	CC2255A4	27	3725000	3838988	-1.0517
Telomeric	Loss	CT2072B6	27	3650000	3838988	-0.9483
Telomeric	Loss	CT2072B8	27	0	325000	-0.9976
Interstitial	Loss	CT2072B8	27	400000	675000	-0.9207
Telomeric	Loss	CT2072B8	27	3575000	3838988	-0.9321
Telomeric	Loss	ES2194B3	27	0	325000	-0.8833
Interstitial	Loss	ES2194B3	27	425000	675000	-0.7038
Telomeric	Loss	ES2194B3	27	3550000	3838988	-0.828
Telomeric	Loss	ES2194C3	27	0	325000	-0.8964
Interstitial	Loss	ES2194C3	27	425000	675000	-0.7279

Table B.7: continued

Anomaly Type	DNA Change	Strain	ODER1 Scaffold	Anomaly Start (mbp)	Anomaly Stop (mbp)	Log2 Coverage
Telomeric	Loss	ES2194C3	27	3550000	3838988	-0.8558
Telomeric	Loss	GR2310A4	27	0	200000	-1.0282
Telomeric	Loss	GR2310A4	27	3625000	3838988	-1.0219
Interstitial	Loss	ES2194B3	30	2875000	3050000	-1.0436
Interstitial	Loss	ES2194C3	30	2875000	3050000	-1.0324
Interstitial	Loss	ES2259A3	31	3525000	3675000	-0.3145
Interstitial	Loss	PK2298C3	31	1325000	1825000	-0.3797
Interstitial	Gain	PK2299B6	31	1075000	1225000	0.91474
Interstitial	Gain	PK2299B6	31	2600000	3200000	0.7802
Interstitial	Gain	PK2299B4	32	400000	2675000	0.48066
Telomeric	Gain	PK2299B6	32	0	150000	0.48069
Interstitial	Gain	PK2299B6	32	675000	1275000	0.59278
Interstitial	Gain	PK2299B6	32	1525000	2375000	0.65952
Interstitial	Gain	ES2194B3	33	300000	550000	0.52442
Interstitial	Gain	ES2194C3	33	450000	550000	0.50946
Interstitial	Loss	PK2298C3	33	500000	725000	-0.474
Interstitial	Gain	PK2299B4	33	2150000	2650000	0.41593
Interstitial	Loss	BP2119A1	34	200000	550000	-0.8307
Interstitial	Loss	GR2308D3	34	200000	550000	-0.7865
Interstitial	Gain	PK2299B6	34	725000	1000000	0.70999

Table B.8: Read statistics

Strain ID	Number of Reads*	Number of aligned reads	Number of aligned reads post-filtering**	Alignment Rate (pre-filtering, %)
CCMP3037	101,414,764	97,950,156	77,144,676	96.58%
UTEX3227	170,066,458	162,923,515	129,478,310	95.80%
CC2192B5	138,509,440	130,194,793	102,180,479	94.00%
UTEX3226	221,704,524	207,209,051	141,542,642	93.46%
CC2192C1	168,980,652	157,311,353	122,615,518	93.09%
PK2298B4	143,241,778	133,269,386	105,248,188	93.04%
CC2192D3	147,208,742	136,952,485	106,689,819	93.03%
PK2297D4	114,717,004	105,662,159	73,872,624	92.11%
PK2298C6	122,450,070	112,077,339	79,144,360	91.53%
CCMP2941	104,400,956	95,339,209	63,331,266	91.32%
RCC3703	106,356,850	97,057,705	58,879,206	91.26%
8020	165,200,684	150,167,143	100,419,147	90.90%
PK2300D3	124,392,604	112,907,847	78,221,794	90.77%
PK2299B3	100,509,876	91,132,378	62,716,696	90.67%
CC2192D2	145,448,058	131,782,306	101,410,843	90.60%
UTEX2797-SC2	165,255,342	149,714,639	104,741,734	90.60%
13A5	228,563,486	206,674,591	139,567,532	90.42%
UTEX2797-SC1	115,143,414	103,994,242	71,029,481	90.32%
UTEX3226-SC1	63,662,436	57,460,459	39,110,900	90.26%
UTEX2797-SC3	122,110,388	110,158,739	75,904,867	90.21%
UTEX3226-SC3	78,977,680	71,206,489	49,272,700	90.16%
UTEX3226-SC2	76,932,478	69,360,491	46,879,202	90.16%
NIES1812	110,040,168	99,097,055	67,050,441	90.06%
ARC81	99,457,688	89,503,320	63,288,259	89.99%
ARC66	114,176,190	102,701,880	73,505,272	89.95%
ARC84	117,665,638	105,641,404	75,460,418	89.78%
ARC83	113,474,842	101,821,061	72,752,813	89.73%
8027	75,879,666	68,080,044	48,191,884	89.72%
UTEX2827	96,988,992	86,868,200	57,778,218	89.57%
CC2192B4	118,215,510	105,818,240	82,885,496	89.51%
UTEX2797	177,699,208	158,109,763	102,287,381	88.98%
UTEX2797-SC4	72,405,100	62,489,365	42,371,869	86.31%
ARC82	95,110,382	74,336,712	53,890,828	78.16%

*Number of input reads, post BlobTools

**Filtering: Reads mapped in proper pairs with a minimum quality of 30

Table B.9 A-type mass spectrometry.

Mass spectrometry data for prymnesins identified in recent A-type Texas strains.

Metabolite name	PK2299B3	CC2192D2	CC2192B5	PK2298C6	PK2298B4	PK2300D3	PK2297D4
PRM-A (3 Cl)	37,352	84,153	81,351	34,654	71,568	56,859	113,391
PRM-A (2 Cl + DB) + pentose	8,478	32,177	18,091	4,722	15,275	1,840	5,531
prymnesin-2	13,598	71,195	87,130	10,195	25,108	4,622	42,405
PRM-A (3 Cl) + pentose + hexose	2,414	14,955	20,223	2,050	5,418	1,322	8,423
prymnesin-1	1,537	7,851	9,356	768	1,869	593	4,177
PRM-B (1 Cl + DB)	202	227	350	163	78	89	107
PRM-B (1 Cl)	133	270	504	211	333	431	167
PRM-B (1 Cl + DB) + pentose	104	434	357	386	265	205	206
prymnesin-B2	224	3,213	1,775	503	2,110	627	634
prymnesin-B1	114	985	742	467	572	631	151
PRM-B (1 Cl) + 2 hexose	142	290	329	229	278	244	226

Table B.10 B-type mass spectrometry

Mass spectrometry data for prymnesins identified in recent B-type Texas strains.

Metabolite name	ES2194A6	CT2072B7	ES2259D3	PK2299B4	PK2298C3	PK2300D1	CC2255D5
PRM-A (3 Cl)	561	267	380	435	352	163	626
PRM-A (2 Cl + DB) + pentose	916	134	664	530	212	319	544
prymnesin-2	1632	172	647	812	419	289	468
PRM-A (3 Cl) + pentose + hexose	369	323	3067	572	319	507	810
prymnesin-1	619	242	296	297	234	122	745
PRM-B (1 Cl + DB)	9567	8351	30233	22167	4337	20069	2298
PRM-B (1 Cl)	108414	57812	145788	137513	105067	206988	50943
PRM-B (1 Cl + DB) + pentose	35282	25832	113879	84281	13816	39835	7937
prymnesin-B2	152842	30558	220860	177471	135509	194503	105522
prymnesin-B1	178796	43332	214255	121550	94188	216739	68564
PRM-B (1 Cl) + 2 hexose	92221	11712	81870	21268	15808	29186	33336

Table B.11: Non-exhaustive list of recognized microbial cryptic species complexes

Species	Lineage	Source
<i>Candida albicans</i>	Ascomycete	(Dudiuk et al. 2017)
<i>Alexandrium tamarense</i>	Dinoflagellate	(Lilly et al. 2007)
<i>Prorocentrum fukuyoi</i>	Dinoflagellate	(Wu et al. 2022)
<i>Akashiwo sanguinea</i>	Dinoflagellate	(Luo et al. 2017a)
<i>Leucocytozoon majoris</i>	Apicomplexan	(Galen, Nunes, et al. 2018)
<i>Pseudochattonella farcimen</i>	Dictyochophycean	(Riisberg and Edvardsen 2008)
<i>Prymnesium parvum</i>	Haptophyte	(Wisecaver et al. 2023)
<i>Pseudo-Nitzschia delicatissima</i>	Diatom	(Giulietti et al. 2021)
<i>Pseudo-Nitzschia pseudodelicatissima</i>	Diatom	(Ruggiero et al. 2015)
<i>Chaetoceros socialis</i>	Diatom	(Degerlund et al. 2012)
<i>Chaetoceros tenuissimus</i>	Diatom	(Grzebyk et al. 2022)
<i>Gephyrocapsa huxleyi</i>	Haptophyte	(Read et al. 2013; Bendif et al. 2014)
<i>Chattonella subsalsa</i>	Raphidophyte	(Klöpfer et al. 2013)
<i>Plasmopara viticola</i>	Oomycete	(Rouxel et al. 2013)
<i>Fusarium graminearum</i>	Ascomycete	(O’Donnell et al. 2004)
<i>Fusarium fujikuroi</i>	Ascomycete	(Proctor et al. 2004)
<i>Chlorokybus atmophyticus</i>	Streptophyte	(Irisarri et al. 2021)
<i>Cryptomonas erosa</i>	Cryptophyte	(Hoef-Emden 2007)
<i>Xiphophagua minuta</i>	Rhizarian	(Goldstein et al. 2010)
<i>Polysiphonia scopulorum</i>	Rhodophyte	(Díaz-Tapia et al. 2020)
<i>Porolithon onkodes</i>	Rhodophyte	(Gabrielson et al. 2018)
<i>Hyalosphenia elegans</i>	Amoebozoan	(Oliverio et al. 2014)
<i>Collodictyon tricilatum</i>	Diphylleid	(Orr et al. 2018)
<i>Caecitellus parvulus</i>	Bigyra	(Hausmann et al. 2006)
<i>Naegleria galeacystis</i>	Discobid	(Zurita-Artaloitia et al. 2023)
<i>Naegleria americana</i>	Discobid	(Zurita-Artaloitia et al. 2023)
<i>Vermamoeba vermiformis</i>	Discobid	(Zurita-Artaloitia et al. 2023)
<i>Giardia duodenalis</i>	Metamonad	(Monis et al. 2003)
<i>Codosiga botrytis</i>	Opisthokont	(Stoupin et al. 2012)
<i>Tetrahymena pyriformis</i>	Ciliate	(Simon et al. 2008)
<i>Paramecium aurelia</i>	Ciliate	(Sonneborn 1975)
<i>Dinenympha porteri</i>	Metamonad	(Igai et al. 2022)
<i>Brachionus plicatilis</i>	Rotifera	(Gómez et al. 2002)
<i>Spirostomum minus</i>	Ciliate	(Shazib et al. 2025)
<i>Galietelia vivans</i>	Foraminiferan	(Morard et al. 2024)
<i>Chaetoceros curvisetus</i>	Diatom	(De Luca et al. 2021)
<i>Aspergillus fumigatus</i>	Opisthokont	(Alastruey-Izquierdo et al. 2014)
<i>Micromonas polaris</i>	Chlorophyte	(Simon et al. 2017)
<i>Micromonas pusila.</i>	Chlorophyte	(Slapeta et al. 2006)
<i>Klebsormidium nitens</i>	Chlorophyte	(Škaloud and Rindi 2013)
<i>Amoebophrya ceratii</i>	Dinoflagellate	(Cai et al. 2020)
<i>Nebela tinctoria</i>	Amoebozoan	(Gilbert et al. 2003)
<i>Euchlanis dilatata</i>	Rotifera	(Kordbacheh et al. 2017)
<i>Pinnularia acidicola</i>	Diatom	(Kollár et al. 2026 Mar 18)
<i>Globigerina bulloides</i>	Foraminiferan	(Stewart et al. 2001)
<i>Turborotalita quinqueloba</i>	Foraminiferan	(Stewart et al. 2001)
<i>Eimeria lata</i>	Apicomplexan	(Blake et al. 2021)
<i>Eimeria nagambie</i>	Apicomplexan	(Blake et al. 2021)
<i>Eimeria zaria</i>	Apicomplexan	(Blake et al. 2021)
<i>Closterium ehrenbergii</i>	Chlorophyte	(Ichimura 1981)
<i>Scenedesmus deserticola</i>	Chlorophyte	(Zou et al. 2016)

Table B.12: Cell counts of cultures prior to chemotyping

Concentrations of isolates upon harvesting for chemotyping in this study, each sample represents 200 mL of harvested culture. Cell concentrations were determined via manual cell counting using a phase counting chamber hemocytometer.

ID	Cells/mL
CC2192B5-1	990000
CC2192B5-2	835000
CC2192B5-3	1940000
CC2192D2-1	1105000
CC2192D2-2	1400000
CC2192D2-3	965000
CC2255D5-1	430000
CC2255D5-2	510000
CC2255D5-3	515000
CT2072B7-1	1015000
CT2072B7-2	1175000
CT2072B7-3	1050000
ES2194A6-1	920000
ES2194A6-2	600000
ES2194A6-3	455000
ES2259D3-1	500000
ES2259D3-2	1445000
ES2259D3-3	1785000
PK2297D4-1	315000
PK2297D4-2	295000
PK2297D4-3	450000
PK2298B4-1	1070000
PK2298B4-2	285000
PK2298B4-3	800000
PK2298C3-1	585000
PK2298C3-2	540000
PK2298C3-3	475000
PK2298C6-1	650000
PK2298C6-2	965000
PK2298C6-3	985000
PK2299B3-1	205000
PK2299B3-2	385000
PK2299B3-3	430000
PK2299B4-1	745000
PK2299B4-2	1330000
PK2299B4-3	1075000
PK2300D1-1	350000
PK2300D1-2	510000
PK2300D1-3	840000
PK2300D3-1	700000
PK2300D3-2	900000
PK2300D3-3	700000

Table B.13: B-type Heterozygosity

Heterozygosity estimates of B-type strains from recent Texas blooms as estimated by K-mer counting via KMC and modeling via GenomeScope2.

ID	Heterozygosity (%)	KMC Ploidy
BP2119A1	0.5047	2
CC2255A4	0.6939	2
CC2255C5	0.4497	2
CC2255C6	0.5144	2
CC2255D5	0.4894	2
CR2175C4	0.659	2
CT2072B4	0.4345	2
CT2072B6	0.4136	2
CT2072B7	0.3849	2
CT2072B8	0.9487	2
CT2072D8	0.4883	2
CT2072F6	0.6062	2
CT2072G1	0.5162	2
ES2194A4	0.4505	2
ES2194A5	0.5174	2
ES2194A6	0.5489	2
ES2194B2	0.5602	2
ES2194B3	0.5116	2
ES2194B5	0.5046	2
ES2194C2	0.4544	2
ES2194C3	0.5701	2
ES2194C4	0.7852	2
ES2194D1	0.508	2
ES2194D2	0.4815	2
ES2259A3	0.6079	2
ES2259D3	0.5166	2
ES2259D4	0.4897	2
GR2306A4	6.9254	2
GR2308D3	0.6253	2
GR2309C5	0.5647	2
GR2310A4	0.521	2
K0081	1.20	3
MC2191C1	0.464	2
PK2294D2	0.5555	2
PK2296B5	0.3834	2
PK2298C3	0.5123	2
PK2299B4	0.6006	2
PK2299B6	0.4538	2
PK2300D1	0.5065	2

Table B.14: Total B-type chromosomal anomalies

Strain	Total
ES2194B3	18
PK2299B6	16
ES2194C3	9
BP2119A1	9
PK2299B4	7
CT2072B8	6
GR2310A4	5
PK2296B5	5
PK2298C3	5
MC2191C1	5
GR2309C5	4
GR2308D3	4
CT2072B7	3
CT2072F6	3
CC2255A4	3
PK2300D1	3
ES2194A5	3
ES2194D1	2
CC2255C6	2
CC2255D5	2
CR2175C4	2
CT2072B6	2
ES2259A3	2
GR2306A4	2
CC2255C5	1
CT2072B4	1
CT2072D8	1
ES2194D2	1
ES2259D3	1
PK2294D2	1
CT2072C4	0
CT2072G1	0
ES2194A4	0
ES2194A6	0
ES2194B2	0
ES2194B5	0
ES2194C2	0
ES2194C4	0
ES2259B4	0
ES2259D4	0
PK2299D2	0

PUBLICATIONS

- [1] **Watervoort NF**, Jeje T, Dilkes BP, Wisecaver JH. 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum* sensu lato species complex. *In Review*. Preprint doi:[10.64898/2026.06.03.730016](https://doi.org/10.64898/2026.06.03.730016)
- [2] Wisecaver J, Jeje T, **Watervoort NF**. 2026. Letter to the Editor regarding “Long-read genome sequencing provides novel insights into the harmful algal bloom species *Prymnesium parvum*” by Jian et al. (2024). *bioRxiv* doi: [10.64898/2026.01.21.699772](https://doi.org/10.64898/2026.01.21.699772)
- [3] Fallon TR, Shende VV, Wierzbicki IH, Pendleton AL, **Watervoort NF**, Auber RP, Gonzalez DJ, Wisecaver JH, Moore BS. 2024. Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins. *Science*. 385(6709):671–678. doi:[10.1126/science.ad03290](https://doi.org/10.1126/science.ad03290).
- [4] Wisecaver JH, Auber RP, Pendleton AL, **Watervoort NF**, Fallon TR, Riedling OL, Manning SR, Moore BS, Driscoll WW. 2023. Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. *Current Biology*. 33(11):2246-2259.e8. doi:[10.1016/j.cub.2023.05.003](https://doi.org/10.1016/j.cub.2023.05.003)

VITA

Education and Training

2022-Present Ph.D. Candidate Biochemistry Purdue University
GPA: 3.83 Anticipated Graduation: December 2026
2017-2021 B.S. With Distinction Biochemistry, Bioinformatics Minor Purdue University
GPA: 3.91

Research and Professional Experience

2025-Present Research Technologist, Washington State University Pullman, Advisor: Jennifer Wisecaver
2022-Present Graduate Research Assistant, Purdue University West Lafayette, Advisor: Jennifer Wisecaver
2021-2022 NSF REPS Postbaccalaureate Fellow, Purdue University West Lafayette, Advisor: Jennifer Wisecaver
2019-2021 Undergraduate Researcher, Purdue University West Lafayette, Advisor: Jody Banks and Yun Zhou
2018 MASI Undergraduate Research Fellow, Purdue University West Lafayette, Advisor: Jody Banks
2018-2020 Undergraduate Researcher, Purdue University West Lafayette, Advisor: Jody Banks

Publications

1. **Watervoort NF**, Jeje T, Dilkes BP, Wisecaver JH. 2026. Hybridization between toxic bloom-forming algae of the *Prymnesium parvum* sensu lato species complex. *In Review*. Preprint doi:[10.64898/2026.06.03.730016](https://doi.org/10.64898/2026.06.03.730016)
2. Wisecaver J, Jeje T, **Watervoort NF**. 2026. Letter to the Editor regarding “Long-read genome sequencing provides novel insights into the harmful algal bloom species *Prymnesium parvum*” by Jian et al. (2024). *bioRxiv* doi: [10.64898/2026.01.21.699772](https://doi.org/10.64898/2026.01.21.699772)
3. Fallon TR, Shende VV, Wierzbicki IH, Pendleton AL, **Watervoort NF**, Auber RP, Gonzalez DJ, Wisecaver JH, Moore BS. 2024. Giant polyketide synthase enzymes in the

biosynthesis of giant marine polyether toxins. *Science*. 385(6709):671–678. doi:[10.1126/science.ado3290](https://doi.org/10.1126/science.ado3290).

4. Wisecaver JH, Auber RP, Pendleton AL, **Watervoort NF**, Fallon TR, Riedling OL, Manning SR, Moore BS, Driscoll WW. 2023. Extreme genome diversity and cryptic speciation in a harmful algal-bloom-forming eukaryote. *Current Biology*. 33(11):2246-2259.e8. doi:[10.1016/j.cub.2023.05.003](https://doi.org/10.1016/j.cub.2023.05.003)
5. **Watervoort NF** & Jeje T, Wisecaver JH. 2026. Harmful algal blooms of *Prymnesium parvum* in Texas consist of multiple cryptic species. *In Prep*.
6. **Watervoort NF**, Wisecaver JH. 2026. Don't judge a book by its cover: cryptic diversity should be an expectation in microbial eukaryotes. *In Prep*.

Teaching Experience

- | | |
|------------|--|
| 2025, 2026 | BIOL/MBIOS301 “General Genetics” Graduate TA, Washington State University (Fall Semester) |
| 2025 | BCHM307 “Biochemistry” Guest Lecturer, “Chemical Diversity Across the Tree of Life”, Purdue University (Spring Semester) |
| 2024 | BCHM307 “Biochemistry” Graduate TA, Purdue University (Fall Semester) |

Awards, Honors, and Indicators

- | | |
|------------------|--|
| 2026 | Society for Freshwater Science Best Oral Presentation of Applied Science |
| 2024 | Purdue Biochemistry Bird Stair Research Fellowship |
| 2024 | Purdue Center for Plant Biology Poster Competition First Place |
| 2023 | Purdue Office of Interdisciplinary Graduate Programs Travel Award |
| 2023 | Phycological Society of America Hoshaw Travel Award |
| 2022 | Purdue PULSe Lynn First Year Graduate Student Fellowship |
| 2021 | NSF Research Experience for Postbaccalaureate Students Fellowship |
| 2018, 2019, 2020 | Edwin T. Mertz Memorial Scholarship |
| 2020 | Elsie L. Gruel Memorial Scholarship |
| 2019 | Martin Agricultural Research Fund Scholarship |
| 2019 | Rex Hall Memorial Scholarship |
| 2018 | Robert J. Woods Scholarship |
| 2018 | Purdue Office of Undergraduate Research Scholarship |

Professional and Synergistic Activities

2026	Society for Freshwater Sciences Member
2026	American Society of Plant Biology Member
2025	Society for the Study of Evolution Member
2024	Purdue University AGR490 Gateways to Graduate School Graduate Student Panelist
2024	Purdue Undergraduate Research Conference Judge
2024-2025	Purdue Center for Plant Biology Trainee Association Social Chair
2024	Purdue Biochemistry Undergraduate Poster Exhibit Judge
2024	Purdue PULSe Spring Reception Poster Judge
2023-Present	Phycological Society of America Member
2023	International Society for Evolutionary Protistology Member
2023	Botanical Society of America Member
2022	Purdue PULSe Fall Open House Volunteer

Mentoring

2025	Washington State University SBS Peer Mentor
2024	Undergraduate Mentoring in the Wisecaver Lab (1 undergraduate)
2023	Purdue PULSe First Year Mentor
2023	Mentored an REU student through Purdue PURE-PD
2023	Mentored a Purdue University First Year Rotation Student (1 graduate student)
2023	Undergraduate Mentoring in the Wisecaver Lab (1 undergraduate)

Oral Presentations

1. Society for Freshwater Science, Spokane, WA, May 2026. **Watervoort, NF.** “Hybridization in the Toxic Bloom-Forming Algae *Prymnesium parvum*.”
2. School of Biological Sciences Symposium, Moscow, ID, Feb 2026. **Watervoort, NF.** “Hybridization and cryptic diversity in *Prymnesium parvum* s.l.”
3. Evolution, Athens, GA, June 2025. **Watervoort, NF.** “Hybridization in the Toxic Bloom-Forming Microalga *Prymnesium parvum*”
4. PULSe Student Seminar Series, West Lafayette, IN, March 2025. **Watervoort, NF.** “Heterozygosity and Genomic Diversity of Texas *Prymnesium parvum* (Haptophyta)”

5. Purdue Center for Plant Biology Trainee Association Half Baked Seminars, West Lafayette, IN, October 2024. **Watervoort, NF**. “Preheating The Oven: What to Do with Genomes from Multiple Cryptic Species?”
6. PSA-ISOP-ISEP Joint Meeting, Seattle, WA, July 2024. **Watervoort, NF**. “Genomic and Metabolomic Diversity of Recent *Prymnesium parvum* (Haptophyta) Blooms in Texas”
7. Purdue Biochemistry Graduate Postdoc Seminar, West Lafayette, IN, January 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
8. Phycological Society of America, Providence, RI, June 2023. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”

Poster Presentations

1. ASPB-CSPB Joint Meeting, Ottawa, ON, Canada, July 2026. **Watervoort, NF**, Jeje, T, Dilkes, BP, Wisecaver, JH. “Hybridization in the Toxic Bloom-Forming Algae *Prymnesium parvum*.”
2. School of Biological Sciences Symposium, Moscow, ID, Feb 2026. **Watervoort, NF**. “Hybridization and cryptic diversity in *Prymnesium parvum* s.l.”
3. ESE Symposium: Future of Waters, West Lafayette, IN, Oct 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
4. PULSe 20th Anniversary Symposium, West Lafayette, IN, Sep 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
5. Purdue Center for Plant Biology Symposium, West Lafayette, IN, Sep 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
6. Purdue Center for Plant Biology Retreat, West Lafayette, IN, May 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
7. American Society of Plant Biology Midwest Meeting, West Lafayette, IN, March 2024. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
8. Purdue Plant Science Poster Symposium, West Lafayette, IN, September 2023. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”

9. Purdue Department of Biochemistry Retreat, West Lafayette, IN, August 2023. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
10. Botany, Boise, ID, July 2023. **Watervoort, NF**, Wisecaver, JH. “Recent Harmful Algal Blooms of *Prymnesium parvum* in Texas Consist of Multiple Cryptic Species”
11. Center for Plant Biology Retreat, West Lafayette, IN, May 2023. **Watervoort, NF**, Wisecaver, JH. “Recent harmful algal blooms of *Prymnesium parvum* in Texas consist of multiple cryptic species”
12. PULSe Spring Reception, West Lafayette, IN, April 2023. **Watervoort, NF**, Wisecaver, JH. “Recent harmful algal blooms of *Prymnesium parvum* in Texas consist of multiple cryptic species”
13. International Society for Evolutionary Protistology, Online, January 2023. **Watervoort, NF**, Auber, RP, Pendleton, AL, Riedling, OL, Wisecaver, JH. “Genome Size Variation in the Cryptic Species Complex *Prymnesium parvum*”
14. Center for Plant Biology Retreat. West Lafayette, IN, May 2022. **Watervoort, NF**, Riedling, O, Auber, RP, Wisecaver, JH. “Extreme genome diversity of *Prymnesium parvum* in Texas”
15. Joint Aquatic Sciences Meeting. Grand Rapids, Michigan, May 2022. **Watervoort, NF**, Riedling, O, Auber, RP, Wisecaver, JH. “Extreme genome diversity of *Prymnesium parvum* in Texas”
16. Purdue Undergraduate Research Symposium, West Lafayette, IN, April 2019. **Watervoort, NF**, Banks, JA. “Determining the Structure of a Sex-Determining Pheromone”
17. Purdue Plant Science Poster Symposium, West Lafayette, IN, September 2018. **Watervoort, NF**, Banks, JA. “Optimizing the Purification of a Sex-Determining Pheromone”
18. Corteva Agriscience, Indianapolis, IN, July 2018. **Watervoort, NF**, Banks, JA. “Optimizing the Purification of a Sex-Determining Pheromone”