TWN
Third World Network
Website: www.twn.my

Briefing Paper
August 2021

Medicinal plant sequence information in the INSDC and its country of origin

Results for a sample of Latin American, African, and Asia-Pacific plants

Edward Hammond

Widespread failure to label digital sequence information (DSI) with its country of origin in research databases has emerged as an important issue in international discussions on access and benefit sharing for genetic resources. The lack of country-of-origin labelling is a problem that must be corrected in order for benefits from DSI to be shared fairly and equitably, particularly if Parties to the Convention on Biological Diversity fail to develop a multilateral benefit-sharing system that avoids the need for track and trace of most individual DSI accessions. And even within a multilateral system for sharing DSI benefits, consistent labelling of DSI with its country of origin is likely to be an important measure to enable implementation and monitoring.

It is well-known and has been previously noted that the large majority of sequence information in the largest international “open access” DSI database, the International Nucleotide Sequence Database Consortium (INSDC), lacks a country-of-origin tag.[1] But defenders of the INSDC, which functions as a US-European-Japanese triumvirate, nevertheless cite the voluntary country-of-origin tag available to researchers who upload sequences as a purportedly useful mechanism for benefit sharing. This optional tagging has even been described as “an existing traceability system”.[2] The INSDC’s defenders further maintain that when the country-of-origin tag is used, the country information entered by laboratories uploading DSI is generally accurate.

But can a voluntary tag that is mostly not used be called a “traceability system”? And is it true that the INSDC country-of-origin tag, when it is used, contains accurate information?

Results

To assess use and non-use of the INSDC’s country-of-origin tag, INSDC entries for 30 species of medicinal plants endemic to developing countries were reviewed. The results confirm that few INSDC entries have a country-of-origin tag: Of the 726 INSDC records analyzed, 577 (79.5%) were found to have no country-of-origin information. One hundred and seven records (14.7%) were found to have an accurate country-of-

---

[1] CBD/DSI/AHTEG/2020/1/4, page 4.
[2] CBD/DSI/AHTEG/2020/1/4, page 5.

Third World Network (TWN) is an independent non-profit international research and advocacy organisation involved in bringing about a greater articulation of the needs, aspirations and rights of the peoples in the South and in promoting just, equitable and ecological development.

Address: 131 Jalan Macalister, 10400 Penang, MALAYSIA **Tel:** 60-4-2266728/2266159 **Fax:** 60-4-2264505
Email: twn@twnetwork.org **Website:** www.twn.my

origin tag. The proportion of accurate tags found in the study sample is slightly lower than the 16% rate of tag use that has been reported in background studies for the Convention on Biological Diversity.

The study further found that 40% of the species examined (12/30) did not have a single INSDC record with a country-of-origin tag correctly completed, while for only 10% of species (3/30) was the country of origin entered for all accessions. Put differently, there were four times as many medicinal plant species from developing countries that had no country-of-origin information at all as there were species with complete information.

The study also finds that, contrary to assertions made in the INSDC's defence, when the country-of-origin tag is used, there are problems with the accuracy of the information provided. Erroneous country-of-origin tags were identified for five species (17%). In total, 42 country-of-origin tags were found to be incorrect while 107 were accurate, meaning that the rate of inaccuracy of the country-of-origin tag in the study sample was a very high 28.2% (42/149).

The study also sought to quantify from where unlabelled sequence information is being uploaded into the INSDC database. While developing-country scientists upload some unlabelled data, the vast majority comes from the North. Of the 577 medicinal plant sequences without country-of-origin information, 84.7% (489/577) were uploaded from developed countries, versus 15.1% (87/577) from developing countries.[3]

Europe was the largest source of unlabelled sequences, accounting for 37.4% (216) of the total. It was closely followed by the US, from which 35.5% (205) of the unlabelled uploads were made. Other parts of the North (Japan, South Korea, Canada, Australia, and Taiwan) accounted for a further 11.8% (68) of the total.

It bears emphasizing here that the above figures refer to uploads of unlabelled sequence information of medicinal plants ***from developing countries***, and that the study method sought to select medicinal plants that are not endemic to any Northern country and not known to have naturalized there (see "Method" section below). Thus, in all or nearly all cases, the 489 sequences without country-of-origin information that were uploaded from the North are sequences of material obtained in the South (including via *ex-situ* collections).

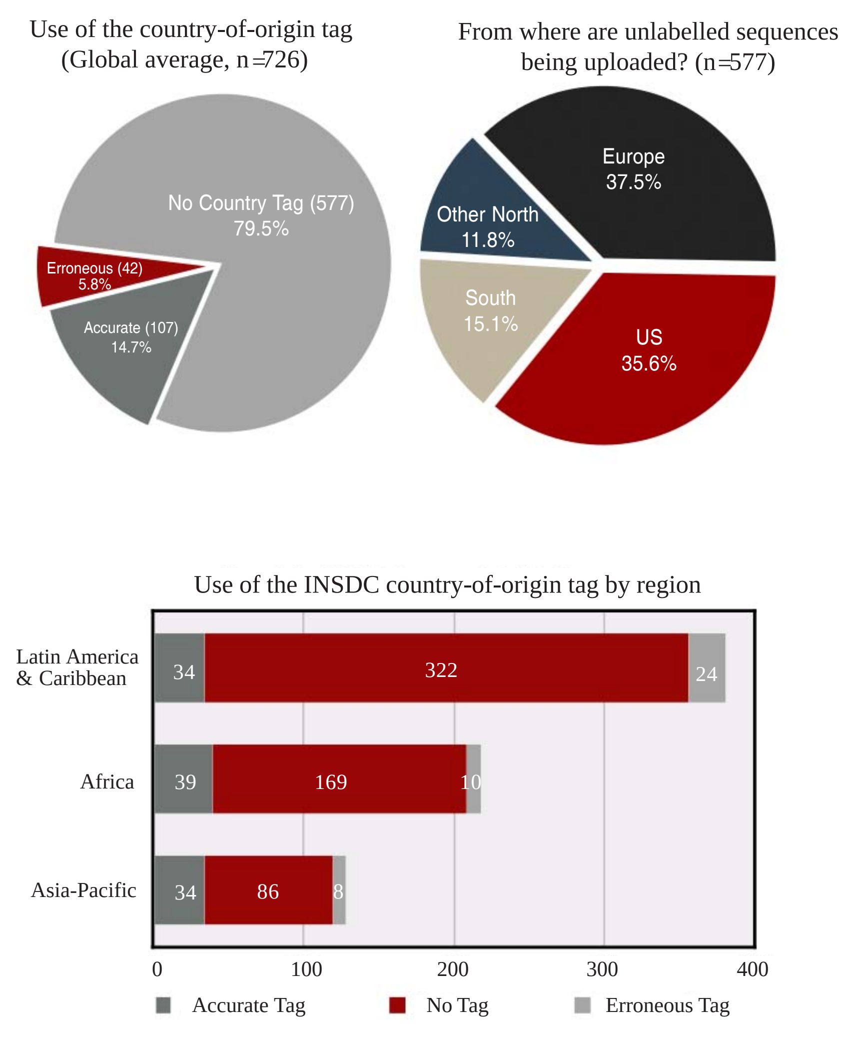
More than three-quarters (77%) of the plant species examined had none, or fewer than one-third, of sequence accessions in the INSDC database labelled. Between one- and two-thirds of accessions were labelled for four species. None had more than two-thirds of accessions labelled except three species in which all had a country-of-origin tag. An interesting observation is that in two of the three cases in which a species was fully labelled, all of the accessions were uploaded by a single lab. In the third case, 13 of 14 accessions were uploaded and labelled by a single lab.

These exceptions show that decisions by individual labs can make a difference, but the weight of the results demonstrates that so long as use of the INSDC country-of-origin tag remains voluntary and thus not widely practised, and its accuracy is suspect and limited by what the uploading lab wishes to enter, the tag will not be of major use for access and benefit-sharing purposes. It is therefore not credible to claim that "an existing traceability system" exists in the INSDC.

In a bilateral approach, the country of origin must be known for all DSI accessions in order for there to be fair and equitable benefit sharing. While a multilateral system for benefit sharing may reduce or even eliminate the need for tracking and tracing for most individual DSI accessions, it is still unlikely that such a system can be implemented and monitored without consistent, mandatory association of DSI with its country of origin.

---

[3] A single unlabelled sequence was uploaded by the International Potato Center (CIP) in Lima, Peru, and was considered to have been uploaded by an international institution.



Use of the country-of-origin tag (Global average, n=726)

From where are unlabelled sequences being uploaded? (n=577)

Use of the INSDC country-of-origin tag by region

Medicinal Plant Sequence Information in the INSDC Database and Country-of-Origin Tag Use

Species	Number of sequences in INSDC	Sequences with a country of origin	Sequences with no country of origin	Sequences with an erroneous country of origin
Latin America and Caribbean				
Justicia pectoralis	7	0	7	
Banisteriopsis caapi	8	0	8	
Sambucus peruviana	5	0	5	
Arracacia xanthorrhiza	58	0	58	
Capsicum rhomboideum	34	0	34	
Bulnesia arborea	86	1	85	
Dioon spinulosum	67	3	64	
Lepidium meyenii	55	9	22	24
Hippeastrum puniceum	19	3	16	
Simarouba amara	41	18	23	
Subtotal	**380**	**34**	**322**	**24**
Percent		**8.9%**	**84.7%**	**6.3%**
Africa				
Hypoxis hemerocallidea	13	0	13	
Pelargonium sidoides	74	0	74	
Macrosphyra longistyla	14	0	14	
Tabernanthe iboga	20	0	15	5
Berlinia confusa	20	1	19	
Voacanga africana	14	2	7	5
Trichilia emetica	21	6	15	
Erythrina senegalensis	9	4	5	
Mondia whitei	19	12	7	
Momordica angustisepala	14	14	0	
Subtotal	**218**	**39**	**169**	**10**
Percent		**17.9%**	**77.5%**	**4.6%**
Asia - Pacific				
Pothos scandens	10	0	10	
Homalanthus nutans	16	0	15	1
Canarium patentinervium	4	0	4	
Decaspermum fruticosum	7	1	6	
Litsea garciae	15	3	12	
Drynaria bonii	17	5	12	
Catunarengam spinosa	39	12	27	
Melodinus cochinchinensis	11	4	0	7
Alphitonia zizyphoides	6	6	0	
Marsypopetalum modestum	3	3	0	
Subtotal	**128**	**34**	**86**	**8**
Percent		**26.6%**	**67.2%**	**6.3%**
Total	**726**	**107**	**577**	**42**
Percent	**100.0%**	**14.7%**	**79.5%**	**5.8%**

Method

Publications on medicinal plants from developing countries were identified by an internet search (e.g., "medicinal plants of Southeast Asia"). One or two species were then arbitrarily selected from peer-reviewed publications matching the internet search.

Plants that are not native to the region where their medicinal use is noted were excluded, as were plants whose native range substantially expands into multiple regions or extends to multiple regions through widespread naturalization. While it is not possible to know in every case the full geographic extent of populations of every plant, based on taxonomic resources, primarily Kew Gardens' Plants of the World,[4] an effort was made to select plants that are solely endemic to a country or countries within the region with which they are associated.

A query on the INSDC's GenBank interface[5] for DNA/RNA sequence information of each species was performed.[6] Species with fewer than three INSDC accessions were excluded, as were those with more than 100 (the latter due to time limitations).

The 100-accession limit resulted in a number of frequently studied medicinal plants, particularly those associated with Indian and Chinese traditional medicine, being excluded from the sample because the number of INSDC sequences of those plants frequently sums into several hundred or thousands. This exclusion is a reason why the total number of accessions from Asia in the study is lower than for Africa and Latin America and the Caribbean.

The use or non-use of the country-of-origin tag for each accession of each species meeting the selection criteria was recorded, as well as the country from which each unlabelled accession was submitted. Once 10 medicinal plants meeting the selection criteria from a region were identified, queries conducted, and results recorded, the process moved to the next region.

Edward Hammond *directs Prickly Research (www.pricklyresearch.com), a research and writing consultancy based in Austin, Texas, USA. He has worked on biodiversity and infectious disease issues since 1994. From 1999 to 2008, Hammond directed the Sunshine Project, an international non-governmental organization specializing in biological weapons control. Hammond was Programme Officer for the Rural Advancement Foundation International (now the ETC Group) from 1995 to 1999. He holds MS and MA degrees from the University of Texas at Austin, where he was an Inter-American Foundation Masters Fellow.*

---

[4] http://plantsoftheworldonline.org/

[5] https://www.ncbi.nlm.nih.gov/genbank/

[6] e.g., "Banisteriopsis caapi"[Organism] AND plants[filter]

Annex: Detailed Results

Below are the complete results of the 30 species and 726 accessions analyzed. The results are listed in order of the percentage of INSDC sequences in which the country-of-origin tag correctly identifies the origin of the genetic resources, 0% meaning none, and 100% meaning all.

0% (0/13)

Hypoxis hemerocallidea African star grass
Southern Africa

Thirteen (13) *H. hemerocallidea* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from South Korea (5), Switzerland (5), Poland (1), the US (1), and the UK (1).

0% (0/10)

Pothos scandens climbing aroid
South Asia, Southeast Asia, New Guinea, Madagascar

Ten (10) *P. scandens* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from the US (4), the UK (2), Mexico (2), France (1), and Canada (1).

0% (0/7)

Justicia pectoralis tilo
Tropical Latin America and Caribbean

Seven (7) *J. pectoralis* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from the US (7).

0% (0/74)

Pelargonium sidoides pelargonium
Southern Africa

Seventy-four (74) *P. sidoides* sequences were found in the INSDC database. None of the sequences had an associated country of origin. Seventy-one (71) of the unlabelled sequences were uploaded from the Netherlands, and the remainder from Germany (2) and Spain (1). The lack of use of the country-of-origin tag for *P. sidoides* seems especially notable given the widely publicized biopiracy controversy over patent claims on this plant by a German pharmaceutical company.[7]

0% (0/20, five sequences incorrectly labelled)

Tabernanthe iboga iboga
West Africa

Twenty (20) *T. iboga* sequences were found in the INSDC database. Five of the sequences had an associated country of origin, all of which were incorrect. In these five sequences the country of origin indicated was

---

[7] African Center for Biodiversity (2010). Major breakthrough in the fight against biopiracy: Pelargonium patents (news release), 26 April. https://www.acbio.org.za/major-breakthrough-fight-against-biopiracy-pelargonium-patents

Belgium, apparently because the sequence is of a sample from the Meise Botanical Garden near Brussels. The unlabelled sequences were uploaded from Germany (10), the UK (2), the US (2), and Sweden (1).

0% (0/14)

Macrosphyra longistyla various
West Africa

Fourteen (14) *M. longistyla* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from Sweden (14).

0% (0/8)

Banisteriopsis caapi ayahuasca
Amazon Basin

Eight (8) *B. caapi* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from the US (6), Japan (1), and Brazil (1).

0% (0/58)

Arracacia xanthorrhiza arracacha
Andes

Fifty-eight (58) *A. xanthorrhiza* sequences were found in the INSDC database. None had an associated country of origin. The unlabelled sequences were uploaded from the US (44) and France (14).

0% (0/16, one sequence incorrectly labelled)

Homalanthus nutans mamala
Pacific

Sixteen (16) *H. nutans* sequences were found in the INSDC database. One of the sequences had an associated country, Australia. The unlabelled sequences were uploaded from the US (10) and Japan (5).

Australia is outside the native range of *H. nutans*, a plant closely associated with Samoa. Determining that the Australian attribution was incorrect involves a complicated story. US and Australian authors uploaded the sequence in a DNA barcoding study.[8] The sample was reportedly collected in Queensland and deposited at the Queensland Herbarium. But a search of the Queensland Herbarium's index, hosted by the Australasian Virtual Herbarium (AVH), returns no entries for *H. nutans.*

The AVH also indexes New Zealand herbaria, and it considers *H. nutans* to be a synonym of *H. polyandrus*, a different species that is endemic to New Zealand. The AVH directs *H. nutans* queries to the New Zealand species, potentially mixing two distinct plants under the same name.

None of the AVH's 50 *H. polyandrus* records come from Australia. One, however, is indicated as having been collected in 2006 north of New Zealand and close to *H. nutans'* native range. The coordinates of that location, however, are in the middle of the Pacific Ocean off the island of Futuna, about 600 km west of Samoa. This "Futuna" record is attributed to a different collector and herbarium than that cited by the US and Australian authors. Moreover, metadata suggests that it is from the Kermadec Islands, a New Zealand possession whose northernmost point is 1700 km south of Futuna.[9]

---

[8] Shapcott A et al. (2015). https://doi.org/10.1371/journal.pone.0122164

[9] See: https://avh.ala.org.au/occurrences/277129f0-ed43-4b62-a50f-0bd4235907bc

The circumstances of taxonomic confusion and incorrect records strongly suggest that the sequence in the INSDC database identified as *H. nutans* from Australia is erroneous, either because it is not from Australia or because it is not an *H. nutans* sequence.

0% (0/5)

Sambucus peruviana
sauco
Central and South America

Five (5) *S. peruviana* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from Belgium (4) and the US (1).

0% (0/34)

Capsicum rhomboideum
various
Mexico, Central America, Northern Andes

Thirty-four (34) *C. rhomboideum* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from Argentina (21), Austria (9), and the US (4).

0% (0/4)

Canarium patentinervium
kedondong (Malaysia)
Southeast Asia

Four (4) *C. patentinervium* sequences were found in the INSDC database. None of the sequences had an associated country of origin. The unlabelled sequences were uploaded from the US (4).

1% (1/86)

Bulnesia arborea
guayacán de bola
Northern South America

Eighty-six (86) *B. arborea* sequences were found in the INSDC database. Just one of the sequences had an associated country of origin, Venezuela. The unlabelled sequences were uploaded from the US (77), the UK (3), Argentina (3), and Germany (2).

4% (3/67)

Dioon spinulosum
palma de chicle/gum palm
Mexico

Sixty-seven (67) *D. spinulosum* sequences were found in the INSDC database of this cycad endemic to a small area of Mexico (and sold as an ornamental elsewhere). Just three of the 67 sequences had an associated country of origin, Mexico. These were uploaded by Mexican scientists. The unlabelled sequences were uploaded from the US (16), Taiwan (15), China (12), Japan (9), Mexico (5), Thailand (4), Australia (1), Italy (1), and Germany (1).

5% (1/20)

Berlinia confusa
various
West and Central Africa

Twenty (20) *B. confusa* sequences were found in the INSDC database. One sequence was labelled with a country of origin, Gabon. The unlabelled sequences were uploaded from Canada (10) and the UK (9).

14% (2/14, five sequences incorrectly labelled)

Voacanga africana various
Tropical Africa

Fourteen (14) *V. africana* sequences were found in the INSDC database. Seven sequences were labelled with a country of origin, five of which were incorrect. One sequence was from Mozambique and another from Cameroon, but five sequences incorrectly indicated Belgium as the country of origin. The sequences appear to be of material from the Meise Botanical Garden near Brussels. The unlabelled sequences were uploaded from China (3), the US (2), Nigeria (1), and the Philippines (1).

14% (1/7)

Decaspermum fruticosum tailor tree
Pacific, parts of Southeast Asia

Seven (7) *D. fruticosum* sequences were found in the INSDC database. One was labelled with a country of origin, which was Indonesia. The unlabelled sequences were uploaded from the UK (6).

16% (3/19)

Hippeastrum puniceum various
Tropical Americas

Twenty-one (21) *H. puniceum* sequences were found in the INSDC database. Five were labelled with a country of origin, two of which are not in *H. puniceum*'s natural range. Three were from Guyana, and two indicated Thailand. *H. puniceum* has naturalized in some areas outside the Americas, and in the absence of further information about the Thailand-labelled accessions, they were excluded from the sample. The unlabelled sequences were uploaded from South Korea (8), the US (4), and Denmark (4).

16% (9/55, 24 sequences incorrectly labelled)

Lepidium meyenii maca
Peru

Fifty-five (55) *L. meyenii* sequences were found in the INSDC database. Thirty-three were labelled with a country of origin, 24 of which appear to be incorrect. Maca was largely unknown outside of Peru until the 1990s. Nine of the 33 labelled samples indicated Peru as the country of origin. Twenty-four of the labelled samples indicated China as the country of origin. The unlabelled sequences were uploaded from the US (8), China (7), Germany (4), Japan (2), and international (1 – International Potato Center, Lima).

20% (3/15)

Litsea garciae various
Southeast Asia, Taiwan

Fifteen (15) *L. garciae* sequences were found in the INSDC database. Three were labelled with a country of origin, all of which were labelled Malaysia. The unlabelled sequences were uploaded from Japan (12).

29% (5/17)

Drynaria bonii (Aglaomorpha bonii) various
Eastern South Asia, Southeast Asia, Southern China

Seventeen (17) *D. bonii* sequences were found in the INSDC database. Five were labelled with a country of origin; all five were labelled as China. The unlabelled sequences were uploaded from China (9) and Germany (3).

29% (6/21)

Trichilia emetica various
Africa, parts of Arabian Peninsula

Twenty-one (21) *T. emetica* sequences were found in the INSDC database. Six were labelled with a country of origin, two each from Mozambique, Zimbabwe, and South Africa. The 15 unlabelled sequences were uploaded from the US (4), the UK (4), Austria (3), Germany (2), Switzerland (1), and South Africa (1). The lack of a country-of-origin tag on the latter sequence appears to be an oversight, as sequence information from the same sample was labelled as being from Zimbabwe.

31% (12/39)

Catunaregam spinosa mountain pomegranate
South and Southeast Asia

Forty-one (41) *C. spinosa* sequences were found in the INSDC database. Fourteen (14) were labelled with a country of origin. There were five from China, four from India, three from Taiwan, and two from South Africa. Kew Gardens and other sources consider South and Southeast Asia as the native range of *C. spinosa*, though it appears to be naturalized in East and Southern Africa. In the absence of further information about the two South-Africa-labelled sequences, they were excluded from the sample. The unlabelled sequences were uploaded from Sweden (15), China (8), France (2), and India (2).

36% (4/11, seven sequences labelled incorrectly)

Melodinus cochinchinensis mountain orange
Eastern South Asia, Southeast Asia, Southern China

Eleven (11) *M. cochinchinensis* sequences were found in the INSDC database. All 11 were labelled with a country of origin, but seven of these labels are incorrect. Four sequences were labelled as being from China. The sequences that were incorrectly labelled were uploaded from Switzerland (5) and Brazil (2). These were tagged as being from Belgium. All seven appear to be generated from *M. cochinchinensis* samples from the Meise Botanical Garden near Brussels.

44% (18/41)

Simarouba amara aceituno
Central America, Caribbean, Northern South America

Forty-one (41) *S. amara* sequences were found in the INSDC database. Eighteen (18) were labelled with a country of origin. There were nine from Costa Rica, four from French Guiana, three from Puerto Rico, and one each from Nicaragua and Panama. The unlabelled sequences were uploaded from Germany (14), Brazil (5), and the US (4).

44% (4/9)

Erythrina senegalensis Senegal coraltree
West Africa and the Sahel

Nine (9) *E. senegalensis* sequences were found in the INSDC database. Four were labelled with a country of origin. All four were labelled Nigeria and were uploaded by Nigerian scientists. The five unlabelled sequences were uploaded from the US.

63% (12/19)

Mondia whitei mukombero (Kenya)
Southern and East Africa

Nineteen (19) *M. whitei* sequences were found in the INSDC database. Twelve sequences had an associated country of origin, five from Kenya, four from Tanzania, and three from South Africa. The Kenyan sequences were uploaded and labelled from Kenya. The unlabelled sequences were uploaded from the Czech Republic (3), the US (2), and South Africa (2).

100% (14/14)

Momordica angustisepala various
West Africa

Fourteen (14) *M. angustisepala* sequences were found in the INSDC database. All of them had an associated country of origin – Nigeria (7), Cameroon (6), and Ghana (1). This unusual result appears to stem from the fact that 13 of the 14 sequences were uploaded by a single botany research group in Germany. Atypical among its peers, this group properly labelled the origin of the sequences. Such attention to country of origin regrettably appears to be quite rare.

100% (6/6)

Alphitonia zizyphoides various
Pacific Islands

Six (6) *A. zizyphoides* sequences were found in the INSDC database. All had an associated country of origin, two each from Vanuatu, Tahiti, and Wallis & Futuna. The latter two are administered by France. All six sequences were uploaded by the University of Hawaii (US).

100% (3/3)

Marsypopetalum modestum various
Southeast Asia

Three (3) *M. modestum* sequences were found in the INSDC database. All three had an associated country of origin, Thailand. The sequences were uploaded by the University of Leiden (Netherlands).