Genetic identification of invasive walking catfish, *Clarias batrachus*, intermingled with African catfish, *C. gariepinus*, in South Africa

J. Paul Grobler[1*], Sinobongo Ndyogolo[1], James Barasa[2], Romulus Abila[3], Hesmari Bindeman[1] & Andre F.J. Schlemmer[4]

[1]*Department of Genetics, University of the Free State, P.O. Box 339, Bloemfontein, 9300 South Africa*

[2]*Department of Fisheries and Aquatic Sciences, University of Eldoret, P.O. Box 1125-30100, Eldoret, Kenya*

[3]*Department of Fisheries Management and Aquaculture Technology, South Eastern Kenya University, P.O. Box 170-90200, Kitui, Kenya*

[4]*Free State Department of Economic Development, Tourism and Environmental Affairs, Private Bag X20801, Bloemfontein, 9300 South Africa*

Received 12 November 2013. To author for revision 3 March 2014. Accepted 15 May 2014

Molecular genetic techniques were used to determine the identity of two catfish individuals that could not be identified as indigenous *Clarias gariepinus* after phenotypic analysis. The unidentified catfish were compared to 126 reference samples, using sequences of the mitochondrial DNA control region. The reference dataset included African catfish from South Africa and Kenya, as well as GenBank records for this species and for three species of Southeast Asian catfish, including the potentially invasive walking catfish, *C. batrachus*. Visual inspection of sequences showed parallel polymorphisms in the unidentified catfish and *C. batrachus* that did not occur in *C. gariepinus*. A bioinformatics approach using the Basic Local Alignment Search Tool (BLAST) of the National Center for Biotechnology Information confirmed that the unidentified catfish show more similarity to catfish of Asian origin compared to *C. gariepinus*. Finally, a phylogenetic approach showed that the catfish analysed group according to continent of origin (Africa and Asia) with strong bootstrap support, and with the unidentified fish contained in the Asian cluster. We conclude that the unidentified catfish are most likely imported *C. batrachus*.

Key words: *Clarias gariepinus, Clarias batrachus*, invasive, mtDNA control region, forensics, phylogenetics.

INTRODUCTION

African catfish, *Clarias gariepinus*, is endemic to Africa (Agnese & Teugels, 2005). The aquaculture potential of this species has long been recognized in Africa, including South Africa (Van der Waal, 1978; Hoffman, Swart & Brink, 2000). The species is also cultivated outside the continent, for example in the Netherlands (Huisman & Richter, 1987) that uses stock sourced from Nigeria, the Central African Republic and Cameroon (Roodt-Wilding, Swart & Impson, 2010). Various members of the genus *Clarias* are also endemic to Southeast Asia, including *C. batrachus*. This walking catfish is widely utilized as food source in that region and is also popular in the ornamental fish trade outside Southeast Asia. Unfortunately, this species has proven to be highly adaptable to new habitats and it is considered an invasive species (Lowe, Browne, Boudjelas & De Poorter, 2004). In the United States of America (U.S.A.), this species was introduced accidentally in the state of Florida from Thailand. It is now found in 25% of rivers and threatens a number of important aquaculture species (Courtenay, 1978). Due to its status and efficacy as an invasive species, possession or import of *C. batrachus* is now banned in the U.S.A.

During an inspection of a South African fish farm by Compliance and Law Enforcement personnel from the Free State Department of Economic Development, Tourism and Environmental Affairs (DETEA), seven light and mottled juvenile individuals were observed in a population claimed to be

*To whom correspondence should be addressed.
E-mail: groblerjp@ufs.ac.za

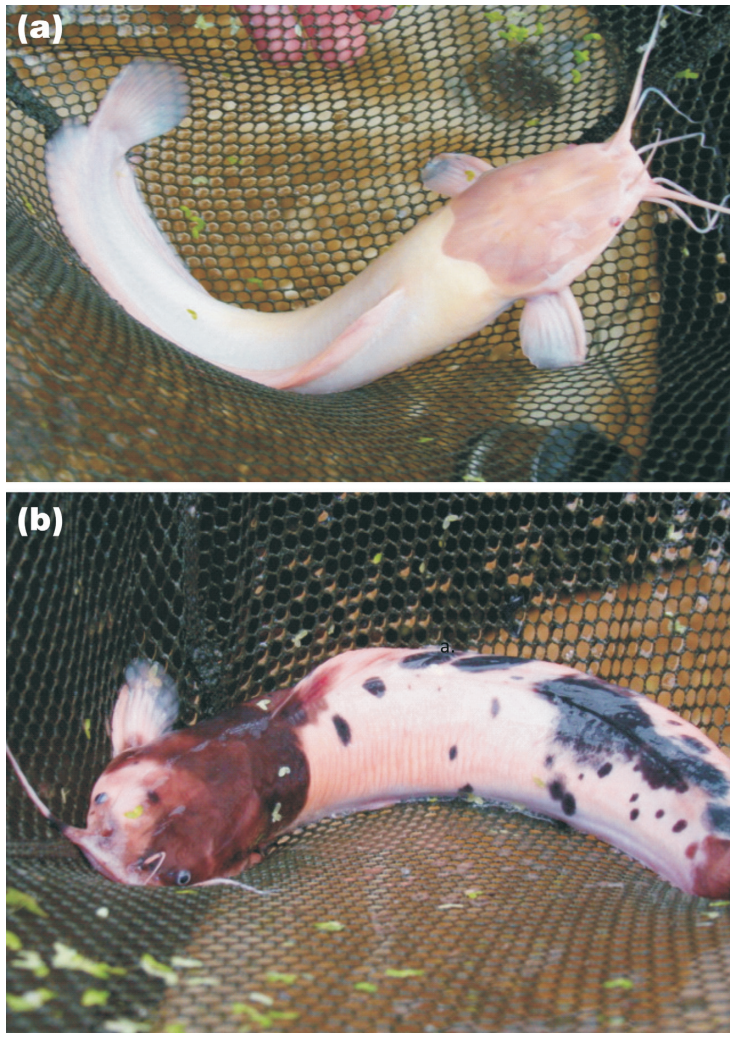

Fig. 1. Albino (**a**) and mottled (**b**) individuals of uncertain taxonomic status, observed in a population of *Clarias gariepinus*.

C. gariepinus (Fig. 1a,b). Red (albino) and mottled phenotypes of *C. gariepinus* have long been known in South Africa (Prinsloo, Schoonbee & Theron,1989; Prinsloo, Schoonbee & Hoffman, 1990). However, the colouration of these mottled individuals was subjectively different compared to the description of this phenotype in *C. gariepinus* (Skelton, 2001), while showing similarities with mottled varieties of *C. batrachus*. Since morphological examination in the field and by consulted specialists could not conclusively prove or disprove the status of these individuals as *C. gariepinus*, genetic screening was considered necessary.

Forensic application of genetic techniques to southern African fish samples were previously reported by Flint *et al*., 1998 (involving orange roughy *Hoplostethus atlanticus*) and Grobler, Kotze, Swart & Hallerman (2005) (misappropriated koi carp *Cyprinus carpio*). African catfish have been the subject of a relatively large number of genetic studies, with published papers on the genetic structure of wild populations (Van der Bank, Grobler & Du Preez, 1992; Roodt-Wilding *et al*., 2010) as well as captive populations (*cf.* Grobler, Du Preez & Van der Bank,1992, and subsequently many others). In this study, we used mitochondrial DNA (mtDNA) markers to assign two of the unknown catfish to their most likely species.

MATERIALS AND METHODS

We removed muscle samples from two of the culled unidentified fish, and designated these *Clarias* x. albino and *C.*x. mottled. DNA was isolated using the High Pure PCT Template Preparation kit (Roche Diagnostics) and quantified using a Nanodrop Lite spectrophotometer (Thermo Scientific). For the genetic characterization of the two catfish samples, variation at the mitochondrial control region was investigated. This region was chosen because a significant amount of reference material for this gene is available from the GenBank database, for several catfish species. Furthermore, this region is regularly sequenced in an on-going catfish genetics project in the laboratories at the University of the Free State, and techniques are thus well established and the region is known to be polymorphic. The D-loop region was amplified using the primers L16473 (5′ -CTA AAA GCA TCG GTC TTG TAA TCC-3′) and H355 (5′-CCT GAA ATG AGG AGG AAC CAG ATG- 3′) (Nazia, Suzana, Azhar, Thuy Nguyen & Siti Azizah, 2010). PCR was performed using KAPA 2G Robust HotStart Readymix (Kapa Biosystems), 0.5 µM of each primer and 50 ng of DNA, with cycling conditions as described by Nazia *et al*. (2010). Sequencing reactions were performed with the ABI Big Dye Terminator v3.1 kit (Life Technologies) and standard protocols. An Applied Biosystems 3130 Genetic Analyzer was used to determine the nucleotide composition of the two sequences. MEGA (Tamura, Dudley, Nei & Kumar, 2007) and Geneious software were used for the visual inspection of sequences, alignment and trimming of sequences (Geneious version 5.4, created by Biomatters; available at http://www.geneious.com).

To establish the identity of the two unknown catfish, we first compared the D-Loop sequences generated to individuals contained on the GenBank database using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For catfish, GenBank contains a significant number of sequences for reference species, including African catfish (*C. gariepinus)* and several Asian species. Best presented among the latter are *C. batrachus*, the broad-head catfish (*C. macrocephalus*) and the Hong Kong catfish (*C. fuscus*).

MEGA software was used to construct a phylogenetic tree showing the relationship between the unknown samples and the available reference material. In this analysis, we compared the two unknown samples to the following reference samples: (i) 41 farmed *C. gariepinus* from the Netherlands, sourced from Africa (these represents all *C. gariepinus* sequences available on GenBank at the time of analysis); nine *C. gariepinus* from Kenya; sequences of five random *C. gariepinus* individuals from across South Africa, supplied by a researcher at Stellenbosch University; and three *C. gariepinus* from Bloemhof Dam, South Africa, sequenced at UFS. We also included 51 *C. batrachus,* 15 *C. macrocephalus* and one *C. fuscus* from GenBank (all available individuals for these species on GenBank). We used a maximum likelihood approach to construct phylogenetic trees, with 1000 bootstrap replications, after determining the model that best fit the model of nucleotide substation in our dataset (from MEGA). A haplotype of the bullhead catfish *Ameiurus natalis*, a catfish species from the North American continent, was used as outgroup (GenBank accession no. KF-410960.1).

To quantify the level of genetic differentiation between populations or species, the two unknown samples were also compared to various reference groupings using standard molecular indexes of differentiation. The groups were: all *C. gariepinus* from South Africa, all *C. gariepinus* from GenBank, all *C. gariepinus* from Kenya, all *C. batrachus*, all *C. macrocephalus* and the single *C. festus*. Genetic differentiation among species was determined as the average number of nucleotide substitutions per site between populations (D_{xy}), the number of net nucleotide substitutions per site between populations (D_a), and the corrected average number of nucleotide differences between populations (Pi_{xY}), using DNA SP (Rozas *et al*., 2003) and Arlequin (Excoffier, Laval & Schneider, 2005) software.

RESULTS

Approximately 350 base-pairs of the mtDNA control region were amplified in each individual. This was trimmed to 281 base-pairs to allow alignment with reference sequences. The sequences between the albino and mottled individuals were very similar, and differed at only two out of 281 bases (with the number of base differences per site equal to 0.004). A visual inspection of sequence diversity can be informative and in this case, scrutiny suggests that there are notable differences between the unknown samples and *C. gariepinus*, combined with significant simi-

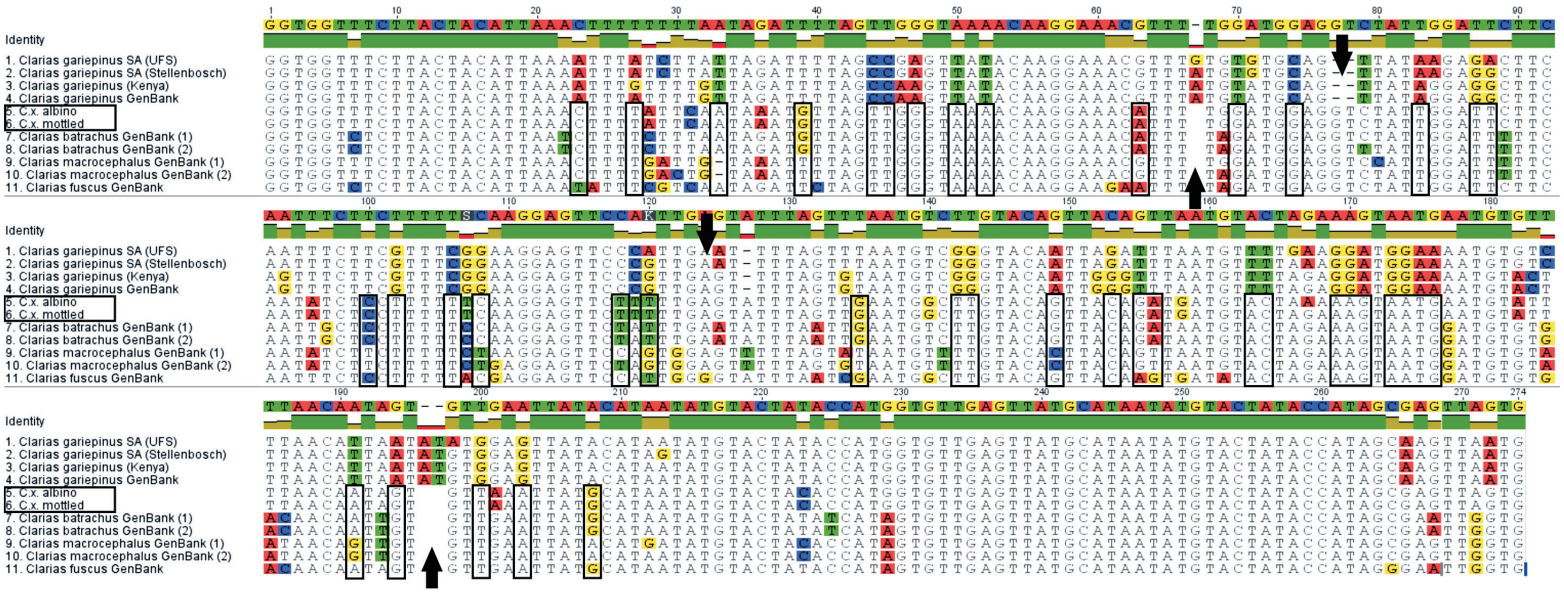

Fig. 2. Overall view of 11 of sequences used, from Geneious software, with haplotypes of four representatives of *Clarias gariepinus* above, followed by two unknown individuals, and with five Southeast Asian catfish species below. Arrows at positions 67, 77–78, 127 and 196–197 show positions with insertions/deletions present in Asian catfish and the unknown individuals, but absent from *C. gariepinus*. Areas in vertical rectangular blocks show 42 positions where the unknown individuals share a base with *C. batrachus*, with the relevant base always absent from *C. gariepinus*.

Table 1. Most likely identity of two unidentified *Clarias* individuals, based on matching of D-Loop sequences to entries the GenBank database, using the Basic Local Alignment Search Tool (BLAST).

Unknown	Match / Rank	Identity	Species
Albino	1	94%	*C. macrocephalus*
	2–51	89–90%	*C. batrachus*
	52–68	83–88%	*C. macrocephalus / C. fuscus*
	69–100	77%	*C. gariepinus*
Mottled	1	95%	*C. macrocephalus*
	2–51	89–90%	*C. batrachus*
	52–68	84–88%	*C. macrocephalus / C. fuscus*
	69–100	76–78%	*C. gariepinus*

larity between the unknown samples and the Asian catfish. Figure 2 shows a snapshot from Geneious software, based on 11 of the sequences used, including five haplotypes of the three Southeast Asian catfish species, the two unknown individuals and four representatives of *C. gariepinus* (from South Africa and Kenya and a sequence from GenBank). Insertions/deletions present in Asian catfish and the unknown individuals, but absent from *C. gariepinus*, occurred at six positions. Bases shared between the unknowns and *C. batrachus* but always absent from *C. gariepinus* occurred at 42 positions. This pattern of polymorphism persisted in the full dataset of 128 haplotypes. It is evident that there are large overlaps in areas of polymorphism between the unknown individuals and *C. batrachus*, distinct from polymorphisms specific for *C. gariepinus*.

Results of the BLAST comparison with sequences on GenBank, showing the most likely matches, are summarized in Table 1. For the albino individual, the best BLAST matches were: *C. macrocephalus* (83–94% identity), *C. batrachus* (89–90%), *C. fuscus* (87%) and *C. gariepinus* (76–77%). Results for the mottled individual were almost identical, with *C. macrocephalus* (84–95% identity), *C. batrachus* (89–90%), *C. fuscus* (88%) and *C. gariepinus* (76–78%).

The pattern of substitution in the dataset was best described by a Hasegawa-Kishino-Yano (HKY) model of nucleotide substitution. In the resulting phylogram, the 126 individuals group into two distinct clusters, reflecting geographic origin from Africa or Asia. Figure 3 shows a condensed phylogram, with the major clusters collapsed to produce a phylogram of manageable size suitable for publication. One cluster (with 100% bootstrap support) contains exclusively known *C. gariepinus* individuals, whereas the second cluster (with 90% bootstrap support) contains all the examples of the three Southeast Asian catfish species. In the latter, all individuals of *C. batrachus* ($n = 51$) cluster together with strong bootstrap support (95%), while 13 out of 15 haplotypes for *C. macrocephalus*

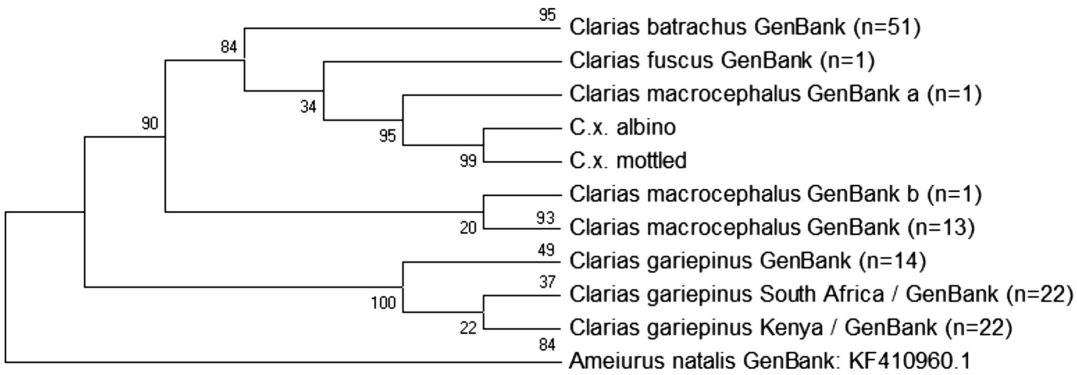

Fig. 3. Collapsed maximum likelihood tree based on unknown, African and Southeast Asian catfish. Asian and African catfish cluster together with 90% and 100% bootstrap support, respectively. The position of unknown individuals *C.*x. albino and *C.*x. mottled suggest close identity with Asian catfish.

Table 2. Genetic differentiation between (i) pairwise combinations of haplotypes of the unknowns and six groups of African and Asian catfish; and (ii) between south African catfish and the remaining five groups. Differentiation is expressed as average number of nucleotide substitutions per site between populations (D_{xy}), number of net nucleotide substitutions per site between populations (D_a), and corrected average number of nucleotide differences between populations (Pi_{XY}).

Population pair		D_{xy}	D_a	Pi_{XY} corrected
Unknown with:	*C. gariepinus*, South Africa	0.262	0.244	69.714
	C. gariepinus GenBank	0.263	0.231	66.593
	C. gariepinus Kenya	0.266	0.259	73.583
	C. batrachus	0.102	0.099	26.940
	C. macrocephalus	0.156	0.129	50.633
	C. fustus	0.162	0.159	58.000
C. gariepinus South Africa with:	*C. gariepinus* GenBank	0.057	0.012	3.270
	C. gariepinus Kenya	0.069	0.048	12.936
	C. batragus	0.270	0.253	72.199
	C. macrocephalus	0.216	0.175	66.223
	C. fustus	0.256	0.239	82.589

form a cluster with 93% bootstrap support. The two unknown samples group with the Asian group, interspersed between the *C. batrachus* and *C. macrocephalus* clusters.

To gauge the accuracy of GenBank records, we screened the phylogram for unexpected groupings. Two of the haplotypes for *C. macrocephalus*, labelled *C. macrocephalus* a & b in Fig. 3, appear to be outliers. Visual inspection of these sequence show many polymorphisms not found in any other haplotype for *C. macrocephalus*. The number of base pair differences per site between these two individuals and the remaining 13 haplotypes for *C. macrocephalus* also exceed the number within the 13 haplotypes that cluster together with high bootstrap support by an order of magnitude (0.145 *vs* 0.016).

Pairwise comparisons between the two unknowns and the reference populations to elucidate genetic differentiation (Table 2) showed that the unknowns were more similar to *C. batrachus* (D_{xy} = 0.102), *C. macrocephalus* (0.156) and *C. fuscus* (0.162) than to any group of *C. gariepinus* (D_{xy} = 0.262–0.266). To test this finding, known South African catfish were also compared to the same panel. As expected, these individuals showed low genetic differentiation from all other groupings of *C. gariepinus* (D_{xy} = 0.057–0.069) but much more differentiation compared with Southeast Asian catfish (D_{xy} = 0.216–0.270). Similar trends were observed for the other measures of differentiation used, D_a and Pi_{XY}.

DISCUSSION

Taking all the results from diverse analytical approaches into consideration, we believe that there is little evidence to support the claim that the two unknown samples are *C. gariepinus*. The conclusion of Asian origin or ancestry is firstly based on the outcome of the BLAST approach, where (i) the first 68 matches in each case were not *C. gariepinus*; and (ii) the level of identity with Asian catfish was 83–95% for compared to 76–78% for *C. gariepinus*. A comparable approach based on sequenced samples, GenBank records and the BLAST tool was recently used successfully by Cawthorn, Steinman & Hoffman (2013) to identify species substitution and mislabelling in the South African meat industry. In catfish, Wong *et al*. (2011) used a barcoding approach based on the cytochrome oxidase I (COI) gene and the GenBank and Barcode of Life Data Systems (BOLD) databases. Our approach to species identification is thus valid and the outcome lends strong support to the conclusion that the unidentified fish are not *C. gariepinus*.

We note that the entries on GenBank may not all be accurate due to possible sequencing errors (*cf.* Harris, 2003) and species misidentification. In the current analyses, GenBank entries grouped according to continent and species in the phylogram, with the only outliers being the haplotypes labelled *C. macocephalus* a & b in Fig. 3. Further investigation revealed many unique polymorphisms in these two haplotypes and a high number of base pair differences per site between these two individ-

uals and the remaining haplotypes for *C. macrocephalus*. We thus conclude that these individuals were sequenced or identified incorrectly, or alternatively represent individuals with significant genetic drift compared to the main dataset for the species. Nevertheless, we submit that the rest of the data, excluding the two outliers, is sufficiently accurate for comparative purposes.

Results of the phylogenetic analysis supported trends from BLAST-based analysis. In this analysis, the 126 unidentified and control individuals grouped into two distinct clusters, with strong bootstrap support for these two clusters (90–100%). Memberships of the two clusters are strictly according to geographic origin, with one cluster populated exclusively by *C. gariepinus* from Africa, and the second cluster containing only examples of the three Southeast Asian catfish species. The two unknown samples group with the Southeast Asian samples. Additional evidence for the unknowns not being *C. gariepinus* comes from an analysis of genetic differentiation, based on pairwise comparisons between the two unknowns and the reference populations from Africa and Southeast Asia. Results showed that the unknowns were more similar to Asian catfish than to local C*. gariepinus*, based on commonly used statistical measures of genetic differentiation. Finally, a visual screening of D-Loop sequences shows that there are major differences between patterns of polymorphism of the unknown samples and *C. gariepinus*, along with substantial similarity between the unknown samples and the Asian catfish.

Overall, there is thus strong evidence that the two unidentified fish belong to a Southeast Asian species, probably *C. batrachus*. The latter species is identified as the most likely identity based on the visual inspection of sequence diversity, the coefficients of differentiation used and the pattern of clustering in the phylogram; but also due to the known popularity of *C. batrachus* in the ornamental fish trade. It is also possible that the unidentified individuals could be hybrids between male *C. gariepinus* and female *C. batrachus*, with the mtDNA haplotype reflecting the maternal parents. Hybridization between *C. gariepinus* and *C. batrachus* (as well as *C. gariepinus* and *C. macrocephalus)* is possible and has been implemented intentionally for stock improvement of Asian catfish, as described by Rahman, Bhadra, Begum, Islam & Hussain (1995) and Senanan, Kapuscinski, Na-Nakorn & Miller (2004). Nuclear markers will be necessary to fully investigate the possibility of hybridization. We note that the two unknown samples group close to *C. batrachus*, but are not intermingled with this species. In the visual comparison of sequences (Fig. 2), it is also evident that there are also some unique polymorphisms in the unknowns compared to other Asian catfish. This is most likely due to genetic drift during the domestication history of the ancestors of these specific catfish. A degree of uniqueness does, however, not detract from the fact that the unknowns show much higher identity to Asian catfish compared to their identity with *C. gariepinus*.

As far as could be ascertained, this is the first attempt to identify catfish in South Africa using sequences of the mtDNA control region in a forensic application. The results from this project show that the danger of *C. batrachus* entering South Africa and invading its rivers should not be underestimated. The results also confirm that routine use of regular sequencing techniques and data available on GenBank can make a valuable contribution to the detection of invasive species. Alternatively, an approach based on CO1-based barcodes and the BOLD database could potentially also contribute to detecting invasive *C. batrachus*.

ACKNOWLEDGEMENTS

We thank Rouvay Roodt-Wilding from the University of Stellenbosch who kindly provided haplotypes for a number of South African individuals; and the University of the Free State for funding and facilities used. We thank Bettine Janse van Vuuren and an anonymous referee for valuable comments on an earlier version of this paper.

REFERENCES

Agnese, J.F. & Teugels, G.G. (2005). Insight into the phylogeny of African Clariidae (Teleostei, Siluriformes): Implications for their body shape evolution, biogeography, and taxonomy. *Molecular Phylogenetics and Evolution*, 36, 546–553.

Courtenay, W.R. (1978). Additional range expansion in Florida of the introduced walking catfish. *Environmental Conservation*, 5, 273–275.

Cawthorn, D-M., Steinman, H.A. & Hoffman, L.C. (2013). A high incidence of species substitution and mislabelling detected in meat products sold in South Africa. *Food Control*, 32, 440–449.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin, version 3.0: an integrated software package for population genetics data analysis. Computational and Molecular Population Genetics Laboratory, Institute of Zoology, University of Berne, Switzerland.

Flint, N.S., van der Bank, F.H., Theron, P.J. & Staby, A. (1998). Genetic variation in two populations of the orange roughy (*Hoplostethus atlanticus*) from Namibia. *South African Journal of Aquatic Science*, 24, 71–83.

Grobler, J.P., Du Preez, H.H. & Van der Bank, F.H. (1992). A comparison of growth performance and genetic traits between four selected groups of African catfish (*Clarias gariepinus*). *Comparative Biochemistry and Physiology*, 102A, 373–377.

Grobler, J.P., Kotze, A., Swart, H. & Hallerman, E.M. (2005). The application of microsatellite DNA markers for forensic analysis of koi carp (*Cyprinus carpio*). *South African Journal of Science*, 101, 19–21.

Harris, D.J. (2003). Can you bank on GenBank? *Trends in Ecolology and Evolution*, 18, 317–319.

Hoffman, L.C., Swart, J.J. & Brink, D. (2000). The 1998 production and status of aquaculture in South Africa. *Water SA*, 26, 133–135.

Huisman, E.A. & Richter, C.J.J. (1987). Reproduction, growth, health control and aquaculture potential of the African catfish, *Clarias gariepinus* (Burchell 1822). *Aquaculture*, 63, 1–14.

Lowe, S., Browne, M., Boudjelas, S. & De Poorter, M. (2004). 100 of the world's worst invasive alien species: a selection from the Global Invasive Species Database. Invasive Species Specialist Group (ISSG), a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN) (first published as special lift-out in Aliens 12, December 2000, updated and reprinted November 2004).

Nazia, A.K., Suzana, M., Azhar, H., Thuy Nguyen, T.T. & Siti Azizah, M.N. (2010). No genetic differentiation between geographically isolated populations of *Clarias macrocephalus* Gunther in Malaysia revealed by sequences of mtDNA cytochrome-b and D-loop gene regions. *Journal of Applied Ichthyology*, 26, 568–570.

Prinsloo, J.F., Schoonbee, H.J. & Theron, J. (1989). The use of a red strain of the sharp tooth catfish *Clarias gariepinus* Burchell in the evaluation of cannibalism among juveniles of this species. *Water SA*, 15, 179–184.

Prinsloo, J.F., Schoonbee, H.J. & Hoffman, L.C. (1990). A comparison of the fecundity of two strains of the sharptooth catfish *Clarias gariepinus*. *South African Journal of Wildlife Research*, 20, 100–103

Rahman, M.A., Bhadra, A., Begum, N., Islam, M.S. & Hussain, M.G. (1995). Production of hybrid vigor through cross breeding between *Clarias batrachus* Lin. and *Clarias gariepinus* Bur. *Aquaculture*, 138, 125–130

Roodt-Wilding, R., Swart, B.L. & Impson, N.D. (2010). Genetically distinct Dutch-domesticated *Clarias gariepinus* used in aquaculture in southern Africa. *African Journal of Aquatic Science*, 35, 241–249.

Senanan, W., Kapuscinski, A.R., Na-Nakorn, U. & Miller, L.M. (2004). Genetic impacts of hybrid catfish farming (*Clarias macrocephalus* × *C. gariepinus*) on native catfish populations in central Thailand. *Aquaculture*, 235, 167–184.

Skelton, P. (2001). A complete guide to the freshwater fishes of southern Africa. Struik Publishers, Cape Town.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, 1596–1599.

Van der Bank, F.H., Grobler, J.P. & Du Preez, H.H. (1992). A comparative biochemical genetic study of 3 populations of domesticated and wild african catfish (*Clarias gariepinus*). *Comparative Biochemistry and Physiology*, 101b, 387–390.

Van der Waal, B.C.W. (1978). Some breeding and production experiments with *Clarias gariepinus* (burchell) in the transvaal. *South African Journal of Wildlife Research*, 8, 13–17.

Wong, L.L., Peatman, E., Lu, J., Kucuktas, H., He, S., Zhou, C. Na-Nakorn, U. & Liu, Z. (2011). DNA barcoding of catfish: species authentication and phylogenetic assessment. *PLOS ONE*, 6, e17812.

Responsible Editor: L.C. Hoffman