

[ncbi.nlm.nih.gov/nuccore/KT956910](https://www.ncbi.nlm.nih.gov/nuccore/KT956910) KT959344
<https://www.ncbi.nlm.nih.gov/nuccore/KT959344>
KT960977 <https://www.ncbi.nlm.nih.gov/nuccore/KT960977> KX711882 <https://www.ncbi.nlm.nih.gov/nuccore/KX711882> KX711883 <https://www.ncbi.nlm.nih.gov/nuccore/KX711883> KX711884
<https://www.ncbi.nlm.nih.gov/nuccore/KX711884>
KX711885 <https://www.ncbi.nlm.nih.gov/nuccore/KX711885>

Funding: The inaugural Urban Barcode Project was funded by the Alfred P. Sloan Foundation (<https://sloan.org/>). DNA Subway was developed for the CyVerse (former iPlant) Collaborative, funded by a grant from the National Science Foundation Plant Cyberinfrastructure Program (#DBI-0735191, DBI-1265383, <https://www.nsf.gov/>). This study took place in collaboration with NYCDOE that funded the Urban Barcoding Project Institutes at the Genovesi Environmental Study Center (<http://schools.nyc.gov/default.htm>). Funding was received by David Andrew Micklos, CSHL, DNA Learning Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

12 novel DNA barcodes on GenBank. Students also identified 7 ant species and demonstrated the potential of DNA barcoding for identification of this especially diverse group when coupled with traditional taxonomy using morphology. Here we outline how DNA barcoding allows citizen scientists to make preliminary taxonomic identifications and contribute to modern biodiversity research.

Introduction

DNA barcoding, or sequence-based specimen identification, was developed by Paul Hebert in 2003 to identify a broad range of taxa by sequencing a standardized short DNA fragment, the “DNA barcode” [1,2]. This technique has enabled the construction of the International Barcode of Life (www.iBOL.org) project, which has so far identified unique genetic barcodes for over 592,000 of the estimated 1 to 6 billion species on Earth [3,4]. For animals, the standard region to create a DNA barcode is the variable 5’ half of mitochondrial gene cytochrome *c* oxidase 1 (*COI*) [1,5]. A region of the chloroplast gene ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) is used for barcoding plants [6], and fungi are barcoded using the nuclear internal transcribed spacer (ITS) region [7,8].

DNA barcoding has many practical applications including identification of fraud in consumer products, furthering the study of taxonomy, and monitoring and accounting for Earth’s biodiversity. In this time of major biodiversity loss, there is a necessity to identify and catalogue organisms to establish the baseline biodiversity. This helps to monitor biodiversity and potentially counteract the disappearance of species, and this can be rapidly accomplished through DNA barcoding [9,10]. An organism can often be sampled non-invasively, with minimal damage to any voucher sample. DNA barcoding can also be used to discover marketplace replacements and identify products taken from endangered species [11–18]. For example, DNA barcodes were previously used to find that five of 23 samples of caviar purchased in New York City (NYC) were mislabeled, including three from threatened sturgeon species [19]. Genetic marketplace monitoring has continued to be a promising tool for detecting fraudulent black caviar present in the NYC area [20].

DNA barcoding is not just a powerful research tool, but the technique also allows students to complete authentic research projects in a short time period. The educational potential of DNA barcoding appeared when students from Trinity High School that documented mislabeling of seafood items purchased in NYC were featured in a media article [21]. Expanding on these efforts, the DNA Learning Center (DNALC) designed the *Urban Barcode Project* in 2011 to demonstrate that large numbers of high school students can work with their classroom teachers to complete independent research projects centered on DNA barcoding. The streamlined DNA barcoding curriculum (published online: www.dnabarcoding101.org) provides the infrastructure needed for teachers to replicate the DNA barcoding laboratory in schools or during workshops. The curriculum is largely based on the student-centered and discovery-based framework provided by the DNALC (see Fig 1).

DNA barcoding programs offer the opportunity to bring students into natural areas, and students who participated in these workshops all spent time collecting samples as part of both a bioblitz and a targeted ant survey in Marine Park (Fig 2). Managed by the NYC Department of Parks and Recreation, Marine Park is the largest public park in Brooklyn and is protected as a Forever Wild Preserve. With more than 530 acres of grassland and salt marsh, and supplied with freshwater from Gerritsen Creek, Marine Park’s unique land (both natural and man-made) and water features provide an opportunity to study the effects of urbanization on wildlife in Brooklyn’s green areas [22].

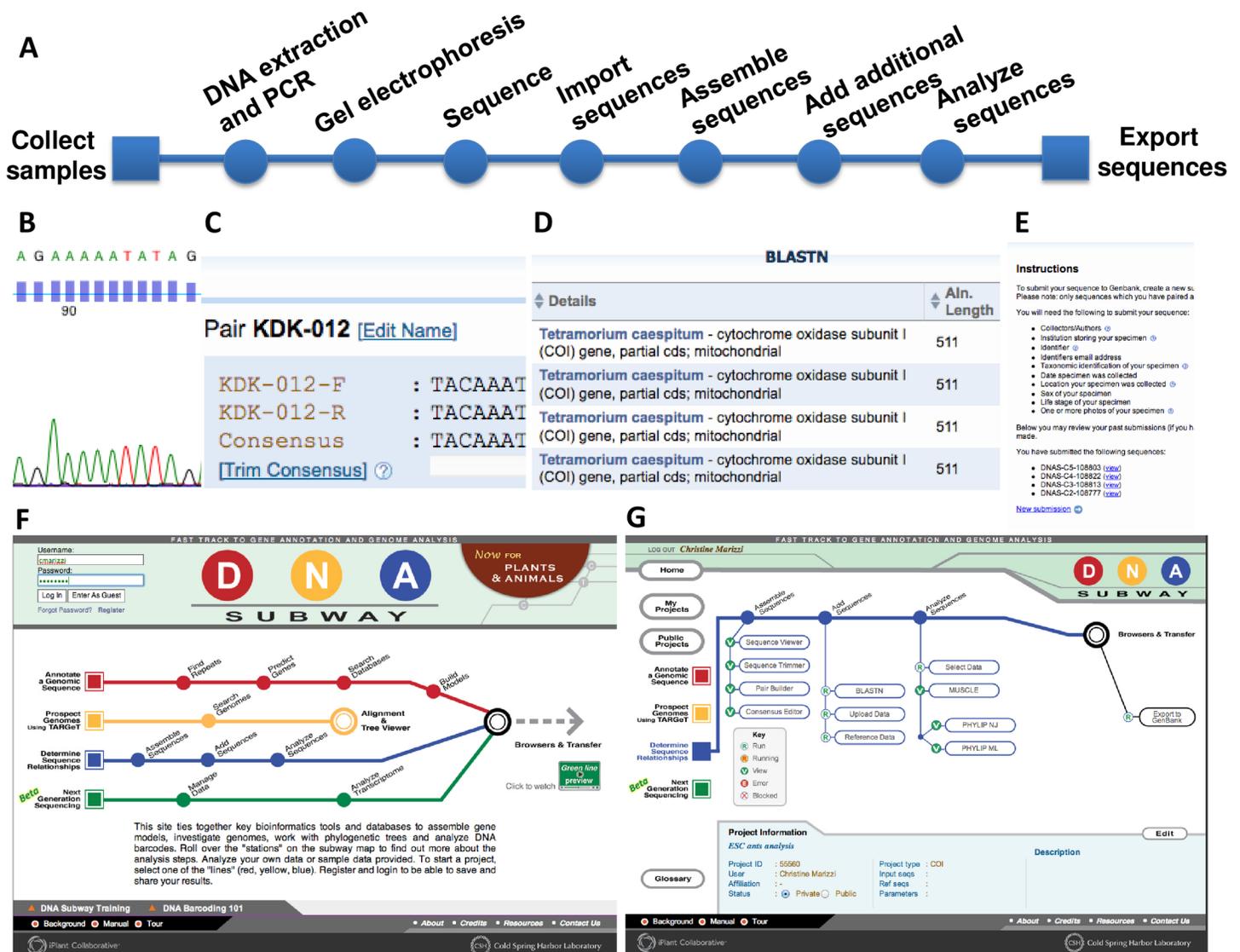


Fig 1. The DNA barcoding workflow. Biochemical protocols at the website *DNA Barcoding 101* (www.dnabarcoding101.org) and bioinformatics tools at *DNA Subway* (www.dnasubway.org) support all parts of the workflow. (A) Participants collected samples and extracted genomic DNA, generated DNA barcodes using PCR, verified the correct size by gel electrophoresis and sent amplicons for automated sequencing. Sequence data was uploaded to the internet-based DNA barcoding bioinformatics pipeline *DNA Subway*, and participants assembled contigs, compared them to additional sequences and analyzed sequence data for potential export to GenBank. (B) Trace file of a DNA sequence from a local ant sample. (C) DNA reads are paired and manually edited to create consensus sequence. (D) Top BLAST hits for an ant consensus sequence identifying it as *Tetramorium caespitum*. (E) *DNA Subway* integrated export function for novel DNA barcodes to GenBank. (F) User-friendly *DNA Subway* interface. (G) *DNA Subway*'s "Blue Line" for DNA barcoding and phylogenetics allows for sequence editing, performing a BLAST search, and phylogenetic tree building on an intuitive, open-source platform.

<https://doi.org/10.1371/journal.pone.0199015.g001>

While DNA barcoding introduces students to laboratory methods that will serve as a foundation to future research in academic labs, it also helps students develop an awareness and concern for biodiversity loss in NYC. Previous studies have suggested that this can be accomplished by inspiring a connection with nature [23,24] and by allowing students to develop a relationship with the species that surround them [25]. DNA barcoding in an educational setting achieves this because students are the lead investigators of their projects [26–29], and in doing so, become citizen scientists who are cataloging species biodiversity in NYC. This is an



Fig 2. Sampling sites for this study. Samples were obtained at Marine Park (Brooklyn, New York, USA) in two localities: at a restored, more natural habitat (# 1 above) and an unrestored, disturbed habitat (#2 above). The figure was prepared based on data available from the U.S. Geological Survey Department of the Interior (USGS) public domain (USGS-viewer.nationalmap.gov) and edited by Chun-hua Yang, DNALC.

<https://doi.org/10.1371/journal.pone.0199015.g002>

especially valuable lesson for urban students, like those in NYC, who are often disconnected from wildlife and the natural environment [30,31].

In partnership with the NYC Department of Education (NYCDOE) Genovesi Environmental Study Center (GESC), the DNALC's barcoding curriculum was implemented during an immersive 10-day summer institute for high school teachers and their students in 2014 and 2015. Nine teachers from nine different NYC public schools, along with their team of students (3–4 students/teacher; 39 students total), were selected to participate in the institute held at GESC in Bergen Beach, Brooklyn. This institute provided teachers with a unique opportunity to attend professional development with their students. Together, they worked through the

labs, data collection, and data analysis. This allowed teacher and student teams to start their research during the summer and continue it during the school year.

Teachers and their student teams began their DNA barcoding research project by exploring the local environment at Marine Park and the importance of the salt marsh ecosystem. The students completed research projects that fell into two categories: 1) a Marine Park bioblitz, where citizen scientist students collected all invertebrate and plant samples they found on designated sampling dates, and 2) the ants of Marine Park, where the focus was on collecting and cataloging ant species found in the park. For the ant project, the students worked with myrmecologist Dr. Sean McKenzie (Rockefeller University) to morphologically identify the samples, and then compared the morphological results to those found with DNA barcoding. Here we report all the results of these pilot projects where DNA barcoding was used as a foundation for introducing urban high school students to modern biodiversity research.

Materials and methods

Sample collection

A sampling permit was first obtained from the NYC Department of Parks and Recreation. Guided by Brooklyn Urban Park Rangers, as well as DNALC and GESC staff, students first visited the site and were taught to systematically document information as samples were collected. Sampling occurred July 14th, July 15th, July 25th, July 28th, July 29th and August 4th of 2014 and August 4th, August 5th and August 10th of 2015 (9 days total) for two hours each day. At the time of collection, all individual organisms were photographed, global positioning system (GPS) coordinates and locality information were recorded, and preliminary species identifications were made using field guides [32,33]. Students collected no more than five individuals per species and locality to prevent any detrimental impacts on the environment, species compositions, and population sizes. Ants were collected using Keebler® Sandies® Pecan Shortbread as bait, following School of Ants guidelines (<http://schoolofants.org>).

To generate a long-term record, sample metadata were entered into the DNALC Sample Database (<https://sampledb.dnalc.org/barcoding>), which functions as an online laboratory manual that can be accessed by the whole team to update and modify data (e.g., for data submissions to GenBank). Students were trained in preserving specimen vouchers for long-term storage. Vouchers were either frozen at -20°C directly (plants) or preserved in 100% ethanol (EtOH) [34] and subsequently frozen at -20°C (invertebrates). All samples were frozen within four hours of collection to minimize DNA degradation.

DNA extraction, polymerase chain reaction (PCR), and sequencing

DNA extraction, PCR, and sequencing protocols are published under dnabarcoding101.org for teachers and professors to use worldwide. Here, we summarize these methods. For DNA extraction, a small portion of each sample (animal or plant) was placed in a 1.5 mL centrifuge tube, and ground using a plastic pestle in 300 µL lysis solution (6 M guanidine hydrochloride (Sigma-Aldrich, USA)). After incubating the sample for ten minutes at 65°C, samples were centrifuged for one minute at 6000 rcf and 150 µL of the supernatant was transferred to a fresh tube. Three µL silica resin solution (25 g of silicon dioxide (Sigma-Aldrich, USA, m.w. = 60.08) in 50 mL total volume of distilled water) was added to bind genomic DNA, followed by a five minutes incubation at 57°C. The samples were centrifuged again for one minute at 6000 rcf, the supernatant discarded and the silica pellet washed twice with 500 µL ice-cold wash buffer (20mM Tris-HCl pH 7.4, 50mM NaCl, 1mM ethylenediaminetetraacetic acid (EDTA), 50% EtOH). Finally the supernatant was discarded and DNA eluted in 100 µL distilled water by incubating samples for five minutes at 57°C.

The *COI*, *rbcL* and ITS barcode regions were amplified via PCR using illustra PuReTaq Ready-To-Go PCR Beads (GE Healthcare Life Sciences, USA). All PCR reactions were performed in a total volume of 25 μ L, including 2.5 μ L of template DNA, 1 μ L of each primer (0.26 picomoles/ μ L), 2.5 units of Taq DNA polymerase, 10mM Tris-HCl (pH 9.0), 50mM KCl, 1.5mM MgCl₂, 200 μ M of each dNTP, 13.8% sucrose, 0.0081% Cresol Red (Sigma-Aldrich, USA) and 22.5 μ L of distilled water.

Invertebrate samples were amplified with DNA primers of the mitochondrial *COI*, LCO1490 (5' -TGTA AACGACGGCCAGTGGTCAACAAATCATAAAGATATTGG-3'), and HCO2198 (5' -CAGGAAACAGCTATGACTAACTTCAGGGTGACCAAAAAATCA-3') [5]. The PCR conditions for the *COI* gene using LCO1490 and HCO2198 primers were as follows: initial denaturation at 94°C for one minute, 30 cycles of denaturation at 95°C for 30 seconds, annealing at 50°C for 30 seconds and extension at 72°C for 45 seconds. Plant samples were amplified with DNA *rbcL* gene primers *rbcLaf* (5' -TGTA AACGACGGCCAGTATGTCAC-CACAAACAGAGACTAAAGC-3') and *rbcLarev* (5' CAGGAAACAGCTATGACGTAAAAAT-CAAGTCCACCRCG-3') [6,8]. The PCR conditions for the *rbcL* gene were as follows: initial denaturation at 94°C for one minute, 35 cycles of denaturation at 94°C for 15 seconds, annealing at 54°C for 15 seconds, and extension at 72°C for 30 seconds. Fungi samples were amplified with DNA primers ITS1F (5' -TGTA AACGACGGCCAGTCCGTAGGTGAACCTGCGG-3') and ITS4 (5' -CAGGAAACAGCTATGACTCCTCCGCTTATTGATATGC-3') targeting the nuclear ITS region [7,35,36]. The PCR conditions were as follows for ITS1F and ITS4: initial denaturation at 94°C for one minute, followed by 35 cycles of denaturation at 94°C for one minute, annealing at 55°C for one minute and extension at 72°C for two minutes. Successful PCR products were confirmed by 2% agarose gel electrophoresis and stained with GelRed™ (Biotium, USA). To streamline sequencing the primers used in this experiment, we incorporated a universal M13 primer sequence at the 5' end (M13F(-21): TGTA AACGACGGCCAGT and M13R(-27): CAGGAAACAGCTATGAC) [37]. Sequencing was performed with an Applied Biosystems ABI 3730xl DNA Analyzer (ThermoFisher Scientific, USA) automatic sequencer at Genewiz (South Plainfield, NJ, USA).

Data analysis

Sequences were analyzed using the “Blue Line” feature on *DNA Subway*, which supports phylogenetic and DNA barcode analyses [38,39]. A series of steps were used to ensure that only high quality data were used in subsequent analyses. First, DNA trace files were analyzed by Phred software [40,41], which calls a nucleotide (A, G, C, or T) for each peak. Each nucleotide is also assigned a “Phred” quality score that corresponds to a logarithmic error probability that the nucleotide call is wrong, or conversely, to the accuracy of the call (see <http://www.dnabarcoding101.org/lab/bioinformatics.html>). These quality scores for individual base calls are represented by blue bars on the *DNA Subway* sequence viewer. Nucleotides were only called if the Phred scores met or exceeded the quality cutoff (Phred score of 20, or greater than 99% accuracy, within a sliding window of 18, indicated with a blue horizontal line placed across the blue bars). Sequences with average Phred scores below 20 were flagged as “low quality”. Although this process is automated on *DNA Subway*, students were also taught to manually inspect each sequence to ensure that they agreed with the calls made by *DNA Subway*.

Next, sequences were trimmed in bulk by the default trimming function that removed “Ns” on the 5' and 3' prime ends (leaving in a sliding window of 12 nucleotides only one “N”) and automatically filtered out low quality sequences. Consensus sequences were built in *DNA Subway* using Merger—from EMBOSS (<http://embooss.sourceforge.net/apps/release/6.6/embooss/apps/merger.html>), and then all sequences were further manually edited to ensure primer removal.

After all of the above cleaning steps were completed, sequences were identified with a BLAST search on *DNA Subway* using default parameters, and then results were confirmed using a BLAST search directly on the GenBank website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The top BLAST result was determined using the highest percent identity score. In the event of a tie between two identity scores, the lowest e-value was used as a deciding-factor. If there were identical e-values and identity scores for multiple species, individuals were only identified to the class, family or genus level, to avoid any potential for incorrect taxonomic identifications.

Each sample was defined as “identified” if the highest percent identity score was $\geq 95\%$, and as a “potential new barcode” if the percent identity score was $<95\%$. For sequences with $\geq 95\%$, after identification to genus and species-level on *DNA Subway*, higher-level plant taxonomic ranks (order, family and class) were determined from Tropicos (<http://tropicos.org>), and higher-level animal taxonomic ranks (order, class, and phylum) were determined from the Integrated Taxonomic Information System (<http://www.itis.gov>). Percentages of identified taxonomic groups were displayed using pie charts. For sequences of $< 95\%$ identity, each was carefully assessed to make sure that both reads were of high quality, with substantial overlap in the consensus sequence, and proper reading frames. If these characteristics were met, the sequences were submitted directly to GenBank as “novel barcodes” through the *DNA Subway* export function.

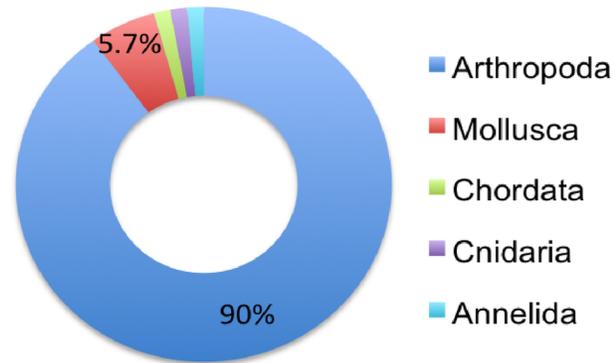
We also assessed the status of the identified genera and species as either native, invasive, or introduced to Marine Park. We did this using several guides by the New York State Department of Environmental Conservation (<http://www.dec.ny.gov/animals/99141.html>), the New York State Department of Agriculture and Markets [42], the Brooklyn Botanic Garden [43] and the USDA Forest Service [44]. An invasive species is a non-native species whose introduction does or is likely to cause economic or environmental harm or harm to human, animal, or plant health. Introduced or exotic species are also not native to the region, but differ from invasive species in that they do not cause significant damage to the ecosystem by outcompeting native species. (<https://www.invasivespeciesinfo.gov/docs/council/isacdef.pdf>). Additionally, morphological identification of selected ants collected in Marine Park was carried out by myrmecologist Sean McKenzie (Rockefeller University) using standard taxonomic keys [45,46]. Jeffry Petracca (Long Island Aquarium) morphologically identified any samples published as novel DNA barcodes.

Results

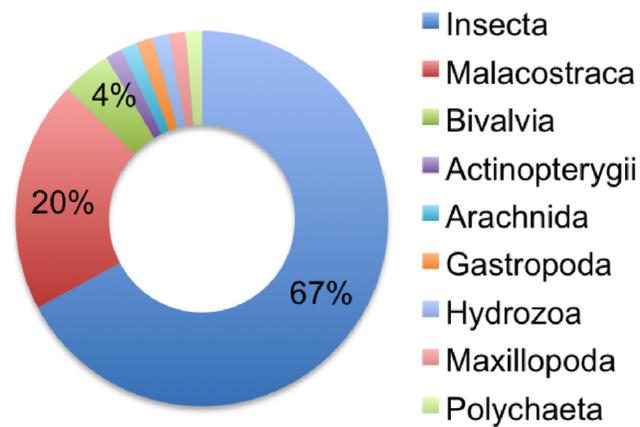
Assessment of species collected in Marine Park

Over the two summers included in this study (2014 and 2015), student teams collected and identified 187 samples (1 fungus, 70 animals, and 116 plants; a full list of the samples, along with identification and sequence information can be found in [S1 Table](#)). Overall, we identified 87 genera and 98 unique haplotypes from the 187 samples collected in Marine Park ([S2 Table](#)). The sole fungus sample was identified as *Simocybe serrulata*, and an overview of the animal and plant samples are provided below. For the animal samples (see [Fig 3](#)), only one chordate was sequenced (sample KQB-011)—the fish *Fundulus heteroclitus* (mummichog, a small killifish species), provided by a park ranger—because collection was restricted to invertebrate samples. Ninety percent of the remaining samples were invertebrates belonging to the phylum Arthropoda ($n = 63$; 90% of collected animals). Of those 63 samples, most were from the Insecta class ($n = 47$; 67.1% of collected animals) in the Hymenoptera ($n = 30$; 42.8% of collected animals) and Diptera ($n = 9$; 12.8% of collected animals) orders ([Fig 3](#)). The second most abundant samples in the Arthropoda phylum were from the Malacostraca class ($n = 14$; 19.7% of all collected animals), of which ten (14.1% of total animals) were in the Decapoda

Animal Phyla



Animal Classes



Animal Orders

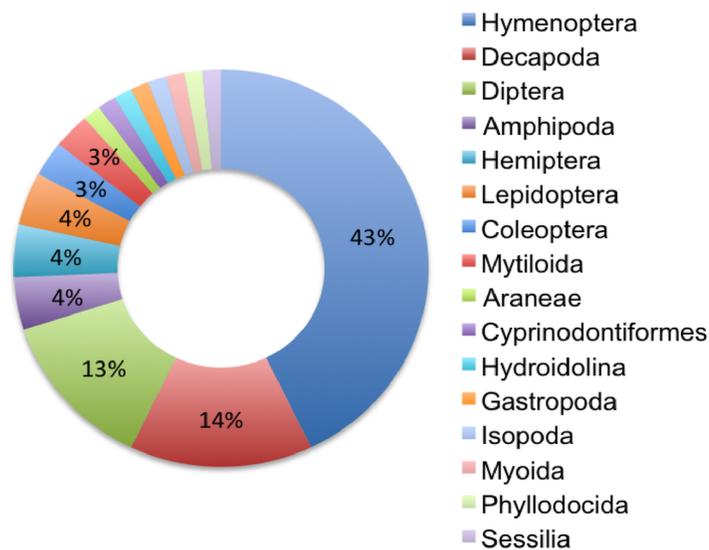


Fig 3. Overview of invertebrates identified by this project, organized by taxonomic level. The vast majority of the animal samples belonged to the phylum Arthropoda (n = 63; 90% of collected animals). Of those 63 samples, most were from the Insecta class (n = 47; 67.1% of collected animals) in the Hymenoptera (n = 30; 42.8% of collected animals) and Diptera (n = 9; 12.8% of collected animals) orders. Taxonomic groups in legends are organized from largest to smallest.

<https://doi.org/10.1371/journal.pone.0199015.g003>

order, three (4.2% of total animals) were Amphipods, and one (1.4% of total animals) was an Isopod. The remaining 8.6% of the total animals were from the phyla Mollusca, Cnidaria and Annelida.

Of the 116 collected and identified plants, the majority were in the Asterales (n = 32; 27.6% of total plants) and Poales (n = 28; 24.1% of total plants) orders, with the highest number of species within each order falling in the Asteraceae (n = 32; 27.6% of total plants) and Poaceae families (n = 28; 24.1% of total plants; see Fig 4). We also found 24 plants in the Rosales and Carophyllales orders (each having n = 12; 10.3% of total plants).

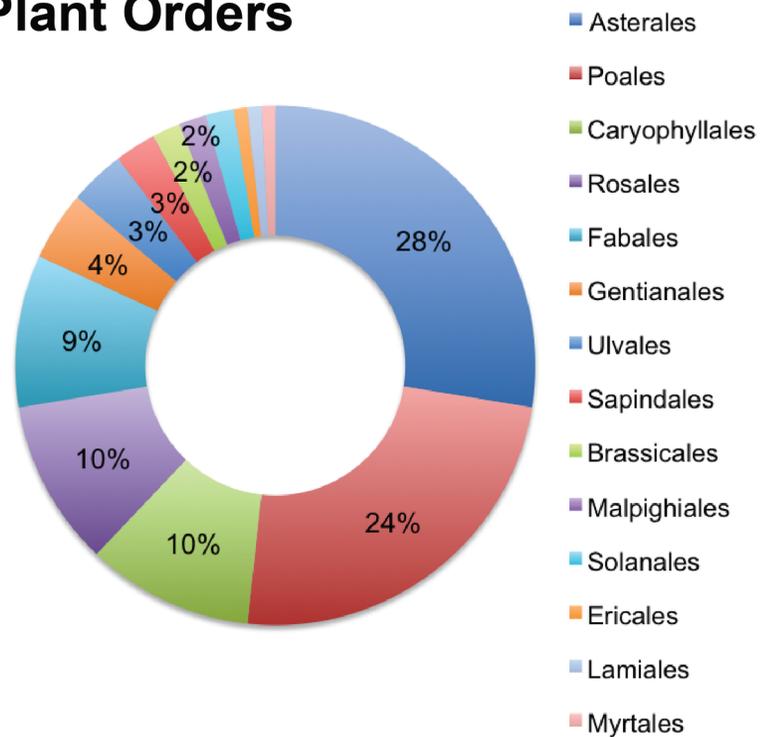
The majority of the sequences analyzed (168/187) received high similarity scores (> 95%) on GenBank. From the remaining 19 samples with low similarity scores (<95%), 12 new barcodes were published as GenBank entries with participants as authors (see Tables 1 and S1 for full sample information), and each of these submissions were verified by a taxonomist. In most instances, we were conservative in our identifications and only identified samples to higher-order ranks (e.g. families or classes). Genus and species-level identification was only made when, in addition to no ties between identify and e-value scores on GenBank, an expert taxonomist confirmed that identity. The seven remaining samples with low similarity scores were not published as novel barcodes because the sequence quality was low and these samples were excluded from these analyses.

We also assessed whether the species we collected in Marine Park were native, introduced, or invasive based on our preliminary taxonomic identifications. From our collected animals, we found 19 native, five introduced, and one invasive species (the ailanthus webworm *Atteva aurea*—see S2 Table); the rest could not be classified based on the guides we used or the taxonomic level (e.g. some only identified to genus-level). Examples of native animal species to the U.S. Northeast region include the ground spider *Gnaphosa* sp., and the soft-shell clam *Mya arenaria*. Introduced animal species included the coccinellid beetle *Harmonia axyridis* and the amphipod *Grandidierella japonica*. From our collected plants, we found 20 native, 14 introduced, and four invasive species; the remainder could not be classified for the same reasons given above for the animal samples. Examples of abundant invasive plant species include the China-sumac *Ailanthus altissima*, the invasive genotype of the common reed *Phragmites australis* and the Autumn olive *Elaeagnus umbellata*. All other species that could not be classified into these categories are marked as unknown in S2 Table.

The utility of barcoding genes

The ability of the traditional barcoding genes (*COI*, *rbcl*, and *ITS*) to identify samples varied across taxonomic groups when we were making our preliminary taxonomic assessments using GenBank. The one fungus sample (GQB-045) was identified to the species-level with *ITS*. In arthropods, 43/63 (68.2%) of samples could be identified to the species-level (see S1 Table). All four mollusk haplotypes, the one annelid, and the one chordate haplotype were sufficient to identify those samples to the species-level. The one cnidarian haplotype could only identify that sample to the genus level. Across all 70 animal samples, 49 samples (or 70%) were identified to the species level using preliminary taxonomic identifications based on GenBank searches.

Plant Orders



Plant Families

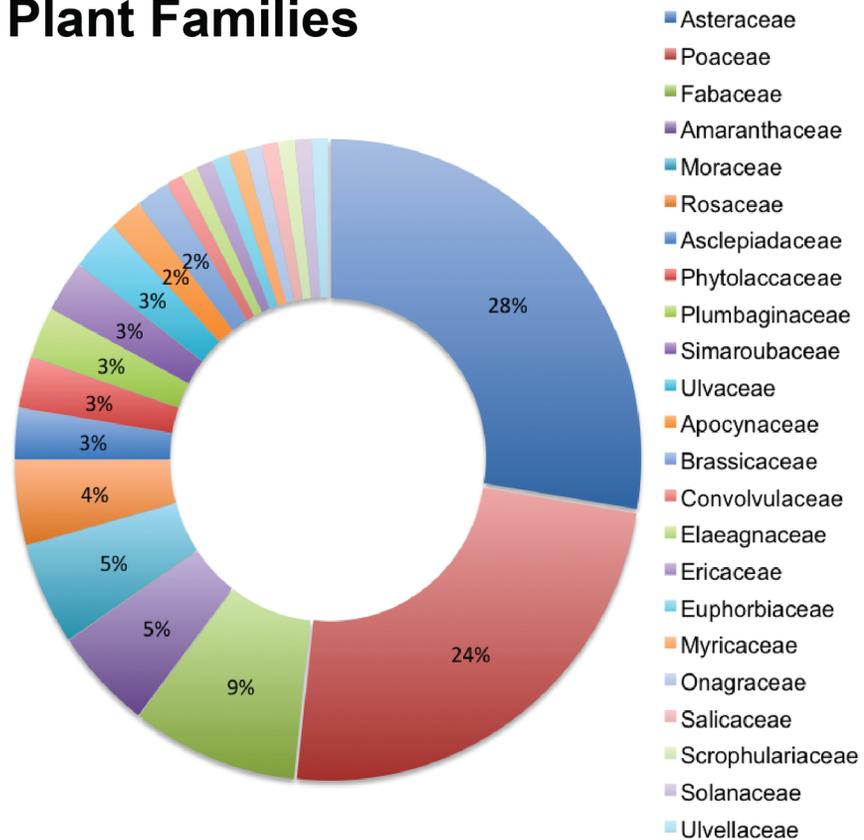


Fig 4. Overview of plants identified by this project, organized by taxonomic level. Of the 116 collected and identified plants the majority were in the Asterales (n = 32; 27.6% of total plants) and Poales (n = 28; 24.1% of total plants) orders, with the highest number of species within each order falling in the Asteraceae (n = 32; 27.6% of total plants) and Poaceae families (n = 28; 24.1%). Taxonomic groups in legends are organized from largest to smallest.

<https://doi.org/10.1371/journal.pone.0199015.g004>

Plant identification with *rbcL* was generally less successful than animal identification with *COI*. The largest numbers of samples were found in the Asterales and Poales orders. In Asterales, 14/32 of samples (43.8%) could be identified to the species-level, while in Poales 11/28 (39.3%) of samples could be identified to the species-level (see [S1 Table](#)). In all instances, sequence lengths were standard for DNA barcoding (>500 bp) with few nucleotide ambiguities. Across all 116 plant samples, 49 samples (or 42.2%) were identified to the species level using preliminary taxonomic identifications—see [S1 Table](#).

Barcoding ants in Marine Park

We collected 29 ant samples from seven ant species (all in the Formicidae family), and four of these were submitted as novel barcodes from the two ant genera *Nylanderia* and *Monomorium* ([Table 1](#)). We were interested in whether the DNA barcodes yielded identifications that were congruent with morphological identifications, and examined this using a subset of 20 samples ([Table 2](#)). The DNA barcode and morphological identifications were congruent in each case, but in some cases either barcoding or traditional taxonomy returned species-level identifications when the other method only identified the sample to the genus-level.

Each sample was identified both by performing a BLAST search on GenBank and then verified by a taxonomist. KHN samples: Barcodes generated by students from Forest Hills High School, Queens. KDK samples: Barcodes generated by students from High School for International Studies, Staten Island. GQB samples: Barcodes generated by students from Frank McCourt High School, Manhattan. KDW sample: Barcode generated by Christine Marizzi, DNALC.

In all cases, amateur high school taxonomists were able to use DNA barcoding to identify the same genus as the morphologist. In nine instances, DNA barcodes were unable to definitively identify the ant to the species-level.

Table 1. Overview of the twelve novel DNA barcodes generated by the participants of the Urban Barcoding Project Institute at GESC in the summers of 2014 and 2015.

Sample name	Closest matching taxon name	Taxonomic level	Gene	GenBank accession code
GQB-003	Ephydriidae	Family	<i>COI</i>	KU682771
GQB-004	Carabidae	Family	<i>COI</i>	KU682774
GQB-007	<i>Gnaphosa</i> sp.	Genus	<i>COI</i>	KU682773
GQB-064	Bombyliidae	Family	<i>COI</i>	KU682772
GQB-070	Leptothecata	Order	<i>COI</i>	KU682775
KDK-019	<i>Spartina</i> sp.	Genus	<i>rbcL</i>	KT956910
KDK-021	Ulvellaceae	Family	<i>rbcL</i>	KT959344
KDW-016	<i>Mya arenaria</i>	Species	<i>COI</i>	KT960977
KHN-010	<i>Monomorium</i> sp.	Genus	<i>COI</i>	KX711882
KHN-011	<i>Monomorium</i> sp.	Genus	<i>COI</i>	KX711883
KHN-012	<i>Nylanderia</i> sp.	Genus	<i>COI</i>	KX711884
KHN-013	<i>Monomorium</i> sp.	Genus	<i>COI</i>	KX711885

<https://doi.org/10.1371/journal.pone.0199015.t001>

Table 2. List of ant samples, and their corresponding morphological and DNA barcoding identifications.

Sample name	Morphological identification	DNA barcoding identification
KFJ-019	<i>Tetramorium caespitum</i>	<i>Tetramorium caespitum</i>
KFJ-020	<i>Nylanderia cf. flavipes</i>	<i>Nylanderia</i> sp.
KFJ-021	<i>Tapinoma sessile</i>	<i>Tapinoma sessile</i>
KFQ-015	<i>Tetramorium caespitum</i>	<i>Tetramorium caespitum</i>
KFQ-017	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.
KFQ-018	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.
KFQ-019	<i>Tapinoma sessile</i>	<i>Tapinoma sessile</i>
KHN-003	<i>Tetramorium caespitum</i>	<i>Tetramorium caespitum</i>
KHN-004_1	<i>Tapinoma sessile</i>	<i>Tapinoma sessile</i>
KHN-004_2	<i>Nylanderia cf. parvula</i>	<i>Nylanderia cf. parvula</i>
KHN-006_1	<i>Nylanderia cf. parvula</i>	<i>Nylanderia cf. parvula</i>
KHN-006_2	<i>Tetramorium caespitum</i>	<i>Tetramorium caespitum</i>
KHN-007	<i>Tetramorium caespitum</i>	<i>Tetramorium caespitum</i>
KHN-008	<i>Tetramorium caespitum</i>	<i>Tetramorium</i> sp.
KHN-009	<i>Aphaenogaster</i> sp.	<i>Aphaenogaster rudis</i>
KHN-010	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.
KHN-011	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.
KHN-012	<i>Nylanderia cf. flavipes</i>	<i>Nylanderia</i> sp.
KHN-013	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.
KHN-014	<i>Monomorium cf. viride</i>	<i>Monomorium</i> sp.

<https://doi.org/10.1371/journal.pone.0199015.t002>

Discussion and outlook

Mirroring prior DNA barcoding student projects (unpublished data, student project descriptions and talks can be found at dnabarcoding101.org), the majority of novel barcodes never before published on GenBank were from invertebrate samples (10/12 novel barcodes, or 83%, in our study; [Table 1](#)). The most abundant invertebrates collected were insects, which is not surprising since insects have been predicted and found to be among the most abundant and speciose animal groups [46]. We will continue exploring this result, perhaps with curated campaigns looking for specific invertebrates, such as nematodes, which are also known to be very speciose and abundant [47,48]. In our study, the most abundant plants were in the Asteraceae and Poaceae families. Similar to what we found, Asteraceae is the most species rich and largest family of flowering plants [49], with Fabaceae the second most speciose family [50] and Poaceae the third most speciose family in North America (<http://bonap.net/tdc>). Our finding an abundance of Poaceae species in Marine Park is not surprising since grasses are abundant in NYC parks.

An additional classification we made was to determine whether the species we identified were native, introduced, or if their status was unknown. While this is important information to provide a snapshot of the type of species living in an ecosystem, it is also a way to expand upon a traditional barcoding project for more advanced students. Students at the institute were encouraged to specifically look for introduced or invasive species, and several student teams gave presentations focused on these organisms. One example of a presentation done by a team was about the invasive variety of *Phragmites australis*, or common reed, which was found in high abundance in Marine Park. *Phragmites* species do particularly well in warmer wetland habitats, and some studies have suggested that its abundance may increase due to warmer temperatures that result from global climate change [http://www.nyis.info/index.php?action=invasive_detail&id=42]. Students can use DNA barcoding to help monitor and track the abundance of this invasive species over several years to see if this pattern holds true in Marine Park.

While using one universal marker for species-level identification has advantages because it is rapid, simple, and automatable [51], there are limitations. Previous reports showed that molecular analysis using conserved universal primers in the Folmer *COI* region have the potential to recover putative numt (nuclear mitochondrial pseudogene) sequences in addition to orthologous mtDNA, leading to DNA barcoding ambiguity [8,52,53]. We carefully screened for typical indicators of potential poor-quality data, such as multiple bands on a gel, raw sequence ambiguities, background noise and double peaks. Before publishing novel DNA barcodes to GenBank, sequences were translated to verify the absence of indels and stop codons using *DNA Subway* default settings. We found that *COI* was able to identify 70% of samples across all animal groups to the species-level, and the remainder to the genus-level with the exception of one sample. *COI* performed better than *rbcL*, and only 42.2% of plant samples could be identified to the species-level. The absence of a single plant barcode marker with both high amplification success rates and discrimination power poses a difficulty for rapid species identification [6,54]. When a sample cannot be identified to the species-level using a single barcoding gene, we recommend the use of morphology or additional markers, such as 28S rDNA, to confirm and pinpoint species identities [55,56]. Additionally, because *rbcL* does not often allow for species level identification in plants, many plant taxonomists now advocate also using *maturase-K* (*matk*) and nuclear ribosomal internal transcribed spacer (ITS) for barcoding seed plants [54,57,58]. While integrating the use of two loci will increase the cost of outreach programs, this approach would contribute toward the identification of samples to the species-level.

Another note of caution is that the NCBI database results must be carefully examined because some of the published sequences on GenBank are misidentified. For example, a study found that many of the species on GenBank are misidentified as metazoans, when they are in fact bacteria, because the use of universal primers can anneal to bacterial species in error [59]. For fungi, it has been reported that 27% of fungal ITS sequences on GenBank were inadequately identified and 20% were incorrectly labeled [60]. To ensure that any sequences we submitted were correct, we only submitted novel barcodes with two high quality sequence-reads to GenBank and had the identities confirmed by a taxonomist. We were also very conservative when it came to identifying any lower-level classifications, and instead often used higher-level taxonomic ranks (e.g., families or orders). This careful consideration before submission to GenBank is especially important for student-generated data because students may make errors during the wet laboratory and subsequent bioinformatics searches. This is also why establishing and maintaining accurate libraries that will help identify specimen is not only important for research purposes but also for science education.

We found that small-bodied animals that were most often collected by students, and that also yielded novel barcodes, were ants. This is not surprising, since ants are collected easily using cookies as bait (<http://schoolofants.org>), and students are often not morally opposed to euthanizing an ant species for DNA extraction (as opposed to larger invertebrates, like crabs). Students also often sampled close to noon, a time when the majority of ant species were foraging. The ants we found in highest abundance belonged to the *Monomorium*, *Tetramorium*, and *Nylanderia* genera. While the ant genus *Monomorium* has been described worldwide, about 26 of 400 species are found in North America (www.antwiki.org). *Tetramorium caespitum* is also known as the pavement ant and is an introduced pest in North America that is often found in high abundance, and was also found as very abundant in another urban habitat, NYC street medians [61]. The *Nylanderia* genus includes 110 species and has a nearly cosmopolitan distribution, with species inhabiting a wide array of habitats in almost all geographic regions, except Europe (www.antwiki.org) [62,63]. Both key introduced species *Tetramorium caespitum* and *Nylanderia flavipes* have been known to be present in NYC. Interestingly, the majority (5/7) of our collected ant species (*Tapinoma sessile*, *Nylanderia cf. flavipes*,

Aphaenogaster rudis, *Lasius neoniger* and *Tetramorium* spp.) have been previously reported as being present in NYC parks; while two others (*Monomorium* cf. *viridae* and *Nylanderia* cf. *parvula*) have not been found [61,64].

We were especially interested in comparing the DNA barcode ant species results with the identifications made by an expert ant taxonomist, and the results are shown in Table 2. In all cases, the *COI* gene is sufficient for identifying ant genera, and these identifications were congruent to the identifications made by the morphologist. Interestingly, no *Monomorium* samples could be identified to the species-level using DNA barcoding, and this genus was also difficult for morphological identification. For example, we were uncertain if we had collected *Monomorium viride* or another *Monomorium* species. Overall, this case study showed the utility of DNA barcodes for accurate identifications by amateur scientists at least to the genus-level.

The 39 students who participated in the summer institute were required to create presentations to report their findings to their peers and their instructors. Because science practices such as oral and poster presentation skills are a defining element of course-based research experiences (CUREs) [65] and are encouraged at all levels of STEM (science, technology, engineering, and math) education, we recommend incorporating this in all DNA barcoding programs.

Conclusions

DNA barcoding continues to demonstrate its potential as a powerful tool for students to act as citizen scientists to make a real contribution to ecosystem assessment. Our findings support that when given the tools [66], high school students generated high-quality data and identified challenging taxonomic samples such as ants with the same accuracy as experts. In terms of species identification using GenBank alone, we showed that while *COI* was able to identify the 70% of animal samples to the species-level, only 42.2% of plant samples could be identified to the species-level. We recommend building additional plant DNA barcoding markers such as *matK* or ITS into the design of future biodiversity sampling programs to help pinpoint plant species identifications. We also caution about using GenBank alone to make definite species confirmations, and instead encourage the use of traditional taxonomy combined with DNA barcoding when submitting novel sequences to Genbank. Most collected plants were common flowering plants in the Asteraceae and Poaceae families (51.7%), and most animals were insects from the phylum Arthropoda (90%). Additionally, students identified two ant species that were previously unreported in a NYC public park and published 12 novel barcodes on GenBank, underlining that the established DNA barcoding workflow can be used to teach molecular techniques and bioinformatics in a contextual, situated learning design that is relevant personally and to the scientific community.

Supporting information

S1 Table. Complete list of 187 samples included in this study, along with sample names, taxonomic identification, sequence information, and species status.

(XLSX)

S2 Table. Complete list of taxonomic identifications made from the 187 samples included in this study. In some instances, we could not identify the samples to the genus-level.

(XLSX)

Acknowledgments

The inaugural *Urban Barcode Project* was funded by the Alfred P. Sloan Foundation. *DNA Subway* was developed under CyVerse (formerly iPlant Collaborative), a cyberinfrastructure project

supported by the National Science Foundation (DBI-0735191 and DBI-1265383). This study took place in collaboration with NYCDOE that funded the Urban Barcoding Project Institutes at GESC. We would like to extend our gratitude to Jessica Kaplan (NYCDOE) for her generous support hosting the program and for facilitating our collaboration. We would also like to thank Karoline Lake who greatly helped with finalizing all sequence analyses and Jeffry Petracca (Long Island Aquarium) who helped to morphologically identify samples. We would like to thank Novem Auyeung (City of New York Parks and Recreation Natural Resources Group) for granting permits to conduct research in Marine Park. We would also like to thank Chun-hua Yang (DNA Learning Center) for support to finalize figures. Lastly, we would like to thank Dr. Florian Krammer (Icahn School of Medicine at Mount Sinai) and *PLoS ONE* reviewers for constructive suggestions.

Author Contributions

Conceptualization: Christine Marizzi, Antonia Florio.

Data curation: Christine Marizzi, Antonia Florio.

Formal analysis: Christine Marizzi, Antonia Florio.

Funding acquisition: Antonia Florio, David Andrew Micklos.

Investigation: Christine Marizzi, Antonia Florio, Jenna Dorey, Sean McKenzie, Carlo Baria, Ron Bepat, Lena Cosentino, Alexander Dvorak, Amina Gacevic, Cristina Guzman-Moumtzis, Francesca Heller, Nicholas Alexander Holt, Jeffrey Horenstein, Vincent Joralemon, Manveer Kaur, Tanveer Kaur, Armani Khan, Jessica Kuppan, Scott Laverty, Camila Lock, Marianne Pena, Ilona Petrychyn, Indu Puthenkalam, Daval Ram, Arlene Ramos, Noelle Scoca, Rachel Sin, Izabel Gonzalez, Akansha Thakur, Husan Usmanov, Karen Han, Andy Wu, Tiger Zhu.

Project administration: Christine Marizzi, Christine Mazza, Fabiana Cellini.

Software: Mohammed Khalfan, Cornel Ghiban, Bruce Nash.

Supervision: Christine Marizzi, Antonia Florio, Christine Mazza, Fabiana Cellini.

Visualization: Christine Marizzi, Antonia Florio.

Writing – original draft: Christine Marizzi, Antonia Florio.

Writing – review & editing: Christine Marizzi, Antonia Florio, Melissa Lee, Mohammed Khalfan, Cornel Ghiban, Bruce Nash, Jenna Dorey, Sean McKenzie, Christine Mazza, Fabiana Cellini, Carlo Baria, Ron Bepat, Lena Cosentino, Alexander Dvorak, Amina Gacevic, Cristina Guzman-Moumtzis, Francesca Heller, Jeffrey Horenstein, Vincent Joralemon, Manveer Kaur, Tanveer Kaur, Armani Khan, Jessica Kuppan, Scott Laverty, Camila Lock, Marianne Pena, Ilona Petrychyn, Indu Puthenkalam, Daval Ram, Arlene Ramos, Noelle Scoca, Rachel Sin, Izabel Gonzalez, Akansha Thakur, Husan Usmanov, Karen Han, Andy Wu, Tiger Zhu, David Andrew Micklos.

References

1. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. *Proc R Soc B Biol Sci.* 2003 Feb 7; 270(1512):313–21. Available from: <https://doi.org/10.1098/rspb.2002.2218> PMID: 12614582
2. Hebert PDN, Ratnasingham S, DeWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc B Biol Sci.* 2003 Aug 7; 270(Suppl 1):S96–9. Available from: <https://doi.org/10.1098/rsbl.2003.0025> PMID: 12952648

3. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System: Barcoding. *Mol Ecol Notes*. 2007; 7(3):355–64. <https://doi.org/10.1111/j.1471-8286.2007.01678.x> PMID: 18784790
4. Larsen BB, Miller EC, Rhodes MK, Wiens JJ. Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life. *Q Rev Biol*. 2017; 92(3):229–65. Available from: <https://doi.org/10.1086/693564>
5. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*. 1994; 3(5):294–9. PMID: 7881515
6. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. *Proc Natl Acad Sci U S A*. 2009; 106(31):12794–7. Available from: <https://doi.org/10.1073/pnas.0905845106> PMID: 19666622
7. Eberhardt U. Methods for DNA barcoding of fungi. *Methods Mol Biol*. 2012; 858:183–205. Available from: https://doi.org/10.1007/978-1-61779-591-6_9 PMID: 22684957
8. Kress WJ, Erickson DL. DNA barcodes: methods and protocols. *Methods Mol Biol*. 2012; 858:3–8. Available from: https://doi.org/10.1007/978-1-61779-591-6_1 PMID: 22684949
9. Fišer Pečnikar Ž, Buzan E V. 20 years since the introduction of DNA barcoding: from theory to application. *J Appl Genet*. 2014; 55(1):43–52. Available from: <https://doi.org/10.1007/s13353-013-0180-y> PMID: 24203863
10. Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, et al. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol Ecol Resour*. 2014; 14(2):221–32. Available from: <https://doi.org/10.1111/1755-0998.12173> PMID: 24118947
11. Little DP. Authentication of Ginkgo biloba herbal dietary supplements using DNA barcoding. *Genome*. 2014; 57(9):513–6. Available from: <https://doi.org/10.1139/gen-2014-0130> PMID: 25495290
12. Shokralla S, Hellberg RS, Handy SM, King I, Hajibabaei M. A DNA Mini-Barcoding System for Authentication of Processed Fish Products. *Sci Rep*. 2015; 5.
13. Clark LF. The current status of DNA barcoding technology for species identification in fish value chains. *Food Policy*. 2015 Jul; 54:85–94. Available from: <https://doi.org/10.1016/j.foodpol.2015.05.005>
14. Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M, et al. DNA barcoding as a new tool for food traceability. *Food Res Int*. 2013 Jan; 50(1):55–63. Available from: <https://doi.org/10.1016/j.foodres.2012.09.036>
15. Wong EHK, Hanner RH. DNA barcoding detects market substitution in North American seafood. *Food Res Int*. 2008 Oct; 41(8):828–37. Available from: <https://doi.org/10.1016/j.foodres.2008.07.005>
16. Lowenstein JH, Amato G, Kolokotronis S-O. The Real *maccoyii* Identifying Tuna Sushi with DNA Barcodes—Contrasting Characteristic Attributes and Genetic Distances. *PLoS One*. 2009 Nov 18; 4(11):e7866. Available from: <https://doi.org/10.1371/journal.pone.0007866> PMID: 19924239
17. Little DP, Jeanson ML. DNA barcode authentication of saw palmetto herbal dietary supplements. *Sci Rep*. 2013 Dec 17; 3(1):3518. Available from: <https://doi.org/10.1038/srep03518> PMID: 24343362
18. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial Teas Highlight Plant DNA Barcode Identification Successes and Obstacles. *Sci Rep*. 2011 Dec 21; 1(1):42. Available from: <https://doi.org/10.1038/srep00042> PMID: 22355561
19. DeSalle R, Birstein VJ. PCR identification of black caviar. *Nature*. 1996; 381(6579):197–8. Available from: <https://doi.org/10.1038/381197a0>
20. Doukakis P, Pikitch EK, Rothschild A, DeSalle R, Amato G, Kolokotronis S-O. Testing the effectiveness of an international conservation agreement: marketplace forensics and CITES caviar trade regulation. *PLoS One*. 2012 Jul 25; 7(7):e40907. Available from: <https://doi.org/10.1371/journal.pone.0040907> PMID: 22848410
21. Schwartz J. Sushi study finds deception. *New York Times*. 22 Aug 2008. Available from: <http://www.nytimes.com/2008/08/22/world/americas/22iht-fish.1.15539112.html> Cited 1 May 2016.
22. New York City Department of Parks and Recreation. Marine Park. 2017. Available from: <https://www.nycgovparks.org/parks/marine-park>
23. Novacek MJ. Colloquium paper: engaging the public in biodiversity issues. *Proc Natl Acad Sci U S A*. 2008 Aug 12; 105(Suppl 1):11571–8. Available from: <https://doi.org/10.1073/pnas.0802599105> PMID: 18695244
24. Ballard HL, Dixon CGH, Harris EM. Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. *Biol Conserv*. 2017 Apr; 208:65–75. Available from: <https://doi.org/10.1016/j.biocon.2016.05.024>
25. Lindemann-Matthies P. The Influence of an Educational Program on Children's Perception of Biodiversity. *J Environ Educ*. 2002; 33(2):22–31.

26. Harris SE, Bellino M. DNA barcoding from NYC to Belize. *Science*. 2013 Dec 20; 342(6165):1462–3. Available from: <https://doi.org/10.1126/science.1230006> PMID: 24357310
27. Henter HJ, Imondi R, James K, Spencer D, Steinke D. DNA barcoding in diverse educational settings: five case studies. *Philos Trans R Soc B Biol Sci*. 2016; 371(1702):20150340. Available from: <https://doi.org/10.1098/rstb.2015.0340> PMID: 27481792
28. Henter HJ, Mel SF. Effects of Gender on Student Response to Course-Based Research. *J Coll Sci Teach*. 2016; 46(2):17–26. Available from: http://www.nsta.org/store/product_detail.aspx?id=10.2505/4/jcst16_046_02_17
29. Santschi L, Hanner RH, Ratnasingham S, Riconscente M, Imondi R. Barcoding life's matrix: translating biodiversity genomics into high school settings to enhance life science education. *PLoS Biol*. 2013 Jan 29; 11(1):e1001471. Available from: <https://doi.org/10.1371/journal.pbio.1001471> PMID: 23382648
30. Bowler I. Understanding Urban Audiences. Community Workshop Results for Rocky Mountain Arsenal NWR. 2015. Available from: [https://www.fws.gov/urban/workshop_pdfs/Community workshop results for Rocky Mountain Arsenal NWR.dox.pdf](https://www.fws.gov/urban/workshop_pdfs/Community%20workshop%20results%20for%20Rocky%20Mountain%20Arsenal%20NWR.dox.pdf)
31. Karsten L, Vliet W Van. Increasing Children's Freedom of Movement: Introduction. *Child Youth Environ*. 2006; 16(1).
32. Alden P, Cassie B. National Audubon Society field guide. New England. 1st. ed. 1998. 447 p.
33. Day L. Field guide to the natural world of New York City. The Johns Hopkins University Press; 2007. 353 p.
34. King JR, Porter SD. Recommendations on the use of alcohols for preservation of ant specimens (Hymenoptera, Formicidae). *Insectes Soc*. 2004 May 1; 51(2):197–202. Available from: <https://doi.org/10.1007/s00040-003-0709-x>
35. Jackson CJ, Barton RC, Evans EG. Species identification and strain differentiation of dermatophyte fungi by analysis of ribosomal-DNA intergenic spacer regions. *J Clin Microbiol*. 1999 Apr; 37(4):931–6. Available from: <https://doi.org/10.1111/j.1365-294X.1993.tb00005.x> PMID: 10074504
36. Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol Ecol*. 1993 Apr; 2(2):113–8. Available from: <https://doi.org/10.1111/j.1365-294X.1993.tb00005.x> PMID: 8180733
37. Messing J. New M13 Vectors for Cloning. *Methods Enzymol*. 1983; 101(C):20–78.
38. Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, et al. The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Front Plant Sci*. 2011; 2(34):1–16. Available from: <https://doi.org/10.3389/fpls.2011.00034> PMID: 22645531
39. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLoS Biol*. 2016 Jan 11; 14(1): e1002342. Available from: <https://doi.org/10.1371/journal.pbio.1002342> PMID: 26752627
40. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998 Mar 1; 8(3):175–85. Available from: <https://doi.org/10.1101/gr.8.3.186> PMID: 9521921
41. Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res*. 1998 Mar 1; 8(3):175–85. Available from: <https://doi.org/10.1101/gr.8.3.175> PMID: 9521921
42. NYS Department of Environmental Conservation and NYS Department of Agriculture and Markets. New York State Prohibited and Regulated Invasive Animals and Plants. 2014.
43. The Brooklyn Botanic Garden. Invasiveness Assessment Scores and Ranks for 183 Nonnative Plant Species in NYS. 2013.
44. Zheng H, Wu Y, Ding J, Binion D, Fu W, Reardon RC. Forest Health Technology Enterprise Team Invasive Plants of Asian Origin Established in the United States and Their Natural Enemies. *Biol Control*. 2006; 1:160. Available from: <http://www.cabdirect.org/abstracts/20067203583.html>
45. Ross KG, Gotzek D, Ascunce MS, Shoemaker DD. Species Delimitation: A Case Study in a Problematic Ant Taxon. *Syst Biol*. 2010 Mar 1; 59(2):162–84. Available from: <https://doi.org/10.1093/sysbio/syp089> PMID: 20525628
46. Hughes JB, Daily GC, Ehrlich PR. Conservation of Insect Diversity: a Habitat Approach. *Conserv Biol*. 2000 Dec 18; 14(6):1788–97. Available from: <https://doi.org/10.1046/j.1523-1739.2000.99187.x>
47. Bernard EC. Soil nematode biodiversity. *Biol Fertil Soils*. 1992 Oct; 14(2):99–103. Available from: <https://doi.org/10.1007/BF00336257>
48. Potter JW, McKeown AW. Nematode biodiversity in Canadian agricultural soils. *Can J Soil Sci*. 2003 Aug; 83(Special Issue):289–302. Available from: <https://doi.org/10.4141/S01-064>

49. Stevens PF. Angiosperm Phylogeny Website. Version 12. Missouri Botanical Garden. 2001. Available from: <http://www.mobot.org/MOBOT/research/APweb/>.
50. Frodin DG. History and concepts of big plant genera. Vol. 53, Taxon. 2004. p. 753–76.
51. Schindel DE, Miller SE. DNA barcoding a useful tool for taxonomists. Nature. 2005; 435(7038):17–17. Available from: <https://doi.org/10.1038/435017b>
52. Moulton MJ, Song H, Whiting M. Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). Mol Ecol Resour. 2010; 10:615–27. Available from: <https://doi.org/10.1111/j.1755-0998.2009.02823.x> PMID: 21565066
53. Stoeckle MY, Kerr KCR. Frequency matrix approach demonstrates high sequence quality in avian BARCODEs and highlights cryptic pseudogenes. Zuo Z, editor. PLoS One. 2012 Aug 27; 7(8):e43992. Available from: <https://doi.org/10.1371/journal.pone.0043992> PMID: 22952842
54. Hollingsworth PM, Graham SW, Little DP. Choosing and Using a Plant DNA Barcode. Steinke D, editor. PLoS One. 2011 May 26; 6(5):e19254. Available from: <https://doi.org/10.1371/journal.pone.0019254> PMID: 21637336
55. Shylla JA, Ghatani S, Tandon V. Utility of divergent domains of 28S ribosomal RNA in species discrimination of paramphistomes (Trematoda: Digenea: Paramphistomoidea). Parasitol Res. 2013 Dec 6; 112(12):4239–53. Available from: <https://doi.org/10.1007/s00436-013-3616-8> PMID: 24096607
56. Lehmitz R, Decker P. The nuclear 28S gene fragment D3 as species marker in oribatid mites (Acari, Oribatida) from German peatlands. Exp Appl Acarol. 2017 Mar 12; 71(3):259–76. Available from: <https://doi.org/10.1007/s10493-017-0126-x> PMID: 28405837
57. Cheng T, Xu C, Lei L, Li C, Zhang Y, Zhou S. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Mol Ecol Resour. 2016 Jan; 16(1):138–49. Available from: <https://doi.org/10.1111/1755-0998.12438> PMID: 26084789
58. Kress WJ, Erickson DL. A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. Shiu S-H, editor. PLoS One. 2007 Jun 6; 2(6):e508. Available from: <https://doi.org/10.1371/journal.pone.0000508> PMID: 17551588
59. Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C. Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. Syst Biol. 2009 Aug 1; 58(4):445–51. Available from: <https://doi.org/10.1093/sysbio/syp033> PMID: 20525598
60. Raja HA, Baker TR, Little JG, Oberlies NH. DNA barcoding for identification of consumer-relevant mushrooms: A partial solution for product certification? Food Chem. 2017; 214:383–92. <https://doi.org/10.1016/j.foodchem.2016.07.052> PMID: 27507489
61. Savage AM, Hackett B, Guénard B, Youngsteadt EK, Dunn RR. Fine-scale heterogeneity across Manhattan’s urban habitat mosaic is associated with variation in ant composition and richness. Schonrogge K, Orivel J, editors. Insect Conserv Divers. 2015 May; 8(3):216–28. Available from: <https://doi.org/10.1111/icad.12098>
62. Ivanov K, Lockhart OM, Keiper J, Walton BM. Status of the exotic ant *Nylanderia flavipes* (Hymenoptera: Formicidae) in northeastern Ohio. Biol Invasions. 2011 Sep 20; 13(9):1945–50. Available from: <https://doi.org/10.1007/s10530-011-0021-z>
63. Wetterer JK. Worldwide Spread of the Yellow-Footed Ant, *Nylanderia flavipes* (Hymenoptera: Formicidae). Florida Entomol. 2011; 94(July):582–7. Available from: <https://doi.org/10.1653/024.094.0323>
64. Youngsteadt E, Henderson RC, Savage AM, Ernst AF, Dunn RR, Frank SD. Habitat and species identity, not diversity, predict the extent of refuse consumption by urban arthropods. Glob Chang Biol. 2015 Mar; 21(3):1103–15. Available from: <https://doi.org/10.1111/gcb.12791> PMID: 25463151
65. Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, et al. Assessment of course-based undergraduate research experiences: a meeting report. CBE Life Sci Educ. 2014 Mar 1; 13(1):29–40. Available from: <https://doi.org/10.1187/cbe.14-01-0004> PMID: 24591501
66. Micklos D, Lauter S, Nisselle A. Lessons from a Science Education Portal. Science (80-). 2011 Dec 23; 334(6063):1657–8. Available from: <https://doi.org/10.1126/science> PMID: 1197074