DNA barcodes from snake venom: a broadly applicable method for extraction of DNA from snake venoms

Cara Francesca Smith1, Mitchell E McGlaughlin1 & Stephen P Mackessy*.1

1School of Biological Sciences, University of Northern Colorado, 501 20th Street, Greeley, CO 80639–0017, USA

Keywords: c-mos • cyt b • DNA barcoding • ND4 • phylogenetics • snake venoms

DNA barcoding is a simple technique used to develop a large-scale system of classification that is broadly applicable across a wide variety of taxa. DNA-based analysis of snake venoms can provide a system of classification independent of currently accepted taxonomic relationships by generating DNA barcodes specific to each venom sample. DNA purification from dried snake venoms has previously required large amounts of starting material, has resulted in low yields and inconsistent amplification, and was possible with front-fanged snakes only. Here, we present a modified DNA extraction protocol applied to venoms of both front- and rear-fanged snakes that requires significantly less starting material (1 mg) and yields sufficient amounts of DNA for successful PCR amplification of regions commonly used for DNA barcoding.

METHOD SUMMARY
Modifications to a commercial DNA extraction kit (cultured cells protocol) allow purification of DNA from snake venoms for DNA-barcoding efforts. The outlined protocol uses ~100x less venom than previously possible and is applicable to both front- and rear-fanged venomous snakes.
have been sequenced from thousands of animal species and are commonly used as universal markers across taxa [3–6]. Although DNA barcode sequences from these regions may only provide broad initial information, this can be used as a molecular foundation for flagging novel specimens for further analysis for purposes of species discovery, determining phylogenetic relationships and investigating population-level genetic variation [3].

Recent efforts have been made to barcode all reptiles and amphibians systematically, as only a fraction of the total herpetological diversity has been previously examined [7,8]. DNA-barcoding efforts have revealed previously unrecognized diversity due to the presence of cryptic species [8], particularly in areas of high herpetofaunal biodiversity [9]. For example, Pomerantz et al. [10] performed real-time DNA barcoding of reptiles in the Chocó rainforest of Ecuador with portable lab equipment to collect and analyze genetic data on rare, cryptic and endangered species, demonstrating the immediate utility of DNA barcoding efforts for assessing biodiversity. Evaluating global herpetological diversity is critical, as reptiles and amphibians represent two of the most speciose vertebrate groups on earth [8]. To further this initiative, the development of reptile- and amphibian-specific COI primers has allowed for a robust data set of over 3000 species to be compiled [8].

Aiding in the effort to create a large-scale herpetofaunal database, DNA-based analysis of snake venoms allows systematic classification independent of current taxonomy by generating DNA barcodes specific to each venom sample [11]. Barcoding individual venoms provides a standardized way to classify venom samples, to confirm species identification and delineation, and to update sample identification based on taxonomic revisions. Correct identification of venom samples, especially when the source is unknown, is integral to the understanding of inter- and intra-specific venom variation.

### Table 1. Forward (F) and reverse (R) primer sequence.

<table>
<thead>
<tr>
<th>Gene Sequence (5’ to 3’) Amplicon length (bp) Location Ref.</th>
<th>Location Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>500–501 Nuclear [37]</td>
<td></td>
</tr>
<tr>
<td>630 Mitochondrial [21]</td>
<td></td>
</tr>
<tr>
<td>387 Mitochondrial [35]</td>
<td></td>
</tr>
<tr>
<td>651–652 Mitochondrial [36]</td>
<td></td>
</tr>
<tr>
<td>500–501 Nuclear [37]</td>
<td></td>
</tr>
<tr>
<td>630 Mitochondrial [21]</td>
<td></td>
</tr>
<tr>
<td>387 Mitochondrial [35]</td>
<td></td>
</tr>
<tr>
<td>651–652 Mitochondrial [36]</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Qubit™ DNA yield (ng/ml) from venom and sequencing success (✓) or failure (X) at mitochondrial (ND4 and cyt b) and nuclear (c-mos) regions.

<table>
<thead>
<tr>
<th>Viperidae</th>
<th>Qubit™ DNA yield (ng/ml)</th>
<th>DNA yield (ng/mg venom)</th>
<th>ND4</th>
<th>cyt b</th>
<th>c-mos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agkistrodon contortrix pictigaster</td>
<td>27.2</td>
<td>1.13</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Bitis gabonica</td>
<td>9.3</td>
<td>0.42</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Crotalus oreganus concolor</td>
<td>25.2</td>
<td>0.79</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Crotalus oreganus helleri</td>
<td>11.2</td>
<td>0.43</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Crotalus scutulatus scutulatus</td>
<td>12</td>
<td>0.38</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Crotalus simus tibasic</td>
<td>196</td>
<td>9.8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Crotalus viridis viridis</td>
<td>80</td>
<td>2.5</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Daboia russelli</td>
<td>256</td>
<td>12.8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sistrurus catenatus edwardsii</td>
<td>10</td>
<td>0.5</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Vipera ammodytes</td>
<td>60</td>
<td>2.88</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Colubridae</th>
<th>Qubit™ DNA yield (ng/ml)</th>
<th>DNA yield (ng/mg venom)</th>
<th>ND4</th>
<th>cyt b</th>
<th>c-mos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boiga dendrophilia</td>
<td>88.8</td>
<td>4.11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Boiga irregularis</td>
<td>20</td>
<td>0.88</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Drypetes fulgidos</td>
<td>49.6</td>
<td>2.43</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Philodryas baroni</td>
<td>544</td>
<td>26.41</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Elapidae</th>
<th>Qubit™ DNA yield (ng/ml)</th>
<th>DNA yield (ng/mg venom)</th>
<th>ND4</th>
<th>cyt b</th>
<th>c-mos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dendroaspis polylepis</td>
<td>65.5</td>
<td>3.28</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Naja annulifera</td>
<td>70.4</td>
<td>3.52</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Naja atra</td>
<td>3520</td>
<td>176</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Naja kaouthia</td>
<td>496</td>
<td>24.8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Naja naja</td>
<td>93.6</td>
<td>4.68</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ophiophagus hannah</td>
<td>37.6</td>
<td>1.88</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Oxyuranus scutellatus</td>
<td>34.8</td>
<td>1.74</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Pseudechis parryii</td>
<td>169</td>
<td>8.45</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Further, because venom components vary taxonomically, geographically, ontogenetically and with dietary preference [12,13], venomic analysis alone is unlikely to provide definitive species determination of unknown samples. Recent revision or clarification of taxonomic relationships within venomous snakes has led to the invalidation of previous species’ delineations, with the concomitant complication of identification of venom samples collected without voucher reference specimens of the snakes [14–25]. Because snake higher taxonomy is still disputed and warrants further investigation, it is imperative to couple phenotypic data (in this case, venom components and protein variation) with genotypic data [15,21]. These shifts in taxonomic classification may negate or obfuscate previous conclusions made about the influence of phylogeny on venom variation and composition of a species, for example in the C. viridis clade [26,27]. Venom-based DNA barcoding provides a framework that allows for venom classifications to be updated in parallel with (yet independent of) taxonomic reassignments.

The combination of fine-level venom variation and the possibility of species misidentification, especially in light of taxonomic reorganization, can further complicate conclusions about venomic data. Such misidentification can have severe consequences with regard to snakebite treatment and management [28,29]. If these clear identifications to species are available regionally, then misapplication of antivenom therapy can be avoided.

DNA barcoding of venom samples is also useful for the confirmation of species identity of commercially available venoms, which may contain pooled venom samples from multiple species or may occasionally be entirely misidentified [11,29]. In addition, DNA-based analysis of snake venoms allowed for rapid species identification during a forensic investigation when no other information was available about a sample [30]. This analysis provided unequivocal evidence of illegal poaching, possession and smuggling of venom from the Indian cobra (Naja naja) [30]. Populations of Naja naja are in danger of extinction in part because of these illegal practices, and prevention and prosecution of these crimes is of utmost importance to ensure the protection of this species [31].

Despite its theoretical advantages, DNA purification from dried snake venom has produced low yields and inconsistent amplification, and it has yet to be investigated in a wide range of venomous snake species [11,30]. Further, the large quantity of starting material previously required for these analyses (100 mg) made venom DNA barcoding from rare species or species with low venom yields problematic [11,32]. Here we outline a modified method for the extraction of DNA from venom and the consistent amplification of mitochondrial markers from small amounts of venom (~1 mg) collected from both front-fanged (families Viperidae, Elapidae and Atractaspidae) and rear-fanged (families Colubridae, Dipsadidae and others) species. This modified protocol has broad applications for genetic analysis and DNA-barcoding efforts across venomous snake species, and it allows for routine barcoding of commercial samples for unequivocal identification of species.

Materials & methods
Venom collection & storage
All vipers (except Vipera ammodytes and Daboia russelli) and all colubrids analyzed were housed individually in the University of Northern Colorado (CO, USA) Animal Resource Facility (IACUC #
Venom was extracted from viperids and colubrids as previously described [33,34]. All venoms extracted at University of Northern Colorado were centrifuged at 9500 x g for 5 mins, lyophilized and stored at −20°C. V. ammodytes venom was obtained from a commercial source (Jabria B.V., Hierden, Holland) and had been air-dried. Naja naja, N. annulifera, N. kaouthia, Oxyuranus scutellatus, Dendroaspis polylepis, Ophiophagus hannah and D. russelii lyophilized venoms were donated by the Kentucky Reptile Zoo (Slade, KY). Naja atra venom (lyophilized) was acquired from Ventoxin Laboratories Incorporated (MD, USA), and Pseudochis porphyriasus venom (lyophilized) was donated by Venom Supplies Pty. Ltd (Tanunda, South Australia). Dried venoms were stored at −20°C until used.

**DNA extraction & purification**

DNA extraction was based on a modified version of the QIAGEN DNeasy® kit cultured cells protocol (see Supplemental File 1). DNA yields were quantified using a Qubit™ fluorometer (Qubit™ DNA DS HS Assay Kit; Thermo Scientific, MA, USA) and successful DNA extraction was confirmed via PCR of mitochondrial markers.

**Mitochondrial & nuclear marker amplification**

Two mtDNA loci, cyt b [21,35] and ND4 [36], and one nuclear locus (C-mos) [37] were amplified from each sample (Table 1). For cyt b, primers from Kocher et al. [35] were used to amplify all elapid samples, and primers from Pook et al. [21] were used to amplify all other samples.

PCR reactions were carried out on an Eppendorf Mastercycler Gradient S (Hamburg, Germany) in total volumes of 21 µl, containing 2–5 µl extracted DNA, 1 µl (10 mM) of each primer (forward and reverse), 4 µl Promega 5x Go Flexi buffer, 1 µl 25 mM dNTPs, 1 µl 25 mM MgCl₂, 0.3 µl Taq polymerase (Promega Go Flexi), and nuclease-free ddH₂O to bring the reaction volume to 21 µl. A standard PCR protocol was used for amplification: 35 cycles of 1 min at 95°C, a 1 min ramp from 50°C to 65°C (0.3°C/second), and 4 min at 65°C. 1% agarose gels run at 125V for ∼30 mins were used to confirm the presence of amplified PCR products.

PCR amplification products were cleaned using ExoSAP purification: 5 µl PCR product, 0.5 µl exonuclease I, and 1 µl FastAP thermosensitive alkaline phosphatase. These reactions were run on an Eppendorf Mastercycler Gradient S thermal cycler at 37°C for 15 min, followed by another 15 min at 85°C. Purified products were then cycle-sequenced in 10 µl reactions of 2 µl Big Dye 5x dilution buffer, 0.33 µl Big Dye III terminator, 0.8 µl cleaned PCR product, 0.5 µl of forward or reverse primer, and ddH₂O to bring to final volume. Cycle sequencing reactions were run on an Eppendorf Mastercycler Gradient S thermal cycler: 1 min at 96°C; and 30 cycles of 96°C for 15 s, 50°C for 20 s, 60°C for 4 min. Samples were sent to Arizona State University's DNA Laboratory for sequencing. Sequenced sample product reads were imported into GENEIOUS (version 8.1.4; www.geneious.com) [38] and analyzed for nucleotide polymorphisms. In the case of failed amplification or failed sequencing, reaction parameters such as total amount of DNA and amount of polymerase were adjusted.

**Genetic analysis**

Sequences from both mitochondrial regions were concatenated to construct a single mitochondrial data set. Phylogenetic trees were constructed independently for the mitochondrial and nuclear data using
MRBAYES [39] as run in GENEIOUS with a General Time Reversible (GTR) substitution model with a chain length of five million generations, sampling every 1000 generations for four heated chains. GENEIOUS was run with default parameters for the rest of the analysis options, including gamma rate variation with 4 gamma categories, heated chain temperature of 0.2, and a subsampling frequency of 1000. A burn-in length of 1.5 million generations was used for each analysis.

Results & discussion

DNA extraction & purification

Successful DNA extraction from venom generally produced a low yield of DNA (Qubit™ average: 267.1 ng/ml; range 9.3–3520 ng/ml; Table 2) that was not visible on a 1% agarose gel. However, this low DNA yield did not affect the success of mitochondrial region amplification or sequencing (Table 2). ND4 and cyt b regions successfully amplified in all species, with amplicon sequence lengths comparable to published sequences (Tables 1 & 2).[21,30,36]. C-mos was successfully amplified in 64% of samples, with amplicon lengths equivalent to previous studies (Table 2)[37].

DNA exists in venom primarily as a result of cellular content deposition in the lumen of the venom gland following cell death. This cell debris may be ejected along with venom during prey envenomation or during venom extraction.[40,41]. Degradative venom enzymes (nucleases, exonucleases, etc.) exist in the venom, likely in an inactivated state[41], but it is currently unknown if degradation of venom DNA occurs due to their actions. However, in addition to DNA, intact mRNA has been purified from venom[42], indicating overall stability of nucleic acids in the venom gland and in dried venom despite possible degradation via DNases and RNases found in venom.

The presence of high-quality DNA allowed species identification from a single venom sample[30] and phylogeographic analysis of mitochondrial regions[11]. DNA barcoding of venoms has been previously described using large amounts of venom (100–200 mg of lyophilized venom), but even with this amount of starting material, extractions produced very low DNA yields (<1 ng/μl;[11]). Although the exact mass of venom that can be extracted depends on several factors[13,34,41,43,44], for smaller species or individuals it is not possible to collect this amount from a single snake during a single extraction. Therefore, the use of over 100 mg of starting material for genetic analysis is likely not feasible for most research efforts, but the ability to utilize 100-fold less venom (the current method) does make analysis possible.

Pook and McEwing[11] did not specify the instrument used to quantify DNA yields; however, due to potential sensitivity issues with Nanodrop[45], we utilized the highly sensitive Qubit™ fluorometer for DNA quantification. Although our DNA yield values appear significantly lower than those of Pook and McEwing[11], it is likely that Qubit™ yields are more reflective of true DNA concentration than spectrophotometric methods. Further, success at both mitochondrial regions for all samples indicates that amplification using our method can be successful even from samples with very low DNA yields (9.3 ng/ml).

Mitochondrial & nuclear marker amplification

Both mitochondrial regions amplified successfully in all species. In general, high-yield samples amplified successfully without further modifications to the reaction mixture. In samples with low levels of DNA template, the amount of DNA in the reaction mixture was increased if PCR failed initially. C-mos amplification was inconsistent, with a success rate of 64%, and amplification success loosely correlated with DNA yield, with low-yield samples failing repeatedly despite increasing DNA concentration in the reaction mixture.

Low DNA yield did not appear to affect success of mitochondrial marker amplification. For example, low-yield Crotalus samples (C. o. helleri and C. s. scutulatus) amplified successfully for both mitochondrial regions during the first round of PCR. However, the sample with the overall lowest DNA yield (B. gabonica) amplified inconsistently (requiring additional modifications, including increasing the amount of template DNA), and the sample with the highest yield (P. baroni) successfully amplified after only one round of PCR with no modifications. Interestingly, failure of amplification at the c-mos region was inconsistent, regardless of taxonomic family or DNA yield, although in general the samples with higher yields amplified more successfully. The use of c-mos primers in all squamate reptiles[46–48] indicates that the failure of c-mos to amplify consistently may instead be due to poor nuclear DNA quality or concentration, as opposed to primer nonspecificity. In fact, cells undergoing apoptosis have been observed in venom glands[40], so nuclear DNA may degrade differently than mitochondrial DNA, which typically amplified successfully and consistently.

Although c-mos amplification was successful in most larger species (e.g. D. russeli, N. atra), mitochondrial regions will likely be more reliable markers throughout venomous snake taxa, regardless of snake size or venom yield. The high number of mitochondria in venom gland cells[41], leading to higher amounts of...
mitochondrial DNA available for amplification versus genomic DNA, may also account for the more consistent success of mitochondrial marker amplification. Both c-mos amplification inconsistency and microsatellite amplification failure (C Smith, Unpublished Data) indicate that nuclear DNA is less abundant and more degraded than mitochondrial DNA, making nuclear markers less reliable for venom barcoding.

Genetic analysis
Phylogenetic trees built from concatenated mitochondrial sequences (Figure 2A) and the c-mos nuclear region (Figure 2B) concurred with previously established taxonomic relationships, with vipers being the most basal group and colubrids and elapids being more derived [21,35,49]. Successful tree building from venom-derived DNA indicates that sequences obtained from venom were of high enough quality to produce useful phylogenetic information.

Conclusions
The current study describes a simple DNA extraction protocol that requires a small amount of starting material (1 mg venom) and yields sufficient amounts of DNA for consistent PCR amplification of mitochondrial and, in some cases, nuclear genes. This method also showed higher success rates of amplification at both nuclear and mitochondrial markers, in more species of different families, than previous published studies [21,30]. Because the amount of venom acquired from one snake is often limited, and small species may produce only several milligram of venom, this method has practical applications for DNA barcoding efforts on an individual sample scale. Additionally, venoms from rear-fanged snakes, which in general are poorly characterized and of low yield [50], can now also be barcoded, facilitating positive identifications within this speciose group of advanced snakes. Ultimately, this efficient and broadly applicable method for the barcoding of snake venoms has utility in the genetic cataloging of snakes and likely other venom-producing taxa as well, and it has the potential to produce an abundance of genetic data from very small amounts of renewable starting material.

Acknowledgments
The authors thank Kentucky Reptile Zoo (Slade, KY, USA), Ventoxin Laboratories Incorporated (Frederick, MD, USA) and Venom Supplies Pty. Ltd (Tanunda, South Australia) for providing some of the venoms used in this study, and Brenda Jones for her assistance with DNA extractions.

Data accessibility

Financial & competing interests disclosure
Funding for this research was provided by the University of Northern Colorado School of Biological Sciences. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial benefit to the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Author contributions
Research was designed by all authors, and performed by CS using reagents and samples provided by SM and MM. Analysis and interpretation of data was performed by CS. Manuscript drafts were written by CS, with revisions from SM and MM.

Supplementary data
To view the supplementary data that accompany this paper please visit the journal website at www.future-science.com/doi/suppl/10.2144/btn-2018-0096

Open access
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

References


First draft submitted: 2 July 2018; Accepted for publication: 29 August 2018

Address correspondence to: Stephen P Mackessy, School of Biological Sciences, University of Northern Colorado, 501 20th Street, Greeley, CO 80639-0017, USA; Fax: +1 970 351 2335; E-mail: stephen.mackessy@uncou.edu

To purchase reprints of this article contact: s.cavana@future-science.com