

A Rapid Protocol for Generating Arthropod DNA Barcodes Suitable for Use with Undergraduate

Students

Authors: Grant R. Brown^{a*}, Claire Dagen^a, Ben Reilly O'Donnell^a, Jefferson A. Graves^a

Author information:

Grant R. Brown (author for correspondence) : grb31@st-andrews.ac.uk

^aUniversity of St Andrews, St Andrews, Fife, UK. KY16 9TH

Short Title

RAPID DNA BARCODING BY STUDENTS

Abstract

We provide a protocol for rapid DNA extraction from spiders suitable for undergraduate practical sessions. Students who were previously naïve to the theory and laboratory technique of DNA barcoding were successfully able to extract and recover 29 DNA sequences from 16 species of small spiders in the family Linyphiidae. We anticipate that with careful selection of specimens, undergraduate students could participate in sessions which both benefit their professional development and further taxonomic understanding across a variety of organisms.

Keywords: education, citizen science, systematics, taxonomy, interdisciplinary

Introduction

The majority of recent taxonomic papers make use of both morphological and DNA evidence in phylogenies (Pante, Schoelinck, and Puillandre 2015). The synthesis of morphological, behavioural, ecological and genetic evidence is generally considered objectively superior to analyses performed using only a single evidence source (Dayrat 2005), and is likely to increase the efficiency of taxonomic progress by reducing the need for subsequent authors to revise poorly described or placed taxa (Sangster and Luksenburg 2015). Of these additional data sources, DNA sequence data is likely to be the simplest and most efficient to collect (Hajibabaei et al. 2007). DNA barcodes, short sections of the mitochondrial cytochrome oxidase subunit I genome, are the most frequently included DNA sequence data (Hebert and Gregory 2005). Mitochondrial DNA is useful for barcoding as it is very abundant in all eukaryotic cells, and has the property that it evolves rapidly, yet without extensive variation. Hence, these sequences are likely to be identical between all members of the same species, yet will be distinct from members of other species (Hebert and Gregory 2005).

There are many taxa however, for which sequence information is available for only a small proportion of their total species richness. For example, there are currently 46,650 valid species of spider (WSC 2018), but only 12% of these species have DNA barcodes associated with them (Ratnasingham and Hebert 2007). This contrasts significantly with such taxa as the Lepidoptera, for which nearly 55% of the 165,000 recognised species have been barcoded at the mitochondrial cytochrome oxidase 1 gene as of April 2017 (iBOL 2018). Taxonomists have called for wider sampling within taxa, including spiders, to facilitate understanding of their systematics and evolution (Garrison et al. 2016). However, obtaining sufficient funding to barcode every species is likely to be difficult from traditional funding streams (Hebert and Gregory 2005). Since there is real risk that many species will become extinct before they are observed by taxonomists and conservationists, there is a pressing need to document taxa more efficiently (Agnarsson, Kuntner, and Paterson 2007).

A possible solution to sample taxa more broadly and generate sequence data is through integration of these methodologies with undergraduate practical sessions. Henter et al. (2016) recently highlighted the suitability of DNA barcoding for use in a variety of educational settings, and in particular the contributions from the National Science Foundation funded San Diego Biodiversity Project (Butler, Henter, and Mel 2014). Students enrolled in this module prepared local invertebrate samples for the extraction and amplification of mitochondrial DNA for sequencing at an external processing facility (Butler, Henter, and Mel 2014). DNA barcodes assembled by students were eventually hosted on the Barcode of Life Data System (Ratnasingham and Hebert 2007), where they are freely accessible for other researchers.

Student practical sessions of this nature present ideal opportunities to introduce students to modern interdisciplinary research techniques, as well as reinforce existing concepts related to speciation, phylogenetics and taxonomic/species identification. The leading funding body for environmental science in the UK, The Natural Environment Research Council (NERC), recently identified the need for students to gain skills in these areas as vital for further post-graduate employment (NERC 2012). Therefore, there is ample opportunity for teachers to both enhance student learning outcomes relevant to industry needs and facilitate student contributions to wider scientific research goals.

In this paper, we present details of a protocol for a rapid DNA extraction practical suitable for undergraduate workers which we developed and ran in October 2016 at the University of St Andrews, Scotland. This protocol differs from that summarised by Henter et al. (2016) in that the protocol can be completed in a short period of only 2 days, and that both the morphological and DNA sequence samples are carefully curated such that they can form reliable reference barcodes. Students who were previously naïve to the theory and process of DNA barcoding were able to successfully extract and recover DNA sequences of suitable length (600-800bp) to produce voucher

barcodes from small spiders in the family Linyphiidae. We anticipate that with careful selection of specimens, students could be involved in generating data to further the taxonomic understanding of arachnids and other understudied organisms, whilst also benefitting from hands-on laboratory experience that synthesises modern molecular and whole-organism research techniques.

Methods

Source Specimens

Living spiders were collected during the period July – October 2016 at several localities in Fife and Perth, Scotland (collection records for each specimen are provided in supplementary table 1). The majority of specimens were collected by grubbing ground layer leaf-litter deposits, although a few specimens were opportunistically hand-collected from inside offices or homes. Spiders of the family Linyphiidae were the primary focus of the practical, although a few specimens from other families (*Robertus lividus* Blackwall 1836 (Araneae: Theridiidae) and *Cryphoeca silvicola* C.L.Koch 1834 (Araneae: Hahniidae)) were included. These species are frequently confused for linyphiids during field collection (G.R.Brown, unpublished data). Specimens were euthanised by freezing and stored at -86°C until use in October 2016. After removal from the freezer, specimens were immersed in 70% ethanol for all handling steps (i.e. identification and the removal of legs for DNA extraction). All specimens used in the session have been retained and are available on request .

Specimen Identification and Initial DNA Extraction

Students worked individually on two specimens each and attempted identification to species level during the first session of the practical using standard taxonomic resources (Roberts 1987; Roberts 1995). It was not expected that students would successfully identify all specimens to species level, and the emphasis of this aspect of the practical was placed on the recognition of diagnostic features and to gain an appreciation of spider anatomy. Students examined their specimens used Leica MS-5 6.3x-40x zoom stereomicroscopes. Species were identified after the practical by the lead

author using an Amscope stereo inspection microscope (model ZM-1TNW2-80AM-9M) and standard keys (Roberts 1987).

The legs of each specimen were removed at the trochanter-femur joint using a clean scalpel or tungsten needle and placed into Eppendorf tubes containing 100µl of fly squishing buffer (Gloor and Engels 1992) for DNA extraction. Extractions were then incubated overnight at 55°C before being boiled for two minutes to halt any further activity of proteinase K (Gloor and Engels 1992). The extraction of DNA using this approach can be considered a semi-destructive protocol in that it removes only a portion of the specimen and did not require grinding of the leg tissue (Paquin and Vink 2009). There is however a caveat, in that dissection proficiency varied considerably between students and in some instances the specimen was substantially damaged during removal of the legs.

Generation of DNA Barcodes

PCR was performed on leg tissue extractions using the primers LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') (Folmer et al. 1994) and Chelicerate Reverse 2 (abbreviated to CHR-2 hereafter; 5'-GGATGGCCAAAAATCAAATAAATG-3')(Barrett and Hebert 2005). Each reaction tube contained 0.4µl 50µM/ml LCO1490, 0.4µl 50µM/ml CHR-2, 1µl of the DNA extract, 8.2µl mQ H₂O and 10µl Bioline MyTaq™ mix DNA polymerase (item number: BIO-25041). PCR was a single denaturation step at 94°C for 180 seconds, with 40 cycles of 15s at 94°C, 30s at 50°C and 40s at 72°C, and finishing with a single 300s elongation step at 72°C (Miller et al. 2013).

The approximate size of the DNA fragments amplified during PCR was assessed by gel electrophoresis. The PCR product, stained with orange G dye, was compared against a 100 bp ladder (Thermo Scientific GeneRuler 100bp; catalogue number: SM0322). Samples of the correct length (600-700bp) for generating barcode sequences were then cleaned up for sequencing using Illustra™ ExoProStar™ 1-step. Each digest sample contained 2µl ExoProStar™ 1-step™ and 10µl of PCR

product. Samples were then placed in the thermocycler for one cycle of 15min at 37°C and 15min at 80°C.

The concentration of DNA in each sample following the digestion stage was estimated by comparison against four Promega Lambda DNA standards (5, 10, 20 and 50 ng/μl), using 1.5% agarose gels. Samples for sequencing were a total volume of 6μl, consisting of 1μl (6.4μM/ml) primer (forward or reverse) and 5-20ng DNA. The total sample volume was made up to 6μl using mQ H₂O. Where possible, both forward and reverse primer- containing sample were sent for sequencing. Samples were sequenced by Edinburgh Genomics, Edinburgh (Edinburgh Genomics 2018). The turnaround for students to receive their sample chromatograms was approximately one week from sample submission.

Interpretation of Sequence Chromatograms

Chromatograms were interpreted by students using FinchTV (Geospiza 2016) in order to generate barcodes as part of the practical assessment. The lead author checked student chromatogram reads and sequences were compared either to publically available reference barcodes on BOLD(Ratnasingham and Hebert 2007), or searched against the National Center for Biotechnology Information (NCBI) nucleotide database (NCBI 2018) for matches .Chromatogram quality was assessed via visual inspection and Phred scores using CodonCode Aligner (CodonCode Corporation 2018). Phred scores reflect read quality; a score of 10 represents a 1 in 10 probability that the base has been miscalled, whilst a score of 50 represents a 1 in 100,000 change that base has been misidentified. These error probabilities are a reliable indicator of sequence quality, and Phred scores above 20 are typically interpreted as "high quality" calls (Ewing and Green 1998).

Timetabling and Goals for Student Learning and Development

The practical session ran over two days (9-1pm; 9-5pm) and was split into two themes. The first day (9-1pm) focused on morphological identification of spiders to species level and the removal of leg tissue for DNA extraction. This session developed skills in stereomicroscopy, small specimen handling and dissection, and the use of standard dichotomous keys for species identification (Roberts 1987).

The second day made use of the leg tissue extractions and focused on preparing the samples for PCR, digestion and shipping the PCR product to the sequencing facility. The goals for this session were to develop skills and confidence in the use of standard molecular laboratory equipment (such as centrifuges and thermocyclers), the calculation of dilutions and reliable use of Gilson pipettes.

Student Background

Students were from a mixed background of biology-related disciplines entering their third year (of four) towards a BSc in Biological sciences at the University of St Andrews. Prior to the practical students were introduced to the basics of DNA barcoding via a guest lecture delivered by Professor Peter Hollingsworth (University of Edinburgh) but were otherwise naïve to DNA extraction protocols and techniques for the identification of spiders. The protocol as provided to the students, and recipes for all formulations, are given in the supplementary material ("Student Protocol 2016.docx").

Results

Generation of Species Barcodes

Students successfully generated barcodes from 29 specimens representing 16 species of spider (table 1). Sequence quality for successful extractions was generally high and the proportion of chromatogram bases with Phred scores of 20+ ranged between 71 and 98.8% (table 1).

[TABLE 1]

The extraction success on a per sample basis was 46%, with 38 successful extractions out of 81 attempts (supplementary table 1). Of the successful extractions, 76% yielded usable barcode sequences (29 sequences out of 38 sent to the sequencing facility; supplementary table 1). The mean length of successful barcode obtained was 639 ± 20 bp with minimum and maximum lengths of 566bp and 655bp respectively (table 1). There was agreement between morphological identification and sequence identity when compared to other published sequences available via NCBI and BOLD databases for all but one sequence (specimen 17; table 1).

Discussion

This paper provides a technique for an undergraduate practical that can be successfully implemented to introduce students to modern taxonomic research techniques and also contribute sequence data for wider research use. Students who were inexperienced with regards to DNA extraction protocols, and morphology-based techniques in species identification, were able to successfully generate useable DNA barcodes of appropriate length and quality for inclusion in future systematic or taxonomic research (table 1). Sequences produced by students generally had high agreement with cytochrome oxidase subunit I (COI) sequences available via the NCBI and BOLD databases, and with the morphological identifications performed by the lead author (table 1). Sequences were generally of high quality as quantified using Phred scores (table 1), lending support to our assertion that, with suitable oversight, students can contribute reliable data to further research efforts. Our findings complement the experiences of other authors (summarised in Henter et al. 2016; see also Harris and Bellino (2013)) and highlight the potential for taxonomic progress even in the face of an apparent funding and manpower crisis.

Feedback from students during and following the practical was positive. Whilst many students commented on the challenging nature of the practical session, there was enthusiasm over the potential to contribute novel research data for wider research use. Additionally, several students expressed positive feedback regarding the “completeness” of the practical, which introduced students to a clear and self-contained start, middle and end of a small research project. Student feedback in our institution is gathered on a per module basis, and asks that students synthesise their experiences and opinions from lectures, practical sessions, and assessments. One possible limitation of this approach is that it is more difficult to evaluate task specific learning gain (Holloway 2006). The use of pre- and post- session questionnaires evaluating student understanding of the practical learning objectives would be a useful improvement to this protocol.

Henter et al. (2016) provide a succinct account of the challenges in running DNA barcoding sessions in educational settings, which we also encountered in running this laboratory practical session. In particular, there was a need for a significant amount of preparatory time from both academic and technical staff. Aliquoting student consumables prior to the practical was probably the most time-consuming activity, although it is likely this can be trimmed down by placing more emphasis on students completing these tasks themselves than we did initially (suggestions are given in the annotated student protocol file; supplementary material). The field collection and identification of donor specimens by specialists also represents a considerable input of time, although this could be circumvented through the use of specimens collected by other researchers, or by using species which are abundant in museum collections. The protocol used here will successfully extract DNA from spider specimens up to 15 years old post mortem (G.R.Brown, unpublished data), although if the main focus of the practical is to examine old material then it may be more appropriate to use more specific method developed for handling degraded specimens (Miller et al. 2013).

Aside from the challenges for teaching staff, we observed a few areas in which students generally struggled. The most significant effector of the overall barcoding success was undoubtedly the use of Gilson pipettes to transfer small volumes during the PCR amplification and digest stages. When preparing samples for shipment to the sequencing facility, there were a considerable number of samples that did not contain the correct volume of digest product (6 μ l), and these samples then failed to produce useable barcodes. The inclusion of a short training session on the use of Gilson pipettes could be added, which would likely increase the rate of successful barcode production. This could be achieved simply, by asking students to pipette small quantities of water into PCR tubes and assessing their accuracy by weight.

Students also struggled, perhaps not unexpectedly, with the manipulation of the smallest specimens (e.g. *Porrhomma pygmaeum* Blackwall 1834). For these small species there was generally considerable damage to the specimen during leg removal and identification, which would render the specimen of low value for further morphological investigation. Students were provided with species local to Fife, Scotland, and which are also generally common in the UK (SRS 2018). While there was no significant concern regarding loss or extensive damage to specimens in this instance, we would advise caution in the choice of specimens if material is harder to come by. Use of larger species would also be likely to improve the chances that the specimen could be used after the extraction.

A final consideration is the cost of the practical. The total start-up cost for us here in St Andrews was £1050, which included the cost of sequencing 81 samples (£364.50) and the bulk-purchasing of consumables (including primers, PCR tube strips, Illustra™ ExoProStar™ 1-step, etc.).

In comparison costs of \$15 per specimen for the San Diego Biodiversity Project (Henter et al. 2016) our protocol is marginally more expensive per student, although we are perhaps more constrained by the smaller class sizes at St Andrews, and we also considerably over-estimated the need for backup aliquots to account for student error.

Despite these challenges, there is considerable scope for DNA barcoding practical sessions performed by undergraduates to further taxonomic research from the generation of sequence data. With judicious selection of specimens and standard primers, novel barcode sequences could be generated without the need for additional external funding, and the general approach could be extended to generate nuclear gene sequences (e.g. ribosomal subunit 12, 16, 18, etc.) for deeper phylogenetic analysis. Not only would this have considerable benefit to wider phylogenetic and taxonomic research efforts, but it would considerably benefit students by exposing them to modern multidisciplinary techniques in zoology, which are vital for further research or employment ambitions (Natural Environment Research Council 2012).

Acknowledgements

The authors would like to thank the biology technical support staff (Mrs Penny Hood, Mrs Maureen Cunningham, Mr Alasdair Stenhouse and Mrs Tanya Sneddon) at the University of St Andrews for their help setting up the practical and facilitating its smooth completion. We would also like to thank the students who participated in the practical and who engaged well with the material.

Funding

This work was not supported by any external funding.

Supplementary information

The following supplemental material is available online:

[Brown et al. 2018 Supplementary Information.docx](#): An annotated version of the student protocol with suggested modifications based on our experience running the session A list of aliquots required for the practical, and DNA barcodes for each species.

Supplementary Table 1.xlsx: The specimen occurrence data and details of sequencing success, including a list of specimens examined which did not produce useable barcodes.

References

- Agnarsson, I., M. Kuntner, and A. Paterson. 2007. "Taxonomy in a Changing World: Seeking Solutions for a Science in Crisis." *Systematic Biology* 56 (3):531-539. doi: 10.1080/10635150701424546.
- Barrett, R. D. H., and P. D. N. Hebert. 2005. "Identifying spiders through DNA barcodes." *Canadian Journal of Zoology* 83 (3):481-491. doi: 10.1139/z05-024.
- Butler, M., H. Henter, and S. Mel. 2014. "From bugs to barcodes: Using molecular tools to study biodiversity." *Tested Studies for Laboratory Teaching* 35:41-55.
- CodonCode Corporation. 2018. CodonCode Aligner version 8.02. Online at <https://www.codoncode.com>, accessed 27th November 2018.
- Dayrat, B. 2005. "Towards integrative taxonomy." *Biological Journal of the Linnean Society* 85 (3):407-415. doi: 10.1111/j.1095-8312.2005.00503.x.
- Edinburgh Genomics. 2018. University of Edinburgh, Edinburgh. Online at <https://genomics.ed.ac.uk>, accessed 27th November 2018.
- Ewing B. and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Research* 8(3): 186-194.

- Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. "DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates." *Molecular marine biology and biotechnology* 3 (5):294-299.
- Garrison, N. L., J. Rodriguez, I. Agnarsson, J. A. Coddington, C. E. Griswold, C. A. Hamilton, M. Hedin, K. M. Kocot, J. M. Ledford, and J. E. Bond. 2016. "Spider phylogenomics: untangling the Spider Tree of Life." *PeerJ* 4:e1719. doi: 10.7717/peerj.1719.
- Geospiza, Inc. 2016. "FinchTV v.1.4.0." Online at <https://digitalworldbiology.com/FinchTV>, accessed 27th November 2018.
- Gloor, G. B. , and W. R. Engels. 1992. "Single-fly DNA preps for PCR." *Drosophila Information Service* 71:148-149.
- Hajibabaei, M., G. A. C. Singer, P. D. N. Hebert, and D. A. Hickey. 2007. "DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics." *Trends in Genetics* 23 (4):167-172. doi: <http://doi.org/10.1016/j.tig.2007.02.001>.
- Harris, S. E., and M. Bellino. 2013. "DNA Barcoding from NYC to Belize." *Science* 342 (6165):1462-1463. doi: 10.1126/science.1230006.
- Hebert, P. D. N., and T. R. Gregory. 2005. "The Promise of DNA Barcoding for Taxonomy." *Systematic Biology* 54 (5):852-859. doi: 10.1080/10635150500354886.
- Henter, H. J., R. Imondi, K. James, D. Spencer, and D. Steinke. 2016. "DNA barcoding in diverse educational settings: five case studies." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702).
- Holloway, J. H. 2006. "Connecting Professional Development to Student Learning Gains". *Science Educator* 15(1): 37-43.
- iBOL (International Barcode of Life Project). (2018). Lepidoptera Barcode of Life project. Biodiversity Institute of Ontario, University of Guelph. Online at www.lepbarcoding.org, accessed 27th November 2018.

- Miller, J. A., K. K. Beentjes, P. van Helsdingen, and S. Ijland. 2013. "Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol." *ZooKeys* (365):245-261. doi: 10.3897/zookeys.365.5787.
- NCBI (National Center for Biotechnology information). 2018. U.S. National Library of Medicine, Bethesda, USA. Online at <https://www.ncbi.nlm.nih.gov>, accessed 27th November 2018.
- NERC (Natural Environment Research Council). 2012. "NERC postgraduate and professional skills needs review." Online at <https://nerc.ukri.org/skills/postgrad/policy/skillsreview/2012/skills-framework/>: NERC. Accessed 12th June 2018.
- Pante, E., C. Schoelinck, and N. Puillandre. 2015. "From Integrative Taxonomy to Species Description: One Step Beyond." *Systematic Biology* 64 (1):152-160. doi: 10.1093/sysbio/syu083.
- Paquin, P., and C. J. Vink. 2009. "Testing compatibility between molecular and morphological techniques for arthropod systematics: a minimally destructive DNA extraction method that preserves morphological integrity, and the effect of lactic acid on DNA quality." *Journal of Insect Conservation* 13 (4):453-457. doi: 10.1007/s10841-008-9183-0.
- Ratnasingham, S., and P. D. N. Hebert. 2007. "bold: The Barcode of Life Data System (<http://www.barcodinglife.org>)." *Molecular Ecology Notes* 7 (3):355-364. doi: 10.1111/j.1471-8286.2007.01678.x.
- Roberts, M. J. 1987. *The Spiders of Great Britain and Ireland. Volume 2. Linyphiidae and checklist*. Harley Books, Martins, Essex.
- Roberts, M.J. 1995. *Spiders of Britain & Northern Europe*. Harper Collins, London.
- Sangster, G., and J. A. Luksenburg. 2015. "Declining Rates of Species Described per Taxonomist: Slowdown of Progress or a Side-effect of Improved Quality in Taxonomy?" *Systematic Biology* 64 (1):144-151. doi: 10.1093/sysbio/syu069.
- SRS (Spider Recording Scheme). 2018. Online at <http://srs.britishspiders.org.uk>, accessed 27th November 2018.

WSC (World Spider Catalog). 2018. World Spider Catalog. Version 19.5. Natural History Museum Bern, online at <http://wsc.nmbe.ch>, accessed on 27th November 2018. doi: 10.24436/2