Bacterial ecology and evolution converge on seasonal and decadal scales

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‡]

[1]The University of Texas at Austin, Department of Integrative Biology, 2415 Speedway #C0930, Austin, TX 78712, USA
[2]Carl von Ossietzky Universität Oldenburg, Institute for Chemistry and Biology of the Marine Environment (ICBM), 26129 Oldenburg, Germany
[3]Stockholm University, Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, 106 91 Stockholm, Sweden
[4]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[5]University of Wisconsin-Madison, Department of Bacteriology, 1550 Linden Drive, Madison, WI, 53706, USA
[6]University of Wisconsin-Madison, Department of Civil and Environmental Engineering, 1550 Linden Drive, Madison, WI, 53706, USA
[7]The University of Texas at Austin, Department of Marine Science, Port Aransas, Austin, TX 78373, USA

Correspondence: *Robin.Rohwer@gmail.com, †trina.mcmahon@wisc.edu, and ‡brett_baker@utexas.edu

Abstract

Ecology and evolution are often viewed as distinct processes, which interact on contemporary time scales in microbiomes. To observe these processes in a natural system, we collected a two-decade, 471-sample freshwater lake time series, creating the longest metagenome dataset to date. Among 2,855 species-representative genomes, diverse species and strains followed cyclical seasonal patterns, and one in five species experienced decadal shifts in strain composition. The most globally abundant freshwater bacterium had constant species-level abundance, but environmental extremes appeared to trigger a shift 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 processes that interact, we propose a conceptualization where ecology and evolution converge along a continuum to better describe change in diverse microbial communities.

Main Text

Microbial communities allow us to observe eco-evolutionary dynamics in real-time due to the short lifespans and large population sizes of microbes. Real-time evolution was famously observed in the *E. coli* long-term evolution experiment[1], but no similar 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, which allows us to directly observe ecology and contemporary evolution in a natural ecosystem. By reconstructing tens of thousands of metagenome-assembled genomes (MAGs), we found that ecology and evolution both unfold at short, seasonal time scales as well as longer-term decadal time scales. In genomes from the most abundant freshwater bacterium, *Nanopelagicaceae*, evolutionary change coincided with environmental extremes. While species-level abundance remained constant, strain composition shifted coincident with an increase in genes under positive selection. Research on eco-evolutionary dynamics focuses on feedbacks between distinct processes of ecology and evolution[2–4]. In our microbial data, however, these processes were difficult to distinguish; ecological dynamics appeared to occur between strains, but the strains themselves were inferred from observations of genomic change. Consistent with the ambiguity of the microbial species concept[5], our observations suggest that it is not possible to cleanly delineate between ecological and evolutionary processes in microbial communities. Therefore, we propose an adjusted conceptualization, where ecology and evolution converge along a continuum.

The TYMEFLIES dataset

We collected 471 samples over 20 years from Lake Mendota (WI, USA)[6] and obtained shotgun DNA libraries (**Fig. 1A**, 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)[7]. We clustered these 30,389 bins at 96% average nucleotide identity (ANI) and obtained 2,855 clusters from which we chose representative MAGs[8] (Supplementary Data 2). Several previous studies have found an emergent species boundary at similar ANI cutoffs[9–11], and we observed a rapid increase in the number of clusters above the 96% ANI cutoff. In this study, we treat the representative MAGs from each 96% ANI cluster as bacterial species and refer to sub-species delineations identified in the mapped metagenomic reads as strains[12].

The representative MAGs have high estimated completeness (median 86%) and low contamination (median 0.9%) (**Fig. 1B**), 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 timeseries[6] 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[13], difficult to culture[14], and typically has highly streamlined genomes[15].

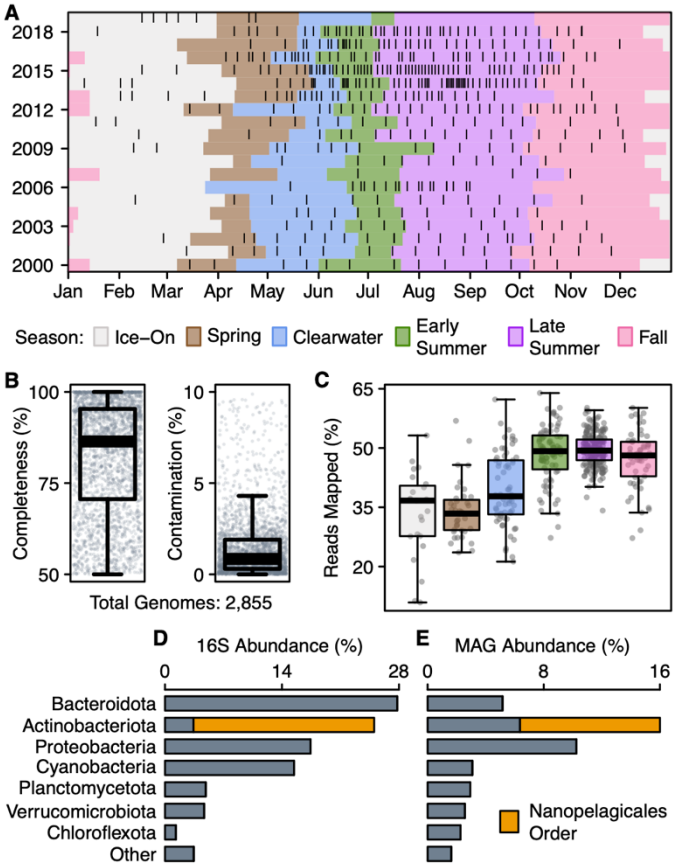

Fig. 1. The TYMEFLIES dataset. A) Metagenome sample dates are indicated by black vertical lines. Microbial seasons[22] are indicated by colored shading. **B)** Quality of the 2,855 representative genomes obtained after clustering to 96% ANI. We treat these genomes as species. **C)** Percent of metagenome reads from each sample that mapped to all reference genomes with an ANI ≥ 93%. Samples are grouped by season to highlight how well the reference genomes reflect each seasonal community. **D)** Rank abundance of phyla as measured by 16S rRNA gene amplicon sequencing[6]. The abundant *Nanopelagicales* order of Actinobacteria is highlighted. **E)** 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.

Seasonal ecology and evolution

Lake Mendota has been the focus of limnological research since the late 1800s and has been routinely sampled since 1984 by the North Temperate Lakes Long-Term Ecological Research program (NTL-LTER)[16]. Microbial sampling began in 2000 as part of an NSF microbial observatory[17]. From this long history of research, we know the lake follows a consistent annual phenology, and that phenological patterns are changing in response to climate change and invasive species[18–21]. Rohwer *et. al*[22] found that these phenological dynamics extend to the bacterial community. To confirm that phenological abundance patterns also exist in our more 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 (*i.e.* 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 single nucleotide variants (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 > 10x), 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 timescales[23,24]. Phenological patterns in sub-species 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 sub-species diversity, we asked if they were correlated. We quantified whether a species' "bloom" in abundance consisted of fewer strains or more strains than its baseline composition. Of the 365 species with seasonal patterns in both abundance and nucleotide diversity (purple bars in **Fig. 2A**), we found both scenarios were common; 21% of these species had less diverse blooms (**Fig. 2C** and yellow bars in **Fig. 2E**), while 19% had more diverse blooms (**Fig. 2D** and green bars in **Fig. 2E**). 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[25] 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[26]. The prevalence of both bloom diversity patterns suggests ecological processes drive changes in allele frequencies.

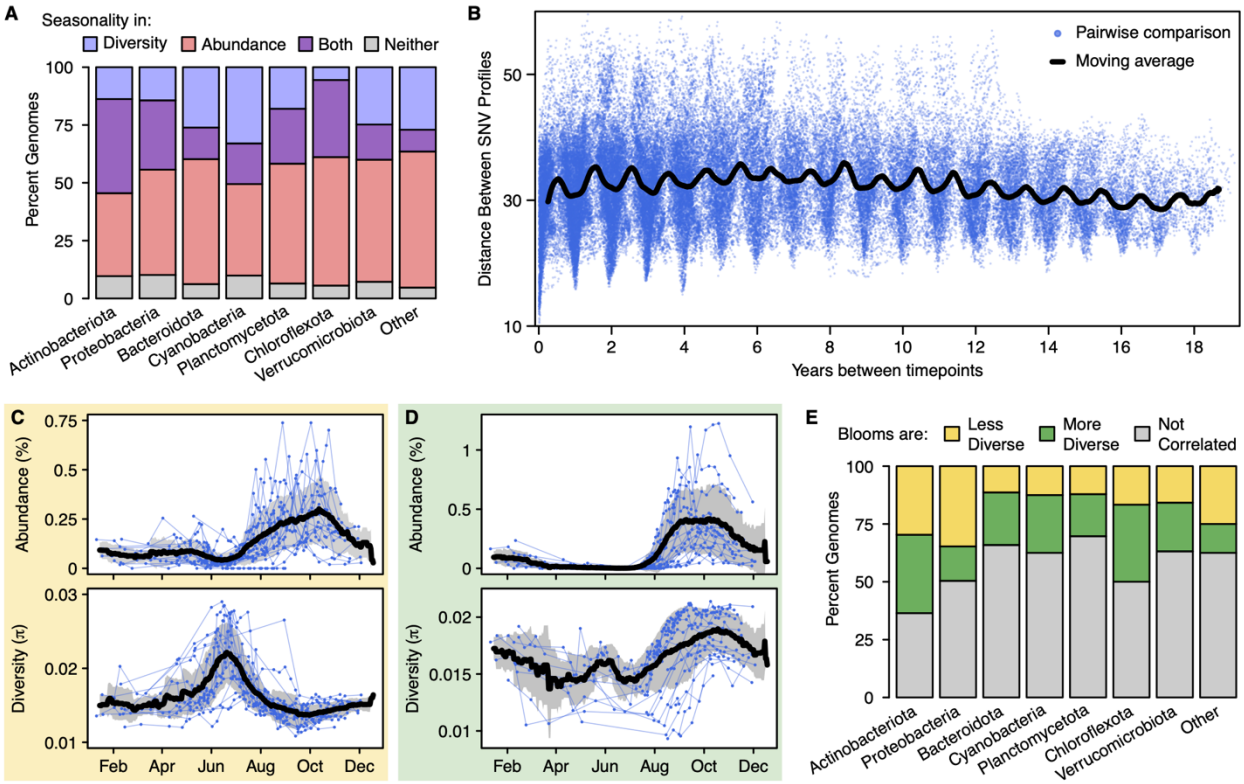

Fig. 2. Bacterial seasonality at the sub-species level. A) The percent of species with seasonality in nucleotide diversity and abundance (a centered log ratio transform was applied to relative abundances). The 1,474 reference species that occurred at least 30 times were included in this analysis. **B)** 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 abundance increases. Displayed is an abundant species (ME2012-08-31_3300044613_group4_bin150) in the *Nanopelagicaceae* family, MAG-120802 genus. **E)** The distribution of bloom diversity patterns across the 365 species that had seasonality in both abundance and nucleotide diversity.

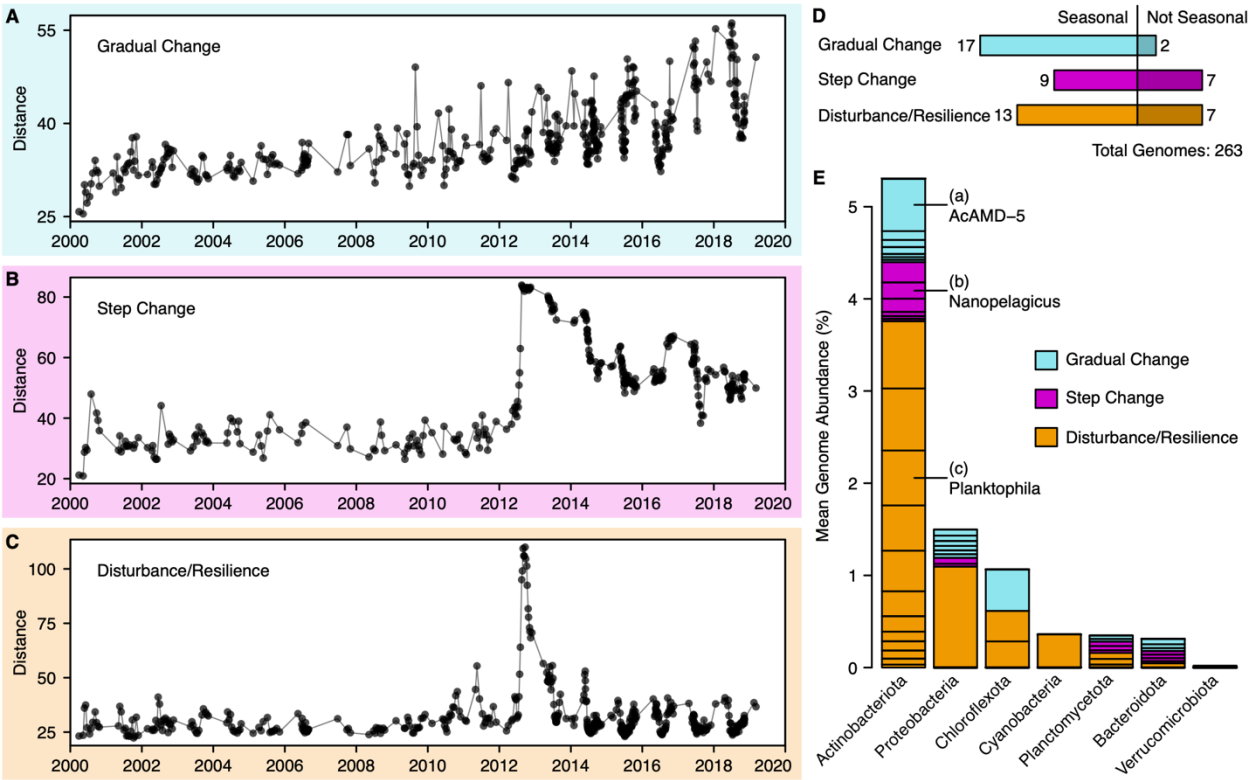

Fig. 3. Long-term changes in strain composition. A) An example of long-term, gradual change in strain composition. Points indicate sample dates, and distance refers to the Euclidean distance between a species' SNV profile on that sample date and its first occurrence in the time series. A species in the *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.

Long-term ecology and evolution

Long-term changes can be masked by seasonal oscillations, lost in what is referred to as the "invisible present"[27]. 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 timeseries. 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** and **C**), which may arise in response to a new stable state after a tipping point, or from a sudden environmental shift[28,29]. 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 vs. 19 species), and resilience was only slightly more common than a lasting step change (20 vs. 16 species) (**Fig. 3D**). The three long-term change patterns were found in many abundant species distributed across phyla (**Fig. 3E**). Many species 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[30]. 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[13], and the 127 *Nanopelagicaceae* species we recovered accounted together for 8% relative abundance on average. Five of these *Nanopelagicaceae* species displayed resilience to the abrupt change, while two experienced lasting step changes in strain composition.

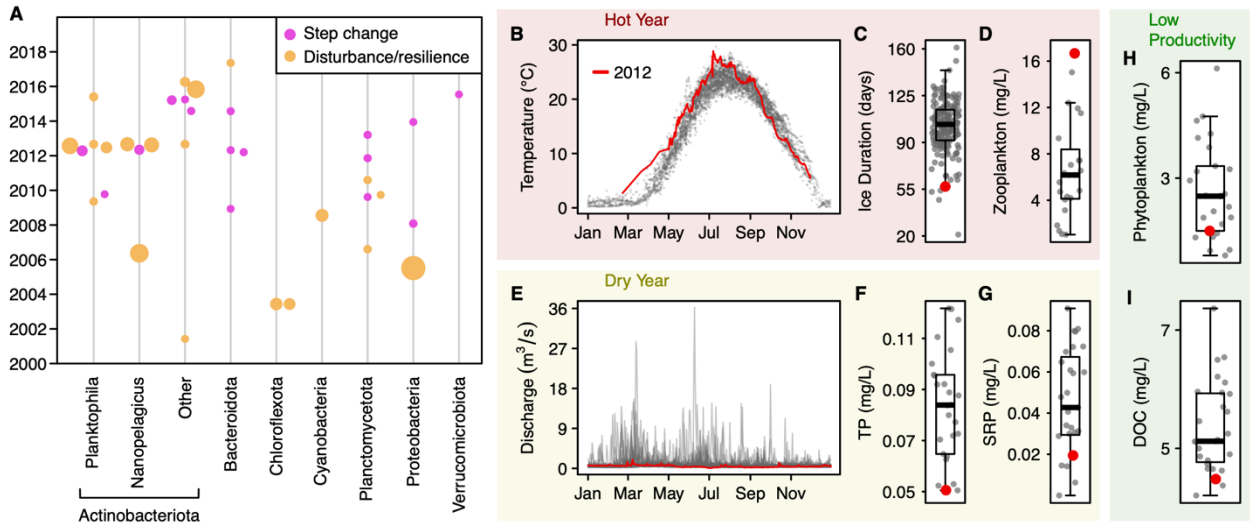

Fig. 4. Abrupt changes in *Nanopelagicaceae* strain composition coincide with environmental extremes in 2012. A) Dates of all abrupt changes in strain composition arranged by phyla. Most changes were isolated events, but multiple species from two abundant genera of Actinobacteriota, *Planktophila* and *Nanopelagicus*, experienced abrupt change in 2012. 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)** Total zooplankton biomass (excluding predatory *Bythotrophes* and *Leptodora*) was unusually high, likely 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 runoff events typical after storms and spring snowmelt (relative to 1989 - 2021). **F)** Total phosphorus, and **G)** soluble reactive phosphorus were low (relative to 1995 – 2021), likely due to low sediment transport. **H)** Low phytoplankton biomass (relative to 1995 – 2020), likely resulting from both high zooplankton grazing and low nutrient availability. **I)** Low dissolved organic carbon (relative to 1996 – 2022), likely a result of low phytoplankton abundance.

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[22] (**Fig. 4B**), the fifth shortest winter ice duration on record since 1856[31] (**Fig. 4C**), the eighth lowest annual discharge from its major tributary on record since 1976 and the second lowest peak discharge[32] (**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[33] (**Fig. 4D**) was likely a result of the mild winter and spring[34] which allowed zooplankton, including the prolific grazer *Daphnia pulicaria,* to establish early. Low total

244 phosphorus and soluble reactive phosphorus (**Fig. 4F-G**) was likely a result of low external nutrient loading associated with mild discharge events[35]. The resulting combination of high
246 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
248 Cyanobacteria during summer[36]. Lake Mendota's dissolved organic carbon (DOC) is primarily provided by phytoplankton[37], consequently DOC was also low in 2012 (**Fig. 4I**). Lake heatwaves
250 are predicted to become hotter and longer with climate change[38], and these observations suggest that the intense epilimnetic heat waves during 2012 had cascading effects on lake biogeochemistry
252 that extended to the level of bacterial strains.

254 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[39]. This major
256 disturbance resulted in a trophic cascade that decreased water clarity[21,40], increased lake anoxia[33], and shifted the bacterial community composition[22]. Although the abrupt changes in strain
258 composition of seven *Nanopelagicaceae* species were not observed until three years later, lag effects are common in complex ecosystems[41]. In contrast to the 2009 species invasion, we did not
260 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
262 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
264 ecological drivers are also drivers of evolutionary change, further emphasizing how ecology and evolution are intertwined.
266

Evolutionary signals in a *Nanopelagicus*

268 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
270 in August 2012 (**Fig. 3B**). An 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
272 adjustment in 2012 and 2013 (**Fig. 5A**).

274 The relative abundance of this species was quite constant throughout our 20-year observation period (**Fig. 5B**), typically with higher abundances during the spring clearwater phase. The step
276 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 with an increase in
278 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
280 large spikes in new SNVs in 2012 (**Fig. 5D**), suggesting that the step change reflects shifts in the relative abundances of existing strains.
282

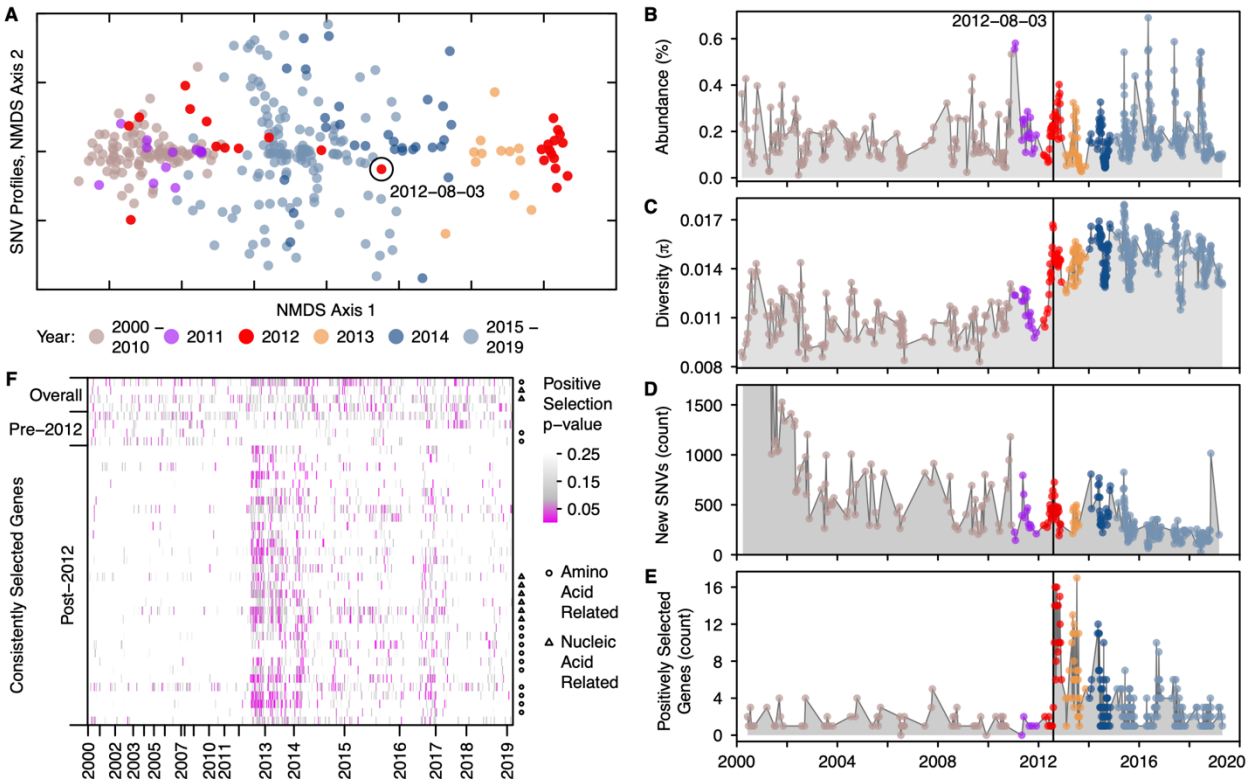

Fig. 5. Step change in strain composition coincides with more genes under selection. A) An abundant *Nanopelagicus* species experienced a step change in strain composition in 2012 (ME2011-09-21_3300043464_group3_bin69, see also **Fig. 3B**). Samples with more similar SNV profiles appear closer on this NMDS plot. Years 2000-2011 cluster together and are distinct from years 2014-2019, which cluster separately. A sudden change in strain composition occurred on August 3, 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 was comprised of 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 F-statistic p-value < 0.05). **F)** Occurrence of consistently selected genes in all the samples, in the pre-2012 period, and in the post-2012 period. X-axis indicates samples over time and Y-axis indicates genes. Shading indicates the significance level of positive selection. 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.

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 timeseries, 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[42] 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[43,44]. In the streamlined *Nanopelagicus*, auxotrophies for various amino acids[15,45] coupled with an enrichment of transporters for many small organic nitrogen compounds, including amino acids[15,46,47] and nucleic acid components[15,45–47] are common. Moreover, the histidine pathway was found split between two different strains of *Nanopelagicus* growing in a mixed culture[45]. 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 (**Fig. 4H**) might indicate lower 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.

Freshwater lakes are focal points on the terrestrial landscape, processing an estimated 70% of net terrestrial carbon production[48]. These ecosystems are stressed by both climate change[49] and invasive species[50], but whether lakes will become net sources or sinks of carbon is uncertain[51,52]. The coincidence of the 2012 evolutionary shifts in *Nanopelagicaceae* with both a species invasion and environmental extremes implicates anthropogenic drivers. Given the foundational role of bacteria in aquatic food webs[53] and the global abundance of *Nanopelagicaceae*[13], its evolution may have wide-ranging impacts on freshwater ecosystems and organic nitrogen compounds may play a central role in freshwater responses to global change.

A continuum of ecology and evolution

The interface between ecology and evolution is delineated by species boundaries, but in bacteria species definitions are hotly debated[5]. 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 like 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 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[54], oceans[55–58], and geysers[59]; and strain-level dynamics have been linked with outcomes such as Cyanobacterial toxicity[60], preterm birth[61], human health[62], and cheese rind aroma[63]. Strains have been described alternately by ecological concepts like metapopulations in the subseafloor[64] and carrying capacity in the human gut[65], or by evolutionary concepts like modes of speciation in global lakes[66]. In pitcher plant microbiomes, strains were ecologically distinct when they differed by only 100 SNVs[67]. Among all these microbiome studies, sometimes strain dynamics are framed as ecology[55–57,59,60,63,65,67] and sometimes as evolution[54,58,61,62,64,66]. However, even in plants and animals speciation is not instantaneous and subspecies population structure creates a blurred line between strains and species[68,69]. Therefore, we propose a shift away from framing eco-evolutionary dynamics around feedbacks between distinct processes[2–4]. 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)[70]. 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

382 study site, sample collection, and DNA extraction procedures is provided by Rohwer and McMahon[6].

384

Metagenome sequencing and assembly

386 Samples were sequenced by the US Department of Energy Joint Genome Institute (JGI) using a NovaSeq 6000 with an S4 flow cell. Sample metadata is available in Supplementary Data 1, and 388 raw sequencing data is available from the NCBI Sequence Read Archive under Umbrella Project accession PRJNA1056043. Individual metagenome SRA accession numbers are listed in 390 Supplementary Data 1. Read filtering was performed using standard JGI protocols[71], which are additionally detailed as metadata paired with each sample through the JGI IMG/M website. Briefly, 392 BBDuk[72] was used to remove adapters and quality trim reads, and BBMap[72] was used to identify and remove common contaminants. In our analyses we treated the resulting filtered fastq files as 394 the metagenome reads. Single-sample assemblies were also generated by JGI with their standard protocol[71] using metaSPAdes[73]. These filtered fastq files and single-sample assemblies are 396 available through the JGI Genome Portal under ITS Proposal ID 504350.

398 **Obtaining and characterizing genomes**

Genomes were binned out of metagenomes using the Texas Advanced Computing Center's 400 Lonestar6 supercomputer and the Launcher utility (version 3.7)[74]. Metagenomic reads were mapped back to sample assemblies using BBMap (version 38.22)[72], sorted BAM files were created 402 using SAMtools (version 1.9)[75], and metagenome-assembled genomes were binned using MetaBAT2 (version 2.12.1)[76]. Metagenomic reads from different samples were cross-mapped back 404 to each assembly. Cross-mapping scales exponentially, so it was performed on assemblies and sample reads broken into approximately 50-sample groups of consecutive sample dates, with 406 samples from the same year grouped together. This resulted in 85,684 genome bins. CheckM2 (version 0.1.3)[7] was used to asses bin quality, including completeness and contamination estimates, 408 and GTDB-tk (version 2.1.1)[77] was used to assign GTDB taxonomy (release 207)[78] to all bins. 30,389 genome bins were at least 50% complete and less than 10% contaminated, and these bins 410 were de-replicated to 96% ANI using dRep (version 3.4.0)[8]. To choose 96% as our ANI cutoff, we ran dRep at ANIs ranging from 90 to 99% and examined the resulting number of de-replicated 412 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 414 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 416 possible, while combining strains that were so closely related they would compete for mapped reads. Applying a 96% ANI cutoff with dRep resulted in 2,855 representative genomes, which we 418 treated as species in this study.

420 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)[72], created

422 sorted BAM files using SAMtools (version 1.9)[75], and calculated relative abundance using coverM
(version 0.6.1)[79]. With the coverM software, we required a minimum read percent identity of 93,
424 proper pairs only, and excluded 1000 bp from each contig end from the calculation. CoverM
calculates relative abundance as the mean coverage divided by the mean coverage across all
426 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
428 sample. A table of representative MAGs along with taxonomy annotations, quality statistics, and
abundance statistics is available as Supplementary Data 2.
430

To further characterize the genomes, we ran inStrain (version 1.7.1)[80] using a minimum read ANI
432 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
434 metrics. To identify genes we ran prodigal (version 2.6.3)[81] on each genome separately. We then
used Kofamscan (version 1.3.0)[82] to assign gene annotations from the Kyoto Encyclopedia of
436 Genes and Genomes (KEGG) database (release 107.1)[42]. Additional custom analyses were
performed using the R programming language (version 4.1.2)[83], and relied extensively on the
438 data.table R package (version 1.14.8)[84], the lubridate R package (version 1.9.3)[85], and GNU
parallel (version 'Chandrayaan')[86].
440

Classifying seasonal and long-term change

442 To classify each species' abundance pattern as seasonal or not, we started with relative abundances
as calculated by coverM (version 0.6.1)[79] and further corrected any abundance to zero if the
444 genome's coverage breadth was 70% or less than its expected breadth, as calculated by inStrain
(version 1.7.1)[80]. We then applied a centered log ratio transformation to the relative abundance
446 values using the compositions R package (version 2.0-6)[87]. After taking a daily linear interpolation
to obtain evenly spaced samples, we detrended the temporal profiles with a cubic fit. Finally, we
448 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
450 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
452 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
454 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.
456

To characterize blooms as more diverse or less diverse, we calculated the Pearson correlation
458 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
460 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 two weeks of lag and used the highest correlation within that window. We chose 0.35 as a reasonable cutoff 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. SNV's were called using inStrain[80], and we only applied this analysis to samples where the species' median coverage was over 10x, 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 medium coverage over 10x 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)[88]. 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 cutoff. 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 timeseries. 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)[89]. 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

502 default boxplot statistics in R. If a date's distance was > 1.5 times the difference between the 3rd
and 1st quartile of observed distances a date was considered an outlier, and if outlier values were
504 maintained for at least a month the species was classified as having a disturbance event with
resilience.
506

Analyzing abrupt change in *Nanopelagicaceae*

508 To place environmental conditions in 2012 in context, historical environmental data was collected
from the North Temperate Lakes Long-Term Ecological Research program (NTL-LTER) through
510 the Environmental Data Initiative (EDI) interface (https://edirepository.org/) and the US
Geological Survey (USGS) Water Data for the Nation (https://waterdata.usgs.gov/nwis) using the
512 USGS dataRetrieval R package (version 2.7.14)[90]. EDI datasets analyzed included ice duration[31];
nutrients, pH, and carbon[91]; major ions[92]; water temperatures combined from multiple datasets[93–
514 97] as described in Rohwer *et al.*[22]; phytoplankton[98]; and zooplankton[99] converted to biomass as
described in Rohwer, Ladwig, *et al.*[33]. River discharge measurements were obtained from the
516 USGS for the Yahara River, the primary tributary into Lake Mendota (site ID: 05427718)[32]. After
exploring all parameters included in these datasets, the occurrence of a hot, dry year with low
518 primary productivity became apparent. Lake heatwaves spanning much of 2012 were confirmed
using the 90th percentile definition from Woolway et al.[38] and the heatwaveR R package (version
520 0.4.6)[100].

522 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
524 identified as SNV positions that were called by inStrain[80] for the first time in a given sample. To
identify dates where an unusual number of new SNVs appeared, possibly indicating the emergence
526 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
528 excluding the initial consecutive dates where new SNVs remained in the 4th quantile. Genes under
selection were identified using dN/dS and pN/pS ratios as calculated by inStrain[80]. A McDonald-
530 Kreitman test[101] was used to identify positively selected genes where the bias of unfixed SNVs to
be nonsynonymous was lower than the bias of fixed SNVs to be nonsynonymous (pNpS/dNdS <
532 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
534 under significant positive selection with high frequency (in the 4th quartile). Consistently selected
genes were identified for the pre-2012 and post-2012 time periods separately.
536

Gene annotations were analyzed in the context of the KEGG pathways[42] they belonged to. For
538 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
540 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.

Data Availability

Metagenome sequences are available from the NCBI Sequence Read Archive under Umbrella Project accession PRJNA1056043. Individual metagenome SRA accession numbers are also listed in Supplementary Data 1. The filtered fastq files and single-sample assemblies used in this study are available through the JGI Genome Portal under ITS Proposal ID 504350. The 2,855 species-representative MAGs are also under the NCBI Umbrella Project accession PRJNA1056043. Individual NCBI Genome IDs are listed in Supplementary Data 2. Environmental data is publicly available through the Environmental Data Initiative (https://edirepository.org/)[31,91–99] and the U.S. Geological Survey's Water Data for the Nation (https://waterdata.usgs.gov/nwis)[32]. Custom scripts used for data processing are available at https://github.com/rrohwer/TYMEFLIES_manuscript and through Zenodo[102].

References

1. Lenski, R. E. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* **11**, 2181–2194 (2017).
2. Bassar, R. D., Coulson, T., Travis, J. & Reznick, D. N. Towards a more precise – and accurate – view of eco-evolution. *Ecology Letters* **24**, 623–625 (2021).
3. Hendry, A. P. *Eco-Evolutionary Dynamics*. (Princeton University Press, 2017).
4. Ware, I. M. *et al.* Feedbacks link ecosystem ecology and evolution across spatial and temporal scales: Empirical evidence and future directions. *Functional Ecology* **33**, 31–42 (2019).
5. Rosselló-Móra, R. & Amann, R. Past and future species definitions for Bacteria and Archaea. *Systematic and Applied Microbiology* **38**, 209–216 (2015).
6. Rohwer, R. R. & McMahon, K. D. Lake iTag measurements over nineteen years, introducing the limony dataset. 2022.08.04.502869 Preprint at https://doi.org/10.1101/2022.08.04.502869 (2022).
7. 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. 2022.07.11.499243 Preprint at https://doi.org/10.1101/2022.07.11.499243 (2022).
8. 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).
9. Varghese, N. J. *et al.* Microbial species delineation using whole genome sequences. *Nucleic Acids Research* **43**, 6761–6771 (2015).
10. 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).

11. Olm, M. R. *et al.* Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems* **5**, 10.1128/msystems.00731-19 (2020).
12. Rossum, T. V., Ferretti, P., Maistrenko, O. M. & Bork, P. Diversity within species: interpreting strains in microbiomes. *Nature Reviews Microbiology* 1–16 (2020) doi:10.1038/s41579-020-0368-1.
13. Chiriac, M.-C., Haber, M. & Salcher, M. M. Adaptive genetic traits in pelagic freshwater microbes. *Environmental Microbiology* **25**, 606–641 (2023).
14. 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).
15. Neuenschwander, S. M., Ghai, R., Pernthaler, J. & Salcher, M. M. Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J* **12**, 185–198 (2018).
16. Magnuson, J. J., Kratz, T. K. & Benson, B. J. Long-Term Dynamics of Lakes in the Landscape: Long-Term Ecological Research on North Temperate Lakes. (Oxford University Press, 2006).
17. Yannarell, A. C., Kent, A. D., Lauster, G. H., Kratz, T. K. & Triplett, E. W. Temporal Patterns in Bacterial Communities in Three Temperate Lakes of Different Trophic Status. *Microb Ecol* **46**, 391–405 (2003).
18. Magee, M. R. & Wu, C. H. Effects of changing climate on ice cover in three morphometrically different lakes. *Hydrol. Process.* **31**, 308–323 (2017).
19. Magee, M. R. & Wu, C. H. Response of water temperatures and stratification to changing climate in three lakes with different morphometry. *Hydrology and Earth System Sciences* **21**, 6253–6274 (2017).
20. Snortheim, C. A. *et al.* Meteorological drivers of hypolimnetic anoxia in a eutrophic, north temperate lake. *Ecological Modelling* **343**, 39–53 (2017).
21. 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.* **n/a**, 1–17 (2020).
22. 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. *Proceedings of the National Academy of Sciences* **120**, e2211796120 (2023).
23. 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).
24. Meziti, A. *et al.* Quantifying the changes in genetic diversity within sequence-discrete bacterial populations across a spatial and temporal riverine gradient. *The ISME Journal* **13**, 767 (2019).
25. Schmidt, M. L. *et al.* Microhabitats are associated with diversity–productivity relationships in freshwater bacterial communities. *FEMS Microbiology Ecology* **96**, fiaa029 (2020).
26. Garcia, S. L. *et al.* Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *The ISME Journal* 1 (2017) doi:10.1038/s41396-017-0001-0.
27. Magnuson, J. J. Long-Term Ecological Research and the Invisible Present. *BioScience* **40**, 495–501 (1990).

28. Turner, M. G. *et al.* Climate change, ecosystems and abrupt change: science priorities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190105 (2020).
29. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
30. Denef, V. J. & Banfield, J. F. In Situ Evolutionary Rate Measurements Show Ecological Success of Recently Emerged Bacterial Hybrids. *Science* **336**, 462–466 (2012).
31. 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).
32. U.S. Geological Survey National Water Information System data available on the World Wide Web. USGS 05427718 Yahara River at Windsor, WI, Daily Discharge 00060. U.S. Geological Survey National Water Information System data available on the World Wide Web (2023).
33. Rohwer, R. R. *et al.* Increased anoxia following species invasion of a eutrophic lake. *Limnology and Oceanography Letters* **n/a**, (2023).
34. Bailey, J. & Hood, J. M. Biotic and thermal drivers alter zooplankton phenology in western Lake Erie. *Limnology and Oceanography Letters* **9**, 219–228 (2024).
35. Carpenter, S. R., Booth, E. G. & Kucharik, C. J. Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnology and Oceanography* **63**, 1221–1233 (2018).
36. 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).
37. 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. *Journal of Geophysical Research: Biogeosciences* **127**, e2021JG006550 (2022).
38. Woolway, R. I. *et al.* Lake heatwaves under climate change. *Nature* **589**, 402–407 (2021).
39. Walsh, J. R., Munoz, S. E. & Vander Zanden, M. J. Outbreak of an undetected invasive species triggered by a climate anomaly. *Ecosphere* **7**, n/a-n/a (2016).
40. Walsh, J. R., Carpenter, S. R. & Zanden, M. J. V. Invasive species triggers a massive loss of ecosystem services through a trophic cascade. *PNAS* **113**, 4081–4085 (2016).
41. Rastetter, E. B. *et al.* Time lags: insights from the U.S. Long Term Ecological Research Network. *Ecosphere* **12**, e03431 (2021).
42. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* **51**, D587–D592 (2023).
43. 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).
44. 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).

45. Garcia, S. L. *et al.* Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. *Mol Ecol* **24**, 4449–4459 (2015).
46. Garcia, S. L. *et al.* Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME Journal* **7**, 137 (2013).
47. 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).
48. Drake, T. W., Raymond, P. A. & Spencer, R. G. M. Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnology and Oceanography Letters* **3**, 132–142 (2018).
49. Grant, L. *et al.* Attribution of global lake systems change to anthropogenic forcing. *Nat. Geosci.* **14**, 849–854 (2021).
50. Roy, H. E. *et al. IPBES Invasive Alien Species Assessment: Summary for Policymakers*. https://zenodo.org/record/8314303 (2023) doi:10.5281/zenodo.8314303.
51. Zwart, J. A. *et al.* Cross-scale interactions dictate regional lake carbon flux and productivity response to future climate. *Geophys. Res. Lett.* 2019GL083478 (2019) doi:10.1029/2019GL083478.
52. Finlay, K. *et al.* Decrease in CO2 efflux from northern hardwater lakes with increasing atmospheric warming. *Nature* **519**, 215–218 (2015).
53. Azam, F. *et al.* The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series* **10**, 257–263 (1983).
54. Ghiotto, G., Zampieri, G., Campanaro, S. & Treu, L. Strain-resolved metagenomics approaches applied to biogas upgrading. *Environmental Research* **240**, 117414 (2024).
55. 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).
56. Ustick, L. J., Larkin, A. A. & Martiny, A. C. Global scale phylogeography of functional traits and microdiversity in Prochlorococcus. *The ISME Journal* **17**, 1671–1679 (2023).
57. Koch, H. *et al.* Genomic, metabolic and phenotypic variability shapes ecological differentiation and intraspecies interactions of Alteromonas macleodii. *Sci Rep* **10**, 809 (2020).
58. 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).
59. Keller, L. M., Colman, D. R. & Boyd, E. S. An active microbiome in Old Faithful geyser. *PNAS Nexus* **2**, pgad066 (2023).
60. 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. *Applied and Environmental Microbiology* **88**, e02464-21 (2022).
61. Liao, J. *et al.* Microdiversity of the vaginal microbiome is associated with preterm birth. *Nat Commun* **14**, 4997 (2023).

62. 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).
63. Niccum, B. A., Kastman, E. K., Kfoury, N., Robbat, A. & Wolfe, B. E. Strain-Level Diversity Impacts Cheese Rind Microbiome Assembly and Function. *mSystems* **5**, 10.1128/msystems.00149-20 (2020).
64. 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).
65. Wolff, R., Shoemaker, W. & Garud, N. Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. *mBio* **14**, e02502-22 (2023).
66. Hoetzinger, M. *et al.* Geographic population structure and distinct intra-population dynamics of globally abundant freshwater bacteria. *The ISME Journal* wrae113 (2024) doi:10.1093/ismejo/wrae113.
67. 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).
68. Sukumaran, J. & Knowles, L. L. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* **114**, 1607–1612 (2017).
69. Coates, D. J., Byrne, M. & Moritz, C. Genetic Diversity and Conservation Units: Dealing With the Species-Population Continuum in the Age of Genomics. *Frontiers in Ecology and Evolution* **6**, (2018).
70. Brock, T. D. *A Eutrophic Lake: Lake Mendota, Wisconsin*. (Springer, New York, NY, 1985). doi:10.1007/978-1-4419-8700-6_7.
71. Clum, A. *et al.* DOE JGI Metagenome Workflow. *mSystems* **6**, 10.1128/msystems.00804-20 (2021).
72. Bushnell, B. *BBMap Short Read Aligner and Other Bioinformatic Tools*. https://www.osti.gov/biblio/1241166 (2014).
73. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).
74. Wilson, L. A. *et al.* Launcher: A simple tool for executing high throughput computing workloads. *Journal of Open Source Software* **2**, 289 (2017).
75. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
76. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
77. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* btz848 (2019) doi:10.1093/bioinformatics/btz848.

78. 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 Research* **50**, D785–D794 (2022).
79. Aroney, S. T. N. *et al.* CoverM: Read coverage calculator for metagenomics. Zenodo https://doi.org/10.5281/zenodo.10531254 (2024).
80. Olm, M. R. *et al.* inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat Biotechnol* **39**, 727–736 (2021).
81. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
82. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
83. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).
84. Barrett, T. et al. Data.Table: Extension of `data.Frame`. (2024).
85. Grolemund, G. & Wickham, H. Dates and Times Made Easy with lubridate. *Journal of Statistical Software* **40**, 1–25 (2011).
86. Tange, O. GNU Parallel 20230822 ('Chandrayaan'). Zenodo https://doi.org/10.5281/zenodo.8278274 (2023).
87. Boogaart, K. G. van den, Tolosana-Delgado, R. & Bren, M. compositions: Compositional Data Analysis. (2023).
88. Oksanen, J. *et al.* vegan: Community Ecology Package. (2022).
89. Zeileis, A. *et al.* strucchange: Testing, Monitoring, and Dating Structural Changes. (2022).
90. DeCicco, L., Hirsch, R., Lorenz, D., Watkins, D. & Johnson, M. *dataRetrieval: R Packages for Discovering and Retrieving Water Data Available from U.S. Federal Hydrologic Web Services*. (U.S. Geological Survey, Reston, VA, 2023). doi:10.5066/P9X4L3GE.
91. 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).
92. 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).
93. 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).
94. 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).

95. 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).
96. 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).
97. 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).
98. 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).
99. 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).
100. Schlegel, R. W. & Smit, A. J. heatwaveR: A central algorithm for the detection of heatwaves and cold-spells. *Journal of Open Source Software* **3**, 821 (2018).
101. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652–654 (1991).
102. 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 Angela Kent, Tony Yannarell, Ashley Shade, Stuart Jones, Ryan Newton, Georgia Wolfe, Todd Miller, Emily Kara Read, Lucas Beversdorf, James Mutschler, and the original Microbial Observatory lead Eric W. Triplett. We thank Sarah Stevens for her early input into the ideas pursued here, Peter Golightly for advice on genes under selection data, Tyler Butts for advice on environmental data, and William Ratcliff and Vincent Denef for advice on framing.

Funding

E. Michael and Winona Foster WARF Wisconsin Idea Fellowship (RRR)
U.S. National Science Foundation (DBI-2011002) (RRR)
The Texas Advanced Computing Center (TACC) at The University of Texas at Austin provided HPC resources that contributed to the research results reported within this paper (http://www.tacc.utexas.edu) (RRR)
U.S. National Institutes of Health (R01-GM116853) (MKirk)
U.S. National Science Foundation (DEB-1831730) (MKirk)

The work (proposal: https://doi.org/10.46936/10.25585/60001198) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231 (MKell)
U.S. Department of Energy Joint Genome Institute (CSP 504350) (KDM)
U.S. Department of Agriculture (WIS01516 and WIS01789) (KDM)
U.S. National Science Foundation (DEB-0702395, DEB-1344254) (KDM)
U.S. National Science Foundation North Temperate Lakes Long-Term Ecological Research program (DEB-9632853, DEB-0217533, DEB-0822700, DEB-1440297, DEB-2025982) (KDM)
U.S. National Science Foundation Microbial Observatory program (MCB-9977903, DEB-0702395) (KDM)
Simons Foundation Investigator in Aquatic Microbial Ecology Award (LI-SIAME-00002001) (BJB)

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. RRR 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. and B.J.B. wrote the final draft incorporating edits provided by M.Kirk., S.L.G., M.Kell., and K.D.M.

Competing interests

Authors declare that they have no competing interests.

Supplementary Information

Supplementary Data 1. TYMEFLIES metagenome metadata. Includes metadata for metagenome samples including JGI, GOLD, and NCBI sample identifiers as well as McMahon Lab identifiers that pair metagenome samples with previous 16S rRNA gene sequencing[6].

Supplementary Data 2. TYMEFLIES MAG metadata. NCBI identifiers corresponding to each species-representative genome, as well as genome quality calculated by CheckM2[7], taxonomy assigned by GTDB-tk[77], and average relative abundance calculated by coverM[79].

Supplementary Data 3. Consistently selected gene annotations. KEGG annotations of consistently positively selected genes in a *Nanopelagicus* species that experienced a step change in strain composition in 2012 (ME2011-09-21_3300043464_group3_bin69). Table row order matches heatmap row order in **Fig. 5F**.