GigaScience

# A single mini-barcode test to screen for Australian mammalian predators from environmental samples

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | GIGA-D-17-00051R1 |
| **Full Title:** | A single mini-barcode test to screen for Australian mammalian predators from environmental samples |
| **Article Type:** | Research |
| **Funding Information:** | Invasive Animals Cooperative Research Centre (Project 1.L.21) — Miss Elodie Modave |
| **Abstract:** | Background: Identification of species from trace samples is now possible through the comparison of diagnostic DNA fragments against reference DNA sequence databases. DNA detection of animals from non-invasive samples, such as predator faeces (scats) that contain traces of DNA from their species of origin, has proved to be a valuable tool for management of elusive wildlife. However, application of this approach can be limited by the availability of appropriate genetic markers. Scat DNA is often degraded, meaning that longer DNA sequences, including standard DNA barcoding markers, are difficult to recover. Instead, targeted short diagnostic markers are required to serve as diagnostic mini-barcodes. The mitochondrial genome is a useful source of such trace DNA markers, because it provides good resolution at species level and occurs in high copy numbers per cell.<br>Results: We developed a mini-barcode, based on a short (178 bp) fragment of the conserved 12S rRNA mitochondrial gene sequence, with the goal of discriminating amongst the scats of large mammalian predators of Australia. We tested the sensitivity and specificity of our primers and can accurately detect and discriminate amongst quolls, cats, dogs, foxes and devils from trace DNA samples.<br>Conclusions: Our approach provides a cost effective, time efficient and non-invasive tool that enables identification of all eight medium-large mammal predators in Australia, including native and introduced species, using a single test. With modification, this approach is likely to be of broad applicability elsewhere. |
| **Corresponding Author:** | Elodie Modave<br>University of Canberra<br>AUSTRALIA |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | University of Canberra |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | Elodie Modave |
| **First Author Secondary Information:** | |
| **Order of Authors:** | Elodie Modave |
| | Anna Jayne MacDonald, BSc, PhD |
| | Stephen Donald Sarre, BAppSc, MAppSc, PhD |
| **Order of Authors Secondary Information:** | |
| **Response to Reviewers:** | Responses to the reviewers<br>all the changes in the text are highlighted in yellow and the line numbers where the changes occurred are noted in the “responses” part of the table below<br><br>reviewer 1: Andrea Galimberti<br><br>- ROWS 111-114: Use "region" instead of "sequence" and the sentence is quite |

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

redundant and somewhat circular. I suggest to rephrase it.

- yes, I agree. I made the sentence shorter. Lines 118-120

- TABLE 1: It is unclear, which criteria were used to adopt the two threshold values. Maybe the authors can calculate a sort of optimum threshold due to minimum cumulative error rate (see Ferri et al. 2009 DOI: 10.1186/1742-9994-6-1or Galimberti et al. 2012 http://dx.doi.org/10.1371/journal.pone.0040122)

- That is what we did, it is based on the thresholds with the lowest cumulative error. I modified the legend to make it clearer.
Lines 193-196
This can be found in the method section, in “Bioinformatic evaluation of the mini-barcode” (lines 454-456) and it is calculated in the additional file 6 for the R code in the section “MODEL” “# Identify the optimal genetic distance threshold for the raw model for "FULL"-“UNIQUE” ”

- ROWS 260-262: This is an important point. What contingency plan the authors propose to overcome this limit? It is unclear from the text.

- I have added “In practice, any such sequences cannot be used to identify the predator with confidence and therefore must be excluded from analysis.”
Lines 293-295

- I also think that the recent review by Galimberti and colleagues (DOI: http://dx.doi.org/10.4404/hystrix-26.1-11347) concerning DNA barcoding on mammalian taxa should be cited.

- thank you for this paper. I included a citation in the “conservation implications” part of the “Discussion” section. I indeed developed this mini-barcode for a management and monitoring purpose thus for a broader application than the “simple” identification purpose. Line 347

reviewer 2: Stephane Boyer

- It is interesting to see that a more relaxed genetic distance threshold may be more appropriate (line 201). The authors used the default 1% threshold in the functions bestCloseMatch and threshID. They seem to base this decision on the graphical representation of threshID (code below).

```
>barplot(t(threshfullMat) [4:5,],
>names.arg=paste((threshfullMat[,1]*100), "%"))
```

The visual reading of this barplot gives some indication of how many false positives/negatives the user may have to tolerate. However, this is somewhat a crude measure of the optimal threshold. A better option is to use the localMinima function in SPIDER, which calculates the most appropriate threshold to use for a given dataset based on pairwise distances only. When running this function on the full dataset (see code below), I obtained a threshold of 0.0335 which seems more appropriate for the data. The authors may want to re-visit their analysis based on that threshold (instead of 1%).

```
>#local minima calculation of optimal species delineation threshold
>Thresh <- localMinima(fullDist)            #Compute the localMinima function
>#Results: 0.0335 ; 0.195
>plot(Thresh, main="localMinima 12S FULL")
```

If the authors choose to use the localMinima function, the optimal threshold should be calculated using the Unique dataset only. As it is not possible to calculate an accurate threshold with this function using singletons only.

- The reviewer is correct, I chose 1% for the FULL database and 4% for the UNIQUE database based on the code lowest cumulative error.
I have not used localMinima, and I thank the reviewer for making us aware of this option. However, this does not seem to provide a sensible output for the unique database in this instance.
As suggested, I have used a threshold of 3.5% (rounding up the localMinima result of 0.335) for the full database. I have incorporated this into the analysis by comparing

results using thresholds of 1% and 3.5%. For example, using best close match, The higher threshold results in a greater number of correct identifications, but also a greater number of incorrect identification. In contrast a 1% threshold has a higher number of "no ID" results. I have amended the discussion to note that the most appropriate threshold will depend on the management context, and the relative importance of false positive identifications / unidentified samples.
The results for the unique database using localMinima are more problematic.
Using

```
uniThresh <- localMinima(uniqueDist)
uniThresh$localMinima[1] *100
plot(uniThresh)
```

the threshold identified is 19%, which seems extremely high in this context. While this threshold does produce perfect results (all samples correctly identified with best close match / threshID) using our unique dataset, my concern is that sequences from taxa that are not well represented in our database will be at a much greater risk of misidentification with such a relaxed threshold. Hopefully as more reference sequences become available from a wider range of Australian mammals it will be possible to improve this analysis. However, in the meantime I would argue that in most management contexts it would be better for a sample to be ambiguously identified, or to have "no ID" than to be incorrectly identified. For example, working with the full database, I see the following results using best close match with thresholds of 1% and 3.5%

```
> table(bestCloseMatch(fullDist, Sppfull, thresh = 0.01))
  correct incorrect     no id
      147         3        24
> table(bestCloseMatch(fullDist, Sppfull, thresh = 0.035))
  correct incorrect     no id
      152         6        16
```

And using threshID with the same thresholds I get:

```
> table(threshID(fullDist, Sppfull, thresh = 0.01))
ambiguous   correct incorrect     no id
        5       142         3        24
> table(threshID(fullDist, Sppfull, thresh = 0.035))
ambiguous   correct incorrect     no id
       12       141         5        16
```

To improve consistency between the two sets of results, I have also amended the text so that analyses with the unique database also use a threshold of 3.5%. This has the same cumulative error as a threshold of 4% (which is what was previously used) and the results are not affected by this change.
text added
Methods: lines 456 + 461-462
Results: lines 183-186, 193-196, 201-202, 216-231
Discussion: lines 319-323
Table 1, Additional file 6, Additional file 7

- I can only commend the authors for providing the annotated R code. The main code works well and is easy to follow. The very last line of code seems incomplete. I think it misses a closing bracket at the very end and another line to query a sequence (as written below)

```
>}
>withinF[[1141]]
```

- The reviewer is correct that the code should end this way. However, in my version of the file this text is not missing.
I have uploaded the file again to make sure that there are no errors.

- I was a little confused with the code for sliding window analysis. I don't understand why the window width was set on 20 bp and why only this particular length was investigated. The authors seem to have used the sliding window analysis to determine the position of potential primers, rather than the position of a suitable mini-barcode region (which was the original purpose of sliding window). If that is the case, then I suppose suitable 'primer windows' must be highly conserved, but what were the other criterion used to select them? It reads as follow on line 343: "…regions up to 200 bp in length, incorporating two primer sites (each of 20 bp in length) that were well-

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

conserved across all taxa but which flanked a region of 100-200 bp that displayed high levels of interspecific variation" What is the threshold for 'well conserved'? What is considered 'high levels of interspecific variation'? Are these based on values obtained from the sliding window analysis?
I would have expected that a range of length, for example from 50 bp up to 200 bp, would have been investigated with the aim of determining the shortest possible mini-barcode region. For example, I ran a sliding window analysis using a width of 150 bp (see code below modified from the authors').

```
>a12SWin <- slidingWindow(a12Sref, width = 150, interval=1)
>length(a12SWin)
>a12SWin[[1]]
>a12SAna <- slideAnalyses(a12Sref, Sppa12S, width = 150, interval =
>"codons", distMeasures = TRUE, treeMeasures = TRUE)
>str(a12SAna)
>plot(a12SAna)
```

Useful variables provided by the sliding window function includes the 'proportion of zero non-conspecific K2P distances'. When this value is 0, the window has enough identification power to tell all species apart. All 150 bp windows starting on base ~90 to ~240 are good picks in this regard. So I do believe the chosen region is probably a good one. But it is unclear why the window starting on position 160 was deemed the best window by the authors

- I agree that this section was unclear and I have amended the manuscript to include more detail. I thank the reviewer for these comments as these have helped to improve my explanation and interpretation.
I did indeed use a wider range of window sizes. I first used larger window sizes (100-175bp) to identify potential mini barcode regions. I then used the shorter window sizes (20-30 bp) to identify conserved sites suitable for primer development within the region of the candidate mini-barcode. The combination of both of these factors (a highly diagnostic sequence and conserved primer sites) are crucial for effective barcode design and adjusting the sliding windows analysis seemed like a good way to identify primer sites.
Using larger windows, I identified regions that may have been good candidate mini-barcodes, except that it was not clear that suitable primer sites were present. By restricting the window size, I was able to clearly identify highly conserved primers as well as the diagnostic mini-barcode regions.
While the broader region (90-240bp) identified by the reviewer using 150bp windows could certainly serve as a mini-barcode, my choice of starting position was driven by the identification of a suitable forward primer sequence, and also with the final length of the amplicon in mind given the location of a suitable reverse primer. While the region from 90 bp is identified as a potential mini-barcode using a window size of 150, I also found a good region between bases ~160-~380. Considering just the smaller window size (for primer design), the region around base 160 was a suitable primer site.
I have updated the text to clarify this and to better explain the approach and criteria.
Methods: lines 375-387, 425-436
Results: lines 163-170
I have also amended the figures to reflect this (Figure 2) and have updated the supplementary R code (additional file 3).

- Now, it is important to note that the actual values on the x-axis on the plots (e.g. Figure 2) are the positions of the first nucleotide of the window. As such, the box drawn on Figure 2 and presented as the 'best candidate site for a short diagnostic amplicon' is slightly misleading because each dot on that graph represents one window. There is also an issue with the positioning of that box as it is clearly not located between positions 160 and 380 as suggested in the legend of Figure 2.

- Indeed, the boxed area was a bit to the right on figure 2, it is fixed now.
Also, I changed the legend in figure 2 to precise that a dot was the window and the x-axis represents the first base of a window.

- Last small comment about the code: I found that on my version of R, there is an issue with object names that start with a number (e.g. 12Sref). Just placing a letter as the first character in the name solves the issue.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

- Additional file 3: This problem is now fixed with all names starting with 12S preceded by “db” (for database)

- Lines 41-55 There is no flow between these sentences. They need to be better linked together. As it stands it is rather laborious to read.

- I changed the text so the sentences flow better
Lines 44-55

- Line 77. it is not clear what you mean by 'barcode tests'

- changed to improve clarity of meaning
Line 77-81

- Lines 113-114 need rephrasing to avoid repetition

- changed
Lines 116-120

- lines 114-120. This paragraph follows few sentences where the authors described their study and their taxa. I think it needs to be more clear that here they are back to general statements. Alternatively, these general statements could be placed before the sentence starting with 'Our goal was…'

- changed. I put the general statements (the two common limiting factors) before summarising the findings in this study (our goal was…)
Lines 109-120

- Line 136. I think it would be useful to include citation [2] here as it is the one describing the sliding window analysis in details.

- changed
Line 137

- Line 144. To create the UNIQUE database, I am guessing that the first step was to remove the singletons and THEN to only keep one sequence per haplotype. It would make sense to write these two steps in the correct order.

- yes, that makes sense. Changed
Lines 145-146

- I was also surprised to see that you had singletons in the FULL dataset, given that line 132-133, it is stated that: "Sequences were obtained from GenBank, with additional targeted sequencing conducted for species under-represented in GenBank." If there were indeed singletons and those species were eliminated, it would be useful to list which species they were

- The singletons referred to non-target species. I focused the additional sequencing on specific target taxa most relevant to wildlife surveys in Australia, in particular the quolls which are poorly represented in sequencing databases. Line 133: to specify that I added sequences from the target animals
I amended Additional file 7 to note all singleton species

- Line 205. Yes, but a 5% distance threshold would have caused much ambiguity for the identification of the other sequences. Any chances one of the sequences for Dasycercus cristicauda was obtained on Genbank and could be either mis-identifiation or a different (cryptic) species?

- the two Dasycercus sequences are from the same team of scientists. The origin of only one sequence (AF009889) was mentioned (the Tanami desert in Northern Territory). As for the second, they don’t know the origin. So it might be an ID error. I put a note in the text
Lines 216-219
Same was true for a western quoll sequence, lines 183-186

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

- Line 208. rather than 'a wide range of Australian mammals', please provide the number of species

- changed to note that 40 species were included, but also to emphasise that these represent a wide taxonomic range (ie not just 40 species from a single order).
Line 234-239

- Line 201. Add "and possibly beyond" to the end of the sentence or something similar to acknowledge that you also successfully used the primers with non-mammalian vertebrates. Alternatively, remove reptile amphibian and bird from the previous sentence, and write a new sentence at the end of the paragraph, stating why the primer was tested on those non-mammalian specimens.

- Changed to “This demonstrates the broad applicability of the primers across the mammalian taxa and their potential applicability to other vertebrate classes”
Lines 234-239

- Table 2. The title for this table could be improved. It does not give much information about what the numbers are. To understand this, the reader need to go to the legend and then guess what 'CT' means or go all the way to the list of abbreviations. Depending on where this list sits in the paper, I would advise to state what CT means in the legend of Table 2.

- Changed. Table 2: Add information in the title, and in the legend: described what CT was and how it is calculated.
Line 253
Lines 254-259

- Line 236. I would replace 'the known predator' by 'known predators'

- changed
Lines 261-262

- Line 254-257. Here the authors highlight how their study brings new knowledge in the subject of DNA-based species detection. This is crucial but not extremely clear. Maybe these sentences need to be restricted to 'studies aiming at identifying predators from scat samples'.

- changed to “Previous studies, based on species identification from scats or hairs, have applied barcoding methods to detect individual species across multiple time points (examples in (Fernández et al. 2006), (McKelvey et al. 2006)). Here we have shown that it is also possible to identify multiple species from a single DNA test, using a straightforward PCR and Sanger sequencing approach”
Lines 279-282

- Line 239. 92% amplification success is quite good. It would have been interesting to compare this to what can be obtained with primers targeting longer DNA fragments. I understand this was not the aim of this particular paper, but in a sense the authors went into all the trouble of designing mini-barcodes because 'regular (longer) barcodes' don't work. It would be good to put this 92% success rate into perspective with the success rate of longer barcodes if there was any such data in the literature. It is eluded to on line 277, but the actual numbers are not provided.

- in [41]: 79% of sequences were amplified using a 134 bp fragment, and in [56]: <70% using regions from 243 bp to 708 bp (different regions for different taxa)
Lines 304-307

- Line 273. I would replace 'by' with 'in'

- changed
Line 300

- Lines 277-282. I would be careful not to inflate the implications of the paper. The

| | |
|---|---|
| | 'approach' used is simply DNA barcoding, the benefits of which have been widely demonstrated elsewhere. The real novelty lies in the primers and the mini-barcode designed for Australian mammals, which does make a very useful tool for managers and scientists. So rather than the 'approach' I would highlight the primers or the mini-barcode here<br><br>- changed to "Using our mini-barcode, DNA can be screened for the presence of multiple Australian predator species in a single and inexpensive test, without the need to develop and apply a set of species-specific primers for each predator of interest. We provide a non-invasive instrument with potential utility for scientists or managers working with endangered or invasive Australian predators, but a similar approach could be used to target predator assemblages in other regions."<br>so more focussed on Australia and the development of the mini-barcode than on the barcoding itself<br>lines 307-312<br><br>- Line 278. Replace 'screen' by 'screened'<br><br>- changed<br>Line 307<br><br>- Line 299. A reference at the end of this sentence would be useful<br><br>- yes, I added 2<br>Line 333<br><br>- Line 329-331. Very interesting potential application<br><br>- yes indeed<br><br>- Line 514. Keith Crandall was editor, not co-author, on that paper. The citation needs to be modified accordingly<br><br>- Changed<br>Line 562<br><br><br>references:<br>Fernández N, Delibes M, Palomares F (2006) Landscape evaluation in conservation: molecular sampling and habitat modeling for the Iberian lynx. Ecol Appl 16:1037–1049.<br>McKelvey KS, Kienast JVON, Aubry KB, et al (2006) DNA analysis of hair and scat collected along snow tracks to document the presence of Canada lynx. Wildl Soc Bull 34:451–455. |
| **Additional Information:** | |
| **Question** | **Response** |
| Are you submitting this manuscript to a special series or article collection? | No |
| **Experimental design and statistics**<br><br><br>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.<br><br><br>Have you included all the information requested in your manuscript? | Yes |

| | |
|---|---|
| **Resources**<br><br>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.<br><br>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist? | Yes |
| **Availability of data and materials**<br><br>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.<br><br>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist? | Yes |

*Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation*

1 **Title:**

2 A single mini-barcode test to screen for Australian mammalian predators from
3 environmental samples

4 **Authors:**

5 Elodie Modave[1], Anna J MacDonald[1] and Stephen D Sarre[1]

6 **Address:**

7 [1] Institute for Applied Ecology, University of Canberra, ACT, 2601, Canberra, Australia

8 **Email addresses:**

9 Elodie Modave: Elodie.modave@canberra.edu.au

10 Anna J MacDonald: Anna.MacDonald@canberra.edu.au

11 Stephen D Sarre: Stephen.Sarre@canberra.edu.au

12 **Corresponding author:**

13 Elodie Modave, Institute for Applied Ecology, University of Canberra, ACT, 2601, Australia.

14 Telephone: +61 2 6201 2267. Fax: +61 2 6201 5305. Email: elodie.modave@canberra.edu.au

15 **Running title:**

16 Mini-barcode for species identification

17 **ABSTRACT**

18 Background: Identification of species from trace samples is now possible through the
19 comparison of diagnostic DNA fragments against reference DNA sequence databases. DNA
20 detection of animals from non-invasive samples, such as predator faeces (scats) that contain

traces of DNA from their species of origin, has proved to be a valuable tool for management of elusive wildlife. However, application of this approach can be limited by the availability of appropriate genetic markers. Scat DNA is often degraded, meaning that longer DNA sequences, including standard DNA barcoding markers, are difficult to recover. Instead, targeted short diagnostic markers are required to serve as diagnostic mini-barcodes. The mitochondrial genome is a useful source of such trace DNA markers, because it provides good resolution at species level and occurs in high copy numbers per cell.

Results: We developed a mini-barcode, based on a short (178 bp) fragment of the conserved 12S rRNA mitochondrial gene sequence, with the goal of discriminating amongst the scats of large mammalian predators of Australia. We tested the sensitivity and specificity of our primers and can accurately detect and discriminate amongst quolls, cats, dogs, foxes and devils from trace DNA samples.

Conclusions: Our approach provides a cost effective, time efficient and non-invasive tool that enables identification of all eight medium-large mammal predators in Australia, including native and introduced species, using a single test. With modification, this approach is likely to be of broad applicability elsewhere.

**Keywords:**

12S rRNA; Dasyurus; DNA barcoding; DNA detection; marsupial; monitoring

## BACKGROUND

The looming biodiversity crisis, referred to by some as the Sixth Mass Extinction [1], has made the conservation of wildlife a rapidly growing concern. There is an urgent need to document the distribution of biodiversity as the foundation for identifying effective solutions

43 to wildlife management issues. The rapid and reliable identification of species at local and
44 regional scales can provide the first step towards determining the distribution of biodiversity
45 in the landscape and changes that might be occurring in that distribution.

46 Advances in genetics and genomics have revolutionized many areas of biology and in
47 particular, the identification of wildlife from trace and environmental samples (e.g. water,
48 soil and faeces, or scats) is now possible through DNA barcoding [2], [3], [4], [5], where the
49 identity of an unknown sample is established by comparing DNA sequences obtained from
50 that sample to an appropriate reference sequence database. The application of DNA
51 barcoding for the identification of species from such environmental DNA (eDNA) samples is
52 useful, particularly when the target species is rare, elusive, difficult to trap or observe
53 without direct interference with live animals, or where morphological identification is
54 problematic [6], [7], [8], [9], [10]. It also makes possible the identification of diet from scats
55 where morphological determinations are likely to be unsuitable for many elements of the
56 diet [11], [12], [13], [14], [15]. Consequently, eDNA analysis from environmental samples
57 collected across a broad spatial and temporal distribution has great potential for enhancing
58 biodiversity management, but is yet to be widely implemented [16], [17].

59 The DNA associated with environmental samples tends to be of low quantity or
60 quality and can be degraded. To ensure that markers for eDNA detection are specific and
61 sensitive, target sequences, also known as mini-barcodes, should be short (i.e. 100-200 base
62 pairs (bp); [18], [19], [20], [2]) and yet have high discriminatory power [21], [22], [23], [24].
63 Marker selection therefore needs to account for the range of species likely to be
64 encountered, as well as discriminating among potential sister taxa. Mitochondrial DNA genes
65 (mtDNA) are usually targeted because they occur in multiple copies in each cell and are

66 therefore more common in trace samples than nuclear sequences, because they can give
67 good resolution of identification at species level, and because their genome is circular, which
68 helps preserving the DNA in some instances. In regions where little is known of the genetic
69 characteristics of the faunal assemblage, identifying the most appropriate DNA sequences to
70 target the fauna present to achieve acceptable levels of accuracy is a challenging exercise
71 and requires a reference database that is sufficiently comprehensive to ensure accurate
72 species assignment [25]. In short, we need DNA barcoding markers that are appropriate to
73 the question being addressed, the ecosystem considered and the taxonomic group studied.
74 Most importantly, if DNA detection is going to be of practical benefit, we need to maximise
75 its effectiveness by developing mini-barcodes that target as many taxa as possible, thus
76 minimising the number of tests that need to be applied. Most DNA barcode studies so far
77 implemented for detection of specific species from terrestrial systems have targeted single
78 species (examples in [7], [9], [26], [27]) to avoid the ambiguity that might arise by attempting
79 to simultaneously identify multiple closely related taxa. Here, we tackle this problem using
80 all extant medium-large Australian mammalian predators as a case study.

81 Australia has a unique assemblage of medium-large mammalian predators, including
82 a suite of marsupials of Gondwanan heritage intermixed with relatively recently arrived
83 eutherian mammals introduced by humans [28], [29]. Here, we develop a DNA mini-barcode
84 to discriminate amongst these key predators, with the goal of species identification using
85 eDNA extracted from scats. We targeted the top native marsupial predators that are likely to
86 produce large easily visible scats including: six species of quoll (four Australian and two New-
87 Guinean; *Dasyurus maculatus*, *D. viverrinus*, *D. geoffroii*, *D. hallucatus*, *D. albopunctatus* and
88 *D. spartacus*), the Tasmanian devil (*Sarcophilus harrisii*), and the extinct thylacine (*Thylacinus*
89 *cynocephalus*), as well as key eutherian mammal predators: the native dingo (*Canis lupus*

*dingo*), and the introduced domestic dog (*Canis lupus familiaris*), red fox, (*Vulpes vulpes),* and domestic cat (*Felis catus*) that are now feral in much of the country. Most of the native marsupial predators have been in decline since, or even before, European settlement in 1788 [30]. Tasmanian devils and the thylacine became extinct on the Australian mainland within the last 3000 to 4000 years [31], but still existed on the island of Tasmania at the time of European settlement. The thylacine has subsequently been hunted to extinction [32], [33] while devil populations have decreased dramatically since the 1990s following the emergence of Devil Facial Tumour Disease [34], [35], [36]. Several species of quoll, together with the dingo, have declined in distribution and abundance on the Australian mainland since European settlement from multiple causes that probably include habitat destruction, hunting, predation by cats and foxes, the spread of cane toads [37], [38], [39] and in the case of dingos, hybridisation with domestic dogs. Although declining or extinct on the mainland, substantial populations of the Tasmanian devil, the spotted-tailed quoll (*D. maculatus)* and the eastern quoll (*D. viverrinus*) remain on the island of Tasmania where they have important ecological roles [40]. However, recent evidence of foxes in Tasmania [41] and potential competition with feral cats [42], [43] compound the issue, and have stimulated an urgent need to understand threats to native predator populations and enable effective management.

Two factors generally limit the application of a DNA barcoding approach. First, short diagnostic sequences that encompass the range of species to be targeted are difficult to find and are likely to be specific to a particular faunal assemblage. Second, the full suite of potential target organisms tends to be poorly known in most natural systems, and reference DNA sequences are not available for many wildlife species, necessitating the development of reference libraries to guide marker selection and interpretation of results. Our goal was to

develop a mini-barcode that can identify all medium to large mammal predators in Australia in a single analysis, including quolls, to species level. This has been difficult to achieve using existing genetic markers because of the high levels of sequence conservation observed between quoll species. We compiled a reference tissue collection and identified a mini-barcode based on the conserved 12S rRNA mitochondrial region that discriminated among taxa with minimal variation within species [44], [45]. We evaluate the specificity and sensitivity of this mini-barcode using the framework outlined in [25] and [46]. By targeting all extant medium to large carnivores in Australia, we aim to produce a mini-barcode that can be applied broadly within continental Australia as well as Tasmania. We demonstrate that despite close homology among some taxa, it is possible to design and implement eDNA markers with high discriminatory power for key continental terrestrial fauna incorporating both marsupials and eutherian mammals. Our approach can be implemented in other parts of the world by targeting appropriate fauna assemblage in the development of the mini-barcode.

## DATA DESCRIPTION

We identified the 12S rRNA gene as a target for development of a mini-barcode marker. We developed a reference DNA database for this gene, including 174 sequences from 24 genera and 41 mammal species. Sequences were obtained from GenBank, with additional targeted sequencing conducted for target species under-represented in GenBank. Sequences were aligned, trimmed to 901 bp, and are provided here in FASTA format (Additional file 1) with additional information on sample and sequence origins in .csv format (Additional file 2).

136 We used the R package SPIDER [47] to conduct a sliding window analysis [2] to identify a
137 short diagnostic region of the 12S rRNA gene suitable for use as a mini-barcode marker. R
138 code for this analysis is provided in text format (Additional file 3).

139 Following design of the *AusPreda_12S* primers, we conducted bioinformatic and
140 laboratory evaluations of the sensitivity and specificity of the mini-barcode. We created two
141 modified versions of our reference 12S rRNA database, trimmed to include only the 178 bp
142 flanked by the mini-barcode *AusPreda_12S* primers. The "FULL" database included all 174
143 sequences from the original database, while the "UNIQUE" database included a subset of 44
144 sequences, where singleton species (species represented by only one haplotype) were
145 removed, and where each remaining haplotype was represented by only a single sequence.
146 These two databases are provided here in FASTA format (Additional files 4 and 5). We used
147 the R package SPIDER to conduct genetic distance based evaluations of the *AusPreda_12S*
148 primers, to identify the risks of incorrect or ambiguous species identifications based on this
149 sequence. R code for these analyses is provided in text format (Additional file 6) and detailed
150 results are provided in .csv format (Additional file 7).

151 We conducted PCRs to evaluate amplification success using the *AusPreda_12S*
152 primers on tissue samples from a range of mammal species. Details of samples used are
153 provided in .csv format (Additional file 8). We also tested amplification success from known-
154 origin scats collected from six different predator species. All PCR products successfully
155 amplified from scats were sequenced to confirm predator of origin: resulting sequences are
156 provided here in FASTA format (Additional file 9).

## 157 RESULTS

### 158 Development of a new mammal mini-barcode



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

We selected the 12S rRNA gene as a promising candidate marker for development of a mini-barcode and developed a 12S rRNA reference sequence database for Australian mammals comprising 174 sequences. Within the 12S rRNA gene, we identified a 178 bp diagnostic mini-barcode region that displayed high levels of inter-specific variation. Within this region, the proportion of zero non-conspecific K2P distances was equal to zero for windows of 175 bp in length, and the number of diagnostic nucleotides per window was high. We identified two potential primer sites with high proportions of zero non-conspecific K2P distances (>0.8) and low numbers of diagnostic nucleotides (0-1 nucleotides per 20bp window). We designed two conserved primers, *AusPreda_12SF* and *AusPreda_12SR*, to amplify this mini-barcode from a range of mammal species. The final PCR product was 218 bp in length, including the primers.

**Bioinformatic evaluation of the mini-barcode**

We used three different genetic distance based analyses to estimate the risks of species mis-identification when using our *AusPreda_12S* primers on samples of unknown origin (Table 1, Additional file 7). These analyses used versions of the 12S rRNA reference sequence database, trimmed to include only the 178 bp mini-barcode region (Additional files 4 and 5). A *nearNeighbour* analysis of all sequences (the “FULL” database) correctly identified 155 sequences and incorrectly identified 19 sequences. All incorrectly identified sequences except one western quoll (*D. geoffroii*) from GenBank originated from species for which only a single reference sequence was available (i.e. singleton species), and thus the nearest neighbour was automatically another species. In most cases this nearest neighbour was a member of the same genus. For example, the nearest neighbour of the only bronze quoll (*D. spartacus*) sequence available was from the western quoll (*D. geoffroii*). This close

genetic similarity has also been shown by Woolley *et al.* [48]. The western quoll incorrectly identified with the nearest neighbour analyses was closely related to the bronze quoll which can indicate that this particular western quoll sequence from GenBank (KJ780027) was possibly mis-identified. Further analyses using a database including only unique haplotypes, from which singleton species were excluded (the “UNIQUE” database) identified correctly all 44 sequences.

Table 1: Summary of results of genetic distance-based evaluations of the *AusPreda_12S* mini-barcode.

| | FULL (1% threshold) | | | | UNIQUE (3.5% threshold) | | | |
|---|---|---|---|---|---|---|---|---|
| | Correct / True | Incorrect / False | Ambiguous | No ID | Correct / True | Incorrect / False | Ambiguous | No ID |
| *Nearest neighbour* | 155 | 19 | - | - | 44 | 0 | - | - |
| *Best close match* | 147 | 3 | 0 | 24 | 42 | 0 | 0 | 2 |
| *Thresh ID* | 142 | 3 | 5 | 24 | 42 | 0 | 0 | 2 |
| | FULL (3.5% threshold) | | | | | | | |
| | Correct / True | Incorrect / False | Ambiguous | No ID | | | | |
| *Best close match* | 152 | 6 | 0 | 16 | | | | |
| *Thresh ID* | 141 | 5 | 12 | 16 | | | | |

Legend: Summary of results of genetic distance-based evaluations of the *AusPreda_12S* mini-barcode conducted using the R package SPIDER to analyse the “FULL” (at 1% and 3.5% thresholds) and “UNIQUE” (at 3.5% threshold) reference sequences databases. The thresholds were calculated based on the minimum cumulative error (Additional file 6) and the 3.5% threshold for the “FULL” database allows for comparison between the two databases. The specified genetic distance thresholds were used for the *bestCloseMatch* and *threshID* analyses.

*BestCloseMatch and ThreshID* analyses, which both assume that sequences from a single species fall within a specified genetic distance threshold, correctly identified 147 and 142 sequences respectively in the “FULL” database using the 1% threshold given by the minimum cumulative error. Three sequences were incorrectly identified in both analyses: *Dasyurus spartacus* (AF009892), *Pseudantechinus macdonnellensis* (EU086642) and *Pseudantechinus roryi* (EU086650) each representing singleton species, and falling within the 1% genetic distance threshold of a congeneric species enabling them to be mistaken for their

close relatives. Five *D. geoffroii* sequences were correctly identified using *BestCloseMatch* but were ambiguously identified in the *ThreshID* analysis because of a close similarity (within the 1% genetic distance threshold) with the single *D. spartacus* sequence. A further 24 sequences could not be identified in either analysis because all other sequences within the reference database were more than 1% different. The majority of these sequences were from singletons, but a more relaxed genetic distance threshold (2%-5%) identified them correctly. *BestCloseMatch* and *ThreshID* analyses of the "UNIQUE" database identified correctly 42 of 44 sequences, but the two remaining sequences, both from *Dasycercus cristicauda*, could not be identified (Table 1; details of results: Additional file 7). As noted previously, these sequences would have been correctly identified if a genetic distance threshold of 5% was used. This represents a high level of divergence between two conspecific sequences, but as both of these sequences were obtained from GenBank, and the origin of one of the samples is unknown, we cannot rule out sample misidentification or sequencing error in this instance.

Using a 3.5% genetic threshold for the "FULL" database, to allow for comparison with the results obtained with the "UNIQUE" database, correctly identified more sequences with the *BestCloseMatch* analysis which was to be expected using a more relaxed genetic threshold allowing for more mismatches among sequences. Nevertheless, six sequences previously resulting in an "No ID" match became correctly identified and two became incorrectly identified. The western quoll (KJ780027) became incorrectly identified using a higher threshold which, once again, lead us to believe that this sequences from GenBank was incorrectly identified to start with. Comparing the *ThreshID* results with the more conservative approach used with the 1% threshold, five sequences that were previously correctly identified became ambiguous and from eight sequences resulting in a "No ID"

match, four became correctly identified, two became incorrectly identified and two had an ambiguous identification.

### Evaluation of the amplification success and sensitivity of the *AusPreda_12S* primers

Our mini-barcode was successfully amplified from all 45 tissue samples tested, including samples from a wide taxonomic range of Australian mammals (40 species), as well as a reptile, an amphibian and a bird (Figure 1, Additional file 8). This demonstrates the broad applicability of the primers across the mammalian taxa and their potential applicability to other vertebrate classes. Because we aimed to target both marsupial and eutherian mammals, we were unable to identify a mini-barcode that amplified only the six target species.

We also successfully amplified our mini-barcode from a wide range of input template DNA concentrations. We set up serial dilutions of DNA from six predator species. Amplification was successful for all three qPCR replicates from all six species for all dilutions from 9 ng / µl to 9 pg / µl inclusive, demonstrating that the primers can amplify from low quantity DNA. Amplification success was less consistent at the highest and lowest DNA concentrations, estimated at 90 ng / µl, 0.9 pg /µl and 0.09 pg / µl (Table 2) indicating that reliability of predator detection from DNA below 9 pg / µl may be poor. Failure to amplify from highly concentrated DNA, despite successful amplification from dilutions of the same DNA extracts, may reflect the presence of PCR inhibitors in these extracts, which were obtained from museum and roadkill specimens.

250 Table 2: Results of qPCR tests conducted to evaluate amplification success of the *AusPreda*_12S mini-
251 barcode from low template DNA. Six DNA samples were serially diluted, with amplification success
252 determined by comparison of CT values[1] for three replicates of each dilution.

| Species | Dilution | Replicate 1 | Replicate 2 | Replicate 3 | CT Mean[2] |
|---|---|---|---|---|---|
| **Cat N22b** | 1 in 10 (9 ng/µl) | 12.444 | 14.281 | 13.373 | 13.366 |
| | 1 in 100 (0.9 ng/µl) | **16.346[3]** | 13.399 | 13.368 | 13.384 |
| | 1 in 1000 (0.09 ng/µl) | 19.252 | 23.382 | 23.994 | 22.209 |
| | 1 in 10 000 (9 pg/µl) | 31.252 | 27.486 | 27.604 | 28.781 |
| | 1 in 100 000 (0.9 pg/µl) | 31.483 | 31.476 | 29.386 | 30.782 |
| | 1 in 1 000 000 (0.09 pg/µl) | **Undetermined** | **Undetermined** | **Undetermined** | - |
| **Dingo AA15020** | 1 in 10 (9 ng/µl) | 14.303 | 13.019 | 15.363 | 14.228 |
| | 1 in 100 (0.9 ng/µl) | 15.879 | 16.791 | 16.623 | 16.431 |
| | 1 in 1000 (0.09 ng/µl) | 19.719 | 19.237 | 17.424 | 18.793 |
| | 1 in 10 000 (9 pg/µl) | 22.652 | 24.957 | 25.196 | 24.268 |
| | 1 in 100 000 (0.9 pg/µl) | **Undetermined** | **Undetermined** | **Undetermined** | - |
| | 1 in 1 000 000 (0.09 pg/µl) | **Undetermined** | **Undetermined** | **Undetermined** | - |
| **Eastern quoll UC1214** | 1 in 10 (9 ng/µl) | 14.128 | 13.509 | 13.449 | 13.695 |
| | 1 in 100 (0.9 ng/µl) | 17.267 | **20.866[3]** | 17.235 | 17.251 |
| | 1 in 1000 (0.09 ng/µl) | 17.662 | 21.523 | 21.385 | 20.190 |
| | 1 in 10 000 (9 pg/µl) | 24.346 | 26.474 | 25.653 | 25.491 |
| | 1 in 100 000 (0.9 pg/µl) | **Undetermined** | **Undetermined** | 34.570 | 34.570 |
| | 1 in 1 000 000 (0.09 pg/µl) | **Undetermined** | **Undetermined** | **Undetermined** | - |
| **Spotted-tailed quoll A3395** | 1 in 10 (9 ng/µl) | 13.460 | 13.928 | 14.048 | 13.812 |
| | 1 in 100 (0.9 ng/µl) | 17.517 | 16.447 | 18.653 | 17.539 |
| | 1 in 1000 (0.09 ng/µl) | 20.374 | 19.540 | 17.003 | 18.972 |
| | 1 in 10 000 (9 pg/µl) | 27.511 | 25.453 | 23.851 | 25.605 |
| | 1 in 100 000 (0.9 pg/µl) | 30.158 | 30.132 | 25.107 | 28.466 |
| | 1 in 1 000 000 (0.09 pg/µl) | **Undetermined** | 35.172 | **Undetermined** | 35.172 |
| **Red fox UC0401** | 1 in 10 (9 ng/µl) | 15.547 | 15.528 | 14.628 | 15.234 |
| | 1 in 100 (0.9 ng/µl) | 19.566 | 17.524 | 16.860 | 17.983 |
| | 1 in 1000 (0.09 ng/µl) | 21.915 | 22.827 | 22.360 | 22.367 |
| | 1 in 10 000 (9 pg/µl) | 26.672 | 25.460 | 25.508 | 25.880 |
| | 1 in 100 000 (0.9 pg/µl) | 31.672 | 30.914 | 28.863 | 30.483 |
| | 1 in 1 000 000 (0.09 pg/µl) | **Undetermined** | 31.601 | **Undetermined** | 31.601 |
| **Tasmanian devil A3357** | 1 in 10 (9 ng/µl) | 15.502 | **16.810[3]** | 14.536 | 15.019 |
| | 1 in 100 (0.9 ng/µl) | 19.736 | 18.729 | 19.702 | 19.389 |
| | 1 in 1000 (0.09 ng/µl) | 23.517 | 22.999 | 21.591 | 22.702 |
| | 1 in 10 000 (9 pg/µl) | 27.216 | 28.006 | 24.130 | 26.451 |
| | 1 in 100 000 (0.9 pg/µl) | 30.876 | 30.734 | 28.977 | 30.196 |
| | 1 in 1 000 000 (0.09 pg/µl) | 32.534 | **Undetermined** | **Undetermined** | 32.534 |

253 [1] Numbers represent observed CT (cycle threshold) values for each replicate qPCR of a series of DNA
254 dilutions. The CT value represents the number of cycles required for the fluorescent signal of a qPCR
255 machine to cross the predetermined threshold, here set at 5000 ΔRn.

256 [2] Undetermined results were excluded when calculating mean CT.

257 [3] Where the qPCR traces were of an irregular shape (three replicates), the replicate was excluded
258 when calculating mean CT.

259 **Evaluation of amplification success from trace samples using known-origin scats**

260 We tested the ability of the *AusPreda_12S* primers to correctly identify known
261 predators by analysing scats from captive animals. 57 scats were tested and amplified
262 product was obtained from 53 samples. We obtained good quality DNA sequences, ranging
263 from 116 bp to 182 bp in length, from 49 (92%) of these 53 scats (Additional file 9). The
264 species of origin was correctly identified for all 49 samples, with scat DNA sequences
265 matched to appropriate GenBank reference sequences with 97-100% sequence identity
266 (Table 3).

267 **DISCUSSION**

268 Non-invasive environmental DNA-based methods can provide a novel approach to
269 the detection of cryptic animals in large-scale surveys [49], with applications to wildlife
270 management. Such DNA approaches can make important contributions to the ability to
271 detect incursions or monitor established invasive species [50], [51], [41] or to detect very
272 rare or declining species of conservation significance [52][8].

273 Here, we report a PCR-based mini-barcode test for medium-large Australian
274 mammalian predators. This test can amplify DNA from and discriminate among the four
275 quoll species found in Australia, as well as the Tasmanian devil (the only other extant large
276 marsupial predator) and introduced mammal carnivores with a high level of accuracy. We
277 expect that these primers will also amplify DNA from both species of New Guinean quoll.
278 Previous studies, aimed at identifying species from scats or hairs, have applied barcoding

methods to detect individual species across multiple time points (examples in [53], [54]). Here we have shown that it is also possible to identify multiple species by implementing a single DNA test, using a straightforward PCR and Sanger sequencing approach. All clear sequences obtained from 49 scats of six target predator species were correctly identified to species level. In the small number of cases where a clear sequence was not obtained from a scat, we found that the sequences obtained were mixed, probably arising from the amplification of two or more species in the same sample. This could arise from cross contamination among samples but is more likely the result of the amplification of prey DNA present in the scat [14], [55]. We have previously observed this phenomenon when using a single species test to detect fox DNA, where rabbit or hare DNA were sometimes erroneously amplified [37]. This demonstrates the need to account for the history of samples analysed (how they were obtained, how fresh they were upon collection, and how samples and DNA extracts were stored) and the importance of a DNA sequencing step in any of these analyses to enable recognition of non-specific PCR amplification. In practice, mixed sequences cannot be used to identify the predator with confidence and therefore such samples must be excluded from analysis. In addition to successful amplification of scat DNA, we demonstrate that our mini-barcode primers can successfully amplify low-template DNA (at least as low as 0.9 pg / µl) from museum samples. This provides further evidence of the utility of this marker for application to eDNA studies.

Whilst DNA metabarcoding may more clearly determine which species are represented in mixed samples, metabarcoding methods are relatively costly and require more specialist equipment, which may not be available to many wildlife managers. In this study, PCR and Sanger sequencing reliably identified the predator of origin for 86% of scat samples, which is likely to be sufficient for many management applications and is a higher

303 success rate than has been reported for several other faecal DNA studies (for example [41], 304 where 79% of sequences were amplified using a 134 bp fragment and [57], where <70% of 305 sequences were amplified using regions ranging from 243 bp to 708 bp according to target 306 taxon). Using our mini-barcode, DNA can be screened for the presence of multiple Australian 307 predator species in a single and inexpensive test, without the need to develop and apply a 308 set of species-specific primers for each predator of interest. We provide a non-invasive 309 instrument with potential utility for scientists or managers working with endangered or 310 invasive Australian predators, but a similar approach could be used to target predator 311 assemblages in other regions.

312 The bioinformatic evaluation of our mini-barcode shows that this marker can reliably 313 discriminate among the eight target predator species (eastern, western, northern and 314 spotted-tail quolls, Tasmanian devils, cats, dogs and foxes) in Australia. The close genetic 315 similarity between the bronze quoll (from New Guinea) and the western quoll (from 316 Australia), described above and supported by [48], may pose some problems for reliable 317 species identification from unknown samples, but the different geographic distributions of 318 these two species will likely provide a clear identification in most cases. The most 319 appropriate threshold to be used will depend on the management context and the relative 320 importance of false positive identifications, but in most cases, an ambiguous or "No ID" 321 identification would be a better result for a sample than to result in a correct identification 322 when this is erroneous.

323 Further development of our reference database, to include additional *D.* 324 *albopunctatus* and *D. spartacus* sequences, will be required to better understand the utility 325 of this test for identification of specimens to species level in New Guinea. Likewise, a better

reference database would improve the relevance of this DNA test for application to historic samples. Sequences from the extinct thylacine could be clearly identified in our initial analyses, but this species could not be included in the UNIQUE database for further bioinformatic analysis because only one 12S rRNA haplotype was available. Finally, because we are working with mitochondrial DNA which is maternally inherited, we cannot currently use this test to distinguish between dogs and dingos, in part because of the prevalence of hybrids in many wild populations [57], [58].

**Considerations when working with scats**

One important consideration for future studies using the *AusPreda_12S* primers is the need to understand the ecological role of the species from which eDNA is detected. Typically, predator DNA is the most abundant in scats, owing to the release of epithelial cells during defecation [59], [60], [61]. However, because there are multiple potential sources of DNA in scat samples, it is also possible that these primers will amplify DNA from prey species. In some cases, this will be obvious, for example where the scats of the prey species detected are clearly morphologically different from carnivore scats. However, other results may be more difficult to interpret, for example where mixed sequences, representing two different predator species which could potentially predate upon one another, are obtained from the same sample.

**Conservation implications**

The *AusPreda_12S* primers provide an opportunity to enhance monitoring of predators across Australia for conservation purposes [63]. For example, western quolls were successfully re-established in Western Australia in 1987 after a recovery plan implemented over 13 years, in areas previously baited with 1080 to remove introduced species [63].

349 Western quolls from Western Australia were also re-introduced to the Flinders Ranges in
350 South Australia in 2014, and that population is now breeding in the wild, with more than 60
351 young born since their relocation [64], [65]. Eastern quolls were re-introduced from
352 Tasmania to Mulligans Flat Woodland Sanctuary, in the Australian Capital Territory, in early
353 2016 [66]. There are also proposals to reintroduce devils to south-eastern mainland Australia
354 to reduce the negative impact that dingo control has on small-mammals through
355 mesopredator release [67], [68], [69], [70]. The development of this mini-barcode now
356 provides a new tool with which to monitor these re-introduced species, and the non-native
357 predators that threaten them, from non-invasive samples.

358 **Future work**

359 In the future, this predator identification tool may be used to model the distribution
360 of predators in Tasmania or mainland Australia, supplementing more traditional data
361 obtained from live trapping and sightings. It is now possible to reliably detect a predator of
362 interest from non-invasive samples. Using the *AusPreda_12S* primers in an initial sample
363 screening step may provide further opportunities to study the diets of each specific
364 predator, by identifying samples to include in targeted metabarcoding studies. This test
365 could also be more broadly useful, with potential application to detection and monitoring of
366 the two New Guinean quoll species.

367 **METHODS**

368 **Selection of a candidate marker gene**

369 We compiled initial reference databases for three mitochondrial genes, 12S rRNA,
370 16S rRNA and ND2, all of which have proven useful for species detection in other studies

[61], [71], [72], [73], [74]. These databases used sequences collected mainly from GenBank (GenBank, RRID:SCR_002760) [75], [76].

We used the R package SPIDER to identify potential mini-barcodes from these initial reference databases. Our criteria were to identify regions of between 100 and 200 bp in length (the maximum that can be reasonably amplified from many eDNA samples) that displayed high levels of inter-specific variation within the region, and that were flanked by primer sites that were well-conserved across all taxa, but particularly across our six key Tasmanian target species. For each gene, we conducted a sliding window analysis with window sizes of 100, 125, 150 and 175 bp, to identify potential mini-barcodes. For each window, we evaluated the number of diagnostic nucleotides per window and the proportion of zero non-conspecific K2P distances, to identify regions with high inter-specific variation, that may be used to discriminate among species. Subsequently, we used further sliding window analyses to identify conserved primer sites adjacent to candidate mini-barcode regions. We used window sizes of 20, 25 and 30 bp to identify potential sites for primer development. Of these, a window size of 20 gave the best results, so we adopted 20 bp as the standard primer length.

We were not able to identify any candidate mini-barcode markers that met all of our criteria from the 16S rRNA and ND2 genes, so all subsequent work was focused on the 12S rRNA gene.

**Development of a reference database for the 12S rRNA gene**

We constructed a reference database for the 12S rRNA gene. This included representatives of native and introduced Tasmanian mammal predators and their potential prey species, their mainland Australian relatives, livestock and other introduced species (i.e.

goat, sheep, horse, wild boar, cow and fallow deer) and humans. Importantly, all six recognised quoll species (four Australian and two New Guinean) were represented (Additional files 1 and 2). The final reference database consisted of 174 sequences representing 41 species from 24 genera. We obtained the majority of sequences from GenBank, but we generated additional sequences from a selection of species that were under-represented in the public database. DNA was extracted from tissue samples from museum specimens, road-killed animals, and western quoll tissues collected during a reintroduction program in the Flinders Ranges (South Australia) involving quolls of Western Australian origin [77]. We used a salting out method [78] with minor modifications as follows. Our lysis buffer included 10% SDS and tissues were digested in a thermomixer for three hours at 56 °C with mixing at 500 rpm. DNA pellets were air dried for 30-60 minutes and re-suspended in 50 µl of ddH$_2$O. Genomic DNA extracts were quantified using a Nanodrop ND1000 spectrophotometer (Thermo Fischer Scientific) and samples were diluted with ddH2O to a final concentration of *ca* 40 ng/ µl. The entire 12S gene region was amplified by PCR using primers 12C and 12gg (Table 4). PCRs of 25 µl final volume contained 0.4 µM of each primer, 1x MyTaq$^{TM}$ red mix (Bioline) and *ca* 3.2 ng/ µl of genomic DNA. Cycling conditions were: 95 °C for 2 mins; ten cycles of 95 °C for 20 s, a touchdown from 60 °C - 50 °C for 20 s, and 72 °C for 1 min; then 35 cycles of 95 °C for 20 s, 50 °C for 20 s, and 72 °C for 1 min; followed by a final extension at 72 °C for 4 mins. PCR products were visualised on a 1.7% TBE agarose gel (Agarose I: Amresco, Solon, OH, USA) run for 40 mins at 90 V. Hyperladder 50 bp (Bioline, Australia) was included to serve as a size reference. Amplicons were cleaned using Diffinity rapid tips (Scientific Specialties, Inc., California, USA) and prepared for sequencing following protocols recommended by the Biomolecular Resource Facility (Australian National University) before being sequenced in both directions on a 96

capillary 3730 DNA Analyzer (Applied Biosystems). Forward and reverse sequences for each sample were manually checked, trimmed of primer sequences and low quality bases at the 3' ends, and aligned using Geneious 8.1.7 (Biomatters, Auckland, New Zealand) [79]. The final alignment was 901 bp in length.

**Table 4: PCR primers used in this study.**

| Marker | Sequence (5' – 3') | Amplicon length | Reference |
|---|---|---|---|
| ***12C & 12GG*** | 12C: AAAGCAAARCACTGAAAATG<br>12GG: TRGGTGTARGCTRRRTGCTTT | 1061 bp | [80] |
| ***AusPreda_12S*** | AusPreda_12SF: CCAGCCACCGCGGTCATACG<br>AusPreda_12SR: GCATAGTGGGGTCTCTAATC | 218 bp | This study |

## Development of primers for the mini-barcode

A sliding window analysis of our 12S rRNA reference database, using the R package SPIDER [47], identified a candidate mini-barcode of 344 bp in length. The proportion of zero non-conspecific K2P distances was equal to zero for bases 66 to 410 of our alignment, using a sliding window analysis with 175 bp windows, and each window included high numbers of diagnostic nucleotides (51-69 per window). Within this candidate mini-barcode, a sliding window analysis using 20 bp windows identified two short, highly conserved regions suitable for primer design (Figure 2 and Additional file 3). These potential primer sites had a high proportion of zero non-conspecific K2P distances (>0.8) and low numbers of diagnostic nucleotides (0-1 per window). Within these regions, we manually designed the primers *AusPreda_12SF* (5'-CCAGCCACCGCGGTCATACG-3') and *AusPreda_12SR* (5'-GCATAGTGGGGTCTCTAATC-3') (Table 4). These primers flank a region of high inter-specific variation and amplify a product of 218 bp in length (178 bp excluding primers).

## Bioinformatic evaluation of the mini-barcode

We used additional functions of the R package SPIDER to estimate the risks of species mis-identification when using our *AusPreda_12S* primers on samples of unknown origin. These analyses were conducted using two versions of our 12S reference database, trimmed to include only the 178 bp of sequences flanked by the *AusPreda_12S* primers. The "FULL" database included all 174 sequences present in the original database (Additional file 4). The "UNIQUE" database was a subset of the "FULL" database in which each haplotype was represented by only a single sequence, and in which singleton species (species represented by only one haplotype) were removed. This included 44 sequences representing 16 species from 12 genera (Additional file 5).

Pairwise genetic distance was calculated for each pair of sequences using the "raw" model. We conducted bioinformatic analyses using the *nearNeighbour*, *bestCloseMatch*, and *threshID* functions to identify the taxa most likely to be misidentified or ambiguously identified using our primers. R code for these analyses is provided in Additional file 6. The *nearNeighbour* function determines, for each sequence in the reference database, whether the most closely related sequence originates from a conspecific, with two outcomes possible: "true" or "false". A genetic distance threshold must be specified for the *bestCloseMatch* and *threshID* functions to account for intra-specific variation. We estimated the most appropriate genetic thresholds to use for the "UNIQUE" and "FULL" databases to be 3.5% and 1% respectively based on the thresholds with the lowest cumulative error. The *bestCloseMatch* analysis identified the most closely related sequence, within the specified genetic distance threshold, and its species of origin, for each query sequence. The *threshID* analysis extended this, to consider species of origin for all sequences within the genetic distance threshold. These analyses had four possible outcomes: "correct", "incorrect",

“ambiguous” and “no identification” [47]. The “FULL” database was also analysed with a 3.5% genetic threshold to allow for comparison with the results of the “UNIQUE” database.

### Evaluation of the amplification success and sensitivity of the *AusPreda_12S* primers

We screened a panel of DNA samples from 45 specimens representing 40 species (Additional file 8) to evaluate amplification success of the *AusPreda_12S* primers. DNA was extracted from tissue samples as described above, and amplified with the *AusPreda_12S* primers using the same cycling conditions as for the 12C and 12gg primers above, with PCR products visualised on a 1.7% TBE agarose gel to determine amplification success (Figure 1).

To test the sensitivity of our primers to detect low template DNA samples, we set up serial dilutions of six DNA extracts originating from museum samples, representing each of the six mammal predators that might be detected in Tasmania (Tasmanian devil, eastern quoll, spotted tail quoll, cat, dog and fox). The DNA concentration of each original DNA extraction was determined using a QuBit Fluorometer and the Qubit dsDNA BR Assay Kit (Thermo Fisher) and diluted with $ddH_2O$ if necessary to obtain a starting concentration of 90 ng / µl. We then set up a series of six 10 X dilutions from each of these “undiluted” (90 ng / µl) samples. For each dilution of each sample, we performed three qPCR replicates, each with a total volume of 25µl including 1X Gold buffer (Applied Biosystems), 2 mM $MgCl^2$, 0.4 mg / ml BSA, 0.4 µM of each primer, 0.6 µl SYBR green (1:2000 Life Technologies nucleic acid gel stain), 0.25 mM of each dNTP, 1 unit of AmpliTaq Gold™ (Applied Biosystems) and 2 µl of the appropriate DNA dilution. qPCRs were conducted using a Viia7 Real-Time PCR system (Thermo Fisher Scientific) with an initial step of 95 °C for 5 mins; followed by 40 cycles of 95 °C for 30 s, 57 °C for 30 s and 72 ° for 30 s. We conducted a comparative CT analysis using

482 the ViiA7 software v1.2.4, with a threshold of 5,000 ΔRn. For each dilution of each DNA
483 sample we calculated the mean CT value and the standard deviation across PCR replicates2.

## 484 Evaluation of amplification success from trace samples using known-origin scats

485 We used previously-extracted DNA from 57 scats of known-origin collected in 2010-
486 2011 from captive animals, including eastern quolls, spotted-tailed quolls, Tasmanian devils,
487 foxes, cats and dogs. DNA was extracted using a combined chelex (Bio Rad Laboratories,
488 Hercules, California, USA) and spin column (Mega quick-spin Total Fragment DNA
489 Purification Kit, Intron Biotechnology) methods [81]. We evaluated amplification success
490 from these samples using the *AusPreda_12S* primers, by conducting PCRs and visualising PCR
491 products by gel electrophoresis as described above.

492 All amplified products were sequenced in both directions using the *AusPreda_12S*
493 primers, following the methods described above for primers 12C and 12gg. Forward and
494 reverse reads were aligned in Geneious 8.1.7 using a global alignment with free end gaps
495 (Geneious alignment) allowing 65% similarity. Primers were trimmed and a consensus
496 sequence was generated for each sample. Consensus sequences were compared against the
497 GenBank database using nucleotide BLAST ( NCBI BLAST, RRID:SCR_004870, MEGABLAST
498 with the “nr” option and a maximum hit of 20) to identify the most likely species of origin.

## 499 Availability of supporting data and material

500 The datasets and R code associated with this article are provided as supporting information.
501 All DNA sequences generated during this study have been submitted to GenBank: accession
502 numbers KX786294 to KX786344. Details on the method used to evaluate the sensitivity of a mini-
503 barcode can also be found in Protocols.io [82].

*Additional file 1*: 12S rRNA reference sequence database used for primer design (FASTA format)

*Additional file 2*: Samples included in the 12S rRNA reference sequence database used for primer design (.csv format)

*Additional file 3*: R code for sliding windows analysis implemented using SPIDER (text format)

*Additional file 4*: Reference database used for genetic distance based evaluation of the *AusPreda_12S* mini-barcode: “FULL” database (FASTA format)

*Additional file 5*: Reference database used for genetic distance based evaluation of the *AusPreda_12S* mini-barcode: “UNIQUE” database (FASTA format)

*Additional file 6*: R code for genetic distance based evaluation of the *AusPreda_12S* mini-barcode implemented using SPIDER (text format)

*Additional file 7*: Detailed results of genetic distance based evaluation of the *AusPreda_12S* mini-barcode (.csv format)

*Additional file 8*: Samples included in the laboratory evaluation of the *AusPreda_12S* mini-barcode (.csv format)

*Additional file 9:* Consensus sequences obtained from 53 known-origin scats by amplification with the *AusPreda_12S* mini-barcode (FASTA format)

**List of abbreviations**

BLAST: Basic Local Alignment Search Tool: Tool available through NCBI to compare an unknown sequence to existing sequences in a public database.

bp: base pairs: pairs of nucleotides in a DNA or RNA strand

CT value: cycle threshold: the number of cycles required for the fluorescent signal of a qPCR machine to cross the predetermined threshold.

DNA: deoxyribonucleic acid

mtDNA: mitochondrial DNA

eDNA: environmental DNA

PCR: polymerase chain reaction, a method used to amplify a target DNA or RNA strand

rRNA: ribosomal ribonucleic acid

TBE: Tris/Borate/EDTA: buffer for gel electrophoresis

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This study was funded by the Invasive Animals Cooperative Research Centre as part of project 1.L.21.

**Authors' contributions**

EM, AM and SS designed the study. EM performed the experiments. EM and AM analysed the data. EM wrote the manuscript and AM and SS provided extensive comments. All authors read and approved the final manuscript.

**Acknowledgements**

Thanks to the following for providing samples: Alison Taylor (Kippax Veterinary Hospital); Kejun Wei (Australian National University); Sally Bryant, Elise Dewar and Michael Driessen (Department of Primary Industries, Parks, Water and Environment, Tasmania); Shane McHugh (Tasmanian Land Conservancy); Judy Dunlop (Department of Parks and Wildlife, Western Australia); Leo Joseph and Robert Palmer (Australian National Wildlife Collection, CSIRO); Melissa Jensen (University of Adelaide); Catriona Campbell and Matthew Young (University of Canberra); Tony Buckmaster (University of Sydney); Shane Jackson (University of Tasmania); Phil Newman (Rugosa Park Alpacas); Conrad Barr; The Australian Museum; Museum Victoria; The Tasmanian Museum and Art Gallery; The Queen Victoria Museum and Art Gallery. Special thanks to Tony Buckmaster, Arthur Georges and members of the Invasive Animals CRC's 2016 Kioloa writing workshop as well as Stephen Harris for helpful suggestions.

## REFERENCES

[1] G. Ceballos, P. R. Ehrlich, A. D. Barnosky, A. García, R. M. Pringle, and T. M. Palmer, "Accelerated modern human–induced species losses: Entering the sixth mass extinction," *Sci. Adv.*, vol. 1, no. 5, p. e1400253, 2015.

[2] S. Boyer, S. D. J. Brown, R. A. Collins, R. H. Cruickshank, M.-C. Lefort, J. Malumbres-Olarte, and S. D. Wratten, "Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA," *PLoS One*, vol. 7, no. 5, 2012.

[3] I. Meusnier, G. A. C. Singer, J.-F. Landry, D. A. Hickey, P. D. N. Hebert, and M. Hajibabaei, "A universal DNA mini-barcode for biodiversity analysis," *BMC Genomics*, vol. 9, no. 1, p. 214, 2008.

[4] A. Alberdi, I. Garin, O. Aizpurua, and J. Aihartza, "The foraging ecology of the mountain long-eared bat Plecotus macrobullaris revealed with DNA mini-barcodes," *PLoS One*, vol. 7, no. 4, p. e35692, 2012.

[5] W. J. Kress, C. García-Robledo, M. Uriarte, and D. L. Erickson, "DNA barcodes for ecology, evolution, and conservation," *Trends Ecol. Evol.*, no. 0, 2015.

[6] O. Berry and S. D. Sarre, "Gel-free species identification using melt-curve analysis," *Mol. Ecol. Notes*, vol. 7, no. 1, pp. 1–4, 2007.

[7] T. M. Wilcox, K. S. McKelvey, M. K. Young, S. F. Jane, W. H. Lowe, A. R. Whiteley, and M. K. Schwartz, "Robust detection of rare species using environmental DNA: the importance of primer specificity," *PLoS One*, vol. 8, no. 3, p. e59520, 2013.

[8] H. C. Rees, B. C. Maddison, D. J. Middleditch, J. R. M. Patmore, and K. C. Gough, “REVIEW: The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology,” *J. Appl. Ecol.*, vol. 51, no. 5, pp. 1450–1459, 2014.

[9] O. Berry, S. D. Sarre, L. Farrington, and N. Aitken, “Faecal DNA detection of invasive species: The case of feral foxes in Tasmania,” *Wildl. Res.*, vol. 34, no. 1, pp. 1–7, 2007.

[10] J. M. Korstian, A. M. Hale, V. J. Bennett, and D. A. Williams, “Using DNA barcoding to improve bat carcass identification at wind farms in the United States,” *Conserv. Genet. Resour.*, pp. 1–8, 2015.

[11] W. Shehzad, T. M. McCarthy, F. Pompanon, L. Purevjav, E. Coissac, T. Riaz, and P. Taberlet, “Prey preference of snow leopard (*Panthera uncia*) in South Gobi, Mongolia,” *PLoS One*, vol. 7, no. 2, p. e32104, 2012.

[12] M. Koester, S. Claßen, and R. Gergs, “Establishment of group-specific PCR primers for the identification of freshwater macroinvertebrates,” *Conserv. Genet. Resour.*, vol. 5, no. 4, pp. 1091–1093, 2013.

[13] S. K. Gupta and A. Kumar, “Molecular identification of man-eating carnivores from scat samples,” *Conserv. Genet. Resour.*, vol. 6, no. 2, pp. 271–274, 2014.

[14] F. Grattarola, S. González, and M. Cosse, “A novel primer set for mammal species identification from feces samples,” *Conserv. Genet. Resour.*, vol. 7, no. 1, pp. 57–59, 2015.

[15] C. Shores, S. Mondol, and S. K. Wasser, “Comparison of DNA and hair-based approaches to dietary analysis of free-ranging wolves (Canis lupus),” *Conserv. Genet. Resour.*, vol. 7, no. 4, pp. 871–878, 2015.

[16] T. Takahara, T. Minamoto, and H. Doi, “Using environmental DNA to estimate the distribution of an invasive fish species in ponds,” *PLoS One*, vol. 8, no. 2, p. e56584, 2013.

[17] A. Fujiwara, S. Matsuhashi, H. Doi, S. Yamamoto, and T. Minamoto, “Use of environmental DNA to survey the distribution of an invasive submerged plant in ponds,” *Freshw. Sci.*, vol. 35, no. 2, p. 0, 2016.

[18] B. E. Deagle, J. P. Eveson, and S. N. Jarman, “Quantification of damage in DNA recovered from highly degraded samples - A case study on DNA in faeces,” *Front. Zool.*, vol. 3, 2006.

[19] M. Hajibabaei, M. Smith, D. H. Janzen, J. J. Rodriguez, J. B. Whitfield, and P. D. N. Hebert, “A minimalist barcode can identify a specimen whose DNA is degraded,” *Mol. Ecol. Notes*, vol. 6, no. 4, pp. 959–964, 2006.

[20] A. Valentini, F. Pompanon, and P. Taberlet, “DNA barcoding for ecologists,” *Trends Ecol. Evol.*, vol. 24, no. 2, pp. 110–117, 2009.

[21] A. R. Bahrmand, H. Madani, V. V Bakayev, M. H. Babaei, G. Samar, and V. Anashchenko, “Polymerase chain reaction of bacterial genomes with single universal primer: application to distinguishing mycobacteria species,” *Mol. Cell. Probes*, vol. 10, no. 2, pp. 117–122, 1996.

[22] R. Meier, K. Shiyang, G. Vaidya, and P. K. L. Ng, “DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success,” *Syst. Biol.*, vol. 55, no. 5, pp. 715–728, 2006.

[23] R. Boutros, N. Stokes, M. Bekaert, and E. C. Teeling, “UniPrime2: a web service providing easier Universal Primer design,” *Nucleic Acids Res.*, p. gkp269, 2009.

[24] E. M. Furlan, D. Gleeson, C. M. Hardy, and R. P. Duncan, “A framework for estimating the sensitivity of eDNA surveys,” *Mol. Ecol. Resour.*, 2015.

[25] A. J. Macdonald and S. D. Sarre, “A framework for developing and validating taxon-specific primers for specimen identification from environmental DNA,” *Mol. Ecol. Resour.*, 2016.

623 [26] R. E. Wheat, J. M. Allen, S. D. L. Miller, C. C. Wilmers, and T. Levi, "Environmental DNA from
1 624 residual saliva for efficient noninvasive genetic monitoring of brown bears (Ursus arctos),"
2 625 *PLoS One*, vol. 11, no. 11, p. e0165259, 2016.
3
4 626 [27] D. J. Morin, M. J. Kelly, and L. P. Waits, "Monitoring coyote population dynamics with fecal
5 627 DNA and spatial capture–recapture," *J. Wildl. Manage.*, vol. 80, no. 5, pp. 824–836, 2016.
6
7 628 [28] J. C. Z. Woinarski, A. A. Burbidge, and P. L. Harrison, "Ongoing unraveling of a continental
8 629 fauna: decline and extinction of Australian mammals since European settlement," *Proc. Natl.*
9 630 *Acad. Sci.*, vol. 112, no. 15, pp. 4531–4540, 2015.
10
11 631 [29] A. A. Burbidge and N. L. McKenzie, "Patterns in the modern decline of Western Australia's
12 632 vertebrate fauna: causes and conservation implications," *Biol. Conserv.*, vol. 50, no. 1–4, pp.
13 633 143–198, 1989.
14
15 634 [30] A. P. Elkin, "Reaction and interaction: a food gathering people and European settlement in
16 635 Australia," *Am. Anthropol.*, vol. 53, no. 2, pp. 164–186, 1951.
17
18 636 [31] O. J. F. Brown, "Tasmanian devil (Sarcophilus harrisii) extinction on the Australian mainland in
19 637 the mid-Holocene: multicausality and ENSO intensification," *Alcheringa An Australas. J.*
20 638 *Palaeontol.*, vol. 30, no. S1, pp. 49–57, 2006.
21
22 639 [32] F. W. King, "Extant unless proven extinct: the international legal precedent," *Conserv. Biol.*,
23 640 vol. 2, no. 4, pp. 395–397, 1988.
24
25 641 [33] R. Paddle, *The last Tasmanian tiger: the history and extinction of the thylacine*. Cambridge
26 642 University Press, 2002.
27
28 643 [34] T. Hollings, M. Jones, N. Mooney, and H. McCallum, "Trophic Cascades Following the Disease-
29 644 Induced Decline of an Apex Predator, the Tasmanian Devil," *Conserv. Biol.*, 2013.
30
31 645 [35] A. Brüniche-Olsen, C. P. Burridge, J. J. Austin, and M. E. Jones, "Disease induced changes in
32 646 gene flow patterns among Tasmanian devil populations," *Biol. Conserv.*, vol. 165, pp. 69–78,
33 647 2013.
34
35 648 [36] H. S. Bender, J. A. Marshall Graves, and J. E. Deakin, "Pathogenesis and molecular biology of a
36 649 transmissible tumor in the Tasmanian devil," *Annu. Rev. Anim. Biosci.*, vol. 2, no. 1, pp. 165–
37 650 187, 2014.
38
39 651 [37] S. Burnett, "Colonizing cane toads cause population declines in native predators: reliable
40 652 anecdotal information and management implications," *Pacific Conserv. Biol.*, vol. 3, no. 1, p.
41 653 65, 1997.
42
43 654 [38] C. A. Belcher, "Demographics of tiger quoll (*Dasyurus maculatus maculatus*) populations in
44 655 south-eastern Australia," *Aust. J. Zool.*, vol. 51, no. 6, pp. 611–626, Jan. 2003.
45
46 656 [39] A. S. Glen, P. J. de Tores, D. R. Sutherland, and K. D. Morris, "Interactions between chuditch
47 657 (Dasyurus geoffroii) and introduced predators: a review," *Australian Journal of Zoology*, vol.
48 658 57, no. 5. pp. 347–356, 2009.
49
50 659 [40] B. A. Fancourt, C. E. Hawkins, and S. C. Nicol, "Evidence of rapid population decline of the
51 660 eastern quoll (Dasyurus viverrinus) in Tasmania," *Aust. Mammal.*, vol. 35, no. 2, pp. 195–205,
52 661 2013.
53
54 662 [41] S. D. Sarre, A. J. Macdonald, C. Barclay, G. R. Saunders, and D. S. L. Ramsey, "Foxes are now
55 663 widespread in Tasmania: DNA detection defines the distribution of this rare but invasive
56 664 carnivore," *J. Appl. Ecol.*, vol. 50, no. 2, pp. 459–468, 2013.
57
58 665 [42] M. Nogales, A. Martín, B. R. Tershy, C. J. Donlan, D. Veitch, N. Puerta, B. Wood, and J. Alonso,
59 666 "A review of feral cat eradication on islands," *Conserv. Biol.*, vol. 18, no. 2, pp. 310–319, 2004.
60
61 667 [43] B. T. Lazenby, N. J. Mooney, and C. R. Dickman, "Effects of low-level culling of feral cats in
62
63


64
65

open populations: a case study from the forests of southern Tasmania," *Wildl. Res.*, vol. 41, no. 5, pp. 407–420, 2015.

[44] A. R. Palumbi and F. Cipriano, "Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation," *J. Hered.*, vol. 89, no. 5, pp. 459–464, 1998.

[45] D. M. Hillis, C. Moritz, B. K. Mable, and R. G. Olmstead, *Molecular systematics*, vol. 23. Sinauer Associates Sunderland, MA, 1996.

[46] A. J. MacDonald and S. D. Sarre, "Species assignment from trace DNA sequences: an in silico assessment of the test used to survey for foxes in Tasmania," *J. Appl. Ecol.*, p. n/a-n/a, Jul. 2015.

[47] S. D. J. Brown, R. A. Collins, S. Boyer, M. LEFORT, J. MALUMBRES-OLARTE, C. J. Vink, and R. H. Cruickshank, "Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding," *Mol. Ecol. Resour.*, vol. 12, no. 3, pp. 562–565, 2012.

[48] P. A. Woolley, C. Krajewski, and M. Westerman, "Phylogenetic relationships within Dasyurus (Dasyuromorphia: Dasyuridae): quoll systematics based on molecular evidence and male characteristics," *J. Mammal.*, vol. 96, no. 1, pp. 37–46, Mar. 2015.

[49] M. K. Schwartz, G. Luikart, and R. S. Waples, "Genetic monitoring as a promising tool for conservation and management," *Trends Ecol. Evol.*, vol. 22, no. 1, pp. 25–33, 2007.

[50] J. A. Darling and M. J. Blum, "DNA-based methods for monitoring invasive species: a review and prospectus," *Biol. Invasions*, vol. 9, no. 7, pp. 751–765, 2007.

[51] A. D. Bastos, D. Nair, P. J. Taylor, H. Brettschneider, F. Kirsten, E. Mostert, E. Von Maltitz, J. M. Lamb, P. Van Hooft, and S. R. Belmain, "Genetic monitoring detects an overlooked cryptic species and reveals the diversity and distribution of three invasive Rattus congeners in South Africa," *BMC Genet.*, vol. 12, no. 1, p. 1, 2011.

[52] C. L. Jerde, A. R. Mahon, W. L. Chadderton, and D. M. Lodge, "'Sight-unseen' detection of rare aquatic species using environmental DNA," *Conserv. Lett.*, vol. 4, no. 2, pp. 150–157, 2011.

[53] N. Fernández, M. Delibes, and F. Palomares, "Landscape evaluation in conservation: molecular sampling and habitat modeling for the Iberian lynx," *Ecol. Appl.*, vol. 16, no. 3, pp. 1037–1049, 2006.

[54] K. S. McKelvey, J. V. O. N. Kienast, K. B. Aubry, G. M. Koehler, B. T. Maletzke, J. R. Squires, E. L. Lindquist, S. Loch, and M. K. Schwartz, "DNA analysis of hair and scat collected along snow tracks to document the presence of Canada lynx," *Wildl. Soc. Bull.*, vol. 34, no. 2, pp. 451–455, 2006.

[55] R. Hausknecht, R. Gula, B. Pirga, and R. Kuehn, "Urine—a source for noninvasive genetic monitoring in wildlife," *Mol. Ecol. Notes*, vol. 7, no. 2, pp. 208–212, 2007.

[56] Y.-C. Cheng and C.-P. Lin, "Dietary Niche Partitioning of Euphaea formosa and Matrona cyanoptera (Odonata: Zygoptera) on the Basis of DNA Barcoding of Larval Feces," *J. Insect Sci.*, vol. 16, no. 1, p. 73, 2016.

[57] P. Savolainen, T. Leitner, A. N. Wilton, E. Matisoo-Smith, and J. Lundeberg, "A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 33, pp. 12387–12390, 2004.

[58] C. G. Radford, M. Letnic, M. Fillios, and M. S. Crowther, "An assessment of the taxonomic status of wild canids in south-eastern New South Wales: phenotypic variation in dingoes," *Aust. J. Zool.*, vol. 60, no. 2, pp. 73–80, 2012.

[59] W. O. C. Symondson, "Molecular identification of prey in predator diets," *Mol. Ecol.*, vol. 11, no. 4, pp. 627–641, 2002.

[60] S. N. Jarman, B. E. Deagle, and N. J. Gales, “Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples,” *Mol. Ecol.*, vol. 13, no. 5, pp. 1313–1322, 2004.

[61] B. E. Deagle, D. J. Tollit, S. N. Jarman, M. A. Hindell, A. W. Trites, and N. J. Gales, “Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions,” *Mol. Ecol.*, vol. 14, no. 6, pp. 1831–1842, 2005.

[62] A. Galimberti, A. Sandionigi, A. Bruno, A. Bellati, and M. Casiraghi, “DNA barcoding in mammals: what’s new and where next?,” *Hystrix, Ital. J. Mammal.*, vol. 26, no. 1, pp. 13–24, 2015.

[63] K. Morris, B. Johnson, P. Orell, G. Gaikhorst, A. Wayne, and D. Moro, “Recovery of the threatened chuditch (Dasyurus geoffroii): a case study,” *Predators with Pouches Biol. Carniv. Marsupials. CSIRO Publ. Melb.*, pp. 435–451, 2003.

[64] Department of the Environment, “Western quolls – Reintroducing the species to the Flinders Ranges (SA),” 2015. [Online]. Available: http://www.environment.gov.au/biodiversity/threatened/publications/factsheet-western-quolls.

[65] S. Katsineris, “Endangered Quolls re-introduced to the flinders ranges,” 2015.

[66] E. Hunt, “Eastern quolls return to Australian mainland after more than 50 years,” *The Guardian*, 02-Mar-2016.

[67] A. S. Glen and C. R. Dickman, “Complex interactions among mammalian carnivores in Australia, and their implications for wildlife management,” *Biol. Rev. Camb. Philos. Soc.*, vol. 80, no. 3, pp. 387–401, 2005.

[68] C. N. Johnson and J. VanDerWal, “Evidence that dingoes limit abundance of a mesopredator in eastern Australian forests,” *J. Appl. Ecol.*, vol. 46, no. 3, pp. 641–646, 2009.

[69] E. G. Ritchie and C. N. Johnson, “Predator interactions, mesopredator release and biodiversity conservation,” *Ecol. Lett.*, vol. 12, pp. 982–998, 2009.

[70] D. O. Hunter, T. Britz, M. Jones, and M. Letnic, “Reintroduction of Tasmanian devils to mainland Australia can restore top-down control in ecosystems where dingoes have been extirpated,” *Biol. Conserv.*, vol. 191, pp. 428–435, 2015.

[71] A. Di Finizio, G. Guerriero, G. L. Russo, and G. Ciarcia, “Identification of gadoid species (Pisces, Gadidae) by sequencing and PCR–RFLP analysis of mitochondrial 12S and 16S rRNA gene fragments,” *Eur. Food Res. Technol.*, vol. 225, no. 3–4, pp. 337–344, 2007.

[72] F. Pompanon, B. E. Deagle, W. O. C. Symondson, D. S. Brown, S. N. Jarman, and P. Taberlet, “Who is eating what: diet assessment using next generation sequencing,” *Mol. Ecol.*, vol. 21, no. 8, pp. 1931–1950, 2012.

[73] P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann, and E. Willerslev, “Towards next-generation biodiversity assessment using DNA metabarcoding,” *Mol. Ecol.*, vol. 21, no. 8, pp. 2045–2050, 2012.

[74] B. E. Deagle, S. N. Jarman, E. Coissac, F. Pompanon, and P. Taberlet, “DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match,” *Biol. Lett.*, vol. 10, no. 9, p. 20140562, 2014.

[75] NCBI, “National Center for Biotechnology Information.” [Online]. Available: http://www.ncbi.nlm.nih.gov/genbank/.

[76] L. Y. Geer, A. Marchler-Bauer, R. C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, and S. H. Bryant, “The NCBI biosystems database,” *Nucleic Acids Res.*, p. gkp858, 2009.

[77] K. E. Moseby, D. E. Peacock, and J. L. Read, "Catastrophic cat predation: A call for predator profiling in wildlife protection programs," *Biol. Conserv.*, vol. 191, pp. 331–340, 2015.

[78] A. J. MacDonald, S. D. Sarre, N. N. FitzSimmons, and N. Aitken, "Determining microsatellite genotyping reliability and mutation detection ability: an approach using small-pool PCR from sperm DNA," *Mol. Genet. Genomics*, vol. 285, no. 1, pp. 1–18, 2011.

[79] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, and C. Duran, "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data," *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, 2012.

[80] M. S. Springer, L. J. Hollar, and A. Burk, "Compensatory substitutions and the evolution of the mitochondrial 12S rRNA gene in mammals.," *Mol. Biol. Evol.* , vol. 12, no. 6, pp. 1138–1150, Nov. 1995.

[81] D. S. L. Ramsey, A. J. MacDonald, S. Quasim, C. Barclay, and S. D. Sarre, "An examination of the accuracy of a sequential PCR and sequencing test used to detect the incursion of an invasive species: the case of the red fox in Tasmania," *J. Appl. Ecol.*, vol. 52, no. 3, pp. 562–570, 2015.

[82] Modave E, MacDonald AJ, Sarre SD (2017): Evaluation of the sensitivity of a mini-barcode. Protocols.io. http://dx.doi.org/10.17504/protocols.io.imecc3e

15
16
17
18
19
20
21
22 779 **TABLE**
23
24 780 Table 3: PCR and DNA sequencing results from 57 known-origin scat samples screened using the *AusPreda_12S* mini-barcode.
25

| Sample | Scientific name | Common name | Amplified | Sequenced | Closest sequence match using BLAST | % ID$^{\alpha}$ | e value$^{\beta}$ |
|---|---|---|---|---|---|---|---|
| 100111-27 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 1.55E-84 |
| 120111-02 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 100 | 6.52E-78 |
| 121010-11 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 1.22E-85 |
| 121010-16 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 98.4 | 2.08E-83 |
| 121010-17 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 1.98E-83 |
| 121010-30 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 5.54E-84 |
| 121010-52 | *Canis lupus familiaris* | Dog | Y | **N** | **NA** | **NA** | **NA** |
| 121010-53 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 98.9 | 2.60E-82 |
| 121010-54 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 1.22E-85 |
| 121010-56 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 98.9 | 7.22E-83 |
| 121110-55 | *Canis lupus familiaris* | Dog | Y | Y | Dog | 99.4 | 5.54E-84 |
| 170211-12 | *Canis lupus familiaris* | Dog | **N** | **NA** | **NA** | **NA** | **NA** |
| 041110-66 | *Dasyurus maculatus* | Spotted-tailed quoll | Y | Y | Spotted-tailed quoll | 98.4 | 2.08E-83 |
| 101110-9 | *Dasyurus maculatus* | Spotted-tailed quoll | Y | Y | Spotted-tailed quoll | 98.2 | 2.33E-72 |
| 170211-25 | *Dasyurus maculatus* | Spotted-tailed quoll | Y | Y | Spotted-tailed quoll | 99.4 | 1.55E-84 |
| 041110-01 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 99.4 | 2.25E-72 |
| 041110-04 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 2.05E-88 |
| 041110-07 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 4.80E-74 |
| 041110-15 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 1.01E-54 |
| 041110-74 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 1.19E-85 |
| 041110-80 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 9.34E-87 |
| 100111-05 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 3.34E-86 |
| 100111-31 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 3.34E-86 |


61
62
63
64
65

| 120111-32 | *Dasyurus viverrinus* | Eastern quoll | **N** | **NA** | **NA** | **NA** | **NA** |
|---|---|---|---|---|---|---|---|
| 120111-33 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 2.61E-87 |
| 170211-14 | *Dasyurus viverrinus* | Eastern quoll | Y | Y | Eastern quoll | 100 | 2.61E-87 |
| 100111-04 | *Felis catus* | Feral cat | Y | Y | Feral cat | 100 | 1.54E-79 |
| 120111-10 | *Felis catus* | Feral cat | Y | Y | Feral cat | 100 | 1.56E-79 |
| 120111-12 | *Felis catus* | Feral cat | Y | Y | Feral cat | 100 | 1.58E-79 |
| 120111-31 | *Felis catus* | Feral cat | Y | **N** | **NA** | **NA** | **NA** |
| 170211-13 | *Felis catus* | Feral cat | Y | Y | Feral cat | 99.2 | 3.36E-60 |
| 170211-21 | *Felis catus* | Feral cat | Y | Y | Feral cat | 100 | 1.61E-79 |
| 170211-22 | *Felis catus* | Feral cat | Y | Y | Feral cat | 100 | 1.55E-79 |
| 041110-42 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 4.02E-80 |
| 041110-47 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 9.34E-87 |
| 041110-48 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 2.61E-87 |
| 041110-53 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 2.47E-82 |
| 041110-59 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 7.32E-88 |
| 121010-06 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 4.02E-80 |
| 121010-22 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 99.4 | 5.58E-84 |
| 200910-24 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 9.34E-87 |
| 200910-25 | *Sarcophilus harrisii* | Tasmanian devil | Y | Y | Tasmanian devil | 100 | 2.61E-87 |
| 080211-04 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 99.4 | 1.22E-85 |
| 080211-05 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 99.4 | 5.54E-84 |
| 080211-06 | *Vulpes vulpes* | Red fox | Y | **N** | **NA** | **NA** | **NA** |
| 080211-07 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 97.2 | 9.35E-61 |
| 080211-08 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 99.4 | 5.54E-84 |
| 080211-09 | *Vulpes vulpes* | Red fox | **N** | **NA** | **NA** | **NA** | **NA** |
| 080211-10 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 100 | 6.52E-78 |
| 080211-11 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 98.9 | 5.66E-84 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 080211-12 | *Vulpes vulpes* | Red fox | Y | **N** | **NA** | **NA** | **NA** |
| 080211-13 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 98.8 | 3.99E-75 |
| 080211-14 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 100 | 6.52E-78 |
| 080211-15 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 99.1 | 2.63E-50 |
| 080211-16 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 100 | 6.52E-78 |
| 080211-17 | *Vulpes vulpes* | Red fox | **N** | **NA** | **NA** | **NA** | **NA** |
| 080211-18 | *Vulpes vulpes* | Red fox | Y | Y | Red fox | 97.8 | 1.23E-80 |

[α] % ID is the percentage pairwise identity between the query sequence and the matching sequence identified using BLAST.

[β] The e-value represents the number of BLAST hits expected by chance. The lower the e-value is, the better.

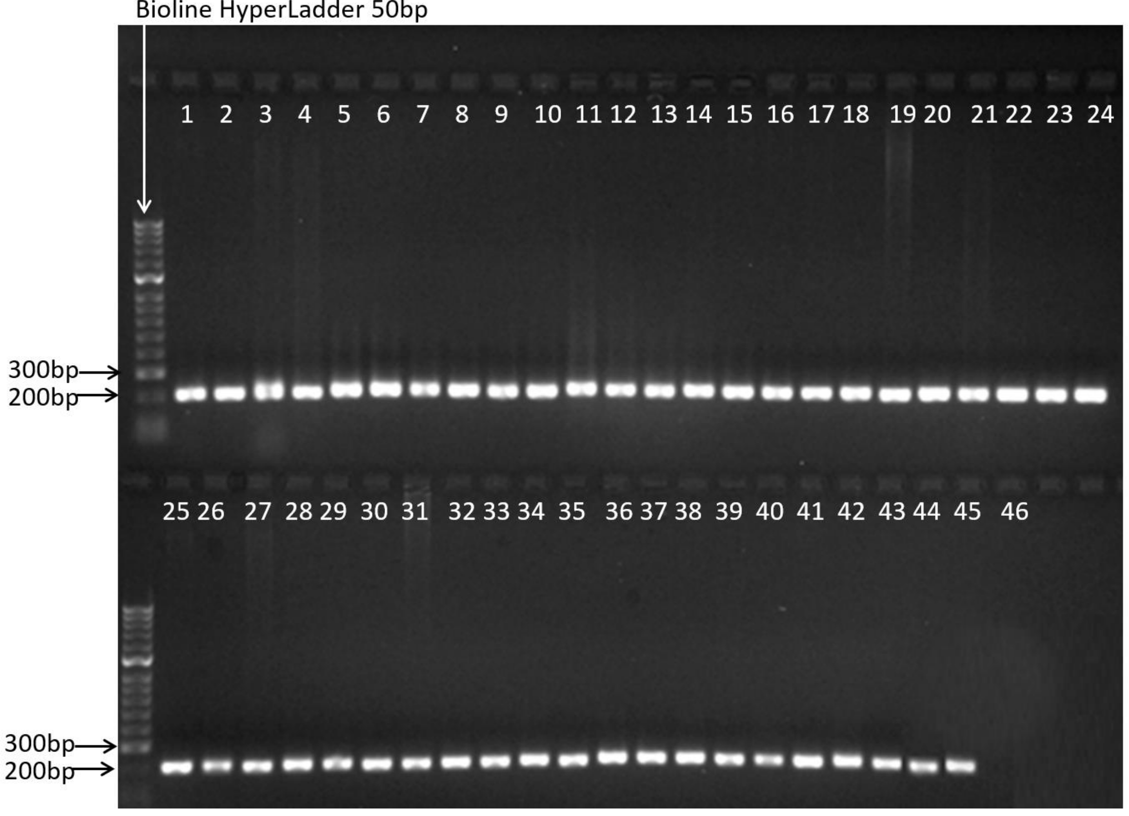
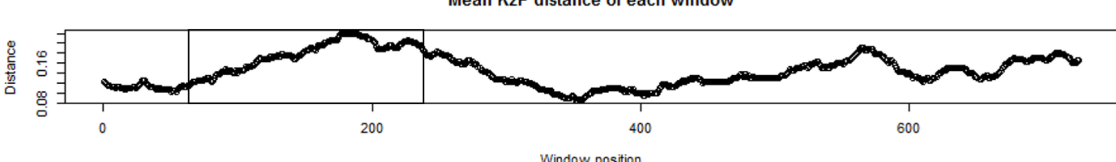


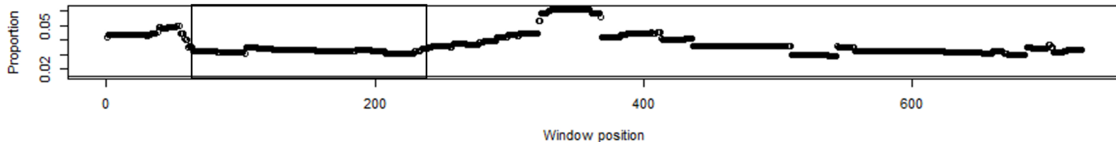

Figure 1: Gel showing amplification success from 45 known tissue samples representing 40 species, using the *AusPreda_12S* mini-barcode primers developed in this study, and a PCR negative. The expected amplicon size is 218bp. Samples are grouped by species as follows: lanes 1 and 2: *Felis catus*, 3: *Canis lupus familiaris*, 4: *Canis lupus dingo*, 5 and 6: *Dasyurus viverrinus*, 7 and 8: *Dasyurus maculatus*, 9 and 10: *Vulpes vulpes*, 11 and 12: *Sarcophilus harrisii*, 13: *Oryctolagus cuniculus*, 14: *Lepus capensis*, 15: *Bos Taurus*, 16: *Ornithorhyncus anatinus*, 17: *Trichosorus vulpecula*, 18: *Petaurus breviceps*, 19: *Tachyglossus aculeatus*, 20: *Potorous tridactylus*, 21: *Bettongia gaimardi*, 22: *Dactylopsila trivirgata*, 23: *Burramys parvus*, 24: *Macropus rufogriseus*, 25: *Thylogale billardierii*, 26: *Pseudomys gracilacaudatus*, 27: *Pseudocheirus peregrinus*, 28: *Antechinus minimus*, 29: *Tiliqua nigrolutea*, 30: *Vombatus ursinus*, 31: *Isoodon obesulus*, 32: *Macropus giganteus*, 33: *Parameles gunnii*, 34: *Sminthopsis leucopus*, 35: *Mus musculus*, 36: *Planigale gilesi*, 37: *Rattus lutreolus velutinus*, 38: *Phascogale tapoatafa*, 39: *Hydromys chrysogaster*, 40: *Macropus rufus*, 41: *Vicugna pacos*, 42: *Dasyurus hallucatus*, 43: *Lathamus discolour*, 44: *Geocrinia laevis*, 45: *Dasyurus geoffroii*, 46: PCR negative.

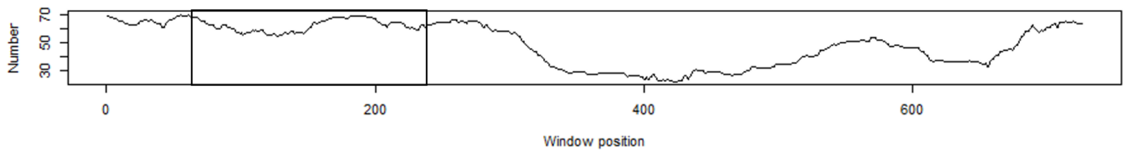
Figure 2

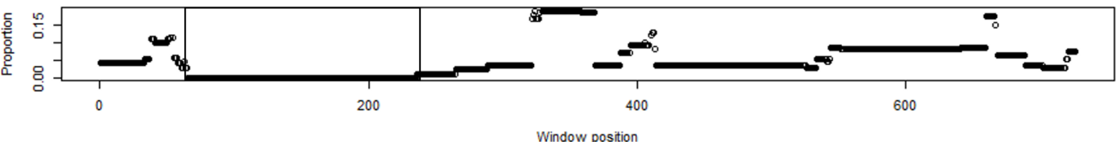
Click here to download Figure Figure 2.docx

a) Window size: 175 bp

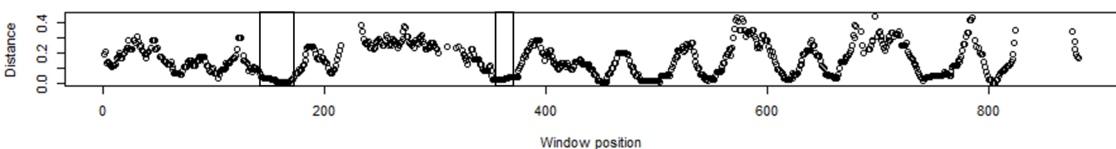


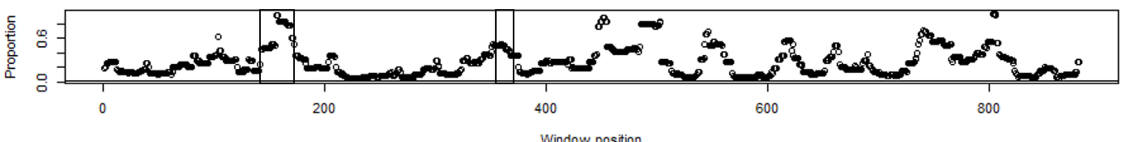



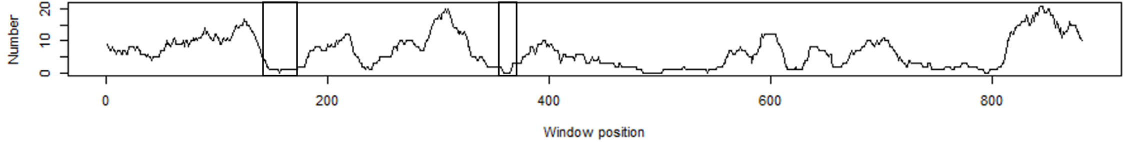




b) Window size: 20 bp



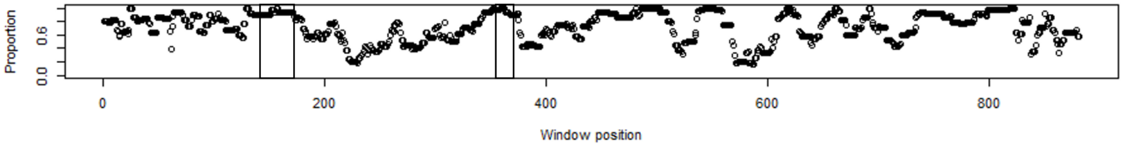



Figure 2: Results of the sliding window analysis conducted using the R package SPIDER for the 12S rRNA gene using window sizes of a) 175 bp and b) 20 bp to identify candidate mini-barcode regions and conserved primer sites respectively. For all panels, the x axes represent the position of each window within the sequence alignment, with each data point marking the position of the first nucleotide of one window. The first (top) panels display the mean K2P distances (a measure of genetic differentiation among species, where a value of zero means that sequences are identical) calculated for each window, with K2P values represented on the y-axes. The second panels represent the proportion of zero cells in the K2P distance matrix. A high proportion of inter-specific genetic distances that are equal to zero indicates sequences that are highly conserved among species. The third panels display the number of nucleotides that are diagnostic among species within each window. The fourth (lowest) panels indicate the proportion of zero non-conspecific K2P distances within each window. When this value is 0, it indicates that the sequence region has high potential to discriminate among species. The area boxed within each panel denotes a) the regions containing the first bases where a mini-barcode of *ca* 175 bp can be developed and b) the regions containing the first bases where conserved primer sites can be developed.

Supplementary Material 1

Click here to access/download

**Supplementary Material**

Additional_file_1_12S_Database_sequences.fasta

Click here to access/download
**Supplementary Material**
Additional_file_2_12S_Database_samples.csv

Click here to access/download
**Supplementary Material**
Additional_file_3_RCode_Window_analysis.txt

Click here to access/download
**Supplementary Material**
Additional_file_4_FULL_Database.fasta

Click here to access/download
**Supplementary Material**
Additional_file_5_UNIQUE_Database.fasta

Click here to access/download
**Supplementary Material**
Additional_file_6_RCode_Tests.txt

Click here to access/download
**Supplementary Material**
Additional_file_7_GeneticDistanceEvaluation_Results.csv

Click here to access/download
**Supplementary Material**
Additional_file_8_LabEvaluation_Samples.csv

Click here to access/download
**Supplementary Material**
Additional_file_9_KnownScat_sequences.fasta