Global high-throughput genotyping of organellar genomes reveals insights into the origin and spread of invasive starry stonewort (*Nitellopsis obtusa*)

Robin S. Sleith · Kenneth G. Karol

Received: 9 April 2020 / Accepted: 1 July 2021 / Published online: 16 July 2021
© The Author(s) 2021

Abstract Aquatic invasive species are damaging to native ecosystems. Preventing their spread and achieving comprehensive control measures requires an understanding of the genetic structure of an invasive population. Organellar genomes (plastid and mitochondrial) are useful for population level analyses of invasive plant distributions. In this study we generate complete organellar reference genomes using PacBio sequencing, then use these reference sequences for SNP calling of high-throughput, multiplexed, Illumina based organellar sequencing of fresh and historical samples from across the native and invasive range of *Nitellopsis obtusa* (Desv. in Loisel.) J.Groves, an invasive macroalgae. The data generated by the analytical pipeline we develop indicate introduction

to North America from Western Europe. A single nucleotide transversion in the plastid genome separates a group of five samples from Michigan and Wisconsin that either resulted from introductions of two closely related genotypes or a mutation that has arisen in the invasive range. This transversion will serve as a useful tool to understand how *Nitellopsis obtusa* moves across the landscape. The methods and analyses described here are broadly applicable to invasive and native plant and algae species, and allow efficient genotyping of variable quality samples, including 100-year-old herbarium specimens, to determine population structure and geographic distributions.

Keywords Characeae · Invasive species · Chloroplast genome · Mitochondria genome

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10530-021-02591-8>.

R. S. Sleith (✉) · K. G. Karol
Lewis B. and Dorothy Cullman Program for Molecular Systematics, New York Botanical Garden, Bronx, NY, USA
e-mail: rsleith@bigelow.org

R. S. Sleith
Department of Biological Sciences, Smith College, Northampton, MA, USA

R. S. Sleith
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA

Introduction

Aquatic invasive species threaten the biodiversity, functioning, and ecosystem services of freshwater ecosystems across the world. It is estimated that damage caused by aquatic invasive species exceeds \$100 million annually and they have contributed directly to the decline of 42% of the threatened and endangered species in the U.S.A. (Early et al. 2016; Pimentel et al. 2005). Appropriate management and

control of an invasive species requires an understanding of the frequency of introduction, the size of the introduced population, and the subsequent pattern of spread (Cristescu 2015; Fitzpatrick et al. 2011). The genetic structure of invasive populations has been shown to have importance for the fitness and invasion potential in the introduced range (Lavergne and Molofsky 2007; Vandepitte et al. 2014). Novel invasive genotypes have the potential to not only increase the fitness of invasive species, but unique genotypes may be more or less susceptible to control treatments (Moody et al. 2008). Furthermore, a beneficial mutation may rapidly spread in a population because of its selective advantage in response to treatment efforts; for example, the aquatic invasive plant *Hydrilla* evolved herbicide resistance through somatic mutations of the gene *pds*, which encodes the enzyme phytoene desaturase, which the herbicide fluridone was initially designed to inhibit (Michel et al. 2004). Invasive species with low genetic diversity can also be successful and become dominant across large areas (e.g. Hagenblad et al. 2015; Hardion et al. 2014). Thus, a clear understanding of the genetic characteristics of damaging aquatic invasive species is essential to their control and eradication.

The macroalga *Nitellopsis obtusa* (Desv. in Loisel.) J.Groves (starry stonewort) has recently become an invasive species of particular concern in North America. *Nitellopsis obtusa* is a species of green algae in the macrophyte family Characeae. *Nitellopsis obtusa* is native to Europe and Asia and is considered rare and endangered in eastern Europe and Japan (Kato 2005; Soulié-Märsche et al. 2002). It anchors itself in the sediment, can grow more than two meters in height, and has specialized star-shaped structures, termed bulbils, that allow prolific vegetative reproduction (Bharathan 1987; Wood and Imahori, 1965). In its native range *Nitellopsis obtusa* occurs in both deep and shallow lakes, rivers, and oxbows (reviewed in Larkin et al. 2018). In both the native and invasive range *N. obtusa* occurs in calcareous, neutral to alkaline, mesotrophic to eutrophic waters, and has been found to tolerate brackish environments in the Baltic Sea (Larkin et al. 2018). The earliest collections of *N. obtusa* in North America were in the Saint Lawrence River in 1974 and 1978, where it was hypothesized to be introduced by ballast water from trans-oceanic shipping (Geis et al. 1981; Karol and Sleith 2017). In 1983 it was recorded in the St. Clair-

Detroit River system in Michigan (Schloesser et al. 1986), and has since spread throughout Michigan's inland lakes, northern Indiana, western New York and in isolated areas of Minnesota, Pennsylvania, Vermont, and Wisconsin (Larkin et al. 2018). It is unclear how *N. obtusa* is moved from lake to lake; both humans and waterfowl have been implicated in its spread (Sleith et al. 2015). In North America, only male individuals have been observed, suggesting that the invasive population is being spread clonally from fragments on boats or trailers, however not all *N. obtusa* specimens display reproductive structures, and the conditions required for formation of female reproductive structures may not be met in the invasive range (Larkin et al. 2018).

The impacts of *Nitellopsis obtusa* on North American ecosystems have yet to be fully understood (Larkin et al. 2018). Few studies have examined the ecology of *N. obtusa* in North American ecosystems, but in a small scale study Brainard and Schulz (2017) found that native plant species richness and biomass decreased with increasing *N. obtusa* abundance. Escobar et al. (2016) examined the climatic niche of *N. obtusa* and found that at a broad, continental scale, large portions of North America may be susceptible to invasion. Sleith et al. (2018) used water chemistry variables to show that *N. obtusa* may have a similar ecological niche to *Chara contraria*, a native species that inhabits hard water sites across North America. These findings indicate that much habitat remains for *N. obtusa* to colonize and demonstrate the need for a clear understanding of the spread vectors and genetic structure of *N. obtusa* across the world.

In this study, organellar genome sequencing is used to determine the genetic signature of invasive and native populations of *Nitellopsis obtusa*. Organellar genomes are typically compact and similar to bacterial genomes, reflecting the endosymbiotic origin of chloroplasts and mitochondria (Gray and Doolittle 1982; Sagan 1967). They have great value in evolutionary, systematic, and population studies of photosynthetic organisms (Straub et al. 2012). Markers such as *rbcL* (chloroplastic) and *cox1* (mitochondrial) have served as barcodes across photosynthetic organisms, including Characeae (Hall et al. 2010). While these protein coding markers can be useful for detecting interspecific variation, they often lack the resolution to discriminate intraspecific variation in green algae (Hall et al. 2010). Advances in high-throughput

sequencing and the high copy number of organellar genomes have allowed cost-effective sequencing of whole organellar genomes (McCormack et al. 2013). These whole genome sequences enable analysis of variable non-coding regions such as introns and intergenic regions that are useful for detecting intraspecific variation. Whole organellar genomes have been used in population studies of model organisms such as rice (Tong et al. 2016), as well as non-model systems such as *Theobroma* and invasive species such as *Ageratina adenophora* (Kane et al. 2012; Nie et al. 2012). Reference-based SNP calling approaches use an identical or closely related reference genome to map re-sequencing reads and call SNPs/indels between the reference and the reads. These techniques have the advantage that variable quality input data (such as > 100-year-old herbarium specimens) can be incorporated into analyses that account for the high rates of missing data specific to historical specimens (Staats et al. 2013).

In this study, reference organellar genomes were generated and used together with Illumina based re-sequencing of 60 samples from across the native (including historical samples) and invasive range to test the number of invasions of *Nitellopsis obtusa* to North America, the origins of invasive populations, and the pattern of spread across the landscape. Specifically, the following hypotheses were addressed: (1) North American *N. obtusa* populations are European in origin (2) a single introduction has occurred, and (3) subsequent spread has been clonal. Finally, we ask whether information from organellar genomes can track the spread of invasive individuals from lake to lake in North America.

Methods

Population sampling

Fresh and preserved tissue from the invasive range were collected following the methods outlined in Sleith et al. (2015, 2018). Samples were returned to the lab in field collected water and cleaned of epiphytes, a fragment was frozen for DNA extraction, and the remaining specimen was mounted on an herbarium sheet, dried, and accessioned in The William and Lynda Steere Herbarium (NY). Additional contemporary samples from the invasive and native range were

sampled with the help of national and international collaborators. To increase the study's geographic and temporal range, historical samples (collections beginning in 1880) including the type specimen, were sampled from The William and Lynda Steere Herbarium (NY). In total these represent 35 samples from the invasive North American range (4 herbarium specimens—collected 2015–2016, and 31 fresh tissue samples—collected 2011–2017), and 25 samples from across the native range of Europe and Asia (18 herbarium specimens—representing collection years 1880–2017, and 7 fresh tissue samples—representing collection years 1974–2016) (Fig. 1c). All sample collection information is listed in Table S1.

DNA extraction

DNA isolation followed the methods previously described by Pérez et al. (2014). Briefly, 0.1 g of tissue was obtained from frozen and herbarium specimens. Genomic DNA was extracted using the Nucleon Phytopure DNA extraction kit (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA), following the manufacturers protocol for small samples.

Reference genomes

Reference genomes were generated from material collected from Upper Little York Lake in Cortland County, New York, U.S.A. and housed at The New York Botanical Garden, deposited in the herbarium under accession NY02666457. Four Pacific Biosciences (PacBio) SMRT cells were run on a PacBio Sequel sequencer to generate 6.6 Gb of sequences. PacBio reads were trimmed and assembled using Canu v. 2.1.1 with genomeSize = 200k, corOutCoverage = 1,000,000, and readSamplingCoverage = 1,000,000 (Koren et al. 2017). Canu contigs were checked for accuracy by aligning Illumina reads from the same sample (average insert size of 300 bp, trimmed using Trimmomatic v. 0.36 with the following settings: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:TRUE SLIDINGWINDOW:4:15 MINLEN:36 [Bolger et al. 2014]), and assessing discrepancies using POLCA v. 3.4.2 (Zimin and Salzberg 2020). Chloroplast and mitochondrial contigs were annotated using Geneious v9 (Kearse et al. 2012) with the *Chara vulgaris* chloroplast (NC_008097.1) and mitochondrial (NC_005255.1)

genomes as a reference. Genome maps were created using OrganellarGenomeDRAW (Lohse et al. 2013) and circos (Krzywinski et al. 2009).

Re-sequencing genomes

Sequence libraries were generated using the Roche KAPA HyperPlus kit, multiplexed at up to 40 samples per lane and sequenced on an Illumina NextSeq500 in PE101 and PE150 High Output mode at Cold Spring Harbor Laboratory.

SNP calling and statistical analyses

Intra-individual chloroplast and mitochondrial polymorphisms have been detected in whole genome sequencing generated from short reads (100–150 bp) (Hazkani-Covo et al. 2010; Scarcelli et al. 2016; Sardina 2017). Scarcelli et al. (2016) determined that genome transfers, from the chloroplast to the mitochondrial or nuclear genome (known as nuclear plastid DNA), are likely responsible for observed intra-individual polymorphisms from whole genome sequencing. The bioinformatic recommendations of Scarcelli et al. (2016) were followed to exclude intra-individual polymorphisms from the SNP calling pipeline (Figure S8). Evidence of intra-individual polymorphism was assessed by assembling Canu corrected and trimmed PacBio reads to final reference genomes using BLASR v1.0 (Chaisson and Tesler 2012). These longer reads allow for clearer elucidation of genome transfers and variation (Scarcelli et al. 2016). As an additional test, DNA was extracted and Illumina sequenced from the male reproductive structure (antheridium), which should have a higher proportion of nuclear DNA to organellar DNA (Maszewski and Kołodziejczyk 1991).

Reads from each sample were aligned to reference genomes using BWA MEM v0.7.12 with -M and -B 4 options (Li and Durbin 2009). Reads with > 10% clipping by BWA were removed from alignments using samjdk v. af51aa30d. SAMTOOLS v1.5 with option -B was used to create an mpileup file (Li et al. 2009). Single Nucleotide Polymorphisms (SNPs) and Insertion-Deletions (indels) were called using VarScan v2.4.3 with -min-var-freq and -min-freq-for-hom set to 0.9 (Koboldt et al. 2012). The resulting Variant Call Format file (vcf) was filtered using VCFtools v0.1.13 (Danecek et al. 2011) to exclude genotypes

called below 50% across all individuals, sites with coverage less than 3, sites with a minor allele count less than 2, and ribosomal regions due to conservation of these regions across the tree of life. To assess the influence of missing data two sets of analyses were run: “liberal” where individuals with more than 75% missing data were excluded from downstream analyses, and “strict” where individuals with more than 10% missing data were excluded from downstream analyses.

Genetic distance trees (UPGMA) with bootstrap support values for 1000 pseudoreplicates were generated using the `aboot` function of the `poppr` package in R (Kamvar et al. 2014, 2015). Discriminant Analysis of Principal Components (DAPC) was performed using the `dapc` function of the R package `ade4` (Jombart 2008). DAPC is a model free approach that involves a principal component analysis (PCA) followed by a discriminate analysis (DA) to maximize discrimination between groups. DAPC was performed on a priori, geography-based, population assignments to test if the dataset matched these population assignments. The number of retained principal components was determined using the cross-validation approach `xvalDapc` (Jombart et al. 2010).

Primer design

As we detected a substitution in the *rps16-odpB* region differentiating samples in the invasive range, we designed primers targeting 944 bases surrounding the T to G base substitution (specifically a pyrimidine to purine transversion). Primers were designed using NCBI Primer-BLAST (Ye et al. 2012). We tested the primers on four samples from the area of Michigan known to have both genotypes, using a gradient of annealing temperatures from 55 to 68 °C; following the PCR conditions described by Pérez et al. (2014). PCR products were visualized with gel electrophoresis and two bands from each sample were sent for Sanger sequencing at Macrogen USA.

Results

Here we present the first sequenced organellar genomes of the genus *Nitellopsis*. The Canu assembly of PacBio reads from a North American sample resulted in one contig that contained the complete

chloroplast genome, and one contig that contained the complete mitochondrial genome. The chloroplast genome (191,757 bp, GenBank Accession: MW556320) was a quadripartite molecule consisting of a large single copy region (LSC 146,109 bp) a small single copy region (SSC 24,456 bp) and two inverted repeats (IR 10,596 bp each) (Figure S1). The mitochondrial genome was 60,237 bp (GenBank Accession: MW556321 Figure S3). Compared to the previously published *Chara vulgaris* genomes (Turmel 2003, 2006) both genomes had identical gene content in the same order (Table S2 and S3).

Illumina sequencing of multiplexed individuals resulted in a variable number of reads per sample, ranging from 3.8 million to 39 million paired reads (Table S1). The number of total reads generated for a sample did not correlate strongly with number of reads mapping to the chloroplast or mitochondria (Table S1). Read numbers mapping to chloroplast vs. mitochondrial genomes were correlated ($y = 0.22x + 20,615$ R^2 0.64). The year in which a specimen was collected was roughly correlated to the number of reads mapping to the organellar genomes, with older specimens having fewer reads, and a lower proportion of organellar reads to total reads (Table S1). For recently collected samples (e.g. 2017 herbarium samples collected in Germany) the number of organellar reads was not correlated with whether a sample was collected from fresh or dried material. Using the chloroplast genome as a reference, VarScan called 515 SNPs/indels; after filtering, 230 SNPs/indels remained across all individuals from the native and invasive range (Figure S2). Fifteen samples with greater than 10% missing data were removed before conducting the “strict” analysis: 9 herbarium samples (representing collection years 1880–2000) and 6 fresh samples (representing collection years 2003–2017) (Table S1). One sample had greater than 75% missing data and was removed before conducting the “liberal” analysis (an herbarium sample collected in 1937).

The mitochondrial genome showed strong evidence of intra-individual polymorphism. When PacBio reads were aligned to the reference genome, regions of polymorphism could be detected when visually inspecting the alignment. Many polymorphisms were found in the *yejV* gene region, and this region was investigated further. The PacBio reads that aligned to the mitochondria ranged in size from 10 to 33 kb, however, the polymorphic sites were only present in

reads with alignment lengths of 3 kb or less, evidence that this gene may have been transferred from the mitochondrial to nuclear genome. Additionally, high levels of intra-individual polymorphism were detected when aligning Illumina reads from the male reproductive structure (a sample with higher nuclear to mitochondrial DNA ratio) to the mitochondrial genome.

Using the mitochondrial genome as a reference VarScan called 153 SNPs/indels; after filtering 45 SNPs/indels remained. Five samples with greater than 10% missing data were removed before conducting the “strict” analysis. No samples had greater than 75% missing data and so the “liberal” analysis included all samples.

The UPGMA tree of chloroplast data recovered distinct groups that corresponded to broad geographic areas (Fig. 1 A [“strict”]; Figure S4 [“liberal”]) Samples from the native range of Europe and Asia generally clustered by region but with some variability (e.g. two samples from Japan are more closely clustered with European samples compared to other samples from Asia). North American samples were most closely clustered to samples from France. North American samples showed near-identical similarity, with five samples from Michigan and Wisconsin separated from all other samples (including other samples collected in MI and WI) by a single T to G transversion in the intergenic region between *rps16* and *odpB* (Fig. 1b).

DAPC of chloroplast data recovered similar patterns as the UPGMA tree. All North American samples clustered tightly together, with samples from Germany and France clustering most closely to North American samples (Fig. 2 [“strict”]; Figure S5 [“liberal”]). Samples from Asia and Sweden were consistently farthest away from North American samples, indicating that these populations were more dissimilar.

The UPGMA tree of mitochondrial data also showed broad geographic patterns, with North American samples clustered together, and samples from Western Europe most closely clustered to the North American samples (Fig. 3 [“strict”]; Figure S6 [“liberal”]). The DAPC of mitochondrial data clustered all North American samples together, with much variability detected in European clusters (Fig. 4 [“strict”]; Figure S7 [“liberal”]).

Primers were designed to amplify the region containing the transversion that separates Midwest

Fig. 2 DAPC of chloroplast genome SNPs, using only samples with 10% or less missing data (strict). Lines and inertia ellipses connect samples from identical populations. All North American samples cluster closely (orange boxes for NorthEast

U.S.A., orange stars for MidWest U.S.A., and blue boxes for Canada). The closest samples to North America are Kazakhstan (red), France (dark green), and Germany (pink)

USA samples (Forward: 5'-AATCACTTGAGCCG-TATGAGTTG-3', Reverse: 5'-AACCAAAT-CAAATTGTAGAAATCGC-3'). Primer efficacy was assessed by gel electrophoresis and Sanger sequencing. All samples amplified, with 62° C identified as the optimal annealing temperature. Only two samples sequenced successfully: sample NY02666588, from Joslin Lake, Washtenaw County, Michigan was found to have the T to G transversion, while sample NY02666590 from Silver Lake, Washtenaw County, Michigan did not.

Discussion

The UPGMA tree and DAPC analyses based on chloroplast and mitochondrial genome data both demonstrated that North American samples are more closely clustered with European samples; a sample from France appears to be most similar, but not identical, to the North American genotype (Fig. 1C). This indicates that the invasive population of *Nitellopsis obtusa* in North America likely resulted from a European introduction, perhaps through trans-oceanic shipping as previously hypothesized (Geis et al. 1981). Historical records of *Nitellopsis obtusa* in brackish coastal environments of Europe indicate that conditions existed for ballast water transportation to be feasible (Larkin et al. 2018). Furthermore, the broad

Fig. 3 UPGMA tree of mitochondrial genome SNPs, using only samples with 10% or less missing data (strict), shows a cluster of identical North American samples (blue segment). Text colors represent geographic areas, including within North America (blue—Canada, Orange—Northeast U.S.A., lighter orange

Midwest U.S.A.). Bootstrap values from 1000 pseudoreplicates (if greater than 70) are indicated by filled circles at nodes. NY barcode 02666457 is the reference genome for the analysis. NY barcode 02023152 is the type specimen for *Nitellopsis obtusa*

range in latitude and longitude recorded for *N. obtusa* in the native range indicates that the species has a relatively broad ecological niche that could make it a successful invader of North America (Escobar et al. 2016; Larkin et al. 2018). Our analysis indicates that regions of France and Western Europe were consistently most closely clustered with North American samples, but further sampling in the native range is needed to confirm which exact native population is most closely related to the invasive population.

Within the invasive range of Canada, the Midwest, and Northeast U.S.A neither the UPGMA tree or the DAPC analyses differentiated samples, with the exception of one substitution (the T to G transversion

separating five samples in Michigan and Wisconsin from all other North American samples). The near identical nature of both the chloroplast and mitochondrial genomes in the invasive range is strong evidence for either: (1) the introduction of a single genotype that has undergone a mutation in North America, or (2) the introduction of two very closely related genotypes (likely from the same general location, due to the observed variability in the native range). The failure to detect the transversion in the eastern invasive range indicates that a second introduction either occurred further west than the previous, or it exists in lower frequency and has evaded detection thus far. Alternatively, this mutation could have arisen in the Midwest

Fig. 4 DAPC of mitochondrial genome SNPs, using only samples with 10% or less missing data (strict). All North American samples cluster close to each other (orange boxes for NorthEast U.S.A., orange stars for MidWest U.S.A., and blue boxes for Canada)

region after an initial introduction and spread of a single population. Following the rise of the mutation this genotype could then have undergone secondary spread in the Michigan and Wisconsin region. Further sampling is needed across this area to fully characterize this phenomenon (Fig. 1b). The use of the primers presented here that target the transversion allow a low-cost (Sanger sequencing) method to detect the genotype, enabling characterization of the distribution of this population across Michigan, Wisconsin and the eastern range. Indeed, two of the samples that were used to validate the primers show that lakes 8 km apart in Washtenaw County Michigan harbor different genotypes of *N. obtusa* as assessed by the transversion.

The organellar genomes of *Nitellopsis obtusa* presented here add a new genus to three previously published organellar genomes from Characeae: *Chara braunii*, *Chara vulgaris*, and *Nitella hyalina*

(Nishiyama et al. 2018; Turmel 2003, 2006). These genomes are conserved in size, gene content and order, despite hundreds of millions of years of evolution separating them. Historical samples as old as 1880 were successfully sequenced, though with proportions of missing data increasing with age. Specimens from herbaria allow increased temporal and spatial sampling and represent a rich trove of information for invasive species research (Vandepitte et al. 2014; Zeng et al. 2018).

Though the SNP calling pipeline accounts for intra-individual variation as described by Scarcelli et al. (2016), we used long reads to assess intra-individual variation in the reference genomes, particularly the mitochondria which showed the most evidence of this phenomenon. Intra-individual polymorphism in the *yejV* region (a gene involved in cytochrome c biogenesis) may be the result of transfer from the

mitochondrial to the nuclear genome, a phenomenon reported in other Charophyte algae (Turmel et al. 2002). The observed polymorphisms likely arise from alignment of short segments of genomic regions that have been transferred from the mitochondrial genome. Once transferred, these segments are released from selective pressure and can accumulate mutations (Leister 2005). In this study, the regions of *yeyV* that aligned were flanked by regions with no homology to the mitochondrial or chloroplast genome, signifying that these regions have likely been transferred to the nuclear genome. This, combined with the increased polymorphisms observed from the male reproductive structure, indicates that intra-individual polymorphism likely arises from transfer of DNA from organellar genomes to the nucleus. When reconstructing the evolutionary history of organellar genomes, it is therefore important to carefully examine sites showing evidence of intra-individual polymorphism.

Compared to other plant and algae species the organellar genomes of *Nitellopsis obtusa* are remarkably conserved across the planet. This could be due to its ability to readily self-propagate from fragments and bulbils, but could also be due to high dispersal rates from migratory waterfowl in the native range (Proctor 1962, 1968). The observed low diversity in the native range and very low diversity in the invasive range may indicate that *N. obtusa* has high phenotypic plasticity, a phenomenon observed in other aquatic invasive species (Riis et al. 2010), allowing two near-identical genotypes to colonize a large portion of North America. Low diversity populations or clones are not uncommon among invasive species, a single clone of the marine alga *Caulerpa taxifolia* dominates large areas of the Mediterranean (Wiedenmann et al. 2001). Terrestrial examples include *Arundo donax* (Hardion et al. 2014) and *Phragmites* (Saltonstall 2002), each with invasive haplotypes that have become invasive across large regions. The low diversity of *N. obtusa* in the invasive range also has management implications: invasive species that have low diversity or are clones may be better candidates for treatments (biological or chemical) that rely on genetic factors (Ward et al. 2008). Though *N. obtusa* may have strong phenotypic plasticity, the low observed diversity in the invasive range could impact how the species responds to climate change in North America, as genetic diversity and local adaptation are important determinants of species responses to climate change (Savolainen et al.

2013). Finally, methods that survey many neutral loci such as whole nuclear genome sequencing or restriction site associated sequencing have more ability to detect population structure among low diversity samples compared to organellar genomes, and represent promising methods to determine if there is fine-scale population structure of *N. obtusa* in North America.

Acknowledgements We would like to thank Victoria D. Davis, Alyssa D. Garvey, Sarah E. Glass, Stephen D. Gottschalk, Catharina R. Grubaugh, Amy J. Havens, Samuel C. Lazlo, Damon P. Little, Robert (A) Stewart, Kam M. Truhn, and Sarah (B) Whorley for field, laboratory, and computational assistance. José D. Anadón, Michael J. Hickerson, John D. Wehr, and Richard M. McCourt provided valuable feedback on drafts of the manuscript. Aurélie Boissezon, Jonathan D. Midwood, Hendrik Schubert, and members of the International Research Group on Charophytes provided specimens from across the native range of *Nitellopsis*. This work was supported by the Sarah K. de Coizart Article TENTH Perpetual Charitable Trust, a Northeast Algal Society Student Grant to Support Research, an International Phycological Society Paul (C) Silva Student Grant, a Phycological Society of America Grant-in-Aid of Research, a Northeast Aquatic Plant Management Society Graduate Scholarship, a Wisconsin Department of Natural Resources grant: Desiccation, Freezing and Dispersal of *Nitellopsis obtusa*: AIRD10716, and a City University of New York Doctoral Student Research Grant. This research was further supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. This material is also based upon work supported by the National Science Foundation under Grant Numbers DEB-1020660, DEB-1036466, DEB-1701691, CNS-0521433, and DBI-1348920.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Data accessibility Reads are available at NCBI SRA under BioProject PRJNA698590.

References

- Babraham Bioinformatics: FastQC A Quality Control tool for High Throughput Sequence Data (2018). <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 19 July 2018
- Bharathan S (1987) Bulbils of some charophytes. *Proc Plant Sci* 97(3):257–263
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
- Brainard AS, Schulz KL (2017) Impacts of the cryptic macroalgal invader, *Nitellopsis obtusa*, on macrophyte communities. *Freshw Sci* 36(1):55–62
- Chaisson MJ, Tesler G (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC*
- Cristescu ME (2015) Genetic reconstructions of invasion history. *Mol Ecol* 24(9):2212–2225
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158
- Early R, Bradley B, Dukes J et al (2016) Global threats from invasive alien species in the twenty-first century and national response capacities. *Nat Commun* 7:12485. <https://doi.org/10.1038/ncomms12485>
- Escobar LE, Qiao H, Phelps NBD, Wagner CK, Larkin DJ (2016) Realized niche shift associated with the Eurasian charophyte *Nitellopsis obtusa* becoming invasive in North America. *Sci Rep* 6:29037
- Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2011) What can DNA tell us about biological invasions? *Biol Invasions* 14.2:245–253
- Geis JW, Schumacher GJ, Raynal DJ, Hyduke NP (1981) Distribution of *Nitellopsis obtusa* (Charophyceae, Characeae) in the St Lawrence River: a new record for North America. *Phycologia* 20(2):211–214
- Gray MW, Doolittle WF (1982) Has the endosymbiont hypothesis been proven? *Microbiol Rev* 46(1):1–42
- Hagenblad J, Hülskötter J, Acharya KP et al (2015) Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. *BMC Genet* 16:103. <https://doi.org/10.1186/s12863-015-0242-8>
- Hardion L, Verlaque R, Saltonstall K, Leriche A, Vila B (2014) Origin of the invasive *Arundo donax* (Poaceae): a trans-Asian expedition in herbaria. *Ann Bot* 114(3):455–462. <https://doi.org/10.1093/aob/mcu143>
- Hall JD, Fucikova K, Lo C, Lewis LA, Karol KG (2010) An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie Algologie* 31.4:529–555
- Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLOS Genet* 6(2):e1000834
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11):1403–1405
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94
- Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281
- Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet* 6
- Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, Cronk Q (2012) Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Am J Bot* 99.2:320–329
- Karol KG, Sleith RS (2017) Discovery of the oldest record of *Nitellopsis obtusa* (Charophyceae, Charophyta) in North America. *J Phycol* 53:1106–1108
- Kato (2005) Rediscovery of the Wild-Extinct Species *Nitellopsis obtusa* (Charales) in Lake Kawaguchi, Japan. http://www.tsumura.co.jp/english/kampo/plant/080/080_02.html#p84. Accessed 17 March 2015
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27(5):722–736
- Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res*. <https://doi.org/10.1101/gr.092759.109>
- Larkin DJ, Monfils AK, Boissezon A, Sleith RS, Skawinski PM, Welling CH, Cahill BC, Karol KG (2018) Biology, ecology, and management of starry stonewort (*Nitellopsis obtusa*; Characeae): a red-listed Eurasian green alga invasive in North America. *Aquat Bot* 148:15–24
- Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc Natl Acad Sci* 104(10):3883–3888
- Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet* 21(12):655–663. <https://doi.org/10.1016/j.tig.2005.09.004>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25.14:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41(W1):W575–W581

- Maszewski J, Kołodziejczyk P (1991) Cell cycle duration in antheridial filaments of *Chara* spp. (Characeae) with different genome size and heterochromatin content. *Pl Syst Evol* 175:23–38. <https://doi.org/10.1007/BF00942143>
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 66(2):526–538
- Michel A, Arias RS, Scheffler BE, Duke SO, Netherland M, Dayan FE (2004) Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). *Mol Ecol* 13.10:3229–3237
- Moody ML, Les DH, Ditomaso JM (2008) The role of plant systematics in invasive aquatic plant management. *J Aquat Plant Manag* 46:7
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS ONE* 7(5):e36869
- Nishiyama T, Sakayama H, de Vries J et al (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* 174.2:448–464.e24
- Pérez W, Hall JD, McCourt RM, Karol KG (2014) Phylogeny of North American Tolypella (Charophyceae, Charophyta) based on plastid DNA sequences with a description of *Tolypella ramosissima* sp. nov. *J Phycol* 50:776–789. <https://doi.org/10.1111/jpy.12219>
- Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ* 52.3:273–288
- Proctor VW (1962) Viability of *Chara* oospores taken from migratory water birds. *Ecology* 43(3):528–529
- Proctor VW (1968) Long-distance dispersal of seeds by retention in digestive tract of birds. *Science* 160(3825):321–322
- R Core Team (2015) R: a language and environment for statistical computing
- Riis T, Lambertini C, Olesen B, Clayton JS, Brix H, Sorrell BK (2010) Invasion strategies in clonal aquatic plants: are phenotypic differences caused by phenotypic plasticity or local adaptation? *Ann Bot* 106(5):813–822. <https://doi.org/10.1093/aob/mcq176>
- Sagan L (1967) On the origin of mitosing cells. *J Theor Biol* 14.3:255–274
- Saltonstall K (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *PNAS* 99(4):2445–2449. <https://doi.org/10.1073/pnas.032477999>
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nat Rev Genet* 14:807–820. <https://doi.org/10.1038/nrg3522>
- Sardina JA (2017) Intra- and inter-population comparison of *Chara brittonii* Allen ex Rob. organellar genomes. Presented at the 56th northeast algal symposium
- Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, Berthouly-Salazar C, Vigouroux Y (2016) Intra-individual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it? *Mol Ecol Resour* 16.2:434–445
- Schloesser DW, Hudson PL, Nichols SJ (1986) Distribution and habitat of *Nitellopsis obtusa* (Characeae) in the Laurentian Great Lakes. *Hydrobiologia* 133(1):91–96
- Sleith RS, Havens AJ, Stewart RA, Karol KG (2015) Distribution of *Nitellopsis obtusa* (Characeae) in New York, U.S.A. *Brittonia* 67(2):166–172
- Sleith RS, Wehr JD, Karol KG (2018) Untangling climate and water chemistry to predict changes in freshwater macrophyte distributions. *Ecol Evol* 8(5):2802–2811
- Soulié-Märsche I, Benammi M, Gemayel P (2002) Biogeography of living and fossil *Nitellopsis* (Charophyta) in relationship to new finds from Morocco. *J Biogeogr* 29(12):1703–1711
- Staats M, Erkens RHJ, van de Vossen B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS ONE* 8:e69189
- Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am J Bot* 99:349–364. <https://doi.org/10.3732/ajb.1100335>
- Tong W, Kim T-S, Park Y-J (2016) Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. *Rice* 9
- Turner M (2003) The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell Online* 15(8):1888–1903
- Turner M (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol* 23.6:1324–1338
- Vandepitte K, de Meyer T, Helsen K, van Acker K, Roldán-Ruiz I, Mergeay J, Honnay O (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. *Mol Ecol* 23(9):2157–2164
- Ward SM, Gaskin JF, Wilson LM (2008) Ecological genetics of plant invasion: what do we know? *Invasive Plant Sci Manag* 1:98–109
- Wiedenmann J, Baumstark A, Pillen T et al (2001) DNA fingerprints of *Caulerpa taxifolia* provide evidence for the introduction of an aquarium strain into the Mediterranean Sea and its close relationship to an Australian population. *Marine Biol* 138:229–234 (2001). <https://doi.org/10.1007/s002270000456>
- Wood RD, Imahori K (1965) Monograph of the characeae, vols 1–2, vol 1. Cramer, Weinheim
- Zeng CX, Hollingsworth PM, Yang J et al (2018) Genome skimming herbarium specimens for DNA barcoding and phylogenomics. *Plant Methods* 14:43. <https://doi.org/10.1186/s13007-018-0300-0>
- Zimin AV, Salzberg SL (2020) The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLOS Comput Biol* 16(6):1–8

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