

Conflict of interest statement

This text is related to our recently published paper^[ref. a] in *Communications Earth and Environment* and serves as a response to a critique submitted by Clement *et al.* to the journal's Matters Arising section. Both the critique and our reply underwent peer review. However, the editorial team decided to reject publishing the exchange, largely agreeing with our reply and concluding, based on the reviewers' and editor's assessments, that the critique did not substantially challenge our analyses and results, hence not meeting the standards required for publication.

Despite the editorial decision, the authors of the critique have chosen to disseminate their submission as a preprint.^[ref. b] In response, we feel it is important to share our reply, especially in light of the critique's public availability. While this exchange no longer qualifies as a formal scientific publication, sharing our response allows us to present a more balanced perspective, and affords us to defend our work against the circulated criticisms. No further action is anticipated from our side beyond the release of this response.

[ref. a] Colli-Silva, M., Richardson, J. E., Neves, E. G., Watling, J., Figueira, A. & Pirani, J.R. 2023. Domestication of the Amazonian fruit tree *cupuaçu* may have stretched over the past 8000 years. *Communications Earth & Environment* 4, 401 (2023). <https://doi.org/10.1038/s43247-023-01066-z>.

[ref. b] Clement, C. R., Alves, R. M., Vicentini, A., Balée, W., Epps, P., Magalhães, M. P., Alves-Pereira, A., Carvalho, J. E. U. de, & Ramirez, H. 2024. Finding the origin of domestication of *cupuaçu* requires more than genomics. In SciELO Preprints. <https://doi.org/10.1590/SciELOPreprints.8304>.

Reply to: Finding the origin of domestication of *cupuaçu* requires more than genomics

Matheus Colli-Silva^{*1}, James E. Richardson^{2,3,4,5}, Eduardo G. Neves⁶, Jennifer Watling⁶, Antonio Figueira⁷, & José Rubens Pirani⁸

¹ Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom.

² School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland.

³ Environmental Research Institute, University College Cork, Cork, Ireland.

⁴ Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom.

⁵ Faculty of Natural Sciences, Rosario University, Bogotá, Colombia.

⁶ Museum of Archaeology and Ethnology, University of São Paulo, São Paulo, Brazil.

⁷ Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, Brazil.

⁸ Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, Brazil.

* Author for correspondence (m.collisilva@kew.org)

By invoking published literature, Clement *et al.* attempted to undermine our study (1) that challenged conventional perspectives on the origin and domestication of *cupuaçu* (*Theobroma grandiflorum*) and its relationship to *cupuí* (*T. subincanum*). They claimed we ignored long-term research in taxonomy, history, biogeography, and genetics, yet neglected our recent and ongoing efforts in most of these areas (2–6). They questioned the validity of

our genomic analyses, but their critique was devoid of empirical evidence or additional data to support any of their claims, instead presenting irregular reasoning based on fundamental errors. Furthermore, they relied on outdated sources when discussing *Theobroma* taxonomy which may sound ultracrepidarian when addressing species delimitations. When Clement *et al.* presented linguistic arguments, their interpretation was also—in our perspective—speculative, and could be viewed from a different angle, as we will argue below.

The origin of *cupuaçu* has long been discussed across various fields, and both sides agree that historical and archaeological evidence has only yielded inconclusive scenarios for the species' origin and domestication that lacked testing or empirical demonstration. When looking into the references brought by the authors, Clement *et al.* mention Augusto Ferreira (7) and Adolpho Ducke (8), whose speculations would align with our findings, but they also mention de Candolle, Cuatrecasas (9), and other authors, whose perspective differs from the one we shown in our study. What the authors may have not conceived is that such inconsistency among historical sources actually created an opportunity for further efforts, which is precisely what our study addresses. But even before our study, another work (10) emphasized that the approximate location of *T. grandiflorum* center of diversity in Eastern Pará could not be confirmed solely by analyzing populations from the “native region” defended by the authors of the critique. In a previous, but still more recent revision of Clement *et al.* (11), Eastern Pará was suggested as the center of origin for *cupuaçu*, but, curiously, its status was also questioned according to the authors themselves (“*Cupuaçu?*” in Fig. 1 of ref. (11)).

Recent studies on cacao (*T. cacao*), *cupuaçu*'s close relative, demonstrate how past relationships between humans and nature may have affected areas that appear primary but might not be (2, 12). For *cupuaçu*, Figure 1 underscores the difficulty in categorizing specimens collected in regions that appear to be within primary forests. These examples represent just a small fraction of the complex setting that we have been studying over the past five years in the taxonomy of *Theobroma*. Furthermore, our analysis of preserved specimens found in herbaria worldwide stress that there are no distinct morphological features that reliably would separate the so-called “morphotypes” of *cupuí*, agreeing with Cuatrecasas (9) taxonomic delimitations. This conclusion is also supported by our published work detailing the main patterns observed in this data (4). A recent study from our group (3) would dispel any doubts about sister relationships between *T. grandiflorum* and *T. subincanum*, but the authors incorrectly assume that we consider one or other species as “valid” [sic] based on this study, which is wrong, as we will discuss below.

The fact that “Amazonia is undercollected” should not be used as an argument to invalidate our findings. Cuatrecasas' monograph (9) was published during a time when collections were even more limited than they are today, further emphasizing the need to continually revisit and expand our knowledge of the region. In fact, because Amazonia is undercollected, every new piece of data is a jewel, and our work simply underscores the importance of new data in potentially reshaping our understanding of various topics. That is why we are currently working on a new taxonomic revision of *Theobroma* (see ref. (6)) to update the knowledge that has evolved since Cuatrecasas' monograph on the genus, published almost 60 years ago (9). We anticipate that Clement *et al.* would show evidence or demonstrate expert knowledge on the morphological delimitations of *Theobroma* species when they say “*cupuí* is poorly defined.” Instead, they seem to rely solely on the last taxonomic literature available without critically analyzing it or verifying the primary data sources, and without cross-checking with our own results that show a paraphyletic pattern.

That said, we will move on to a central topic on Clement *et al.* critique: does domestication create new species? The authors resolutely assert that no, but we diverge. Decisions regarding taxonomic ranks are arbitrary, and there is no inherent criterion dictating whether something should be classified as a species, variety, or subspecies (13). When studying domestication, numerous instances arise where “species” are actually considered subspecies, forms, or varieties of the progenitor (see ref. (14) for a review of over two hundred crops). For *cupuaçu*, if we were to adopt *T. grandiflorum* as conspecific to *T. subincanum*, which would follow Clement *et al.* logic, one might expect the names used in cultivation would mirror those in nature. However, this would ignore two hundred years of usage of both names. Also, following plant nomenclature rules (15), the correct name for the combined species (wild form + domesticate) would be *T. grandiflorum* (Willd. ex Spreng.) K.Schum., as its basionym, *Bubroma grandiflorum* Willd. ex Spreng. (1826), predates *T. subincanum* Mart. (1830). This might appear illogical since the domesticated form would serve as the “parent” of the taxon from which it was derived; but this discrepancy exists just because the domesticate was recognized and named before the wild form was. At this point, and for the sake of nomenclatural stability, we are happy to maintain both entities as “valid species” [sic] (incorrect and inappropriate terminology that Clement *et al.* used to refer to what would be a correct name for a species in a rank), regardless of their origin. This decision, however, is out of scope of this discussion, was never a topic of debate in our published study (1), and it would be valid for future taxonomic treatments, as the one we are currently working (see ref. (6)).

In terms of genetics, our results indicate that *cupuí* populations have maintained their natural genetic structure and phenotype over time, exhibiting a paraphyletic assemblage, contrary to the findings for *cupuaçu* in our study. In natural settings, many authors in phylogeography and population genetics suggest that the origin of a new species often involves the paraphyly of its sister species (*e.g.* 14). Nevertheless, it appears that the authors have not considered our results collectively and misinterpreted many of our figures, which demonstrated: (i) low levels of heterozygosity, F_{ST} , and nucleotide diversity in *cupuaçu*, not typical characteristics of wild species; (ii) a highly homogeneous ADMIXTURE sampling, indicating a new genetic group of *cupuaçu*, as depicted in Fig. 2d of the main manuscript; and (iii) high levels of mutation loads, evidenced by fewer genes under much stronger selective pressure in *cupuaçu* compared to *cupuí*, consistent with the “cost-of-domestication” hypothesis (17–19). Considering these results collectively, it is plausible, based on the evidence we have presented, to interpret that the genetic structure revealed by *cupuaçu* would not be due to natural processes like natural selection or genetic drift, but likely reflects artificial selection. To challenge these findings, new samples or analyses would be required. Regrettably, the authors have not introduced any new data or analysis to effectively counter our results, so the discussion is limited to speculation.

Clement *et al.* also write their critics on genetics based on arguments that are simply irregular or inappropriate. For instance, they misinterpreted Fig. S6 of our work, incorrectly alleging low ancestral range probabilities. They persist in suggesting that we should have dated a phylogeny based on Richardson *et al.* (20), employing an illogical method of cross-checking two different phylogenies constructed under different optimization methods to estimate lineage divergence times based on the branch length of one phylogeny with the estimated age of the other, which is simply wrong. Even if we were to entertain their insistence on using such irregular method, we would require either sequences from the branch to calibrate (fossil data of *Theobroma*, which is nonexistent), or utilize secondary calibration based on ages of Richardson *et al.* (20). However, this is exceedingly complex and cannot be reduced to

simple cross-comparison (as suggested by Clement *et al.*), as the calibration points of Richardson *et al.* are distant from *Theobroma*, introducing increased error.

They also critiqued our method of variant calling, stating that we used the reference genome of *T. cacao* (a species from the same genus). However, along with *de novo* assembly, this is a very common approach in population genomics, and some studies actually recommend to use closely related reference genomes for SNP discovery, not to say that this has also been discussed empirically in many instances for RADseq data (21–23), with both the reference-aligned and *de novo* assembly pipelines being able to yield reliable results for species from the same genus (24). While the reference genome of *T. grandiflorum* is still not fully available (25), it would be great to have seen this discussion grounded in empirical analyses or data. These analyses could have empirically demonstrated the potential impact of various parameters on our results, but, disappointingly, these were not made by the authors.

Moving ahead, integrating linguistic insights into our discussion adds a compelling layer to the ongoing discourse. As biologists and archaeologists, we find this approach unexpected, but valuable for understanding the complexities inherent in our subject. However, we note that Clement *et al.* seem fixated on emphasizing the modern dispersal of the Tupi-Guaranis, overlooking the broader context of Macro-Tupi language expansion linked to these populations before European colonization (26). Recent studies, for instance, have revealed a nuanced pattern of Tupi populations living in the Amazon Basin approximately 2,000 years ago (27). Although this timeframe is more recent than the dates suggested by our demographic analyses of *cupuaçu*, it indicates that the linguistic approach may also extend back significantly before colonization, adding at least a layer of complexity to the subject.

Furthermore, our study outlines two distinct “pulses” in the domestication of *cupuaçu*, and we actually do hypothesized that the fruit likely began to be widely dispersed throughout the Amazon basin only within the last 200 years, when it became widely popular. Our demographic reconstruction suggests the initial domestication pulse by 5,000–8,000 years, but this does not necessarily imply that the larger fruits of *cupuaçu* were “created” and “named” at that time. Given that *cupuaçu* has a generation time of 3–5 years (28, 29), it is challenging to determine precisely when the fruit became “large enough” to be referred to as “*cupuaçu*” or to distinguish the linguistic split from the genetic divergence. This ambiguity suggests that the recognition of larger *cupuaçu* fruits may have occurred later or through multiple instances, illustrating the gradual evolution of domestication practices over millennia. Nonetheless, our main conclusions remain valid until new evidence, based on fresh data, either corroborates or refutes our hypotheses. Other arguments based on speculation are essentially opinions or theoretical exercises that align with one classical viewpoint or another. They, however, do little to advance the discussion when not properly based on new data or analyses, instead contributing to confusion, and somehow disrupting the scientific process of knowledge generation.

In conclusion, our published study introduces novel evidence based on genomic data that challenges prevailing perspectives defended by Clement *et al.* arguments, which often lacked conclusive findings in prior literature. We acknowledge the limitations of our study, particularly in geographic sampling, but this has already been stated in the published manuscript (1), where we already advocated for future studies to address these gaps with new data and analyses. While critiques from other disciplines can be pertinent, those made by Clement *et al.* appear to restate a reactionary stance in response to new hypotheses arising from novel evidence. Rather than providing a substantive challenge to the conclusions drawn

from our published study, they simply reinforce established viewpoints. This approach, in our view, does not offer the significant insights or advancements that would warrant a publication in the Matters Arising section.

References

1. M. Colli-Silva, J. E. Richardson, E. G. Neves, J. Watling, A. Figueira, J. R. Pirani, Domestication of the Amazonian fruit tree cupuaçu may have stretched over the past 8000 years. *Commun Earth Environ* **4**, 1–8 (2023).
2. M. Colli-Silva, J. E. Richardson, A. Figueira, J. R. Pirani, Human influence on the distribution of cacao: insights from remote sensing and biogeography. *Biodivers Conserv*, doi: 10.1007/s10531-023-02777-7 (2024).
3. A. M. Bossa-Castro, M. Colli-Silva, J. R. Pirani, B. A. Whitlock, L. T. Morales Mancera, N. Contreras-Ortiz, M. L. Cepeda-Hernández, F. Di Palma, M. Vives, J. E. Richardson, A phylogenetic framework to study desirable traits in the wild relatives of *Theobroma cacao* (Malvaceae). *Journal of Systematics and Evolution* n/a (2024).
4. M. Colli-Silva, J. Richardson, J. Pirani, A taxonomic dataset of preserved specimen occurrences of *Theobroma* and *Herrania* (Malvaceae, Byttnerioideae) stored in 2020. *BDJ* **11**, e99646 (2023).
5. M. Colli-Silva, J. R. Pirani, L. J. Dorr, (2950) Proposal to conserve the name *Cacao sylvestris* (*Theobroma sylvestre*) (Malvaceae: Byttnerioideae) with a conserved type. *TAXON* **72**, 438–440 (2023).
6. M. Colli-Silva, J. E. Richardson, A. M. Bossa-Castro, J. R. Pirani, Phylogenetic evidence reshapes the taxonomy of Cacao and its allies (*Theobroma* and *Herrania*; Malvaceae, Byttnerioideae). *Brittonia*, doi: 10.1007/s12228-024-09783-1 (2024).
7. A. R. Ferreira, *Viagem Filosófica Pelas Capitánias Do Grão Pará, Rio Negro, Mato Grosso e Cuiabá, 1783-1792* (Real Jardim Botânico, Ajuda, 1800).
8. A. Ducke, Plantas de cultura precolombiana na Amazônia brasileira. Notas sobre as espécies ou formas espontâneas que supostamente lhes teriam dado origem. *Boletim Técnico do Instituto Agrônomo do Norte* **8**, 1–24 (1946).
9. J. Cuatrecasas, Cacao and its allies, a taxonomic revision of the genus *Theobroma*. *Contr. U.S. Natl. Herb.* **35**, 379–614 (1964).
10. R. M. Alves, A. M. Sebbenn, A. S. Artero, C. Clement, A. Figueira, High levels of genetic divergence and inbreeding in populations of cupuassu (*Theobroma grandiflorum*). *Tree Genetics & Genomes* **3**, 289–298 (2007).
11. C. R. Clement, W. M. Denevan, M. J. Heckenberger, A. B. Junqueira, E. G. Neves, W. G. Teixeira, W. I. Woods, The domestication of Amazonia before European conquest. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150813 (2015).

12. C. Lanaud, H. Vignes, J. Utge, G. Valette, B. Rhoné, M. Garcia Caputi, N. S. Angarita Nieto, O. Fouet, N. Gaikwad, S. Zarrillo, T. G. Powis, A. Cyphers, F. Valdez, S. Q. Olivera Nunez, C. Speller, M. Blake, F. Valdez, S. Raymond, S. M. Rowe, G. S. Duke, F. E. Romano, R. G. Llor Solórzano, X. Argout, A revisited history of cacao domestication in pre-Columbian times revealed by archaeogenomic approaches. *Sci Rep* **14**, 2972 (2024).
13. J. Bates, Is Domestication Speciation? The Implications of a Messy Domestication Model in the Holocene. *Agronomy* **11**, 784 (2021).
14. R. S. Meyer, A. E. DuVal, H. R. Jensen, Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* **196**, 29–48 (2012).
15. N. J. Turland, J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W.-H. Kusber, D.-Z. Li, K. Marhold, T. W. May, J. McNeill, A. M. Munro, J. Prado, M. J. Price, G. Smith, International Association for Plant Taxonomy, Eds., *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code): Adopted by the Nineteenth International Botanical Congress, Shenzhen, China, July, 2017* (Koeltz Botanical Books, Glashütten, Germany, 2018) *Regnum vegetabile*.
16. L. H. Rieseberg, L. Brouillet, Are many plant species paraphyletic? *TAXON* **43**, 21–32 (1994).
17. Q. Liu, Y. Zhou, P. L. Morrell, B. S. Gaut, Deleterious Variants in Asian Rice and the Potential Cost of Domestication. *Molecular Biology and Evolution* **34**, 908–924 (2017).
18. B. T. Moyers, P. L. Morrell, J. K. McKay, Genetic Costs of Domestication and Improvement. *Journal of Heredity* **109**, 103–116 (2018).
19. J. Lu, T. Tang, H. Tang, J. Huang, S. Shi, C.-I. Wu, The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends in Genetics* **22**, 126–131 (2006).
20. J. E. Richardson, B. A. Whitlock, A. W. Meerow, S. Madriñán, The age of chocolate: a diversification history of Theobroma and Malvaceae. *Front. Ecol. Evol.* **3** (2015).
21. H. E. L. Lischer, K. K. Shimizu, Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics* **18**, 474 (2017).
22. A. B. A. Shafer, C. R. Peart, S. Tusso, I. Maayan, A. Brelsford, C. W. Wheat, J. B. W. Wolf, Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol Evol* **8**, 907–917 (2017).
23. J. Bohling, Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. *Ecology and Evolution* **10**, 7585–7601 (2020).
24. S. Fitz-Gibbon, A. L. Hipp, K. K. Pham, P. S. Manos, V. L. Sork, Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data

using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome* **60**, 743–755 (2017).

25. R. M. Alves, V. A. C. de Abreu, R. P. Oliveira, J. V. dos A. Almeida, M. de M. de Oliveira, S. R. Silva, A. R. Paschoal, S. S. de Almeida, P. A. F. de Souza, J. A. Ferro, V. F. O. Miranda, A. Figueira, D. S. Domingues, A. M. Varani, Genomic decoding of *Theobroma grandiflorum* (cupuassu) at chromosomal scale: Evolutionary insights for horticultural innovation. *bioRxiv* [Preprint] (2023).
<https://doi.org/10.1101/2023.12.23.573069>.
26. F. S. Noelli, “The Tupi Expansion” in *The Handbook of South American Archaeology*, H. Silverman, W. H. Isbell, Eds. (Springer New York, New York, NY, 2008; http://link.springer.com/10.1007/978-0-387-74907-5_33), pp. 659–670.
27. M. A. Castro e Silva, K. Nunes, R. B. Lemes, À. Mas-Sandoval, C. E. Guerra Amorim, J. E. Krieger, J. G. Mill, F. M. Salzano, M. C. Bortolini, A. da C. Pereira, D. Comas, T. Hünemeier, Genomic insight into the origins and dispersal of the Brazilian coastal natives. *Proceedings of the National Academy of Sciences* **117**, 2372–2377 (2020).
28. R. M. Alves, S. F. D. S. Chaves, R. P. D. Oliveira, J. L. Pedroza Neto, A. Sebbenn, Canopy replacement used in the evaluation of cupuassu tree genotypes in the state of Pará. *Rev. Bras. Frutic.* **42**, e-597 (2020).
29. M. D. A. Falcão, E. Lleras, Fundação Universidade do Amazonas, Brasil, Centro Nacional de Recursos Genéticos, Brasil, Instituto Interamericano de Cooperação para a Agricultura, Brasil, Aspectos fenológicos, ecológicos e de produtividade do cupuaçu - *Theobroma grandiflorum* (Willd. ex Spreng.) Schum. *Acta Amaz.* **13**, 725–735 (1983).

ORCIDs

Matheus Colli-Silva: <https://orcid.org/0000-0001-7130-3920>

James E. Richardson: <https://orcid.org/0000-0001-9014-4865>

Eduardo G. Neves: <https://orcid.org/0000-0002-2830-2735>

Jennifer Watling: <https://orcid.org/0000-0001-7897-5105>

Antonio Figueira: <https://orcid.org/0000-0001-8641-2556>

José Rubens Pirani: <https://orcid.org/0000-0001-7984-4457>

Authors contribution

Matheus Colli-Silva prepared the initial response. All other authors have reviewed, edited and added their comments and suggestions.

Abstract

By invoking published literature, Clement *et al.* attempted to undermine our study that challenged conventional perspectives on the origin and domestication of *cupuaçu* (*Theobroma grandiflorum*) and its relationship to *cupuí* (*T. subincanum*). They claimed we ignored long-term research in taxonomy, history, biogeography, and genetics, yet neglected

our recent and ongoing efforts in most of these areas. They questioned the validity of our genomic analyses, but their critique was devoid of empirical evidence or additional data to support any of their claims, instead presenting irregular reasoning based on fundamental errors. Furthermore, they relied on outdated sources when discussing *Theobroma* taxonomy which may sound ultracrepidarian when addressing species delimitations. When Clement *et al.* presented linguistic arguments, their interpretation was also, in our perspective, speculative and could be viewed from a different angle, as we will argue in this piece.

Keywords: botany, cacao, domestication, matters arising, species delimitations, *Theobroma*.

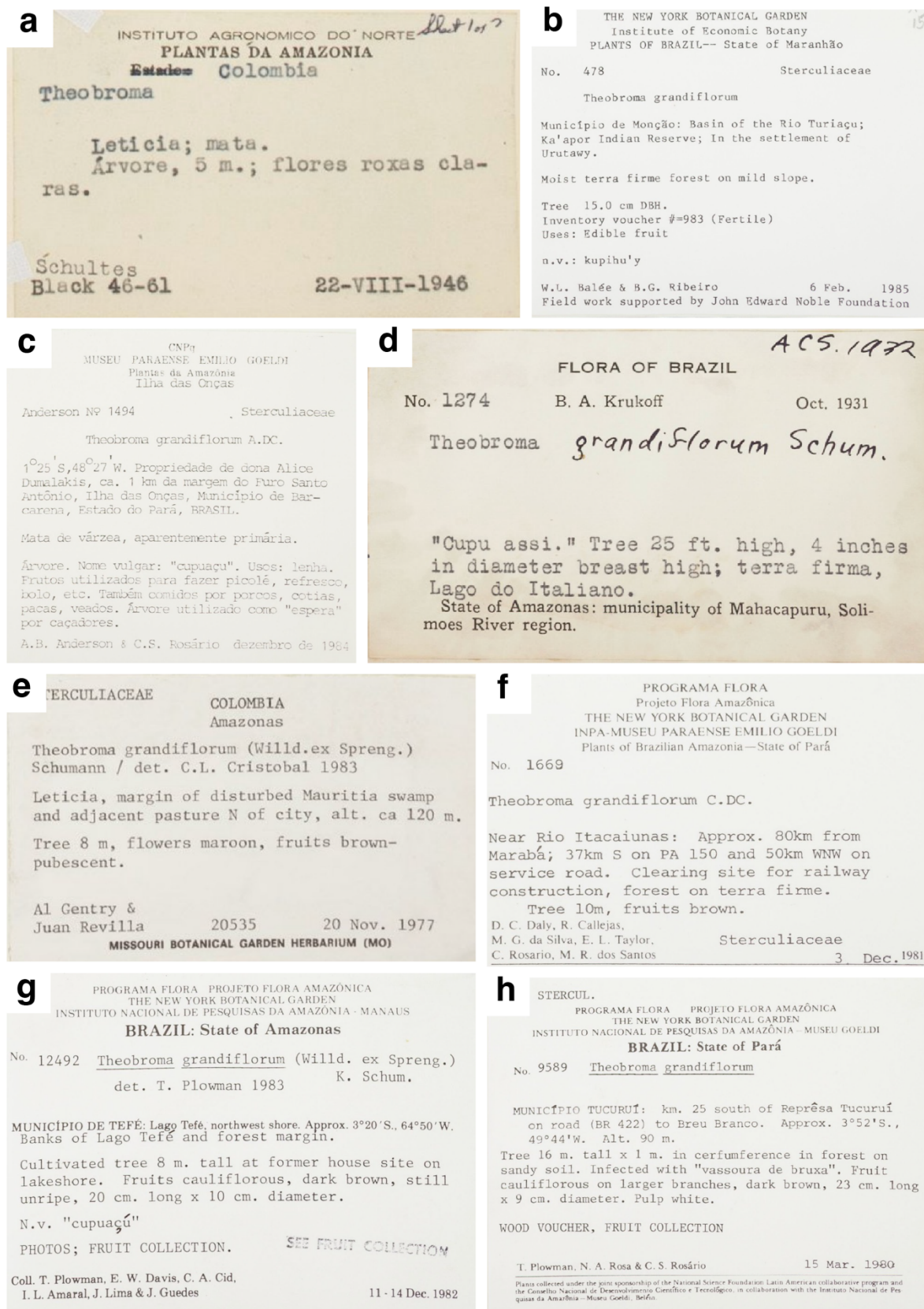


Figure 1. Selected specimen labels of *Theobroma grandiflorum* preserved collections from various regions of the Amazon basin, illustrating the challenge of distinguishing between “wild” and “cultivated” specimens. Even when cultivation information is not explicitly provided (b, f, h, e), the distribution of the voucher locations is consistently linked to human

settlements. Specimens were collected from multiple locations in both the Western (a, d, e, g) and Eastern Amazon basin (b, c, f, h), including the region where Clement *et al.* propose as the “native area” of *cupuaçu*. These and other specimens were physically examined during our ongoing taxonomic research and are preserved in herbaria (e.g. refs. 2–5). Specimen identities have been confirmed, and this work is part of our broader efforts to study the taxonomy and diversity of *Theobroma* species. We believe that these efforts were neglected by the authors of the critique.