CORRESPONDENCE

Comment on Falade et al. (2016) DNA-barcoding of *Clarias gariepinus*, *Coptodon zillii* and *Sarotherodon melanotheron* from Southwestern Nigeria [version 1; referees: 1 approved, 1 approved with reservations]

Jeremy A. Miller¹, Isolde van Riemsdijk¹,², Mohd Zacaery Khalik¹,², David J. Scager¹,², Menno Schilthuizen¹,²

¹Naturalis Biodiversity Center, Leiden, Netherlands
²Institute Biology Leiden, Leiden University, Leiden, Netherlands

Abstract

A publication by Falade et al. was selected for discussion by a Naturalis Biodiversity Center-Leiden University Journal Club. The study focused on the identification of fish from Southwestern Nigeria using a DNA barcoding approach. Questions raised during the discussion led to a reanalysis and reinterpretation of the data presented. The authors characterize the process of deriving a taxonomic identification from their sequence data as straightforward, but we were concerned that their approach made it nearly impossible to fail to obtain a taxonomic name for each sequence. The value of sophisticated DNA taxonomy, as well as the pitfalls of its naïve application, are discussed. We suggest that journal discussion groups may be an untapped resource for expanding rigorous peer review, particularly for journals that have adopted an open review model.

This article is included in the Phylogenetics channel.
Corresponding author: Menno Schilthuizen (menno.schilthuizen@naturalis.nl)

How to cite this article: Miller JA, van Riemsdijk I, Khalik MZ et al. Comment on Falade et al. (2016) DNA-barcoding of *Clarias gariepinus, Coptodon zillii* and *Sarotherodon melanotheron* from Southwestern Nigeria [version 1; referees: 1 approved, 1 approved with reservations] F1000Research 2016, 5:2654 (doi: 10.12688/f1000research.9829.1)

Copyright: © 2016 Miller JA et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Grant information: The Endless Forms Research Group (Naturalis) budget footed the bill for the journal club drinks at Meneer Jansen in Leiden. IvR is supported by the ‘Nederlandse organisatie voor Wetenschappelijk Onderzoek’ (NWO Open Programme 824.14.014).

Competing interests: No competing interests were disclosed.

DNA sequence data has become widely accepted as a useful tool for taxonomic determination and discovery\cite{1,2}. But the potential pitfalls of DNA taxonomy in operation have been forewarned for some time\cite{3,4}.

The DNA barcode itself is simply a standard region selected to facilitate comparison\cite{5}. A library built of many such sequences and based on a gene evolving at a rate that minimizes variation within and maximizes variation between species becomes a powerful taxonomic resource\cite{6}. But the journey from DNA barcode sequence to species determination still requires critical application, particularly when applied to taxa or regions that are not currently well represented in sequence databases.

Falade et al.\cite{7} obtained DNA sequences for sixteen individual fish from Southwest Nigeria, a region with relatively sparse coverage in sequence databases. Such data are valuable because broad geographic and taxonomic representation provide insight into genetic diversity within taxonomic groups and help us to refine hypotheses of species circumscription and phylogenetic relationships.

Falade et al.\cite{7} sequenced each specimen for the standard animal DNA barcode region cytochrome oxidase I (COI) and a region of the 16S mitochondrial ribosome. The authors queried their sequences against both the BOLD Systems (RRID: SCR_004278; boldsystems.org/index.php/IDS_OpenIdEngine) and NCBI GenBank (RRID: SCR_004860; BLASTN, RRID: SCR_001598; blast.ncbi.nlm.nih.gov/Blast.cgi) databases (because BOLD does not include 16S, these sequences were only compared to GenBank).

Although the authors claim that “this resulted in straightforward identification”, we take a more nuanced view on their results.

The BOLD identification engine and BLASTN comparison with GenBank work differently and were created for different purposes\cite{8,9,10}; only BOLD is specifically intended to be used as a taxonomic identification tool, while BLASTN assesses sequence similarity. BLASTN will always return the most similar sequences in GenBank. BOLD is more discriminating, since it is limited to a handful of specific loci and uses similarity thresholds to assess whether or not a query sequence can be matched to identified sequences in the database with high confidence. BOLD will alert the user when it determines that no confident identification could be made. DNA-based identification is complicated by the fact that both BOLD and GenBank include misidentified sequences\cite{11}.

BOLD failed to identify with confidence any of the sixteen COI sequences. Eight were classified as probably belonging to one of a handful of possible species, while the rest received no hit. From this, we infer that Falade et al. made their taxonomic determinations based almost entirely on BLASTN results. As reported (Table 1), all but one of these were scored as 98–99% identical to their top GenBank hit with the remaining sequence (KX231778; Coptodon_zilli_odooba_1) scoring 86% identical.

To view the results in context, we downloaded from BOLD all COI sequences identified as one of the three species specified by Falade et al. [search ‘Taxonomy’ for Clarias gariepinus, Sarotherodon melanotheron, and Coptodon zilli] (the latter also under the synonym Tilapia zillii). These sequences were combined with the Falade et al. data and initially aligned using MAFFT version 7.187\cite{12} with manual adjustments made using Mesquite version 3.10\cite{13}(mesquite-project.wikispaces.com/). A phylogenetic analysis was performed using RAxML version 8.2.8\cite{14}. Initial alignment and phylogenetic analysis were performed through the CIPRES Science Gateway version 3.3\cite{15}(RRID: SCR_008439; phylo.org/). Alignment required reversing or reverse-complementing some of the sequences from Falade et al. The problematic sequence KX231778 could not be satisfactorily aligned with the others and had to be excluded from the tree. The remaining COI sequences did cluster with other GenBank sequences in such a way as to suggest the remaining taxonomic determinations reported by Falade et al. are credible.

Another anomalous sequence is KX243287 (Clarias_gariepinus_ asejire_12), a 16S sequence approximately twice the length of the others. We have no explanation for this.

The evidence presented by Falade et al. is not sufficient to determine at least the COI sequence KX231778. The method applied by Falade et al. made it nearly impossible to fail to obtain a taxonomic name for each sequence. This is a scientific flaw, and an example of the uncritical application of DNA taxonomy.

This paper was discussed as part of a regular journal discussion group offered by the Endless Forms research group at Naturalis Biodiversity Center, which involves students in the Evolution, Biodiversity, and Conservation program at Leiden University. Similar journal-article-based discussion groups can be found at many universities and Natural History Museums. We support the rationale that behind open review journals (blog.f1000research.com/2014/05/21/what-is-open-peer-review/) and therefore decided to share the sense of our discussion with the broader community. We would like to encourage other journal discussion groups to include open review articles in their literature discussions, and consider sharing summaries of their discussions as article comments. Healthy science literature depends on a robust pool of potential reviewers\cite{16}. We see journal discussion groups as an untapped resource for providing feedback on scientific literature, and also as incubators for developing student-scientists into constructive and rigorous peer reviewers.

**Dataset 1. Aligned COI sequence data**

http://dx.doi.org/10.5256/f1000research.9829.d141383 (FASTA format)
Table 1. Sequences from Falade et al.13 queried using the BOLD and GenBank databases. Top BOLD hit and BOLD identification note summarize results from BOLD. Top Blast hit and Sequence name specify the best match in GenBank (excluding the Falade et al. sequences) according to BLASTN, with the Blast metrics Query cover and Ident. See also Table 2 in Falade et al. Note that BOLD contains no 16S data, so these sequences are listed as NA (not applicable).

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Locus</th>
<th>Specimen voucher no.</th>
<th>Top BOLD hit</th>
<th>Sequence name</th>
<th>Query cover</th>
<th>Ident</th>
<th>Blast note</th>
</tr>
</thead>
<tbody>
<tr>
<td>KX231778</td>
<td>COI</td>
<td>Coptodon zillii odoba_1</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231779</td>
<td>COI</td>
<td>Coptodon zillii odoba_2</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>87</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231780</td>
<td>COI</td>
<td>Coptodon zillii odoba_3</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231781</td>
<td>COI</td>
<td>Coptodon zillii odoba_4</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>87</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231782</td>
<td>COI</td>
<td>Coptodon zillii odoba_5</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231783</td>
<td>COI</td>
<td>Coptodon zillii odoba_6</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>87</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231784</td>
<td>COI</td>
<td>Coptodon zillii odoba_7</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231785</td>
<td>COI</td>
<td>Coptodon zillii odoba_8</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>87</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231786</td>
<td>COI</td>
<td>Coptodon zillii odoba_9</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231787</td>
<td>COI</td>
<td>Coptodon zillii odoba_10</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>87</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231788</td>
<td>COI</td>
<td>Coptodon zillii odoba_11</td>
<td>No Hit</td>
<td>Coptodon zillii mitochondrial genome</td>
<td>86</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

1. Coptodon zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
2. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
3. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
4. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
5. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
6. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
7. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
8. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
9. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
10. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome
11. Tilapia zillii isolate MAB08 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial genome

Note: BOLD contains no 16S data, so these sequences are listed as NA (not applicable).
<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Locus</th>
<th>Specimen voucher no.</th>
<th>Top BOLD hit</th>
<th>BOLD identification note</th>
<th>Top Blast hit</th>
<th>Sequence name</th>
<th>Query cover</th>
<th>Ident</th>
<th>Blast note</th>
</tr>
</thead>
<tbody>
<tr>
<td>KX231789</td>
<td>COI</td>
<td>Clarias_gariepinus_asejire_12</td>
<td>No Hit</td>
<td></td>
<td>KT001082.1</td>
<td>Clarias gariepinus mitochondrion, complete genome</td>
<td>90</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>KX231790</td>
<td>COI</td>
<td>Clarias_gariepinus_asejire_13</td>
<td>Top Hit: Chordata - Siluriformes - Clarias gariepinus (99.84%)</td>
<td>A species level match could not be made, the queried specimen is likely to be one of the following: Clarias gariepinus, Clarias sp., Clarias magur, Clarias cf. stappersii, Clarias ngamensis</td>
<td>JQ699203.1</td>
<td>Clarias gariepinus isolate CLGP5 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial</td>
<td>93</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231791</td>
<td>COI</td>
<td>Clarias_gariepinus_asejire_14</td>
<td>Top Hit: Chordata - Siluriformes - Clarias gariepinus (100%)</td>
<td>A species level match could not be made, the queried specimen is likely to be one of the following: Clarias gariepinus, Clarias sp., Clarias magur, Clarias cf. stappersii, Clarias ngamensis</td>
<td>KX619412.1</td>
<td>Clarias gariepinus cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial</td>
<td>93</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231792</td>
<td>COI</td>
<td>Clarias_gariepinus_asejire_15</td>
<td>Top Hit: Chordata - Siluriformes - Clarias gariepinus (100%)</td>
<td>A species level match could not be made, the queried specimen is likely to be one of the following: Clarias gariepinus, Clarias sp., Clarias magur, Clarias cf. stappersii, Clarias ngamensis</td>
<td>JQ699201.1</td>
<td>Clarias gariepinus isolate CLGP3 cytochrome oxidase subunit I (COI) gene, partial cds; mitochondrial</td>
<td>93</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX231793</td>
<td>COI</td>
<td>Clarias_gariepinus_asejire_16</td>
<td>No Hit</td>
<td></td>
<td>KX619412.1</td>
<td>Clarias gariepinus cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial</td>
<td>92</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>KX243276</td>
<td>16S</td>
<td>Coptodon_zilli_odooba_1</td>
<td>NA</td>
<td></td>
<td>KM658974.1</td>
<td>Coptodon zillii mitochondrion, complete genome</td>
<td>93</td>
<td>99</td>
<td>Top three hits from source publication, fourth hit reported</td>
</tr>
<tr>
<td>KX243277</td>
<td>16S</td>
<td>Coptodon_zilli_odooba_2</td>
<td>NA</td>
<td></td>
<td>KM658974.1</td>
<td>Coptodon zillii mitochondrion, complete genome</td>
<td>93</td>
<td>99</td>
<td>Top hit this sequence, second hit reported</td>
</tr>
<tr>
<td>KX243278</td>
<td>16S</td>
<td>Coptodon_zilli_odooba_3</td>
<td>NA</td>
<td></td>
<td>GQ168017.1</td>
<td>Tilapia aff. zilli 'Kisangani' isolate J72 16S ribosomal RNA gene, partial sequence; mitochondrial</td>
<td>90</td>
<td>99</td>
<td>Top three hits from source publication, fourth hit reported</td>
</tr>
<tr>
<td>KX243279</td>
<td>16S</td>
<td>Sarotherodon_melanotheron_odooba_4</td>
<td>NA</td>
<td></td>
<td>JF894132.1</td>
<td>Sarotherodon melanotheron mitochondrion, complete genome</td>
<td>93</td>
<td>99</td>
<td>Top two hits from source publication, third hit reported</td>
</tr>
<tr>
<td>KX243280</td>
<td>16S</td>
<td>Sarotherodon_melanotheron_odooba_5</td>
<td>NA</td>
<td></td>
<td>JF894132.1</td>
<td>Sarotherodon melanotheron mitochondrion, complete genome</td>
<td>89</td>
<td>99</td>
<td>Top two hits from source publication, third hit reported</td>
</tr>
<tr>
<td>Accession no.</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>KX243281</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
<td>16S</td>
</tr>
<tr>
<td>KX243282</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243283</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243284</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243286</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243287</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243289</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KX243291</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Accession no.**
- KX243281
- KX243282
- KX243283
- KX243284
- KX243285
- KX243286
- KX243287
- KX243288
- KX243289
- KX243290
- KX243291

**Sequence name**
- Clarias gariepinus mitochondrial genome

**Query**
- hit

**Cover**
- NA

**Ident**
- NA

**Blast note**
- NA

**Blast hit**
- Top three hits from source publication, fourth hit reported

**Identification note**
- NA

**Loos**
- 16S

**Specimen voucher no.**
- Clarias gariepinus_odooba_6
- Clarias gariepinus_odooba_7
- Clarias gariepinus_odooba_8
- Clarias gariepinus_odooba_9
- Clarias gariepinus_odooba_10
- Clarias gariepinus_odooba_11
- Clarias gariepinus_asejire_12
- Clarias gariepinus_asejire_13
- Clarias gariepinus_asejire_14
- Clarias gariepinus_asejire_15
- Clarias gariepinus_asejire_16

**Page 6 of 9**
Author contributions
MS, JM, IV, MZK, and DS conceived the study and outlined major points. IVR and JM analyzed the data and wrote initial drafts of the manuscript. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests
No competing interests were disclosed.

Grant information
The Endless Forms Research Group (Naturalis) budget footed the bill for the journal club drinks at Meneer Jansen in Leiden. IVR is supported by the ‘Nederlandse organisatie voor Wetenschappelijk Onderzoek’ (NWO Open Programme 824.14.014).

References

The title is appropriate for the content of the article. However, there is a spelling mistake since the correct genus is *Coptodon*, instead of *Coptedon* as the authors wrote it. The abstract concisely summarizes the ideas presented in the article. The authors explain with clarity their points of view about the paper under study. Conclusions are justified on the basis of the analysis performed. The information given is adequate, and discussed with clarity.

Multiple factors converge on decision making, therefore studies that bring together morphology, life cycles, ecology, genetics, and bioinformatics are desirable to improve our comprehension about species, in particular those that come from understudied localities. Falade *et al.* (2016) identified fish specimens at species level by morphology, later analyzing genes COI and 16S rRNA with the aim of correlating morphologic and genetic data. Miller *et al.* (2017) made an objection to the bioinformatic methodology employed by Falade *et al.*, stating that it was “nearly impossible to fail to obtain a taxonomic name for each sequence”. Miller *et al.* objected particularly one sequence that produced no hits on BOLD database, a problem also addressed by Falade in their original paper. The absence of genetic sequences in public databases from specimens of remote or understudied areas is a problem that researchers from those areas face quite frequently. Even though Miller *et al.* (2017) are correct in addressing the methodology shortness in Falade’s work, it is important to remark that Falade *et al.* made an important contribution in submitting genetic sequences from 3 fish species of the underrepresented country Nigeria to public databases such as GeneBank and BOLD. Hopefully, there will be more interdisciplinary studies on Nigerian fish fauna.

In a more philosophical note, none of the branches of biology can alone answer all the questions, or explain or predict the totality of biological phenomena. In particular, definition of the concept “species” is under discussion even today. Molecular biology and bioinformatics are two of the many tools that are available to elucidate the boundaries between 2 species. For example, to what extent a similarity percent of nearly 100%, based on the study of certain genes in a certain biological group, can be taken as an indicator that two species are different? That percent seems to be different for different taxa, and also varies depending on the genes under study. Bioinformatic tools to analyze genetic data are improving at a fast pace, but it is still important not to underestimate information about morphology, ecology, life cycles, etc. to complete the picture of each taxa. It is also worth noting that the improvement of bioinformatics tools relies on pre-existing information, and when that information is missing there might be a bit of a problem.
We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

**Competing Interests:** No competing interests were disclosed.

---

Shadi Shokralla
Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

The authors discussed many aspects in Falade et al. (2016) article but their explanation didn't convince me with their findings. For example:

- The authors focused on CO1 data and almost ignored 16S data.
- I was expected to see more figures to proof their points.
- "The evidence presented by Falade et al. is not sufficient to determine at least the COI sequence KX231778. The method applied by Falade et al. made it nearly impossible to fail ...." what is the right way in the authors' eyes.
- "The remaining COI sequences did cluster with other GenBank sequences in such a way as to suggest the remaining ....." At which level the clustering parameters were set to? It is a vague expression.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

**Competing Interests:** No competing interests were disclosed.