

# SCIENTIFIC REPORTS



OPEN

## Importance of sampling frequency when collecting diatoms

Naicheng Wu<sup>1,2,3</sup>, Claas Faber<sup>1</sup>, Xiuming Sun<sup>1</sup>, Yueming Qu<sup>1</sup>, Chao Wang<sup>4</sup>, Snjezana Ivetic<sup>5</sup>, Tenna Riis<sup>3</sup>, Uta Ulrich<sup>1</sup> & Nicola Fohrer<sup>1</sup>

Received: 08 June 2016

Accepted: 24 October 2016

Published: 14 November 2016

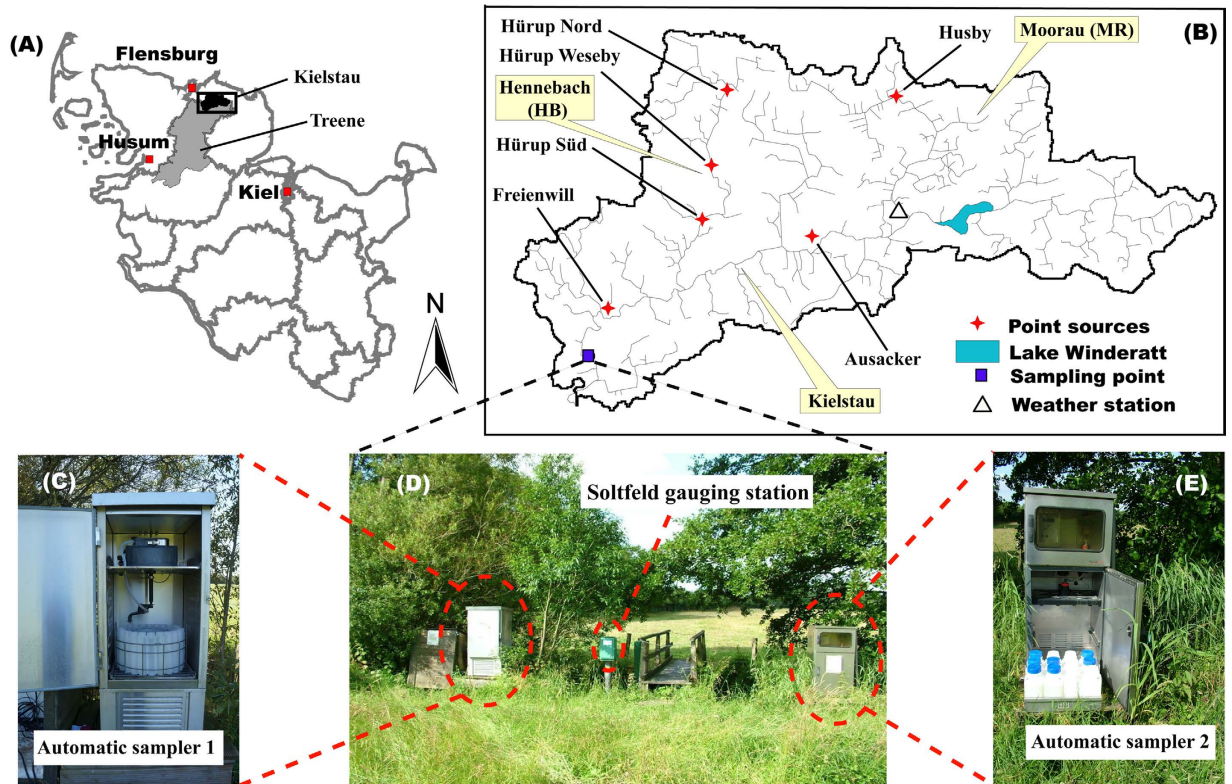
There has been increasing interest in diatom-based bio-assessment but we still lack a comprehensive understanding of how to capture diatoms' temporal dynamics with an appropriate sampling frequency (ASF). To cover this research gap, we collected and analyzed daily riverine diatom samples over a 1-year period (25 April 2013–30 April 2014) at the outlet of a German lowland river. The samples were classified into five clusters (1–5) by a Kohonen Self-Organizing Map (SOM) method based on similarity between species compositions over time. ASFs were determined to be 25 days at Cluster 2 (June–July 2013) and 13 days at Cluster 5 (February–April 2014), whereas no specific ASFs were found at Cluster 1 (April–May 2013), 3 (August–November 2013) (>30 days) and Cluster 4 (December 2013 - January 2014) (<1 day). ASFs showed dramatic seasonality and were negatively related to hydrological wetness conditions, suggesting that sampling interval should be reduced with increasing catchment wetness. A key implication of our findings for freshwater management is that long-term bio-monitoring protocols should be developed with the knowledge of tracking algal temporal dynamics with an appropriate sampling frequency.

Global freshwater resources are widely affected by multiple stressors, such as stream morphological change, point and diffuse sources<sup>1</sup>, which has led to widespread bio-monitoring and bio-assessment. As diatoms strongly respond to environmental changes<sup>2–6</sup>, they are increasingly being employed as bio-indicators to assess ecological conditions in lentic and lotic ecosystems around the world<sup>7–9</sup>. Diatoms are a key component of riverine algae and are often the most important primary producers of stream ecosystems<sup>4,6,10–12</sup>. They are preferred as bio-indicators for a range of reasons besides their sensitivity to subtle changes in environmental conditions due to their short life cycle<sup>2</sup>. For instance, diatoms can be observed in nearly every aquatic environment including fresh- and marine waters, moist terrestrial habitats, such as soils, rock surfaces or epiphytes<sup>13</sup>. They are cosmopolitan, with a wide geographical distribution and well-known autecology of most species<sup>14,15</sup>. They are also easy to sample and induce minimal impact on resident biota during collections. Furthermore, their small cell sizes (commonly between 10–200 µm in diameter or length)<sup>16</sup> allow them to behave similarly to silt particles in stream bodies<sup>17</sup> and easily be transported by water when a rainfall event occurs, making them a potential tracer for hydrological processes<sup>18</sup>.

Research on the factors controlling the abundance and distribution of the diatom community in streams and rivers has thus become an important part of monitoring overall water quality<sup>19</sup>. Generally the abundance of algal communities is controlled by resource availability (e.g. light and nutrients), disturbances and grazing<sup>6,20</sup>. More specifically, studies on the effect of resources and disturbances on diatom communities show that resource factors predominantly govern the processes of biomass gain<sup>21</sup>, while the processes of biomass loss are determined primarily by hydraulic factors<sup>22</sup>. Lange *et al.*<sup>10</sup> found that light, nutrients and grazing interact to determine stream diatom community composition and functional group structure<sup>10</sup>. For example, compared to 'low profile' taxa, 'high profile' and 'motile' taxa became more prevalent at higher resource levels and the benefit of nutrient enrichment for 'motile' diatoms was higher at ambient than at reduced light<sup>10</sup>. Disturbances, for instance caused by extreme fluctuations in stream flows, can also change stream algal communities and functional traits<sup>23,24</sup>.

One of the basic premises of diatom-based monitoring is that the species assemblage has been characterized accurately within the study system across spatial and temporal scales. Up to now, temporal dynamics of diatom

<sup>1</sup>Department of Hydrology and Water Resources Management, Institute for Natural Resource Conservation, Kiel University, 24118 Kiel, Germany. <sup>2</sup>Aarhus Institute of Advanced Studies, Aarhus University, Høegh-Guldbergs Gade 6B, 8000 Aarhus C, Denmark. <sup>3</sup>Department of Bioscience, Aarhus University, Ole Worms Allé 1, 8000 Aarhus C, Denmark. <sup>4</sup>Pearl River Fisheries Research Institute, Chinese Academy of Fishery Science, 510380 Guangzhou, China. <sup>5</sup>Marine Science and Engineering, Red Sea Research Centre, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. Correspondence and requests for materials should be addressed to N.W. (email: nwu@hydrology.uni-kiel.de)



**Figure 1.** The location of the Kielstau catchment (B) in Schleswig-Holstein state (A) and three photos of the Soltfeld gauging station and automatic water samplers for daily mixed samples (C–E). The maps (A,B) are created using ArcGIS 10.3.1 software (<http://www.esri.com/software/arcgis>) and modified with Adobe Photoshop CS6. Photos were taken by Sun (2015).

communities have been widely investigated, since these variations can affect ecosystem processes, functioning and stability, and reflect major shifts in environmental conditions<sup>25,26</sup>. Nevertheless, the temporal sampling intervals across these studies varied greatly (sub-daily to yearly) and mainly depended on the research objectives and spatial sampling scales. For large spatial sampling studies, the researchers normally visited the sampling sites only once (few of them revisited for a second time)<sup>23,27–29</sup>. Seasonal and/or monthly collections have often been applied at small spatial-scale sampling or in regular bio-monitoring projects<sup>6,25,30,31</sup>. Few studies have applied weekly sampling frequencies<sup>32,33</sup>. However, sampling at a daily or sub-daily interval is rare in previous studies<sup>34</sup>. Based on this literature review we raised the following questions:

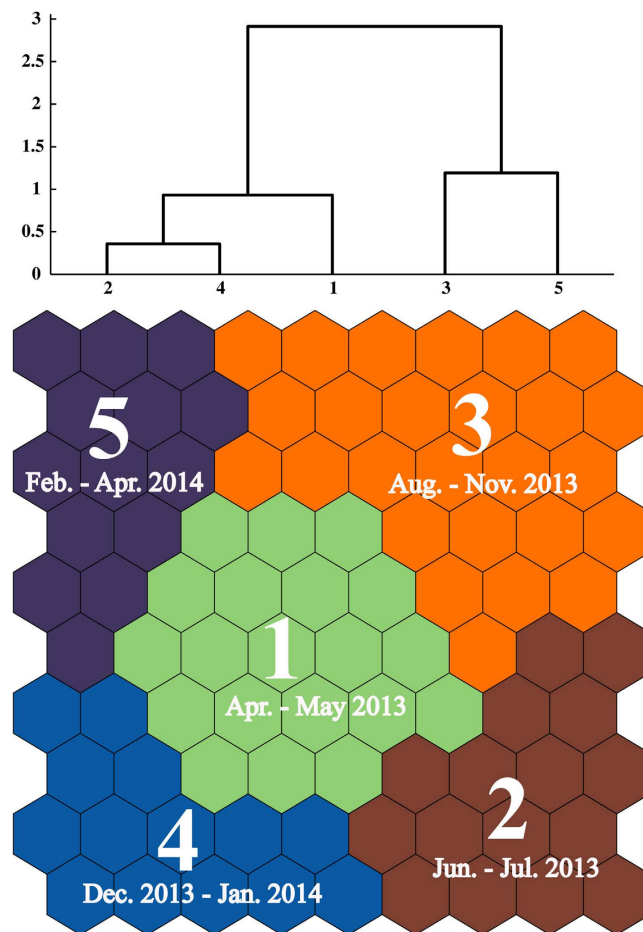
1. How well does a single sample of diatoms represent the time or season of sampling for comparison of diatom communities among multiple sites?
2. What would be the appropriate sampling frequency (ASF) in order to sample and describe the actual riverine diatom communities over time?

To our knowledge, these two questions are yet unanswered and no studies have determined the ASF of diatom communities. Only with suitable sampling frequency we will obtain enough information about community dynamics and at the same time avoid unnecessary duplication of efforts for lab processing and microscopic identification.

In this study, we used riverine pelagic diatom communities as a model system to answer these two questions and fill this research gap. We collected and analyzed the daily samples of riverine diatom communities over a 1-year period (25 April 2013–30 April 2014) at the outlet of a German lowland river (Fig. 1). The objectives were to 1) describe the temporal dynamics of diatom community, and 2) quantify the appropriate sampling frequency (ASF) at different hydrological conditions.

## Results

**Taxonomic composition and diversity.** During the study period (April 2013–April 2014), we observed a total of 113 diatom species belonging to 45 genera (e.g., *Achnanthes*, *Eunotia*, *Gomphonema*, *Navicula*, *Surirella*, *Ulnaria*, etc.; Supplementary material). Within all samples, *Achnantheidium minutissima* and *Navicula lanceolata* were the most abundant species (relative abundances: 39.9% and 15.9% of the total abundance, respectively). Other dominant species with relative abundance >1% were *Planothidium lanceolatum* (6.4%), *Stephanodiscus hantzschii* (4.9%), *Navicula gregaria* (4.8%), *Cocconeis placentula* (3.6%), *Cyclotella meneghiniana* (3.2%), *Nitzschia amphibia* (2.5%), *Diatoma mesodon* (1.7%), *Nitzschia adamata* (1.7%), *Ulnaria biceps* (1.6%),



**Figure 2.** Relationship between each cluster (above) and visualization of the temporal pattern of riverine diatom communities by self-organizing maps (SOMs) (below). Numbers indicate the different clusters. Dates below the numbers are the samples related to each cluster.

*Punctastriata lancettula* (1.2%), *Frustulia viridula* (1.1%) and *Gomphonema olivaceum* (1.0%). The mean diatom density across the whole sampling period was  $1.15 \times 10^6$  cell/L, ranging from a minimum of  $1.97 \times 10^4$  to a maximum of  $6.36 \times 10^6$  cell/L. The mean values ( $\pm$ SD; min-max) of species richness, evenness, Simpson diversity index, Shannon-Wiener index across all samples were 21 ( $\pm 6.13$ ; 7–36), 0.65 ( $\pm 0.17$ ; 0.24–0.96), 0.71 ( $\pm 0.17$ ; 0.19–0.93) and 1.94 ( $\pm 0.52$ ; 0.54–2.95), respectively.

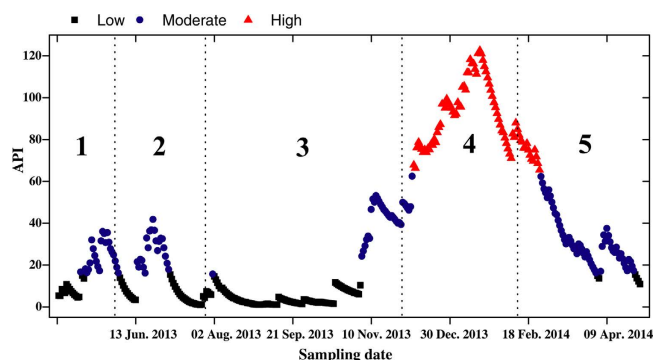
**Classification and visualization of sampling dates.** Based on the Kohonen Self-Organizing Map (SOM) hierarchical cluster analysis, we classified the samples of 348 days (25 April 2013–30 April 2014) into five SOM clusters (Fig. 2). With 31 exemptions (15 in Cluster 4), SOM Clusters 1–5 were composed mainly of samples taken from April–May 2013, June–July 2013, August–November 2013, December 2013–January 2014, and February–April 2014, respectively. To facilitate the following calculation of the appropriate sampling frequency (ASF), we manually categorized exemptions to the closest time groups. For example, although sampling dates (on 15 March 2014, 17 March 2014, 22 March 2014 and 01 April 2014) were classified into Cluster 3, we manually put these dates into Cluster 5 for the above reason. Nevertheless, MRPP analyses indicated significant differences among and between all different SOM clusters ( $p < 0.01$ ) (Table 1).

Using antecedent precipitation indices (APIs) to indicate catchment wetness conditions, we found that Clusters 1–3 were in low wetness conditions (mean APIs: 17.7, 13.6 and 12.7, respectively), Cluster 4 was in high wetness condition (mean API: 91.0) while Cluster 5 was in moderate condition with a mean API of 43.1 (Fig. 3).

**Appropriate sampling frequency (ASF).** Relationships between percentage similarity index (PSI) and sampling intervals at different SOM clusters showed, as expected, that PSIs decreased with increasing sampling intervals (Fig. 4). We found gradual PSI declines at Cluster 2 (June–July 2013) and Cluster 5 (February–April 2014) with ASFs of 25 days and 13 days, respectively (Fig. 4B,E). We were not able to quantify the ASFs of Cluster 1 and 3, which were  $> 30$  days (Fig. 4A,C). In contrast, the ASF of Cluster 4 was  $< 1$  day (Fig. 4D). ASF was negatively correlated with API (Fig. 5). During the low and moderate wetness conditions, ASFs vary greatly. For example, an ASF of 25 days would be suitable when API = 13.6 (Cluster 2). However, we found ASFs  $> 30$  days at API = 12.7 (Cluster 3) and API = 17.7 (Cluster 1).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1				
Cluster 2	0.3362 (0.001)			
Cluster 3	0.1595 (0.001)	0.4005 (0.001)		
Cluster 4	0.1090 (0.001)	0.2819 (0.001)	0.2572 (0.001)	
Cluster 5	0.2697 (0.001)	0.2217 (0.001)	0.3259 (0.001)	0.2060 (0.001)

**Table 1. Multi-Response Permutation Procedure (MRPP) analyses between different SOM clusters.** Numbers are A values. P values are listed in the parentheses. The difference among the five clusters is also significant ( $A = 0.3834$ ,  $p = 0.001$ ).



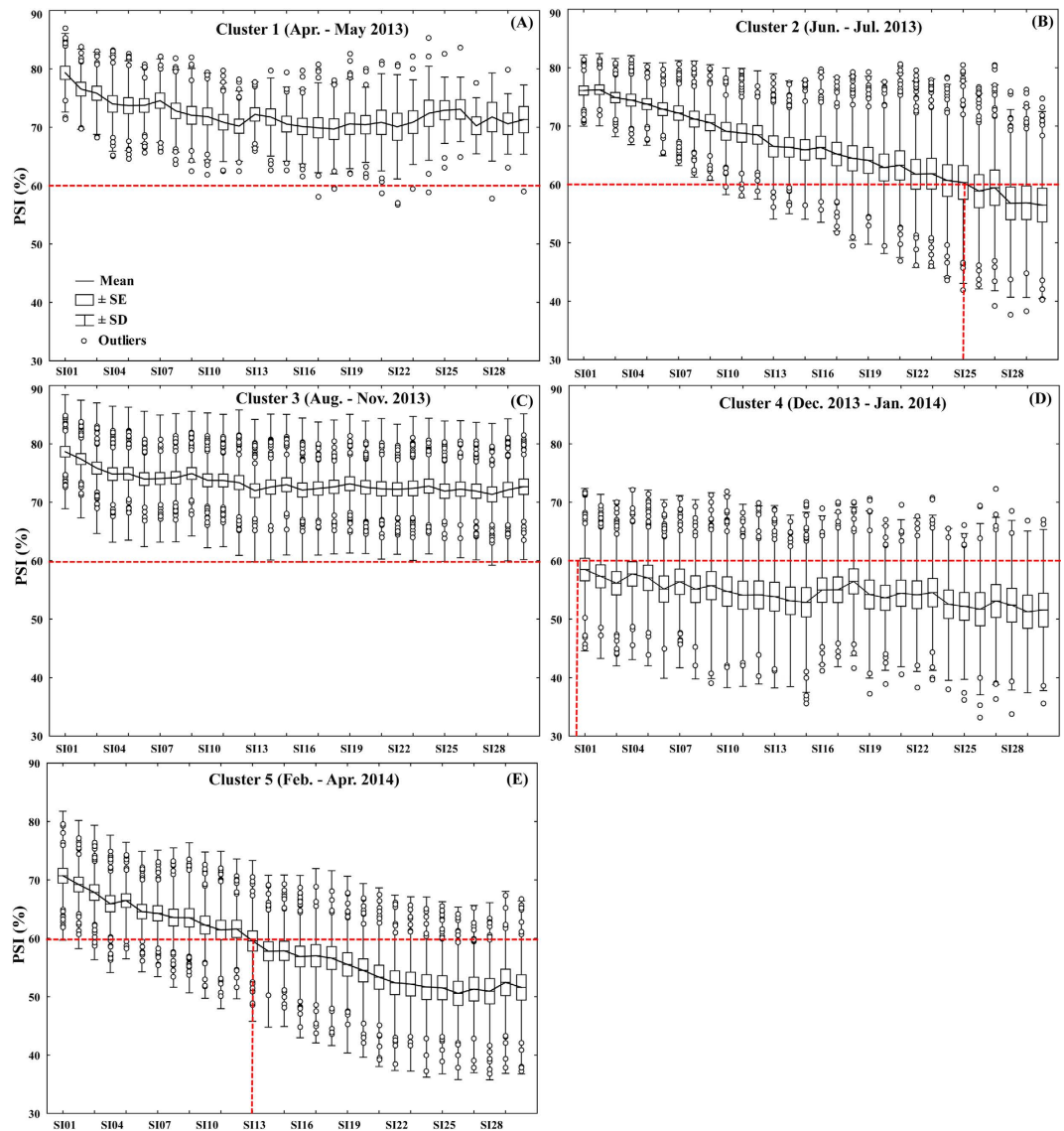
**Figure 3. Temporal dynamics of API (antecedent precipitation index) with relation to different SOM clusters (indicated by the large central numbers).** To clearly show the wetness conditions we classified the whole APIs by quartiles of a long-term daily APIs (October 2010–June 2015): low:  $\leq$  1<sup>st</sup> quartile (API: 15.5); moderate: 1<sup>st</sup>–3<sup>rd</sup> quartile (API: 15.5–63.4); high:  $>$  3<sup>rd</sup> quartile (API: 63.4). Black squares indicate low wetness conditions, blue circles are moderate, and red triangles are high. Vertical dashed lines separate different SOM clusters.

## Discussion

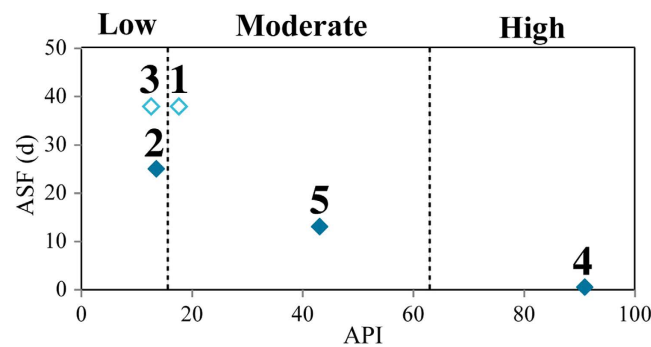
The riverine diatom community showed a dramatic temporal dynamic in this study, in accordance with previous studies<sup>4,6,8,35–37</sup>. Although the relationship between riverine algae community variations and environmental factors (e.g., resources, disturbances, grazers, etc.) has been intensively investigated, there is still no general consensus as to which factors regulate riverine algae community in lotic habitats, and furthermore the contributions of main environmental factors to algae variation remain controversial<sup>6,38</sup>. The lack of consensus between studies might be due to the fact that the community–environment relationship is dependent on spatial scale. For example, studies at large scale (e.g., national scale) often found that the geographical topography (e.g., altitude, latitude, longitude) and climate were the dominant factors regulating diatom variation<sup>11,27,29,39</sup>. In contrast, at small scales the main regulating factors have been shown to be typically hydrological factors (e.g., discharge), habitat (e.g., substrate composition, sediment input and transport), water quality (e.g., nutrient, dissolved oxygen) as well as bio–interaction (e.g., grazing, competition, parasitism)<sup>27,32</sup>.

In this study, we found a close relationship between riverine diatom communities and hydrological conditions (as indicated by antecedent precipitation indices, APIs) as expected (Figs 2 and 3). Given that temperature is an important environmental factor governing algal growth, we might expect that the diatom communities have the potential to change much more quickly during the warmer period than during winter. However, our results showed a contrary phenomenon: PSI community similarities in winter were much lower than those in summer (Fig. 4), indicating the greater contribution of hydrological factors than physicochemical variables to diatom community changes. Hydrological conditions and precipitation are general factors that determine the habitat and affect (directly or indirectly) many other environmental variables that are key factors in diatom community development, such as discharge, residence time, temperature, light availability, and dissolved oxygen<sup>25,32</sup>. Moreover, strong precipitation and associated surface run-off may accelerate the transportation of diatom species from soil surface to stream water via soil macro-pores or surface runoff<sup>18</sup>. This process also potentially affects the composition of stream diatom communities. Martínez-Carreras *et al.*<sup>18</sup> found that the contribution of aerial diatoms in stream water samples collected during rainfall events increased with runoff during all seasons<sup>18</sup>.

Furthermore, storms have been identified as important for the delivery of major nutrients (e.g., P and N) from diffuse agricultural sources<sup>40</sup>, which were a primary factor contributing to variation in phytoplankton assemblages<sup>32</sup>. As a comprehensive index, API integrates not only hydrological processes but also the delivery of nutrients as well as toxic substances such as pesticides and runoff from roads. It can thus be a promising proxy for a large range of environmental factors (e.g., nutrients, hydrological factors) influencing diatom communities. Parallel research (Sun *et al.* unpublished data) found that API was significantly correlated (Pearson,  $p < 0.05$ )



**Figure 4.** Percentage similarity index (PSI) with different sampling intervals (SI) at different SOM clusters (A–E). SE = standard error, SD = standard deviation. Red dashed lines are the critical threshold of PSI = 60%.



**Figure 5.** Relationship between ASFs (appropriate sampling frequencies) and APIs (antecedent precipitation indices). Numbers indicate different SOM cluster (Cluster 4 was assigned with 0.5; for details see main text). API indices were mean values of each SOM cluster. Vertical dashed lines separate different wetness conditions (i.e. low, moderate, high). Clusters 1 and 3 (indicated by open diamond) were >30 days.

with chemical variables (e.g., total phosphorus, orthophosphate phosphorus, nitrite-nitrogen, chloride, silicon, sulphate) and hydrological factors (e.g. precipitation, surface runoff, water level) in the study area.

Although diatom communities have been greatly employed in bio-assessment programs, the basic and critical prerequisite is to track their temporal dynamics accurately. This prerequisite is widely recognized but has not yet been studied until now. Recent research carried out in a Chinese subtropical mountain river network has concluded that long-term and high-frequency (i.e., monthly) bio-monitoring protocols should be developed for accurately capturing the benthic algal temporal dynamics<sup>25</sup>. Prompted by the importance of bio-monitoring studies and the scarcity of investigations on ASFs, this study took a step further based on a valuable daily sampling regime and quantified the ASFs at different wetness conditions (indicated by APIs) (Figs 4 and 5). Considering the ASF variation at low and moderate wetness conditions (Fig. 5), we recommend an ASF of 7 days (i.e. weekly sampling) for low and moderate wetness periods when  $API < 63.4$ , and during high wet conditions ( $API > 63.4$ ) an ASF of daily or sub-daily (e.g., rain-event based sampling) would be necessary to capture the short-term variation of diatom communities.

Currently, the sampling frequencies recommended by commonly used bio-assessment protocols such as the U.S. Environmental Protection Agency are mostly single or seasonally sampling depending on the project goals<sup>41</sup>. In addition, some small spatial scale sampling or regular bio-monitoring projects usually apply seasonal or monthly frequencies<sup>6,25,30,31</sup>, while few studies with weekly sampling frequency<sup>32,33</sup>. By comparison, our results, provide a concrete baseline for the first time and will be beneficial for future planning of diatom based bio-monitoring campaigns.

Some uncertainties from this study need further investigations. Firstly, our conclusion was drawn from a rough estimation because of the low number of ASFs (Fig. 5). To overcome the above drawback and for safety reasons, the solution in this study was to take a much smaller ASF value than the lowest boundary. Nevertheless, as for the bio-assessment especially for long-term monitoring campaigns, this solution is not ideal, since it might waste a great deal of efforts by potentially taking some unnecessary samples. In future studies, therefore, we should try to increase the number of ASFs, which will allow more accurate ASFs to be developed. In particular, more attention should be paid to the very low wetness condition since the ASFs may increase exponentially (rather than linearly) when API gets close to 0. On the other hand, very low wetness conditions may also pose the risk of stream drying, which is not the case in our study region but requires higher sampling frequency during rewetting. Secondly, the ASFs were computed from the Kielstau catchment with specific climate and precipitation types, and may change if we implement these ASFs into other catchments or climate zones directly, because APIs are catchment-specific, and the actual precipitation related to a certain API will vary between catchments. Thus, further testing and assessment of the applicability of ASFs obtained from this study are still needed. Thirdly, the issue we highlighted may not be limited to only diatom or riverine algae communities. This consideration should also be taken into account for other organisms often used as bio-indicators, such as zooplankton, macroinvertebrates, fish, and macrophytes. Nevertheless, ASFs may be dependent on species generation times and mobility, which should be taken into consideration in future research. Organisms with shorter generation times and higher species mobility might need higher ASFs compared to those with longer generation times and lower mobility. Organism-based stream bio-assessment would be greatly benefited from further studies spreading this approach to other organisms and quantifying the appropriate sampling frequency in the target study reaches with habitat heterogeneities.

In conclusion, the temporal dynamics of the diatom community in our study area were remarkable and the similarity between samples decreased with increasing sampling interval. The temporal dynamic showed a close relationship with hydrological conditions (indicated by APIs) and ASFs were clearly negatively correlated with hydrological wetness conditions. To effectively capture variation processes of diatom assemblages, we recommend a higher sampling frequency (i.e., daily or sub-daily sampling) during high wetness condition when  $API > 63.4$  than during low and moderate conditions when  $API < 63.4$  (i.e. weekly sampling). Moreover, we expect that the current approach will ultimately draw the attention of ecologists especially when they prepare long-term bio-monitoring protocols.

## Methods

**Description of the study area.** The Kielstau catchment, a lowland watershed with a drainage area of 50 km<sup>2</sup>, is an UNESCO Demosite for Ecohydrology since 2010 and is located in the northern part of Germany (Schleswig-Holstein). It originates in the upper part of Lake Winderatt and is an important tributary of the Treene River (Fig. 1A). Moorau (MR) and Hennebach (HB) are two main tributaries within the catchment. The drained fraction of agricultural area in the Kielstau catchment is estimated to be 38%<sup>42</sup>. Sandy, loamy and peat soils are characteristic of the catchment. Land use is dominated by arable land and pasture (~55% and ~26%, respectively, of the catchment area)<sup>42</sup>. Six wastewater treatment plants are located in the Kielstau watershed (main stream: Ausacker and Freienwill; Moorau: Husby; Hennebach: Hürup Nord, Hürup Weseby and Hürup Süd) (Fig. 1B). The instream water quality has been considerably influenced by the combination of point sources, artificial drainage systems and diffuse sources in the catchment<sup>4</sup>.

**Sampling methods and primary procedures.** The Soltfeld gauging station, installed at the outlet of the Kielstau catchment (Fig. 1D), is a part of the official gauging network of the Federal State of Schleswig-Holstein. Close to the gauging station are the automatic devices for the collection of mixed daily water samples, which were taken during the period from 25 April 2013 to 30 April 2014. With the automatic sampler (Fig. 1C,E), mixed daily water samples were collected and taken back once a week (normally on Monday) to the lab of the Department of Hydrology and Water Resources Management of Kiel University. We expect that algae present in water samples are a mix of suspended benthic algae, and any phytoplankton communities entered the stream from upstream lakes or wetlands or communities developing in the river. Here we will call this the riverine algae community.

Moreover, since diatoms are the dominant algal group in many rivers, including the Kielstau where we performed this study<sup>6</sup>, we focus on diatom communities in the following analyses. In the lab, the water samples with known volumes (normally 2.5 L) were fixed in 5% non-acetic Lugol's iodine solution<sup>43</sup>. Forty-eight hours later, the supernatant liquid was removed and the retained organisms were transferred into glass containers. After sedimentation again, samples were concentrated to 25 mL for further processing. Sampling was not possible on 23 dates because of low temperatures or other technical factors, and a total of 348 samples were collected.

**Diatom preparation and identification.** Permanent diatom slides were prepared after oxidizing the organic material by the Hydrogen Peroxide Method (30% H<sub>2</sub>O<sub>2</sub> solution) and mounted in Naphrax (Northern Biological supplies Ltd., UK, R1 = 1.74) following Biggs and Kilroy<sup>44</sup>. A minimum of 300 valves were counted for each sample using a Zeiss Axioskop microscope at 1000× under oil immersion. Diatoms were identified to the lowest taxonomic level possible (mainly species level) according to following key books<sup>13,45–47</sup>. Diatom densities were expressed as cell/L.

**Data analyses.** To describe the diatom community, we calculated the diversity indices (i.e. species richness, evenness, Simpson diversity index, Shannon-Wiener index by *diversity* function in R package *vegan*), diatom density and relative abundances of dominant species.

Diatom assemblages were classified using Kohonen Self-Organizing Map (SOM), which is a novel approach to visualize high-dimensional data with unsupervised learning rules<sup>48</sup>. Within the algorithm, the complex statistical relations between high-dimensional data items are converted into simple geometric relationships on a low-dimensional display preserving the most important information from the primary data<sup>49</sup>. In this study, the SOM model described the patterns of temporal variation in diatom species: a total of 113 species were included. Sampling dates with similar species composition and structure were classified into the same neuron or into neighboring neurons, according to the degree of dissimilarity. The output of the SOM was a total of 90 neurons (virtual units), which were arranged into a 10\*9 hexagonal lattice to provide better visualization. The map size was set according to a heuristic equation (5n<sup>1/2</sup>)<sup>50</sup>, and then based on the minimum best values of quantization and topographic errors. The cells of the map were then subdivided into different groups according to the similarity of the weight vectors of the neurons using Ward's linkage method with Euclidian distance<sup>48</sup>. The group numbers were mainly based on the degree of dissimilarity of each SOM cell in the hierarchical clustering. The unified distance matrix (U-matrix)<sup>51</sup> and Davies-Bouldin index<sup>52</sup> were also applied to reinforce the group definition. We carried out all these analyses with Matlab software (Mathworks Inc 2001) using the SOM toolbox<sup>53</sup>. To assess the effectiveness of the hierarchical clustering on the SOM units, we used the Multi-Response Permutation Procedure (MRPP, *mrpp* function in R package *vegan*) to determine whether there was a significant difference between the clusters. We tested null hypothesis that there was no difference among the groups with a Monte Carlo randomization procedure with 999 permutations.

To determine an appropriate sampling frequency (ASF), we calculated the community similarity for each pairing of all possible successive days of 1–30 days within each SOM cluster. We hypothesized that the similarities will decrease with sampling interval increasing (e.g., the similarity between Day<sub>i</sub> vs. Day<sub>i+1</sub> is higher than those of Day<sub>i</sub> vs. Day<sub>i+2</sub>, Day<sub>i</sub> vs. Day<sub>i+3</sub>, and so on). In this study, the Percentage Similarity Index (PSI) was used as follows:

$$PSI = 100 - 0.5 \sum |P_{ik} - P_{jk}| = \sum \text{minimum}(P_{ik}, P_{jk}) \quad (1)$$

where the PSI represents the similarity between communities *i* and *j*, and ranges from 0.0 to 100, with 100 meaning the two communities have identical composition.  $P_{ik}$  and  $P_{jk}$  are the proportions of individuals present in communities *i* and *j*, respectively, that comprise the *k*<sup>th</sup> species. In other words, the PSI is the sum of the lowest percent value of a species between communities *i* and *j*. In spite of its simplicity, the percentage similarity measure is one of the better quantitative similarity coefficients available<sup>54,55</sup>. However, there is still no statistical method which can test the significance of the PSI; thus some scientists suggest a “cut-off” based on mean values from a large number of comparisons to indicate an ecologically-relevant change. For example, Matthews (1998) used the movement or displacement of points over time in ordination space to estimate the amount and directionality of community change. He recommends that a PSI of >0.6 among samples indicates a stable or similar community<sup>56</sup>. Therefore, in this study we also employed this empirical PSI of 60% as the threshold to quantify the ASFs at different SOM clusters.

To estimate the hydrological condition in the catchment, we calculated the API (antecedent precipitation index) on a daily basis (for details, see Fedora and Beschta<sup>57</sup> and Shaw<sup>58</sup>).

$$API_t = (k API_{t-1}) + P_{t-1} \quad (2)$$

where  $API_t$  = antecedent precipitation index (mm) at day<sub>*t*</sub>, and  $P_{t-1}$  = precipitation (mm) at the day<sub>*t-1*</sub>. The precipitation data are from our weather station in Moorau tributary (see Fig. 1). The value of *k* expresses moisture loss and varies seasonally, usually between 0.85 and 0.98, and is dependent on the potential loss of moisture (mainly through evapotranspiration)<sup>58</sup>. In this study, to estimate *k*, we used Penman potential evaporation derived from historical data from the University of Keele (Table 11.1 of Shaw<sup>58</sup>).

## References

- Hering, D. *et al.* Managing aquatic ecosystems and water resources under multiple stress--an introduction to the MARS project. *Science of the Total Environment* **503–504**, 10–21, doi: <http://dx.doi.org/10.1016/j.scitotenv.2014.06.106> (2015).
- Cabecinha, E. *et al.* Multi-scale approach using phytoplankton as a first step towards the definition of the ecological status of reservoirs. *Ecological Indicators* **9**, 240–255 (2009).

3. Domingues, R. & Galvão, H. Phytoplankton and environmental variability in a dam regulated temperate estuary. *Hydrobiologia* **586**, 117–134, doi: 10.1007/s10750-006-0567-4 (2007).
4. Wu, N. C., Schmalz, B. & Fohrer, N. Development and testing of a phytoplankton index of biotic integrity (P-IBI) for a German lowland river. *Ecological Indicators* **13**, 158–167, doi: 10.1016/j.ecolind.2011.05.022 (2012).
5. Zalack, J., Smucker, N. & Vis, M. Development of a diatom index of biotic integrity for acid mine drainage impacted streams. *Ecological Indicators* **10**, 287–295 (2010).
6. Wu, N. C., Schmalz, B. & Fohrer, N. Distribution of phytoplankton in a German lowland river in relation to environmental factors. *Journal of Plankton Research* **33**, 807–820, doi: 10.1093/plankt/FBQ139 (2011).
7. Borics, G. *et al.* A new evaluation technique of potamo-plankton for the assessment of the ecological status of rivers. *Archiv für Hydrobiologie (Supplement)* **161**, 465–486 (2007).
8. Gillett, N. D., Luttenton, M. R. & Steinman, A. D. Spatial and temporal dynamics of phytoplankton communities in a Great Lakes drowned river-mouth lake (Mona Lake, USA). *Journal of Limnology* **74**, 453–466, doi: 10.4081/jlimnol.2015.1055 (2015).
9. Kruk, C. *et al.* A morphological classification capturing functional variation in phytoplankton. *Freshwater Biology* **55**, 614–627, doi: 10.1111/j.1365-2427.2009.02298.x (2010).
10. Lange, K., Liess, A., Piggott, J. J., Townsend, C. R. & Matthaei, C. D. Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. *Freshwater Biology* **56**, 264–278, doi: 10.1111/j.1365-2427.2010.02492.x (2011).
11. Wu, N., Cai, Q. & Fohrer, N. Contribution of microspatial factors to benthic diatom communities. *Hydrobiologia* **732**, 49–60, doi: 10.1007/s10750-014-1843-3 (2014).
12. Wu, N. C. *et al.* Impacts of cascade run-of-river dams on benthic diatoms in the Xiangxi River, China. *Aquatic Sciences* **72**, 117–125 (2010).
13. Round, F. E., Crawford, R. M. & Mann, D. G. *The Diatoms: Biology and Morphology of the genera.* (Cambridge University Press, 1990).
14. van Dam, H., Mertens, A. & Sinkeldam, J. A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands. *Netherlands Journal of Aquatic Ecology* **28**, 117–133 (1994).
15. Wu, N. C., Schmalz, B. & Fohrer, N. A comparison of phytoplankton assemblages generated by two sampling protocols in a lowland catchment, Germany. *Annales de Limnologie/International Journal of Limnology* **47**, 313–323 (2011).
16. Mann, D. G. In *Automatic Diatom Identification* (eds H. du Buf & M. M. Bayer) 9–40 (World Scientific Publishing, 2002).
17. Jones, J. I., Duerdoth, C. P., Collins, A. L., Naden, P. S. & Sear, D. A. Interactions between diatoms and fine sediment. *Hydrological Processes* **28**, 1226–1237, doi: 10.1002/hyp.9671 (2014).
18. Martinez-Carreras, N. *et al.* Hydrological connectivity inferred from diatom transport through the riparian-stream system. *Hydrology and Earth System Sciences* **19**, 3133–3151, doi: 10.5194/hess-19-3133-2015 (2015).
19. Majaneva, M., Autio, R., Huttunen, M., Kuosa, H. & Kuparinen, J. Phytoplankton monitoring: the effect of sampling methods used during different stratification and bloom conditions in the Baltic Sea. *Boreal Environment Research* **14**, 313–322 (2009).
20. Biggs, B. J. F., Stevensen, R. J. & Lowe, R. L. A habitat matrix conceptual model for stream periphyton. *Archiv für Hydrobiologie* **143**, 21–56 (1998).
21. Piggott, J. J., Lange, K., Townsend, C. R. & Matthaei, C. D. Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *Plos One* **7**, e49873 (2012).
22. Biggs, B. J. F., Nikora, V. I. & Snelder, T. H. Linking scales of flow variability to lotic ecosystem structure and function. *River Research and Applications* **21**, 283–298, doi: 10.1002/rra.847 (2005).
23. Lange, K., Townsend, C. R. & Matthaei, C. D. A trait-based framework for stream algal communities. *Ecology and Evolution* **6**, 23–36, doi: 10.1002/ece3.1822 (2016).
24. Pace, M. L., Findlay, S. E. G. & Lints, D. Zooplankton in advective environments: The Hudson River community and a comparative analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 1060–11069 (1992).
25. Tang, T., Jia, X., Jiang, W. & Cai, Q. Multi-scale temporal dynamics of epilithic algal assemblages: evidence from a Chinese subtropical mountain river network. *Hydrobiologia* **770**, 289–299, doi: 10.1007/s10750-015-2603-8 (2016).
26. Suikkanen, S., Laamanen, M. & Huttunen, M. Long-term changes in summer phytoplankton communities of the open northern Baltic Sea. *Estuarine, Coastal and Shelf Science* **71**, 580–592, doi: 10.1016/j.ecss.2006.09.004 (2007).
27. Bae, M.-J. *et al.* Concordance of diatom, macroinvertebrate and fish assemblages in streams at nested spatial scales: Implications for ecological integrity. *Ecological Indicators* **47**, 89–101, doi: 10.1016/j.ecolind.2014.07.030 (2014).
28. Hering, D. *et al.* Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. *Freshwater Biology* **51**, 1757–1785, doi: 10.1111/j.1365-2427.2006.01610.x (2006).
29. Tang, T., Wu, N., Li, F., Fu, X. & Cai, Q. Disentangling the roles of spatial and environmental variables in shaping benthic algal assemblages in rivers of central and northern China. *Aquatic Ecology* **47**, 453–466, doi: 10.1007/s10452-013-9458-8 (2013).
30. Wu, N. C. *et al.* Changes in benthic algal communities following construction of a run-of-river dam. *Journal of the North American Benthological Society* **28**, 69–79, doi: 10.1899/08-047.1 (2009).
31. Centis, B., Tolotti, M. & Salmaso, N. Structure of the diatom community of the River Adige (North-Eastern Italy) along a hydrological gradient. *Hydrobiologia* **639**, 37–42, doi: 10.1007/s10750-009-0019-z (2010).
32. Bussi, G. *et al.* Impacts of climate change, land-use change and phosphorus reduction on phytoplankton in the River Thames (UK). *Science of the Total Environment*, doi: 10.1016/j.scitotenv.2016.1002.1109, doi: 10.1016/j.scitotenv.2016.02.109 (2016).
33. Read, D. S., Bowes, M. J., Newbold, L. K. & Whiteley, A. S. Weekly flow cytometric analysis of riverine phytoplankton to determine seasonal bloom dynamics. *Environ. Sci. Processes Impacts* **16**, 594–603 (2014).
34. Ilyash, L. V., Matori, D. N., Koltsova, T. I. & Sham, H. H. Spatial distribution and daily dynamics of phytoplankton in Nhatrang Bay of the South China Sea. *Oceanology* **44**, 219–229 (2004).
35. Baykal, T., Acikgoz, I., Udoh, A. U. & Yildiz, K. Seasonal variations in phytoplankton composition and biomass in a small lowland river-lake system (Melen River, Turkey). *Turkish Journal of Biology* **35**, 485–501 (2011).
36. Piirsoo, K., Pall, P., Tuvikene, A. & Viik, M. Temporal and spatial patterns of phytoplankton in a temperate lowland river (Emajõgi, Estonia). *Journal of Plankton Research* **30**, 1285–1295, doi: 10.1093/plankt/fbn082 (2008).
37. Basu, B. K. & Pick, F. R. Phytoplankton and zooplankton development in a lowland, temperate river. *Journal of Plankton Research* **19**, 237–253 (1997).
38. Basu, B. K. & Pick, F. R. Longitudinal and seasonal development of planktonic chlorophyll a in the Rideau River, Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 804–815 (1995).
39. Passy, S. I. A distinct latitudinal gradient of diatom diversity is linked to resource supply. *Ecology* **91**, 36–41 (2010).
40. Outram, F. N., Cooper, R. J., Sünnerberg, G., Hiscock, K. M. & Lovett, A. A. Antecedent conditions, hydrological connectivity and anthropogenic inputs: Factors affecting nitrate and phosphorus transfers to agricultural headwater streams. *Science of the Total Environment* **545–546**, 184–199, doi: 10.1016/j.scitotenv.2015.12.025 (2016).
41. Barbour, M., Gerritsen, J., Snyder, B. & Stribling, J. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition.* (EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water, 1999).
42. Fohrer, N., Schmalz, B., Tavares, F. & Golon, J. Ansätze zur Integration von landwirtschaftlichen Drainagen in die Modellierung des Landschaftswasserhaushalts von mesoskaligen Tieflandeinzugsgebieten. *Hydrologie & Wasserbewirtschaftung* **51**, 164–169 (2007).



43. Sabater, S. *et al.* Longitudinal development of chlorophyll and phytoplankton assemblages in a regulated large river (the Ebro River). *Science of Total Environment* **404**, 196–206 (2008).
44. Biggs, B. J. F. & Kilroy, C. *Stream Periphyton Monitoring Manual*. (NIWA, 2000).
45. Bey, M. Y. & Ector, L. Atlas des diatomées des cours d'eau de la région Rhône-Alpes (Tome 1 - Tome 5). (2013).
46. Lange-Bertalot, H. *Iconographia Diatomologica. Annotated Diatom Micrographs* Vol. 7, 9, 15, 18 (Koeltz Scientific Books, 2000a, 2000b, 2005, 2007).
47. Simonsen, R. Atlas and Catalogue of the Diatom Types of Friedrich Hustedt. Vol. 1-3, *Catalog*. (J. Cramer, 1987).
48. Wang, C. *et al.* Patterning and predicting phytoplankton assemblages in a large subtropical river. *Fundamental and Applied Limnology/Archiv für Hydrobiologie* **185**, 263–279, doi: 10.1127/fal/2014/0684 (2014).
49. Kohonen, T. *Self-organizing maps*. (Springer, 2001).
50. Vesanto, J. *Neural network tool for data mining: SOM Toolbox. -Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000)*. (Oulun yliopisto-paino, 2000).
51. Ultsch, A. *Self-organizing neural networks for visualization and classification*. (Springer-Verlag, 1993).
52. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**, 224–227 (1979).
53. Alhoniemi, E., Himberg, J., Parhankangas, J. & Vesanto, J. *SOM Toolbox [online]*—<http://www.cis.hut.fi/projects/somtoolbox>. (2000).
54. Wolda, H. Similarity indices, sample size and diversity. *Oecologia* **50**, 296–302 (1981).
55. Reitz, E. J. & Wing, E. S. *Zooarchaeology*. (Cambridge University Press, 2008).
56. Matthews, W. J. *Patterns in freshwater fish ecology*. (Chapman and Hall, 1998).
57. Fedora, M. A. & Beschta, R. L. Storm runoff simulation using an antecedent precipitation index (API) model. *Journal of Hydrology* **112**, 121–133 (1989).
58. Shaw, E. M. *Hydrology in Practice (third edition)*. 349 (Taylor & Francis, 1994).

## Acknowledgements

This study was supported financially by a DFG grant (FO 301/15-1, FO 301/15-2, WU 749/1-1, WU 749/1-2) (N. Wu, C. Faber), AIAS CO-FUND funding (N. Wu) and China Scholarship Council (CSC) (X. Sun, Y. Qu). We would like to thank Dr. Jessica Barker for the English polishing, Mrs. Monika Westphal, Mrs. Bettina Hollmann and student assistances for their support during the field campaigns. B. Messyasz, M. Kokociński and C. Wetzel helped greatly with the diatom identification. The constructive comments of two anonymous reviewers greatly improved our manuscript.

## Author Contributions

N.W., C.F., T.R., U.U. and N.F. developed the basic ideas. N.W., C.F., X.S. and S.I. performed the sampling collection and processing. N.W., Y.Q. and C.W. performed the data analyses and calculation. N.W. wrote the main manuscript. All authors saw the manuscript before submission.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Wu, N. *et al.* Importance of sampling frequency when collecting diatoms. *Sci. Rep.* **6**, 36950; doi: 10.1038/srep36950 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016