

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337063445>

Evidence for two likely cases of scientific misconduct: plagiarism of complete mitochondrial DNA sequences of a bird and a fish

Technical Report · November 2019

CITATIONS

0

READS

274

1 author:



George Sangster

Naturalis Biodiversity Center

137 PUBLICATIONS 2,508 CITATIONS

SEE PROFILE

Evidence for two likely cases of scientific misconduct: plagiarism of complete mitochondrial DNA sequences of a bird and a fish

(preliminary report)

Author:	Dr George Sangster
Affiliation:	Naturalis Biodiversity Center, Leiden, The Netherlands
Date of version:	28 April 2020
Credentials of the author:	Dr George Sangster holds a PhD in systematics and evolution from Stockholm University (2014). He has been trained in DNA sequencing, has worked on various groups of birds, and has published on molecular systematics in multiple peer-reviewed journals.
Conflicts of interest:	The author of this report received no financial support from any individuals or organizations in preparing this report. The author of this report declares no conflict of interest.
Notes added on 28 April 2020:	<p>The two papers examined in this report have now (April 2020) been retracted by the publisher:</p> <p>https://www.tandfonline.com/doi/full/10.1080/24701394.2020.1758395</p> <p>https://www.tandfonline.com/doi/full/10.1080/24701394.2020.1758393</p> <p>As a consequence, the contents of this report are no longer confidential.</p> <p>A previous version of this report was submitted to the editors of <i>Mitochondrial DNA Part A</i> in November 2019 and formed the motivation for those retractions.</p>

Table of contents

1. Summary	3
2. Initial suspicions.....	3
3. Aims and scope	3
4. Methodology	3
5. Definitions of scientific misconduct and plagiarism	4
6. Background information for non-specialist readers	4
6.1 GenBank.....	4
6.2 Complete mitochondrial genomes (mitogenomes).....	5
6.3 Nomenclature	5
7. Case 1: <i>Anser cygnoides</i> (Aves: Anseriformes)	6
7.1 Data on GenBank.....	6
7.1.1 KJ124555	6
7.1.2 Other <i>Anser cygnoides</i> mitogenomes on GenBank.....	6
7.2 Publication metadata	6
7.2.1 Mu et al. (2014)	6
7.2.2 Liu et al. (2016a)	6
7.3 Contents of the papers	7
7.3.1 Mu et al. (2014)	7
7.3.2 Liu et al. (2016a)	7
7.4 Impact.....	9
7.5 Summary of findings.....	9
7.6 Interpretation of findings.....	10
7.7 Evaluation of alternative explanations.....	11
7.8 Conclusions	11
8. Case 2: <i>Paracorbis variegatus</i> (Actinopterygii: Cipriniformes)	12
8.1 Data on GenBank	12
8.2 Publication metadata	12
8.2.1 Shi et al. (2014)	12
8.2.2 Liu et al. (2016b)	12
8.3 Contents of the papers	13
8.3.1 Shi et al. (2014).....	13
8.3.2 Liu et al. (2016b)	13
8.4 Impact.....	15
8.5 Summary of findings.....	16
8.6 Interpretation of findings.....	17
8.7 Evaluation of alternative explanations.....	17
8.8 Conclusions	18
9. The role of Mitochondrial DNA Part A	18
10. Recommendations for further investigation.....	19
11. References	19

1. Summary

Evidence is presented that Liu et al. (2016a, 2016b) may be guilty of scientific misconduct by (i) publishing two mitogenome announcements as their own, while in fact these mitogenomes were already published by other authors who are demonstrably the original authors of those sequences (Mu et al. 2014, Shi et al. 2014), (ii) re-publishing at least one figure by others as their own work, and (iii) publishing false information in both papers. The evidence indicates that these are not honest errors resulting from sloppiness, but rather that Liu et al. (2016a, 2016b) acted to mislead the editors, referees and readers of their papers.

2. Initial suspicions

While conducting a review of the authenticity of avian mitogenomic DNA sequences, the author of this report noticed that a mitochondrial genome sequence of a Swan Goose *Anser cygnoides* (Aves: Anseriformes) was described in two different papers (Mu et al. 2014, Liu et al. 2016a). GenBank data showed that the sequence was authored by two people that are listed as authors of the first paper. Examination of the papers showed that authorship of these papers does not overlap, and that one figure appeared in both papers, which raised the suspicion of plagiarism by the authors of the second paper.

Another paper by the second team (Liu et al. 2016b) also raised suspicions; it claimed to publish the first complete mitochondrial genome sequence of a species of fish, *Paracoptis variegatus* (Actinopterygii: Cypriniformes), whereas a description of such a sequence was already published by another team (Shi et al. 2014). In addition, the caption of a figure included in Liu et al. (2016b) was evidently incorrect and did not correspond to the figure. These findings prompted the investigation reported upon here.

3. Aims and scope

This report aims to present and interpret the evidence the author has gathered so far about two papers (Liu et al. 2016a, 2016b). This report could serve as a starting point for a formal and exhaustive investigation by an appropriate committee.

This report does not address any claims made in other papers by Chang Zhong Liu or his co-authors.

4. Methodology

Published evidence (scientific papers, GenBank data) was examined to determine who authored the original data, and who may have copied data or figures from others.

The contents of the relevant papers were examined to assess whether there is evidence for deception, i.e. evidence that data, text or figures may have been re-used without giving proper credit, or if data have been fabricated.

Published data (publications) were verified against the data available at GenBank.

The figures and figure captions in relevant papers were examined to assess whether data may have been copied from one paper to another.

MEGA7 (Kumar et al. 2016) was used to calculate the nucleotide composition of the two mitogenomes. The MITOS2 webserver (Berndt et al. 2013) was used to obtain information on

the gene order of the mitogenome of *Paracobitis variegatus*. These data were then compared with statements in the relevant papers.

While preparing this preliminary report, the author has consulted and - where relevant - followed the checklist prepared by the Association for the Promotion of Research Integrity, Japan (APRIN 2019).

5. Definitions of scientific misconduct and plagiarism

Scientific misconduct (also termed '*research misconduct*') includes "fabrication, falsification, and plagiarism," and requires an "intent to deceive" by the authors (Mello & Brennan 2003, Resnick et al. 2015). Fabrication, falsification, and plagiarism have also been classified as 'fraud' (e.g. Catano & Turk 2007).

Honest error (e.g. sloppiness) or differences in interpretation are not considered misconduct (Mello & Brennan 2003, Catano & Turk 2007).

Plagiarism is a less strictly defined term and has many descriptions and definitions (e.g. Liddell 2003, Juyal et al. 2015). A comprehensive definition was proposed by Liddell (2003):

"Plagiarism is the act of using someone else's words, ideas, organization, drawings, designs, illustrations, statistical data, computer programs, inventions or any creative work as if it were new and original to you; this includes real and intellectual property and public domain material." (Liddell 2003: 49).

6. Background information for non-specialist readers

6.1 GenBank

Like other DNA sequences, complete mitochondrial DNA genomes (also called 'mitogenomes') are made available on GenBank, a free online repository. GenBank assigns a unique accession number to each DNA sequence (e.g. KJ124555), which can be used by the authors and other scientists to refer to the original sequence data. Because access to sequences on GenBank is free, anyone can obtain and reuse the sequences deposited there.

Many journals require that authors deposit their DNA sequences at GenBank, and include the GenBank number(s) in their paper. This is a healthy practice because it allows others to verify the relevant conclusions and re-use the sequences to produce new knowledge.

Data listed by GenBank include the author(s) and the date of submission of the DNA sequence, the title of the paper/project that produced the sequence, and various other metadata.

In some cases, the annotations (e.g. data on the position of genes) of the mitogenome sequences are verified by GenBank staff. If a sequence is verified, a new accession is created with a different accession number that begins with "NC_". Thus, for KJ124555, GenBank staff has created another accession, NC_023832. The relevant GenBank page of the latter always mentions the accession number of the original one. This does not represent a second sequence, but merely duplicates the original one which was submitted by the authors. Both accession numbers can be used in the literature.

6.2 Complete mitochondrial genomes (mitogenomes)

Mitogenomes are often – but not always – published as data papers. Such papers are sometimes called ‘mitogenome announcements’. The purpose of such papers is to document the sequence and announce the availability of the sequence to other researchers. Mitogenomes can be used for various purposes, including but not limited to: (i) phylogenetics, (ii) phylogeography, (iii) conservation genetics, (iv) molecular identification, (v) taxonomy, (vi) design of primers for molecular studies, (vii) assembly of next generation sequence fragments, and (viii) studies of molecular evolution.

Mitogenome announcements are descriptive rather than analytical, and typically include details about the length and gene order of the mitogenome, and the length and position of the protein-coding genes, transfer RNA and ribosomal RNA.

Most mitogenome announcements also provide information about the evolutionary relationships of the relevant species. This is usually depicted by a **phylogeny** (with branch lengths corresponding to the number of mutations since the common ancestor) or a **cladogram** (where branches are of equal length and right-justified). Phylogenies can be inferred by various algorithms. The most commonly used methods in scientific papers are Maximum Likelihood analysis (ML) and Bayesian Inference (BI). The reliability of the relationships is typically expressed by bootstrap values, and Posterior Probability values, respectively.

The ease and low cost of obtaining complete mitogenomes, and their importance in various disciplines, has made it attractive to publish mitogenome announcements. Indeed, many hundreds of such announcements are being published each year. The peer-reviewed journal *Mitochondrial DNA* and its successors, *Mitochondrial DNA Part A* and *Mitochondrial DNA Part B* are major outlets for such mitogenome announcements. However, other journals also publish such papers.

6.3 Nomenclature

In the literature, the species name *Anser cygnoides* is sometimes spelled *Anser cygnoid*.

The species *Paracobitis variegatus* is currently placed by taxonomists in the genus *Homatula*. Thus, the names *Paracobitis variegatus* and *Homatula variegata* refer to the same species.

7. Case 1: *Anser cygnoides* (Aves: Anseriformes)

7.1 Data on GenBank

7.1.1 KJ124555

KJ124555 represents a complete mitochondrial genome of the Swan Goose *Anser cygnoides*.

The length of this sequence is 16,739 base pairs.

According to GenBank (accessed 27 March 2019), the authors of the sequence are: “**Mu,C. and Xu,Q**”.

The sequence was “Submitted (09-JAN-2014) College of Animal Science and Technology, Yangzhou University, Wenhui Zhong Road, Yangzhou, Jiangsu 225012, China”

7.1.2 Other *Anser cygnoides* mitogenomes on GenBank

A search on GenBank, using the search string “*Anser cygnoides* AND (mitochondrial OR mitochondrion) AND 14000:20000[Sequence Length]”, produced 12 unique sequences of this species (Table 1).

At the time of writing this report (1 April 2019), KJ124555 was the only mitogenome sequence of *Anser cygnoides* that has a total length of 16,739 base pairs (Table 1).

7.2 Publication metadata

7.2.1 Mu et al. (2014)

Details of this sequence (KJ124555) were published by Mu et al. (2014) in issue 12 of volume 22 of *Journal of Agricultural Biotechnology* (Beijing).

Full authorship of the paper is:

MU Chun-Yu¹, HUANG Zheng-Yang¹, CHEN Yang¹, WANG Bin¹, SU Yan-Hui¹, LI Yang¹, SUN Zhi-Ming², XU Qi¹, ZHAO Wen-Ming¹, CHEN Guo-Hong¹

Affiliations, as listed in the paper, are:

1 Jiangsu Key Laboratory of Animal Genetics, Breeding and Molecular Design, Yangzhou University, Yangzhou 225009, China;

2 Wuxi Zoo, Wuxi 225300, China

Publication history:

The first page includes the statement: “收稿日期: 2014-03-24 接受日期: 2014-04-25”. This translates to: “Received date: 24 March 2014, Accepted: 25 April 2014”.

7.2.2 Liu et al. (2016a)

Details of the same mitogenome sequence (KJ124555) were published by Liu et al. (2016a).

Full authorship of the paper is:

Chang Zhong Liu, Guang Hui Wei, Jian He Hu, and Xing You Liu

The affiliation of these authors, as listed in the paper, is:

College of Animal Science, Henan Institute of Science and Technology, Xinxiang, Henan Province, China

The first page includes the following details about the paper's history:

Received 27 January 2015

Revised 8 March 2015

Accepted 14 March 2015

Published online 29 April 2015

7.3 Contents of the papers

7.3.1 Mu et al. (2014)

The paper states that the GenBank accession number of the sequence is KJ124555.

The sequence length was stated to be 16,739 base pairs, which is identical to the data for KJ124555 on GenBank.

The paper states that a sample of "*Anser cygruoles*" was obtained from Wuxi Zoo. This is obviously a misspelling of the species name *Anser cygnoides*.

The paper includes multiple figures, including three cladograms. One of these (Fig. 5A) is here reproduced as Fig. 1 (see below).

The English caption in Mu et al. (2014: 1490) to this figure is "Maximum likelihood(ML) tree, and the number of branches is the posterior probability values".

7.3.2 Liu et al. (2016a)

The authors of this paper claim to have produced and to report the first mitogenome sequence of *Anser cygnoides*:

"In this study, we undertook the first complete the Swan Goose (*Anser cygnoides* L.) mitochondrial genome." (Liu et al. 2016a: 2427)

Liu et al. (2016a) did not cite Mu et al. (2014) in their paper, neither in the main text, nor in the list of references.

The paper states that the GenBank accession number of the sequence is KJ124555, i.e. the same number as in Mu et al. (2014).

The sequence length was stated to be 16,739 base pairs, which is identical to the data for KJ124555 on GenBank and the data in Mu et al. (2014).

The text in Liu et al. (2016a) on sample collection and tissue storage is as follows ('[sic]' is added where the authors misspelled a word):

"*Anser cygruoles* [sic] samples were collected from Tarim River (Xinjiang, China). All specimens were preserved in 100% ethanol and stored at 4 °C before DNA extraction. Total genomic DNA was obtained by phenol–chloroform extraction from the blood tissue of *A. cygruoles* [sic] and stored at –20 °C." (Liu et al. 2016a: 2427)

This is nearly identical to the relevant text in Jiang et al. (2014), who described their collection and storage of samples of *Aspiorhynchus laticeps*, a species of fish:

“*A. laticeps* samples were collected from Tarim River (Xinjiang, China). All specimens were preserved in 100% ethanol and stored at 4 °C before DNA extraction. Total genomic DNA was obtained by phenol–chloroform extraction from the fin tissue of *A. laticeps* and stored at –20 °C.” (Jiang et al. 2014: 219)

It is also noteworthy that the species name “*Anser cygruoles*” in Liu et al. (2016a: 2427) is the same misspelling as in Mu et al. (2014).

Another noteworthy aspect of their supposed sampling locality is that this locality is in extreme western China outside the (normal) range of *Anser cygnoides* (see Fig. 3, below).

The nucleotide composition of KJ124555 was incorrectly described in Liu et al. (2016a), but correctly described in Mu et al. (2014) (Table 2).

The last paragraph of the main text of Liu et al. (2016a) is as follows:

“Phylogenetic trees were constructed by using the ML methods (Figure 1). As indicated by the tree, different species from the same family clustered together (e.g. *A. cygnoides* Linn. Var *domestica*), and the species from *Galliformes* formed a monophyletic group. Throughout the phylogenetic analysis, *A. cygnoides* has a closer relationship to *Anser anser*, but is distantly related to *Anser fabalis* and *Anser albifrons* which have the higher level of specialization. So far, there has been neither good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of Anserinae diversity because of a wide variety of *Galliformes* (Slack et al., 2003). Thus, the mitochondrial genome data and phylogenetic analysis of the *A. cygnoides* can enrich the evolution research of *Galliformes*.” (Liu et al. 2016a: 2427)

This is clearly a modified copy of a text by Jiang et al. (2014), a paper that Liu et al. (2016b) did not cite:

“Phylogenetic trees were constructed by using the NJ and ML methods (Figs. 2 and 3). As indicated by the tree, different species from the same family clustered together (e.g. Cyprinidae), and the species from *Perciformes* formed a monophyletic group. Throughout the phylogenetic analysis, *A. laticeps* has a closer relationship to *Schizothorax*, but is distantly related to *Xenocypris* and *Hypophthalmichthys* which have the higher level of specialization. So far, there has been neither a good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of cypriniform diversity because of a wide variety of *Cypriniformes* (Saitoh et al., 2006). Thus, the mitochondrial genome data and phylogenetic analysis of the *A. laticeps* can enrich the evolution research of *Cypriniformes*.” (Jiang et al. 2014: 224)

Other issues with this paragraph in Liu et al. (2016a) are that it is incorrect to state that there is “neither [a] good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of Anserinae diversity” (see e.g. Livezey 1996, 1997, Gonzalez et al. 2009), let alone that this is somehow “because of a wide variety of *Galliformes*” (which is a different clade of birds). In addition, the reference to Slack et al. (2003) is inappropriate; this reference does not address the taxonomy or phylogenetics of Anserinae.

The paper by Liu et al. (2016a) includes a single figure: a cladogram with various species of Anseriformes and one species of *Galliformes* (reproduced here as Fig. 2, see below). This figure is nearly identical to Fig. 5A in Mu et al. (2014), except for the removal of Chinese characters, the removal of brackets, the replacement of the species names by a different font (most obviously in the letters *a* and *f*), and the lower quality of the GenBank accession numbers and the numbers on branches.

Figure 1 in Liu et al. (2016a) shares several idiosyncrasies with Fig. 5A in Mu et al. (2014):

- Both cladograms include a sequence of *Gallus gallus* (Galliformes) that should serve as the outgroup to the other species, which are all members of Anseriformes. However, in both cladograms the root is not placed on the *Gallus gallus* branch but is placed incorrectly on the *Anas platyrhynchos* branch.
- In both figures, the numbers on the branches represent the relative distance from root to tip (which, from left to right, ranges from 1 to 0). This is (almost) never shown in avian phylogenies because it is trivial.
- The figures are stated to have been based on Maximum Likelihood analysis, but in both cases the caption notes that the numbers on the branches represent 'posterior probability values'. Posterior probability values are a measure of support for Bayesian analysis, not Maximum Likelihood analysis.
- The captions to both figures use the same phrase "*the number of branches is the posterior probability values*", which is incorrect in terms of both grammar and syntax.
- Both figures contain the same misspelling of the name *Cairina* (i.e. 'Cairna').
- In both figures, the last accession number (KJ124555) is the only one without an extension. Extensions are used on GenBank to indicate the version of the sequence.

The paper includes a 'Declaration of interest' which reads: "This project was funded by the supported by grant from the State of China Agriculture Research System (CARS-589). The authors report that they have no conflicts of interest." (Liu et al. 2016a: 2428).

7.4 Impact

Mu et al. (2014) was not cited at the time of writing (cf. Google Scholar 1 April 2019).

Liu et al. (2016a) was cited by 5 papers at the time of writing (cf. Google Scholar 1 April 2019).

7.5 Summary of findings

Authorship

- (i) The authors of a mitogenomic sequence of *Anser cygnoides* (KJ124555) are also authors of the paper by Mu et al. (2014).
- (ii) The names of the authors of the paper by Liu et al. (2016a) do not contain any of the authors of the sequence (cf. GenBank) or the paper by Mu et al. (2014).

Characteristics of the sequence

- (iii) GenBank accession numbers in both papers are identical.
- (iv) The length of the mitogenome sequence mentioned in both papers is identical, and differs from that of all other sequences of *Anser cygnoides* available on GenBank.

Time line

- (v) The mitogenomic sequence (KJ124555) was submitted before both papers were published.
- (vi) The paper by Mu et al. (2014) was already submitted and accepted (and likely published and available online) before the manuscript of Liu et al. (2016a) was received by *Mitochondrial DNA Part A*.

Main text

- (vii) A sentence about the sample collection locality and aspects of tissue storage in Liu et al. (2016a) are nearly identical to statements in Jiang et al. (2014) about a species of fish.
- (viii) Consequently, the sample collection locality mentioned in Liu et al. (2016a) is false.
- (ix) The sample collection locality mentioned in Liu et al. (2016a) is outside the normal range of *Anser cygnoides*.
- (x) In both Mu et al. (2014) and Liu et al. (2016a), the section on sample collection includes the same, erroneous, species name (i.e. *Anser cygruoles*).
- (xi) The nucleotide composition of KJ124555 was incorrectly described by Liu et al. (2016a), but correctly described by Mu et al. (2014).
- (xii) The main text in Liu et al. (2016a) contains a paragraph on phylogenetic relationships that is almost identical to that in Jiang et al. (2014), without crediting that paper.
- (xiii) A statement in Liu et al. (2016b) about the state of taxonomy and phylogenetics of geese was incorrect and was supported by an irrelevant reference.

Figures

- (xiv) Figure 1 in Liu et al. (2016a) is nearly identical to Figure 5a in Mu et al. (2014).
- (xv) Comparison of branch numbers and GenBank accession numbers shows that the figure in Mu et al. (2014) is of higher quality.
- (xvi) Figure 5a in Mu et al. (2014) and Figure 1 in Liu et al. (2016a) share several idiosyncrasies.
- (xvii) The captions to these figures show the same phrase "*the number of branches is the posterior probability values*", that is incorrect in terms of both grammar and syntax.

Credit

- (xviii) Liu et al. (2016a) failed to mention the paper by Mu et al. (2014).

Impact

- (xix) Liu et al. (2016a) profited from their paper by receiving multiple citations, whereas Mu et al. (2014) did not.

7.6 Interpretation of findings

The combination of:

- (i) identical GenBank accession numbers in both papers,
 - (ii) the non-overlap of author names and affiliations of the two teams,
 - (iii) the availability of the paper by Mu et al. (2014) and the mitogenomic sequence (KJ124555) by the time Liu et al. submitted their manuscript,
 - (iv) the near-identical figure and caption, and
 - (v) failure by Liu et al. (2016a) to mention the paper by Mu et al. (2014),
- demonstrates that Liu et al. (2016a) have published a mitogenomic sequence (KJ124555) as their own and copied and modified a figure from the paper by Mu et al. (2014) and published this as their own work.

By taking credit of the work by Mu et al. (2014), Liu et al. (2016a) so far seem to have been successful in misleading other scientists. Liu et al.'s (2016a) paper has been cited multiple times, whereas Mu et al.'s (2014) paper has not gathered any citations.

In addition, by stating that the sample locality is Tarim River in Xinjiang (outside the normal range of the species), rather than the actual source of the sample (i.e. Wuxi Zoo, Wuxi, eastern China), Liu et al. (2016a) have provided false information that may negatively affect faunistic and phylogeographic research. Liu et al. (2016a) also provided false information on the nucleotide composition of KJ124555.

7.7 Evaluation of alternative explanations

Alternative explanations can be ruled out by the available evidence:

- **Alternative explanation 1:** The original team that worked on the sequence that was registered as KJ124555 on GenBank has split up and went on publishing separately about the same sequence.
 - Counter-evidence: the two papers indicate that there is no overlap of affiliations (past or present);
 - Counter-evidence: there is no indication that any of the authors of the Liu et al. (2016a) paper previously worked at Mu et al.'s institute (or vice versa);
 - Counter-evidence: the authors of the sequence on GenBank (KJ124555) do not include any members of Lui et al's team;
 - Counter-evidence: the plagiarized figure and caption indicate deception, not rivalry.
- **Alternative explanation 2:** There are actually two different sequences but Liu et al. (2016a) made an honest error by listing the accession number of Mu et al.'s (2014) sequence, rather than that of their own sequence.
 - Counter-evidence: the plagiarized and modified figure and caption, and failure to cite the paper by Mu et al. (2014), are inconsistent with sloppiness, but rather indicate deception.
 - Counter-evidence: at the time of writing this report (1 April 2019), no correction of Liu et al. (2016a) has been published in *Mitochondrial DNA Part A*, and no sequence of *Anser cygnoides* by these authors has been made available on GenBank.
- **Alternative explanation 3:** Liu et al. (2016a) are the original authors of the sequence that was registered as KJ124555 on GenBank. They were plagiarized by Mu et al. (2014).
 - Counter-evidence: Authorship of the GenBank sequence does not include any of the members of Liu et al.'s (2016a) team;
 - Counter-evidence: the higher quality of Fig. 5A in Mu et al. (2014: 1490) than that of Fig. 1 in Liu et al. (2016a: 2428) demonstrates that the former is the original and the latter is a (modified) copy.
 - Counter-evidence: The printed publication dates of the two papers show that Mu et al. (2014) was published before Liu et al. (2016a).
- **Alternative explanation 4:** Fig. 5A in Mu et al. (2014: 1490) and Fig. 1 in Liu et al. (2016a: 2428) were made independently by the two groups of authors; the similarity of the figures results from using the same methods and the same data.
 - Counter-evidence: the two figures are not just identical in terms of topology, branch numbers, selection of species and sequences, but also show the same idiosyncrasies.

Thus, none of these alternative explanations is credibly supported by the evidence.

7.8 Conclusions

The evidence indicates that this represents a case of **plagiarism** and **misappropriation of intellectual property**. In addition, the false sample collection locality and the false information on the nucleotide composition provided by Liu et al. (2016a) represent **fabrication**. Liu et al. (2016a) have deliberately misled the editor, referees and readers of the journal *Mitochondrial DNA Part A* for their own benefit.

8. Case 2: *Paracobitis variegatus* (Actinopterygii: Cipriniformes)

8.1 Data on GenBank

GenBank (accessed 29 March 2019) lists a single mitogenome of this species: JX144893.

The length of this sequence is 16,571 bp.

According to GenBank (accessed 29 March 2019), the authors of this sequence are:

Shi, Y.R., Yang, M., Chen, T. and You, P.

The sequence was "Submitted (07-JUN-2012) School of Life Sciences, Shaanxi Normal University, 199 South Chang'an Road, Xi'an, Shaanxi 710062, China".

8.2 Publication metadata

8.2.1 Shi et al. (2014)

JX144893 was formally published by Shi et al. (2014) in Volume 42, issue 4 (July 2014) of *Journal of Shaanxi Normal University (Natural Science Edition)*.

Full authorship of the paper is:

SHI Yanru¹, XING Lianxi², CHEN Tao³, YANG Min¹, YOU Ping¹

Affiliations, as listed in the paper, are:

1 College of Life Science, Shaanxi Normal University, Xi'an 710119, Shaanxi, China;

2 College of Life Science, Northwest University, Xi'an 710069, Shaanxi, China;

3 Department of Asset Management, Guilin University of Technology, Guilin 541004, Guangxi, China.

The first page includes the statement: "收稿日期 : 2 0 1 2 - 0 9 - 1 5". This translates to: "Received date: 15 September 2012".

8.2.2 Liu et al. (2016b)

Details of a mitogenome sequence of *Paracobitis variegatus* were also published by Liu et al. (2016b).

Full authorship of the paper is:

Chang Zhong Liu, Guang Hui Wei, Jian He Hu, and Xing You Liu

The affiliation of these authors, as listed in the paper, is:

College of Animal Science, Henan Institute of Science and Technology, Xinxiang, Henan Province, China

The first page includes the following details about the paper's history:

Received 1 February 2015

Revised 8 March 2015

Accepted 14 March 2015
Published online 29 April 2015

8.3 Contents of the papers

8.3.1 Shi et al. (2014)

The paper states that the GenBank accession number of the sequence is JX144893.

The sequence length was stated to be 16,571 base pairs, which is identical to the data for JX144893 on GenBank.

The paper includes a statement about where the sample was collected:

“标本于2009年5月采自陕西省柞水县营盘镇老林村（33°08'N；108°41'E）”，
which translates to:

“The specimen was collected from Laolin Village, Yingpan Town, Lishui County, Shaanxi Province in May 2009 (33°08'N; 108°41'E).”

The paper includes several figures but no cladogram/phylogeny.

8.3.2 Liu et al. (2016b)

Liu et al. (2016b) claim that their paper reports the first mitogenomic sequence of *Paracobitis variegatus*:

“In this study, the complete mitochondrial genome sequence of the *Paracobitis variegatus* [sic] was first reported.” (Liu et al. 2016b: 2421)

Liu et al. (2016b) do not cite the paper by Shi et al. (2014), neither in the main text nor in the references.

The paper does not state the GenBank accession number of the sequence being described.

The total length of the mitogenome is stated to be 16,571 bp, which is identical to the length of JX144893, and the data in Shi et al. (2014).

The paper does not include any details about the locality where the sample was collected.

The opening sentence of the paper is about *Paragyrodactylus* (Platyhelminthes), an ectoparasite of *Paracobitis variegatus*:

“*Paragyrodactylus* is a genus of Gyrodactylidae comprising three nominal species, *Paragyrodactylus iliensis*, *Paragyrodactylus barbatuli*, and *Paragyrodactylus variegatus* (Bakke et al., 2007; Elson & Lightowlers, 2006; You et al., 2014), all of which infect river loaches (Nemacheilidae) inhabiting streams in central Asia.” (Liu et al. 2016b: 2421)

This is copied almost verbatim from a paper by Ye et al. (2014), without citing that paper:

“*Paragyrodactylus* Gvosdev and Martechov, 1953 is a genus of Gyrodactylidae comprising three nominal species, *Paragyrodactylus iliensis* Gvosdev and Martechov, 1953 (= *P. dogieli* Osmanov, 1965), *Paragyrodactylus barbatuli* Ergens, 1970 and *Paragyrodactylus variegatus* You, King, Ye and Cone, 2014, all of which infect river loaches (Nemacheilidae) inhabiting streams in central Asia [8].” (Ye et al. 2014: 2)

The Elson & Lightowlers (2006) reference was improperly cited by Liu et al. (2016b). Elson & Lightowlers (2006) did not write anything about the genus *Paragyrodactylus* but rather discussed the clonality of mitochondrial DNA. Ye et al. (2014) cited this reference in a proper context.

The second sentence of the main text is as follows:

“Here, the first mitochondrial genome for *Paragyrodactylus*, *P. variegatus*, is sequenced and characterized.” (Liu et al. 2016b: 2421)

Note that *Paragyrodactylus variegatus* is the name of a flatworm species that infects *Paracobitis variegatus*.

The nucleotide composition of JX144893 was incorrectly described in Liu et al. (2016b), but correctly described in Shi et al. (2014) (Table 2).

Describing the gene order of *P. variegatus*, Liu et al. (2016b) stated that:

“The arrangement of rRNA and protein coding genes of *P. variegatus* is typical for gyrodactylids. However, the gene order of some tRNA genes is different: there are three tRNAs genes (tRNA^{Gln}, tRNA^{Phe}, and tRNA^{Met}) between ND4 and the major non-coding region and five tRNAs genes (tRNA^{Tyr}, tRNA^{Leu}, tRNA^{Ser}, tRNA^{Leu}, and tRNA^{Arg}) between ND6 and ND5 in *P. variegatus*, while *Gyrodactylus* spp. have one tRNA (tRNA^{Phe}) and seven tRNAs genes (tRNA^{Tyr}, tRNA^{Leu}, tRNA^{Gln}, tRNA^{Met}, tRNA^{Ser}, tRNA^{Leu}, and tRNA^{Arg}) in the same location.” (Liu et al. 2016b: 2421)

This is an almost exact copy of a portion of text in Ye et al. (2014):

“The arrangement of rRNA and protein coding genes of *P. variegatus* is typical for gyrodactylids. However, the gene order of some tRNA genes is different: there are three tRNAs (tRNA^{Gln}, tRNA^{Phe}, tRNA^{Met}) between ND4 and the major non-coding region and five tRNAs (tRNA^{Tyr}, tRNA^{Leu1}, tRNA^{Ser2}, tRNA^{Leu2}, tRNA^{Arg}) between ND6 and ND5 in *P. variegatus*, while *Gyrodactylus* spp. have one tRNA (tRNA^{Phe}) and seven tRNAs (tRNA^{Tyr}, tRNA^{Leu1}, tRNA^{Gln}, tRNA^{Met}, tRNA^{Ser2}, tRNA^{Leu2}, tRNA^{Arg}) in the same location, respectively.” (Ye et al. 2014: 4)

Note, however, that the '*P. variegatus*' in Ye et al. (2014) refers to *Paragyrodactylus variegatus*, which is a flatworm (Platyhelminthes), whereas the '*P. variegatus*' in the paper by Liu et al. (2016b) is supposed to refer to *Paracobitis variegatus*, a fish of the order Cypriniformes.

Perhaps not surprisingly, the description of the gene order in Liu et al. (2016b) is incorrect for *Paracobitis variegatus*. In JX144893, there are not three tRNAs between ND4 and the major non-coding region, and there are not five but zero tRNAs between ND6 and ND5 (Table 3).

The last paragraph of the main text of Liu et al. (2016b) is as follows ('[sic]' is added wherever the authors misspelled a word):

“Phylogenetic trees were constructed by using the ML methods [sic] (Figure 1). As indicated by the tree, different species from the same family clustered together, and the species from Nemaehelina [sic] formed a monophyletic group. Throughout the phylogenetic analysis, *Paracobitis variegates* [sic] has a closer relationship with *Barbatula toni*, but is distantly related to *Lefua echigonia* and *Schistura balteata* which have the higher level of specialization. So far, there has been neither good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of cypriniform diversity because of a wide variety of Nemaehelina [sic] (Gissi et al., 2008; Kritsky & Boeger, 2003). Thus, the mitochondrial genome data and phylogenetic analysis of the *Paracobitis variegates* [sic] can enrich the evolution research of Cyprinoides [sic].” (Liu et al. 2016b: 2421)

This is clearly a modified copy of a text by Jiang et al. (2014), a paper that Liu et al. (2016b) did not cite:

“Phylogenetic trees were constructed by using the NJ and ML methods (Figs. 2 and 3). As indicated by the tree, different species from the same family clustered together (e.g. Cyprinidae), and the species from Perciformes formed a monophyletic group. Throughout the phylogenetic analysis, *A. laticeps* has a closer relationship to *Schizothorax*, but is distantly related to *Xenocypris* and *Hypophthalmichthys* which have the higher level of specialization. So far, there has been neither a good reference taxonomy nor a comprehensive phylogenetic study that encompasses the whole spectrum of cypriniform diversity because of a wide variety of Cypriniformes (Saitoh et al., 2006). Thus, the mitochondrial genome data and phylogenetic analysis of the *A. laticeps* can enrich the evolution research of Cypriniformes.” (Jiang et al. 2014: 224)

Ironically, note that almost every place where Liu et al. (2016b) replaced one or more words, it included a misspelling.

In addition, the two references cited by Liu et al. (2016b) in support of their penultimate sentence have no relevance to that sentence. Thus, neither Gissi et al. (2008) nor Kritsky & Boeger (2003) deal with the phylogenetics or diversity of cypriniform fishes. Gissi et al. (2008) discussed the molecular characteristics of animal mitochondrial genomes in general, whereas Kritsky & Boeger (2003) deal with the phylogeny of a group of flatworms (Platyhelminthes). These references were used by Ye et al. (2014) in a proper context.

Fig. 1 in Liu et al. (2016b) represents a cladogram of 30 species of fish of the order Cypriniformes. The figure has a low resolution.

The caption to Fig. 1 in Liu et al. (2016b) is as follows (‘[sic]’ is added where the authors misspelled a word):

“Figure 1. The consensus phylogentic [sic] relationship of *Paracobitis variegatus* with the other species from maximum likelihood (ML) analyses. The numbers on the branches are bootstrap values for ML. GenBank accession numbers from top to bottom of different species are HM010670, HM010665, JN837667, JN837669, JF340418, JF340429, JF340432, JF340421, JF340424, HM010606, JN837663, JF304030, HM010609, HM010612, HM010613, JN827658, HM010682, JN837662, JN837646, HM010569, HM010559, JN837646, JN837650, HM010676, HM010582, HM010584, JF340389, HM010513, JF340419, and JF340432.” (Liu et al. 2016b: 2422)

These 30 GenBank accession numbers do not correspond to the 30 species in the cladogram (Table 4, see below). First, accession numbers in the caption refer to only 12 unique species, whereas Fig. 1 in Liu et al. (2016b) shows 30 different species. Second, there is no overlap of species between the caption and the figure. Third, some accession numbers do not even correspond to fishes (JF304030 corresponds to a species of copepod, JN827658 corresponds to a moth). Fourth, the 30 sequences represent a mixture of COI, cyt b and RAG1 sequences and thus cannot have produced this cladogram.

The paper includes a ‘Declaration of interest’ which reads: “This project was funded by the supported by grant from the State of China Agriculture Research System (CARS-589). The authors report that they have no conflicts of interest.” (Liu et al. 2016b: 2422)

8.4 Impact

Liu et al. (2016b) was not cited at the time of writing this report (cf. Google Scholar 1 April 2019). However, the paper has received 63 views according to the relevant webpage of the paper (cf. Taylor & Francis 1 April 2019).

8.5 Summary of findings

Authorship

- (i) The authors of the mitogenome sequence of *Paracobitis variegatus* (JX144893) are also authors of the paper by Shi et al. (2014).
- (ii) The names of the authors of the paper by Liu et al. (2016b) do not contain any of the authors of the sequence (cf. GenBank) or the paper by Shi et al. (2014).

Characteristics of the sequence

- (iii) Liu et al. (2016b) did not mention a GenBank accession number.
- (iv) At the time of writing (April 2019) only one mitogenomic sequence of *Paracobitis variegatus* was available on GenBank (i.e. JX144893).
- (v) The length of the mitogenome sequence mentioned in both papers is identical to that of JX144893 (cf. GenBank).

Time line

- (vi) The mitogenomic sequence (JX144893) was submitted to GenBank before both papers were received by the editors of the relevant journals.
- (vii) The paper by Shi et al. (2014) was already submitted (and likely published and available online) before the manuscript of Liu et al. (2016b) was received by *Mitochondrial DNA Part A*.

Main text

- (viii) The main text in Liu et al. (2016b) contains a sentence modified from Ye et al. (2014) without crediting that paper.
- (ix) The main text in Liu et al. (2016b) contains an entire paragraph (about phylogenetics) modified from Jiang et al. (2014) without crediting that paper.
- (x) Another sentence claims that the first mitochondrial genome for *Paragyrodactylus variegatus* is sequenced; this is a species that infects *Paracobitis variegatus*.
- (xi) The nucleotide composition of JX144893 was incorrectly described in Liu et al. (2016b). It was described partially but correctly in Shi et al. (2014).
- (xii) The text on the gene order of JX144893 was copied from Ye et al. (2014), who described a mitogenome of a species in a different phylum.
- (xiii) Consequently, the information on the gene order of JX144893 is false.
- (xiv) Two statements in Liu et al. (2016b) were supported by three completely irrelevant references. These references were seemingly taken from Ye et al. (2014) who used these in a proper context.

Figures

- (xv) Figure 1 in Liu et al. (2016b) may not be original work by the authors but efforts by the author of this report to find the original source of the figure so far have been unsuccessful.
- (xvi) There are multiple problems with the caption to Figure 1 in Liu et al. (2016b). The GenBank accession numbers in the caption have nothing to do with either the subject of the paper (the mitogenome of *Paracobitis variegatus*) or the cladogram of Fig. 1 in Liu et al. (2016b).

Credit

- (xvii) Liu et al. (2016b) failed to mention the paper by Shi et al. (2014) which first described the mitogenome of *Paracobitis variegatus*.

Impact

- (xviii) Neither Shi et al. (2014) nor Liu et al. (2016b) had been cited by the time this report was prepared.

8.6 Interpretation of findings

First, the combination of:

- (i) only one mitogenome of *Paracobitis variegatus* being available on GenBank (which is authored by members of the paper by Shi et al. 2014),
 - (ii) the non-overlap of author names and affiliations of the two teams,
 - (iii) the availability of the paper by Shi et al. (2014) and the mitogenomic sequence (JX144893) by the time Liu et al. (2016b) submitted their manuscript,
 - (iv) failure by Liu et al. (2016b) to mention the paper by Mu et al. (2014),
 - (v) the claim made by Liu et al. (2016b) in their abstract that they report the first mitogenomic sequence of *Paracobitis variegatus*,
- demonstrates that Liu et al. (2016b) **have published a mitogenomic sequence (JX144893), previously published by Shi et al. (2014), as their own.**

Second, the erroneous caption to Fig. 1 in Liu et al. (2016b), which shows no actual relationship to Fig. 1, demonstrates that Liu et al. (2016b) have **deliberately included false information in their paper**. Similarly, by copying a text from Ye et al. (2014), which described the mitogenome of a species in a different phylum, Liu et al. (2016b) have carelessly provided false information that may negatively affect studies of the gene order in fishes.

Third, major portions of the main text by Liu et al. (2016b) are (modified) copies of texts by Ye et al. (2014) and Jiang et al. (2014), without giving credit to those papers. This demonstrates that Liu et al. (2016b) have **plagiarized at least two other papers**.

Fourth, the improper referencing of three other works, and the curious statement in the main text that the mitogenome of *Paragyrodactylus variegatus* (a parasite of *Paracobitis variegatus*) is sequenced, demonstrate that Liu et al. (2016b) **made little or no effort to produce an original scientific paper**.

8.7 Evaluation of alternative explanations

Alternative explanations for the findings reported here are very similar to those of Case 1 but can be ruled out by the available evidence:

- **Alternative explanation 1:** The original team that worked on the sequence that was registered as JX144893 on GenBank has split up and went on publishing separately about the same sequence.
 - Counter-evidence: the two papers indicate that there is no overlap of affiliations (past or present);
 - Counter-evidence: there is no indication that any of the authors of the Liu et al. (2016b) paper previously worked at Shi et al.'s institute (or vice versa);
 - Counter-evidence: the authors of the sequence on GenBank (JX144893) do not include any members of Liu et al.'s team;
- **Alternative explanation 2:** There are actually two different sequences but Liu et al. (2016b) made an honest error by listing the accession number of Shi et al.'s (2014) sequence, rather than that of their own sequence.
 - Counter-evidence: the nonsensical caption to Figure 1 in Liu et al. (2016b), and failure to cite the paper by Shi et al. (2014), are inconsistent with sloppiness, but rather indicate deception.
 - Counter-evidence: at the time of writing this report (1 April 2019), no correction of Liu et al. (2016b) has been published in *Mitochondrial DNA Part A*, and no sequence of *Paracobitis variegatus* by these authors has been made available on GenBank.

- **Alternative explanation 3:** Liu et al. (2016b) are the original authors of the sequence that was registered as JX144893 on GenBank. They were plagiarized by Shi et al. (2014).
 - Counter-evidence: Authorship of the GenBank sequence does not include any of the members of Liu et al.'s (2016b) team;
 - Counter-evidence: The printed 'received' dates of the two papers show that the paper by Shi et al. (2014) was prepared and submitted before that of Liu et al. (2016b).

Thus, none of these alternative explanations is credibly supported by the evidence.

8.8 Conclusions

The evidence indicates that Liu et al. (2016b) represents a case of **plagiarism** and **misappropriation of intellectual property**. In addition, the multiple types of errors in the caption to Fig. 1 in Liu et al. (2016b) and the false description of nucleotide composition are the result of **fabrication**. Liu et al. (2016b) have deliberately misled the editor, referees and readers of the journal *Mitochondrial DNA Part A* for their own benefit.

9. The role of Mitochondrial DNA Part A

While plagiarism, fabrication and falsification are often difficult to detect, in these two cases the problems are quite obvious just from reading the papers or checking the main claims therein against publicly available data. This suggests that quality control at *Mitochondrial DNA Part A* was insufficient.

Specifically, the following issues could have been flagged as suspect by the referees or editors:

Both cases:

- The main claim (publication of the first mitogenome) could have been falsified by a simple search on Google Scholar; this would have revealed that a sequence of the relevant species was already published by others in 2014.

Case 1: *Anser cygnoides* (Liu et al. 2016a)

- A simple search on GenBank for accession number KJ124555 would have revealed that the authors of the manuscript are not the authors of that sequence.

Case 2: *Paracobitis variegatus* (Liu et al. 2016b)

- The accession number is not mentioned in the published paper (and thus presumably also lacking from the submitted manuscript). This is against current standards.
- Whereas the accession number was (likely) not included in the manuscript, a simple search on GenBank for a complete mitogenome of *Paracobitis variegatus* would have revealed that a mitogenomic sequence of this species was available and that the authors of the manuscript are not the authors of that sequence.
- Just verifying (on GenBank) one of the accession numbers listed in the caption to Fig. 1 would have revealed that the caption does not correspond to the figure.
- The opening sentence on a genus of parasites of *Paracobitis variegatus* is hardly relevant. The second sentence, however, is problematic because it suggests that one of these parasites was sequenced by Liu et al. (2016b). This clearly corresponds neither to the title of the paper nor to the abstract.

10. Recommendations for further investigation

Further verification of these cases can, and should be attempted. One avenue for investigation is the laboratory work. If mitogenomes of *Anser cygnoides* and *Paracobitis variegatus* were sequenced by Liu et al. (2016a) and Liu et al. (2016b), respectively, one would expect that laboratory records/materials can be made available that verify this. Such records/materials may include, but are not limited to:

- (i) Evidence that tissue samples of *Anser cygnoides* and *Paracobitis variegatus* were obtained, either by the authors in the field, from captivity or from a museum, or by others and then forwarded to the authors;
- (ii) Remaining tissue sample material;
- (iii) Remaining DNA extract;
- (iv) Records about which primers were used to amplify the DNA extract;
- (v) Remaining material of these primers;
- (vi) Laboratory records documenting the PCR amplification of DNA fragments of *Anser cygnoides*;
- (vii) Records that show that sequencing was carried out at the institute, or alternatively, communications that show that sequencing was delegated to an external institute/company, and
- (viii) Trace files of the relevant DNA fragments.

In addition, attempts should be made to verify other published works by these authors.

11. References

- Association for the Promotion of Research Integrity (APRIN), Japan 2019. How to investigate allegations of research misconduct: A checklist. Available at: <https://retractionwatch.com/2019/01/08/how-to-investigate-allegations-of-research-misconduct-a-checklist/> [accessed 27 March 2019]
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. & Stadler, P.F. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69: 313–319.
- BirdLife International & Handbook of the Birds of the World 2017. *Bird species distribution maps of the world*. Version 7.0. Available at <http://datazone.birdlife.org/species/requestdis>
- Catano, V.M. & Turk, J. 2007. Fraud and misconduct in scientific research: a definition and procedures for investigation. *Medicine & Law* 26: 465–476.
- Elson, J.L. & Lightowlers, R.N. 2006. Mitochondrial DNA clonality in the dock: can surveillance swing the case? *Trends in Genetics* 22: 603–607.
- Gonzalez, J., Düttmann, H. & Wink, M. 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *Journal of Zoology* 279: 310–318.
- Jiang, G.T., Lin, Q., He, P., Zhang, X., Yun, L., Li, X., Liu, G. et al. 2016. The complete mitochondrial genome of the Wugangtong white goose. *Mitochondrial DNA* 27: 3890–3892.
- Jiang, M., Yang, C., & Wen, H. 2014. The complete mitochondrial genome of *Aspiorhynchus laticeps* and its phylogenetic analysis. *Meta Gene* 2: 218–225.
- Juyal, D., Thawani, V. & Thaledi, S. 2015. Plagiarism: an egregious form of misconduct. *North American Journal of Medical Sciences* 7: 77–80.
- Kumar, S., Stecher, G. & Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.
- Lee, M.Y., Jeon, H.S., Choi, Y.S., Joo, S. & An, J. 2017. Sequencing and analyzing complete mitochondrial genome of *Anser cygnoides* (Anserini: Anserinae). *Mitochondrial DNA Part B* 2: 228–229.
- Liddell, J. 2003. A comprehensive definition of plagiarism. *Community & Junior College Libraries* 11: 43–52.

- Lin, Q., Cao, R., Jiang, G.T., Liu, G., Dai, Q.Z., Zhang, S.R., Hou, D.X., et al. 2016c. Determination and analysis of the complete mitochondrial genome sequence of Yanling white goose. *Mitochondrial DNA Part A* 27: 1012–1013.
- Lin, Q., Cao, R., Jiang, G.T., Qiu, L., Hu, G.B., Dai, Q.Z., Zhang, S.R., et al. 2016a. The complete mitochondrial genome of the Xupu goose. *Mitochondrial DNA Part A* 27: 1010–1011.
- Lin, Q., Jiang, G.T. & Dai, Q.Z. 2018. The complete mitochondrial genome of the *Anser cygnoides* Linnaeus, 1758 breed Daozhou and its phylogenetic analyses. *Russian Journal of Genetics* 54: 1493–1497.
- Lin, Q., Jiang, G.T., Cao, R., Yun, L., Li, G.J., Dai, Q.Z., Zhang, S.R., et al. 2016b. Determination and analysis of the complete mitochondrial genome sequence of Wugangtong grey goose. *Mitochondrial DNA Part A* 27: 1008–1009.
- Liu, C.Z., Wei, G.H., Hu, J.H., & Liu, X.Y. 2016a. Complete mitochondrial genome of the Swan Goose (*Anser cygnoides* L.) and its phylogenetic analysis. *Mitochondrial DNA Part A* 27: 2427–2428.
- Liu, C.Z., Wei, G.H., Hu, J.H. & Liu, X.Y. 2016b. Complete mitochondrial genome of *Paracobitis variegates* [sic] and its phylogenetic analysis. *Mitochondrial DNA Part A* 27: 2421–2422.
- Livezey, B.C. 1996. A phylogenetic analysis of geese and swans (Anseriformes: Anserinae), including selected fossil species. *Systematic Biology* 45: 415–450.
- Livezey, B.C. 1997. A phylogenetic classification of waterfowl (Aves: Anseriformes), including selected fossil species. *Annals of the Carnegie Museum* 66: 457–496.
- Mello, M.M. & Brennan, T.A. 2003. Due process in investigations of research misconduct. *New England Journal of Medicine* 349: 1280–1286.
- Mu, C., Huang, Z., Chen, Y., Wang, B., Su, Y., Li, Y., Sun, Z, Xu, Q., Zhao, W. & Chen, G. 2014. [Completely sequencing and gene organization of the *Anser cygnoides* mitochondrial genome.] *Journal of Agricultural Biotechnology* 22(12): 1482–1493. DOI: 10.3969/j.issn.1674-7968.2014.12.003. [In Chinese, English summary.]
- Ren, T., Liang, S., Zhao, A. & He, K. 2016. Analysis of the complete mitochondrial genome of the Zhedong White goose and characterization of NUMTs: reveal domestication history of goose in China and Euro. *Gene* 577: 75–81.
- Resnik, D.B., Neal, T., Raymond, A. & Kissling, G.E. 2015. Research misconduct definitions adopted by US research institutions. *Accountability in Research* 22: 14–21.
- Shi, Y., Xing, L., Chen, T., Yang, M. & You, P. 2014. [Sequencing and analysis of the complete mitochondrial genome of *Paracobitis variegatus*.] *Journal of the Shaanxi Normal University* 42: 50–56. [In Chinese, English summary.]
- Slack, K.E., Janke, A., Penny, D. & Arnason U. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. *Gene* 302: 43–52.
- Ye, F., King, S.D., Cone, D.K. & You, P. 2014. The mitochondrial genome of *Paragyrodactylus variegatus* (Platyhelminthes: Monogenea): differences in major non-coding region and gene order compared to *Gyrodactylus*. *Parasites & Vectors* 7(1), 377.
- Zhu, H., Li, B., Li, L. & Zhou, L. 2016. Complete mitochondrial genome of Swan goose *Anser cygnoides* (Anseriformes: Anatidae). *Mitochondrial DNA Part A* 27: 3177–3178.

Table 1. Complete mitochondrial DNA sequences of *Anser cygnoides* on GenBank (accessed 28 March 2019). Note that KJ124555 remains the only sequence with a length of 16,739 base pairs.

GenBank Acc. No.	Length (bp)	Publication / authors
KJ124555, NC_023832	16,739	Mu et al. (2014)
KP026178	16,741	Jiang et al. (2016)
KJ794188	16,742	Lin et al. (2016a)
KJ794189	16,740	Lin et al. (2016b)
KJ778677	16,742	Lin et al. (2016c)
KT427463	16,742	Ren et al. (2016)
KP238480	16,740	Zhu et al. (2016)
KY767671	16,740	Lee et al. (2017)
KU211647	16,741	Lin et al. (2018)
MK133022	16,741	Li,C., Lin,Q., Jiang,G. and Dai,Q. (unpublished)
KP881611	16,740	Zhang,Y.F., Xie,Z.X., Liu,J.B., Deng,X.W., Xie,Z.Q., Huang,L., Huang,J.L., Zeng,T.T. and Wang,S. (unpublished)
KP943133	16,742	Zhang,Y.F., Xie,Z.X., Liu,J.B., Deng,X.W., Xie,Z.Q., Huang,L., Huang,J.L., Zeng,T.T. and Wang,S. (unpublished)

Table 2. Nucleotide composition of mitogenomes of *Anser cygnoides* and *Paracobitis variegatus*. The MEGA7 source refers to an analysis performed by the author of this report using the relevant GenBank sequence. Note that the data in the original papers were correct, whereas those in the papers by Liu et al. (2016a, 2016b) were incorrect.

Sequence	Source	A	T	C	G
KJ124555 (<i>Anser cygnoides</i>)	MEGA7	30.2%	22.5%	32.2%	15.1%
	Mu et al. (2014)	30.21%	22.49%	32.24%	15.06%
	Liu et al. (2016a)	29.2%	22.2%	32.8%	15.8%
JX144893 (<i>Paracobitis variegatus</i>)	MEGA7	29.5%	26.1%	27.1%	17.3%
	Shi et al. (2014)		55.6%	-	-
	Liu et al. (2016b)	30.4%	45.8%	9.5%	14.2%

Table 3. Information on the gene order of mitogenome JX144893 (*Paracobitis variegatus*), as generated by the MITOS2 webserver (Berndt et al. 2013). Protein-coding genes are in red, transfer RNA in blue, and ribosomal RNA in green.

	Start	Stop	Strand	Length	Overlap/noncoding	Start/stop codons
tRNA-Phe (gaa)	1	69	+	69	0	
rRNA-S	70	1018	+	949	2	
tRNA-Val (tac)	1021	1092	+	72	18	
rRNA-L	1111	2765	+	1655	0	
tRNA-Leu2(taa)	2766	2840	+	75	0	
ND1	2841	3815	+	975	7	ATG/TAA
tRNA-Ile (gat)	3823	3894	+	72	-2	
tRNA-Gln (ttg)	3893	3963	-	71	1	
tRNA-Met (cat)	3965	4033	+	69	0	
ND2	4034	5080	+	1047	-2	ATG/TAG
tRNA-Trp (tca)	5079	5148	+	70	1	
tRNA-Ala (tgc)	5150	5218	-	69	1	
tRNA-Asn (gtt)	5220	5292	-	73	1	
OL	5294	5324	+	31	-2	
tRNA-Cys (gca)	5323	5388	-	66	-1	
tRNA-Tyr (gta)	5388	5456	-	69	1	
CO1	5458	7008	+	1551	0	GTG/TAA
tRNA-Ser2 (tga)	7009	7079	-	71	3	
tRNA-Asp (gtc)	7083	7155	+	73	13	
CO2	7169	7859	+	691	0	ATG/T(AA)
tRNA-Lys (ttt)	7860	7935	+	76	1	
ATP8	7937	8104	+	168	-10	ATG/TAA
ATP6	8095	8778	+	684	-1	ATG/TAA
CO3	8778	9562	+	785	-1	ATG/TA(A)
tRNA-Gly (tcc)	9562	9634	+	73	0	
ND3	9635	9985	+	351	-2	ATG/TAG
tRNA-Arg(tcg)	9984	10053	+	70	0	
ND4L	10054	10350	+	297	-7	ATG/TAA
ND4	10344	11722	+	1379	3	ATG/TA(A)
tRNA-His (gtg)	11726	11795	+	70	0	
tRNA-Ser1 (gct)	11796	11862	+	67	1	
tRNA-Leu1 (tag)	11864	11936	+	73	0	
ND5	11937	13775	+	1839	-4	ATG/TAA
ND6	13772	14293	-	522	0	ATG/TAG
tRNA-Glu (ttc)	14294	14362	-	69	4	
Cyt b	14367	15507	+	1141	0	ATG/T(AA)
tRNA-Thr (tgt)	15508	15579	+	72	-2	
tRNA-Pro (tgg)	15578	15648	-	71	7	
OH	15656	16400	+	745	171	

Table 4. Comparison of the caption to Fig. 1 with the species included in Fig. 1 of Liu et al. (2016b). Abbreviations are RAG-1, Recombination activating gene 1; Cyt b = cytochrome b; COI = Cytochrome oxidase subunit I.

GenBank accession numbers listed in the caption to Fig. 1 of Liu et al. (2016b)	Corresponding DNA marker (cf. GenBank)	Corresponding species name (cf. GenBank)	Species included in Fig. 1 of Liu et al. (2016b), from top to bottom
HM010670	RAG-1	<i>Paracobitis anguillioides</i>	<i>Apteronotus albifrons</i>
HM010665	RAG-1	<i>Paracobitis acuticephala</i>	<i>Gyrinocheilus aymonieri</i>
JN837667	RAG-1	<i>Schistura macrotaenia</i>	<i>Pseudorasbora pumila</i>
JN837669	RAG-1	<i>Triplophysa stenura</i>	<i>Mylopharyngodon piceus</i>
JF340418	RAG-1	<i>Schistura cryptofasciata</i>	<i>Ctenopharyngodo idella</i>
JF340429	RAG-1	<i>Homatula potanini</i>	<i>Cyprinus carpio</i>
JF340432	RAG-1	<i>Homatula potanini</i>	<i>Carassius cuvieri</i>
JF340421	RAG-1	<i>Homatula potanini</i>	<i>Erimyzon oblongus</i>
JF340424	RAG-1	<i>Homatula potanini</i>	<i>Moxostoma congestum</i>
HM010606	RAG-1	<i>Schistura fasciolata</i>	<i>Xyrauchen texanus</i>
JN837663	RAG-1	<i>Schistura caudofurca</i>	<i>Ictiobus bubalus</i>
JF304030	COI	<i>Acartia tonsa</i>	<i>Myxocyprinus asiaticus</i>
HM010609	RAG-1	<i>Paracobitis</i> sp. 1	<i>Misgurnus anguillicaudatus</i>
HM010612	RAG-1	<i>Paracobitis</i> sp. 1	<i>Misgurnus nikolskyi</i>
HM010613	RAG-1	<i>Paracobitis</i> sp. 1	<i>Cobitis choii</i>
JN827658	COI	<i>Hylesia murex</i>	<i>Cobitis striata</i>
HM010682	RAG-1	<i>Paracobitis anguillioides</i>	<i>Cobitis sinensis</i>
JN837662	RAG-1	<i>Homatula pycnolepis</i>	<i>Acantopsis choirorhynchus</i>
JN837646	Cyt b	<i>Homatula pycnolepis</i>	<i>Pangio anguillaris</i>
HM010569	Cyt b	<i>Paracobitis anguillioides</i>	<i>Schistura balteata</i>
HM010559	Cyt b	<i>Paracobitis anguillioides</i>	<i>Lefua echigonia</i>
JN837646	Cyt b	<i>Homatula pycnolepis</i>	<i>Paracobitis variegatus</i>
JN837650	Cyt b	<i>Homatula pycnolepis</i>	<i>Barbatula toni</i>
HM010676	RAG-1	<i>Paracobitis anguillioides</i>	<i>Sinogastromyzon puliensis</i>
HM010582	Cyt b	<i>Paracobitis anguillioides</i>	<i>Homaloptera leonardi</i>
HM010584	Cyt b	<i>Paracobitis anguillioides</i>	<i>Vaillantella maassi</i>
JF340389	Cyt b	<i>Homatula potanini</i>	<i>Leptobotia mantschurica</i>
HM010513	Cyt b	<i>Schistura fasciolata</i>	<i>Botia macracantha</i>
JF340419	RAG-1	<i>Homatula potanini</i>	<i>Chalceus macrolepidotus</i>
JF340432	RAG-1	<i>Homatula potanini</i>	<i>Phenacogrammus interruptus</i>

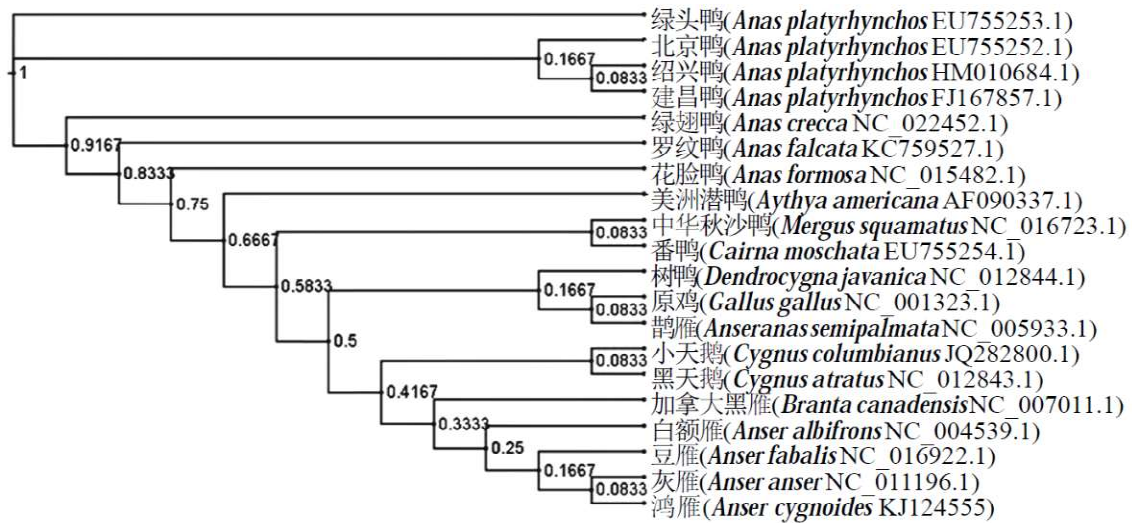


Figure 1. Reproduction of Fig. 5A in Mu et al. (2014: 1490). (Copied from pdf at 300%). The English caption in Mu et al. (2014: 1490) to this figure is: “Maximum likelihood(ML) tree, and the number of branches is the posterior probability values”.

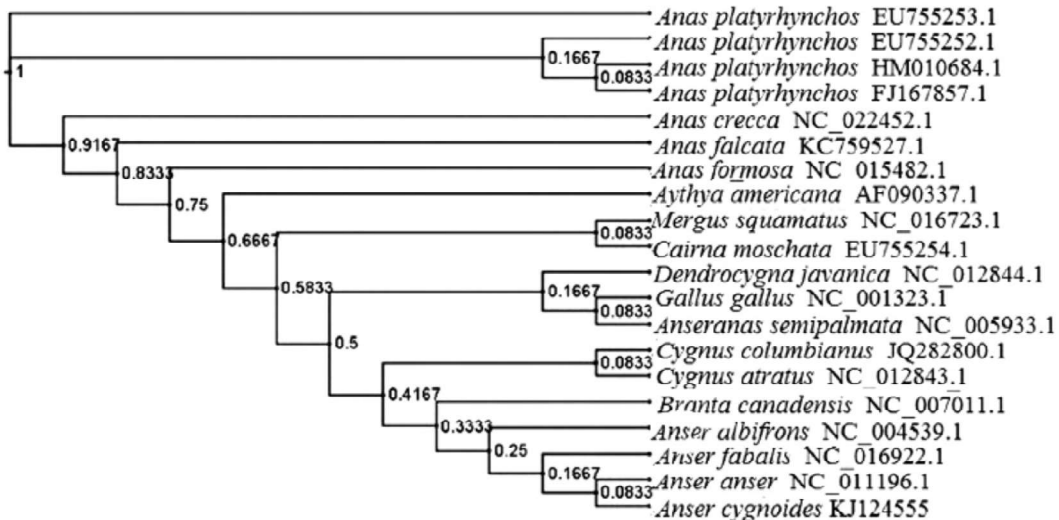


Figure 2. Reproduction of Fig. 1 in Liu et al. (2016a: 2428). (Copied from pdf at 300%). The caption in Liu et al. (2016a: 2428) to this figure is: “The consensus phylogentic relationship of *Anser cygnoides* L. with the other species from maximum likelihood (ML) analyses, and the number of branches is the posterior probability values.”

Note that this figure is almost identical to that in Mu et al. (2014), except for the removal of Chinese characters, the removal of brackets, the different font of the species names (most obviously in the letters *a* and *f*), and the lower quality of the GenBank accession numbers and the numbers on branches.

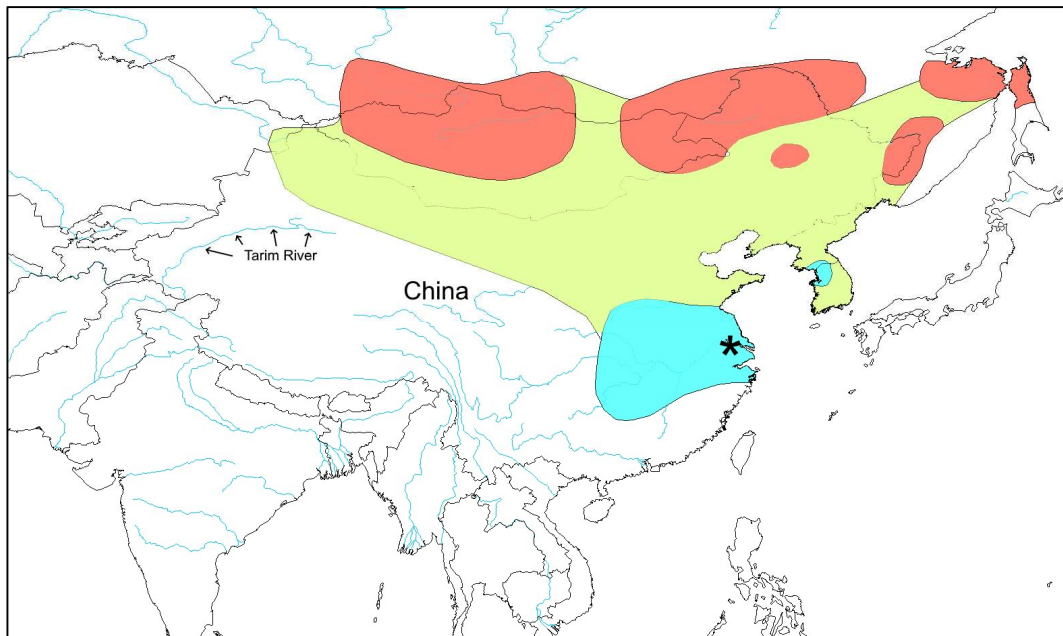


Figure 3. Range map of *Anser cygnoides* based on BirdLife International & Handbook of the Birds of the World (2017), with major rivers based on data downloaded from the DIVA-GIS website (<http://www.diva-gis.org/gdata>). Red = breeding range, green = migratory range, blue = wintering range. The locality where Mu et al. (2014) stated their sample was collected (Wuxi Zoo) is indicated by an asterisk. The locality where Liu et al. (2016a) stated their sample of *Anser cygnoides* was taken, “Tarim River in Qijiang”, is indicated by small arrows. Note that this is outside the known range of *Anser cygnoides*.