

Deadline of Submission of Comments: 22-May-23

Document Number: ANSI/ASB Std 180

Document Title: Standard for the Use of GenBank for Taxonomic Assignment of Wildlife

#	Section	Type of Comment (E-Editorial, T-Technical)	Comments	Proposed Resolution	Final Resolution
1	1	T	Concern that the term "conceptually" is too broad	However, relevant criteria can be applied to other sequence databases.	Reject: Many criteria involve terminology specific to GenBank, such as an "e-value." Other databases, such as BOLD, may not have e-values, so the criteria citing them would not be "relevant." BOLD does, however, provide some measure of goodness of fit, so the criteria could be used conceptually with the existing language.
2	1	T	In Section 4.1 it states minimum requirements, however this is not captured in the scope. This should be addressed up front.	This standard provides the minimum requirements.....	Accept
3	5.1	T	Issue with the term "close enough" in the last sentence which is arguably subjective. Additionally, the additional details provided specifically for 5.6 is out of place given the brief descriptors given for 5.1-5.5.	Remove the following sentence: "The evaluation of the hit(s) should take into account the sequence quality, whether it is attributed to the correct species, and whether the hit(s) is a close enough match for the taxon in question."	Reject: Whether a hit is "close enough" will depend on use of validated methods, training, and experience (all 4.3), as well as evaluation of inter- and intra-specific distances and completeness of sampling for the taxon in question. 5.1 is meant to generally outline the considerations to come, not provide specific guidance. One of the most important considerations is whether a hit is "close enough," (see 5.6.1 b) 3 and 4, with additional guidance given in the following note). Also see 4.1.2, which explicitly states that there is no single threshold for "close enough" across all genes and the tree of life. "Close enough" is a critical consideration, and each taxon and locus will require independent assessment to determine that value
4	5.2b	T	These seems to be repetitious and vague with the use of quality criteria.	"shall meet quality criteria as identified by the laboratory (see ASB048)." Remove the following text "including the following: - clearly defined criteria, ensuring their - analysts adhere to those defined criteria;"	Accept with Modification: "5.2 b) shall meet quality criteria as defined by the laboratory (see ASB048), "
5	5.2c	T	For non-coding regions, there is no need to translate to check for premature stop codons (i.e., control region sequences). Be specific here with which sequences this applies to. Also this wording should be consistent with that given in 5.3f	"shall ensure the hit(s) from a protein coding region does not contain premature stop codons."	Accept with Modification: 5.2c now reads "c) from a protein coding region shall be examined to ensure it does not contain premature stop codons (e.g., by translation)." 5.2c is in reference to the laboratory's query sequence, not about hits retruned. Stop codons in non-coding regions are not "premature," but we will further clarify
6	5.3b	T	If you are not using data, that should be recorded in the case record (as per ASB 3125 7.5.1.5). That detail should be added here.	"shall not use any hit(s) that is an anomaly among the returned results and shall be noted in the case record"	Reject: AR3125 7.5.1.5, is in relation to technical records, states that "If an observation, data, or calculation is rejected, the reason, the identity of the individual(s) taking the action and the date shall be recorded in the technical record." BLAST hits are sequences that are not generated by the laboratory, and the composition of hits from a subsequent search of the same seequence can change as new data are uploaded to Genbank, or as original data are revised by submitters. Therefore, they are not Technical Records, and AR 3125 7.5.1.5 does not apply
7	5.4.1c (2)	T	5.4.1.c2 is essentially decribing a voucher, and it is very difficult to glean from published literature who completed the identification and whether or not that person is an expert.	Remove 5.4.1.c2.	Reject: It can be difficult to determine if data was derived from a voucher, but it is not always impossible. If the hit on which conclusions are based was from a voucher, it offers strong support. If one can't determine voucher status, it confers moderate support (5.4.2a)

8	5.4.2a	T	To be consistent with the suggested removal of 5.4.1c2 need remove cross reference to 5.4.1c	Remove the following text: "(5.4.1c)".	Reject: It can be difficult to determine if data was derived from a voucher, but it is not always impossible. If the hit on which conclusions are based was from a voucher, it offers strong support. If one can't determine voucher status, it confers moderate support (5.4.2a)
9	5.5	T	This section should say something about what it actually is.	Add immediately after 5.5 the following text: "This section is to assess whether locus selection and taxonomic representation is appropriate."	Accept with Modification: Section revised to read: "This section is to assess whether locus selection and taxonomic representation is appropriate, and if the taxa in question are well-separated."
10	5.6.1b(2)	E	The text "(as defined in 5.4.2)" is in an awkward place compared to the rest of the document	Move the text "(as defined in 5.4.2)" to the end of the sentence (after the text "support criteria") to match the format of the rest of the document	Reject: (5.4.2) should follow "moderate" closely, so it can't be read as about both "strong and moderate criteria"
11	5.2. b)	T	There is the potential for different quality criteria across labs.	Delineate more clearly what quality criteria are central and how the lab ensures that the analyst adheres to those criteria	Reject: Quality criteria differ across laboratory laboratories, due to different instrumentation and methodology. Quality criteria are also beyond the scope of this document
12	5.3. e)	T	What constitutes an ambiguous base?	Operationalize/define more clearly what constitutes an ambiguous base.	Reject: Practitioners with appropriate knowledge and training recognize ambiguous bases. It is a common term in genetics
13	5.4.1 c) & d)	E	What if there are several empirical/scientific studies with potentially conflicting results?	Replace with wording that highlights the consensus or majority of empirical studies	Reject: Familiarity with and interpretation of phylogenetic and population genetic literature rests on a practitioner's training and experience. Additionally, when targeting markers with good resolution at the species level, there is seldom major disagreement in the literature about the tips of the tree branches (species-level arrangement).
14	8.5	E	How is the "relevant taxonomic authority" to be determined? E.g., Morphological or genetic analysis and/ or reference to peer-reviewed publications? This should be specified/made more explicit Relatedly, what suggests that the relevant taxonomic authority confers validity? If there is an authoritative source or standard supporting this assertion, it should be cited here.	See suggestions in the comment.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
15	5.1	E	Without more specific guidance, it's unclear how to establish "close enough."	Provide criteria that establishes when something is "close enough."	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
16	5.2b	E	It's not clear what constitutes "clearly defined criteria." Are there specific areas that criteria must cover, such as quantity of DNA or ability to type all nucleotides in a particular sequence? Moreover, laboratories may define criteria that do not meet standards of "quality," even if those criteria are clearly defined. Finally, there is no indication of how laboratories shall "ensure their analysts adhere" to the criteria. More specificity and explanation is required on all these points.	The standard should specify what constitutes "quality criteria," what and how much detail must be provided to be deemed "clearly defined criteria," and how laboratories "ensure" their analysts are adhering to those criteria.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
17	5.3a	E	There is no indication of how laboratories shall "ensure the hits belong to..."	The standard should specify the process for ensuring that hits belong to a taxonomic category (e.g., hits must be documented and reviewed).	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
18	5.3a NOTE	E		The Note should be deleted and this information integrated into the standard by specifying what analysts should do, e.g, "Analysts must disclose situations involving a complete unknown, where it is not possible to assess the broader taxonomic group."	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
19	5.3b	E	There is no indication of how laboratories shall "ensure the hits...."	The standard should specify the process for ensuring that hits belong to a taxonomic category. Also, the standard should make clear that any anomalies should be documented and logged, for use in quality control and also possible disclosure in discovery in a criminal case.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.

20	5.3b	E	The sentence structure makes the meaning unclear. Also, is the only indication of anomaly that the species is "interleaved among many"? It seems doubtful that this is the only signal of an anomaly in the returned results.	Rewrite: "shall ensure...." this could be an indication of human error... An anomaly would be indicated by being the only representative... [and other things....]	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
21	5.3e	E	There is no indication of how the review for ambiguous bases should proceed or what should happen in ambiguous bases are discovered.	This should detail the steps required to "review" and also make clear that ambiguity must be documented and reported.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
22	5.3f	E	There is no indication of how laboratories shall "ensure the hits...."	The standard should specify the process for ensuring that hits belong to a taxonomic category.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
23	5.4	E	The standard assumes the "correct species" is known/knowable.	It should make clear that the criteria make it likely that the species has been identified, but not imply that the correct species is a certain fact. E.g., "The criteria listed below shall be used to determine the strength of the conclusion that the sequence may be attributed to the identified species...."	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
24	5.4.1b	E	What is the level of concordance that would give rise to a reasonable conclusion with respect to species, and might it be more precisely quantified? And what about expectations subject to debate/controversy in the literature?	The standard should explain what is meant by "concordant" and, further, require reports to note disagreements within the literature (e.g., no cherry-picking the studies that confirm the conclusion while ignoring those that cast doubt on it.)	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
25	5.6.1 + NOTE 1	E	The standard requires 99-100% sequence identity but the Note confuses this requirement by asserting that "lower than 99% may still warrant a species level identification."	The standard should clarify the scientific basis for the 99%-100% standard. It should delete the NOTE and make it an additional subsection: "A species level identification may be made for identities lower than 99% when [list criteria and scientific justification]..."	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
26	5.6.1 NOTE	E	This note identifies a critical consideration which should be specified as part of the standard.	Delete the NOTE and make a new subsection that requires identification of both percent identity and length of the match.	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.
27	5.6.2 d	E	Presumably "interspecific distance" refers to a statistic. The statistical procedure should be specified and the standard should make clear how one accounts for sampling error in the statistic. Further, how is one to assess whether "there is little definitive information on interspecific distance"? Is there some information? Is it possible to obtain more?	Specify the acceptable statistical procedure. Clarify how one assesses the presence/absence of "definitive information." Moreover, this should require documentation and reporting of <99% identical and any information as to interspecific distance (d.1) and any discrepancies discovered at a lower taxonomic level (d.2).	The comments were received after the comment deadline. They were provided to the consensus body for review. The consensus body voted (procedural vote) to not address the comments on June 22, 2023.