scientific reports

OPEN

MATTERS ARISING

Best BLAST hit alone cannot be used as evidence of fraud

Natalia Díaz-Arce ✉ & Naiara Rodríguez-Ezpeleta ✉

In a recent study, Blanco-Fernandez et al.[1] applied molecular tools to authenticate fish products and conclude evidence of "worrying international fraud". They revealed mislabeling in recognizable and unrecognizable fish products labeled as anchovy, hake and tuna commercialized by European companies. Their analyses consisted of extracting DNA from the fish product to be authenticated followed by amplification and sequencing of a suite of DNA markers and comparing the resulting sequences to the GenBank sequence database using BLAST (Basic Tool Alignment Search Tool) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). By carefully reanalyzing their data, we identify errors in their identification of tuna species and conclude that best BLAST hit alone cannot be used as evidence of fraud.

Seafood product traceability is essential to detect intentional or unintentional mislabeling and thus helps reduce unreported, unregulated and illegal fishing while enhancing consumer safety[2]. Genetic methods have shown to be powerful for seafood product traceability[3], in particular, when morphological characteristics cannot be confidently used such as in young age specimens (e.g., juveniles of bigeye and yellowfin tunas are very difficult to distinguish[4]) or in closely related species (e.g. black and white anglerfish[5]), and especially in processed products (e.g. filleted and canned), where anatomical traits important for fish identification (e.g. head, fins, skin) are absent. Accurate genetic based seafood product traceability requires developing approaches that unequivocally discriminate between species, for which it is essential to understand intra-specific variability as well as each species' evolutionary context. In their study, Blanco-Fernandez, et al.[1] use the best hit of a BLAST search against GenBank to assign species to the sample to be authenticated. Here, by analyzing their BLAST results considering other information, we show that using BLAST alone can lead to erroneous species assignments and thus to conclude fraud when there is not.

One of the fraud detections from Blanco-Fernandez et al. is the substitution of albacore (*Thunnus alalunga*) by Atlantic bluefin tuna (*Thunnus thynnus*), which is more than twice as expensive as albacore tuna (https://www.eumofa.eu/es/home). The authors explain this potential substitution by over-quota-caught bluefin tuna being sold as another species. This is a strong claim that needs clear evidence to be made. We thus examined the sequences corresponding to those albacore-labelled tuna products (MW557512, MW557513, MW557514) claimed to be mislabelled bluefin tuna due to a best BLAST hit with sequence EU562888, belonging to *T. thynnus* according to GenBank. Our hypothesis was that the mislabeling, rather than in the seafood products, is in the sequence in GenBank due to the mitochondrial introgression reported between *T. alalunga* and *T. thynnus*[6]. Indeed, it has been estimated that approximately 2–3% of *T. thynnus* individuals have the so-called "alalunga-like" mitochondrial DNA[7], which has often misled mitochondrial based phylogenetic inferences of the genus *Thunnus*[8]. This hypothesis was confirmed by a phylogenetic inference including the putatively mislabelled sequences from Blanco-Fernandez, et al.[1] and their best BLAST hit, as well as representative sequences from *T. alalunga*, *T. thynnus* (including those labelled as "alalunga-like") and *T. albacares* (Fig. 1). The tree shows two clearly differentiated clades: one is exclusively composed by *T. thynnus* sequences, while the other includes *T. alalunga*, *T. thynnus* "alalunga-like" sequences as well as the sequences from the putatively mislabelled products and their best BLAST hit. These results refute the mislabeling of *T. thynnus* products as *T. alalunga* reported by Blanco-Fernandez et al.[1].

We have not found any other potential misidentification in Blanco-Fernandez et al.[1], and therefore, their claim about existing fraud in highly appreciated fish still holds. Yet, this clear and easy to detect case in tunas could be the tip of the iceberg of errors made in food fraud studies relying solely on best BLAST hits to report mislabeling. Indeed, the bluefin tuna and albacore case is not an isolated one, and instances of genetic introgression that could lead to misidentification are increasingly reported in teleost fishes (e.g.[11,12]). Additionally, other factors such as traditionally used morphological characters for species assignment not being diagnostic can also occur and lead to false conclusions regarding mislabelling. This is the case of the black and white anglerfish for which mislabelling was reported[13] based on the colour of their peritoneum as species diagnostic character, whereas it has recently been discovered that black anglerfish can have white peritoneum[14] and, thus, reported mislabelling

AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Txatxarramendi Ugartea Z/G, 48395 Sukarrieta, Bizkaia, Spain. ✉email: ndiaz@azti.es; nrodriguez@azti.es
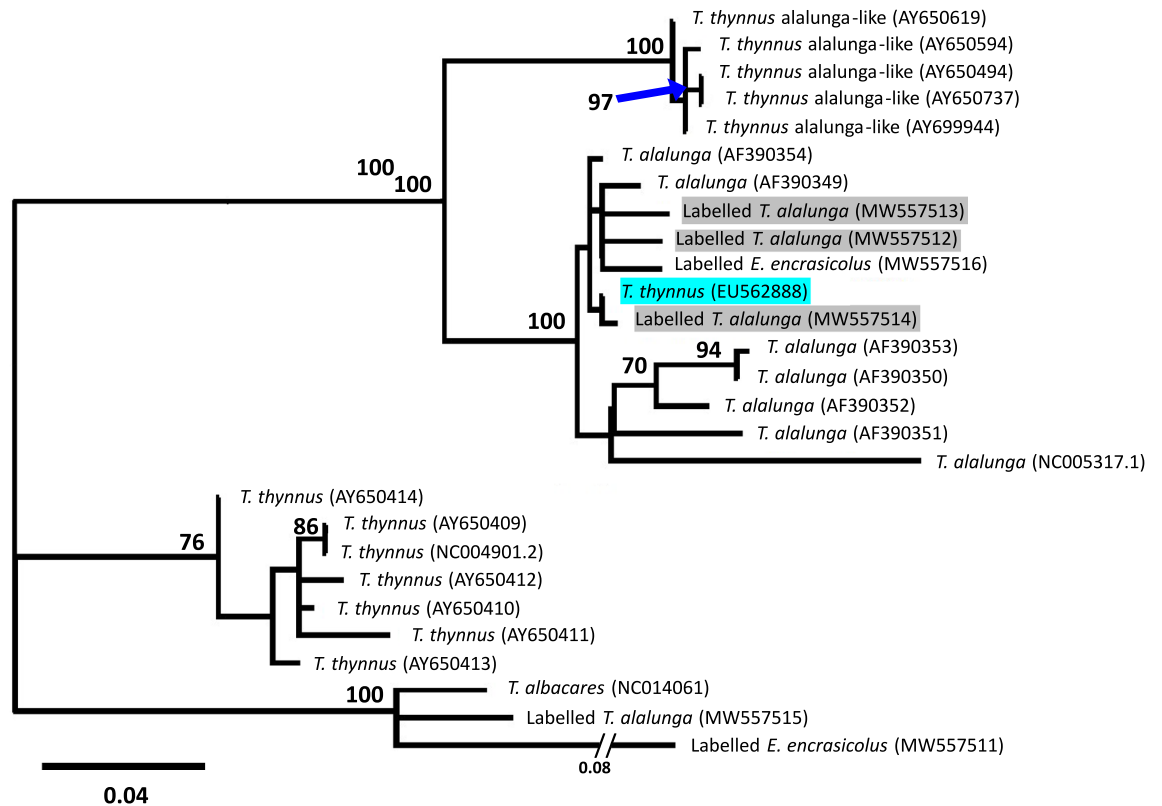
Figure 1. Maximum likelihood phylogenetic tree built using the GTRCAT model as implemented in RAxML version 8.1.21[9] based on the mitochondrial control region (positions with less than 50% missing data between 5 to 414 of NC005317 sequence) aligned using Clustal W[10] (default parameters) including all the available tuna identified sequences from Blanco-Fernandez et al. (MW557511-16) with the ones labelled *T. alalunga* matching to *T. thynnus* labelled in grey, the best BLAST hit *T. thynnus* sequence (EU562888), shaded in blue, and reference sequences from *T. albacares* (NC014061), *T. alalunga* (AF390349-54), *T. thynnus* (AY650409-14 and NC004901) and *T. thynnus* "alalunga-like" (AY650494, AY650594, AY650619, AY650737, AY699944). The tree is unrooted and branch support was assessed by a 100-replicate rapid-bootstrap analysis. Bootstrap values of those branches over or equal to 70% are indicated.

was most likely not so. As shown above, questioning our understanding of the evolutionary history of the species under investigation is essential for seafood fraud studies and could avoid errors such as the one made by Blanco-Fernandez et al. when they report mislabeling of albacore products.

Additionally, BLAST results deeply depend on the accuracy of the reference databases, and presence of contaminated sequences has been reported in GenBank[15]. Blanco-Fernandez et al.'s work increase the contaminations in GenBank as they have contributed sequences whose taxonomic assignment has relied on best BLAST hit. As consequence of this, there are now three sequences of *T. alalunga* in GenBank labelled *T. thynnus*. These sequences, as well as those not obtained from morphologically identifiable specimens, should be retracted from GenBank to avoid further ramifications.

In summary, we conclude that best BLAST hit cannot be used as evidence of fraud, and that studies on seafood authentication should consider the evolutionary context of the species under study. Not doing so can result in serious consequences as illustrated by the problems we found in the work of Blanco-Fernandez et al., who base their claims on tuna mislabeling trends in Spain on erroneous taxonomic assignments. Finally, besides using phylogenetic inference instead of BLAST search for genetic-based seafood authentication, we propose the generation of custom, tailored and curated reference sequence libraries specific for each case study that should be made publicly available. We recommend checking these reference libraries using phylogenetic quality control to detect misidentified or dubious sequences, and testing for adequate coverage for important species. In addition, we advise to review existing literature reporting known cases of interspecific hybridization or haplotype sharing involving included or closely related species. Finally, in order of avoid the inclusion of spurious sequences in public databases, we recommend the submission of sequence data produced only from identified species or to upload them as environmental samples otherwise.

Data availability
All data analysed for this reply have been downloaded from GenBank using the accession numbers provided in the figures.

References

1. Blanco-Fernandez, C. *et al.* Fraud in highly appreciated fish detected from DNA in Europe may undermine the development goal of sustainable fishing in Africa. *Sci. Rep.* **11**, 11423. https://doi.org/10.1038/s41598-021-91020-w (2021).
2. Lewis, S. G. & Boyle, M. The expanding role of traceability in seafood: Tools and key initiatives. *J. Food Sci.* **82**, A13–A21. https://doi.org/10.1111/1750-3841.13743 (2017).
3. Hellberg, R. S., Pollack, S. J. & Hanner, R. H. In *Seafood Authenticity and Traceability* (eds Amanda M. Naaum & Robert H. Hanner) 113–132 (Academic Press, 2016).
4. Nakamura, I. & Séret, B. Field identification key of tunas of the genus Thunnus. *Cybium* **26**, 141–145 (2002).
5. Caruso, J. Lophiidae. *Fishes N. East. Atl. Mediterr.* **3**, 1362–1363 (1986).
6. Viñas, J. & Tudela, S. A validated methodology for genetic identification of tuna species (genus Thunnus). *PLoS ONE* **4**, e7606 (2009).
7. Alvarado Bremer, J. R., Viñas, J., Mejuto, J., Ely, B. & Pla, C. Comparative phylogeography of Atlantic bluefin tuna and swordfish: The combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Mol. Phylogenet. Evol.* **36**, 169–187. https://doi.org/10.1016/j.ympev.2004.12.011 (2005).
8. Díaz-Arce, N., Arrizabalaga, H., Murua, H., Irigoien, X. & Rodríguez-Ezpeleta, N. RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. *Mol. Phylogenet. Evol.* **102**, 202–207. https://doi.org/10.1016/j.ympev.2016.06.002 (2016).
9. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 (2014).
10. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680. https://doi.org/10.1093/nar/22.22.4673 (1994).
11. Horoiwa, M. *et al.* Mitochondrial introgression by ancient admixture between two distant lacustrine fishes in Sulawesi Island. *PLOS ONE* **16**, e0245316. https://doi.org/10.1371/journal.pone.0245316 (2021).
12. Shum, P. & Pampoulie, C. Molecular identification of redfish (genus Sebastes) in the White Sea indicates patterns of introgressive hybridisation. *Polar Biol.* **43**, 1663–1665. https://doi.org/10.1007/s00300-020-02718-y (2020).
13. Espiñeira, M., González-Lavín, N., Vieites, J. M. & Santaclara, F. J. Authentication of anglerfish species (Lophius spp.) by means of polymerase chain reaction–restriction fragment length polymorphism (PCR–RFLP) and forensically informative nucleotide sequencing (FINS) methodologies. *J. Agric. Food Chem.* **56**, 10594–10599. https://doi.org/10.1021/jf801728q (2008).
14. Aguirre-Sarabia, I. *et al.* Evidence of stock connectivity, hybridization, and misidentification in white anglerfish supports the need of a genetics-informed fisheries management framework. *Evol. Appl.* **14**, 2221–2230. https://doi.org/10.1111/eva.13278 (2021).
15. Steinegger, M. & Salzberg, S. L. Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* **21**, 115. https://doi.org/10.1186/s13059-020-02023-1 (2020).

Acknowledgements

Author contributions

N.D.A. and N.R.E. conceived the idea, conducted the analyses, and wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.D.-A. or N.R.-E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.