

Long-term stability of the genome structure of the cyanobacterium, *Dolichospermum* in a deep German lake

Author

Woodhouse, JN, Burford, MA, Neilan, BA, Jex, A, Tichkule, S, Sivonen, K, Fewer, DP, Grossart, H-P, Willis, A

Published

2024

Journal Title

Harmful Algae

Version

Version of Record (VoR)

DOI

[10.1016/j.hal.2024.102600](https://doi.org/10.1016/j.hal.2024.102600)

Rights statement

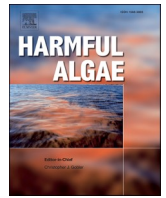
Crown Copyright © 2024 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Downloaded from

<https://hdl.handle.net/10072/438567>

Griffith Research Online

<https://research-repository.griffith.edu.au>



Original Article

Long-term stability of the genome structure of the cyanobacterium, *Dolichospermum* in a deep German lake

J.N. Woodhouse^a, M.A. Burford^b, B.A. Neilan^c, A. Jex^{d,e}, S. Tichkule^{d,f}, K. Sivonen^g, D.P. Fewer^g, H-P Grossart^{a,h}, A. Willis^{b,1,*}

^a Department of Plankton and Microbial Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), 16775 Stechlin, Germany

^b Australian Rivers Institute, and School of Environment and Science, Griffith University, Brisbane, Australia

^c School of Environmental and Life Sciences, The University of Newcastle, Callaghan 2308, NSW, Australia

^d Population Health and Immunity, Walter and Eltza Hall Institute of Medical Research, Parkville, VIC, Australia

^e Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC, Australia

^f Department of Medical Biology, University of Melbourne, Melbourne, Australia

^g Department of Microbiology, University of Helsinki, Viikinkaari 9, FI-00014 Helsinki, Finland

^h Department of Biochemistry and Biology, Potsdam University, 14469 Potsdam, Germany

ARTICLE INFO

Edited by Dr. C. Gobler

Keywords:

ADA cluster

Akinetes

Anabaena

Dolichospermum

Phylogenomics

Sediment core

Single-cell genomes

ABSTRACT

Dolichospermum is a cyanobacterial genus commonly associated with toxic blooms in lakes and brackish water bodies worldwide, and is a long-term resident of Lake Stechlin, northeastern Germany. In recent decades, shifts in the phosphorus loading and phytoplankton species composition have seen increased biomass of *Dolichospermum* during summer blooms from 1998, peaking around 2005, and declining after 2020. Cyanobacteria are known to rapidly adapt to new environments, facilitated by genome adaptation. To investigate the changes in genomic features that may have occurred in Lake Stechlin *Dolichospermum* during this time of increased phosphorus loading and higher biomass, whole genome sequence analysis was performed on samples of ten akinetes isolated from ten, 1 cm segments of a sediment core, representing a ~45-year period from 1970 to 2017. Comparison of these genomes with genomes of extant isolates revealed a clade of *Dolichospermum* that clustered with the ADA-6 genus complex, with remarkable genome stability, without gene gain or loss events in response to recent environmental changes. The genome characteristics indicate that this species is suited to a deep-chlorophyll maximum, including additional light-harvesting and phosphorus scavenging genes. Population SNP analysis revealed two sub-populations that shifted in dominance as the lake transitioned between oligotrophic and eutrophic conditions. Overall, the results show little change within the population, despite diversity between extant populations from different geographic locations and the in-lake changes in phosphorus concentrations.

1. Introduction

Cyanobacterial harmful algal blooms (cyanoHABs) form when there is a rapid proliferation of cyanobacterial biomass in the upper surface waters. Numerous factors have been proposed to explain the dominance of cyanobacteria in freshwaters, including accessibility of different organic and inorganic nutrients, resistance from grazing, and buoyancy regulation to control their position in the water column (for a review see Burford et al., 2016). In addition, cyanobacteria can adapt rapidly to local conditions through intraspecific variation in a population, genetic

drift, or the acquisition of new genetic material from local environs (via viruses and horizontal gene transfer). These genome changes can occur within two years (Willis et al., 2020). When sustained over long periods, these adaptive mechanisms can reinforce stochastically acquired mutations and lead to speciation (for reviews see, Willis and Woodhouse, 2020; Dvořák et al., 2023). *Dolichospermum* (basonym *Anabaena*) is a genus of globally occurring harmful bloom-forming diazotrophic cyanobacterium, with strains capable of producing toxins belonging to five distinct classes, microcystins, saxitoxins, cylindrospermopsins, anatoxins and guanitoxins (D'Agostino et al., 2016; Li et al., 2016; Lima

* Corresponding author.

E-mail address: anusuya.willis@csiro.au (A. Willis).

¹ currently at: Australian National Algae Culture Collection, CSIRO, Hobart, Australia

et al., 2022). There is a global distribution and existence of toxic and non-toxic strains of *Dolichospermum*, with high diversity suggesting localized adaptation (Steffen et al., 2012). However, how strains or populations adapt over time, or change in response to local conditions, has not been explored. Understanding the process of local adaptation is essential for understanding the potential invasion and ultimate success of cyanobacteria.

Dolichospermum is commonly found in lakes and reservoirs across Germany and has been reported in Lake Stechlin over the past 40 years, with the earliest accounts in 1982 and 1983 (Casper et al., 1985, Padisak et al., 2003a, Padisak et al., 1998). Lake Stechlin has, for most of the last 50 years, been an oligotrophic lake. However, in the past 10-20 years, phosphorus levels have rapidly increased within the lake (Selmezy et al., 2019). Increases in bioavailable phosphate have coincided with drastic change in phytoplankton and zooplankton biomass (Selmezy et al., 2019). Specifically, increases in cyanobacterial biomass, particularly the large filamentous form of *Dolichospermum*, and in a decrease in otherwise abundant large cladocerans, have potential impacts on nutrient cycling and ecosystem health.

Reports during the 1990s of *Dolichospermum* in Lake Stechlin suggested that filaments matched the taxonomic description of *Dolichospermum* (ex. *Anabaena*) *lemmermannii* and several *Dolichospermum*-like filaments lacking heterocysts (Padisak et al., 2003a). Between 1994 and 2008, the biomass of *Dolichospermum* in the lake (averaged over the upper 25 m of the water column) was considered an 'occasional dominant' species (Padisak et al., 2010). Surface blooms of *Dolichospermum* formed infrequently and were short-lived, contributing <5% of the annual phytoplankton biomass, despite their occasional dominance. This biomass might be understated, as the precise vertical distribution of *Dolichospermum* over this time is unclear. However, there is a deep chlorophyll maximum at 7-10 m, historically dominated by *Planktothrix* and *Cyanobium* (Padisak et al., 2003b). The summers of 1998 and 1999 represent a sudden shift in this pattern with *Dolichospermum* biomass exceeding 200 µg (fresh weight) L⁻¹ (Padisak et al., 2003a). Since 2005, increasing densities of *Dolichospermum* and *Aphanizomenon* have been reported, both within the upper mixed layer and forming a distinct deeper chlorophyll maximum (Selmezy et al., 2015, Giling et al., 2017, Selmezy et al., 2019). An exceptional bloom of *Dolichospermum* was reported in 2011, following storm Otto, which appeared to be due to the wind-induced thermocline deepening (Kasprzak et al., 2017); a response confirmed experimentally (Giling et al., 2017).

Strains of cyanobacteria species have been found to differ in their optimal environmental conditions. For example, in *Microcystis*, a non-diazotrophic cyanoHAB genus, high-phosphorus adapted genotypes were shown to cluster into distinct clades, occurring across different eutrophic lakes, and low phosphorus-adapted genotypes were found in both eutrophic and oligotrophic lakes and co-occurring with high-phosphorus adapted genotypes (Jackrel et al., 2019). Similarly, changes in the ratio of toxic and non-toxic genotypes in a *Raphidiopsis* population occurred in response to phosphorus (Burford et al., 2014). It is unknown if similar patterns might be observed for *Dolichospermum* genotypes or species. Genome comparisons of *Dolichospermum* species from a variety of sources have revealed high genomic diversity. *Dolichospermum* species are polyphyletic, with strains found within the clades of the closely related genera *Aphanizomenon* (A), *Dolichospermum* (D) and *Anabaena* (A), termed the 'ADA cluster' (Driscoll et al., 2018). Phylogenomic analysis currently supports the occurrence of at least 10 distinct species within the ADA cluster (Osterholm et al., 2020; Dreher et al., 2021a). Despite gross morphological characteristics being similar, species within the ADA cluster exhibit a high level of morphological and genomic diversity even within a single ADA genus complex (Cirés and Ballot, 2016, Driscoll et al., 2018, Osterholm et al., 2020, Dreher et al., 2021a, Dreher et al., 2021b). There have not yet been any genomic studies suggesting co-occurring genotypes of *Dolichospermum* species, although different ADA species are known to co-occur (Shan et al., 2019).

ADA species produce akinetes, specialized cells that function as resting spores, allowing regeneration of populations under favorable conditions (Kaplan-Levy, et al., 2010). Sedimentation rates in lakes can bury akinetes, and particularly the deep center can capture dominant communities (Weisbrod et al., 2020). Akinetes are resistant to degradation and can persist in the sediment for upwards of 100 years, with akinetes isolated from sediment cores having been revived (Wood et al., 2009; Legrand et al., 2016). Thus, akinetes in a sediment core can then represent a genetic history of a population (Hobbs et al., 2021; Wood et al., 2009). Additionally, akinetes often have diacritic features used for morphological identification of ADA species, however, typical *Aphanizomenon* akinetes have not been found in the Lake Stechlin sediment (Ramm et al. 2012). The previous identification of *Aphanizomenon* and *D. circinale* in Lake Stechlin, might represent morphological changes of the population, misidentification, or a shift in the population between species or strains.

Here, we sought to understand the genetic history of *Dolichospermum* in Lake Stechlin, through whole genome sequencing of akinetes obtained from a sediment core. Specifically, we sought to i) to evaluate the extent to which genomes of *Dolichospermum* have changed in response to increasing phosphorous concentrations, ii) to identify genes that promote the dominance of *Dolichospermum* within Lake Stechlin, and iii) understand the phylogenomic placement of Lake Stechlin *Dolichospermum* in a global context.

2. Methods

2.1. Sediment sampling

A sediment core (D = 6 cm, L = 36 cm) was collected from the deepest point of Lake Stechlin (53°9'13" N, 13°1'49"E) at 69 m depth, on 6 June 2017, using an UwiTech corer (<http://www.uwitec.at/html/corer.html>). The core was immediately returned to the lab and sliced into 1 cm segments (volume = 28.26 cm³). Each segment was scraped into a 50 ml Falcon tube and stored at 4°C until used. Sedimentation rates within Lake Stechlin have been previously determined via sedimentation traps and using Cesium¹³⁷ (Fuchs et al., 2015, Wurzbacher et al., 2017). These measurements estimate a sedimentation rate of around 5 mm per year for the last ~10 years, compared to 2 - 3 mm per year from ~1995 - ~2010, and ~1 mm per year prior to 1973. Based on these sedimentation rates, the 10 cm sediment core provides a history from around 1973 to 2017, or ~45 years (Fig. 1). However, these sedimentation rates restrict the resolution of timescales, with each 1 cm segment integrating from seven to two years of *Dolichospermum* akinete populations. Additionally, annual mixing during December can disturb the upper 1 cm of the sediment surface, although previous studies have found minimal sedimentary mixing occurs (Fuchs et al., 2015, Wurzbacher et al., 2017). This may impact the heterogeneity of akinetes throughout the sediment. With this in mind, downstream analysis focused on trends occurring through time, or within groups of sediment layers, rather than emphasizing differences occurring within a single layer.

2.2. Akinete enumeration in sediment

To determine the counts of intact akinetes and empty akinete cases, the top 10 cm (sliced in 1 cm portions) of the sediment core were resuspended in diH₂O to 50 mL volume; then a 1 mL sample was diluted 6-fold. Samples were counted using a Sedgewick-Rafter counter under an inverted microscope (Eclipse Ti2; Nikon) at 100x magnification. A total of 100 squares were counted for each sample. Counts were calculated as intact akinetes or akinete cases per cm³ sediment.

2.3. Akinete isolation

Visually intact akinetes were isolated via micro-pipette from the first

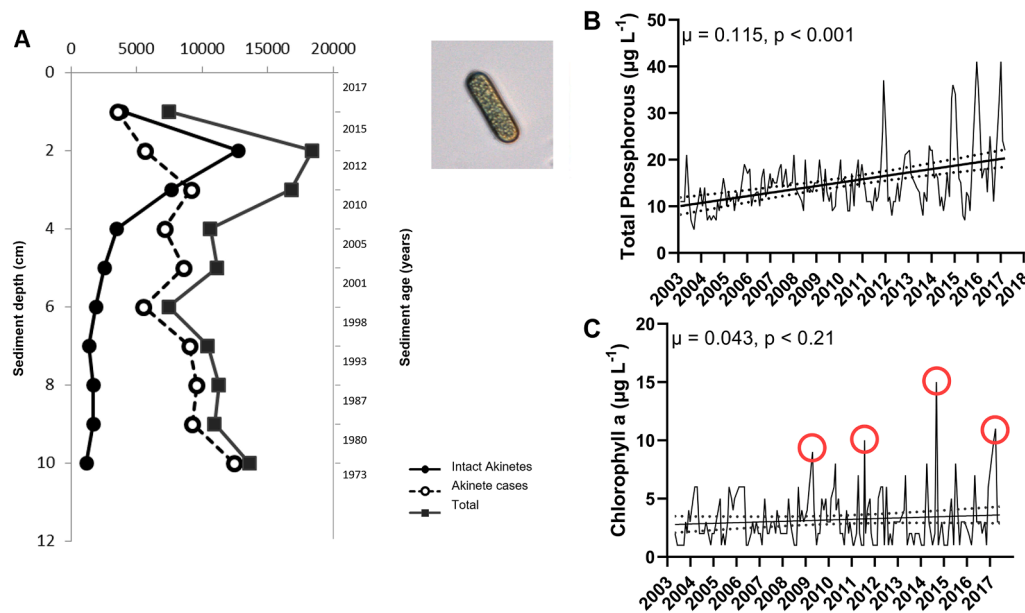


Fig. 1. A) Number of straight rod-shaped akinetes per cm^3 in the top ten sediment segments, each a 1 cm slice. Intact akinetes (solid circles), akinete cases (empty circles) and total akinetes (squares). Insert shows a light microscope image of a *Dolichospermum* sp. akinete, compound microscope (Eclipse Ti2; Nikon) at 200x magnification. B) Total phosphorous and C) total chlorophyll *a* concentrations measured from within the upper mixed layer from Lake Stechlin between from 2003-2017. Red circles highlight periods when visible accumulations of *Dolichospermum* were present.

ten 1 cm sediment segments. A small amount of sediment was diluted in diH_2O in a 6 cm plastic petri dish. Akinetes were isolated manually using a micro-pipette under an inverted microscope (Eclipse Ti2; Nikon), washed once in diH_2O and twice in PBS buffer. Ten akinetes from each of the top 10 segments (ST_sed01-10) were isolated for genome amplification, pooled in a PCR tube and stored frozen. In addition, two *Dolichospermum* filaments (ST_Chy and ST_Epi) were collected from an active bloom occurring in the water column of Lake Stechlin in the days following the sediment core sampling. Each of these filaments was processed in the same manner as the akinetes. Microscopic analysis of one filament (ST_Chy) showed an active infection by an unknown fungal-like parasite.

2.4. DNA isolation from akinetes

PCR tubes, each containing 10 intact akinetes, were centrifuged and the PBS buffer removed via a pipette. The remaining liquid was removed via speed-vac. The dried akinetes were resuspended in 3 μL lysis buffer, and a single silica-zirconium bead (1 mm diameter) was added. DNA from the 10 cells was isolated using an Illustra Single Cell GenomiPhy™ DNA amplification kit (Merck) with modifications to the manufacturer's protocol. Briefly, following the addition of lysis solution, tubes were frozen for 1 h at -70°C , and immediately thawed at 37°C for 10 min. After that, cells were vortexed for 30 s, followed by the addition of the lysis neutralization buffer. The remainder of the protocol was conducted as per the manufacturer's specifications. The presence of cyanobacteria gDNA was confirmed by PCR amplification using cyanobacteria 16S rRNA gene primers (CYA106F/CYA741R, Nübel et al., 1997).

2.5. Genomic DNA sequencing and assembly

Nextera DNA libraries were generated for each of the 12 samples and sequenced using a $2 \times 150\text{bp}$ Nextera Kit on the Illumina NextSeq instrument at the Genome Sequencing Facility, Walter and Eliza Hall Institute, Australia. Quality control of demultiplexed paired-end sequence libraries, including removing contaminating adapter sequences and low-quality base-pairs, was performed using bbdutk.sh (<https://sourceforge.net/projects/bbmap/>). Given the likelihood for

contaminating sequences either not fully removed during washing or introduced during genome amplification, we followed a metagenomic pipeline for genome assembly. Raw reads were assembled into contiguous sequences with MEGAHIT v1.2.8 (Li et al., 2015). DasTool (Sieber et al., 2018) was used to evaluate bins obtained using MaxBin2 (Wu et al., 2016), MetaBAT (Kang et al., 2019) and CONCOCT (Alneberg et al., 2014) binning algorithms. Genome completeness and contamination were evaluated using CheckM (Parks et al., 2015), genome statistics with Quast (Gurevich et al., 2013), and taxonomic classification using GTDB-TK (Chaumeil et al., 2020). A single *Dolichospermum* bin was obtained for each sample, and, except for ST_sed10, completeness was $> 90\%$. Annotation of *Dolichospermum* genomes was performed using Prokka (Seemann, 2014).

2.6. Comparative genome analysis

A phylogenomic analysis of both the akinete and single filament Lake Stechlin *Dolichospermum* genomes was conducted using GTDB-TK (Chaumeil et al., 2020) and the Anvi'o platform (Eren et al., 2015). Using the GTDB-TK package, we inferred a *de novo* phylogenomic tree containing genomes isolated from Lake Stechlin and publicly available genomes from the ADA (*Anabaena*, *Dolichospermum*, *Aphanizomenon*) cluster previously described (Dreher et al., 2021b) based on the bac120 single copy marker gene set. The same genomes were used to determine the ADA specific pangenome and core genome using Anvi'o. Genome-wide similarity between genomes was determined using the pyANI function as calculated across all single-copy core genes within the ADA cluster, and gANI (genome-wide Average Nucleotide Identity) was determined using FastANI (Jain et al., 2018). As previously reported, the existence of discrete species groups was confirmed using an gANI cut-off between 95-96%. Roary (Page et al., 2015) was used to define core, accessory and species-specific orthologous genes from Prokka generated gff genome files for each species, and the accessory and species-specific genes from each species clusters were manually compared to identify functional differences. For each species cluster, using results from CheckM, Quast, Roary, and Anvi'o, we calculated additional metrics including genome size, %GC, gene number and density, in addition to the %GC content for core, accessory and species-specific orthologous

gene (OG) families.

2.7. Population genome analysis

Population genomic analysis of the akinete genomes was performed using sample ST_sed9 as reference genome, initially error corrected with Pilon v1.23 (Walker et al., 2014), which represented the oldest sample where a near-complete genome was available. Raw reads of samples 1-8 were filtered using Trimmomatic v.0.36 (Bolger et al., 2014) and aligned to the ST_sed9 reference genome by using Burrows – Wheeler Aligner (BWA) v.0.7 (Li and Durbin, 2009) with default parameters. Single Nucleotide Polymorphisms (SNPs) were detected by using Genome Analysis Toolkit (GATK) v3.7.0 and performed default hard filtering according to GATK's best practices workflow to avoid false positives (Van der Auwera et al., 2013).

All identified SNPs (695) were used to perform population structure analyses. To assess the clustering among samples 1 through 9, we conducted Principal Component Analysis (PCA) employing the 'prcomp' function in the R. This analysis utilized all SNPs and was executed with the default parameters. An unrooted phylogenetic tree was generated based on the neighbour-joining algorithm using SNP occurrence between genomes in Molecular Evolutionary Genetics Analysis 7 (MEGA) software (Stecher et al., 2020) with default parameters. Furthermore, we utilized a quantitative clustering method to calculate the proportion of shared ancestry among the isolated genomes, a process executed using the STRUCTURE software v.2.3 (Pritchard et al., 2000). The STRUCTURE analysis was conducted with a burn-in period of 20,000 iterations followed by 100,000 Monte Carlo Markov Chain (MCMC) iterations using the admixture model. We performed 20 runs for each value of K, ranging from 1 to 10, where K is the number of genetic clusters in the population. The best value for K was estimated by using Clumpak (Kopelman, et al., 2015). Population genetic analyses were performed using R package PopGenome (Pfeifer et al., 2014) to calculate nucleotide diversity, fixation index (*F_{st}*) i.e. variance, genetic divergence (*D_{xy}*), and Tajima's D. Nonsynonymous and synonymous substitution ratio was calculated by using KaKs Calculator (Zhang et al., 2006), with the Nei and Gojobori method. The results were plotted with the R statistical programming language (version 3.6.1).

3. Results

3.1. Akinete isolation and enumeration

In the sediment core obtained from the deepest point of Lake Stechlin, akinetes were present as a straight (*D. circinalis*) rod shape (Fig. 1ab), consistent with the current dominance of *D. circinalis* in the epilimnic water of the lake (Selmeczy et al., 2019). We approximated time periods for each sediment segment using published values of sedimentation rates that used measurements of Cesium¹³⁷ (Fuchs et al., 2015; Wurzbacher et al., 2017). Lake Stechlin is very well characterised, and the sediment layering is known to be stable without annual mixing beyond the top centimetre (Fuchs et al., 2015; Wurzbacher et al., 2017).

The average number of all akinetes were calculated for each of the top ten sediment segments (1 cm depth each) (Fig. 1a). A peak in total (including intact and empty cases) akinetes was measured in segments ST_sed2 and ST_sed3 (2010-2015), below which the number of akinetes cm⁻³ decreased. This period coincides with an increase in total phosphorous concentrations in the epilimnion (Fig. 1b) and an increase in chlorophyll *a* concentration including a greater number of *Dolichospermum* dominated events (Fig. 1c). Akinete concentrations in the sediment appear stable throughout the period 1970-2010 with a minimum occurring between 1998-2001. Deeper sediment layers, integrate a greater number of years and therefore likely overrepresent the annual export of akinetes to the sediment.

3.2. Lake Stechlin akinete genomes

Genomes were assembled with > 98% completeness for ST_sed1 – ST_sed9, 55% for ST_sed10, 55% for ST_chy and 70% for ST_Epi (Table 1). Each genome represents a 'population' genome assembled from DNA from the ten akinetes isolated from each sediment segment. The genomes are available at NCBI Genbank with BioProject number PRJNA787667 and for accession numbers, see Table 1.

3.3. Comparative genomics of Lake Stechlin cyanobacterial akinetes

We performed a phylogenomic analysis of our Lake Stechlin akinete genomes, in the context of other publicly available ADA genomes using the bac120 single copy gene marker set implemented in GTDB-TK (Chaumeil et al., 2020). The Lake Stechlin genomes formed a monophyletic clade with the ADA-6 genus complex, including the isolate *Dolichospermum* sp. DET69 (Dreher et al., 2021b) (Fig. 2).

This clade of *Dolichospermum* akinetes and single filaments grouped with a > 98% ANI and formed a distinct species cluster (Supplementary Fig. 1), with between species ANI < 95%. Genomes derived from the single-filaments exhibited a greater dissimilarity to akinete derived genomes. However, this likely arose due to reductions in genome coverage of the single-filament genomes, a result of genome completeness of only ~70%, compared to the genome completeness of ~98-99% of the akinete genomes (see Table 1, Supplementary Fig. 1).

The genome characteristics of *Dolichospermum* derived from Lake Stechlin with *Dolichospermum* sp. DET69 forming ADA-6 were most similar to those of the ADA-3 lineage (Fig. 2, blue, Supplementary Fig. 1). The Lake Stechlin/ADA-6, and ADA-3 lineages possessed similarly sized genomes (~5.8 Mbp), similar numbers of OG families and a similar coding density. However, the %GC content, which can also be a marker of species differentiation, of the two lineages was different, with the %GC of the Stechlin/ADA-6 lineage estimated at ~37.5% and the ADA-3 lineage estimated at 38.5%. In this respect, the %GC of Lake Stechlin/ADA-6 were among the lowest within the ADA group, similar to the ADA-1 (Fig. 2, red) lineage, but not as low as that of the ADA-4 lineage, which had a %GC content of around 37% (Table 2).

3.4. Comparison of Lake Stechlin akinetes with the ADA cluster

To investigate the basis of the differences in %GC content, we compared genome sequences for the Stechlin genomes and the closest ADA genome clusters (ADA-2, ADA-3, ADA-4) (Table 2). We calculated the %GC content of the pan-genome, the shared core, shared accessory, and clade-specific OGs (Table 2, Supplementary Table 1). The %GC content of the shared core genome was similar for all groups (~39.55 ± 0.65%). We also found no discernible variation in %GC content of accessory OGs shared within the Stechlin and three closest ADA clusters. We identified a total of 816 OGs specific to the ADA-Stechlin lineage (with 45 OGs present across all 12 genomes), with an average %GC of 34.3%, therefore, the difference in genome GC% is a result of clade-specific OGs.

Comparison of the accessory genomes of ADA clusters identified functional genes that varied in copy number or presence/absence between clusters (Fig. 2, Supplementary Table 1). Genome assemblies using short-read sequencing can be prone to gene absences as a result of incomplete assemblies, or contamination via errors in metagenome binning, thus we report the follow average differences between genome clusters acknowledging this. The Lake Stechlin akinete genomes compared to other ADA genomes have an additional copy of *KaiB* and *KaiC* circadian clock regulator proteins, expansion of *Cph1*, *Cph2* and *BchB* light harvesting proteins, an additional copy of the *bacA*, vitamin-B12 uptake protein, and an additional copy of the *PhnD* (organic phosphate import protein). In the opposite direction, we observed an absence of Cas viral modifying enzymes, and *NrtA*, *NrtB* nitrate uptake proteins, and a reduction in the copies of *GypA*, *GypC* gas vesicle proteins.

Table 1

Statistics of akinete genome assemblies of Lake Stechlin, each genome includes DNA from ten akinetes, from each sediment segment (ST_sed1 – ST_sed10) and the extant sample (ST_Epi, calculated with QUASt Gurevich et al., 2013) and CheckM (Parks et al., 2015).

Akinete Genome Assembly (sample number)	No. contigs	Total length (bp)	GC (%) [*]	N50* (bp)	Completeness (%)	Contamination (%)	Strain heterogeneity [§]	Coding sequences (CDS)	Accession number (Genbank)
ST_sed1	647	6020366	37.56	15674	99.85	1.67	0	5276	JAKOGR000000000
ST_sed2	525	5748588	37.5	19684	99.44	0.44	0	4096	JAKOGS000000000
ST_sed3	782	6304092	37.6	16590	99.89	2	0	4942	JAKOGT000000000
ST_sed4	706	5873098	37.5	18889	98.78	1.22	22.22	5271	JAKOGU000000000
ST_sed5	1020	6622813	37.51	10561	97.26	13.78	97.1	5259	JAKOGV000000000
ST_sed6	611	5942300	37.62	16863	99.56	3.33	90.91	5163	JAKOGW000000000
ST_sed7	633	5767940	37.58	15016	98.5	0.3	50	4999	JAKOGX000000000
ST_sed8	532	5781196	37.57	18532	99.33	0.44	50	5195	JAKOGY000000000
ST_sed9	1078	7103190	37.27	14776	99.22	5.56	83.33	5264	JAKOGZ000000000
ST_sed10	779	4218483	36.82	8838	55.33	0	0	5313	JAKOHA000000000
ST_Epi	1086	4448791	36.51	6570	70.06	1.49	0	4867	JAKOHB000000000

^{*}QUAST: statistics are based on contigs of size ≥ 500 bp.

[§] Strain heterogeneity: an index between 0 (no strain heterogeneity) and 100 (all markers present, >1 appear to be from closely related organisms). If the SH index is high (ideally 100%), it suggests most of the contamination is from very similar species and thus any contamination is likely from the pangenome of the species being considered. Alternatively, if the SH index is very low (ideally 0%) this indicates all the contamination is likely from other species.

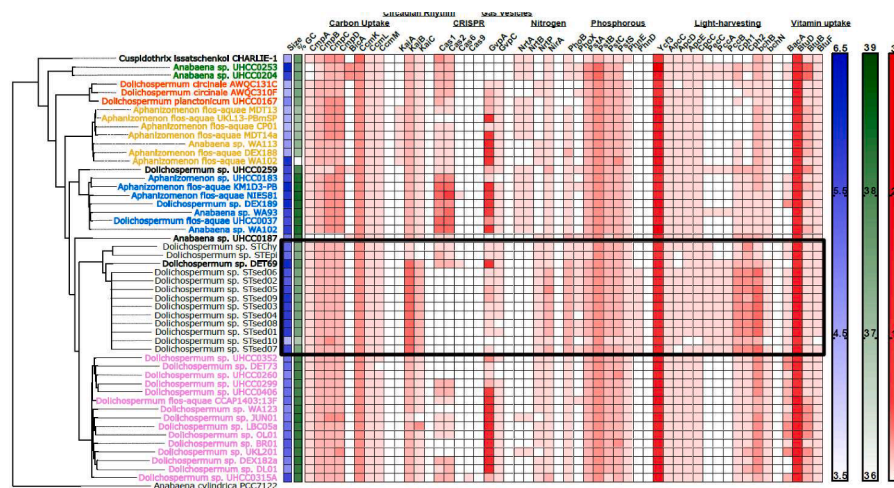


Fig. 2. Phylogenomic analysis of Lake Stechlin *Dolichospermum* genomes in the context of other ADA genome clades. Phylogenomic tree was generated using GTDB-TK and is based on the bac120 single copy marker gene set. Genomes are colored according to the ADA clades based in Driscoll et al., (2018) and Dreher et al. (2021a) and are as follows: ADA-1 = red, ADA-2 = pink, ADA-3 = blue, ADA-4 = yellow, ADA-7 = green, ADA-5, 6, and Lake Stechlin, and 8-10 = black, the Lake Stechlin genomes are surrounded by the black box. STsed01 -10 refers to akinete genomes, STEpi and STChy to the two filaments. Differences in genome characteristics and content are represented as heatmaps with genome size (blue, Mbp), % GC content (green), and gene copies (red) of various ecological functions, the three columns on the right show the range of difference represented by the color change.

Table 2

Overview of the differences in the size and %GC content of the pangenome, core and accessory genomes of ADA-Stechlin and the closest ADA genome clusters.

Clade	Pan-genome		Shared Core		Shared Accessory	Clade-specific Core		Clade-specific Accessory		
	# OGs	% GC	# OGs	% GC		# OGs	% GC	# OGs	% GC	
ADA	15,190	38.15								
Stechlin	5931	37.5	920	39.4	4919	37.1	45	34.3	816	34.3
ADA-2	7237	38.16	920	39.9	4838	37.24	0		1264	37.7
ADA-3	6618	38.35	920	40.1	4685	37.38	9	39.54	1003	37.5
ADA-4	5664	37.41	920	38.8	3169	37	36	34.54	1538	42.4

3.5. Population genome analysis

Approximately 700 polymorphic sites (SNPs) were identified across the nine high-quality Stechlin akinete population genomes (low-quality ST_sed10 excluded). Genetic differences were not correlated significantly with sample depth (cm) (Fig. 3A). Rather, SNP based phylogenetic analysis identified two clades: samples ST_sed1 – ST_sed4, ST_sed8 – ST_sed9, created one clade, and samples ST_sed5 – ST_sed7, a second

clade (Fig. 3B). The presence of population sub-structuring was further corroborated by Principal Component Analysis (PCA), with PC1 accounting for 54% of the variability between the two clades (Supplementary Fig. 3). Consistent with these findings, the STRUCTURE analysis, when set to the best K = 4, delineates two primary populations. The first population comprises isolates ST_sed1 – ST_sed4 and ST_sed8, all of which possess a distinct genomic background indicated in blue. This background is unshared by the isolates ST_sed5 – ST_sed7, which

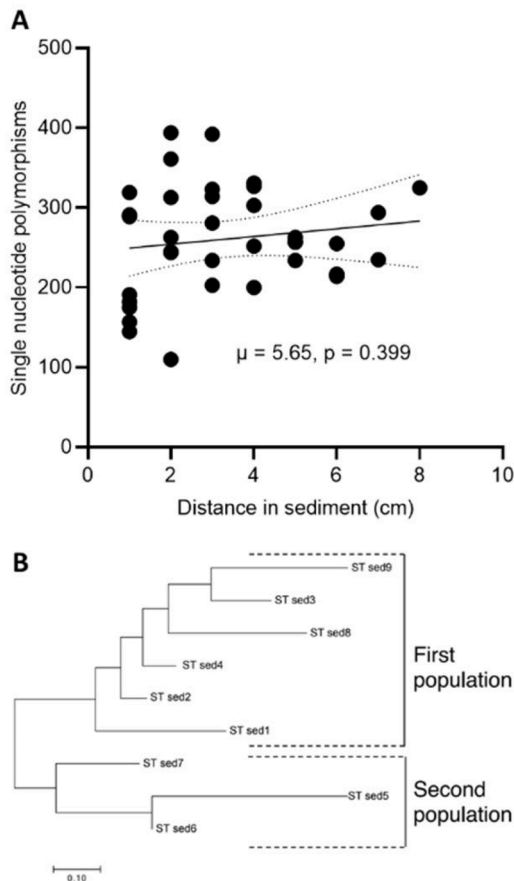


Fig. 3. Population genome analysis of akinete genome sequences. a) The plot illustrate the number of SNPs between each pair of genomes collected from different sediment depths. The x-axis represents the distance between the depths of each pair of genomes, while the y-axis represents the number of SNPs between that pair. Positive slope of the regression line between pairwise SNP difference and sample depth in sediment (age) of the akinete genomes ($R^2 = 0.02088$, $p\text{-value} = 0.399$), μ and dotted lines represents the linear trend line with standard deviation, with $p\text{-value}$ from $t\text{-test}$; b) Sub-populations based on phylogenomic clustering of Lake Stechlin akinete population genomes based on whole genome SNPs.

constitute a separate population and display a unique genomic background, highlighted in green (Supplementary Fig. 4). Fifty-five polymorphic genes (Nucleotide diversity (π) > 0) were detected between the two populations (Supplementary Table 2). Only, two genes, 01883 (Serine/threonine protein kinase) and 01166 (domain of unknown function, DUF29), appeared to be responsible for the divergence (high D_{xy}) and differentiation (high variance, F_{st}) of the two populations, these genes were also under-balancing selection (indicated by a positive Tajima's D) (Fig. 4). The first sub-population contained samples from two time periods, samples ST_sed9 – ST_sed8 represent a period from ~1950 – 1990, when warmer water from the adjacent eutrophic Lake Nehmitz was mixed with Lake Stechlin, reducing the lake water retention time from ca. 60 years to only 300 days. The samples ST_sed4 – ST_sed1 correspond to the current eutrophic period beginning ~2010, with a lake water retention time of ca. 60 years and an increase in phosphorus concentration (Fig 1b, Selmečzy et al., 2019). The second sub-population containing samples ST_sed7 – ST_sed5 corresponds to a time frame of 1990s - 2010, occurring between the time periods of the first population samples, and which correlates with an oligotrophic period with a lake water retention time of ca. 60 years,

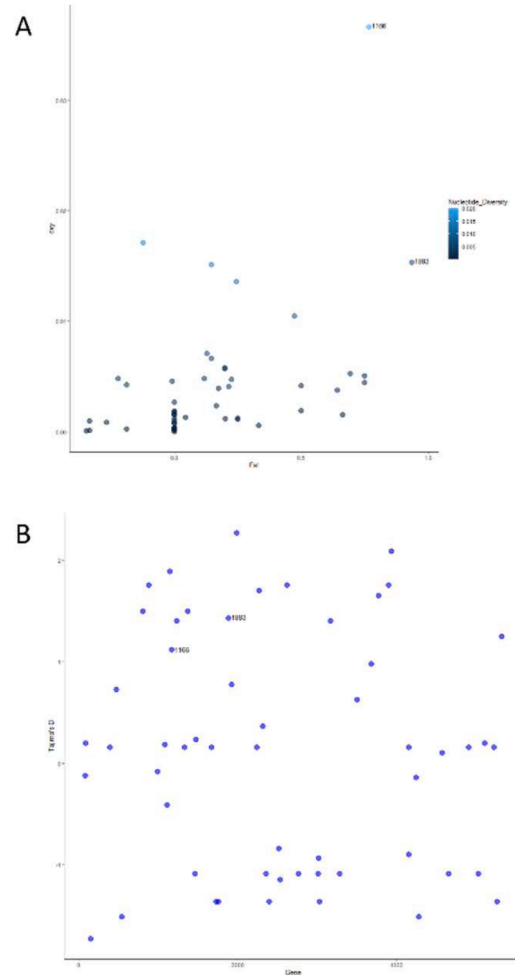


Fig. 4. Population genetic measurements for each gene identified with SNPs, a) scatter plot with F_{st} (variance) vs D_{xy} (divergence), and nucleotide diversity indicated by the colour gradient, and b) Tajima's D; with genes 01883 and 01166 labelled.

4. Discussion

Genomes of *Dolichospermum* akinetes isolated from 1cm segments of the top 10 cm of a sediment core from Lake Stechlin had high genome similarity, to one another and the extant pelagic population. Each genome was assembled from DNA from ten akinetes isolated from each 1cm sediment segment, with this upper 10 cm representing a time-period of around 45 years, as confirmed by previous radioisotope dating (Fuchs et al. 2015, Wurzbacher et al., 2017). During this time period, closure of the adjacent nuclear power plant in 1990 increased water retention times from one to sixty years, and total phosphorous levels increase from ~ 15 $\mu\text{g l}^{-1}$ in the early 2000s to 40 $\mu\text{g l}^{-1}$ in 2016. The increase in total phosphorus has not coincided with a known external phosphorus source and is likely to have resulted from climate change-induced stronger and longer-lasting stratification that has led to the altered coupling of the hypo- and epilimnion, and release of phosphorus from the sediment (Selmečzy et al., 2019). The sediment has been shown to have a high phosphorus content of ~1.29 g kg^{-1} dry weight sediment (Gonsiorczyk et al., 1998). This change in phosphorus levels correlated with increases in the total biomass of *Dolichospermum* sp. and *Aphanizomenon flos-aqua*, and a decrease in zooplankton biomass (Selmečzy et al., 2019; Üveges et al., 2012). However, no akinetes or genomes were found of *Aphanizomenon* spp. potentially indicating an historic misidentification. Thus, the stability of the Lake Stechlin *Dolichospermum* genomes over this time illustrate a population of a single

genotype, without evidence of adaptation to the increasing phosphorus loads during the past ~20 years or changes in genomic elements that might explain its expansion in recent years.

Interestingly, population analysis based on SNP occurrence in the akinete genomes revealed two discrete intraspecific *Dolichospermum* populations. The occurrence of the two *Dolichospermum* populations changed over time and correlated with changes in the lake's trophic status (Fig. 1). One population occurred twice, firstly during the warm/mixing period (1950s – 1990) and secondly returning during the eutrophic period (~2010 – 2017). The second population occurred once during the cooling oligotrophic period (1990s – ~2010) (Selmeczy et al., 2019). Based on high variance (*Fst*) and divergence (*Dxy*) (Fig. 4) two genes appeared responsible for the divergence of the two populations: a serine/threonine-protein kinase (gene 01883) and a hypothetical protein characterized as a 'domain of unknown function', DUF29 (gene 01166). Serine/threonine-protein kinases control phosphorylation and are important regulators of protein expression and thus cell metabolism (Zhang et al., 2007, Yang et al., 2013). The DUF29 has been found across many cyanobacteria, and its structure suggests a possible role in DNA helicase activity. The SNPs in these two genes were responsible for slight changes in amino-acids in the two proteins. Although these amino-acid changes did not appear in predicted functional domains, they may affect the solubility of the proteins (Islam et al., 2012), which in turn may affect their rates of activity (Han et al., 2020), perhaps an important difference between oligotrophic and eutrophic lake conditions. Despite no clear functional change in the two populations, their shift over time does suggest environmental influences on intraspecific population dynamics.

The stability of the *Dolichospermum* genomes across the environmental changes, has two potential implications for their adaptation. In the first case, the genome is sufficiently expansive, possessing a suite of proteins that allow for growth under a range of varying environmental conditions, or adaptation has occurred at a transcriptional/translational level that is undetected through genomics. This is supported by the finding that some of the genes undergoing positive selection are involved in post-translation modifications (Fig. 4, Table S1). In the second case, *Dolichospermum* did not adapt to the environmental changes, but rather that the changes led to an environment more suited to buoyant nitrogen-fixing cyanobacteria: which can move between illuminated surface waters and phosphorus-enriched bottom waters, use luxury phosphorus storage, and supplement low dissolved nitrogen availability through nitrogen fixation, combining to a competitive advantage over other phytoplankton (Carey et al., 2012). Thus, the environmental changes favored *Dolichospermum* and resulted in higher biomass.

Phylogenomic analysis showed that the Lake Stechlin *Dolichospermum* species formed a clade with *Dolichospermum* DET69, isolated from the USA, which has been described as ADA-6 (Dreher et al., 2021b), and is closely related to the ADA-3 *Dolichospermum* clade. The ANI of < 96% (Supplementary Fig 1.), and difference in GC content, indicates ADA-6 and ADA-3 are distinct species. The ADA-3 clade contains mostly *Aphanizomenon* morphotypes and produces the secondary metabolites, anatoxin-a and anabaenopeptin, while the Lake Stechlin/ADA-6 clade contain the gene cluster only for anabaenopeptin. This is in line with previous analysis of Lake Stechlin water samples that found no evidence of toxins produced by *Aphanizomenon* or *Dolichospermum* spp. (Dadheech et al., 2001).

Compared to other ADA species clusters, the genomes of the Lake Stechlin akinetes contains additional genes for light-harvesting and increased phosphorus scavenging capacity (Fig. 2). This is in line with their occupancy in the deep chlorophyll maxima of an oligotrophic lake (Leach et al., 2018). The increase in vitamin B12 transport gene copies suggest interdependence with co-occurring bacteria, and the low incidence of CRISPR-Cas elements suggests infrequent encounters with cyanophages. Together with the overall reduction in %GC content of Lake Stechlin genotypes, these functional genome traits, particularly the

phosphorous and vitamin B12 metabolism, are similar to differences observed amongst oligotrophic and eutrophic *Microcystis* genotypes (Jackrel et al., 2019). Both ADA and *Microcystis* clades have an extensive pangenome and exhibit global distribution (Driscoll et al., 2018). Parallels existing between how particular ADA and *Microcystis* clades adapt to oligotrophic and eutrophic conditions may suggest common mechanisms by which cyanobacteria inhabit these different environmental conditions.

Sequences of ADA genomes are still underrepresented in databases and literature, with fewer than 100 representing at least 10 identifiable species clusters (Dreher et al., 2021b). Whether the number of ADA species continues to expand as more genomes become available is unclear. Strains from geographically distinct locations are present together within various ADA clades, indicating that ADA taxa are not inherently restricted in their migration. The emergence of Lake Stechlin *Dolichospermum* sp. as an ADA-6 species clustering with USA strain DET69, suggests that additional sequencing may reveal this as a global species that occurs in a distinct environmental niche: with a deep chlorophyll maxima. Lake Stechlin is a groundwater-fed lake, and although connected by channels to nearby Lake Nehmitz, it has remained isolated since its formation ~10,000 years ago following the retreat of the glaciers across Europe. The fish *Coregonus fontanae* that is endemic to Lake Stechlin supports this physical isolation. Ultimately, extensive genome sequencing of ADA taxa, particularly in oligotrophic-mesotrophic lakes, will reveal how this species is unique.

In conclusion, we sequenced the genomes of 10 akinetes from each 1 cm sediment segment, representing 45 years of the historical population of *Dolichospermum* to understand its increasing dominance within Lake Stechlin. However, comparative genomics revealed a single ADA-6 genotype with remarkable genome stability, without gene gain or loss events across 45 year period or in response to recent environmental changes. The genome's core characteristics imply occupancy of the deep-chlorophyll maxima are supported by additional light-harvesting capacity and phosphorus scavenging genes. Analysis of SNPs revealed two intraspecific populations, that shifted in dominance, corresponding to the oligotrophic and eutrophic states of the lake. Overall, our findings support the stable occurrence of a species with the capacity to proliferate both in the deep-chlorophyll maxima of oligo-mesotrophic lakes and the epilimnion of eutrophic lakes.

CRedit authorship contribution statement

J.N. Woodhouse: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **M.A. Burford:** Funding acquisition, Resources, Supervision, Writing – review & editing. **B.A. Neilan:** Funding acquisition, Resources, Supervision, Writing – review & editing. **A. Jex:** Methodology, Resources, Supervision, Writing – review & editing. **S. Tichkule:** Data curation, Formal analysis, Investigation, Writing – review & editing. **K. Sivonen:** Resources, Writing – review & editing. **D.P. Fewer:** Resources, Writing – review & editing. **H-P Grossart:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **A. Willis:** Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was funded by ARC-Linkage grants LP130100311 and LP18010083 (BN and AJ), the IGB-Berlin, Department for Experimental Limnology for a visiting scholarship (AW), the German Science Foundation grant (1540/30-1, HPG), and the Victorian State Government Operational Infrastructure Support and Australian Government National Health and Medical Research Council Independent Research Institute Infrastructure Support Scheme awarded to AJ..

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.hal.2024.102600](https://doi.org/10.1016/j.hal.2024.102600).

References

- Alneberg, J., Bjarnason, B., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11 (11), 1144–1146.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120.
- Burford, M.A., Beardall, J., Willis, A., Orr, P., Magalhaes, V., Rangel, L., Azevedo, S., Neilan, B., 2016. Understanding the winning strategies used by the toxic cyanobacterium *Cylindrospermopsis raciborskii*. *Harmful Algae* 54, 44–53.
- Burford, M.A., Davis, T., Orr, P., Rati, S., Willis, A., Neilan, B., 2014. Nutrient-related changes in ecotype dominance of the cyanobacterium *Cylindrospermopsis raciborskii*. *FEMS Microbiology Ecology* 89 (1), 135–148.
- Carey, C.C., Ibelings, B.W., Hoffmann, E.P., Hamilton, D.P., Brookes, J.D., 2012. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. *Water Research* 46 (5), 1394–1407.
- Casper, S.J., Schönborn, W., 1985. *Diffugia limnetica* (Levander) Penard (Protozoa: Testacea) as indicator organism of calcite precipitation in Lake Stechlin, GDR. *Archiv für Protistenkunde* 130 (4), 305–311.
- Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36 (6), 1925–1927.
- Cirés, S., Ballot, A., 2016. A review of the phylogeny, ecology and toxin production of bloom-forming *Aphanizomenon* spp. and related species within the Nostocales (cyanobacteria). *Harmful Algae* 54, 21–43.
- Dadheech, P.K., Selmecky, G.B., Vasas, G., Padišák, J., Arp, W., Tapolczai, K., Casper, P., Krienitz, L., 2001. Presence of potential toxin-producing cyanobacteria in an oligo-mesotrophic lake in Baltic Lake District, Germany: an ecological, genetic and toxicology survey. *Toxins* 6 (6), 2912–2931.
- D'Agostino, P.M., Woodhouse, J.N., Makower, A.K., Yeung, A.C.Y., Ongley, S.E., Micallef, M.L., Moffitt, M.C., Neilan, B.A., 2016. Advances in genomics, transcriptomics and proteomics of toxin-producing cyanobacteria. *Environmental Microbiology Reports* 8 (1), 3–13.
- 2nd Dreher, T.W., Davis, E.W., Mueller, R.S., 2021a. Complete genomes derived by directly sequencing freshwater bloom populations emphasize the significance of the genus level ADA clade within the Nostocales. *Harmful Algae* 103, 102005.
- 2nd Dreher, T.W., Davis, E.W., Mueller, R.S., Otten, T.G., 2021b. Comparative genomics of the ADA clade within the Nostocales. *Harmful Algae* 54, 21–43.
- Driscoll, C.B., Meyer, K.A., Šulcius, S., Brown, N.M., Dick, G.J., Cao, H., Gasiunas, G., Timinskas, A., Yin, Y., Landry, Z.C., Otten, T.G., Davis, T.W., Watson, S.B., Dreher, T. W., 2018. A closely-related clade of globally distributed bloom-forming cyanobacteria within the Nostocales. *Harmful Algae* 77, 93–107.
- Dvořák, P., Jahodářová, E., Stanojković, A., Skoupy, S., Casamatta, D.A., 2023. Population genomics meets the taxonomy of cyanobacteria. *Algal Research* 72, 103128.
- Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T. O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3 e1319.
- Fuchs, A., Selmecky, G.B., Kasprzak, P., Padišák, J., Casper, P., 2015. Coincidence of sedimentation peaks with diatom blooms, wind, and calcite precipitation measured in high resolution by a multi-trap. *Hydrobiologia* 763, 329–344.
- Giling, D.P., Nejtgaard, J.C., Berger, S.A., Grossart, H.-P., Kirillin, G., Penske, A., Lentz, M., Casper, P., Sareyka, J., Gessner, M.O., 2017. Thermocline deepening boosts ecosystem metabolism: evidence from a large-scale lake enclosure experiment simulating a summer storm. *Global Change Biology* 23 (4), 1448–1462.
- Gonsiorczyk, T., Casper, P., Koschel, R., 1998. Phosphorus-binding forms in the sediment of an oligotrophic and an eutrophic hardwater lake of the Baltic Lake District (Germany). *Water Science and Technology* 37 (3), 51–58.
- Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8), 1072–1075.
- Hobbs, W.O., Dreher, T.W., Davis, E.D., Vinebrooke, R.D., Wong, S., Weissman, T., Dawson, M., 2021. Using a lake sediment record to infer the long-term history of cyanobacteria and the recent rise of an anatoxin producing *Dolichospermum* sp. *Harmful Algae* 101, 101971.
- Han, X., Ning, W., Ma, X., Wang, X., Zhou, K., 2020. Improving protein solubility and activity by introducing small peptide tags designed with machine learning models. *Metabolic Engineering Communications* 11, e00138.
- Islam, M.M., Khan, M.A., Kuroda, Y., 2012. Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant. *Biochimica et Biophysica Acta* 1824 (10), 144–1150.
- Jackrel, S.L., White, J.D., Evans, J.T., Buffin, K., Hayden, K., Sarnelle, O., Deneff, V.J., 2019. Genome evolution and host-microbiome shifts correspond with intraspecific niche divergence within harmful algal bloom-forming *Microcystis aeruginosa*. *Molecular Ecology* 28 (17), 3994–4011.
- Jain, C., Rodriguez, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9, 5114.
- Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7, e7359.
- Kaplan-Levy, R.N., Hadas, O., Summers, M.L., Rucker, J., Sukenik, A., 2010. Akinetes: Dormant Cells of Cyanobacteria. In: Lubzens, E., Cerda, J., Clark, M. (eds). In: *Dormancy and Resistance in Harsh Environments*. Topics in Current Genetics, 21. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-12422-8-2>.
- Kasprzak, P., Shatwell, T., Gessner, M.O., Gonsiorczyk, T., Kirillin, G., Selmecky, G., Padišák, J., Engelhardt, C., 2017. Extreme Weather Event Triggers Cascade Towards Extreme Turbidity in a Clear-water Lake Ecosystems. *Ecosystems* 20, 1407–1420.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15, 1179–1191.
- Leach, T.H., Beisner, B.E., Carye, C.C., Pernica, P., Rose, K.C., Huot, Y., Brentrup, J.A., Domaizon, I., Grossart, H.-P., Ibelings, B.W., Jacquet, S., Kelly, P.T., Rusak, J.A., Stockwell, J.D., Straile, D., Verburg, P., 2018. Patterns and drivers of deep chlorophyll maxima structure in 100 lakes: the relative importance of light and thermal stratification. *Limnology and Oceanography* 66 (2), 628–646.
- Legrand, B., Lamarque, A., Sabart, M., Latour, D., 2016. Characterization of akinetes from cyanobacterial strains and lake sediment: a study on their resistance and toxic potential. *Harmful Algae* 59, 42–50.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25 (14), 1754–1760.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31 (10), 1674–1676.
- Li, X., Dreher, T.W., Li, R., 2016. An overview of diversity, occurrence, genetics and toxin production of bloom-forming Dolichospermum (Anabaena) species. *Harmful Algae* 54, 54–68.
- Lima, S.T., Fallon, T.R., Cordoza, J.L., Chekan, J.R., Delbaje, E., Hopiavuori, A.R., Alvarenga, D.O., Wood, S.M., Juhavaya, H., Baumgartner, J.T., Dorr, F.A., Etcheagaray, A., Pinto, E., McKinnie, S.M.K., Fiore, M.F., Moore, B.S., 2022. Biosynthesis of Guanitoxin enables global environmental detection in freshwater cyanobacteria. *Journal of the American Chemical Society* 144, 9372–9379.
- Nübel, U., Garcia-Pichel, F., Muyzer, G., 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology* 63 (8), 3327–3332.
- Osterholm, J., Popin, R.V., Fewer, D.P., Sivonen, K., 2020. Phylogenomic analysis of secondary metabolism in the toxic cyanobacterial genera *Anabaena*, *Dolichospermum* and *Aphanizomenon*. *Toxins* 12 (4), 248.
- Padišák, J., Krienitz, L., Scheffler, W., Koschel, R., Kristiansen, J., Grigorszky, I., 1998. Phytoplankton succession in the oligotrophic Lake Stechlin (Germany) in 1994 and 1995. *Hydrobiologia* 369, 179–197.
- Padišák, J., Scheffler, W., Kasprzak, R., Koschel, R., Krienitz, L., 2003a. Interannual variability in the phytoplankton composition of Lake Stechlin (1994–2000). *Archiv für hydrobiologie, Special Issues Advances in Limnology* 58, 101–133.
- Padišák, J., Barbosa, F., Koschel, R., Krienitz, L., 2003b. Deep layer cyanoprokaryota maxima in temperate and tropical lakes. *Archiv für hydrobiologie, Special Issues Advances in Limnology* 58, 175–199.
- Padišák, J., Hajnal, E., Krienitz, L., Lakner, J., Uveges, V., 2010. Rarity, ecological memory, rate of floral change in phytoplankton – and the mystery of the Red Cock. *Hydrobiologia* 653, 45–64.
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31 (22), 3691–3693.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25 (7), 1043–1055.
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular biology and evolution* 31 (3), 1929–1936.
- Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetic* 155 (2), 945–959.
- Ramm, J., Lupu, A., Hadas, O., Ballot, A., Rucker, J., Wiedner, C., Sukenik, A., 2012. A CARD-FISH protocol for the identification and enumeration of cyanobacterial akinetes in lake sediments. *FEMS Microbiology Ecology* 82 (1), 23–36.
- Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30 (14), 2068–2069.
- Selmecky, G.B., Tapolczai, K., Casper, P., Krienitz, L., Padišák, J., 2015. Spatial- and niche segregation of DCM-forming cyanobacteria in Lake Stechlin (Germany). *Hydrobiologia* 764, 229–240.
- Selmecky, G.B., Abonyi, A., Krienitz, L., Kasprzak, P., Casper, P., Telcs, A., Somogyvári, Z., Padišák, J., 2019. Old sins have long shadows: climate change weakens efficiency of trophic coupling of phyto- and zooplankton in a deep oligo-

- mesotrophic lowland lake (Stechlin, Germany)—a causality analysis. *Hydrobiologia* 831, 101–117.
- Sieber, C.M.K., Probst, A.J., Sharrar, A., Brian, T.C., Hess, M., Tringe, S.G., Banfield, J.F., 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* 3 (7), 836–843.
- Shan, K., Song, L., Li, L., Liu, L., Wu, Y., Jia, Y., Zhou, Q., Peng, L., 2019. Analysis of environmental drivers influencing interspecific variations and associations among bloom-forming cyanobacteria in large, shallow eutrophic lakes. *Harmful Algae* 84, 84–94.
- Stecher, G., Tamura, K., Kumar, S., 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37 (4), 1237–1239.
- Steffen, M.M., Li, Z., Effler, T.C., Hauser, L.J., Boyer, G.L., Wilhelm, S.W., 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLOS One* 7, e44002.
- Üveges, V., Tapolczai, K., Krienitz, L., Padisák, K., 2012. Photosynthetic characteristics and physiological plasticity of an *Aphanizomenon flos-aquae* (Cyanobacteria, Nostocaceae) winter bloom in a deep oligo-mesotrophic lake (Lake Stechlin, Germany). *Hydrobiologia* 698, 263–272.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* 43 (1110), 11.10.11–33.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sathikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9 (11), e112963.
- Weisbrod, B., Wood, S.A., Steiner, K., Whyte-Wilding, R., Puddick, J., Laroche, O., Dietrich, D.R., 2020. Is a central sediment sample sufficient? Exploring spatial and temporal microbial diversity in a small lake. *Toxins* 12, 580.
- Willis, A., Bent, S.J., Jameson, L.D., 2020. Morphological changes and genome evolution in *Raphidiopsis raciborskii* CS-506 after 23 years in living culture. *Applied Phycology* 3 (1), 189–198.
- Willis, A., Woodhouse, J.N., 2020. Defining Cyanobacterial Species: Diversity and Description through Genomics. *Critical Reviews in Plant Sciences* 39 (2), 101–124.
- Wood, S.A., Jentzsch, K., Rueckert, A., Hamilton, D., Craig, C., 2009. Hindcasting cyanobacterial communities in Lake Okaro with germination experiments and genetic analyses. *FEMS Microbiology and Ecology* 67, 252–260.
- Wu, Y.-W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32 (4), 605–607.
- Wurzbacher, C., Fuchs, A., Attermeyer, K., Frindte, K., Grossart, H.-P., Hupfer, M., Casper, P., Monaghan, M.T., 2017. Shifts among Eukaryota, Bacteria, and Archaea define the vertical organization of a lake sediment. *Microbiome* 5 (1), 41.
- Yang, M.-K., Qiao, Z.-X., Zhang, W.-Y., Xiong, Q., Zhang, J., Li, T., Ge, E.F., Zhao, J.-D., 2013. Global phosphoproteomic analysis reveals diverse functions of Serine/Threonine/Tyrosine phosphorylation in the model cyanobacterium *Synechococcus* sp. strain PCC 7002. *Journal of Proteome Research* 12 (4), 1909–1923.
- Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K., Yu, J., 2006. KaKs Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4 (4), 259–263.
- Zhang, X., Zhao, F., Guan, X., Yang, Y., Liang, C., Qin, S., 2007. Genome-wide survey of putative Serine/Threonine protein kinases in cyanobacteria. *BMC Genomics* 8 (8), 395.