Diversity and Distributions, (Diversity Distrib.) (2017) 1–13

BIODIVERSITY RESEARCH

Environmental DNA genetic monitoring of the nuisance freshwater diatom, *Didymosphenia geminata*, in eastern North American streams

Stephen R. Keller[1]*, Robert H. Hilderbrand[1], Matthew K. Shank[2] and Marina Potapova[3]

[1]*Department of Plant Biology, University of Vermont, 63 Carrigan Dr., Burlington, VT 05405, USA,* [2]*Susquehanna River Basin Commission, 4423 North Front Street, Harrisburg, PA 17110, USA,* [3]*Academy of Natural Sciences, Drexel University, 1900 Benjamin Franklin Parkway, Philadelphia, PA 19103, USA*

*Correspondence (current address): Stephen R. Keller, Department of Plant Biology, University of Vermont, 63 Carrigan Dr., Burlington, VT 05405, USA.
E-mail: srkeller@uvm.edu

ABSTRACT

Aim Establishing the distribution and diversity of populations in the early stages of invasion when populations are at low abundance is a core challenge for conservation biologists. Recently, genetic monitoring for environmental DNA (eDNA) has become an effective approach for the early detection of invaders, especially for microscopic organisms where visual detection is challenging. *Didymosphenia geminata* is a globally distributed freshwater diatom that shows a recent emergence of nuisance blooms, but whose native versus exotic status in different areas has been debated. We address the hypothesis that the distribution and genetic diversity of *D. geminata* in eastern North America is related to the recent introduction of non-native lineages, and contrast that with the alternative hypothesis that *D. geminata* is cryptically native to the region (i.e. at low abundance) and only forms nuisance blooms when triggered by a change in environment.

Location The Mid-Atlantic region of North America.

Methods We analysed 118 stream samples for *D. geminata* eDNA, validated our results for a subset of sites using direct visual enumeration by microscopy and used molecular cloning to sequence *D. geminata* from two sites where eDNA was detected.

Results (1) *D. geminata* eDNA was detected at seven spatially unique sites, six of which were previously documented to contain recent *D. geminata* blooms. (2) Sites where *D. geminata* eDNA was detected exhibited no difference in environmental conditions compared to sites with no-detected *D. geminata* eDNA. (3) Sequencing of *D. geminata* eDNA showed that blooms were composed of multiple genetic lineages, closely related to those sampled elsewhere across the globe.

Main conclusions We interpret these results as most consistent with the hypothesis that *D. geminata* is an exotic invader in the Mid-Atlantic region, still in its early stages of invasion; thus, genetic monitoring and management efforts may still be effective at controlling its spread.

Keywords

diatom, Didymo, environmental DNA, genetic diversity, harmful algal blooms, invasive species.

INTRODUCTION

The study of biological invasions is replete with examples of long-distance dispersal of a non-native species to a new biogeographic region, where it increases in abundance and expands its range (Wilson *et al.*, 2009; Blackburn *et al.*, 2011). While less common, it is also recognized that invasiveness may emerge in native species due to a change in

DOI: 10.1111/ddi.12536
http://wileyonlinelibrary.com/journal/ddi

ecological conditions or interspecific interactions in its historic range, often the result of human-mediated environmental change (Valéry *et al.*, 2008; Buczkowski, 2010; Carey *et al.*, 2012). Distinguishing between these scenarios is key to understanding the biogeography of invasiveness and its ecological underpinnings, and may have direct impact on how invasions are managed.

The introduction of a non-native exotic versus the ecological release of a native endemic is likely to produce differences in the distribution, abundance and genetic diversity of newly emerging invasions. If invasiveness is due to the recent introduction of a non-native, we would expect established populations near the introduction site(s) as well as other suitable sites that have not yet been colonized, and genetic diversity at invaded sites that is shared with distant geographic regions. On the other hand, if invasiveness is due to a subset of native populations increasing in density or aggressiveness in response to a change in environment, we would expect a wider distribution of occupied sites where environments are suitable, with higher abundance 'invasive' populations at sites where environmental conditions promote the transition to invasiveness. Depending on the phylogeographic history of the species, a native species transitioning to higher abundance is also more likely to contain geographically restricted or endemic genetic diversity. Both scenarios underscore the need for distributional data coupled with genetic monitoring to identify the likely origins of invasiveness.

For newly emerging invasions, the ability to rapidly and accurately monitor the presence of the invader is essential. However, native populations at low abundance or introduced populations at the earliest stages of invasion can be difficult to detect with traditional monitoring such as visual assessments. Recently, the spread of invasive species has been facilitated by genetic monitoring for environmental DNA (eDNA; e.g. DNA sloughed off by organisms into their environment) (Lodge *et al.*, 2012). Monitoring of eDNA has become an effective approach for the early detection of invaders such as the Asian carp in aquatic ecosystems (Jerde *et al.*, 2011), as it can detect the presence of the invader even when the population is small and visual detection may be difficult (Darling & Mahon, 2011).

Didymosphenia geminata (Lyngbye) M. Schmidt 1899 is a globally distributed freshwater diatom that is an emerging invasive species in rivers and streams. Its invasiveness stems from the ability to form nuisance blooms in cool, nutrient-poor waters that are prime habitat for trout and other game and non-game fish (Spaulding & Elwell, 2007; Sundareshwar *et al.*, 2011). Impacts to base trophic levels (Gillis & Lavoie, 2014) and primary consumers (Rost & Fritsen, 2014) have also been documented, which together represent a conservation threat to the diversity, functioning, and recreational use of aquatic ecosystems.

The native range of *D. geminata* is considered to be Asia, northern Europe and North America (Blanco & Ector, 2009), but its status as an introduced invader in different geographic regions is debated, with direct implications for how *D. geminata* is managed (Taylor & Bothwell, 2014, 2015; Bergey & Spaulding, 2015). In New Zealand, recent blooms are clearly the result of human introductions, possibly dispersed by anglers (Kilroy *et al.*, 2009). In North America, species distribution models based on bioclimatic data support the Pacific Northwest and large regions of the Rocky Mountains as suitable habitat for *D. geminata*, but models predict a narrower suitable region in eastern North America (Kumar *et al.*, 2009), calling into question whether *D. geminata* is historically native to this region or more recently introduced and spreading. Historical surveys and palaeolimnological evidence from sediment cores support the presence of *D. geminata* dating back to the late 19th and early 20th centuries in the Pacific Northwest and Rocky Mountains (Kumar *et al.*, 2009; Taylor & Bothwell, 2014). Palaeolimnological records from Quebec show absence of *D. geminata* prior to ~1970, followed by rapid increase in abundance to present time (Lavery *et al.*, 2014). In the Mid-Atlantic region, historical presence of *D. geminata* was noted from sediment deposits of the Delaware River in the vicinity of Philadelphia, Pennsylvania (Boyer, 1916, 1927). Modern blooms of *D. geminata* were first observed in 2006 in Virginia, and blooms have since been reported in Maryland (Gunpowder Falls, Lower Savage River, North Branch Potomac River, Big Hunting Creek), Pennsylvania (Delaware and Youghiogheny rivers, Pine Creek), Virginia (Smith, Jackson and Pound rivers) and West Virginia (Elk River, Glady Fork, Gandy and Seneca creeks).

In this study, we establish a genetic monitoring programme for *D. geminata* to determine the distribution and genetic diversity of this bloom-forming diatom in eastern North America, where its native status is unclear. Using a combination of eDNA monitoring and visual enumeration of *D. geminata* from benthic samples, we address the hypothesis that *D. geminata* invasion is caused by introduction of non-native populations that form nuisance blooms where they have established. This would indicate that the current distribution of *D. geminata* is not in equilibrium with the availability of suitable habitat, and thus has opportunity to continue spreading. We contrast this with the alternative hypothesis that *D. geminata* is native to the region and is widespread but at low abundance (i.e. below visual detection), and only forms nuisance blooms under favourable environmental conditions. In addition, we establish the genetic diversity of *D. geminata* in the region to assess whether blooms (1) contain diversity endemic to the region, (2) reflect the occurrence and spread of a clonal invasive genotype or (3) contain a diverse assemblage of genotypes that is shared with other *D. geminata* blooms globally.

METHODS

Field sampling

IWe collected stream water samples for eDNA analysis from mid-Atlantic streams in Maryland and Pennsylvania. Field

sampling of 76 Maryland stream sites occurred during March, April and May of 2014. We sampled first- through third-order streams across Maryland, spanning diverse physiographic regions (Appalachian Plateau, Ridge and Valley, Eastern Piedmont and Atlantic Coastal Plain geographic provinces). Among these were three-third-order streams known previously from visual surveys to currently or in the recent past contain *D. geminata*: Gunpowder Falls, Big Hunting Creek in Catoctin Mountain Park and the lower Savage River below Savage River Reservoir (R. Klauda, K. Hana, & T. Gardeur, unpublished data). Additional field sampling of 26 Pennsylvania stream sites (third through seventh order) occurred in 2014 (16 samples) and 2015 (26 samples). Sixteen of the 42 total samples collected in Pennsylvania were collected in April 2014 from sites in the Pine Creek watershed. Those same 16 sites were sampled in the Pine Creek watershed again in April 2015, as were eight additional sites spread throughout the Pennsylvania portion of the Susquehanna River drainage. We also sampled two sites in the Ohio River drainage in south-west Pennsylvania from the Youghiogheny River in July 2015.

Field sampling for eDNA consisted of deploying a plankton drift net with 35-μm mesh, fitted with a 250-μm-mesh pre-filter, in the centre of the stream channel for sufficient duration to filter a target volume of > 10,000 L (Fig. 1) (Cary *et al.*, 2006). Duration time was calculated based on flow velocity measurements taken at the depth of the sampling net. The resulting sample was preserved in 500 mL of 70% ethanol and brought back to the laboratory within 48 h where it was stored at −20 °C until extraction. The entire field sampling apparatus was disinfected between sites with 5% bleach following established standards (Cary *et al.*, 2006), and waders were soaked in a 2% Virkon solution for at least 1 min.

In addition to plankton drift net samples for eDNA, we collected water chemistry data that included total phosphorus, soluble reactive phosphorus (SRP), pH, nitrate, sulphate, chloride and specific conductance, which are thought to be predictive of *D. geminata* distribution or abundance (Kilroy & Bothwell, 2011; Bothwell *et al.*, 2014; Bray *et al.*, 2016). At Maryland sites, a 1 L grab sample of stream water was collected for laboratory chemical analyses at the Appalachian Laboratory (University of Maryland Center for Environmental Science), using standard protocols for each analyte (American Public Health Association (APHA), 2005). At most Pennsylvania sites, depth-integrated water samples were collected across a transect perpendicular to flow. Samples were composited into a churn-splitter, bottled and sent to ALS Environmental Laboratory (Middletown, Pennsylvania) for chemical analysis within 24 h. Specific conductance was measured using a YSI 6600 data sonde. For compatibility with the Maryland water chemistry data, we restricted analysis of Pennsylvania water chemistry to the sampling that occurred closest in time to the *D. geminata* eDNA sample for a given site.

Figure 1 Images of field sampling for *Didymosphenia geminata*. Upper: plankton drift net deployed to sample 10,000 L of water for eDNA (photo credit: S. Keller); lower: substrate colonized by *D. geminata* in stream (photo credit: M. Shank).

eDNA quantitative PCR assay of *D. geminata* presence/absence

All eDNA samples were processed at the Appalachian Laboratory using a strict QC pipeline and implementing a *D. geminata*-specific eDNA protocol based on quantitative real-time PCR (hereafter, qPCR) (Cary *et al.*, 2014). The *D. geminata* qPCR protocol is highly sensitive and capable of detecting *D. geminata* abundance as low as 1 cell per mL of plankton tow net sample (Cary *et al.*, 2006). We implemented stringent procedures designed to minimize the risk of both false positives (*D. geminata* being detected when it was in fact absent) and false negatives (failure to detect *D. geminata* when it is in fact present) in eDNA testing. Samples were assigned internal tracking codes for blind processing without regard to sample origin, and each round of extraction included a 'reagent blank' negative control. We also controlled for variation in the efficiency of DNA extraction from environmental samples and the presence of PCR inhibitors (Uyua *et al.*, 2014; Jones *et al.*, 2015) by including the addition of exogenous control DNA, which allowed us to

assess the efficiency of the extraction procedure (Cary *et al.*, 2014). After testing several protocols, we used the MoBio PowerSoil DNA isolation kit for all samples, which produced 100% amplification efficiency of exogenous control DNA in all reactions (see section "Results").

During extraction, eDNA samples were homogenized by shaking, and a 1 mL aliquot was transferred to a 1.5-mL microcentrifuge tube. Tubes were spun at 8000 *g* for 4 min to pellet cells, and residual ethanol was decanted. Pellets were washed with 1 mL of sterile water, re-centrifuged and decanted as above and resuspended in 100 μL of sterile water. The sample pellet was mixed thoroughly and then transferred to a MoBio bead tube for disruption by vortexing for 20 min at maximum speed. eDNA was then extracted following the manufacturer's instructions, with two additional modifications. First, we spiked each sample with control plasmid DNA mixed in with the MoBio C1 buffer solution. Following Cary *et al.* (2014), this control DNA consisted of 300 ng per sample of pGEM-3Z plasmid vector (Promega, Madison, WI, USA). Second, to obtain optimal cell lysis following sample disruption, we added 20 μL of proteinase K (20 mg mL^{-1}) and incubated with shaking (170 rpm) at 55 °C for 60 min. Extracted samples were quantified for total eDNA concentration using the Qubit BR assay (Invitrogen, Carlsbad, CA, USA), and stored at −80 °C until further analysis.

Extracted eDNA was used as template for qPCR using *D. geminata*-specific primers and probe based on the nuclear 18S ribosomal RNA gene (rDNA) (Cary *et al.*, 2014). All qPCR assays were performed using two technical replicates for each target (*D. geminata* and pGEM; a total of four qPCRs per sample), which consisted of duplicate reactions originating from the same eDNA extraction. Reactions were performed in 20 μL volumes consisting of 10 μL 2X TaqMan Fast Universal PCR Master Mix, 1.8 μL of 10 μM forward primer D602F (5′-GTTGGATTTGTGATGGAATTTGAA-3′), 1.8 μL of 10 μM reverse primer D753R (5′-AATACATTC ATCGACGTAAGTC-3), 0.5 μL of 10 μM custom TaqMan *D. geminata* probe D641FAM (5′-FAM-CACCCACGGATG ACAGTTTCTGA-MGB-3′), 0.4 μL ROX dye and 2 μL of purified DNA template. Equivalent reactions were set up for the pGEM exogenous DNA controls, substituting the following primers: forward primer M13 (5′-CCCAGTCACGAC GTTGTAAAACG-3′), reverse primer pGEMR (5′-TGTGTG GAATTGTGAGCGGA-3′) and a custom TaqMan pGEM probe (5′-6FAM-CACTATAGAATACTCAAGCTTGCATGCC TGCA-MGBNFQ-3′). Primer and probe sequences were based on Cary *et al.* (2007).

Each qPCR was run on a StepOnePlus qPCR thermal cycler (Life Technologies, Carlsbad, CA, USA) using the following settings: 50 °C for 2 min, 95 °C for 20 s, followed by 40 cycles of 95 °C for 3 s and 60 °C for 30 s. Negative controls (sterile water) were included for each run, as well as a positive control stream sample (SAVA-REF1-14) that showed amplification of *D. geminata* eDNA during initial testing. A site was considered as *D. geminata*-positive based on eDNA detection when the relative fluorescence (R_N) of the reporter exceeded the baseline threshold ($R_N > 0.1$ on a log scale; Fig. 2).

Benthic diatom sampling and direct cell counts

To determine the correspondence between *D. geminata* detection based on eDNA versus direct cell counts under microscopy, we collected diatom samples using a modified Rapid Bioassessment Protocol (Barbour & Gerritsen, 1999). At each site, 11 pieces of natural substrate were removed from the stream across a transect perpendicular to flow. Benthic algae within a 12-cm^2 section from each substrate piece (132 cm^2 in total) was disturbed using a soft brush, rinsed into a composite sample bottle, and sample volume was recorded. A 50 mL subsample was then extracted and preserved with formaldehyde. All diatom sample processing procedures followed The Academy of Natural Sciences protocol (Charles *et al.*, 2002), and presence/absence of *D. geminata* was determined for each sample using microscopic examination. Benthic sampling overlapped eDNA sampling at 19 sites, consisting of 23 samples collected in the Susquehanna drainage in Pennsylvania in 2014 and 2015. While eDNA and benthic samples were spatially co-located at these sites, the sampling time often differed for logistical reasons; time between eDNA and benthic sampling ranged from 0 to 187 days, with a mean of 122 days. This effect would be expected to decrease any association between eDNA and microscopic analysis of *D. geminata*, making an observed association conservative.

Molecular cloning and DNA sequencing of positive sites

We cloned and sequenced *D. geminata* eDNA from two Maryland samples (Lower Savage River and Gunpowder Falls) that successfully amplified *D. geminata* during qPCR. The objective here was to confirm the genetic identity of *D. geminata* in eDNA samples that tested positive, and estimate their diversity.

Diversity was assessed by molecular cloning and sequencing *D. geminata* for the 18S and 5.8S subunits of rDNA, including the internal transcribed spacer (ITS1 and ITS2) variable regions. We followed the protocol of Cary *et al.* (2007) and used the primers D602F (5′-GTTGGATTTGT-GATGGAATTTGAA-3′) and ITS4R (5′-TCCTCCGCTTATT-GATATGC-3′) to PCR-amplify from DNA extractions that tested positive during qPCR. PCR products were then cloned using the TOPO-TA cloning kit (Invitrogen). Positive clones were selected, PCR-amplified, cleaned with ExoSAP and sequenced on an ABI 3130xl at the West Virginia University genomics core using the pair of PCR primers as well as two additional internal sequencing primers (D1659F: 5′-GC TGGGGATTGCAGCTA-3′; and D1670R: 5′-CACCAGTAA AGGCATTAGCTG-3′) to generate a maximum overlapping region of approximately 1700 bp.

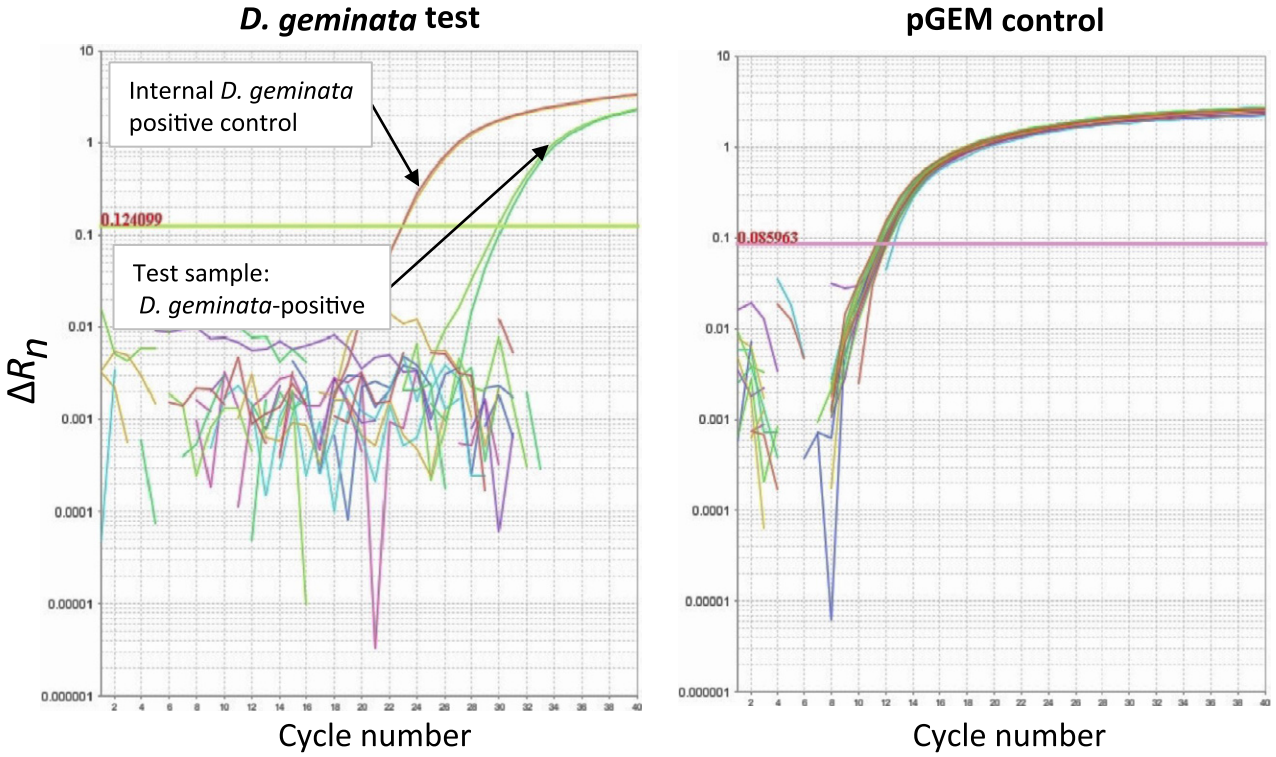

Figure 2 Example qPCR results for *D. geminata* eDNA testing. Plots show change in relative fluorescence (ΔR_N) as a function of cycle number. Note the log scale on the *y*-axis. Results in the left panel show an example amplification of two *D. geminata*-positive samples: an internal *D. geminata*-positive control (SAVA-REF1-14) and a test sample (Pine Blackwell, PA). The right panel shows amplification of the exogenous control DNA (pGEM plasmid) from the same extraction as the test samples. The horizontal line near $\Delta R_N = 0.1$ in both plots shows the fluorescence threshold applied to determine the threshold cycle number (C_T value) when the amplification of the target rises above the baseline. Samples negative for the qPCR assay display baseline fluorescence that never amplifies above this threshold.

Analysis of the sequences proceeded by trimming all reads of low-quality bases. Contigs were assembled among overlapping reads, and multiple sequence alignment was carried out using the Muscle algorithm (Edgar, 2004) implemented in GENEIOUS v.7.1.7 (Kearse *et al.*, 2012). Singletons (polymorphisms present in only a single sequenced clone) and alignment gaps due to indels were treated as missing data. We used maximum likelihood (ML) in the MEGA7 software (Kumar *et al.*, 2016) to construct a phylogeny of our *D. geminata* sequences that overlapped with previously published sequences of *Didymosphenia* and closely related diatom species used as outgroups (Table 1). Of the 1700 bp sequenced, a 951-bp section of the alignment overlapped with the accessions listed in Table 1 and was analysed for phylogenetic relationships. We selected the Tamura 3-parameter + gamma model of molecular evolution based on the minimum Bayesian Information Criterion from model testing. Support for nodal relationships was assessed with 500 bootstrap replicates. We also assessed relationships among sequences using haplotype networks constructed with the statistical parsimony method using the PEGAS package in R (Paradis, 2010). We made one network on the same 951-bp alignment used above for phylogeny building, but restricted our inclusion of outgroups to the closely related *Cymbella proxima* (GenBank accession AM502017). A second network was constructed using 662 bp of the more highly variable 3′ end of the alignment, for which only data from sequenced clones generated during this study were available. Lastly, nucleotide diversity within *D. geminata*-positive sites from the Lower Savage River (SAVA) and Gunpowder Falls (GUN) was computed on the 662-bp alignment in MEGA based on the Tamura 3-parameter + gamma model, with standard errors estimated from 1000 bootstrap replicates. All new sequences have been submitted to GenBank (accession numbers KY421425-KY421452).

Didymosphenia presence in relation to water chemistry

We used the eDNA results to determine whether *D. geminata*-positive sites were characterized by differences in water chemistry. Due to the small number of sites containing *D. geminata*, we implemented a bootstrap resampling approach to test for associations with water chemistry. We ran 10,000 iterations that randomly subsampled the *D. geminata*-negative sites with replacement to extract a sample size equal to that of *Didymosphenia*-positive sites, and calculated the mean value of each variable per iteration. This generated

Table 1 GenBank accessions of rDNA sequences from *Didymosphenia geminata* and closely related diatom genera used in this study.

Species	Strain	Collection locale	GenBank accession	Study
Didymosphenia geminata	TCC777	Trentino Riviere de Brusago, Italy	KT072999	Keck *et al.* (2015)
Didymosphenia geminata	CH058	Colorado, USA	KJ011636	Nakov *et al.* (2014)
Didymosphenia geminata	n/a	Bull River, New Zealand	JN680079	Cary *et al.* (2014)
Didymosphenia geminata	B40	Lake Baikal, Russia	KJ011637	Nakov *et al.* (2014)
Didymosphenia dentata	B547	Lake Baikal, Russia	KJ011635	Nakov *et al.* (2014)
Cymbella proxima	AT-210Gel13	Germany	AM502017	Bruder and Medlin (2007)
Cymbopleura naviculiformis	AT-177.04	Germany	AM501997	Bruder and Medlin (2007)
Gomphonema micropus	AT-117.09	Germany	AM501964	Bruder and Medlin (2007)

a distribution of 10,000 mean values for the randomly subsampled sites. If the observed mean value for *D. geminata*-positive sites fell within the 95% quantiles of the randomized distribution, we concluded that *D. geminata*-positive sites did not differ in water chemistry from negative sites.

RESULTS

Distribution of *D. geminata* eDNA

We collected 118 samples of stream water from 102 unique sites across our study region (Table S1). This represented > 1.8 million L of filtered stream water (range across sites: 8494–29,620 L; mean per site = 15,681 L) to test for the distribution of *D. geminata* eDNA in Mid-Atlantic streams. From this extensive regional assessment, we identified only seven positive sites (i.e. where *D. geminata* eDNA was detected) (Fig. 3). Of these, six sites were known previously to form visual blooms (Klauda *et al.*, 2013; Shank *et al.*, 2016), while one site tested positive for eDNA with no prior history of observed blooms (Trout Run; see below). An additional site, Big Hunting Creek, MD, was known to form blooms previously (2012), but was negative for *D. geminata* eDNA in 2014; importantly, visual blooms were also not observed at Big Hunting Creek during 2014 (L. Donaldsen, personal communication).

Two of the sites that tested positive for *D. geminata* eDNA were in Maryland, and both were in streams where visual blooms have been observed: the Lower Savage River below the Savage River reservoir, and Gunpowder Falls below Prettyboy Reservoir (Klauda *et al.*, 2013). Within Pennsylvania samples, we detected *D. geminata* eDNA at Pine Creek 13.7 km upstream from the town of Blackwell (2014 and 2015), the West Branch of Pine Creek (2014 and 2015) and two samples from the Youghiogheny River: (1) downstream of the dam in Confluence, PA, and (2) the Ramcat boat launch downstream of the confluence with the Casselman River (both sampled in 2015 only). An additional Pennsylvania site, Trout Run (only sampled in 2015), tested positive for eDNA in one of two sample duplicates. The qPCR assay was repeated for Trout Run a second time and similarly resulted in one of two sample duplicates testing positive (for a total of two out of four reactions amplifying across both rounds of analysis). We interpret this as presence of *D. geminata* in Trout Run at very low abundance, given the stringent QC procedures to ensure lack of false positives, but also emphasize that this result should be interpreted cautiously until further sampling can confirm *D. geminata* presence.

Didymosphenia presence in benthic samples

Of 23 benthic samples from 19 sites collected across the Susquehanna drainage in Pennsylvania, *D. geminata* was detected in four samples from two sites in the Pine Creek watershed, all of which also tested positive for *D. geminata* eDNA (Table S1). In 22 of the 23 samples (96%), benthic and eDNA techniques were in accordance, either both positive or both negative for the presence of *D. geminata*. The exception was Trout Run, where eDNA suggested *D. geminata* presence, but benthic sampling was negative.

Didymosphenia geminata association with water chemistry

Sites positive for *D. geminata* eDNA did not differ from negative sites for any water chemistry variables (Fig. 4). Unlike previous findings that showed an association between *D. geminata* blooms and SRP levels < 1–2 ppb (Bothwell *et al.*, 2014), our analysis showed that total phosphorous and SRP levels of *D. geminata*-positive sites were near the middle of the randomized distributions, and ≫2 ppb. Sulphate was the only variable in *D. geminata*-positive streams that was near the tail of the randomized distribution, but the location of the observed mean in the lower quantile was opposite to the prediction that *D. geminata* benefits from sulphate during blooms (Rost *et al.*, 2011).

Genetic diversity of *D. geminata*-positive sites

We sequenced a fragment of the rDNA locus from two sites positive for *D. geminata* eDNA in MD (Lower Savage River, SAVA, and Gunpowder Falls, GUN). For each site, we used molecular cloning to isolate and sequence samples (e.g. 'clones') of amplified *D. geminata* eDNA, totalling 28 sequenced clones across both sites. We obtained full-length sequences of 1700 bp for six clones from each sampling site

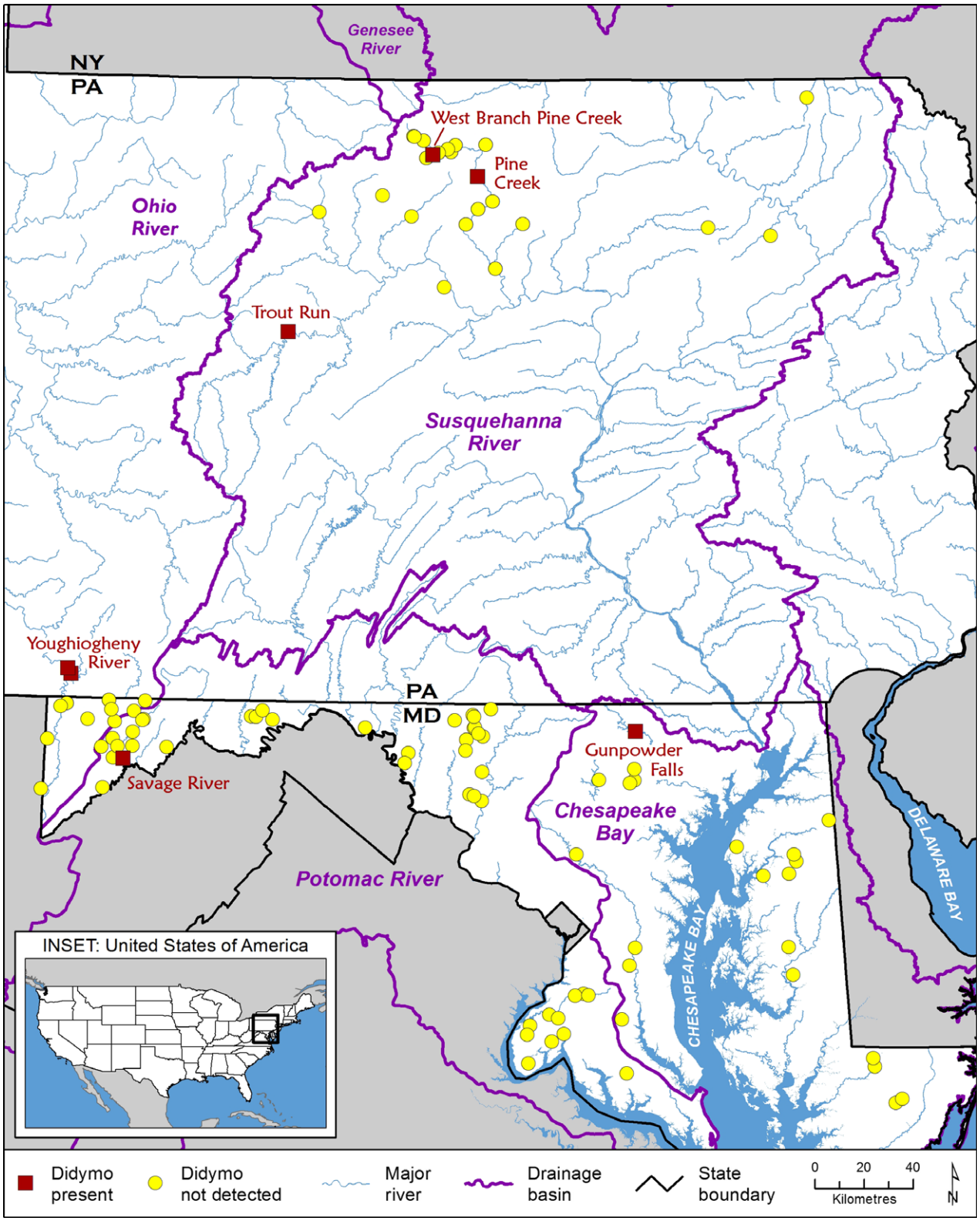

Figure 3 Map of sampling sites with *D. geminata* present (red squares and labelled) and absent (yellow circles). Drainage basins and major rivers within the study region are indicated. Sites labelled as 'Didymo Present' (red squares) indicate where *D. geminata* eDNA was detected. In all such sites, *D. geminata* has also been observed forming visual blooms, except for Trout Run (eDNA detected, but no visual blooms).

(12 in total). We compared these sequences to a *D. geminata* sequence deposited in GenBank (accession number KT072999) for the 951-bp where there was sequence overlap. The average pairwise identity between our sequences and the GenBank accession was 99.99%, whereas the average identity between our sequences versus an outgroup (*Cymbella* proxima; AM502017) was 97.86%, confirming that our eDNA samples are most likely *D. geminata* and not another related diatom.

The six fully sequenced clones overlapped for 951 bp with other published sequences from *Didymosphenia* and closely related diatoms (Table 1). The ML phylogeny for this set of samples returned a monophyletic *Didymosphenia* group with strong bootstrap support (98%; Fig. 5a). Intraspecific diversity within *Didymosphenia* was present, but topological relationships were not strongly supported (bootstrap values 64–65%). Mid-Atlantic samples (GUN and SAVA) clustered with sequences from Colorado, Italy and New Zealand (Fig. 5a). These relationships were also seen in the haplotype network, in which a *D. geminata* accession from Italy shared a haplotype with eight clones, while *D. geminata* accessions from New Zealand and Colorado, USA, shared a haplotype with three clones (Fig. 5b). Haplotypes of *D. geminata* and *D. dentata* from Russia were not shared with any other samples.

For all 28 GUN and SAVA clones, we were able to obtain sequence for a 662-bp subset of the alignment. Genetic diversity for this subset consisted of four single nucleotide polymorphisms (SNPs) present in > 1 sequence, with minor allele frequencies ranging from 0.107 to 0.429. These SNPs defined a network of six haplotypes, which showed that *D. geminata* does not occur as a single, clonal strain in our region, but rather constitutes multiple genetic strains across the two sites (Fig. 6). Sites had similar levels of haplotype richness (five haplotypes per site; four haplotypes shared across both sites; Fig. 6) and nucleotide diversity (GUN: $\pi = 0.002 \pm 0.001$; SAVA: $\pi = 0.001 \pm 0.001$).

DISCUSSION

In eastern North America, recent nuisance blooms have called into question whether *D. geminata* is a human-introduced exotic or a native member of the indigenous diatom community that has recently become problematic due to a change in environment or emergence of a novel genetic strain. We conducted an extensive sampling of

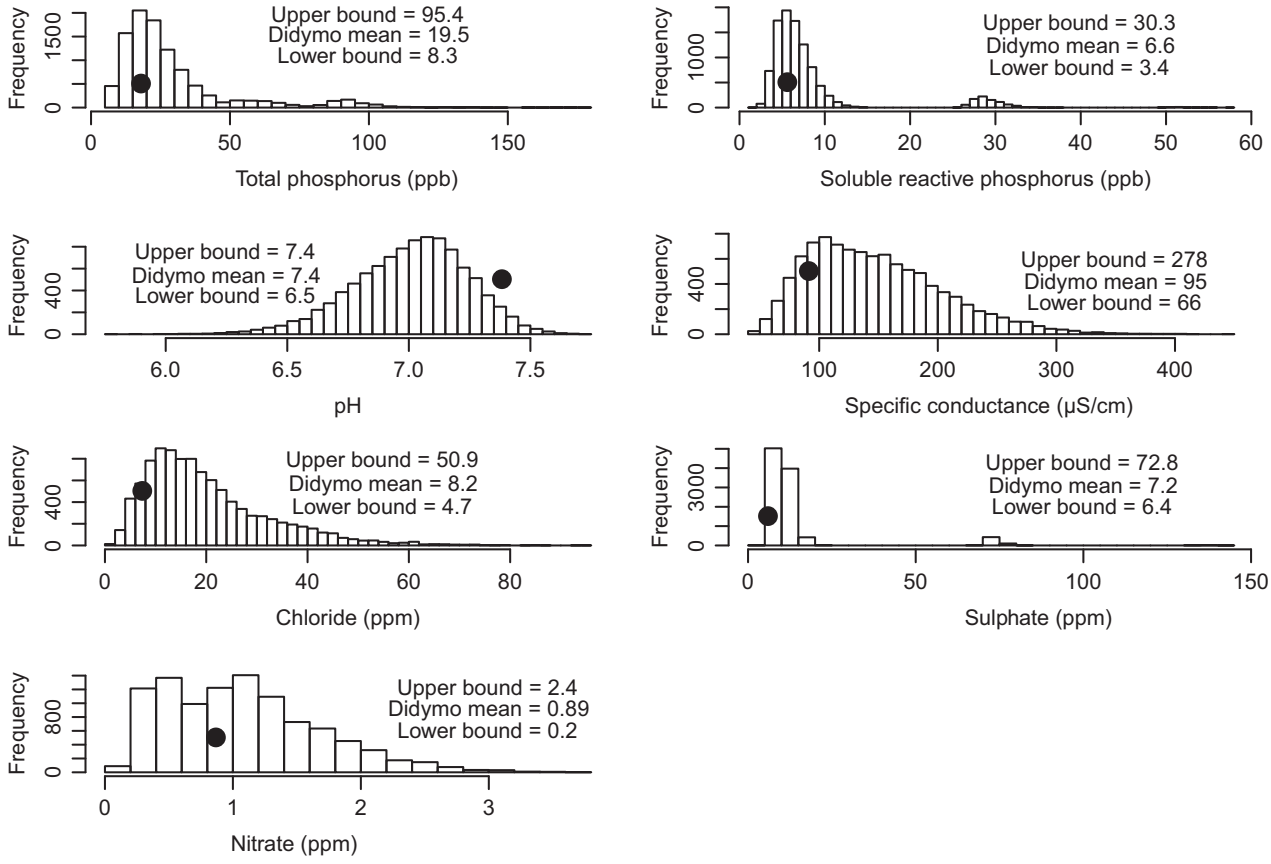

Figure 4 Resampled distributions of environmental variables for all sampled streams compared to the mean value (solid circle) of those found to contain *D. geminata*. The distributions for all streams represent the means of 10,000 random subsamples with size equal to those for *D. geminata*-containing streams and sampled with replacement. We list the lower and upper 95% bounds for the distributions.

streams to determine the distribution, environmental associations and genetic diversity of *D. geminata* in the Mid-Atlantic. Our dual approach of eDNA analysis for rapid determination of *D. geminata* presence, along with more traditional but labour-intensive microscopic analysis of benthic samples, returned highly consistent results. We have shown that *D. geminata* is *not* currently widespread at low abundance, nor does it generally occur in streams that lack evidence of blooms, as would be predicted for an indigenous species that only increases in abundance under favourable conditions. Rather, *D. geminata* is confined to a few sites mostly known previously from visual detection of nuisance blooms, and its absence from sites where conditions appear favourable (cool, nutrient-poor waters) suggests its distribution is not at equilibrium in the region. Furthermore, we demonstrate that *D. geminata* consists of multiple strains that are phylogenetically related to *D. geminata* sampled in Colorado, Italy and New Zealand. We discuss these results in the light of how genetic monitoring and diversity data can help inform questions about the distributional status of newly emerging nuisance species.

Is *D. geminata* an exotic invasive species in eastern North America?

While not definitive on the question of invasive status, we interpret our results as most consistent with *D. geminata* being an exotic invader in the Mid-Atlantic. Both eDNA and benthic data clearly indicate that *D. geminata* is not widespread in the region, despite availability of environmental conditions within its tolerance levels (Kumar *et al.*, 2009; Bothwell *et al.*, 2014). This is less consistent with the hypothesis that *D. geminata* is a native species that only forms blooms when triggered by a shift in environment. Rather, eDNA, microscopic analysis of benthic samples and previous visual surveys all are highly congruent in showing that *D. geminata* occurs sporadically at present, and when it does occur, it generally forms nuisance blooms.

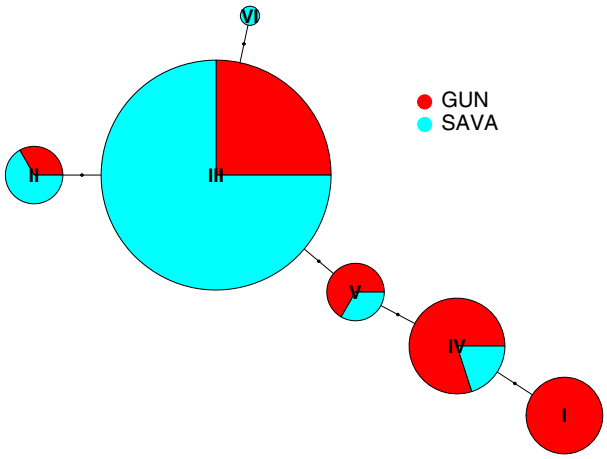

Figure 6 Haplotype network of 28 *Didymosphenia geminata* clones from a 662-bp alignment of rDNA sequences. Sites are Gunpowder Falls (GUN) and the Lower Savage River (SAVA).

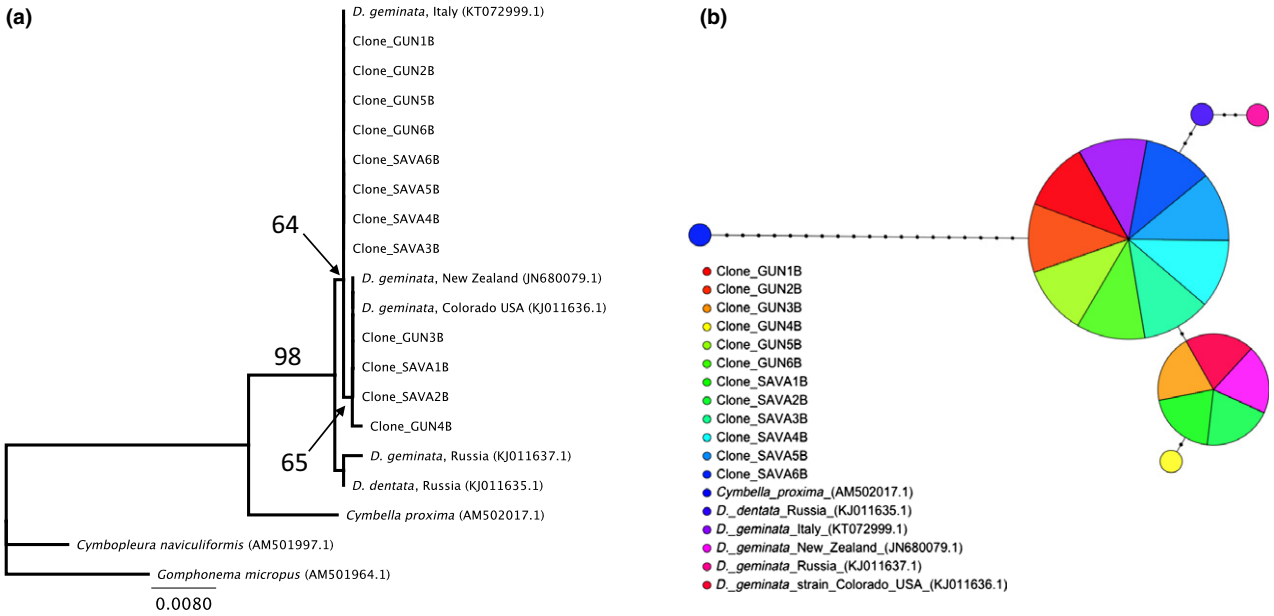

Figure 5 Genetic relationships among *Didymosphenia* lineages and closely related taxa based on a 951-bp alignment of rDNA (18S, 5.8S and ITS1 and 2). (a) Maximum-likelihood phylogeny for the tree with the highest log likelihood (−1726.58) showing relationships between *D. geminata* sequences from this study (SAVA and GUN), previously published sequences obtained from GenBank, and outgroup taxa. Support values are from 500 bootstrap replicates. (b) Haplotype network of *Didymosphenia* samples and outgroup *Cymbella proxima*.

The sequencing of clones from two positive sites in Maryland showed that *D. geminata* occurs as multiple genetic strains within sites rather than as a single clonal strain, and that haplotypes are shared between sites. Thus, nuisance blooms do not appear to result from the spread of a novel genetic variant. Rather, shared genetic diversity and lack of endemism may indicate high propagule pressure into the Mid-Atlantic from elsewhere. The identity of the source population(s) is not known, and *Didymosphenia* lineages were phylogenetically related to *D. geminata* collected from across the globe. One group of Mid-Atlantic samples was most closely related to *D. geminata* from Colorado, USA, and New Zealand, while the other set of Mid-Atlantic samples was associated with *D. geminata* from Italy (Fig. 5a,b). These two groups formed unique haplotypes in the haplotype network; however, bootstrap values in the ML phylogeny provided only moderate support.

The phylogeographic structure of *D. geminata* is not known, so it is unclear whether the shared relatedness of samples from distant geographic sources represents human-mediated dispersal or the natural pattern of diversity in different parts of *D. geminata*'s historical range. Biogeographic studies of diatoms have often assumed that species are ubiquitous in distribution with no obvious dispersal limitation (i.e. 'everything is everywhere'), and thus are predicted to lack regional patterns of endemism or genetic structure. However, recent studies have challenged this notion by showing that appreciable rates of endemism, dispersal limitation and/or intraspecific phylogeographic structure exists in diatoms (Kilroy *et al.*, 2007; Vanormelingen *et al.*, 2008; Kermarrec *et al.*, 2013). For example, the freshwater alga *Synura petersenii* is composed of regionally endemic yet cryptically diverse genetic populations, as well as one clade that is globally distributed (Boo *et al.*, 2010), and a very similar pattern has recently been reported for the diatom *Gomphonema parvulum,* a relative of the *Didymosphenia* genus (Kermarrec *et al.*, 2013). In the case of *D. geminata*, the pattern of few sites containing *D. geminata* coupled with multiple lineages co-existing within sites suggests one of three possible causes: (1) *D. geminata* is native to the region where it occurs sporadically as uncommon yet genetically diverse populations; (2) *D. geminata* is an exotic species early in its stages of invasion and its genetic diversity results from introduction from a high-diversity source region; or (3) it is an exotic species with diversity reflecting multiple introductions originating from genetically distinct sources. Additional sequencing from sites across the globe will greatly help clarify the biogeographic history of *D. geminata*, and potentially aid in identifying the source regions of introduced genotypes.

It is important to keep in mind the limitations of the current analysis for determining *D. geminata* genetic diversity and phylogenetic relationships. Here, we used molecular cloning and sequencing of nuclear ribosomal DNA (rDNA) to characterize *D. geminata* diversity. While the polymorphism we uncovered shows that *D. geminata* forms several haplotypes with global distributions, and that nuisance blooms are not restricted to a single clonal strain, the low levels of polymorphism limit additional insight into the genetic structure of *D. geminata*. Thus, inference of relatedness between Mid-Atlantic *D. geminata* samples and those from previous studies is necessarily constrained by analysis of a single locus. Future studies would greatly benefit from expanded sampling of geographic sites and additional sequence loci to produce a more refined picture of the phylogeographic structure of *D. geminata.*

Didymosphenia geminata occurs as regionally rare, locally invasive nuisance blooms

Contrary to our prediction if *D. geminata* is a widespread indigenous species at low abundance, we find evidence of only seven *D. geminata*-positive sites out of 102 spatially unique sites surveyed (6.8%). Our finding of *D. geminata* occurring rarely but at nuisance levels contrasts with other recent studies of *D. geminata* occupancy across regional watersheds. For example, in Alberta, Canada, 67 (88%) of 76 were *D. geminata*-positive sites with 34% of these showing bloom formation (Jackson, 2015). In New Zealand, Bray *et al.* (2016) report 45 (82%) of 55 sites positive for *D. geminata,* with 25% of sites occurring as blooms.

Of the seven Mid-Atlantic sites that tested positive for eDNA, six were already known to contain *D. geminata* from previous visual observations of nuisance blooms, while the remaining site (Trout Run) is a new record of *D. geminata* occurrence. Water chemistry at Trout Run was similar to other *D. geminata* -containing streams in our sample, except that pH (5.7) was relatively low relative to other streams (Fig. 4). Two benthic samples collected in Trout Run analysed with microscopy failed to detect *D. geminata*, despite apparent presence of eDNA in the water column. Disparity between *D. geminata* presence on the substrate and in the water column has been documented in previous studies. For instance, in a recent study in New Zealand, 27% of sites confirmed *D. geminata* suspended in the water column, but not in benthic sampling (Bray *et al.*, 2016). Additionally, Kilroy and Dale (2006) demonstrated an ability to detect *D. geminata* cells using drift net sampling techniques like those employed here at a distance up to 14 miles downstream of sparse colonies. Therefore, the most likely explanation is that *D. geminata* was present in the water column at Trout Run, yet not present on the substrate. Interestingly, another site where blooms have been previously documented, Big Hunting Creek, did not test positive for *D. geminata* eDNA during our sampling (April 2014). This corresponds with an absence of *D. geminata* blooms during monthly visual surveys conducted by National Park Service staff during 2014 at Big Hunting Creek (L. Donaldsen, personal communication), suggesting that local outbreaks may be episodic and possibly subject to extinction events (Simberloff & Gibbons, 2004).

The issue of detectability is a concern for determining occupancy when species are rare. If *D. geminata* is indeed non-native, then it is important to identify new infestations

before they become difficult to manage (Simberloff *et al.*, 2013). The minimum detection threshold for the qPCR assay used here has been determined previously by Cary *et al.* (2007), Cary *et al.* (2014). They reported sensitivity down to 1 *D. geminata* cell mL^{-1} of drift net sample. Given that we filtered an average of ~15,000 L per site concentrated into 500 mL, and then subsampled 1 mL for eDNA analysis, this equates to a lower detection limit of approximately 1 *D. geminata* cell per 30 L of unfiltered stream water. Thus, even with the relatively high sensitivity of our eDNA assay, it is possible that populations at low abundance went undetected. However, we also note the near-perfect correspondence between *D. geminata* detection based on eDNA and from microscopic investigation of benthic samples (22/23 samples in agreement, the exception being Trout Run), suggesting that the eDNA approach does not suffer from lower false-negative rates than would be expected from time-intensive microscopy investigation.

Environmental association with *D. geminata* occurrence

We found no association between *D. geminata* and water chemistry. Previous studies found that *D. geminata* is more likely to form blooms when SRP is low, generally < 2 ppb (Bothwell *et al.*, 2014). Our seven positives sites were not biased towards streams with low total P or SRP (Fig. 4), nor were *D. geminata*-positive sites confined to cold tailwaters with regulated base flow, as has also been suggested (Bray *et al.*, 2016). None of the Pennsylvania streams in the Susquehanna watershed containing *D. geminata* (Pine Creek, West Branch Pine Creek, Trout Run) are tailwater-influenced. *D. geminata* - positive sites trended towards lower sulphate, although the mean was still within the 95% confidence interval of the re-sampled distribution (Fig. 4). Previous studies hypothesized a connection between *D. geminata* abundance and sulphate, but in the opposite direction, with sulphate additions tending to increase *D. geminata* densities (Rost *et al.*, 2011). The small number of streams containing *D. geminata* limits more precise inference of environmental associations; thus, further work is necessary to determine the environmental preferences and tolerances of *D. geminata* in these waters, and whether they differ from other regions where stream chemistry has been associated with *D. geminata* blooms.

It is interesting to note that many of the sites that tested negative for *D. geminata* eDNA seem to fall roughly within the habitat suitability model projected by Kumar *et al.* (2009), suggesting that the distribution of *D. geminata* may not be at equilibrium with the availability of suitable habitat. Other studies have similarly shown spread of *D. geminata* through time, and caution that environmental association analyses early during invasion often do not reveal the potential distribution of the species (Kilroy & Unwin, 2011; Montecino *et al.*, 2016). Indeed, recent evidence of *D. geminata* expansion has been observed locally within the Pine Creek watershed. At the West Branch of Pine Creek at Crippen Run (WPIN CRIPPEN; Table S1), located ~1.2 km upstream of a known *D. geminata* population (WPIN RWQMN), *D. geminata* was not observed in two eDNA samples (April 2014 and 2015; this study) and five benthic samples analysed by microscopy (November 2013–May 2015; Shank *et al.*, 2016). However, *D. geminata* was later observed from benthic samples in November 2015 at this location (*D. geminata* density: 104.5 cells cm^{-2}; Shank *et al.*, 2016), suggesting that the species probably expanded its range during the period of observation.

CONCLUSIONS

Extensive sampling and analysis using both eDNA and microscopy revealed that *D. geminata* primarily occurs in the Mid-Atlantic where nuisance blooms have already been noted. We interpret its patchy occurrence in streams with blooms, along with DNA sequencing showing that Mid-Atlantic *D. geminata* are genetically diverse and closely related to lineages elsewhere across the globe, as most consistent with the hypothesis that *D. geminata* is an exotic species in our region. If correct, these results suggest that invasion by *D. geminata* is still in its early stages in the Mid-Atlantic region and not at equilibrium with the availability of suitable habitat, and that management efforts may still be effective at controlling its spread.

Our study also contributes to the general knowledge of biological invasions, and whether emerging nuisance species are native and responding to an anthropogenic change in environment, or rather are exotic and invasive due to recent introduction and spread. Genetic monitoring and population genetic analysis provide powerful additions to the suite of tools needed to address these conservation challenges.

ACKNOWLEDGEMENTS

We thank Jason Cessna and Nathan Reinhart for sample collection, Regina Trott for genetic analyses, Jeff Zimmerman for G.I.S. analyses and the Maryland Biological Stream Survey for stream chemistry and land use data. Jane Molofsky, Ron Klauda and three referees provided helpful comments on the manuscript. This study was funded by the Mid-Atlantic Panel on Aquatic Invasive Species (S.K. and R.H.), and the Pennsylvania SeaGrant (M.S. and M.P).

REFERENCES

American Public Health Association (APHA) (2005) *Standard methods for the examination of water and wastewater*. American Public Health Association, Washington, DC, USA.

Barbour, M. & Gerritsen, J. (1999) Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, 2nd edition. U.S. Environmental Protection Agency, Office of Water, EPA 841-B-99-002, U.S., Washington, D.C.

Bergey, E. & Spaulding, S. (2015) Didymosphenia: it's more complicated. *Aquatic Invasions*, **65**, 225–225.

Blackburn, T., Pyšek, P. & Bacher, S. (2011) A proposed unified framework for biological invasions. *Trends in Ecology and Evolution*, **26**, 333–339.

Blanco, S. & Ector, L. (2009) Distribution, ecology and nuisance effects of the freshwater invasive diatom *Didymosphenia geminata* (Lyngbye) M. Schmidt: a literature review. *Nova Hedwigia*, **88**, 347–422.

Boo, S.M., Kim, H.S., Shin, W., Boo, G.H., Cho, S.M., Jo, B.Y., Kim, J.H., Kim, J.H., Yang, E.C., Siver, P.A., Wolfe, A.P., Bhattacharya, D., Andersen, R.A. & Yoon, H.S. (2010) Complex phylogeographic patterns in the freshwater alga *Synura* provide new insights into ubiquity vs. endemism in microbial eukaryotes. *Molecular Ecology*, **19**, 4328–4338.

Bothwell, M.L., Taylor, B.W. & Kilroy, C. (2014) The Didymo story: the role of low dissolved phosphorus in the formation of *Didymosphenia geminata* blooms. *Diatom Research*, **29**, 229236.

Boyer, C.S. (1916) *The Diatomaceae of Philadelphia and vicinity*. J.B. Lippincott Co., Philadelphia, PA.

Boyer, C.S. (1927) Supplement: Synopsis of North American Diatomaceae. Part II. Naviculatae, Surirellatae. *Proceedings of the Academy of Natural Sciences Philadelphia*, **79**, 229–583.

Bray, J., Harding, J.S., Kilroy, C., Broady, P. & Gerbeaux, P. (2016) Physicochemical predictors of the invasive diatom *Didymosphenia geminata* at multiple spatial scales in New Zealand rivers. *Aquatic Ecology*, **50**, 1–14.

Bruder, K. & Medlin, L.K. (2007) Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). *I. The genus Placoneis. Nova Hedwigia*, **84**, 331–352.

Buczkowski, G. (2010) Extreme life history plasticity and the evolution of invasive characteristics in a native ant. *Biological Invasions*, **12**, 3343–3349.

Carey, M.P., Sanderson, B.L., Barnas, K.A. & Olden, J.D. (2012) Native invaders – challenges for science, management, policy, and society. *Frontiers in Ecology and the Environment*, **10**, 373–381.

Cary, S.C., Hicks, B.J., Coyne, K.J., Rueckert, A., Gemmill, C.E.C., Barnett, C.M.E. (2007) A sensitive genetic-based detection capability for Didymosphenia geminata (Lyngbye) M. Schmidt: Phases Two and Three. CBER Contract Report 62. Client Report prepared for MAF Biosecurity New Zealand.. Centre for Biodiversity and Ecology Research, University of Waikato, Hamilton. 87 pp

Cary, S., Coyne, K., Rueckert, A. & Wood, S. (2014) Development and validation of a quantitative PCR assay for the early detection and monitoring of the invasive diatom *Didymosphenia geminata*. *Harmful Algae*, **36**, 63–70.

Charles, D.F., Knowles, C. & Davis, R.S. (2002) Protocols for the analysis of algal samples collected as part of the US Geological Survey National Water-Quality Assessment Program.

Darling, J.A. & Mahon, A.R. (2011) From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, **111**, 978–988.

Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **19**, 1792–1797.

Gillis, C.-A. & Lavoie, I. (2014) A preliminary assessment of the effects of *Didymosphenia geminata* nuisance growths on the structure and diversity of diatom assemblages of the Restigouche River basin, Quebec, Canada. *Diatom Research*, **29**, 281–292.

Jackson, L. (2015) Environmental constraints on *Didymosphenia geminata* occurrence and bloom formation in Canadian Rocky Mountain lotic systems. *Canadian Journal of Fisheries and Aquatic Sciences*, **73**, 964–972.

Jerde, C.L., Mahon, A.R., Chadderton, W.L. & Lodge, D.M. (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.

Jones, L., Uyua, N. & Manrique, J. (2015) The peril of PCR inhibitors in environmental samples: the case of *Didymosphenia geminata*. *Biodiversity and Conservation*, **24**, 1541–1548.

Kearse, M., Moir, R., Wilson, A. & Stones-Havas, S. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Keck, F., Rimet, F., Franc, A. & Bouchez, A. (2016) Phylogenetic signal in diatom ecology: perspectives for aquatic ecosystems biomonitoring. *Ecological Applications*, **26**, 861–872.

Kermarrec, L., Bouchez, A., Rimet, F. & Humbert, J.F. (2013) First evidence of the existence of semi-cryptic species and of a phylogeographic structure in the *Gomphonema parvulum* (Kutzing) Kutzing complex (Bacillariophyta). *Protist*, **164**, 686–705.

Kilroy, C. & Bothwell, M. (2011) Environmental control of stalk length in the bloom-forming, freshwater benthic diatom *Didymosphenia geminata* (Bacillariophyceae). *Journal of Phycology*, **47**, 981–989.

Kilroy, C. & Unwin, M. (2011) The arrival and spread of the bloom-forming, freshwater diatom, *Didymosphenia geminata*, in New Zealand. *Aquatic Invasions*, **6**, 249–262.

Kilroy, C., Biggs, B.J.F. & Vyverman, W. (2007) Rules for macroorganisms applied to microorganisms: patterns of endemism in benthic freshwater diatoms. *Oikos*, **116**, 550–564.

Kilroy, C.A. & Dale, M. (2006) A comparison of sampling methods for the detection of the invasive alga Didymosphenia geminata in New Zealand rivers. National Institute of Water and Atmospheric Research Client Report no. CHC2006-078.

Kilroy, C., Larned, S.T. & Biggs, B.J.F. (2009) The non-indigenous diatom *Didymosphenia geminata* alters benthic communities in New Zealand rivers. *Freshwater Biology*, **54**, 1990–2002.

Kumar, S., Spaulding, S.A., Stohlgren, T.J., Hermann, K.A., Schmidt, T.S. & Bahls, L.L. (2009) Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. *Frontiers in Ecology and the Environment*, **7**, 415–420.

Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.

Lavery, J., Kurek, J., Ruhland, K., Gillis, C., Pisaric, M. & Smol, J. (2014) Exploring the environmental context of recent *Didymosphenia geminata* proliferation in Gaspésie, Québec, using paleolimnology. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 1–11.

Lodge, D.M., Turner, C.R., Jerde, C.L., Barnes, M.A., Chadderton, L., Egan, S.P., Feder, J.L., Mahon, A.R. & Pfrender, M.E. (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, **21**, 2555–2558.

Montecino, V., Molina, X., Bothwell, M., Muñoz, P., Carrevedo, M.L., Salinas, F., Kumar, S., Castillo, M.L., Bizama, G. & Bustamante, R.O. (2016) Spatio temporal population dynamics of the invasive diatom *Didymosphenia geminata* in central-southern Chilean rivers. *Science of the Total Environment*, **568**, 1135–1145.

Nakov, T., Ruck, E.C., Galachyants, Y., Spaulding, S.A. & Theriot, E.C. (2014) Molecular phylogeny of the Cymbellales (Bacillariophyceae, Heterokontophyta) with a comparison of models for accommodating rate variation across sites. *Phycologia*, **53**, 359–373.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, **26**, 419–420.

Rost, A. & Fritsen, C. (2014) Influence of a tributary stream on benthic communities in a *Didymosphenia geminata* impacted stream in the Sierra Nevada, USA. *Diatom Research*, **29**, 249–257.

Rost, A., Fritsen, C. & Davis, C. (2011) Distribution of freshwater diatom *Didymosphenia geminata* in streams in the Sierra Nevada, USA, in relation to water chemistry and bedrock geology. *Hydrobiologia*, **665**, 157–167.

Shank, M., Potapova, M., Maloney, K., Honeyfield, D. & Spooner, D. (2016) Didymosphenia geminata *in Pennsylvania: an investigation of current and historic distribution, habitat suitability, and nutrient content*. Final Project Report to the Pennsylvania, Sea Grant.

Simberloff, D. & Gibbons, L. (2004) Now you see them, now you don't! – population crashes of established introduced species. *Biological Invasions*, **6**, 161–172.

Simberloff, D., Martin, J. & Genovesi, P. (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology and Evolution*, **28**, 58–66.

Spaulding, S. & Elwell, L. (2007) Increase in nuisance blooms and geographic expansion of the freshwater diatom *Didymosphenia geminata*: recommendations for response. White Paper, 1–33.

Sundareshwar, P.V., Upadhayay, S., Abessa, M., Honomichl, S., Berdanier, B., Spaulding, S.A., Sandvik, C. & Trennepohl, A. (2011) *Didymosphenia geminata*: algal blooms in oligotrophic streams and rivers. *Geophysical Research Letters*, **38**, L10405.

Taylor, B.W. & Bothwell, M.L. (2014) The origin of invasive microorganisms matters for science, policy, and management: the case of *Didymosphenia geminata*. *BioScience*, **64**, 531–538.

Taylor, B.W. & Bothwell, M.L. (2015) Response to Bergey and spaulding. *BioScience*, **65**, 226–226.

Uyua, N., Manrique, J. & Jones, L. (2014) An optimized DNA extraction protocol for benthic *Didymosphenia geminata*. *Journal of Microbiological Methods*, **104**, 12–18.

Valéry, L., Fritz, H., Lefeuvre, J. & Simberloff, D. (2008) In search of a real definition of the biological invasion phenomenon itself. *Biological Invasions*, **10**, 1345–1351.

Vanormelingen, P., Verleyen, E. & Vyverman, W. (2008) The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, **17**, 393–405.

Wilson, J., Dormontt, E. & Prentis, P. (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology and Evolution*, **24**, 136–144.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Site data and results for Didymosphenia geminata monitoring in Maryland (MD) and Pennsylvania (PA), USA.

BIOSKETCH

Stephen R. Keller's research focuses on the interaction between genomic variation in natural populations and environmental change, especially genetic aspects of species invasions and climate change adaptation (http://www.uvm.edu/~srkeller/). The interdisciplinary team of co-authors in this study reflect our individual backgrounds in genetics, aquatic ecology and the ecology and taxonomy of diatoms.

Author contributions: S.K. and R.H. conceived the idea and designed the sampling and analysis, with M.S. contributing additional samples; S.K., R.H., M.S. and M.P. collected the data; S.K. and R.H. conducted the statistical analyses; S.K. led the manuscript writing, with contributions from R.H., M.S. and M.P.

Editor: Aibin Zhan