



A DNA Barcoding Protocol for Authenticating Traded Medicinal Plant Raw Materials

Ignacio Andersen^{1*}, Ivana Girma², Zarina Forde³

¹Department of Life Sciences, Pearl River Graduate Institute of Life Sciences, Chile

²Department of Life Sciences, Fjordland University Centre for Life Sciences, Italy

³Department of Life Sciences, Nejd Graduate Institute of Life Sciences, New Zealand

Article Info

Volume 1, Issue 1, 2026

doi: [10.62715/ijbsr2.2026.v1.i1.pp13-18](https://doi.org/10.62715/ijbsr2.2026.v1.i1.pp13-18)

Received: 27 January 2026

Accepted: 17 March 2026

Published: 30 May 2026

Abstract

Dried roots, barks, leaves and seeds sold as medicinal raw materials are hard to identify once they have been cut, powdered or mixed, and substitution of one species for another is a recognised problem in the herbal trade. DNA barcoding, the identification of a specimen from a short standard region of its genome, can resolve many such cases where morphology and chemistry cannot. Published studies differ widely, however, in the markers they use, in how reference sequences are obtained and in the rules by which a match is declared, and few describe a procedure that a quality-control laboratory could adopt as written. This article proposes one. The protocol has eight steps: defining the target species and its likely adulterants, building a reference library from vouchered specimens, sampling the consignment, extracting DNA from dried tissue, amplifying a primary and a supporting locus, sequencing and quality control, assigning identity by rules fixed in advance, and reporting the result with its limits. For each step the article gives the reasoning, the controls required and the record to be kept. The protocol is assembled from the published literature and from accepted laboratory practice. It has not been validated on market samples here, and the article sets out what such validation would need to show.

Keywords: DNA barcoding, Medicinal plants, Herbal drug authentication, ITS2, Adulteration, Reference library

© 2026 Ignacio Andersen et al. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

A trader's sack labelled with the name of a medicinal plant may hold the species named, a close relative, an unrelated plant that shares a vernacular name, or a blend of these. The causes are well known to pharmacognosists: collectors identify by local names that cover several species, scarce species are replaced by common ones, and material passes through many hands before it reaches a manufacturer. Conventional quality control relies on macroscopic and microscopic examination and on chromatographic fingerprints (WHO, 2011). These methods are indispensable, yet they struggle with fragmented material from closely related species whose anatomy and chemistry overlap.

* Corresponding author: Ignacio Andersen, Department of Life Sciences, Pearl River Graduate Institute of Life Sciences, Chile. E-mail: ignacio.andersen@pearl.edu

© 2026 Ignacio Andersen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

DNA barcoding was introduced as a way of identifying organisms from a short, standardised gene region compared against a library of sequences from identified specimens (Hebert *et al.*, 2003). Its application to herbal materials has grown quickly, and reviews agree both that it works for raw botanicals and that it has been reported with very uneven rigour (de Boer *et al.*, 2015; Parveen *et al.*, 2016; Mishra *et al.*, 2016). The dispute that followed the market survey of Newmaster *et al.* (2013) showed how much depends on method: critics noted that processed products may contain little intact DNA, so that a failure to detect a species need not mean that it is absent.

This article sets out a protocol for the simpler and more tractable case of crude raw material, meaning dried plant parts that have not been extracted. It is a proposal, built from published methods, and it has not been applied to samples in the course of this work. The aim is a procedure definite enough to be tested, criticised and improved.

2. Background: Choice of Marker and Its Consequences

No single region of the plant genome does for plants what the mitochondrial gene *COI* does for animals. After comparing candidates, the CBOL Plant Working Group recommended two plastid coding regions, *rbcL* and *matK*, as the core barcode for land plants (CBOL Plant Working Group, 2009). The pair is robust and well represented in databases, but it often fails to separate closely related species. Kress *et al.* (2005) had earlier proposed the plastid spacer *trnH-psbA* together with the nuclear ribosomal internal transcribed spacer (ITS). A large comparative study later concluded that ITS should be added to the core barcode for seed plants (China Plant BOL Group, 2011).

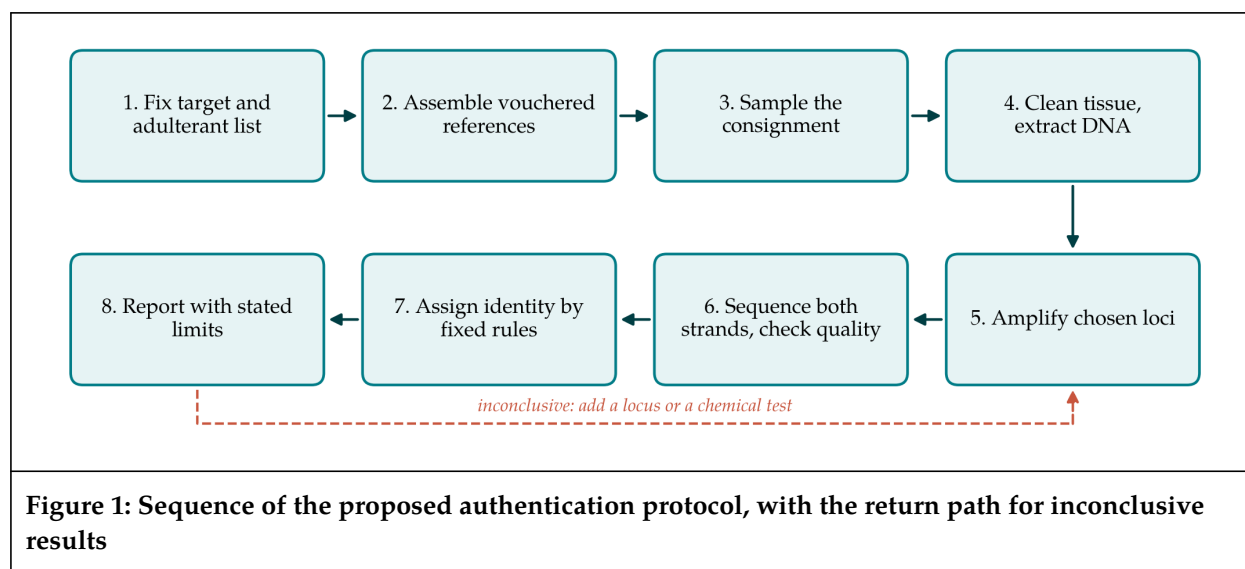
For medicinal plants the most influential proposal has been that of Chen *et al.* (2010), who tested several regions on a large set of medicinal species and recommended ITS2, the second spacer of the ITS region, as a standard. ITS2 is short, which matters greatly for dried and stored material in which DNA is broken into small fragments. Hollingsworth *et al.* (2011) review the trade-offs between markers in practical terms. Table 1 condenses the properties that bear on the present protocol.

Table 1: Candidate barcode regions, their strengths and weaknesses for dried raw material, and the role assigned to each in the proposed protocol			
Region	Strength	Weakness	Role in protocol
ITS2	Short; variable between species; suited to degraded DNA (Chen <i>et al.</i> , 2010)	Fungal sequences may amplify; multiple copies can differ	Primary locus
<i>rbcL</i>	Amplifies and aligns reliably; broad database coverage	Often identical among congeners	Supporting locus; confirms family and genus
<i>trnH-psbA</i>	Highly variable spacer (Kress <i>et al.</i> , 2005)	Length varies; repeats disturb sequencing	Alternative supporting locus
<i>matK</i>	More variable than <i>rbcL</i> ; part of core barcode	Primers fail in many groups; long fragment	Used only if primers are proven for the genus
Full ITS	More characters than ITS2 alone	Too long for badly degraded samples	Used for reference vouchers
Mini-barcode	Very short fragment recoverable from old material	Few characters; must be designed per group	Fallback when standard loci fail

Two consequences follow. A protocol for traded material should lead with a short nuclear region and support it with a plastid region of independent inheritance, so that hybridisation or contamination affecting one locus is exposed by the other. And the choice of marker cannot be separated from the reference library, since a region is useful only if verified sequences for the target and its substitutes exist.

3. The Proposed Protocol

The protocol is arranged in eight steps, shown in order in Figure 1. Steps 1 and 2 are done once for each commodity and then maintained. Steps 3 to 8 are repeated for each consignment.



3.1. Step 1: Define the target and its likely substitutes

The laboratory begins by writing down the accepted scientific name of the species the material should be, with the pharmacopoeial or regulatory source of that name, and a list of the species known or suspected to be traded in its place. The list is compiled from pharmacopoeial monographs, the ethnobotanical and market-survey literature, and the experience of buyers. It should include congeners growing in the source region and unrelated plants sold under the same vernacular name. This list defines what the test must be able to tell apart, and every later step is judged against it.

3.2. Step 2: Build a vouchered reference library

Public databases contain misidentified and unverified sequences, and an identification is only as good as the reference it matches. For each species on the list the laboratory should therefore hold sequences obtained from specimens identified by a taxonomist and deposited in a recognised herbarium, with several individuals per species from different localities so that variation within the species is represented. Sequences and voucher details are recorded in a local library and, preferably, submitted to the Barcode of Life Data System, which links each record to its specimen and trace files (Ratnasingham and Hebert, 2007). Sequences taken from public databases without a traceable voucher are treated as supporting evidence only.

3.3. Step 3: Sample the consignment

A consignment is seldom uniform. Samples are drawn from several containers and from several positions within each, following the sampling provisions used for other quality tests on herbal materials (WHO, 2011). Whole or coarsely cut material is sorted by eye first, and pieces that differ in appearance are tested as separate items. Each item receives a code that follows it through the laboratory, and a portion is retained. For powders, replicate subsamples are taken, with the caveat stated in Step 7 about mixtures.

3.4. Step 4: Clean the tissue and extract DNA

Surface contamination by fungi, soil and dust from other plant material is the main source of misleading sequences. Pieces are brushed, wiped or briefly washed, and tissue is taken from the interior where the part allows it. Extraction follows the CTAB method of Doyle and Doyle (1987) or a commercial silica-column kit, with the modifications commonly needed for tissues rich in polysaccharides and phenolic compounds, such as added polyvinylpyrrolidone and extra washing. An extraction blank is

processed with every batch. Extraction and amplification are set up in a room separate from the one in which amplified products are handled.

3.5. Step 5: Amplify a primary and a supporting locus

ITS2 is amplified first, with published plant primers of the kind derived from the universal ribosomal primers of [White et al. \(1990\)](#). A plastid region, *rbcL* by default, is amplified in parallel. Every run includes a negative control without template and a positive control from a reference specimen of a species that is not on the target list, so that carry-over of the positive control would be recognised for what it is. If neither locus amplifies after dilution of the extract to reduce inhibitors, a mini-barcode within the same region is attempted before the item is reported as having yielded no result.

3.6. Step 6: Sequence and check quality

Products are sequenced in both directions by the Sanger method. The two reads are assembled, primer sequences are trimmed, and the chromatograms are inspected. A consensus is accepted only where both strands agree and base calls are of high quality across the region compared. Double peaks along a read are recorded and not edited away, because they may signal a mixed sample or divergent ITS copies. Each consensus is screened against a general database to confirm that it is of plant origin, which removes the fungal sequences that ITS primers sometimes recover.

3.7. Step 7: Assign identity by rules fixed in advance

The query is compared with the local reference library and, as a secondary check, with public databases using BLAST ([Altschul et al., 1990](#)). The decision rule is written before testing begins. This protocol adopts the logic of the best close match criterion of [Meier et al. \(2006\)](#): the item is assigned to a species if its nearest reference sequences all belong to that species and lie within a distance threshold derived from the variation observed within species in the library. If the nearest sequences belong to more than one species the result is ambiguous, and if nothing lies within the threshold there is no match. The supporting locus must agree at least at the level of genus. Disagreement between loci is reported as such.

Sanger sequencing reads the dominant template. It cannot be relied on to reveal a minor component in a powder, and a clean sequence of the correct species does not prove the absence of an admixture. Where mixtures are the concern, sequencing of many amplicons in parallel, as in DNA metabarcoding ([Taberlet et al., 2012](#)), is the appropriate tool, with the added cautions about quantification discussed by [Raclariu et al. \(2018\)](#).

3.8. Step 8: Report the result and its limits

The report gives, for each item, the loci attempted, the outcome at each, the reference sequences matched with their voucher numbers, the decision rule and the conclusion in one of four forms: consistent with the declared species, identified as another named species, ambiguous among named species, or no result. It states that the test identifies the plant from which DNA was recovered and says nothing about the content of active constituents, the plant part or the presence of contaminants that carry no plant DNA. Chromatograms and control results are archived with the report.

4. Fitting the Protocol to a Working Laboratory

The protocol is meant to sit beside pharmacognostic and chemical tests. Morphology and microscopy confirm the plant part and detect foreign matter; chromatography addresses constituents; the barcode addresses species identity. An inconclusive barcode result is referred back to these methods, and where they disagree with the sequence the conflict is investigated before any conclusion is issued. A laboratory without its own sequencer can perform Steps 1 to 5 and send products to a sequencing service, which keeps the equipment needed to that of an ordinary molecular biology bench.

The heaviest cost is the reference library. It is incurred once per commodity and grows in value as more consignments are tested against it. Laboratories serving the same trade would gain from pooling vouchered sequences, and public deposition makes that possible without formal agreements.

5. Limitations and What Remains to Be Shown

The protocol has not been validated here, and several of its claims are expectations drawn from the literature. Validation would need to establish, for each commodity, that the chosen loci separate the target from every species on the substitute list, using reference specimens that were not used to set the threshold; that the procedure gives the same answer when repeated by different analysts on different days; and how often dried material of realistic age yields a usable sequence. Blind testing with deliberately substituted samples would give the most direct measure of error rates.

Some limits are inherent. Recently diverged species and hybrids may share sequences at any practical number of loci. Heavily processed or very old material may yield no DNA. Taxonomic disagreement about species boundaries carries through to the reference library. In these situations the protocol is designed to return an ambiguous or null result, which is preferable to a confident error.

6. Conclusion

Authentication of crude medicinal plant material by DNA barcoding is within reach of an ordinary quality-control laboratory if the work is organised around a defined list of substitutes, a vouchered reference library, two independent loci, routine controls and a decision rule set down beforehand. The eight steps proposed here supply that organisation. Their performance on real consignments is an empirical matter that remains to be established, commodity by commodity.

Conflict of interest

To be declared by the author(s) before publication.

Funding

To be declared by the author(s) before publication.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). [Basic local alignment search tool](https://doi.org/10.1016/S0022-2836(05)80360-2). *Journal of Molecular Biology*, 215(3): 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- CBOL Plant Working Group (2009). [A DNA barcode for land plants](https://doi.org/10.1073/pnas.0905845106). *Proceedings of the National Academy of Sciences of the United States of America*, 106(31): 12794-12797. <https://doi.org/10.1073/pnas.0905845106>
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010). [Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species](https://doi.org/10.1371/journal.pone.0008613). *PLoS ONE*, 5(1): e8613. <https://doi.org/10.1371/journal.pone.0008613>
- China Plant BOL Group (2011). [Comparative analysis of a large dataset indicates that internal transcribed spacer \(ITS\) should be incorporated into the core barcode for seed plants](https://doi.org/10.1073/pnas.1104551108). *Proceedings of the National Academy of Sciences of the United States of America*, 108(49): 19641-19646. <https://doi.org/10.1073/pnas.1104551108>
- de Boer, H.J., Ichim, M.C. and Newmaster, S.G. (2015). [DNA barcoding and pharmacovigilance of herbal medicines](https://doi.org/10.1007/s40264-015-0306-8). *Drug Safety*, 38(7): 611-620. <https://doi.org/10.1007/s40264-015-0306-8>
- Doyle, J.J. and Doyle, J.L. (1987). [A rapid DNA isolation procedure for small quantities of fresh leaf tissue](#). *Phytochemical Bulletin*, 19: 11-15.
- Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003). [Biological identifications through DNA barcodes](https://doi.org/10.1098/rspb.2002.2218). *Proceedings of the Royal Society B: Biological Sciences*, 270(1512): 313-321. <https://doi.org/10.1098/rspb.2002.2218>
- Hollingsworth, P.M., Graham, S.W. and Little, D.P. (2011). [Choosing and using a plant DNA barcode](https://doi.org/10.1371/journal.pone.0019254). *PLoS ONE*, 6(5): e19254. <https://doi.org/10.1371/journal.pone.0019254>
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005). [Use of DNA barcodes to](#)

- identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23): 8369-8374. <https://doi.org/10.1073/pnas.0503123102>
- Meier, R., Shiyang, K., Vaidya, G. and Ng, P.K.L. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology*, 55(5): 715-728. <https://doi.org/10.1080/10635150600969864>
- Mishra, P., Kumar, A., Nagireddy, A., Mani, D.N., Shukla, A.K., Tiwari, R. and Sundaresan, V. (2016). DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal*, 14(1): 8-21. <https://doi.org/10.1111/pbi.12419>
- Newmaster, S.G., Grguric, M., Shanmughanandhan, D., Ramalingam, S. and Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine*, 11: 222. <https://doi.org/10.1186/1741-7015-11-222>
- Parveen, I., Gafner, S., Techen, N., Murch, S.J. and Khan, I.A. (2016). DNA barcoding for the identification of botanicals in herbal medicine and dietary supplements: Strengths and limitations. *Planta Medica*, 82(14): 1225-1235. <https://doi.org/10.1055/s-0042-111208>
- Raclariu, A.C., Heinrich, M., Ichim, M.C. and de Boer, H. (2018). Benefits and limitations of DNA barcoding and metabarcoding in herbal product authentication. *Phytochemical Analysis*, 29(2): 123-128. <https://doi.org/10.1002/pca.2732>
- Ratnasingham, S. and Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7(3): 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8): 2045-2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, 315-322.
- WHO (2011). *Quality Control Methods for Herbal Materials*. World Health Organization, Geneva.