

Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems biomonitoring

François Keck Frédéric Rimet Alain Franc Agnès Bouchez

January 14, 2016

Abstract

Diatoms include a great diversity of taxa and are recognized as powerful bioindicators in rivers. However using diatoms for monitoring programs is costly and time consuming because most of the methodologies necessitate species-level identification. This raises the question of the optimal tradeoff between taxonomic resolution and bioassessment quality. Phylogenetic tools may form the bases of new more efficient approaches for biomonitoring if relationships between ecology and phylogeny can be demonstrated. We estimated the ecological optima of 127 diatom species for 19 environmental parameters using count data from 2119 diatom communities sampled during 8 years in eastern France. Using uni- and multivariate analyses, we explored the relationships between freshwater diatom phylogeny and ecology (*i.e.* the phylogenetic signal). We found a significant phylogenetic signal for many of the ecological optima that were tested, but the strength of the signal varied significantly from one trait to another. Multivariate analysis also showed that the multidimensional ecological niche of diatoms can be strongly related to phylogeny. The presence of clades containing species that exhibit homogeneous ecology suggests that phylogenetic information can be useful for aquatic biomonitoring. This study highlights the presence of significant patterns of ecological optima for freshwater diatoms in relation to their phylogeny. These results suggest the presence of a signal above the species level, which is encouraging for the development of simplified methods for biomonitoring survey.

This is a post-print version of an article originally published in *Ecological Applications*. ([link editor](#)). Please cite: Keck F., Rimet F., Franc F. & Bouchez A. (In press) Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems biomonitoring. *Ecological Applications*.

Contents

1	Introduction	2
2	Material and Methods	4
2.1	Genetic database	4
2.2	Phylogeny	4
2.3	Ecological optima	5
2.4	Phylogenetic signal	5
2.5	Phylogenetic Principal Component Analysis	8
2.6	Statistical Packages	8
3	Results	8
3.1	Ecological and Phylogenetic data	8
3.2	Phylogenetic signals	10
3.3	Phylogenetic principal component analysis	10
4	Discussion	12
4.1	Phylogenetic signal in ecological niches of diatoms	12
4.2	Ecological optima patterns in the phylogeny	13
4.3	Implications for biomonitoring	14
5	Acknowledgments	15
	References	16

1 Introduction

The monitoring of water quality using biological indicators is commonly used by environmental managers worldwide (Ibáñez et al., 2010). This approach is based on the view that biological communities' composition is directly affected by environmental conditions and human activities (Chapman, 1996). Algae and especially diatoms, a particular group of microalgae, are well suited to monitor environmental health, because they include groups exhibiting a wide range of ecological optima and their communities respond rapidly to changes in habitat quality (Lowe and Pan, 1996; Stevenson and Smol, 2003). Despite these advantages, ecological assessments using diatoms are less common than those using macroinvertebrates or fishes (Gallacher, 2002; Carter et al., 2006; Resh, 2007). This may be related to the huge diversity of diatoms, a group of algae containing approximately 12 000 described species (Guiry, 2012) and potentially 100 000 existing ones (Mann and Vanormelingen, 2013). Such diversity makes microscopic identification at species-level a challenge (Kelly, Penny, et al., 1995; Kociolek, 2005; Besse-Lototskaya et al., 2006). In fact, the majority of diatom indices is based on species or sub-species levels (Rimet, 2012) and requires highly qualified staff and resources.

Two ways have been explored to avoid this problem. First, some authors have proposed to reduce the number of species used to calculate bioassessment indices (Lenoir and Coste, 1996; Lavoie, Dillon, et al., 2009). In the same vein, results of DeNicola (2000) also suggested that resolution of taxonomic synonymies may reduce

the number of species to evaluate. Secondly, it was proposed to work at a lower taxonomic resolution. Studies comparing bioassessment performances of different taxonomic levels showed that some higher taxa (in particular genus level) might be relatively precise and efficient compared to species level (Kelly, Penny, et al., 1995; Chessman et al., 1999; Grown, 1999; Hill et al., 2001; Wunsam, Cattaneo, et al., 2002; Raunio and Sojinen, 2007; Rimet and Bouchez, 2012b). However, some authors stressed the risk of oversimplification and emphasized the importance to work at the most precise taxonomic level (Round, 1991; Patrick and Palavage, 1994; Kocielek, 2005; Ponader and Potapova, 2007; Bennett et al., 2014). Thus, it is still unclear what constitutes an appropriate group for biomonitoring. An alternative and interesting way might be to use mixed taxonomic levels, as suggested by Jones (2008) for macroinvertebrates. The idea behind this is to adapt the taxonomic detail to the information content for each clade. For diatoms, moving from genus to species may require much more effort and special training. Reducing, where possible, the level of taxonomic identification, could both reduce cost and errors. Bioassessment protocols should be able to benefit from such approaches, but this raises methodological questions about how to select biomonitoring groups and to test their efficiency.

Recently, new frameworks have been proposed to link species assemblage, species traits and phylogeny (Webb et al., 2002; Mouquet et al., 2012). The integration of phylogenetics and community ecology aims to disentangle the role of phylogeny in species interactions, community structures and ecological processes. But such an approach could also be interesting for conservation purposes and applied ecology, especially when managers have to deal with complex biological indicators with important diversity, complicated taxonomy and arduous determination. The set of tools provided by phylogenetics seems promising for exploring the ecological variability within and between taxonomic levels and to make new proposals for efficient biomonitoring groups (Carew et al., 2011; Larras et al., 2014). However, using phylogeny in an environmental assessment context requires the ability to demonstrate the relationships existing between phylogenetic position of species and their ecology (Carew et al., 2011). This can be assessed by analysis of the phylogenetic signal, since there is a tendency for closely related species to possess more similar trait values than more distantly related species. For instance, Buchwalter et al. (2008) explored the phylogenetic signal for physiological traits related to cadmium bioaccumulation, compartmentalization and susceptibility for 21 macroinvertebrates species. Carew et al. (2011) developed a similar approach, but focused on more general traits reflecting chironomids and mayflies' sensitivity to pollution, and Larras et al. (2014) focused on the sensitivity of 14 diatoms species to 4 different herbicides. These studies showed the existence of phylogenetic signals that were more or less pronounced depending on the considered trait and biological group. Above all, they highlighted the potential of including a phylogenetic component in the development of ecotoxicological and bioassessment tools. The next step in the integration of phylogeny into biomonitoring tools using diatom communities is to pursue efforts for characterizing, measuring and testing phylogenetic signal on large databases including a significant number of species and their ecological optima.

Here, we analyze jointly the phylogeny and the ecology of 127 freshwater diatom taxa included in river biomonitoring tools. The aim of this study was to take a

preliminary view of phylogenetic signals of diatoms' ecological optima for a variety of environmental parameters. These parameters (related to mineralization, organic matter, nutrients, *etc.*) mainly reflect the chemical status of water bodies. They define, at least partially, the fundamental habitat niche of diatoms and are recognized to influence directly the water quality and the ecological potential of freshwaters. We provide a phylogenetic tree based on 18S and *rbcL* markers for 127 species. For all these species, we estimated ecological optima for 19 different environmental parameters affecting the quality of freshwaters. Using univariate and multivariate approaches we explore the phylogenetic signal and discuss the implications of our findings for biomonitoring.

2 Material and Methods

2.1 Genetic database

We collected DNA sequences of diatoms from the open access GenBank database (NCBI 2013) and the Thonon Culture Collection (TCC, INRA 2013), which provides a nucleotide sequences database for indexed micro-algae species. We focused on the nuclear gene coding for the small subunit 18S rRNA and the chloroplast *rbcL* gene coding for the RuBisCO enzyme. These markers are among the most popular and they are available for an important diversity of diatom species (Theriot et al., 2011). DNA sequences of 18S and *rbcL* genes were extracted from GenBank using **Geneious Basic 5.6.3** software with the following query: (18S OR *rbcL*) AND (diatom OR Bacillariophyta). The query was rerun every two months in 2013 to keep the database updated. Data extracted from GenBank and TCC were manually cleaned. Each sequence was blasted against the GenBank database to check for taxonomic inconsistencies. In case of problems, we checked for taxonomic synonyms using the Catalogue of Diatom Names (Fourtanier and Kociolek, 2011) and if the error persisted, we dropped the sequence. After cleaning, our database contained 1236 sequences for the 18S gene and 1084 sequences for the *rbcL* gene.

2.2 Phylogeny

Many species were sequenced several times from different strains and sequences were derived from many different sources. Therefore, the phylogenetic reconstruction was divided into different steps presented in Figure 1. Nucleotide sequences were first aligned using the **Muscle** algorithm (Edgar, 2004) provided through the **SeaView** graphical user interface (Gouy et al., 2010). The 18S alignment presented 634 complete sites (without gaps or undetermined nucleotides) and the *rbcL* 467 complete sites. Tree topologies and branch lengths were computed separately for the two markers with maximum likelihood method (ML) in **PhyML 3.0** (Guindon et al., 2010). We used the **MrAIC** software (Nylander, 2004) to select the best substitution models using the Akaike Information Criterion (AIC). After reconstruction, each tree was pruned so that there was only one remaining strain per species or subspecies. Some species appeared to be polyphyletic, which can result from taxonomic assignment errors, inaccurate tree reconstruction or a mismatch between taxonomy and phylogeny. To limit the impact of such problems, pruning was done manually with the following rules: for species with more than one available strain, isolated strains –

in the phylogenetic tree – or sequences not published in peer reviewed journals were deleted first. When only a monophyletic group remained, we selected one of the strains; prioritizing the working scale of the source publication (a study focusing on a specific genus is preferred to a broad taxonomic range study). After pruning, 411 unique species or subspecies remained on the 18S phylogenetic tree and 425 for the *rbcL* phylogenetic tree (the accession numbers of the selected sequences are provided in Appendix A, Table A1). However, 18S and *rbcL* taken separately have a limited ability to recover phylogenetic relationships within diatoms (Theriot et al., 2011). Therefore, previously selected 18S and *rbcL* sequences were concatenated for those species for which both data were available (286 species). This alignment was used to estimate a more robust and accurate tree (Tree_\cap) using a partitioned ML analysis (20 runs; 100 bootstraps) in RAxML 7.2.8 (Stamatakis, 2006). Finally, 18S and *rbcL* sequences were concatenated even for species for which only one marker was available (550 species). This alignment was used to estimate a tree with a higher diversity (Tree_\cup) using a partitioned ML analysis (20 runs; 100 bootstraps) with Tree_\cap as a topological constraint in RAxML. All the trees were dated in relative time using a semi-parametric method based on penalized likelihood (Sanderson, 2002).

2.3 Ecological optima

Ecological optima for each species were estimated from results of biomonitoring surveys conducted as part of the French program Réseau de Contrôle de Surveillance (RCS). Samples were collected in rivers and streams in eastern France between 2001 and 2008. The database connects the diatoms relative species abundance (estimated by counting 400 individuals) to the chemical and physical conditions for 2119 samples. Environmental parameters and their abbreviations are presented in Table 1. For each species and each environmental parameter, we calculated an optimum value (*i.e.* the species optimum for the parameter). We chose the weighted averaging method (WA, Ter Braak and Looman, 1986) to estimate species optima. For a given species (k), a given environmental parameter (x) and a set of sampled sites ($i = 1, \dots, n$), WA corresponds to the mean of the parameter x weighted by y_{ik} , the abundance of the species k at each corresponding sampled site (Equation 1). This is a simple, reliable and robust statistic with a long tradition of being used in diatom ecology (*e.g.* Salden, 1978; Charles, 1985; Oksanen et al., 1988), paleolimnological studies (Birks, 2010) and biomonitoring (Ector and Rimet, 2005).

$$\text{WA}_k = \frac{\sum_{i=1}^n y_{ik} x_i}{\sum_{i=1}^n y_{ik}} \quad (1)$$

To promote statistical robustness, we chose to estimate WA only for species occurring in at least 10 different samples. Diatoms counts have been log-transformed and environmental parameters have been log-transformed and standardized before WA computation.

2.4 Phylogenetic signal

Phylogenetic signal is “the tendency for related species to resemble each other more than they resemble species drawn at random from the tree” (Blomberg and Garland, 2002). From a statistical point of view, this is non-independence of traits values

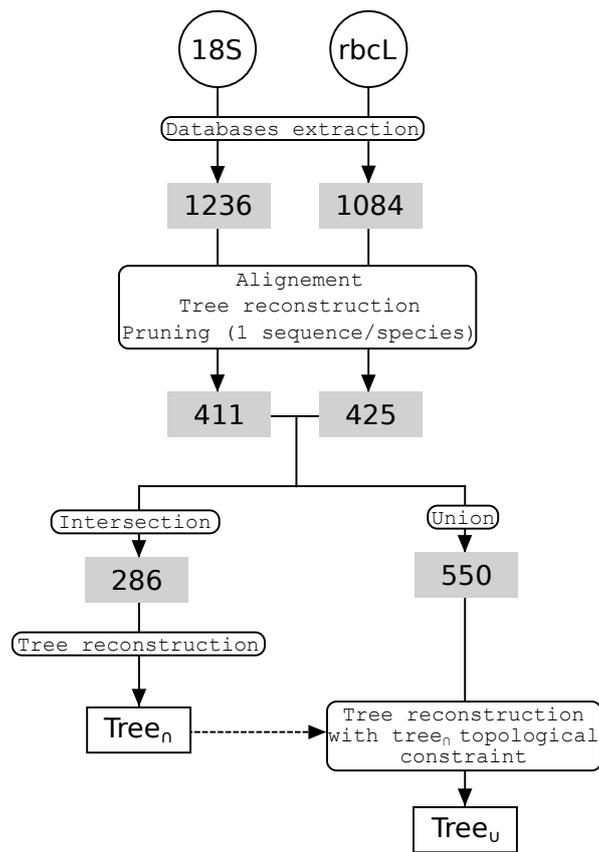


Figure 1: Flowchart describing the main steps of phylogenetic reconstruction. Numbers in grey frames represent the number of sequences available after each operation.

	Unit	Min	1 st Quartile	Median	Mean	3 rd Quartile	Max
NO ₃ ⁻	mg.L-1	0.100	2.660	4.800	6.442	8.500	69.700
NO ₂ ⁻	mg.L-1	0.010	0.020	0.053	0.126	0.120	2.960
NH ₄ ⁺	mg.L-1	0.028	0.050	0.070	0.269	0.150	17.200
NKJ	mg.L-1 (N)	0.213	0.703	1.000	1.049	1.000	14.200
PO ₄ ³⁻	mg.L-1	0.000	0.030	0.128	0.344	0.323	13.100
TP	mg.L-1 (P)	0.000	0.020	0.096	0.180	0.205	5.674
DOC	mg.L-1 (C)	0.200	1.500	2.600	2.867	3.663	19.720
DOM	mg.L-1	1.000	3.600	7.425	16.783	14.317	3890.000
BOD	mg.L-1	0.500	0.900	1.867	1.985	2.750	20.000
O ₂ p	%	0.000	83.600	93.500	91.369	101.000	202.200
Na ⁺	mg.L-1	0.000	5.200	10.300	31.661	22.050	863.500
Cl ⁻	mg.L-1	0.000	7.900	15.350	51.844	31.050	1411.000
Ca ²⁺	mg.L-1	0.000	45.000	76.000	82.603	103.000	478.000
K ⁺	mg.L-1	0.000	1.300	2.400	4.165	4.600	224.000
Mg ²⁺	mg.L-1	0.000	3.750	6.800	12.863	14.038	136.600
SO ₄ ²⁻	mg.L-1	0.000	12.533	31.700	89.182	84.725	1380.000
Conductivity	µS.cm-1	18.700	334.000	464.500	615.848	656.250	4818.000
pH	-	6.300	7.653	7.900	7.828	8.100	8.800
Temperature	°C	0.200	14.337	16.720	16.166	18.700	30.000

Table 1: Summary of the environmental parameters present in the dataset. NKJ = Kjeldahl nitrogen; TP = total phosphorus; DOC = dissolved organic carbon; DOM = dissolved organic matter; BOD = biological oxygen demand; O₂p = pressure of oxygen.

among species as a direct consequence of their phylogenetic relatedness. They are numerous ways to measure and test the phylogenetic signal described in the literature (see Revell et al., 2008). Münkemüller et al. (2012) reviewed some of the most popular indices to measure the phylogenetic signal and showed that the performance of these statistics can vary strongly, depending on the size and structure of the phylogeny, and the complexity of traits evolution model.

In this study, we estimated the phylogenetic signal for ecological optima using Pagel’s λ (Pagel, 1999) and Abouheif’s C_{mean} (Abouheif, 1999). Pagel’s λ ranges between 0 (no signal, traits are distributed randomly) and 1 (signal, traits are distributed following a Brownian motion model). It is a branch-length transformation method, estimated by maximum likelihood and tested by likelihood ratio test (Pagel, 1999). Abouheif’s C_{mean} is an alternative way to test for phylogenetic signal. Contrary to Pagel’s λ , it does not rely on an evolutionary model but measures autocorrelation among tips with a particular matrix of phylogenetic proximities (Pavoine et al. 2008). It can be tested with randomized permutations. Both Pagel’s λ and Abouheif’s C_{mean} have been singled out by Münkemüller et al. (2012) as powerful and reliable methods to measure and test the phylogenetic signal and have been previously used on ecological traits (Freckleton et al., 2002; Comte et al., 2014).

Since phylogenetic signal measurement strongly depends on the phylogenetic tree used as input, the robustness of the results was assessed by repeating analyses with 100 bootstrap trees. This procedure allows getting a distribution of signal measures and tests with respect to both the topology and branch lengths.

2.5 Phylogenetic Principal Component Analysis

Phylogenetic principal component analysis (pPCA) extends the classical PCA to the analysis of phylogenetic structures in biological traits (Jombart, Pavoine, et al., 2010). This is a dimension-reduction method which aims to uncover the main phylogenetic structures of a set of traits by finding combinations of them that are phylogenetically autocorrelated. To this end, Jombart, Pavoine, et al. (2010) defined two types of phylogenetic patterns: global and local structures. Global structures occur at large scales in the phylogeny and define general patterns of traits similarities among taxa. They exhibit positive autocorrelation and are directly related to the idea of phylogenetic signal. By contrast, local structures occur in specific parts of the phylogeny between closely related taxa which exhibit strong variability. Local structures are characterized by negative autocorrelation and reflect traits overdispersion (*i.e.* antesignal). We applied pPCA directly on the standardized matrix of ecological optima to inspect Global and Local structures along the phylogeny with respect to the complete set of environmental variables. To this end, Global and Local scores of species were mapped onto the phylogeny to uncover interesting patterns. A species with a high score on a Global principal component shares similar traits values with its neighbors. On the other hand, a high score on a Local principal component means strong differences with its neighbors. We also used lagged scores (*i.e.* for one species, the mean value of its neighbors weighted by their distance). Lagged scores smooth scores for a given component and make visualizing patterns easier.

2.6 Statistical Packages

We performed all the statistical analyses with R 3.0.2 software (R Development Core Team, 2013). Phylogenies were handled and dated with the `ape` package (Paradis et al., 2004). Pagel’s λ statistics and tests were computed with the package `phytools` (Revell, 2012). The pPCA and Abouheif C_{mean} tests were performed using the `adephylo` package (Jombart, Balloux, et al., 2010).

3 Results

3.1 Ecological and Phylogenetic data

In the 2119 sampled sites, 909 different diatoms species (or varieties) were identified. Using only species occurring in more than 10 different sites restricted the number of taxa to 398. Finally, crossing ecological dataset with the phylogenetic tree ($Tree_{\cup}$) led to a subset of 127 species for the analyses. All the following results are based on this subset. For both 18S and *rbcL* markers, the most complex evolutionary model (*i.e.* GTR+I+G) was selected according to AIC. The final phylogenetic tree ($Tree_{\cup}$) based both on 18S and *rbcL* but representing only the set of species included in the signal analyses (127 species) is presented in Figure 2 (and with bootstrap support values in Appendix B, Figure B1) whereas the complete $Tree_{\cup}$ (550 species) and $Tree_{\cap}$ (286 species) are provided in Appendix B, Figures B2 and B3. The most recent branches are well supported by bootstraps overall whereas some deep branches exhibit low support values.

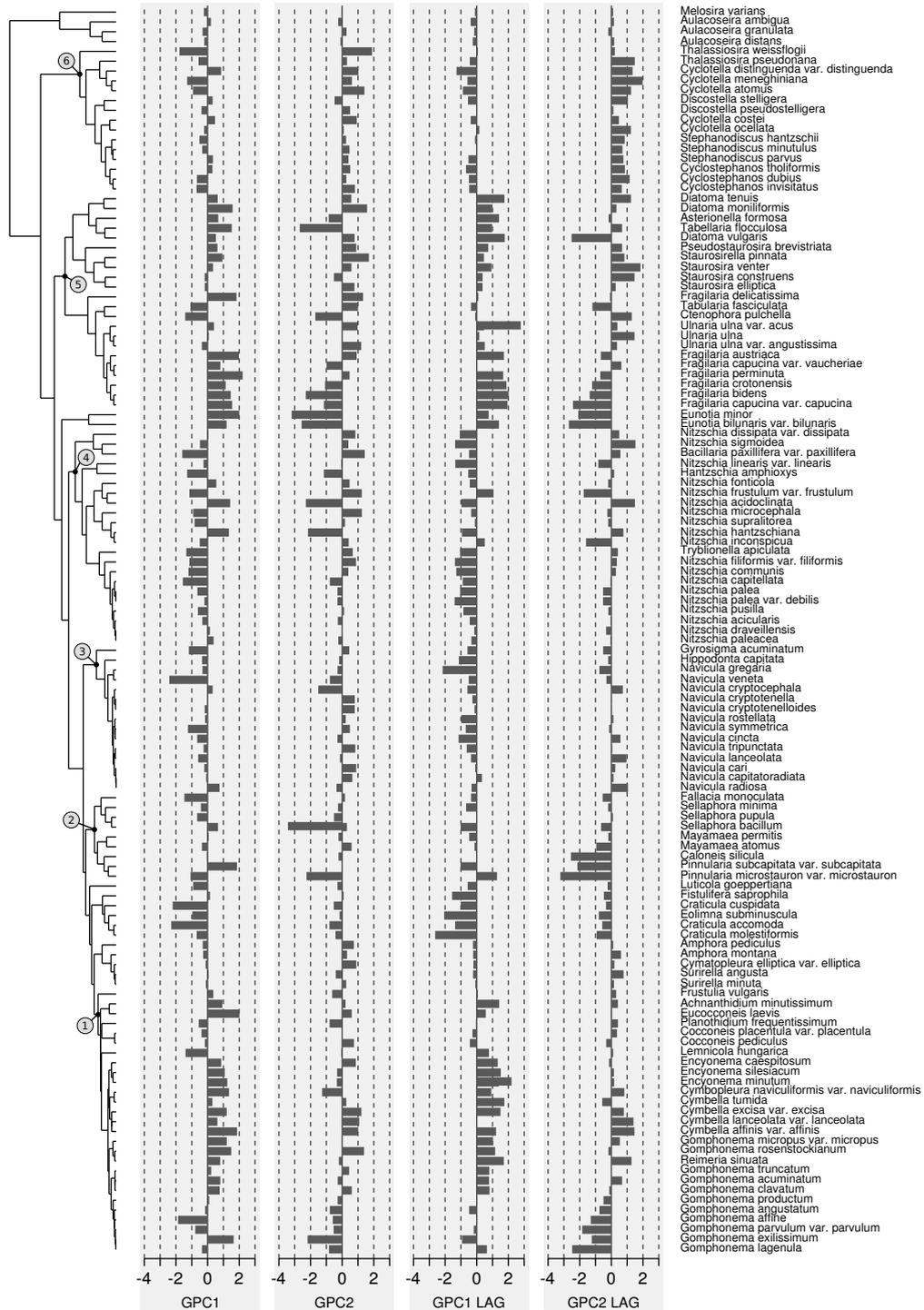


Figure 2: Phylogenetic tree reconstructed by maximum likelihood from 18S and *rbcL* DNA sequences for 127 different diatom taxa. The bars give the score of each taxon on the 1st (GPC1) and 2nd (GPC2) Global Principal Component of the pPCA. Lag vector scores for each of these components are also provided (GPC1 LAG and GPC2 LAG). The different clades discussed in the body of the text are directly labeled on the tree.

	λ	p-value	C_{mean}	p-value
NO_3^-	0.740	<0.001	0.404	<0.001
NO_2^-	0.116	1	0.185	0.002
NH_4^+	0.000	1	0.158	0.012
NKJ	0.000	1	0.057	0.149
PO_4^{3-}	0.667	<0.001	0.370	<0.001
TP	0.647	<0.001	0.352	<0.001
DOC	0.682	<0.001	0.330	<0.001
DOM	0.561	<0.001	0.286	<0.001
BOD	0.514	<0.001	0.266	<0.001
O_2P	0.684	<0.001	0.352	<0.001
Na^+	0.784	0.001	0.238	<0.001
Cl^-	0.776	<0.001	0.275	<0.001
Ca^{2+}	0.697	0.008	0.191	0.005
K^+	0.653	<0.001	0.286	<0.001
Mg^{2+}	0.713	0.016	0.146	0.012
SO_4^{2-}	0.674	0.026	0.159	0.003
Conductivity	0.723	0.002	0.206	0.003
pH	0.545	1	0.167	0.009
Temperature	0.250	0.356	0.17	0.009

Table 2: Measurements and tests of the phylogenetic signal for 19 ecological optima with two methods (Pagel’s λ and Abouheif’s C_{mean}). Statistics are in bold if their p-value < 0.05. For abbreviations references see Table 1.

3.2 Phylogenetic signals

We found statistical evidence of phylogenetic signal for many tested traits (Table 2 and Figure 3). Except for SO_4^{2-} and Mg^{2+} , phylogenetic signal measures and tests appeared to be robust with respect with phylogenetic tree reconstruction uncertainty (Figure 3). Using the most likely tree, both Pagel’s λ and Abouheif’s C_{mean} led to the conclusion of absence of a signal for NKJ. The Pagel’s λ also concluded that there was an absence of phylogenetic signal for NO_2^- , NH_4^+ , pH and temperature. With both methods the strength of the phylogenetic signal was found to be variable depending on the trait considered. The Pagel’s λ scaled between <0.001 (NKJ and NH_4^+) and 0.784 (Na^+) whereas Abouheif’s C_{mean} scaled between 0.057 (NKJ) and 0.404 (NO_3^-).

3.3 Phylogenetic principal component analysis

We kept the two first positive axes of the pPCA, which expressed the presence of a very strong Global structure (Figure 4A). By contrast, Local structure (negative eigenvalues) was very small and a test for negative autocorrelation on all variables separately found no statistical support for such pattern (p-values ≥ 0.875). Thus, we do not include results on Local structure in this paper. The first global principal component (GPC1) reflects essentially a gradient of organic matter, nutrients and minerals on the one hand and oxygen on the other (Figure 4B). Species values for this component clearly exhibited patterns in the phylogeny (Figure 2, GPC1 and GPC1-Lag). Clusters 3, 4 and 6 are related to high values of organic matter, nutrients and minerals and low value of oxygen, while clusters 1 and 5 seem to characterize the opposite kind of water media. The second global principal component (GPC2) discriminates nutrients from minerals and also expresses the strong negative correlation between nutrient levels and oxygen in the water (Figure 4B). GPC2 clarified information given by GPC1, in particular that cluster 2 is more related to

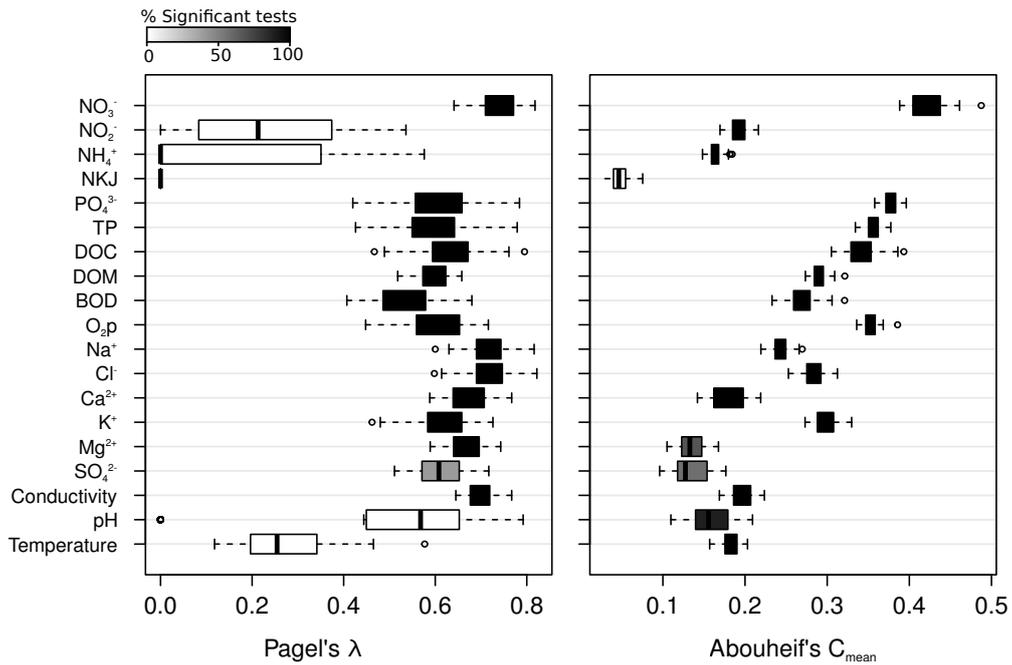


Figure 3: Phylogenetic signal measured and tested for 19 ecological optima with two methods (Pagel's λ and Abouheif's C_{mean}) over 100 bootstrap trees. Boxes summarize the distribution of statistic values for the 100 phylogenetic trees. Boxes color indicates the percentage of associated significant tests (p-value < 0.05) over the 100 trees.

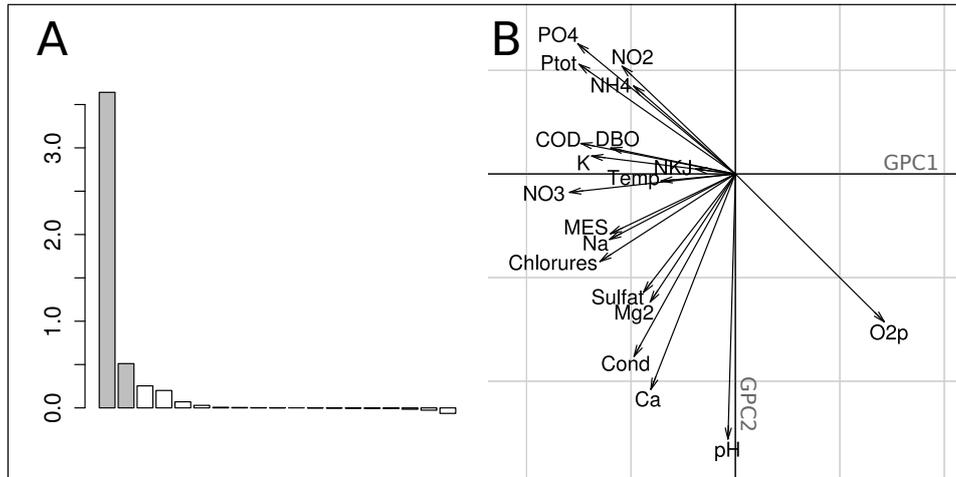


Figure 4: **A.** Eigenvalues extracted by the pPCA. **B.** Loadings for the 1st (GPC1) and 2nd (GPC2) Global Principal Component of the pPCA (grid mesh = 0.2). For abbreviations reference see Table 1.

nutrients than cluster 6 (Figure 2, GPC2 and GPC2-Lag).

4 Discussion

4.1 Phylogenetic signal in ecological niches of diatoms

Ecologists are increasingly interested in the evolutionary perspective that sister species are ecologically similar (Losos, 2008). This has obvious implications for ecological studies and applied monitoring (Wiens et al., 2010). In the case of freshwater diatoms, most of the ecological optima showed a strong phylogenetic signal using Pagel's λ , Abouheif's C_{mean} and pPCA. Ecological optima which exhibit the strongest signal can be divided in three components. The first is ions (Na^+ , K^+ , Cl^- , Ca^{2+}), which directly impact the salinity, the alkalinity and the conductivity of water, the second is inorganic nutrients (PO_4^{3-} , TP, NO_3^-) and the third is indicators of organic matter (DOC, DOM, BOD). These traits, which seem to be well conserved throughout evolution, are also known to be important factors in determining diatom assemblages (Van Dam et al., 1994; Leland and Porter, 2000; Kelly, 2003; Rimet, 2009) as they are involved in environmental filtering and species competition. But none of these studies made a connection between ecological traits and phylogeny. It is possible that a number of the ecological preferences we studied here are to some extent correlated with species' cell/body size which is well conserved in diatom phylogeny (Nakov et al., 2014). However, the impact of cell size on diatoms ecological optima is not clear. Previous studies have shown contrasting results (Lavoie, Lento, et al., 2010; Berthon et al., 2011).

Conversely, some parameters showed low or no phylogenetic signal. This was the case for three of the four forms of nitrogen (NKJ , NO_2^- and NH_4^+), but also pH and temperature. At this point, it seems difficult to formulate any evolutionary hypothesis about the lability of these traits. As far as we know, they cannot be

considered as neutral: NKJ , NO_2^- and NH_4^+ are important sources of nitrogen for diatoms (Schoeman, 1973; Patrick, 1977; Berges et al., 2002), while pH and temperature are also known to be important factors in the determination of communities' composition (e.g. Renberg and Hellberg, 1982; Coste and Ector, 2000; Zampella et al., 2007). The lack of signal may be the consequence of overdispersion due to recent evolutionary events like convergent evolution or character displacement. For NKJ, the signal is probably blurred by the fact that nitrogen heterotrophic species given in Patrick (1977) are rare and dispersed in the phylogeny (e.g. *Gomphonema parvulum* in cluster 1, *Nitzschia frustulum*, *N. fonticola* and *N. palea* in cluster 4 and *Cyclotella meneghiana* in cluster 6). It has also been shown that very closely related taxa (cryptic and pseudo-cryptic species) can exhibit ecological niche differentiation through exclusive competition when co-occurring in sympatry (Vanellander et al., 2009).

Methodological and data artifacts may also, at least partially, explain the absence of signal for ecological parameters. It is a possibility that the very restricted range of NO_2^- and NH_4^+ values in the database (with many of them below the detection threshold) make the detection of the signal impossible. The absence of a signal for pH is surprising but may also be due to a restricted range of variation in the database, since pH values range from 7.04 to 8.3 in the 5th-95th percentiles interval. Acid rivers are not well represented in this dataset and it implies that acidophilic diatoms are under-represented in the analyses (e.g. only two *Eunotia* and one *Frustulia* species). Additional work including more acidophilic species should be done to investigate further the phylogenetic signal of pH optima. Finally, temperature measurements in the field are often not representative of the real river temperature, but mostly on the season and the time of the day when the sampling was done. Nonetheless, it is always difficult to interpret the phylogenetic signal from an evolutionary perspective (Revell et al., 2008) and, as stressed by Losos (2008), phylogenetic signal is a necessary condition, but it does not systematically imply phylogenetic niche conservatism with active constraints.

4.2 Ecological optima patterns in the phylogeny

Traits of interest may be more conserved in some clades compared to others. Despite the large diversity (more than 6800 taxa) of the epipellic motile Naviculaceae (Round et al., 1990), for instance, this family is considered a robust indicator of slightly eutrophic conditions in generic diatom indices (e.g. Rumeau and Coste, 1988). Accordingly, our analyses showed that species of cluster 3, which corresponded to the Naviculaceae, displayed exactly this kind of ecology. Other clusters, as cluster 1 and 5 showed also a homogeneous ecology, but on the opposite side of the gradient (rivers with low nutrients and organic matter concentrations). Cluster 5 encompasses araphid diatoms, which are non-motile, most of them being attached to hard substrates by a short stalk (*Ulnaria*, *Fragilaria*, *Diatoma*, *Staurosira*) (Round et al., 1990; Rimet and Bouchez, 2012a). Similarly, in cluster 1, many genera also can be attached to substrata by stalks (*Gomphonema*, *Cymbella*) or live in mucous tubules attached to substrates (*Encyonema*). These different life-forms confer very different advantages, especially, a better competition efficiency for nutrient uptake (Passy, 2007; Berthon et al., 2011), and generic diatom indices consider them as indicator of low pollution level (Rumeau and Coste, 1988; Wu, 1999). Here again, taxa in the

tree represented homogeneous ecology.

In contrast, other clades exhibited phylogenetic and ecological signals which are much fuzzier, especially along the 2nd axis of the pPCA. Cluster 4, for example, is composed of taxa of the Bacillariaceae. This family is also very diverse taxonomically and only a few of the freshwater taxa are represented in our study. The genus *Nitzschia*, which is well represented within clade 4, has recently been reported to be polyphyletic (Rimet, Kermarrec, et al., 2011). While generic diatom indices consider this epipelagic motile genus (Round et al., 1990) as an indicator of polluted rivers (Rumeau and Coste, 1988; Wu, 1999), the clade including *Nitzschia* also encompasses several marine planktonic genera (*e.g.* *Pseudo-nitzschia*, *Cylindrotheca*, *Fragilariopsis*, but pruned before the statistical analyses) that are very distinct from *Nitzschia* in ecology and morphology. Moreover, cluster 4 encompasses a paraphyletic species (*Nitzschia inconspicua*), which showed different salinity responses in culture depending on the genotype (Rovira, 2013). Such studies are unfortunately too rare, but explain why ecological variability can be observable in the phylogeny.

Both Pagel’s λ and Abouheif’s C'_{mean} led to very similar conclusions about the phylogenetic signal. Despite low support of deep branches, both methods appeared to be robust in bootstrap analyses. The more accurate the phylogeny, the better results are likely to be. However in the present case, it seems like the phylogenetic signal can be detected even if deep nodes are hard to resolve. Thus, we argue that emphasis should be placed on maximizing the number of species in the phylogeny. It is important to have as much diversity as possible with a good representation of every clade to correctly estimate phylogenetic signal and to extract interesting subgroups (Losos, 2008).

4.3 Implications for biomonitoring

The existence of a phylogenetic signal for ecological traits is promising for the development of simpler bioassessment protocols. Using species-level analyses in environmental monitoring, when the phylogenetic signal is strong at higher level, means that a part of the extracted information is redundant (Carew et al., 2011). Avoiding such redundancy can probably save time and effort. Here, the environmental parameters which exhibit the strongest signal are of particular relevance in a context of ecological monitoring for freshwaters. Diatom indices have been historically developed to monitor organic pollution (Sládeček, 1986; Kobayasi and Mayama, 1989) and eutrophication (Kelly and Whitton, 1995; Torrisi and Dell’Uomo, 2006). Interestingly, the positions of diatom species on gradients of organic matter and nutrients seem to be highly dependent on the phylogeny. Moreover, the loadings of the first component of the pPCA clearly describe a gradient of general water pollution, whereas the second component discriminates factors associated with eutrophication. The fact that phylogeny is linked, at least partially, to the distribution of ecological traits between species should enable the development of simplified biomonitoring methods by merging species into categories of known ecological values. A simple way to achieve this would be to use taxonomical groups, assuming that phylogeny and taxonomy are strongly related (even if we showed that some taxa were polyphyletic). An obvious advantage of taxonomy is that there is no need to reconstruct a phylogeny to test the depth of the signal; standard nested ANOVA can be used directly with taxonomic categories (Gittleman and Luh, 1992). Furthermore, a tool based

on the currently recognized taxonomy should be more readily and widely adopted by environmental managers and field operators. On the other hand, phylogeny can be more flexible and in some cases can fit better to ecological traits than taxonomy. We showed this for the genus *Nitzschia*, which is probably polyphyletic. Another typical case here is the genus *Cyclotella* that we found divided in two clades with different optimums. The ecological heterogeneity of the *Cyclotella* species included in the analyses is recognized: *Cyclotella costei* and *C. ocellata* are commonly found in oligotrophic waters (Wunsam, Schmidt, et al., 1995; Rimet, Druart, et al., 2009) whereas *C. meneghiniana* and *C. atomus* are associated with high nutrient concentrations and eutrophic waters (Weckström and Juggins, 2006; Duong et al., 2007).

The other interesting prospect that arises when phylogenetic signal is strong enough is the possibility to use phylogeny to make predictions for new species. It can be simpler nowadays to sequence a species for an individual marker than performing an experimental estimation of its ecological optima (Shokralla et al., 2012). Knowing the DNA sequence of a given species, it is feasible to estimate its position in a phylogenetic tree and to estimate any kind of traits value from the knowledge we have of other related species (Guénard, Ohe, et al., 2011; Guénard, Legendre, et al., 2013). Such approaches have been only tested on macroinvertebrates and under laboratory conditions. Thus, it is not yet apparent whether results can be extended to the natural habitat but since diatoms are extremely diverse, it is impossible to measure traits for every species and models could be of great help. Moreover many species are rare and the estimation of their ecological profile can be underpowered and strongly biased by very low relative abundances in sampling data. With the democratization of molecular biology, such modeling approaches appear promising in ecotoxicology (Guénard, Ohe, et al., 2011; Larras et al., 2014), but could probably be relevant in many other ecological applications.

Both ecological prediction and biomonitoring simplification require strong phylogenetic signal. We demonstrated here that such signal exists in diatoms. However, statistical patterns in the phylogeny are not sufficient to guarantee the efficiency of these methods in routine analysis. Moreover, intraspecific variability and phenotypic plasticity may degrade the signal. Further investigations will be necessary with additional species and data for validation. There is no doubt that the development of large ecological traits databases and large genetic databases and phylogenetic trees for diatoms will permit such exciting applications in the foreseeable future.

5 Acknowledgments

We thank Sylvain Guyot for his scientific assistance and Elliot Shubert for language editing. The authors would like to thank three anonymous reviewers for their very helpful comments. This program was funded by ONEMA (French National Office for Water and Aquatic Ecosystems) in the context of the 2013-2015 “Phylogeny and Bioassessment” program. UMR CARRTEL is member of the ECOTOX INRA network. UMR CARRTEL and UMR BIOGECO are members of the R-Syst INRA network.

References

- Abouheif, E. (1999). “A method for testing the assumption of phylogenetic independence in comparative data”. In: *Evolutionary Ecology Research* 1.8, pp. 895–909.
- Bennett, J. R., D. R. Sisson, J. P. Smol, B. F. Cumming, H. P. Possingham, and Y. M. Buckley (2014). “Optimizing taxonomic resolution and sampling effort to design cost-effective ecological models for environmental assessment”. In: *Journal of Applied Ecology* 51.6, pp. 1722–1732.
- Berges, J. A., D. E. Varela, and P. J. Harrison (2002). “Effects of temperature on growth rate, cell composition and nitrogen metabolism in the marine diatom *Thalassiosira pseudonana* (Bacillariophyceae)”. In: *Marine Ecology Progress Series* 225, pp. 139–146.
- Berthon, V., A. Bouchez, and F. Rimet (2011). “Using diatom life-forms and ecological guilds to assess organic pollution and trophic level in rivers: a case study of rivers in south-eastern France”. In: *Hydrobiologia* 673.1, pp. 259–271.
- Besse-Lototskaya, A., P. F. M. Verdonshot, and J. A. Sinkeldam (2006). “Uncertainty in diatom assessment: sampling, identification and counting variation”. In: *Hydrobiologia* 566.1, pp. 247–260.
- Birks, H. J. B. (2010). “Numerical methods for the analysis of diatom assemblage data”. In: *The Diatoms: Applications for the Environmental and Earth Sciences*. Ed. by J. P. Smol and E. F. Stoermer. 2nd. Cambridge: Cambridge University Press, pp. 23–54.
- Blomberg, S. P. and T. Garland (2002). “Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods”. In: *Journal of Evolutionary Biology* 15.6, pp. 899–910.
- Buchwalter, D. B., D. J. Cain, C. A. Martin, L. Xie, S. N. Luoma, and T. Garland Jr (2008). “Aquatic insect ecophysiological traits reveal phylogenetically based differences in dissolved cadmium susceptibility”. In: *Proceedings of the National Academy of Sciences* 105.24, pp. 8321–8326.
- Carew, M. E., A. D. Miller, and A. A. Hoffmann (2011). “Phylogenetic signals and ecotoxicological responses: potential implications for aquatic biomonitoring”. In: *Ecotoxicology* 20.3, pp. 595–606.
- Carter, J. L., V. H. Resh, D. M. Rosenberg, and T. B. Reynoldson (2006). “Biomonitoring in North American rivers: a comparison of methods used for benthic macroinvertebrates in Canada and the United States”. In: *Biological Monitoring of Rivers: Applications and Perspectives*. 1st ed. Chichester, England: Ziglio G., Siligardi M., Flaim G., Eds., pp. 203–228.
- Chapman, D. V. (1996). *Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring*. E & Fn Spon London.
- Charles, D. F. (1985). “Relationships between Surface Sediment Diatom Assemblages and Lakewater Characteristics in Adirondack Lakes”. In: *Ecology* 66.3, pp. 994–1011.
- Chessman, B., I. Growns, J. Currey, and N. Plunkett-Cole (1999). “Predicting diatom communities at the genus level for the rapid biological assessment of rivers”. In: *Freshwater Biology* 41.2, pp. 317–331.

- Comte, L., J. Murienne, and G. Grenouillet (2014). “Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes”. In: *Nature Communications* 5.5023.
- Coste, M. and L. Ector (2000). “Diatomées invasives exotiques ou rares en France: principales observation effectuées au cours des dernières décennies”. In: *Systematic and Geography of Plants* 70.2, pp. 373–400.
- DeNicola, D. M. (2000). “A review of diatoms found in highly acidic environments”. In: *Hydrobiologia* 433, pp. 111–122.
- Duong, T. T., A. Feurtet-Mazel, M. Coste, D. K. Dang, and A. Boudou (2007). “Dynamics of diatom colonization process in some rivers influenced by urban pollution (Hanoi, Vietnam)”. In: *Ecological Indicators* 7.4, pp. 839–851.
- Ector, L. and F. Rimet (2005). “Using bioindicators to assess rivers in Europe: An overview”. In: *Modelling community structure in aquatic ecosystems*. Lek S., Scardi M., Verdonshot P.F.M., Descy J.-P., Park Y.S. (eds), pp. 7–19.
- Edgar, R. C. (2004). “MUSCLE: multiple sequence alignment with high accuracy and high throughput”. In: *Nucleic Acids Research* 32.5, pp. 1792–1797.
- Fourtanier, E. and J. P. Kociolek (2011). *Catalogue of Diatom Names, On-Line Version*. California Academy of Sciences. URL: <http://research.calacademy.org/research/diatoms/names/index.asp>.
- Freckleton, R. P., P. H. Harvey, and M. Pagel (2002). “Phylogenetic analysis and comparative data: a test and review of evidence”. In: *The American Naturalist* 160.6, pp. 712–726.
- Gallacher, D. (2002). “The application of rapid bioassessment techniques based on benthic macroinvertebrates in East Asian rivers (a review).” In: *Proceedings of the International Association of Theoretical and Applied Limnology*. 27th Congress in Dublin 1998. Vol. 27. Williams W. D., Ed., pp. 3503–3509.
- Gittleman, J. L. and H.-K. Luh (1992). “On Comparing Comparative Methods”. In: *Annual Review of Ecology and Systematics* 23, pp. 383–404.
- Gouy, M., S. Guindon, and O. Gascuel (2010). “SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building”. In: *Molecular Biology and Evolution* 27.2, pp. 221–224.
- Growns, I. (1999). “Is genus or species identification of periphytic diatoms required to determine the impacts of river regulation?” In: *Journal of Applied Phycology* 11.3, pp. 273–283.
- Guénard, G., P. Legendre, and P. Peres-Neto (2013). “Phylogenetic eigenvector maps: a framework to model and predict species traits”. In: *Methods in Ecology and Evolution* 4.12, pp. 1120–1131.
- Guénard, G., P. C. v. d. Ohe, D. de Zwart, P. Legendre, and S. Lek (2011). “Using phylogenetic information to predict species tolerances to toxic chemicals”. In: *Ecological Applications* 21.8, pp. 3178–3190.
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel (2010). “New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0”. In: *Systematic Biology* 59.3, pp. 307–321.
- Guiry, M. D. (2012). “How many species of algae are there?” In: *Journal of Phycology* 48.5, pp. 1057–1063.

- Hill, B. H., R. J. Stevenson, Y. Pan, A. T. Herlihy, P. R. Kaufmann, and C. B. Johnson (2001). "Comparison of correlations between environmental characteristics and stream diatom assemblages characterized at genus and species levels". In: *Journal of the North American Benthological Society* 20.2, pp. 299–310.
- Ibáñez, C., N. Caiola, P. Sharpe, and R. Trobajo (2010). "Ecological indicators to assess the health of river ecosystems". In: *Handbook of Ecological Indicators for Assessment of Ecosystem Health*. Ed. by S. E. Jørgensen, L. Xu, and R. Costanza. 2nd. Boca Raton, Florida: CRC Press, pp. 447–464.
- Jombart, T., F. Balloux, and S. Dray (2010). "adephylo: new tools for investigating the phylogenetic signal in biological traits". In: *Bioinformatics* 26.15, pp. 1907–1909.
- Jombart, T., S. Pavoine, S. Devillard, and D. Pontier (2010). "Putting phylogeny into the analysis of biological traits: a methodological approach". In: *Journal of Theoretical Biology* 264.3, pp. 693–701.
- Jones, F. C. (2008). "Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates". In: *Environmental Reviews* 16, pp. 45–69.
- Kelly, M. G. (2003). "Short term dynamics of diatoms in an upland stream and implications for monitoring eutrophication". In: *Environmental Pollution* 125.2, pp. 117–122.
- Kelly, M. G., C. J. Penny, and B. A. Whitton (1995). "Comparative performance of benthic diatom indices used to assess river water quality". In: *Hydrobiologia* 302.3, pp. 179–188.
- Kelly, M. G. and B. A. Whitton (1995). "The trophic diatom index: a new index for monitoring eutrophication in rivers". In: *Journal of Applied Phycology* 7.4, pp. 433–444.
- Kobayasi, H. and S. Mayama (1989). "Evaluation of river water quality by diatoms". In: *The Korean Journal of Phycology* 4.2, pp. 121–133.
- Kocielek, J. P. (2005). "Taxonomy and ecology: further considerations". In: *Proceedings of the California Academy of Sciences* 56.1, pp. 99–106.
- Larras, F., F. Keck, B. Montuelle, F. Rimet, and A. Bouchez (2014). "Linking Diatom Sensitivity to Herbicides to Phylogeny: A Step Forward for Biomonitoring?" In: *Environmental Science & Technology* 48.3, pp. 1921–1930.
- Lavoie, I., P. J. Dillon, and S. Campeau (2009). "The effect of excluding diatom taxa and reducing taxonomic resolution on multivariate analyses and stream bioassessment". In: *Ecological Indicators* 9.2, pp. 213–225.
- Lavoie, I., J. Lento, and A. Morin (2010). "Inadequacy of size distributions of stream benthic diatoms for environmental monitoring". In: *Journal of the North American Benthological Society* 29.2, pp. 586–601.
- Leland, H. V. and S. D. Porter (2000). "Distribution of benthic algae in the upper Illinois River basin in relation to geology and land use". In: *Freshwater Biology* 44.2, pp. 279–301.
- Lenoir, A. and M. Coste (1996). "Development of a practical diatom index of overall water quality applicable to the French National Water Board Network". In: *Use of Algae for Monitoring Rivers II*. International symposium, Volksbildungsheim Grillhof Vill, AUT, 17-19 September 1995. Universität Innsbruck: Whitton, B.A., Rott, E., Eds., pp. 29–43.

- Losos, J. B. (2008). “Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species”. In: *Ecology Letters* 11.10, pp. 995–1003.
- Lowe, R. L. and Y. Pan (1996). “Benthic algal communities as biological monitors”. In: *Algal Ecology: Freshwater Benthic Ecosystems*. San Diego, California, USA: Stevenson R. J., Bothwell M. L., Lowe R. L., Eds., pp. 705–739.
- Mann, D. G. and P. Vanormelingen (2013). “An inordinate fondness? The number, distributions, and origins of diatom species”. In: *Journal of Eukaryotic Microbiology* 60.4, pp. 414–420.
- Mouquet, N., V. Devictor, C. N. Meynard, F. Munoz, L. F. Bersier, J. Chave, P. Couteron, A. Dalecky, C. Fontaine, and D. Gravel (2012). “Ecophylogenetics: advances and perspectives”. In: *Biological Reviews*.
- Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schifffers, and W. Thuiller (2012). “How to measure and test phylogenetic signal”. In: *Methods in Ecology and Evolution* 3.4, pp. 743–756.
- Nakov, T., E. C. Theriot, and A. J. Alverson (2014). “Using phylogeny to model cell size evolution in marine and freshwater diatoms”. In: *Limnology and Oceanography* 59.1, pp. 79–86.
- Nylander, J. (2004). *MrAIC.pl*. Version 1.4.5. Uppsala University.
- Oksanen, J., E. Läärä, P. Huttunen, and J. Meriläinen (1988). “Estimation of pH optima and tolerances of diatoms in lake sediments by the methods of weighted averaging, least squares and maximum likelihood, and their use for the prediction of lake acidity”. In: *Journal of Paleolimnology* 1.1, pp. 39–49.
- Pagel, M. (1999). “Inferring the historical patterns of biological evolution”. In: *Nature* 401.6756, pp. 877–884.
- Paradis, E., J. Claude, and K. Strimmer (2004). “APE: analyses of phylogenetics and evolution in R language”. In: *Bioinformatics* 20.2, pp. 289–290.
- Passy, S. I. (2007). “Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters”. In: *Aquatic Botany* 86.2, pp. 171–178.
- Patrick, R. (1977). “Ecology of Freshwater Diatoms - Diatom Communities”. In: *The Biology of Diatoms*. Botanical Monographs 13. Berkeley and Los Angeles, California: Werner D., Ed., pp. 284–332.
- Patrick, R. and D. M. Palavage (1994). “The value of species as indicators of water quality”. In: *Proceedings of the Academy of Natural Sciences of Philadelphia* 145, pp. 55–92.
- Ponader, K. C. and M. G. Potapova (2007). “Diatoms from the genus *Achnanthis* in flowing waters of the Appalachian Mountains (North America): Ecology, distribution and taxonomic notes”. In: *Limnologica* 37.3, pp. 227–241.
- Raunio, J. and J. Soininen (2007). “A practical and sensitive approach to large river periphyton monitoring: comparative performance of methods and taxonomic levels.” In: *Boreal environment research* 12.1.
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Renberg, I. and T. Hellberg (1982). “The pH history of lakes in southwestern Sweden, as calculated from the subfossil diatom flora of the sediments.” In: *AMBIO: A Journal of the Human Environment* 11.1, pp. 30–33.

- Resh, V. H. (2007). “Multinational, freshwater biomonitoring programs in the developing world: Lessons learned from African and Southeast Asian river surveys”. In: *Environmental Management* 39.5, pp. 737–748.
- Revell, L. J. (2012). “phytools: an R package for phylogenetic comparative biology (and other things)”. In: *Methods in Ecology and Evolution* 3.2, pp. 217–223.
- Revell, L. J., L. J. Harmon, and D. C. Collar (2008). “Phylogenetic signal, evolutionary process, and rate”. In: *Systematic Biology* 57.4, pp. 591–601.
- Rimet, F. and A. Bouchez (2012a). “Life-forms, cell-sizes and ecological guilds of diatoms in European rivers”. In: *Knowledge and Management of Aquatic Ecosystems* 406.1, pp. 1–14.
- Rimet, F. (2009). “Benthic diatom assemblages and their correspondence with ecoregional classifications: case study of rivers in north-eastern France”. In: *Hydrobiologia* 636.1, pp. 137–151.
- (2012). “Recent views on river pollution and diatoms”. In: *Hydrobiologia* 683.1, pp. 1–24.
- Rimet, F. and A. Bouchez (2012b). “Biomonitoring river diatoms: Implications of taxonomic resolution”. In: *Ecological Indicators* 15.1, pp. 92–99.
- Rimet, F., J.-C. Druart, and O. Anneville (2009). “Exploring the dynamics of plankton diatom communities in Lake Geneva using emergent self-organizing maps (1974–2007)”. In: *Ecological Informatics* 4.2, pp. 99–110.
- Rimet, F., L. Kermarrec, A. Bouchez, L. Hoffmann, L. Ector, and L. K. Medlin (2011). “Molecular phylogeny of the family Bacillariaceae based on 18S rDNA sequences: focus on freshwater Nitzschia of the section Lanceolatae”. In: *Diatom Research* 26.3, pp. 273–291.
- Round, F. E. (1991). “Diatoms in river water-monitoring studies”. In: *Journal of Applied Phycology* 3.2, pp. 129–145.
- Round, F. E., R. M. Crawford, and D. G. Mann (1990). *The diatoms: biology and morphology of the genera*. Cambridge, UK. 760 pp.
- Rovira, L. (2013). “The ecology and taxonomy of estuarine benthic diatoms and their use as bioindicators in a highly stratified estuary (Ebro Estuary, NE Iberian Peninsula): a multidisciplinary approach”. PhD thesis. IRTA San Carlos de la Ripita (Spain): University of Barcelona. 1-295.
- Rumeau, A. and M. Coste (1988). “Initiation à la systématique des diatomées d’eau douce”. In: *Bulletin Français de la Pêche et de la Pisciculture* 309, pp. 1–69.
- Salden, N. (1978). *Beiträge zur Ökologie der Diatomeen (Bacillariophyceae) des Süßwassers*. Naturhistorischer Verein der Rheinlande und Westfalens.
- Sanderson, M. J. (2002). “Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach”. In: *Molecular Biology and Evolution* 19.1, pp. 101–109.
- Schoeman, F. R. (1973). *A systematical and ecological study of the diatom flora of Lesotho with special reference to the water quality*. Pretoria, South Africa: National Institute for Water Research, p. 355.
- Shokralla, S., J. L. Spall, J. F. Gibson, and M. Hajibabaei (2012). “Next-generation sequencing technologies for environmental DNA research”. In: *Molecular Ecology* 21.8, pp. 1794–1805.
- Sládeček, V. (1986). “Diatoms as Indicators of Organic Pollution”. In: *Acta Hydrochimica et Hydrobiologica* 14.5, pp. 555–566.

- Stamatakis, A. (2006). “RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models”. In: *Bioinformatics* 22.21, pp. 2688–2690.
- Stevenson, R. and J. Smol (2003). “Use of algae in environmental assessments”. In: *Freshwater Algae in North America: Ecology and Classification*. San Diego, California, USA: Wehr J. D., Sheath R. G., Eds., pp. 775–804.
- Ter Braak, C. J. F. and C. W. N. Looman (1986). “Weighted averaging, logistic regression and the Gaussian response model”. In: *Vegetatio* 65.1, pp. 3–11.
- Theriot, E. C., E. Ruck, M. Ashworth, T. Nakov, and R. K. Jansen (2011). “Status of the pursuit of the diatom phylogeny: Are traditional views and new molecular paradigms really that different?” In: *The Diatom World*. Ed. by J. Seckbach and J. Kocielek. New York, USA: Springer, pp. 119–142.
- Torrise, M. and A. Dell’Uomo (2006). “Biological Monitoring of Some Apennine Rivers (central Italy) Using the Diatom-Based Eutrophication / Pollution Index (epi-D) Compared to Other European Diatom Indices”. In: *Diatom Research* 21.1, pp. 159–174.
- Van Dam, H., A. Mertens, and J. Sinkeldam (1994). “A coded checklist and ecological indicator values of freshwater diatoms from the Netherlands”. In: *Netherlands Journal of Aquatic Ecology* 28.1, pp. 117–133.
- Vanelslander, B., V. Créach, P. Vanormelingen, A. Ernst, V. A. Chepurnov, E. Sahan, G. Muyzer, L. J. Stal, W. Vyverman, and K. Sabbe (2009). “Ecological Differentiation Between Sympatric Pseudocryptic Species in the Estuarine Benthic Diatom *Navicula Phyllepta* (bacillariophyceae)”. In: *Journal of Phycology* 45.6, pp. 1278–1289.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue (2002). “Phylogenies and community ecology”. In: *Annual Review of Ecology and Systematics* 33, pp. 475–505.
- Weckström, K. and S. Juggins (2006). “Coastal Diatom–Environment Relationships from the Gulf of Finland, Baltic Sea”. In: *Journal of Phycology* 42.1, pp. 21–35.
- Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, T. Jonathan Davies, J. A. Grytnes, and S. P. Harrison (2010). “Niche conservatism as an emerging principle in ecology and conservation biology”. In: *Ecology Letters* 13.10, pp. 1310–1324.
- Wu, J.-T. (1999). “A generic index of diatom assemblages as bioindicator of pollution in the Keelung River of Taiwan”. In: *Hydrobiologia* 397, pp. 79–87.
- Wunsam, S., A. Cattaneo, and N. Bourassa (2002). “Comparing diatom species, genera and size in biomonitoring: a case study from streams in the Laurentians (Quebec, Canada)”. In: *Freshwater Biology* 47.2, pp. 325–340.
- Wunsam, S., R. Schmidt, and R. Klee (1995). “Cyclotella-taxa (Bacillariophyceae) in lakes of the Alpine region and their relationship to environmental variables”. In: *Aquatic Sciences* 57.4, pp. 360–386.
- Zampella, R. A., K. J. Laidig, and R. L. Lowe (2007). “Distribution of diatoms in relation to land use and pH in blackwater coastal plain streams”. In: *Environmental Management* 39.3, pp. 369–384.