Title

Two decades of bacterial ecology and evolution in a freshwater lake

Authors

Rohwer, Robin R
Kirkpatrick, Mark
Garcia, Sarahi L
et al.

Publication Date

2025-01-03

DOI

10.1038/s41564-024-01888-3

Peer reviewed

nature microbiology

Resource

https://doi.org/10.1038/s41564-024-01888-3

Two decades of bacterial ecology and evolution in a freshwater lake

Received: 28 April 2024

Accepted: 14 November 2024

Published online: 3 January 2025

Check for updates

Robin R. Rohwer[1] ✉, Mark Kirkpatrick[1], Sarahi L. Garcia[2,3], Matthew Kellom[4], Katherine D. McMahon[5,6] ✉ & Brett J. Baker[1,7] ✉

Ecology and evolution are considered distinct processes that interact on contemporary time scales in microbiomes. Here, to observe these processes in a natural system, we collected a two-decade, 471-metagenome time series from Lake Mendota (Wisconsin, USA). We assembled 2,855 species-representative genomes and found that genomic change was common and frequent. By tracking strain composition via single nucleotide variants, we identified cyclical seasonal patterns in 80% and decadal shifts in 20% of species. In the dominant freshwater family *Nanopelagicaceae*, environmental extremes coincided with shifts in strain composition and positive selection of amino acid and nucleic acid metabolism genes. These genes identify organic nitrogen compounds as potential drivers of freshwater responses to global change. Seasonal and long-term strain dynamics could be regarded as ecological processes or, equivalently, as evolutionary change. Rather than as distinct interacting processes, we propose a conceptualization of ecology and evolution as a continuum to better describe change in microbial communities.

Microbial communities allow us to observe eco-evolutionary dynamics in real time due to the short lifespans and large population sizes of microbes[1,2]. Real-time evolution was famously observed in the *Escherichia coli* long-term evolution experiment[3], but few long-term observations exist for natural, ecologically complex systems. Here, we introduce a two-decade, 471-sample microbial time series from a freshwater lake, the TYMEFLIES dataset[4], which allows us to directly observe ecology and contemporary evolution in a natural ecosystem. The Lake Mendota (Wisconsin, USA) microbial observatory[5] is part of the North Temperate Lakes Long-Term Ecological Research programme[6], which builds on limnological research dating back to the late 1800s. Long-term and abrupt change in Lake Mendota are well documented and linked to multiple interacting drivers including climate[7–9], land use[10,11] and invasive species[12,13], and these drivers are also impacting the lake's microbial communities[14,15].

The dynamism of freshwater and marine bacterial communities, especially in response to seasonal drivers, is evident in several long-term time series where 16S ribosomal RNA genes were used to define species-like units[16–18], as well as in Lake Mendota[14]. However, a genome-resolved approach is necessary to incorporate evolution into our understanding of microbial community change. Selective pressures change gene frequencies, which manifest as genomic diversity[19]. This microdiversity can be measured across samples by mapping short metagenomic reads against reference genomes and identifying single nucleotide variants (SNVs) in the mapped reads[20,21]. Such strain-resolved approaches have identified both the ecological relevance of strains[22,23] and a variety of evolutionary strategies shaping them[24–27].

Here, we describe community-wide strain-resolved bacterial change over 20 years. By reconstructing tens of thousands of

[1]Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA. [2]Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany. [3]Science for Life Laboratory, Department of Ecology, Environment, and Plant Sciences, Stockholm University, Stockholm, Sweden. [4]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI, USA. [6]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, WI, USA. [7]Department of Marine Science, The University of Texas at Austin, Austin, TX, USA. ✉e-mail: Robin.Rohwer@gmail.com; trina.mcmahon@wisc.edu; brett_baker@utexas.edu

https://doi.org/10.1038/s41564-024-01888-3

metagenome-assembled genomes (MAGs), we found that inter- and intraspecific changes unfold at short, seasonal time scales as well as longer-term decadal time scales, in some cases coinciding with environmental extremes. Research on such eco-evolutionary dynamics usually focuses on feedbacks between distinct processes of ecology and evolution[28–30]. In our microbial data, however, these processes were difficult to distinguish. Ecological dynamics appeared to occur between strains within a population, but the strains themselves were inferred from observations of genomic change. Consistent with the ambiguity of the microbial species concept[31], our observations suggest that it is not possible to cleanly delineate between ecological and evolutionary processes in natural microbial communities. Therefore, we propose an adjusted conceptualization, where ecology and evolution converge along a continuum.

Results

The TYMEFLIES dataset

We collected 471 samples over 20 years from Lake Mendota (Wisconsin, USA)[4] and obtained shotgun DNA libraries (Fig. 1a and Supplementary Data 1). We refer to these 'Twenty Years of Metagenomes Exploring Freshwater Lake Interannual Eco/evo Shifts' as the TYMEFLIES dataset. By cross-mapping reads from ~50 metagenomes to each single-sample metagenome assembly, we obtained a total of 85,684 genome bins, 30,389 of which were medium or high quality (>50% completeness and <10% contamination)[32]. We clustered these 30,389 bins at 96% average nucleotide identity (ANI) and obtained 2,855 clusters from which we chose representative MAGs[33] (Supplementary Data 2). Several previous studies have found an emergent species boundary at similar ANI cut-offs[34–36], and we observed a rapid increase in the number of clusters above the 96% ANI cut-off. In this study, we treat the representative MAGs from each 96% ANI cluster as bacterial species and refer to subspecies delineations identified in the mapped metagenomic reads as strains[20].

The representative MAGs have high estimated completeness (median 86%) and low contamination (median 0.9%) (Fig. 1b and Supplementary Data 2) and reflect the abundant members of the lake's bacterial community, especially in well-sampled seasons (Fig. 1c). Using a 16S rRNA gene amplicon dataset from the same time series[4] as a reference for the expected community composition (Fig. 1d), we found that our representative MAGs comprise most of the abundant taxa (Fig. 1e). Moreover, we obtained 168 representative MAGs from the Nanopelagicales order, which is the most abundant order in Lake Mendota and accounts for 22% of the amplicon reads and 10% of the mapped metagenomic reads. Similar to SAR11 bacteria in the oceans, this freshwater lineage is abundant in lakes globally[37], difficult to culture[38] and typically has highly streamlined genomes[39].

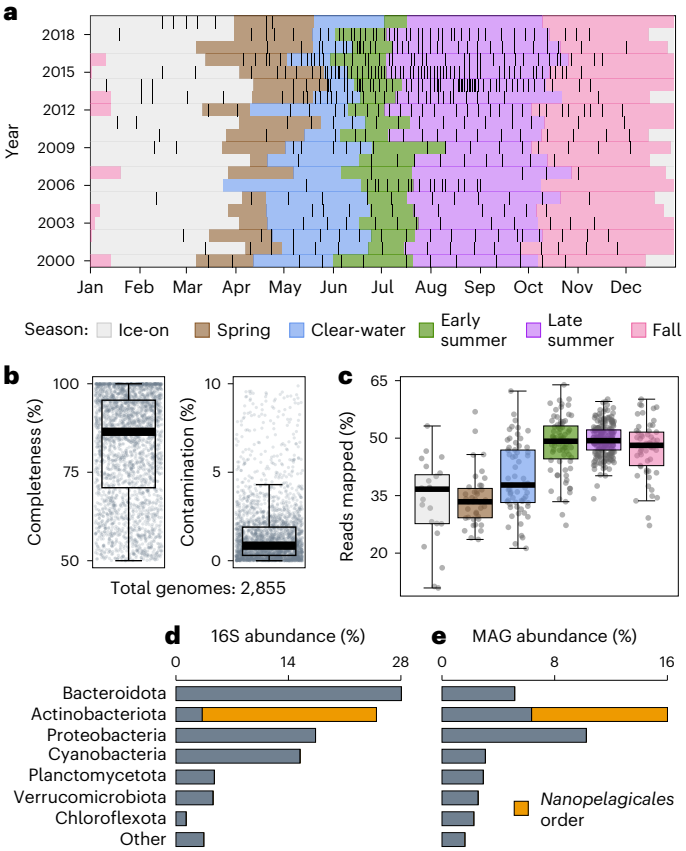

Fig. 1 | The TYMEFLIES dataset. a, The metagenome sample dates are indicated by black vertical lines, and microbial seasons[14] are indicated by coloured shading. Ice-on indicates contiguous ice cover; spring, a diatom bloom; clear-water, a phase of intense zooplankton grazing and high water clarity; early summer, a switch to cyanobacterial dominance; late summer, a period of strong thermal stratification; and fall, an unstratified period after fall mixing. **b**, The quality of the 2,855 representative genomes obtained after clustering to 96% ANI. We treat these genomes as species. **c**, The percent of metagenome reads from each sample that mapped to all reference genomes with an ANI ≥93%. The samples are grouped by season to highlight how well the reference genomes reflect each seasonal community. **d**, The rank abundance of phyla as measured by 16S rRNA gene amplicon sequencing[4]. The abundant *Nanopelagicales* order of Actinobacteria is highlighted. **e**, The abundance of phyla in the TYMEFLIES reference genomes, quantified as the mean relative abundance normalized by genome size and sequencing depth. The *Nanopelagicales* order is again highlighted. The box plots indicate Q1 − 1.5 × interquartile range (IQR), Q1, median, Q3 and Q3 + 1.5 × IQR.

Seasonal ecology and evolution

From a century of limnological research, we know that Lake Mendota follows a consistent annual phenology and that phenological patterns are changing in response to climate change and invasive species[40–43]. These seasonal dynamics are evident in bacterial[14], viral[44] and protistan[45] community composition. To confirm that phenological abundance patterns also exist in our finely resolved bacterial species, we identified annual peaks in species relative abundance using periodograms (magnitude of Fourier transforms). After limiting this temporal analysis to the subset of 1,474 species that occurred at least 30 times over at least 10 years, we found that 72% of them have consistent seasonal abundance patterns (Fig. 2a).

To determine whether evolutionary dynamics (that is, changes in allele frequency within the species) also unfold seasonally, we mapped reads from each sample against each species' reference genome and identified shifts in strain composition from changes in nucleotide diversity (π) and allele frequencies at SNVs. We found that 33% of the 1,474 species displayed consistent seasonal nucleotide diversity patterns (Fig. 2a). To gain greater resolution of the strain composition of the 236 species abundant enough over time to reliably call SNVs (median coverage >10×), we created a 'SNV profile' for each date with the frequencies of the reference alleles. For each species, we calculated the Euclidean distance between every date's SNV profile (Fig. 2b). We found that 80% of these 236 abundant species had consistent phenological patterns in their strain composition. This demonstrates that phenological patterns evident in the bacterial community extend to the finest possible taxonomic resolution. Several short-term freshwater studies have also observed changes in strain composition on seasonal time scales[46,47]. Phenological patterns in subspecies strains similar to those at the species level suggest ecological processes may shape bacterial strain composition, but these changes are evidenced by intraspecific genomic change and could thus also be interpreted as seasonal evolution.

Given the ubiquity of seasonal patterns in both species abundance and subspecies diversity, we asked whether they were correlated. We quantified whether a species' 'bloom' in abundance consisted of fewer

https://doi.org/10.1038/s41564-024-01888-3

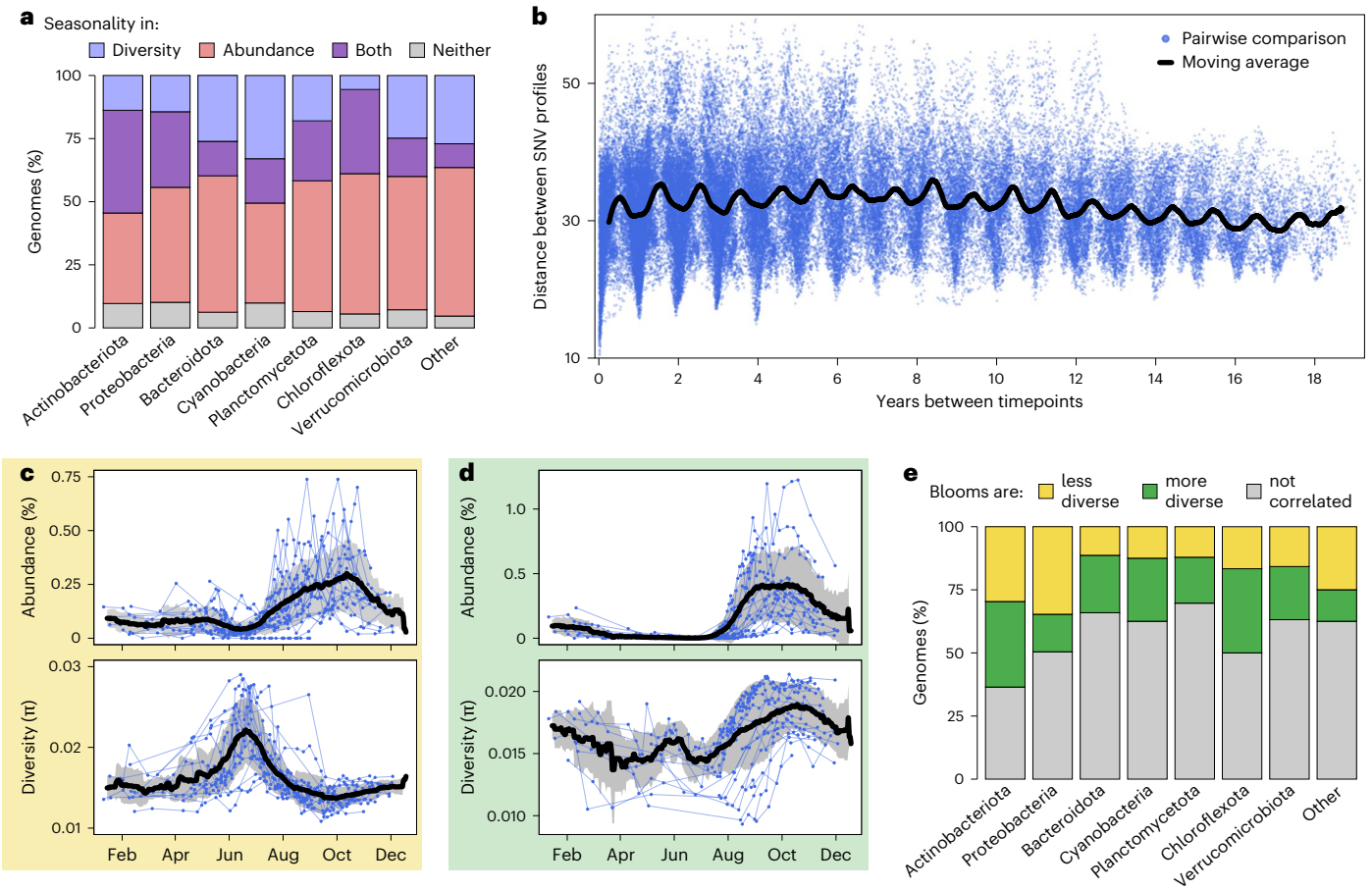

Fig. 2 | Bacterial seasonality at the subspecies level. a, The per cent of species with seasonality in nucleotide diversity and abundance (a centred log ratio transform was applied to relative abundances). The 1,474 species that occurred at least 30 times were included in this analysis. **b**, A time-decay plot of the Euclidean distances between the SNV profiles of an abundant species in the *Nanopelagicus* genus (ME2017-06-13_3300043469_group7_bin14). A smaller distance between SNV profiles indicates that the strain composition is more similar. Each blue point represents a pairwise comparison between two sample dates, with the time between those dates on the *x* axis. The black line is a 6 month moving average, drawn to highlight the annual periodicity of strain similarities. **c**, An example of a less diverse bloom, where nucleotide diversity decreases as relative abundance increases. Displayed is an abundant species in the *Planktophila* genus (ME2011-09-04_3300044729_group3_bin142). **d**, An example of a more diverse bloom, where nucleotide diversity increases as relative abundance increases. Displayed is an abundant species in the *Nanopelagicaceae* family, MAG-120802 genus (ME2012-08-31_3300044613_group4_bin150). The thin blue lines represent individual years, and thick black lines with shading represent mean ± s.d. **e**, The distribution of bloom diversity patterns across the 365 species that had seasonality in both abundance and nucleotide diversity.

strains or more strains than its baseline composition. Of the 365 species with seasonal patterns in both abundance and nucleotide diversity (Fig. 2a, purple bars), we found that both scenarios were common; 21% of these species had less diverse blooms (Fig. 2c,e, yellow bars), while 19% had more diverse blooms (Fig. 2d,e, green bars). Further, all abundant phyla demonstrated an even mix of both bloom types (Fig. 2e). A lower-diversity bloom suggests that a subset of strains outcompeted the others, while a higher-diversity bloom suggests that micro-niches allowed rarer strains to gain abundance, resulting in higher strain diversity[48] due to a more even strain composition. This is in agreement with a previous study that found both overlapping and distinct niches within freshwater bacterial species[25]. The prevalence of both bloom diversity patterns suggests ecological processes drive changes in allele frequencies.

Long-term ecology and evolution

Long-term changes can be masked by seasonal oscillations, lost in what is referred to as the 'invisible present'[49]. The unprecedented length of the TYMEFLIES metagenome dataset provides a unique lens into the invisible present, enabling the identification of overlayed long-term patterns. To find long-term changes in strain composition, we developed a classifier trained on the distance between each date's SNV profile and the SNV profile of that species' first occurrence in the time series. We trained this classifier on 11 examples of manually identified temporal patterns and then applied it to all 263 species with sufficient abundance to reliably call SNVs. Our classifier identified gradual change (Fig. 3a), which may arise from genetic drift or in response to a slow press disturbance. It also identified abrupt change (Fig. 3b,c), which may arise in response to a new stable state after a tipping point or from a sudden environmental shift[50,51]. Among instances of abrupt change, we identified step changes (Fig. 3b), where the new strain composition persisted during the remainder of our time frame, as well as patterns of disturbance with resilience (Fig. 3c), where the strain composition recovered to baseline.

We found that 21% of the most abundant species experienced one kind of long-term change in their SNV profiles during our 20 year study period, and these changes overlayed both seasonal and acyclical short-term dynamics (Fig. 3d). Abrupt change was almost twice as common as gradual change (seen in 36 versus 19 species), and resilience was only slightly more common than a lasting step change (20 versus 16 species) (Fig. 3d). The three long-term change patterns were found in many abundant species distributed across phyla (Fig. 3e). Many species

https://doi.org/10.1038/s41564-024-01888-3

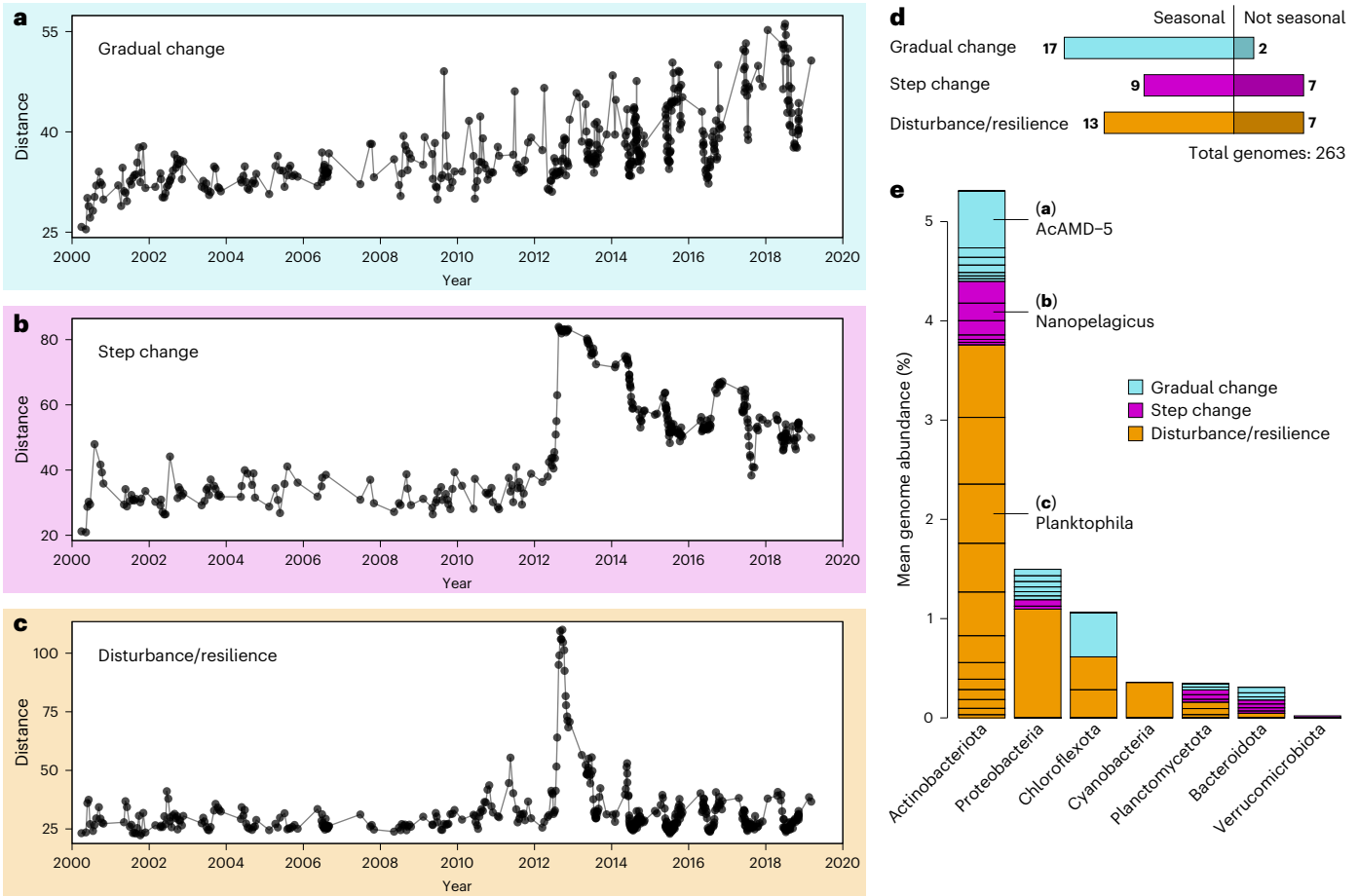

Fig. 3 | Long-term changes in strain composition. a, An example of long-term, gradual change in strain composition. The points indicate sample dates and distance refers to the Euclidean distance between a species' SNV profile on that sample date and its first occurrence in the time series. A species in the Nanopelagicales order, AcAMD-5 family is shown (ME2005-06-22_3300042363_group2_bin84). **b**, An example of an abrupt step change in strain composition in a species in the *Nanopelagicus* genus (ME2011-09-21_3300043464_group3_bin69). **c**, An example of a disturbance/resilience pattern, where an abrupt change in strain composition is followed by recovery to the original strain composition, in a species in the *Planktophila* genus (ME2015-07-03_3300042555_group6_bin161). **d**, Long-term change patterns often overlayed seasonal patterns. Of the 263 species abundant enough to observe their SNV profiles, 39 had both long-term and seasonal patterns, while 16 had only long-term patterns. **e**, The distribution of long-term patterns across phyla. Each species that underwent long-term change is indicated by a section of the phyla's bar, scaled by the mean abundance of that species. The sections corresponding to the examples highlighted in **a**–**c** are labelled.

in the Actinobacteriota phylum were abundant enough to include in this analysis, providing a detailed view of change in these common freshwater heterotrophs. Long-term changes in SNV profiles reflect shifts in intraspecific strain composition, which is typically attributed to evolutionary processes[52]. The fact that during our observation period over a fifth of the species experienced long-term changes in their SNV profiles emphasizes the importance of including contemporary evolutionary change in our understanding of microbial ecology.

Abrupt changes in *Nanopelagicaceae*

In general, related species did not change in unison with each other, suggesting that the drivers of evolutionary change are highly specific (Fig. 4a). One exception is an abrupt change event that impacted seven species within the *Nanopelagicaceae* family (acI) in 2012, specifically species in the *Nanopelagicus* and *Planktophila* genera (acI-B and acI-A). This is the most abundant family in Lake Mendota and in freshwaters globally[37], and the 127 *Nanopelagicaceae* species we recovered together accounted for 8% of the relative abundance on average. Five of these *Nanopelagicaceae* species displayed resilience to an abrupt change in 2012, while two experienced lasting step changes in strain composition.

A myriad of possible environmental variables could have driven this event. A leading candidate is extreme weather as Lake Mendota was unusually warm and dry in 2012. The lake experienced high epilimnion water temperatures during spring and summer, with the hottest July on record since 1894[14] (Fig. 4b), the fifth shortest winter ice duration on record since 1856[53] (Fig. 4c), the eighth lowest annual discharge from its major tributary on record since 1976 and the second lowest peak discharge[54] (Fig. 4e). These environmental conditions led to top-down and bottom-up controls on the lake's primary productivity. The highest spring zooplankton abundance since measurements began in 1994[55] (Fig. 4d) was probably a result of the mild winter and spring[56], which allowed zooplankton, including the prolific grazer *Daphnia pulicaria*, to establish early. Low total phosphorus and soluble reactive phosphorus (Fig. 4f,g) was probably a result of low external nutrient loading associated with mild discharge events[57]. The resulting combination of high zooplankton grazing and low phosphorus, typically the limiting nutrient in lakes, may be responsible for low phytoplankton biomass (Fig. 4h), which in Lake Mendota is dominated by cyanobacteria during summer[58]. Lake Mendota's dissolved organic carbon (DOC) is primarily provided by phytoplankton[59], consequently DOC was also low in 2012 (Fig. 4i). Lake heatwaves are predicted to become hotter and longer with climate change[60], and these observations suggest that the intense epilimnetic heatwaves during 2012 had cascading effects on lake biogeochemistry that extended to the level of bacterial strains.

https://doi.org/10.1038/s41564-024-01888-3

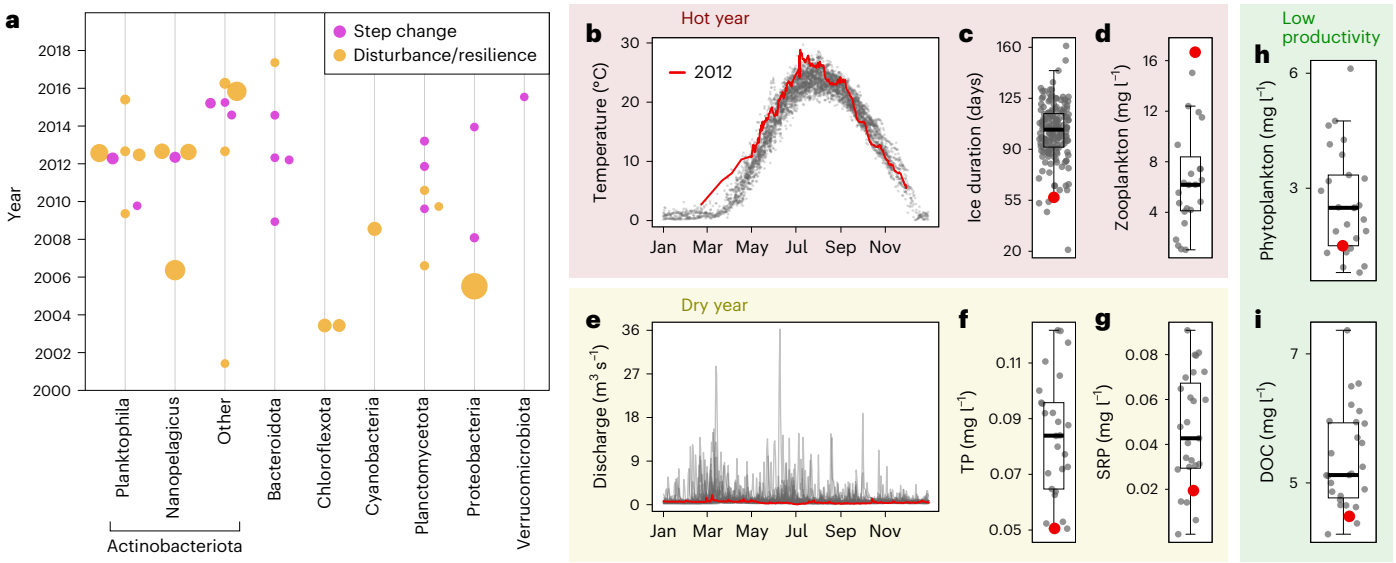

Fig. 4 | Abrupt changes in *Nanopelagicaceae* strain composition coincide with environmental extremes in 2012. a, Dates of all abrupt changes in strain composition arranged by phyla. Most changes were isolated events, but multiple species from two abundant genera of Actinobacteriota, *Planktophila* and *Nanopelagicus*, experienced abrupt change in 2012. The point size is scaled by species abundance. **b**, Unusually high epilimnion water temperatures during spring and summer 2012 (relative to 1894–2019). **c**, The preceding winter had an unusually short ice duration (relative to 1853–2023). **d**, The total zooplankton biomass (excluding predatory *Bythotrephes* and *Leptodora*) was unusually high, probably enabled by warm early spring temperatures (relative to 1995–2018). **e**, Discharge from the Yahara River, the main tributary to Lake Mendota, was unusually low and lacked high run-off events typical after storms and spring snowmelt (relative to 1989–2021). **f**,**g**, The total phosphorus (TP) (**f**) and soluble reactive phosphorus (SRP) (**g**) were low (relative to 1995–2021), probably due to low sediment transport. **h**, Low phytoplankton biomass (relative to 1995–2020), probably resulting from both high zooplankton grazing and low nutrient availability. **i**, Low dissolved organic carbon (DOC) (relative to 1996–2022), probably a result of low phytoplankton abundance. The box plots indicate Q1 − 1.5 × IQR, Q1, median, Q3 and Q3 + 1.5 × IQR.

Another possible driver is the irruption of the invasive zooplankton spiny water flea (*Bythorephes cederströmii*) in 2009, which itself was driven by an unusually cool summer[61]. This major disturbance resulted in a trophic cascade that decreased water clarity[13,43], increased lake anoxia[55] and shifted the bacterial community composition[14]. Although the abrupt changes in strain composition of seven *Nanopelagicaceae* species were not observed until 3 years later, lag effects are common in complex ecosystems[62]. In contrast to the 2009 species invasion, we did not see bacterial community-level shifts corresponding to the 2012 extreme weather, but environmental drivers of strain dynamics may be highly specific. Ecosystem-wide drivers like these two disturbances can have cascading and interacting effects on nutrient and carbon dynamics, which in turn impact bacteria. The observed long-term intraspecific changes suggest that such ecological drivers are also drivers of evolutionary change, further emphasizing how ecology and evolution are intertwined.

Evolutionary signals in a *Nanopelagicus*

To understand the dynamics of abrupt evolutionary change, we further examined one of the abundant species, a *Nanopelagicus* (acI-B), that experienced a step change in strain composition in August 2012 (Fig. 3b). A non-metric multidimensional scaling (NMDS) ordination of its SNV profiles indicated the strain composition changed abruptly at that time and settled into a new composition after a period of adjustment in 2012 and 2013 (Fig. 5a).

The relative abundance of this species was quite constant throughout our 20 year observation period (Fig. 5b), typically with higher abundances during the spring clear-water phase. The step change in strain composition (Fig. 3b) coincided with one in genome-wide nucleotide diversity (Fig. 5c). These patterns could result from the introduction of a new strain or from an increase in the evenness of existing strain abundances. To distinguish between these hypotheses, we counted the number of previously unobserved SNVs in the mapped reads of every sample. We did not see large spikes in new SNVs in 2012 (Fig. 5d), suggesting that the step change reflects shifts in the relative abundances of existing strains.

This interpretation is consistent with a dramatic increase in the number of genes under positive selection that occurred at this time (Fig. 5e). As the relative abundances of some strains increase, alleles specific to them appear to undergo partial (or 'soft') selective sweeps. If strain composition re-equilibrated, this signal would die out. However, the increase in the number of genes under selection persisted (Fig. 5e). This could arise from continuing fluctuations in strain abundances, consistent with the larger distances between SNV profiles seen after the step change (Fig. 5a). To identify candidate loci that reflect the phenotypic differences between strains driving adaptations, we sought genes that consistently showed signs of being positively selected over the entire time series, only during the pre-2012 period and only during the post-2012 period. Four genes were consistently selected both pre- and post-2012, four genes were consistently selected pre-2012 and 33 genes were consistently selected post-2012. We used gene functional predictions[63] to identify their potential metabolic pathways. Of the 33 consistently selected genes post-2012, ten are involved in amino acid metabolism or aminoacylation and six are involved in nucleic acid synthesis or degradation (Fig. 5f).

Previously, the absence of biosynthesis or auxotrophies for amino acids and nucleotides has been highlighted for microorganisms with streamlined genomes[64,65]. In the streamlined *Nanopelagicus*, auxotrophies for various amino acids[39,66] coupled with an enrichment of transporters for many small organic nitrogen compounds, including amino acids[39,67,68] and nucleic acid components[39,66–68], are common. Moreover, the histidine pathway was found split between two different strains of *Nanopelagicus* growing in a mixed culture[66]. Our observation of consistent selection on amino acid and nucleic acid metabolism suggests that these genes differentiate the post-2012 strains. Additionally, the low phytoplankton biomass in 2012 (Fig. 4h) might indicate lower

https://doi.org/10.1038/s41564-024-01888-3

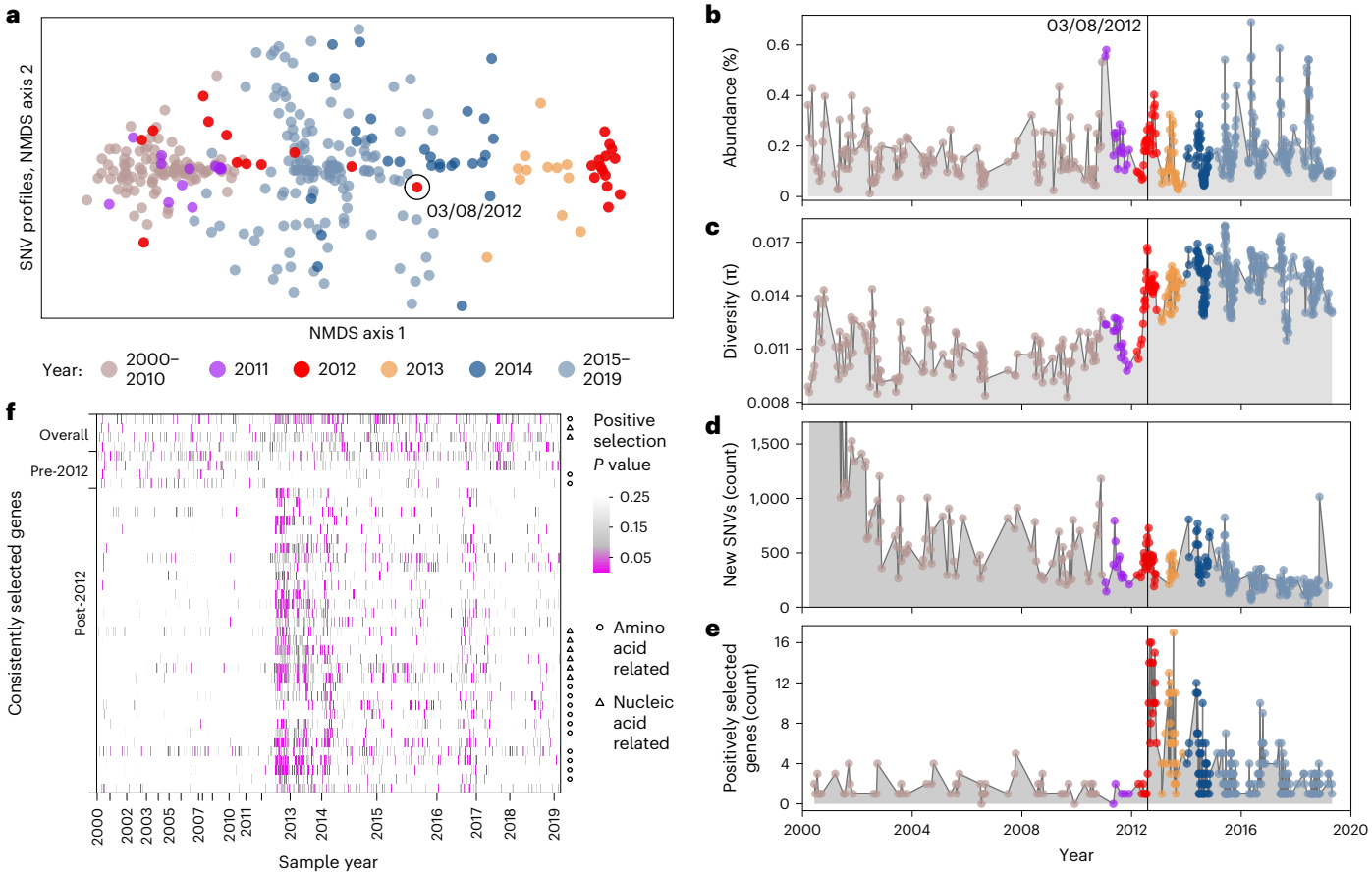

Fig. 5 | Step change in strain composition coincides with more genes under selection. a, An abundant *Nanopelagicus* species experienced a step change in strain composition in 2012 (ME2011-09-21_3300043464_group3_bin69, see also Fig. 3b). Samples with more similar SNV profiles appear closer on this NMDS ordination. The years 2000–2011 cluster together and are distinct from years 2014–2019, which cluster separately. A sudden change in strain composition occurred on 3 August 2012. **b**, Despite the abrupt change in strain composition, the relative abundance of this species remained constant over time. **c**, Concurrent with the shift in strain composition, nucleotide diversity increased and then remained high, indicating that the new equilibrium comprised a more diverse assemblage of strains. **d**, The absence of a spike in the number of new SNVs suggests that an increase in the evenness of existing strains occurred, rather than the introduction of new strains. **e**, Concurrent with the shift in strain composition, the number of genes under positive selection also increased (McDonald–Kreitman two-sided *F*-statistic *P* value <0.05). **f**, The occurrence of consistently selected genes in all the samples, in the pre-2012 period and in the post-2012 period. The *x* axis indicates samples ordered consecutively and the *y* axis indicates genes. The shading indicates the significance level of positive selection (McDonald–Kreitman two-sided *F*-statistic *P* value). Amino acid-related genes and nucleic acid-related genes are indicated on the right axis. Full annotations are available in Supplementary Data 3. Note that the *x* axis is evenly spaced by sample, so that years with more samples take up more space.

influx of fixed nitrogen into the system, which could have cascading effects on the processing of organic nitrogen in abundant microorganisms. Therefore, it appears that biosynthesis, use and reuse of small organic nitrogen compounds are key in the ecology and evolution of these globally abundant lake bacteria.

Discussion

Freshwater lakes are focal points on the terrestrial landscape, processing an estimated 70% of net terrestrial carbon production[69]. These ecosystems are stressed by both climate change[70] and invasive species[71], but whether lakes will become net sources or sinks of carbon is uncertain[72,73]. The foundational role of bacteria in aquatic food webs[74] makes understanding their responses to global change a pressing question[75]. The coincidence of the 2012 shifts in *Nanopelagicaceae* strains with both a species invasion and environmental extremes implicates anthropogenic drivers. Given the global abundance of *Nanopelagicaceae*[37], changes in its strain composition may have wide-ranging impacts on freshwater ecosystems, and organic nitrogen compounds may play a central role in freshwater responses to global change. However, it is ambiguous whether such shifts in strain composition reflect ecological or evolutionary change.

The interface between ecology and evolution is delineated by species boundaries, but in bacteria species definitions are hotly debated[31]. Using a commonly chosen definition for microbial species boundaries, we found interspecific ecological dynamics mirrored intraspecific evolutionary dynamics, with no emergent boundary delineating ecology from evolution. Should interactions such as competition and niche differentiation between strains be considered ecology, or does the fact that they were inferred from observations of genomic change place them in the realm of evolution? Should positive selection of organic nitrogen metabolism genes be considered evolution, or are soft selective sweeps simply evidence of ecological shifts between phenotypically distinct strains? Can we differentiate ecological from evolutionary processes when they occur on the same time scales, in response to the same likely environmental drivers, and across unclear species delineations?

Our two-decade TYMEFLIES dataset, its associated 2,855 species-representative MAGs and decades of North Temperate Lakes Long-Term Ecological Research program (NTL-LTER) environmental data raise these questions again and again. We identified seasonal and decadal strain dynamics that could be considered alternately ecology or evolution across diverse and abundant phyla.

Other microbiome studies have similarly identified microdiversity at the strain level as key to understanding microbial change. Strains have displayed distinct environmental preferences in anaerobic digesters[76], oceans[22,23,26,77,78] and geysers[79], and strain-level dynamics have been linked with outcomes such as cyanobacterial toxicity[80], preterm birth[81], human health[82] and cheese rind aroma[83]. Strains have been described alternately by ecological concepts such as metapopulations in the subseafloor[84] and carrying capacity in the human gut[85], or by evolutionary concepts such as modes of speciation in lakes[24,27]. In pitcher plant microbiomes, strains were ecologically distinct when they differed by only 100 SNVs[86]. Among all these microbiome studies, sometimes strain dynamics are framed as ecology[23,77–80,83,85,86] and sometimes as evolution[22,26,27,76,81,82,84]. However, even in plants and animals speciation is not instantaneous and subspecies population structure creates a blurred line between strains and species[87,88]. Therefore, we propose a shift away from framing eco-evolutionary dynamics around feedbacks between distinct processes[28–30]. To better encompass microbial communities, we should frame change as converging along a continuum of ecology and evolution.

Methods

Lake Mendota samples

Lake Mendota is a eutrophic temperate lake located in Madison, Wisconsin (USA)[89]. Integrated samples were collected from the upper 12 m at a 25 m deep location referred to as the central 'deep hole' (43° 05′58.2″ N, 89° 24′16.2″ W). During the summer stratified months, these 12 m samples span the epilimnion layer. Bacteria were collected on 0.2 µm polyethersulfone filters (Pall Corporation), stored at −80 °C, and DNA was extracted by a single person after randomizing sample order in 2018–2019 using FastDNA Spin kits (MP Biomedicals). A detailed description of the study site, sample collection and DNA extraction procedures is provided by Rohwer and McMahon[4].

Metagenome sequencing and assembly

Sample DNA was sequenced by the US Department of Energy Joint Genome Institute (JGI) using a paired library layout with a NovaSeq 6000 Sequencing System and an S4 flow cell (Illumina Inc.). Samples were sequenced to a depth of 80 ± 20 million reads and 23 ± 6 billion bases per sample. Sample metadata are available in Supplementary Data 1 and raw sequencing data are available from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Umbrella Project accession PRJNA1056043. Individual metagenome SRA accession numbers are listed in Supplementary Data 1. Read filtering was performed using standard JGI protocols[90] (IMG Pipeline version 5, minor releases listed in Supplementary Data 1), which are additionally detailed as metadata paired with each sample through the JGI IMG/M website. Briefly, BBDuk[91] was used to remove adaptors and quality trim reads, and BBMap[91] was used to identify and remove common contaminants. In our analyses we treated the resulting filtered fastq files as the metagenome reads. Single-sample assemblies were also generated by JGI with their standard protocol[90] (IMG Pipeline version 5, minor releases listed in Supplementary Data 1) using metaSPAdes[92]. These filtered fastq files and single-sample assemblies are available through the JGI Genome Portal under ITS Proposal ID 504350.

Obtaining and characterizing genomes

Genomes were binned out of metagenomes using the Texas Advanced Computing Center's Lonestar6 supercomputer and the Launcher utility (version 3.7)[93]. Metagenomic reads were mapped back to sample assemblies using BBMap (version 38.22)[91], sorted BAM files were created using SAMtools (version 1.9)[94] and MAGs were binned using MetaBAT2 (version 2.12.1)[95]. Metagenomic reads from different samples were cross-mapped back to each single-sample assembly to perform differential coverage binning. Cross-mapping scales exponentially, so it was performed on assemblies and sample reads broken into approximately 50-sample groups of consecutive sample dates, with samples from the same year grouped together. This resulted in 85,684 genome bins. CheckM2 (version 0.1.3)[32] was used to assess bin quality, including completeness and contamination estimates, and the Genome Taxonomy Database Toolkit (GTDB-Tk) (version 2.1.1)[96] was used to assign GTDB taxonomy (release 207)[97] to all bins. 30,389 genome bins were at least 50% complete and less than 10% contaminated and these bins were dereplicated to 96% ANI using dRep (version 3.4.0)[33]. To choose 96% as our ANI cut-off, we ran dRep at ANIs ranging from 90% to 99% and examined the resulting number of dereplicated bins, as well as the number of bins from the same assembly that were combined. We chose 96% ANI because very few (one) of the 30,389 bins were combined into an ANI group with a bin created from the same assembly, and because 96% ANI was generally located right before a sudden increase in the total number of genome groups. Our goal was to separate as many species as possible while combining strains that were so closely related they would compete for mapped reads. Applying a 96% ANI cut-off with dRep resulted in 2,855 representative genomes, which we treated as species in this study. These MAGs are available from the NCBI SRA under BioProject accession PRJNA1158976 and their associated metadata is detailed in Supplementary Data 2.

To quantify the relative abundance of each species in every sample, we mapped all sample reads against the concatenated 96% ANI reference genomes using BBMap (version 38.22)[91], created sorted BAM files using SAMtools (version 1.9)[94] and calculated relative abundance using coverM (version 0.6.1)[98]. With the coverM software, we required a minimum read percent identity of 93, proper pairs only, and excluded 1,000 bp from each contig end from the calculation. CoverM calculates relative abundance as the mean coverage divided by the mean coverage across all genomes multiplied by the proportion of reads that mapped to the genome, thus normalizing by recovered genome size to estimate the fraction of cells that belong to a given species in each sample. A table of representative MAGs along with taxonomy annotations, quality statistics and abundance statistics is available as Supplementary Data 2.

To further characterize the genomes, we ran inStrain (version 1.7.1)[21] using a minimum read ANI of 93%, as recommended by the inStrain documentation given our previous choice of 96% ANI to dereplicate genomes. This software called SNVs and calculated nucleotide diversity, among other metrics. To identify genes, we ran prodigal (version 2.6.3)[99] on each genome separately. We then used Kofamscan (version 1.3.0)[100] to assign gene annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (release 107.1)[63]. Additional custom analyses were performed using the R programming language (version 4.1.2)[101] and relied extensively on the data.table R package (version 1.14.8)[102], the lubridate R package (version 1.9.3)[103] and GNU parallel (version 'Chandrayaan')[104].

Classifying seasonal and long-term change

To classify each species' abundance pattern as seasonal or not, we started with relative abundances as calculated by coverM (version 0.6.1)[98] and further corrected any abundance to zero if the genome's coverage breadth was 70% or less than its expected breadth, as calculated by inStrain (version 1.7.1)[21]. We then applied a centered log ratio transformation to the relative abundance values using the compositions R package (version 2.0–6)[105]. After taking a daily linear interpolation to obtain evenly spaced samples, we detrended the temporal profiles with a cubic fit. Finally, we performed a periodogram analysis by computing the magnitude of the fast Fourier transform. If a peak occurred within 30 days of 365 days, we considered it an annual oscillation, and if any of the top five peaks corresponded to an annual period, we classified the species as having a seasonal abundance pattern. We applied this analysis only to the 1,474 species that occurred on least 30 dates over at least 10 years. To classify each species' nucleotide diversity pattern as seasonal or not, we similarly performed a fast Fourier transform on its inStrain-calculated nucleotide diversity over time.

We used the same periodogram analysis to classify it as having seasonal nucleotide diversity or not, and we applied this analysis to the same subset of 1,474 species.

To characterize blooms as more diverse or less diverse, we calculated the Pearson correlation between centered log ratio-transformed relative abundance and nucleotide diversity for the 365 species that had both seasonal abundance and seasonal nucleotide diversity annual oscillations. We considered it a positive correlation (more diverse blooms) if the Pearson correlation was at least 0.35 and a negative correlation (less diverse blooms) if the Pearson correlation was less than or equal to −0.35. We repeated this analysis with up to 2 weeks of lag and used the highest correlation within that window. We chose 0.35 as a reasonable cut-off after manual examination of the first 150 species' correlations.

To calculate SNV profiles for each species, we created vectors corresponding to every SNV position in its genome, where the value of each element was the percent of mapped reads that matched the reference genome base at that position in each sample. SNVs were called using inStrain[21], and we only applied this analysis to samples where the species' median coverage was over 10×, as at coverages less than that we observed a drop in the total SNVs called. Therefore, for both long-term and seasonal analysis of SNV profiles, we included only species that had median coverage over 10× on at least 30 dates over at least 10 years, which resulted in a subset of 263 species. To identify changes in SNV profiles, we created a distance matrix for each species based on Euclidean distances between each sample's SNV profile using the vegan R package (version 2.6–4)[106]. From this, we created a table of time elapsed and Euclidean distance between each sample date.

To identify seasonal patterns in each species' SNV profiles, we created a daily linear interpolation of pairwise distances between all samples, taking the mean when multiple sample pairs occurred with the same time interval. After detrending with a cubic fit, we performed a periodogram analysis to identify annual oscillations and the presence of seasonal patterns using the same criteria as with our abundance and nucleotide diversity annual oscillation analysis.

To identify long-term change patterns, we subset our pairwise distance table to the distance of each sample from the first sample. We developed a classifier for these temporal profiles of distances between SNV profiles using 11 manually chosen species. We chose our training set to encompass examples of each pattern of change including no change, and to include both high and low numbers of observations. Our classifier criteria was hierarchical: first gradual change was identified, then step change was identified and finally disturbance/resilience patterns were identified. After training, the classifier was applied to all 263 species above the abundance cut-off. Gradual change was identified if a linear fit to the daily linearly interpolated distances, excluding dates closer than a month to the starting date, resulted in an adjusted R^2 of at least 0.55. Dates closer than a month to the starting date were excluded because they tended to be highly similar, and a linear interpolation was applied to account for uneven sampling dates, particularly the high frequency of summer sampling in the latter decade of the time series. Possible step change locations were identified after excluding dates closer than a month to the starting date and applying an F test to the linearly interpolated distances using the strucchange R package (version 1.5–3)[107]. If a breakpoint was identified by the F test, the means of measured (as opposed to interpolated) before and after distances were different (two-sided Mann–Whitney P value < 0.01), and the step resulted in a new mean at least 33% higher than the previous mean, a step change pattern was identified. Disturbance/resilience patterns were then identified using outlier distances calculated by the default box plot statistics in R. If a date's distance was >1.5 times the difference between the third and first quartile of observed distances, a date was considered an outlier, and if outlier values were maintained for at least a month, the species was classified as having a disturbance event with resilience.

Analysing abrupt change in *Nanopelagicaceae*

To place environmental conditions in 2012 in context, historical environmental data was collected from the NTL-LTER through the Environmental Data Initiative (EDI) (https://edirepository.org/) and the US Geological Survey (USGS) Water Data for the Nation (https://waterdata.usgs.gov/nwis) using the USGS dataRetrieval R package (version 2.7.14)[108]. EDI datasets analysed included ice duration[53]; nutrients, pH, and carbon[109]; major ions[110]; water temperatures combined from multiple datasets[111–115] as described in Rohwer et al.[14]; phytoplankton[116]; and zooplankton[117] converted to biomass as described in Rohwer et al.[55]. River discharge measurements were obtained from the USGS for the Yahara River, the primary tributary into Lake Mendota (site ID: 05427718)[54]. After exploring all parameters included in these datasets, the occurrence of a hot, dry year with low primary productivity became apparent. Lake heatwaves spanning much of 2012 were confirmed using the 90th percentile definition from Woolway et al.[60] and the heatwaveR R package (version 0.4.6)[118].

Relative abundance and nucleotide diversity of the *Nanopelagicus* MAG ME2011-09-21_3300043464_group3_bin69 were calculated as for the seasonal analysis. New SNVs were identified as SNV positions that were called by inStrain[21] for the first time in a given sample. To identify dates where an unusual number of new SNVs appeared, possibly indicating the emergence of a new strain, the new SNV counts were compared across all sample dates. Initially, high numbers of new SNVs are expected, so outlier dates were identified among the remaining samples after excluding the initial consecutive dates where new SNVs remained in the fourth quantile. Genes under selection were identified using the ratio of nonsynonymous to synonymous SNVs in relation to the reference genome (dN/dS) and the ratio of nonsynonymous to synonymous SNVs when at least two alleles were present (pN/pS) as calculated by inStrain[21]. A McDonald–Kreitman test[119] was used to identify positively selected genes where the bias of unfixed SNVs to be non-synonymous was lower than the bias of fixed SNVs to be non-synonymous, that is, when (pN/pS)/(dN/dS) <1, and positive selection was considered statistically significant when the two-sided Fisher P value was less than or equal to 0.05. A gene was considered consistently selected if it appeared under significant positive selection with high frequency (in the fourth quartile). Consistently selected genes were identified for the pre-2012 and post-2012 time periods separately.

Gene annotations were analysed in the context of the KEGG pathways[63] they belonged to. For each potential pathway, all genes present in the genome were visualized with KEGG Pathway Maps (https://www.genome.jp/brite/br08901). When multiple genes that surrounded the selected gene existed in the genome, that pathway was considered a likely annotation. When likely pathways involved amino acid metabolism or aminoacylation, they were considered amino acid related. When likely pathways involved purine or pyrimidine metabolism, they were considered nucleic acid related.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Metagenome and MAG sequences are available from the NCBI SRA under Umbrella Project accession PRJNA1056043. Individual metagenome SRA accession numbers are listed in Supplementary Data 1 and individual MAG SRA accession numbers are listed in Supplementary Data 2. Most MAGs are available under the NCBI BioProject accession PRJNA1158976, but a few, detailed in Supplementary Data 2, are available from the Open Science Framework[120]. The filtered fastq files and single-sample assemblies used in this study are available through the JGI Genome Portal under ITS Proposal ID 504350. Environmental data is publicly available through the EDI (https://edirepository.org/)[53,109–117]

and the US Geological Survey's Water Data for the Nation (https://waterdata.usgs.gov/nwis)[54].

Code availability

Custom scripts used for data processing are available via GitHub at https://github.com/rrohwer/TYMEFLIES_manuscript and via Zenodo at https://doi.org/10.5281/zenodo.10663021 (ref. 121).

References

1. Brennan, G. L. & Logares, R. Tracking contemporary microbial evolution in a changing ocean. *Trends Microbiol.* **31**, 336–345 (2023).
2. Martiny, J. B. H. et al. Investigating the eco-evolutionary response of microbiomes to environmental change. *Ecol. Lett.* **26**, S81–S90 (2023).
3. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J.* **11**, 2181–2194 (2017).
4. Rohwer, R. R. & McMahon, K. D. A two-decade microbial time series from a freshwater lake, introducing the limony and TYMEFLIES datasets. Preprint at https://doi.org/10.1101/2022.08.04.502869 (2024).
5. Kane, M. D. Microbial observatories: exploring and discovering microbial diversity in the 21st century. *Microb. Ecol.* **48**, 447–448 (2004).
6. Magnuson, J. J., Kratz, T. K. & Benson, B. J. (eds). *Long-Term Dynamics of Lakes in the Landscape: Long-Term Ecological Research on North Temperate Lakes* (Oxford Univ. Press, 2006).
7. Magee, M. R., Wu, C. H., Robertson, D. M., Lathrop, R. C. & Hamilton, D. P. Trends and abrupt changes in 104 years of ice cover and water temperature in a dimictic lake in response to air temperature, wind speed, and water clarity drivers. *Hydrol. Earth Syst. Sci.* **20**, 1681–1702 (2016).
8. Kucharik, C. J., Serbin, S. P., Vavrus, S., Hopkins, E. J. & Motew, M. M. Patterns of climate change across wisconsin from 1950 to 2006. *Phys. Geogr.* **31**, 1–28 (2010).
9. Robertson, D. M., Ragotzkie, R. A. & Magnuson, J. J. Lake ice records used to detect historical and future climatic changes. *Clim. Change* **21**, 407–427 (1992).
10. Lathrop, R. C. Perspectives on the eutrophication of the Yahara lakes. *Lake Reserv. Manag.* **23**, 345–365 (2007).
11. Carpenter, S. R., Booth, E. G., Kucharik, C. J. & Lathrop, R. C. Extreme daily loads: role in annual phosphorus input to a north temperate lake. *Aquat. Sci.* **77**, 71–79 (2015).
12. Spear, M. J. et al. Early changes in the benthic community of a eutrophic lake following zebra mussel (*Dreissena polymorpha*) invasion. *Inland Waters* **12**, 311–329 (2022).
13. Walsh, J. R., Carpenter, S. R. & Zanden, M. J. V. Invasive species triggers a massive loss of ecosystem services through a trophic cascade. *Proc. Natl Acad. Sci. USA* **113**, 4081–4085 (2016).
14. Rohwer, R. R., Hale, R. J., Vander Zanden, M. J., Miller, T. R. & McMahon, K. D. Species invasions shift microbial phenology in a two-decade freshwater time series. *Proc. Natl Acad. Sci. USA* **120**, e2211796120 (2023).
15. Beal, M. R. W., O'Reilly, B., Hietpas, K. R. & Block, P. Development of a sub-seasonal cyanobacteria prediction model by leveraging local and global scale predictors. *Harmful Algae* **108**, 102100 (2021).
16. Gilbert, J. A. et al. Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
17. Yeh, Y.-C. & Fuhrman, J. A. Effects of phytoplankton, viral communities, and warming on free-living and particle-associated marine prokaryotic community structure. *Nat. Commun.* **13**, 7905 (2022).
18. Ferrera, I. et al. Seasonal and interannual variability of the free-living and particle-associated bacteria of a coastal microbiome. *Environ. Microbiol. Rep.* **16**, e13299 (2024).
19. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
20. Rossum, T. V., Ferretti, P., Maistrenko, O. M. & Bork, P. Diversity within species: interpreting strains in microbiomes. *Nat. Rev. Microbiol.* https://doi.org/10.1038/s41579-020-0368-1 (2020).
21. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).
22. Delmont, T. O. et al. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *eLife* **8**, e46497 (2019).
23. Ilicic, D., Ionescu, D., Woodhouse, J. & Grossart, H.-P. Temperature-related short-term succession events of bacterial phylotypes in Potter Cove, Antarctica. *Genes* **14**, 1051 (2023).
24. Bendall, M. L. et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* **10**, 1589–1601 (2016).
25. Garcia, S. L. et al. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *ISME J.* **12**, 742–755 (2018).
26. Hwang, Y. & Girguis, P. R. Differentiated evolutionary strategies of genetic diversification in Atlantic and Pacific thaumarchaeal populations. *mSystems* **7**, e01477-21 (2022).
27. Hoetzinger, M. et al. Geographic population structure and distinct intra-population dynamics of globally abundant freshwater bacteria. *ISME J.* **18**, wrae113 (2024).
28. Bassar, R. D., Coulson, T., Travis, J. & Reznick, D. N. Towards a more precise—and accurate—view of eco-evolution. *Ecol. Lett.* **24**, 623–625 (2021).
29. Hendry, A. P. *Eco-Evolutionary Dynamics* (Princeton Univ. Press, 2017).
30. Ware, I. M. et al. Feedbacks link ecosystem ecology and evolution across spatial and temporal scales: empirical evidence and future directions. *Funct. Ecol.* **33**, 31–42 (2019).
31. Rosselló-Móra, R. & Amann, R. Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* **38**, 209–216 (2015).
32. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
33. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
34. Varghese, N. J. et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* **43**, 6761–6771 (2015).
35. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 1–8 (2018).
36. Olm, M. R. et al. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* https://doi.org/10.1128/msystems.00731-19 (2020).
37. Chiriac, M.-C., Haber, M. & Salcher, M. M. Adaptive genetic traits in pelagic freshwater microbes. *Environ. Microbiol.* **25**, 606–641 (2023).
38. Kim, S., Kang, I., Seo, J.-H. & Cho, J.-C. Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical 'helper' catalase. *ISME J.* **13**, 2252–2263 (2019).

39. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* **12**, 185–198 (2018).
40. Magee, M. R. & Wu, C. H. Effects of changing climate on ice cover in three morphometrically different lakes. *Hydrol. Process.* **31**, 308–323 (2017).
41. Magee, M. R. & Wu, C. H. Response of water temperatures and stratification to changing climate in three lakes with different morphometry. *Hydrol. Earth Syst. Sci.* **21**, 6253–6274 (2017).
42. Snortheim, C. A. et al. Meteorological drivers of hypolimnetic anoxia in a eutrophic, north temperate lake. *Ecol. Modell.* **343**, 39–53 (2017).
43. Matsuzaki, S.-I. S. et al. Climate and food web effects on the spring clear-water phase in two north-temperate eutrophic lakes. *Limnol. Oceanogr.* **66**, 30–46 (2020).
44. Zhou, Z. et al. A 20-year time-series of a freshwater lake reveals seasonal dynamics and environmental drivers of viral diversity, ecology, and evolution. *Nat. Microbiol.* https://doi.org/10.1038/s41564-024-01876-7 (2025)
45. Krinos, A. I. et al. Time-series metagenomics reveals changing protistan ecology of a temperate dimictic lake. *Microbiome* **12**, 133 (2024).
46. Okazaki, Y., Nakano, S., Toyoda, A. & Tamaki, H. Long-read-resolved, ecosystem-wide exploration of nucleotide and structural microdiversity of lake bacterioplankton genomes. *mSystems* **7**, e00433-22 (2022).
47. Meziti, A. et al. Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. *ISME J.* **13**, 767–779 (2019).
48. Schmidt, M. L. et al. Microhabitats are associated with diversity–productivity relationships in freshwater bacterial communities. *FEMS Microbiol. Ecol.* **96**, fiaa029 (2020).
49. Magnuson, J. J. Long-term ecological research and the invisible present. *BioScience* **40**, 495–501 (1990).
50. Turner, M. G. et al. Climate change, ecosystems and abrupt change: science priorities. *Philos. Trans. R. Soc. B* **375**, 20190105 (2020).
51. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
52. Denef, V. J. & Banfield, J. F. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science* **336**, 462–466 (2012).
53. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. North temperate lakes LTER: ice duration—Madison lakes area 1853–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/69B3391E13955392587413ECBFC7C298 (2023).
54. *USGS 05427718 Yahara River at Windsor, WI, Daily Discharge 00060* (US Geological Survey National Water Information System, 2023).
55. Rohwer, R. R. et al. Increased anoxia following species invasion of a eutrophic lake. *Limnol. Oceanogr. Lett.* **9**, 33–42 (2024).
56. Bailey, J. & Hood, J. M. Biotic and thermal drivers alter zooplankton phenology in western Lake Erie. *Limnol. Oceanogr. Lett.* **9**, 219–228 (2024).
57. Carpenter, S. R., Booth, E. G. & Kucharik, C. J. Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnol. Oceanogr.* **63**, 1221–1233 (2018).
58. Carey, C. C., Hanson, P. C., Lathrop, R. C., Amand, S. & L, A. Using wavelet analyses to examine variability in phytoplankton seasonal succession and annual periodicity. *J. Plankton Res.* **38**, 27–40 (2016).
59. Berg, S. M., Peterson, B. D., McMahon, K. D. & Remucal, C. K. Spatial and temporal variability of dissolved organic matter molecular composition in a stratified eutrophic lake. *J. Geophys. Res. Biogeosci.* **127**, e2021JG006550 (2022).
60. Woolway, R. I. et al. Lake heatwaves under climate change. *Nature* **589**, 402–407 (2021).
61. Walsh, J. R., Munoz, S. E. & Vander Zanden, M. J. Outbreak of an undetected invasive species triggered by a climate anomaly. *Ecosphere*https://doi.org/10.1002/ecs2.1628 (2016).
62. Rastetter, E. B. et al. Time lags: insights from the US Long Term Ecological Research Network. *Ecosphere* **12**, e03431 (2021).
63. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
64. Ramoneda, J., Jensen, T. B. N., Price, M. N., Casamayor, E. O. & Fierer, N. Taxonomic and environmental distribution of bacterial amino acid auxotrophies. *Nat. Commun.* **14**, 7608 (2023).
65. Castelle, C. J. et al. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* **16**, 629–645 (2018).
66. Garcia, S. L. et al. Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. *Mol Ecol* **24**, 4449–4459 (2015).
67. Garcia, S. L. et al. Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J.* **7**, 137–147 (2013).
68. Hamilton, J. J. et al. Metabolic network analysis and metatranscriptomics reveal auxotrophies and nutrient sources of the cosmopolitan freshwater microbial lineage acI. *MSystems* **2**, e00091-17 (2017).
69. Drake, T. W., Raymond, P. A. & Spencer, R. G. M. Terrestrial carbon inputs to inland waters: a current synthesis of estimates and uncertainty. *Limnol. Oceanogr. Lett.* **3**, 132–142 (2018).
70. Grant, L. et al. Attribution of global lake systems change to anthropogenic forcing. *Nat. Geosci.* **14**, 849–854 (2021).
71. Roy, H. E. et al. IPBES invasive alien species assessment: summary for policymakers. *Zenodo* https://zenodo.org/record/8314303 (2023).
72. Zwart, J. A. et al. Cross-scale interactions dictate regional lake carbon flux and productivity response to future climate. *Geophys. Res. Lett.* https://doi.org/10.1029/2019GL083478 (2019).
73. Finlay, K. et al. Decrease in CO_2 efflux from northern hardwater lakes with increasing atmospheric warming. *Nature* **519**, 215–218 (2015).
74. Azam, F. et al. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
75. Tiedje, J. M. et al. Microbes and climate change: a research prospectus for the future. *mBio* https://doi.org/10.1128/mbio.00800-22 (2022).
76. Ghiotto, G., Zampieri, G., Campanaro, S. & Treu, L. Strain-resolved metagenomics approaches applied to biogas upgrading. *Environ. Res.* **240**, 117414 (2024).
77. Koch, H. et al. Genomic, metabolic and phenotypic variability shapes ecological differentiation and intraspecies interactions of *Alteromonas macleodii*. *Sci. Rep.* **10**, 809 (2020).
78. Ustick, L. J., Larkin, A. A. & Martiny, A. C. Global scale phylogeography of functional traits and microdiversity in Prochlorococcus. *ISME J.* **17**, 1671–1679 (2023).
79. Keller, L. M., Colman, D. R. & Boyd, E. S. An active microbiome in Old Faithful geyser. *Proc. Natl Acad. Sci. USA Nexus* **2**, pgad066 (2023).
80. Yancey, C. E. et al. Metagenomic and metatranscriptomic insights into population diversity of microcystis blooms: spatial and temporal dynamics of mcy genotypes, including a partial operon that can be abundant and expressed. *Appl. Environ. Microbiol.* **88**, e02464-21 (2022).
81. Liao, J. et al. Microdiversity of the vaginal microbiome is associated with preterm birth. *Nat. Commun.* **14**, 4997 (2023).

82. Ma, C. et al. Cross-cohort single-nucleotide-variant profiling of gut microbiota suggests a novel gut-health assessment approach. *mSystems* **8**, e00828-23 (2023).
83. Niccum, B. A., Kastman, E. K., Kfoury, N., Robbat, A. & Wolfe, B. E. Strain-level diversity impacts cheese rind microbiome assembly and function. *mSystems* https://doi.org/10.1128/msystems.00149-20 (2020).
84. Anderson, R. E., Graham, E. D., Huber, J. A. & Tully, B. J. Microbial populations are shaped by dispersal and recombination in a low biomass subseafloor habitat. *mBio* **13**, e00354-22 (2022).
85. Wolff, R., Shoemaker, W. & Garud, N. Ecological stability emerges at the level of strains in the human gut microbiome. *mBio* **14**, e02502–e02522 (2023).
86. Goyal, A., Bittleston, L. S., Leventhal, G. E., Lu, L. & Cordero, O. X. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. *eLife* **11**, e74987 (2022).
87. Sukumaran, J. & Knowles, L. L. Multispecies coalescent delimits structure, not species. *Proc. Natl Acad. Sci. USA* **114**, 1607–1612 (2017).
88. Coates, D. J., Byrne, M. & Moritz, C. Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Front. Ecol. Evol.* **6** (2018).
89. Brock, T. D. *A Eutrophic Lake: Lake Mendota, Wisconsin* (Springer, 1985); https://doi.org/10.1007/978-1-4419-8700-6_7
90. Clum, A. et al. DOE JGI metagenome workflow. *mSystems* https://doi.org/10.1128/msystems.00804-20 (2021).
91. Bushnell, B. BBMap short read aligner and other bioinformatic tools. *OSTI.GOV* https://www.osti.gov/biblio/1241166 (2014).
92. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
93. Wilson, L. A. et al. Launcher: a simple tool for executing high throughput computing workloads. *J. Open Source Softw.* **2**, 289 (2017).
94. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
95. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
96. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btz848 (2019).
97. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2022).
98. Aroney, S. T. N. et al. CoverM: read coverage calculator for metagenomics. *Zenodo* https://doi.org/10.5281/zenodo.10531254 (2024).
99. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
100. Aramaki, T. et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
101. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
102. Barrett, T. et al. Data.table: extension of 'data.Frame'. R package version 1.16.99 (2024); https://Rdatatable.gitlab.io/data.table
103. Grolemund, G. & Wickham, H. Dates and times made easy with lubridate. *J. Stat. Softw.* **40**, 1–25 (2011).
104. Tange, O. GNU Parallel 20230822 ('Chandrayaan'). *Zenodo* https://doi.org/10.5281/zenodo.8278274 (2023).
105. van den Boogaart, K. G., Tolosana-Delgado, R. & Bren, M. compositions: compositional data analysis (v.2.0-6). (2023).
106. Oksanen, J. et al. vegan: community ecology package (v.2.6-4). (2022).
107. Zeileis, A., Leisch, F., Hornik, K. & Kleiber, C. strucchange: An R Package for Testing for Structural Change in Linear Regression Models. *J. Stat. Soft.* **7**, 1–38 (2002).
108. DeCicco, L., Hirsch, R., Lorenz, D., Watkins, D. & Johnson, M. dataRetrieval: R packages for discovering and retrieving water data available from US Federal Hydrologic web services (v.2.7.14). https://doi.org/10.5066/P9X4L3GE (2023).
109. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. North temperate lakes LTER: chemical limnology of primary study lakes: nutrients, pH and carbon 1981–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/325232E6E4CD1CE04025FA5674F7B782 (2023).
110. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. North temperate lakes LTER: chemical limnology of primary study lakes: major ions 1981–current. *Environmental Data Initiative* https://doi.org/10.6073/pasta/bb563f16c7338fdb3ddf82057ef43cc6 (2023).
111. Rohwer, R. R. & McMahon, K. D. Lake Mendota microbial observatory temperature, dissolved oxygen, pH, and conductivity data, 2006–present. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/7E533C197ED8EBD27777A89A2C8D7DFE (2022).
112. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. North temperate lakes LTER: physical limnology of primary study lakes 1981–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/316203040EA1B8ECE89673985AB431B7 (2021).
113. Magnuson, J., Carpenter, S. & Stanley, E. North temperate lakes LTER: high frequency water temperature data—Lake Mendota Buoy 2006–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/8CEFF296AD68FA8DA6787076E0A5D992 (2020).
114. Robertson, D. Lake Mendota water temperature secchi depth snow depth ice thickness and meterological conditions 1894–2007. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/F20F9A644BD12E4B80CB288F1812C935 (2016).
115. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. Lake Mendota multiparameter sonde profiles: 2017–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/5F15BF453851987FC030B2F07A110B21 (2021).
116. Magnuson, J. J., Carpenter, S. R. & Stanley, E. H. North Temperate Lakes LTER: phytoplankton—Madison Lakes Area 1995–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/43D3D401AF88CC05C6595962BDB1AB5C (2022).
117. Magnuson, J., Carpenter, S. & Stanley, E. North Temperate Lakes LTER: zooplankton—Madison Lakes Area 1997–current. *Environmental Data Initiative* https://doi.org/10.6073/PASTA/D5ABE9009D7F6AA87D1FCF49C8C7F8C8 (2022).
118. Schlegel, R. W. & Smit, A. J. heatwaveR: a central algorithm for the detection of heatwaves and cold-spells. *J. Open Source Softw.* **3**, 821 (2018).
119. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
120. Rohwer, R. TYMEFLIES Robinning MAGs not included in NCBI SRA BioProject PRJNA1158976. *OSF* https://doi.org/10.17605/OSF.IO/TX2JG (2024).
121. Rohwer, R. R. rrohwer/TYMEFLIES_manuscript: v1.0. *Zenodo* https://doi.org/10.5281/zenodo.10663021 (2024).

Acknowledgements

Long-term datasets such as TYMEFLIES rely on researchers who contribute a portion of their time and effort to future projects they may not be involved in. This work would not be possible without the

generosity of many, including Lake Mendota sampling leads A. Kent, T. Yannarell, A. Shade, S. Jones, R. Newton, G. Wolfe, T. Miller, E. K. Read, L. Beversdorf, J. Mutschler and the original Microbial Observatory lead E. W. Triplett. We thank S. Stevens for her early input into the ideas pursued here, P. Golightly for advice on genes under selection data, T. Butts for advice on environmental data, and W. Ratcliff and V. Denef for advice on framing. R.R.R. acknowledges support from the E. Michael and Winona Foster Wisconsin Alumni Research Foundation (WARF) Wisconsin Idea Fellowship, the US National Science Foundation (NSF) (DBI-2011002), and the Texas Advanced Computing Center at The University of Texas at Austin that provided high performance computing resources that contributed to the research results reported within this paper (http://www.tacc.utexas.edu). M. Kirk acknowledges support from the US National Institutes of Health (NIH) (R01-GM116853) and the US NSF (DEB-1831730). M. Kell acknowledges that the work (proposal: https://doi.org/10.46936/10.25585/60001198) conducted by the US Department of Energy (DOE) Joint Genome Institute (JGI) (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the US DOE operated under contract no. DE-AC02-05CH11231. K.D.M. acknowledges support from the US DOE JGI (CSP 504350), the US Department of Agriculture (USDA) (WIS01516 and WIS01789), the US NSF (DEB-0702395, DEB-1344254) and the US NSF Microbial Observatory program (MCB-9977903, DEB-0702395). B.J.B. acknowledges support from the Simons Foundation Investigator in Aquatic Microbial Ecology Award (LI-SIAME-00002001). This work would not be possible without the long-term support of the US NSF NTL-LTER (DEB-9632853, DEB-0217533, DEB-0822700, DEB-1440297 and DEB-2025982).

Author contributions

R.R.R. and K.D.M. conceptualized the research and obtained initial funding. K.D.M. and B.J.B. provided resources. R.R.R. conducted field and laboratory work and curated data. R.R.R. performed analyses and created visualizations. M. Kirk. advised statistical approaches. S.L.G., M. Kell., K.D.M. and B.J.B. advised analysis approaches. R.R.R. wrote the first draft, and R.R.R., K.D.M. and B.J.B. wrote the final draft incorporating edits provided by M. Kirk., S.L.G. and M. Kell.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-024-01888-3.

Correspondence and requests for materials should be addressed to Robin R. Rohwer, Katherine D. McMahon or Brett J. Baker.

Peer review information *Nature Microbiology* thanks Timothy Ghaly, David Pearce and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Robin Rohwer

Last updated by author(s): Nov 3, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected by hand from the lake, no software was used in data collection.
Data analysis	Launcher utility (version 3.7), IMG Pipeline (version 5), BBMap (version 38.22), SAMtools (version 1.9), MetaBAT2 (version 2.12.1), CheckM2 (version 0.1.3), GTDB-tk (version 2.1.1), dRep (version 3.4.0), coverM (version 0.6.1), inStrain (version 1.7.1), prodigal (version 2.6.3), Kofamscan (version 1.3.0), R programming language (version 4.1.2), data.table R package (version 1.14.8), lubridate R package (version 1.9.3), GNU parallel (version 'Chandrayaan'), compositions R package (version 2.0-6), vegan R package (version 2.6-4), strucchange R package (version 1.5-3), USGS dataRetrieval R package (version 2.7.14), heatwaveR R package (version 0.4.6), and custom analysis scripts available at https://github.com/rrohwer/TYMEFLIES_manuscript

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Metagenome and MAG sequences are available from the NCBI Sequence Read Archive (SRA) under Umbrella Project accession PRJNA1056043. Individual metagenome SRA accession numbers are listed in Supplementary Data 1, and individual MAG SRA accession numbers are listed in Supplementary Data 2. The filtered fastq files and single-sample assemblies used in this study are available through the JGI Genome Portal under ITS Proposal ID 504350. Environmental data is publicly available through the Environmental Data Initiative (https://edirepository.org/)41,101–109 and the U.S. Geological Survey's Water Data for the Nation (https://waterdata.usgs.gov/nwis)42.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this observational study, water samples were collected from Lake Mendota from 2000 to 2019.
Research sample	Lake water was collected from the approximate epilimnion layer using a 12 m sample tube. Samples were collected at the central, deepest place on Lake Mendota (WI, USA). Bacteria were collected on 0.2 um filters, bulk DNA was extracted, and shotgun sequencing on an Illumina instrument was used to generate metagenome samples.
Sampling strategy	Samples were taken over 20 years from the same location. Sample frequency was determined largely by funding and feasibility. Samples were taken approximately twice a month for the first 10 years, and 1-2 times a week for the second 10 years.
Data collection	Water samples were filtered and filters were placed in a -80 C freezer until extraction and sequencing in 2018-2020. Sample metadata was recorded in lab and field notebooks.
Timing and spatial scale	Samples were taken over 20 years from the same location. Sample frequency was determined largely by funding and feasibility. Samples were taken approximately twice a month for the first 10 years, and 1-2 times a week for the second 10 years.
Data exclusions	Some analyses required only 1 sample from each date. In these 6 instances, the sample with the most standard collection was chosen to represent that date. These sample choices are indicated in the "Rohwer.Unduplicated.Dates.Sample.Choice" column of supplemental Data 1.
Reproducibility	All analysis scripts were sourced and ran without errors. Scripts are available and software versions are reported.

Randomization	Samples were grouped by season and by year for various analyses. The season was determined by either month, or using season definitions defined from environmental variables in Rohwer et al 2023, PNAS.
Blinding	Sample order was randomized and samples were blinded for DNA extraction and sequencing.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	This observational study spanned 20 years and all seasons in south-central Wisconsin.
Location	Lake Mendota (Madison, WI, USA), central deep hole (43°05'58.2"N 89°24'16.2"W)
Access & import/export	No permits were required for the publicly accessible lake.
Disturbance	No significant disturbance was created as this lake frequently hosts recreational boat/ice traffic and collecting water samples is a similar activity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA