Taxonomy and Systematics of *Urophyllum* (Rubiaceae) in Thailand and Indochina

A thesis submitted by

Sawita Yooprasert

For the Degree of Doctor of Philosophy

University of Reading Herbarium,

School of Biological Sciences, University of Reading

May 2021

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Sawita Yooprasert

Acknowledgements

Firstly, I would like to thank my sponsors, the Royal Thai Government and Department of National Parks, Wildlife and Plant Conservation, for providing the scholarship and letting me work on this project.

I am deeply grateful to my supervisors: A. Culham and T. Utteridge for all the valuable advice, support, and encouragement. I would also like to thank M. Christodoulou and K. Könyves for their advice, training, and continued support.

I would like to extend my gratitude to my examiners: M. Newman and J. Hawkins for their patience and kindness in going through my thesis.

My deepest thanks go to my senior colleagues in BKF: K. Chayamarit, R. Pooma, S. Suddee, V. Chamchumroon, and N. Pattharahirantricin for their support and guidance. I would like to express my gratitude to M. Poopath, V.D. Nyugen and K.S. Nyugen who assisted in helping to organise fieldwork and collection permits. I would also like to thank T. Jonganurak and his colleagues for taking care of my studentship during the study.

Special thanks to the curators and herbarium staff of AAU, ABD, C, QSBG and SING for sending herbarium specimens on loan to the University of Reading, and to BK, BM, CMUB, E, HN, HNU, K, KKU, PSU and VNM for providing accesses to the collections in the herbaria. My deepest appreciation goes to the colleagues in BKF who helped dealing with herbarium collections and to all the members of the Culham Research Group for moral support and interesting discussion. Particular thanks (diolch) to Tomos and Nattanon for their friendship.

I owe a very important debt to my mum and my sisters for their love, constant support and encouragement. Last, but certainly not least, to my partner Jordan. Words are not enough to express how grateful I am for your love, help, encouragement, and patience when I have got a cob on. Nice one!

Abstract

The genus *Urophyllum* Wall. is a taxonomically problematic genus of Rubiaceae in Thailand due to the lack of a recent taxonomic revision and identification key to the species. This has led to confusion in the identification of species and no conservation status assessments for the genus. The aim of this thesis, therefore, is to produce a taxonomic revision of *Urophyllum* with an identification key to the species focused on Thailand that will contribute towards the Rubiaceae account in the Flora of Thailand which remains unwritten. An integrative taxonomic approach is applied in this thesis, using both morphological and molecular data for species delimitation. A rigorous approach of combining linear and geometric morphometric data with machine learning was applied to 130 specimens of 13 *Urophyllum* taxa to test groups (taxa). The results based upon linear morphometric data revealed that 86% of specimens were accurately classified based on the data supplied, with seven taxa classified with 100% accuracy. Geometric morphometrics performed worse with only 60% of specimens being accurately classified. However, a combined data approach improved the accuracy, with 91% of specimens being successfully classified, including 10 taxa with a perfect accuracy. Phylogenetic analyses were also performed on 18 species (39 samples) of *Urophyllum* to provide support for the morphological classification. Whole plastid genomes and nrDNA cistron sequences were mined from genome skimming data. The results of the phylogenetic analyses reveal an incongruent relationship between plastid and nrDNA. Combining morphometric data with molecular results reveal that nrDNA largely supports the classification of taxa using morphometrics, whilst plastid DNA reveals a geographic pattern. Moreover, the three misclassified taxa from the combined morphometric data were resolved to their own clade using nrDNA. The results from the morphological and molecular investigation were used to inform a taxonomic revision of *Urophyllum* in Thailand and as the basis to publish new species in Cambodia and Vietnam. This thesis therefore directly contributes towards the Flora of Thailand, and the Flora of Cambodia, Laos and Vietnam. The application of methods in this study also provides a robust framework for testing the species delimitation in a wider taxonomic context.

Table of Contents

Chapter 1 Introduction

1.1. Taxonomy and conservation

Descriptive taxonomy is the fundamental basis of other disciplines in biology including molecular biology, ecology, and conservation, and crucially serves as the essential knowledge of world biodiversity (Wilson, 2004). A plant name allows for the effective communication of knowledge about that particular plant, and therefore is the essential foundation for further studies. Taxonomy is particularly important, during this time of accelerated biodiversity loss, due to human activity as without information of plant species conservation assessments and strategies cannot be made (Mace, 2004).

The need for plant conservation action was outlined by the Global Strategy for Plant Conservation (GSPC) of the UN Convention on Biological Diversity, with the first fundamental target to have an online flora of all known plants (https://www.cbd.int/gspc/, accessed 1 May 2021). In order to meet this feat, the World Flora Online Consortium (WFO) was launched to outline a World Flora Online, and to form an international consortium of institutions that can collaborate and provide the essential content and knowledge. The aim is to build upon existing knowledge and published Floras, checklists, and taxonomic revisions, but essentially there is the requirement to collect and generate new data for poorly understood groups of plants. As part of this collaboration, the Flora of Thailand project which has been working on the taxonomic studies of vascular plants in Thailand for over 50 years will now contribute towards this global target.

1.2. The Flora of Thailand project: mission to revise *Urophyllum*

The Flora of Thailand project has been the driving force behind taxonomic studies in Thailand for over 50 years. The project was launched in 1967 with collaboration between Thai and international botanists, the primary aim of the project was to produce a comprehensive floristic treatment for the native vascular plants found in Thailand. To date, there are 13 volumes, with up to 4 parts for each volume (data from https://www.dnp.go.th/botany/, accessed 29 April 2021). An e-Flora of Thailand is in production and is due to be launched in July 2021, initially covering Volumes 2–9 of the Flora of Thailand. Furthermore, the Board of the Flora of Thailand has recently agreed to

be a member of the World Flora Online Consortium (WFO) (R. Pooma - Director of BKF, pers. comm.).

There is an ambitious aim to complete the entire Flora of Thailand by 2024 (Middleton *et al.*, 2019), therefore many Thai botanists are involved in the contribution to this project by targeting the remaining plant groups that do not have an account. The family Rubiaceae is one of the remaining plant groups that requires study for the Flora of Thailand and is one of the five largest groups in angiosperms (Bremer and Eriksson, 2009). In Thailand, the number of genera and species are expected to be close to 110 genera with approximately 600 species (Puff, Chayamarit and Chamchumroon, 2005). This thesis is a contribution towards the Rubiaceae account, focussing upon the genus *Urophyllum* Wall., as well as developing techniques that can be applied across the remaining accounts for the Flora of Thailand project to provide robust identification and classification.

Identification of *Urophyllum* species within the genus, in particular, is challenging due to the lack of recent taxonomic accounts for the genus and limited precise diagnostic characters reported (usually leaf size, lateral vein pairs or hair density) (Puff, Chayamarit and Chamchumroon, 2005; pers. obs.). The genus *Urophyllum* is often misidentified as *Lasianthus* Jack due to both genera having axillary inflorescences and similar leaf venation (Bremekamp, 1940; pers. obs.). However, *Lasianthus* is distinguished from *Urophyllum* as plants are hermaphrodite, fresh leaves usually have an unpleasant smell when bruised, campanulate flowers, and drupe fruits, usually blue to purple with pyrenes (1–9) inside (Bremekamp, 1940; Hua, 2002), whereas members of *Urophyllum* are dioecious plants, fresh leaves lack an unpleasant smell when bruised, urceolate flowers, and baccate fruits, usually yellowish orange to red with numerous small seed inside (Bremekamp, 1940; Puff, Chayamarit and Chamchumroon, 2005; pers. obs.).

The taxonomic study of Thai plants should not be limited to the occurrence of those plants within the country. As Thailand is located in between four major biogeographical regions: the Himalayas, China, Indochina, and Sundaland, the flora is therefore largely integrated with that of Indochinese, Indo-Burmese and Malesian regions (Van Welzen *et al.*, 2011). The study of *Urophyllum* provided in this thesis therefore includes the species

in Thailand and their occurrence in Indochina and Peninsular Malaysia in order to provide accurate conservation status assessment.

1.3. Taxonomic overview

The genus *Urophyllum* is a member of tribe Urophylleae Bremek. ex Verdc. in the Rubiaceae family. The genera in Urophylleae usually are woody plants, with simple to fimbriate stipules, a bilocular to plurilocular ovary, and indehiscent fruits, often fleshy with numerous seeds (Verdcourt, 1958; Bremer and Manen, 2000; Smedmark *et al.*, 2008; Smedmark and Bremer, 2011). To date, there are 15 genera and approximately 240 species of Urophylleae mostly found in the Palaeotropics (except for two genera found in the Neotropics: *Amphidasya* Standl. and *Raritebe* Wernham). Two of the largest genera: *Urophyllum* and *Pauridiantha* Hook.f. are found in Asia and Africa, respectively (Smedmark *et al.*, 2008; Smedmark and Bremer, 2011).

Urophyllum comprises approximately 120 species distributed predominantly in the wet tropical regions of Asia (Taylor *et al.*, 2011; Wong *et al.*, 2019; Govaerts *et al.*, 2020). *Urophyllum* is a shrub or treelet, often found as an understorey plant (Bremekamp, 1940) within tropical evergreen forests, typically near streams. The genus name comes from Greek: *Uro-* meaning 'tail' and *-phyllum* meaning 'leaf', the name therefore represents the character of leaves found in the type species of the genus, *Urophyllum villosum* Wall., and several other species in the genus. The main morphological characters that identify the genus are shown in Figure 1.1 and include: a shrub or treelet habit with opposite decussate branches, young branches are flat usually with a ridge along the middle part, stipules are usually lanceolate in shape and often more than 1 cm long, inflorescences are axillary, plants are dioecious, usually with yellow to orange berry-like fruits with an annular disc on top and numerous seeds with alveolate testa (pers. obs.). Some species have been recorded as medicinal plants used in parturition or treating fever in Indonesia (Java) and Malaysia (Pahang) including *U. arboreum* (Reinw. ex Blume) Korth., *U. glabrum* Wall. and *U. hirsutum* (Wight) Hook.f. (Priyadi *et al.*, 2010).

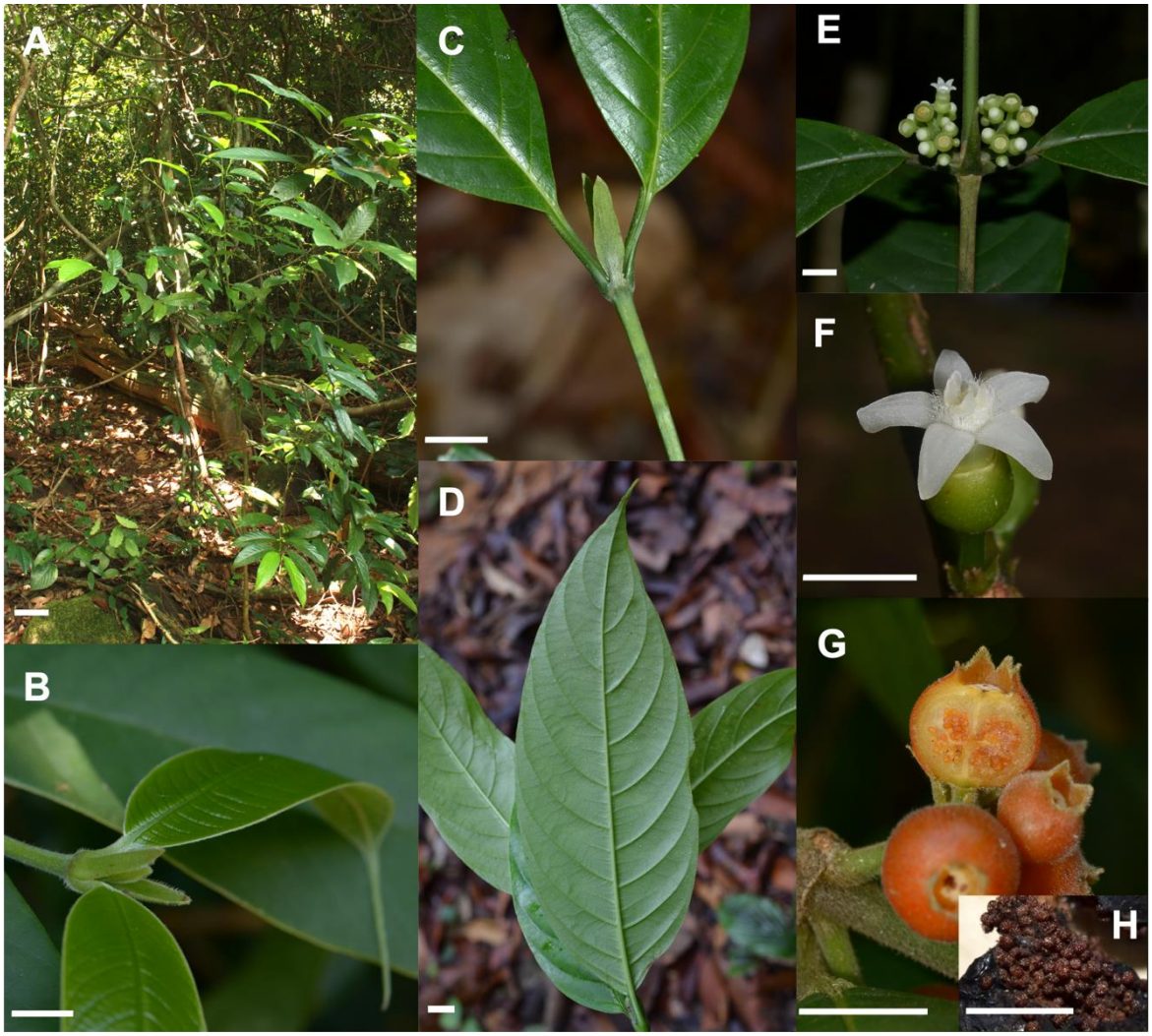


Figure 1.1 Characters of genus *Urophyllum*. **A** Habit (*U. glabrum*). **B** stipule and young leaf with long tail apex (*U. villosum*). **C** stipule (*U. glabrum*). **D** leaf with long acuminate apex (*U. glabrum*) **E** axillary inflorescences bearing staminate flowers (*U. longifolium* (Wight) Hook.f.). **F** pistillate flower (*U. chinense* Merr. & Chun). **G** fruits, cross-section reveals seeds (*U. villosum*). **H** seeds (*U. longifolium*). Scale bar 10 cm (A); 1 cm (B–G); 0.5 mm (H). Photographs by Manop Poopath (B, C, E, G); Nattanon Meeprom (F); Sawita Yooprasert (A, D, H).

Taxonomic revisions of *Urophyllum* during 1800s–1900s were based largely upon morphological studies providing diagnostic characters and/or identification keys (Hooker, 1849, 1880; King and Gamble, 1904; Ridley, 1923, 1932; Bremekamp, 1940; Wong, 1989). Morphological based taxonomy can be controversial as it relies on the accuracy of character definition, which can lead to ambiguity of work when characters are poorly defined (Christodoulou, Clark and Culham, 2020). In the case of *Urophyllum*, species in Malaya were reported to be challenging to separate due to the overlap in morphological

characters in many species, this lead to an incomplete identification key with many species assigned per choice (Ridley, 1932). Bremekamp (1940) used morphological characters to split *Urophyllum* into smaller genera of which four were monotypic. These divisions were based largely upon inflorescence type, the insertion position of hairs on the corolla throat and the hair density. However, Wong (1989) did not recognise the segregate genera and argued that the morphological characters that define the proposed genera are not exclusive, instead they should be recognised as variation amongst species within one genus. A phylogenetic study by Smedmark and Bremer (2011) on tribe Urophylleae revealed that two genera *sensu* Bremekamp (1940) were nested inside the *Urophyllum* clade, therefore they were synonymised to the *Urophyllum* (Smedmark and Bremer, 2011). These studies highlight that traditional taxonomic study, based upon diagnostic characters alone or focused on characters of one specimen (typological species concept) may not reflect the relationship within *Urophyllum*.

In order to provide robust characters for identification, the quantitative approach of morphometrics can play an important role. Morphometrics refers to the quantitative analysis of form, and has long been used in numerical taxonomy for over a century (Sneath and Sokal, 1973). There are two categories of morphometrics: traditional and geometric. Traditional morphometrics is a tool to quantify the distance measurements with explicit interpretation, and data can commonly be easily obtained (Christodoulou, Clark and Culham, 2020). This has resulted in the popular usage in classification, and traditional morphometrics has been used to resolve the relationships in many plant groups, even for cryptic species (Nobis *et al.*, 2016; Macfarlane, Sokoloff and Remizowa, 2017; Di Pietro *et al.*, 2020). In cases of subtle changes in shape, that might not be detected by the traditional morphometrics, more modern geometric morphometric approaches are more appropriate. Geometric morphometrics involves analysing the change of organismal shape as a whole, using landmark coordinates (Christodoulou, Clark and Culham, 2020). The development of morphometric approaches and their application to the classification of *Urophyllum* species can be found in Chapter 2 of this thesis.

1.4. Taxonomy of *Urophyllum*

Urophyllum was first published by Wallich in Roxburgh (1824), who described two species based on William Jack's manuscript together with his own observations: *U. villosum*, and

U. glabrum. These two species are not the first recognised species of this group as the species *Wallichia arborea* Reinw. ex Blume was described by Blume (1823). However, the genus *Wallichia* Reinw. ex Blume is an illegitimate name as it was previously used for a genus in Arecaceae (Roxburgh, 1820). Blume (1826) renamed the genus to *Axanthes* Blume without being aware that the name *Urophyllum* had been published earlier and thus takes priority (Bremekamp, 1940). *Axanthes* had been recognised as its own genus for a few decades, with around 10 species described (Blume, 1826; Wight, 1847) before it was considered to be included in *Urophyllum* by Hooker and Bentham (1849) and then further recognised by both Korthals (de Vriese, Dozy and Molkenboer, 1851) and J.D. Hooker (1880), who formally synonymised the genus to *Urophyllum*.

Several taxonomic revisions of *Urophyllum* occurred during the 1800s (Candolle, 1830; Hooker, 1849, 1880) but none of these works provided an identification key to the species. This meant species identification is difficult. Many studies in the 1900s focused upon regional level accounts, especially in Peninsular Malaysia and Singapore with many new species published; an identification key is generally included in all publications but are limited geographically (King and Gamble, 1904; Ridley, 1923, 1932; Wong, 1989). *Urophyllum sensu* Bremekamp (1940) is fundamentally different from other accounts, due to a proposed split of *Urophyllum* based upon inflorescence type, insertion of the hairs on corolla throat, and their density. Bremekamp (1940) recognised eight new genera based upon these morphological characters; the diagnostic characters for each proposed genus are summarised in Table 1.1. Wong (1989) found that the diagnostic characters were inconsistent and not exclusive to each of the proposed genera. Furthermore, phylogenetic studies using both nuclear ribosomal (nrDNA: ITS and ETS) and plastid (*rps*16 intron and *trn*T–F) DNA regions did not support the splitting circumscription and revealed that five genera are nested within the *Urophyllum* clade (Figure 1.1) (Smedmark *et al.*, 2008; Smedmark and Bremer, 2011; Obico and Alejandro, 2012). Three of the genera (*Maschalocorymbus* Bremek., *Pleiocarpidia* K.Schum. and *Pravinaria* Bremek.) have been combined to *Urophyllum sensu lato* based upon molecular studies (Smedmark and Bremer, 2011). Therefore, *Urophyllum sensu lato* is preferred in this study, which also follows morphological classifications by Hooker (1880), King and Gamble (1904), Ridley (1923) and Wong (1989).

Table 1.1 Diagnostic characters for *Urophyllum s.str.* and its allied genera based on Bremekamp (1940). Modified from Smedmark and Bremer (2011).

Genus	Geographic distribution	Diagnostic characters
Antherostele Bremek.	Philippines	Leaves with domatia (also found in *Urophyllum*, pers. obs.), corolla with velvety hairs on the inside, and anthers linear, syngenesious.
Crobylanthe Bremek.	Borneo	Hairs in corolla throat inserted on two scales at the base of each lobe.
Didymopogon Bremek.	Sumatera	Hairs in corolla tube forming two rings, one in the throat and one at the base.
Lepidostoma Bremek.	Sumatera	Hairs in corolla throat inserted on a scale at the base of each lobe.
Leucolophus Bremek.	Western Malesia	Hairs in corolla throat forming a ring and stipules glabrous inside.
Maschalocorymbus Bremek.*	Vietnam to Malesia	Inflorescences trichotomously corymbose and hairs in corolla throat forming a ring.
Pleiocarpidia K. Schum.*	Myanmar to Malesia	Inflorescences trichotomously corymbose or paniculate, hairs in corolla throat moniliform from base, and stigma peltate.
Pravinaria Bremek.*	Borneo	Inflorescences single-flowered, axillary with one involucel.
Praravinia Korth.	Malesia	Inflorescences with two involucels, corolla with more numerous segments than the calyx, and corolla throat densely covered with stiff, white hairs.
Rhaphidura Bremek.	Borneo	Hairs in corolla throat forming a ring and stipules appressed pubescent inside.
Stichianthus Valeton	Borneo	Cauliflorous, solitary flowers borne in rows along internodes.
Urophyllum Wall.	Tropical & Sub-tropical Asia	Hairs in the corolla throat sparse, attached at the base of corolla lobes and style branches erect or ascending, acute, or obtuse.

* Genus now included within *Urophyllum sensu lato*

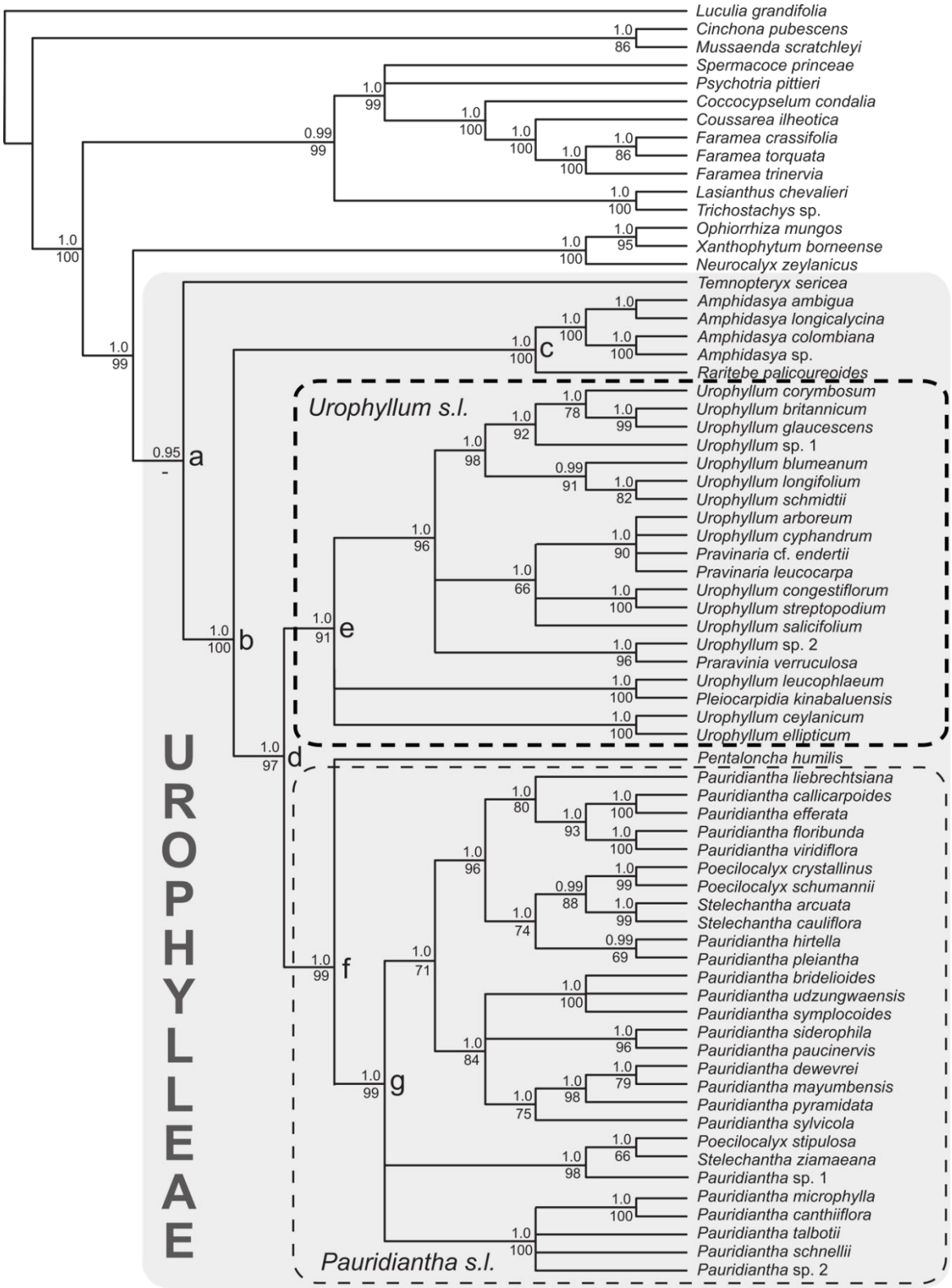


Figure 1.2 Phylogenetic tree showing *Urophyllum sensu lato* with *Pleiocarpidia*, *Praravinia* and *Pravinaria* nested inside (source: Smedmark and Bremer (2011).

1.4.1. *Urophyllum* species in Thailand and mainland Indochina

Taxonomic knowledge of *Urophyllum* species in Thailand has largely relied on the works of Craib (1931, 1932) with specimens from Thailand collected by Kerr (Kerr, 1933; Middleton *et al.*, 2019). Craib (1931, 1932) enumerated 12 species of *Urophyllum* in Thailand of which seven were newly described taxa. However, species identification is challenging as a key to the species was never produced and limited precise diagnostic characters were provided (usually leaf size, lateral vein pairs or hair density) (Puff, Chayamarit and Chamchumroon, 2005; pers. obs.).

Classification of *Urophyllum* from other countries in mainland Indochina including Cambodia, Laos, Myanmar and Vietnam is very limited. In the Flora of Myanmar, only one species *U. longifolium* is recorded (Hooker, 1880). Pitard (1923) in the Flora of Vietnam described five taxa endemic to Vietnam. This means that knowledge of the genus is limited in the whole region, and other than regional accounts the genus has not been revisited for many years. All the species recorded from Thailand and Indochina to date are shown in Table 1.2.

The phylogenetic study of tribe Urophylleae using nrDNA and plastid regions (Smedmark and Bremer, 2011) included four species of *Urophyllum* found in Thailand (*U. blumeanum* Hook.f.,
U. longifolium, *U. schmidtii* C.B.Clarke and *U. streptopodium* Wall. ex Hook.f.) with no species sampled from Cambodia, Laos, or Vietnam. The lack of phylogenetic study of species within *Urophyllum* in Thailand and Indochina has not been completed to date, therefore, one of the aims of this research is to undertake a phylogenetic study to investigate the relationship of *Urophyllum* species in the Indochina region.

Table 1.2 *Urophyllum* species in Indochina as recorded from the literature.

No.	Species	Country
1	*U. aequale* Craib	Peninsular Thailand
2	*U. argenteum* Pit.	Vietnam
3	*U. blumeanum* Hook.f.	Peninsular Thailand
4	*U. crassum* Craib	Peninsular Thailand
5	*U. fuscum* Craib	Peninsular Thailand
6	*U. hirsutum* (Wight) Hook.f.	Peninsular Thailand
7	*U. lecomtei* Pit.	Vietnam
8	*U. longifolium* Hook.f.	Myanmar and Thailand
9	*U. longifolium* var. *pilosum* Craib	Peninsular Thailand
10	*U. longipes* Craib	Peninsular Thailand
11	*U. oblongum* Craib	Peninsular Thailand
12	*U. olivaceum* Craib	South-eastern Thailand
13	*U. schmidtii* C.B.Clarke	South-eastern Thailand
14	*U. streptopodium* Wall. ex Hook.f.	Vietnam
15	*U. talangense* Craib	Peninsular Thailand
16	*U. tonkinense* Pit.	Vietnam
17	*U. trifurcum* H.Pearson ex King	Peninsular Thailand
18	*U. villosum* Wall.	Peninsular Thailand and Vietnam

1.5. Aims and outline of thesis

The primary aim of this thesis is to provide a taxonomic revision of genus *Urophyllum* in Thailand and species that occur in mainland Indochina. This includes: **1)** a re-evaluation of the species concepts used in *Urophyllum*, based on novel molecular and morphological data and techniques; **2)** to evaluate the level of endemism of *Urophyllum* in Thailand, and provide accurate conservation status to form the basis of further studies of the genus; **3)** to develop a taxonomic toolkit for the taxonomic revision of *Urophyllum.* A taxonomic toolkit includes useful plastid gene regions to resolve species relationships within *Urophyllum*, together with the application of morphometrics and machine learning approaches for species classification. It is envisaged that the techniques applied to *Urophyllum* in this thesis can be applied more widely to other genera in Rubiaceae, as the family account is written for the Flora of Thailand.

Below is an outline of the work contributing to the aims of this thesis:

- Chapter 2: Use of morphological characters for species identification in *Urophyllum*, applying morphometric techniques combined with supervised machine learning for species classification.
- Chapter 3: Use of whole plastid genome and nrDNA data to study the phylogenetic relationships of *Urophyllum* species in Thailand and Indochina.
- Chapter 4: Use evidence from Chapters 2 and 3 to produce a taxonomic account of *Urophyllum* in Thailand, in the form of a synopsis.
- Chapter 5: Present five new species of *Urophyllum* from Cambodia and Vietnam, as a precursor for the Flora of Cambodia, Laos and Vietnam.

Chapter 2 Morphometric classification of *Urophyllum*

2.1. Introduction

2.1.1. Morphometrics in plant classification

Morphological study is the most common approach in plant taxonomy. It usually serves as the first tool to identify a species (and lower ranks), by comparing a specimen to a description and the type specimen stated in the protologue of that species. Since morphological data are easy to obtain and distance measurements explicit (Christodoulou, Clark and Culham, 2020), they have been used in traditional or multivariate morphometric studies for over a century (Sneath and Sokal, 1973)

Morphometrics can be combined with molecular tools to study species complexes and hybrids (Hansen, Elven and Brochmann, 2000; Vigalondo *et al.*, 2015; Wei *et al.*, 2015). The analytical choices depend on the data and objectives of the study. In general, they can be categorised by the criteria considering whether *a priori* knowledge of sample groups is provided or not (Henderson, 2005). Two widely used methods with no input prior knowledge include cluster analysis (CA) and ordinations (e.g., principal component analysis (PCA)). CA is typically used as an exploratory tool to assess the association between objects, and groups them according to similarity of information presented in the data. The objects within a group are similar to each other and they are different from other groups (Tan *et al.*, 2014). In ordinations, PCA is also used as an exploratory tool for data analysis and known to help reduce dimensions of a dataset to principal components (PC) presenting a score plot between each pair of PC axes (Jolliffe and Cadima, 2016). The first PC accounts for the largest amount of variation found in the dataset. The second, third and so forth PCs are perpendicular to the previous PC axes and account for the residual variations (Henderson, 2006).

If there is prior knowledge of a group (hypotheses of what the groupings are), traditional multivariate statistics can be used to classify objects, such as: discriminant analysis (DA) (for two groups data) or canonical variate analysis (CVA) (for multiclass data) (Henderson, 2005). The downside of these analyses is the assumption that the data have a normal

distribution. This assumption is difficult to achieve from qualitative data that are usually collected from plant characteristics.

Data collections in traditional morphometrics usually measure distances between two points, ratios and angles (Adams, Rohlf and Slice, 2013). The analysis of these data can be referred to as linear morphometrics (Christodoulou, Clark and Culham, 2020). Although, they include measurements of shape in the form of ratios, the ratio is inferred shape relative to size (Zelditch, Swiderski and Sheets, 2012). This means that an analysis on ratio data cannot be interpreted to the shape difference alone which leads to the use of a more modern method called geometric morphometrics.

2.1.2. Geometric morphometrics

Geometric morphometrics have been used as a tool for studying shape change for more than two decades (Adams, Rohlf and Slice, 2004). It has been widely used in zoological shape studies (Ibañez, Cowx and O'Higgins, 2007; Hedrick and Dumont, 2018; Cox *et al.*, 2020). For plant morphology, it has been used to study the variation in leaf shape in plants such as: *Quercus* L. (Viscosi *et al.*, 2009), *Potentilla* L. (Klingenberg *et al.*, 2012), and *Uvaria* L. (Meade and Parnell, 2003) and flower shape studies of orchids (Shipunov and Bateman, 2005; O'hanlon, Li and Norma-Rashid, 2013). It has also been applied to study fruit shape in apple cultivars (with overall accuracy >70%) (Christodoulou, Battey and Culham, 2018), this demonstrates the benefits of geometric morphometrics in studies where there are limited number of external morphological characters (Christodoulou, Battey and Culham, 2018).

The most common datatypes used in geometric analysis include landmark and outline data (Webster and Sheets, 2010). Landmark data are a set of coordinates that summarise the shape of an object. Thus, choosing appropriate landmark points is the first crucial step for geometric analysis (Christodoulou, 2015). Zelditch *et al.* (2012) described five criteria to consider during landmark selection, that includes repeatability, consistency of relative position, adequacy, homology and coplanarity of landmarks. The first criterion can be quantified by measuring the digitisation error, meaning that landmark digitisation is repeated multiple times, and their differences are compared (Christodoulou, 2015). Consistency of the position means that the landmarks are accurate with no switch of

locations between them in different specimens (Zelditch, Swiderski and Sheets, 2012). This can easily be detected when performing landmark digitisation on a specimen. Adequacy is one of the most difficult criteria to quantify, it refers to the number of landmarks being enough to summarise the overall shape structure. One method to quantify this is by looking at the landmark configuration by removing the background photograph to determine whether all the landmarks represent the form of the subject (Zelditch, Swiderski and Sheets, 2012). Homology, in this case, refers to the correspondence of the same landmark on two or more objects (Zelditch, Swiderski and Sheets, 2012). It is important that each landmark corresponds to the same location of an object, thus the variation in shape can be compared (Christodoulou, 2015). The fifth criterion is specific to 2D landmark digitisation that comes from 3D objects, where distortion can occur. To make sure that all landmarks lie within the same plane, objects must be photographed in the same orientation (Zelditch, Swiderski and Sheets, 2012).

An outline method has gained popularity due to the fact that landmarks on smooth curved objects are difficult to define (MacLeod, 1999). Outline data can be analysed in the form of semi-landmarks, eigenshape analysis or Fourier analysis (Klingenberg, 2008). As outline coordinates require superimposition analysis, as in the homologous landmarks method, this means that points in outline data are also treated to some degree as homologous landmarks (Christodoulou, 2015). However, the homology assumption is not explicit in outline data, as outline methods assume homology purely based upon location. Furthermore, different results can be obtained from the different methods of data treatment and how the corresponding points are defined to overcome this issue (Klingenberg, 2008).

2.1.3. Machine learning

Computing performance and power have been improved in the last decade; this has led to machine learning (ML) being applied in biological studies. Machine learning is a subset of artificial intelligence (AI) that has been widely used in biomedical disciplines (Kourou *et al.*, 2015; Vamathevan *et al.*, 2019; Stamate *et al.*, 2020). To date it has not been widely applied in plant biology; however, Christodoulou *et al.* (2018) used machine learning to identify 27 apple cultivars using only external fruit features. Machine learning refers to topics which

apply a statistical model to facilitate pattern recognition, classification and prediction based on existing datasets (Tarca *et al.*, 2007). Like morphometrics, ML can broadly be divided into two categories depending on the prior knowledge of the data group (or class in this sense), these include supervised and unsupervised learning (Tarca *et al.*, 2007; Christodoulou, Clark and Culham, 2020).

Supervised learning is a method that uses characters/features from known labelled class objects to develop a model that can classify them and can be used for predicting unknown class objects (Tarca *et al.*, 2007). If the classes represent taxa, this method can be referred to as identification (Christodoulou, Clark and Culham, 2020). In contrast, in unsupervised learning, there are no predefined class labels provided. Thus, the aim of an analysis is to explore the data and discover the objects natural similarity (Tarca *et al.*, 2007; Fogel, 2008). This kind of grouping can lead to classification in biological data (Fogel, 2008; Christodoulou, Clark and Culham, 2020). Although, there is a popular subset of machine learning, called deep learning, that is mainly intended for the analyses of large multivariate datasets (Angermueller *et al.*, 2016; Christodoulou, Clark and Culham, 2020). In botany, plant classification and identification usually deal with small number of specimens per taxon, in this case deep learning is typically not suitable for analysis of this datatype. In this chapter, I focus upon supervised learning as the tool for identifying *Urophyllum* taxa.

There are many types of supervised learning classifiers for analysing biological data. To select an appropriate classifier, one will need to train different classifiers and choose one with the best performance (Christodoulou, Clark and Culham, 2020). As described in the "No free lunch" theorem (Wolpert and Macready, 1997), the average performance of all classifiers is equal when all possible problems are considered. Based upon the "No free lunch" theorem, it can be inferred that there is no classifier that always outperforms others for every problem; each classifier will reveal the best performance in a particular problem or dataset (Christodoulou, Clark and Culham, 2020). Therefore, training multiple classifiers is the best way to find the most suitable for a particular problem.

Machine learning includes three steps: training, validating and testing (Christodoulou, Clark and Culham, 2020). In the first step, data are used to train the classifiers. If the whole

dataset is used in this step, it will cause model overfitting and therefore will not provide real performance metrics due to the data being reused in later steps. To avoid this problem, data can be partitioned into separate training (include validation set) and testing sets prior to an analysis (Christodoulou, Clark and Culham, 2020). In order to evaluate the performance of a classifier on a dataset, the validating step is typically performed; to avoid overfitting in this step, cross-validation takes place. One of the most common cross-validation methods used in machine learning for biological data is k-fold cross-validation (Olden, Lawler and Poff, 2008). When a training dataset is divided into k equal subsets, the k-1 subsets are used in training and one subset is left for validating. This will be repeated k times until all possible subsets have been trained and validated (Christodoulou, Clark and Culham, 2020). This cross-validation is modified further to have m repetitions of k-fold cross-validation (Christodoulou, Clark and Culham, 2020). The last step of machine learning is to use the best performing classifier taken from the previous steps to predict (or test) group membership on a test dataset. The prediction can then be reported in the form of a confusion matrix (Tarca *et al.*, 2007).

2.1.4. Preliminary classification and identification of *Urophyllum* species

In this study, the morphological and typological species concepts have been used to initially classify *Urophyllum* species. The classical method of morphological investigation has been performed on both herbarium specimens and newly collected plants from field collections. This includes observing and gathering characters that show both similar and different traits, then classifying samples into groups of similar specimens. The features in each group are compared with the features described in the protologue of each species and with the type specimen (where access was possible) to apply a species name to a particular group. The groups that remain unmatched to any published species name are recognised as unknown species and a number is assigned to that group (e.g., species1, species2 etc.). For morphologically similar taxa, for example *Urophyllum longifolium* var. *longifolium*, *U. longifolium* var. *pilosum* and *U. talangense*, it is difficult to provide identification using characters from the initial description and type specimens. Only hair density and angle of hairs on the stipules show differences for these taxa (pers. obs.). Therefore, these specimens are grouped using these hair characters into three forms (f) of *U. longifolium* followed by letters alphabetically for each form (*U. longifolium* f.A, f.B and

f.C, respectively), to test if they can be classified as separate groups. *Urophyllum talangense* samples are initially labelled as *U. longifolium* forms due to the high morphological similarity.

In order to test that identified *Urophyllum* taxa (and/or species) are actually different in morphological characters and that these characters can be used for species identification morphometric analyses were performed with supervised machine learning on both linear and geometric morphometric datasets.

2.2. Aim and objectives

The main aim of this chapter is to use morphological characters to identify *Urophyllum* species in mainland Southeast Asia. To reach this aim, the three objectives listed below were achieved.

1. Investigate morphological characters in different *Urophyllum* taxa. The characters are classified as quantitative (measurement/count) and qualitative (binary/multistate characters) data. This dataset will represent the linear morphometric data, where size and shape cannot be interpreted separately.
2. Find possible shape differences using geometric morphometric analysis. This method can be used for the study of shape separately from size.
3. Perform supervised machine learning of different classifiers and find the classifier that provides the best prediction of *Urophyllum* taxa in the study area.

2.3. Materials and methods

All supplemental data (tables and figures start with S before the number) in this chapter can be found in Appendix B unless stated otherwise.

2.3.1. Plant materials

Urophyllum species used in this study were obtained from both loaned herbarium specimens from AAU, BKF, FU and new collections from fieldwork (Appendix A). Samples were selected only if they have all of the required characters including at least an intact stipule, undamaged-mature leaf, and inflorescence. This provided 130 specimens for 13 taxa (Table 2.1 and Table S2.1). The characters were chosen based upon literature sources such as descriptions of the species or identification keys where available (Craib, 1932; Ridley, 1932; Tan *et al.*, 1995). Characters from literature sources were supplemented from personal observations during field collections and of herbarium materials.

2.3.2. Data acquisition

2.3.2.1. Linear morphometrics

Morphological characters were measured by either observation by eye, under a stereomicroscope, or using ImageJ v1.50e (Schneider, Rasband and Eliceiri, 2012). A total of 27 characters were measured and can be categorised into 15 quantitative and 12 qualitative characters that include both vegetative and reproductive parts (Table 2.1).

Measurement using ImageJ

The specimens were photographed using a Nikon D810 camera with a Nikon AF-S 50 mm 1:1.8G lens on a Kaiser Fototechnik R1 system copy stand. Photographs were then used to measure characters 1–8 and 11–12 listed in Table 2.2. When possible, each character was measured at a maximum of three replicates per sample. One exception was the pedicel length (character 15) where a maximum of six replicates were measured from 2–3 inflorescences; as specimens typically have more than one inflorescence bearing multiple flowers, therefore measuring the pedicel length from different inflorescences will remove bias. Replicate measurements were used for calculating an arithmetic mean. For characters 9 and 10 (measurement of angle of tertiary veins to midrib), a photograph was taken of an entire abaxial leaf using a Nikon D5100 with a Nikon AF-S 40 mm Micro f/2.8 DX G lens on

the same copy stand as above. These images were used to measure characters 9 and 10 in ImageJ setting the Region of Interest (ROI) at 1 cm in diameter. The ROI was placed in the middle area between two lateral veins at the base and the middle of the leaf. Qualitative data were observed under a Leica S6D stereomicroscope with Leica L2 light. Leaf descriptions are defined using terminology following Hickey (1979).

Vegetative characters differences between pistillate and staminate plants

To test whether different sexes in *Urophyllum* species can be identified using vegetative characters, two species with samples from both pistillate and staminate plants: *U. glabrum* (GL) and *U. longifolium* f.C (LC) were selected. The proportion of pistillate and staminate plants in GL and LC are 17:12 and 14:10, respectively. Eleven vegetative characters (quantitative data in Table 2.2) were collected and analysed using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Estimate of minimum sample size

Even though the number of samples used in this study could not be changed, the estimation of minimum sample size at which quantitative characters will be effective to show the differences among taxa, was considered to select characters. To do this, quantitative data were tested for normality using Shapiro-Wilk test on R (R Core Team, 2019). The non-normally distributed data were transformed using logarithms. Data collected for character 4 (percentage) were transformed using arcsine. Then an ANOVA was performed on the data for each character to calculate the effect size (η^2), as in Christodoulou (2015). The effect size is used to estimate minimum sample size in "pwr" package (Champely *et al.*, 2020) on R. The results reveal that the minimum sample for six characters (4, 6, 7, 13, 14 and 15) was >25 samples which was five times higher than the number of *U. chinense* samples (the smallest number of samples in this study) (Table 2.1). Therefore, the data from these five characters were omitted from the study. The minimum sampling size for each quantitative character is summarised in Table S2.2. In total, 21 characters of linear morphometric data are used in the machine learning section. All data collected can be found in Table S2.3.

Table 2.1 List of taxa, the acronyms and number of specimens and leaves used in this study.

Taxa	Sample Code	Number of specimens	Number of specimens in train dataset (60%)	Number of specimens in test dataset (40%)	Number of leaves for landmarks digitisation
U. argenteum	AR	7	4	3	11
U. chinense	CH	5	3	2	6
U. crassum	CR	7	4	3	11
U. glabrum	GL	29	17	12	45
U. hirsutum	HI	7	4	3	10
U. lecomtei	LE	8	4	3	11
U. longifolium f.B	LB	7	4	3	11
U. longifolium f.C	LC	24	14	10	38
U. longipes	LG	8	5	3	10
U. streptopodium	ST	6	4	2	8
U. villosum	VI	7	4	3	9
U. sp.1 (*U. chinense* subsp. *latistipulum* sp. nov.)	S1	8	5	3	11
U. sp.2 (*U. bidoupense* sp. nov.)	S2	7	4	3	11
	Total	**130**	**77**	**53**	**192**

Table 2.2 A list of characters measured for linear morphometric analysis. Background colours indicate type of data: grey shaded rows are qualitative data; unshaded rows are quantitative data; green shaded rows indicate characters that were omitted from the datasets.

Plant parts	No.	Characters	Measurement / Character states
Petiole	1	Petiole length	mm
Leaf	2	Leaf width	
	3	Leaf length	
	4	The widest point of the leaf to total leaf length (calculated by length from leaf base to the widest part × 100 /total leaf length)	
	5	Lateral vein number	count
	6	Angle at 10% leaf length from the base	degree
	7	Angle at 25% leaf length from the base	
	8	Angle at 25% leaf length from the apex	
	9	Angle of tertiary veins to midrib at the base of a leaf (number of tertiary veins within 1 cm diameter)	
	10	Angle of tertiary veins to midrib at the mid of a leaf (number of tertiary veins within 1 cm diameter)	
Stipule	11	Stipule length	mm
Inflorescence	12	Primary peduncle length	
	13	Rachis length	
	14	Secondary peduncle length	
	15	Pedicel length (up to six flowers)	

Table 2.2 (continued) A list of characters measured for linear morphometric analysis. Background colours indicate type of data: shaded rows are qualitative data; unshaded rows are quantitative data.

Plant parts	No.	Characters	Measurement / Character states
Petiole	16	Hairs on petiole	0 = absent; 1 = present
	17	Hairs inside petiole channel (canaliculate)	0 = absent; 1 = present
Leaf shape	18	Leaf base shape	0 = cuneate; 1 = convex; 2 = round; 3 = concave; 4 = mixed
	19	Leaf shape	0 = elliptic; 1 = ovate; 2 obovate; 3 = oblong
Adaxial leaf	20	Adaxial leaf surface hair distribution	0 = hairless; 1 = only midrib;
	20	Adaxial leaf surface hair distribution (continued)	2 = up to lateral vein; 3 = up to tertiary vein; 4 = young leaf up to lateral vein, mature only midrib; 5 = young leaf up to lateral vein, mature only base of midrib; 6 = hairy only base of midrib (ca. 1 cm)
Abaxial leaf	21	Abaxial leaf surface hair distribution	1 = only midrib; 2 = up to lateral vein; 3 = up to tertiary vein; 4 = up to quaternary vein; 5 = hairy lower than quaternary vein
	22	Abaxial vein protrude	1 = up to quaternary vein; 2 = lower than quaternary vein
Abaxial leaf	23	Pocket domatia at axil of branching between lateral veins and the midrib	0 = absent; 1 = present every angle on a leaf; 2 = present in some angles; 3 = angle cover with dense hairs and cannot see clearly after remove

Table 2.2 (continued) A list of characters measured for linear morphometric analysis. Background colours indicate type of data: shaded rows are qualitative data; unshaded rows are quantitative data.

Plant parts	No.	Characters	Measurement / Character states
Abaxial leaf (continued)	24	Hairs at axil of branching between lateral veins and the midrib	0 = glabrous/no hair; 1 = hairy, denser than other areas; 2 = hairy but not show any degree of denser than other areas
Stipule	25	Stipule hair distribution	1 = hairy all over; 2 = hairy at margin only; 3 = hairy around midline only
Inflorescence	26	Abaxial corolla hair distribution	0 = glabrous/scaly around apex; 1 = hairy all over; 2 = scaly all over
	27	Calyx lobe	0 = truncate/toothed; 1 = lobed

2.3.2.2. Geometric morphometrics

a. Landmark configuration

Mature leaves were chosen from the previous zoomed-in photographs used in the linear dataset for digitising landmarks. A maximum of two leaves per sample were chosen when possible, resulting in 192 leaves in total per replicate (Table 2.1). Each photograph was renamed to the codes seen in Table S2.1, the first two characters indicate taxon; the next two digits identify the specimen; a letter represents an individual leaf; and finally, a number indicates replicates. To start digitising, a .tps file containing all image data was created using TPSUtil v1.78 (Rohlf, 2019). The nine landmarks were digitised in the same order in TPSDig2 v2.31 (Rohlf, 2017). A study area on the abaxial side of a leaf was chosen by dividing the total number of lateral veins by two, working on the right-hand side of the midrib. If the resulting number was a decimal with a .5 value, a higher whole number was chosen. This provided the middle lateral vein for the study area, from here one lateral vein above, and one below was selected to add landmarks. Such an example is shown in Figure 2.1.

All landmarks selected are illustrated in Figure 2.1 and the raw co-ordinates for each leaf are given in Table S2.4. Landmarks 1, 2 and 3 exhibit the point where the lateral vein attaches to the midrib arranged from leaf base to leaf apex, respectively. Landmarks 4-5 and 7-8 are points where the lateral vein forms a closed loop to the lateral vein above. Landmarks 6 and 9 are pseudo-landmarks and they were not selected by the landmark criteria. However, they are points at the leaf margin with reference to landmarks 4 and 7 in 90-degree angle to the midrib. The digitisation was performed twice (two replicates) to evaluate digitisation error, shuffling the order of images (using TPSUtil) and one week apart, to reduce the effect of muscle memory.

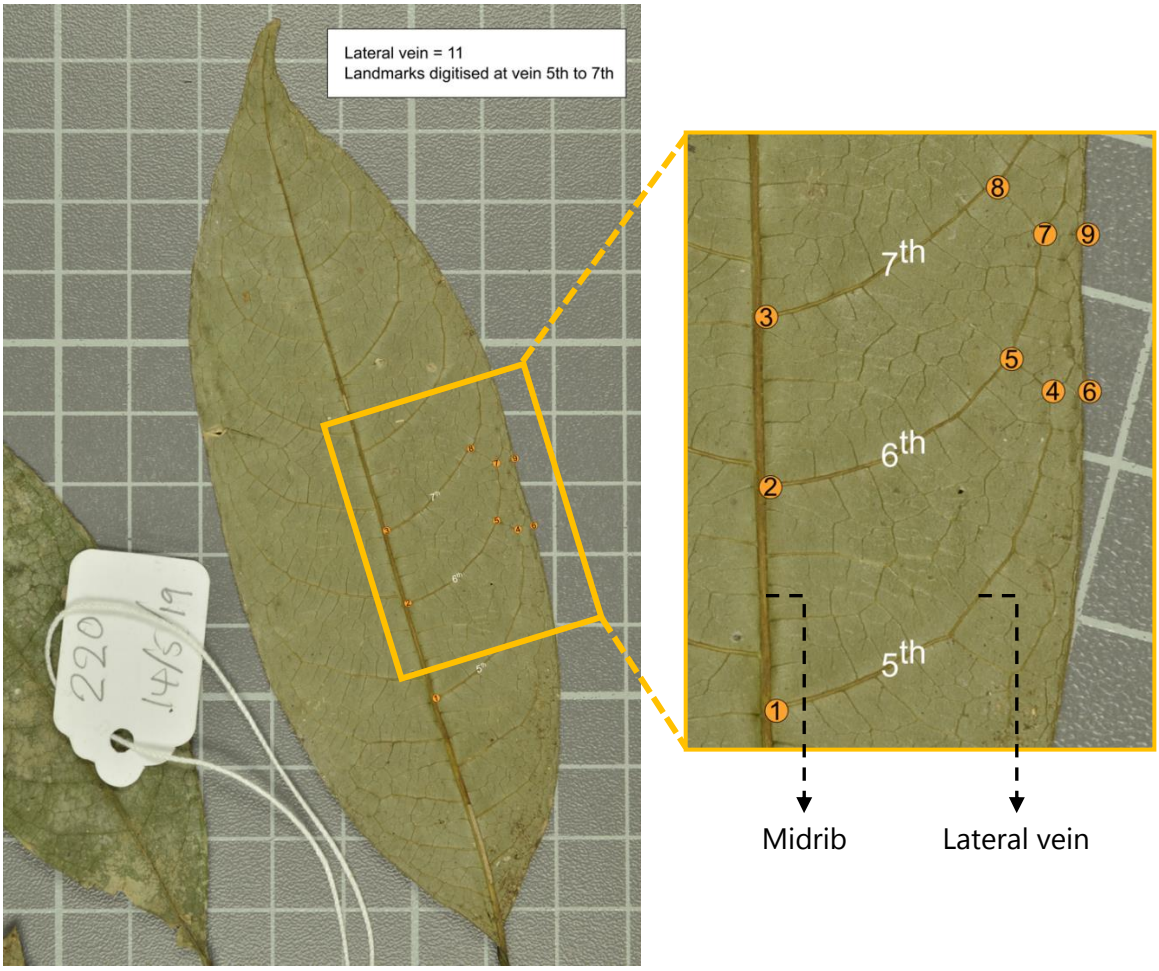


Figure 2.1 Example of Landmark digitisation on adaxial leaf side of *Urophyllum longipes* (LG). Orange circle with number indicates landmark point. White letters indicate lateral vein number.

b. Measuring size, performing Procrustes superimposition and testing variation within species, specimens, and leaves (Procrustes ANOVA)

After finishing landmark digitisation, the .tps file was used to perform further analyses in MorphoJ (Klingenberg, 2011). The processes are similar to the steps provided in the software tutorials on MorphoJ User's Guide webpage (https://morphometrics.uk/MorphoJ_page.html). The landmark coordinates were imported to MorphoJ and a Procrustes superimposition was performed.

Procrustes superimposition is a method to remove the effect of size from the shape and align samples into the same plane which involves scaling and rotating samples associated to the shape centroid (Bookstein *et al.*, 1999; Zelditch, Swiderski and Sheets, 2012). Therefore, measuring the size needed to be performed beforehand. The size commonly used in geometric morphometrics is the centroid size (Christodoulou, 2015). Centroid size is calculated by the square root sum of the squared distances of every landmark from shape centroid (see the equation below). This calculation can be done on MorphoJ by performing Procrustes superimposition.

$$\sqrt{\sum_{i=1}^{m} x_1^2 + x_2^2 + \cdots x_m^2}$$

x = distances between shape centroid and landmarks (1, 2, 3, ..., m)

After Procrustes superimposition, Procrustes coordinates were recorded, and any outliers were inspected using the built-in function "Find Outliers …". The classifiers used to group samples by categories during analyses were set to taxon, specimen, individual leaf, and replicate (error).

A hierarchical ANOVA was performed on the dataset to test not only variation within taxa, specimens, and leaves but also the digitisation error from recording any landmarks. The built-in function of Procrustes ANOVA in MorphoJ can only select one random (individual) effect, which means a comparison of the specimen to the taxon level could not be made. Therefore, the F ratio was computed manually following Viscosi and Cardini (2011) to modify the results from MorphoJ.

c. Testing taxon differences using Permutational ANOVA and multivariate analyses

Results from Procrustes ANOVA were observed and analysed to inform the next step. Favoured results show that the variation among leaves (for an individual), regardless of taxa and specimens, must be significantly larger than the digitisation error. This meant the digitisation error would explain a lower percentage of variance and can be negligible. The second result that could be evaluated is leaf variation within and among specimens (significant level in "Specimens" effect). If the leaves between specimens had a greater difference than within a specimen (statistically significant) then replicate leaves from a specimen can be averaged and used for further analyses. Therefore, differences between taxa will be tested in the next stage.

To test the size differences between taxa, centroid size and log centroid size data from Procrustes superimposition were averaged at the specimen level. Then the normality test was done using Shapiro-Wilk test in R. From the test, centroid size data were not normally distributed with P = 0.019. Therefore, the log centroid size (P = 0.631) was chosen. The data distribution was plotted using *boxplot* in R. In order to use ANOVA testing on group means' differences, not only the assumptions of normally distributed data must be met but also homogeneity of variances. Therefore, Bartlett's test was performed in R. The result accepted the null hypothesis with P = 0.489 meaning that the variances among groups are equal. Permutational one-way ANOVA was performed to evaluate whether there was a significant size difference among taxa. If this was the case, then Tukey's pairwise post-hoc tests were used to evaluate differences between each pair of taxa. These statistical analyses are performed on PAST3 (Hammer, Harper and Ryan, 2001).

Shape differences between taxa were tested using Principal Component Analysis and Canonical Variates Analysis with 1000 pairwise permutation test in MorphoJ.

2.3.3. Machine learning

Machine learning methods are performed for the identification of taxa. The data include: 1) twenty-one quantitative and qualitative characters of linear morphometric data; and 2) log centroid size and PC scores to represent size and shape changes of geometric morphometric data (Table S2.5). The correlation of all variables was tested prior to further analyses. All processes and analyses (Figure 2.2) in this section are performed using R.

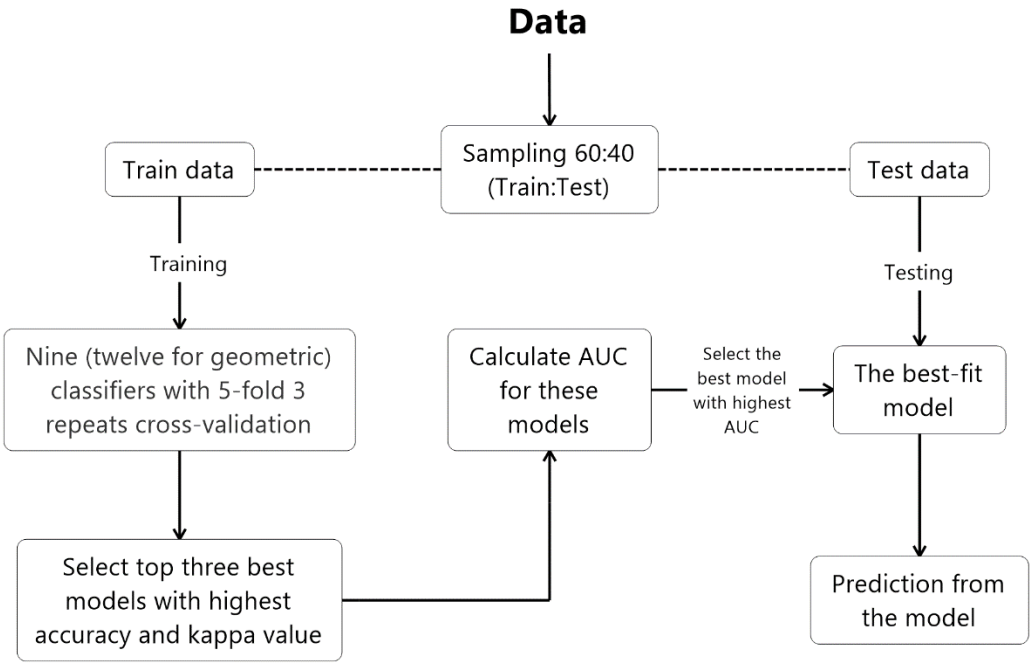


Figure 2.2 Flowchart showing workflow steps in machine learning.

2.3.3.3. Sampling training and testing sets

Random sampling of the datasets was performed to partition data into training and testing sets in the proportion 60:40, respectively. In order to evaluate the effect of specimen selection on the classifiers' performance, the sampling was repeated three times with the "set.seed" function, this meant that the datasets were partitioned identically for both the linear and geometric data.

2.3.3.4. Training different classification models (classifiers) to find the best choice and testing the model with testing data

The training sets were used in the validating step to train nine and 12 different classifiers for linear and geometric data, respectively (Table 2.3) (detailed description of classifiers can be found in Christodoulou *et al.* (2020)). The three repeats of 5-fold cross-validation were performed to evaluate a classifier's performance. Accuracy and kappa metrics were recorded to make a preliminary selection of the top three classifiers. Then, the Area Under the ROC Curve (AUC) was calculated for the classifiers using the testing sets. The best classifier was that with the highest AUC value and this was used for predicting the test

data, confusion matrices were plotted to observe the prediction. All steps were performed on R using the "caret" package (Kuhn, 2008) with additional packages required for nine classifiers presented in Table 2.3. AUC calculation was performed using "pROC" package (Robin *et al.*, 2011).

Table 2.3 Classifier descriptions. Classifier types summarised from Christodoulou *et al.* (2020).

Classifiers	Abbreviation	Classifier type	Additional packages
Bagged Classification and Regression Tree	BCART	Estimating boundaries	-
C4.5	c4.5	Estimating boundaries	RWeka (Hornik, Buchta and Zeileis, 2009)
Classification and Regression Tree	CART	Estimating boundaries	-
Conditional Inference Random Forest	cForest	Estimating boundaries	party (Hothorn, Hornik and Zeileis, 2008)
k-Nearest Neighbour	KNN	Non-parametric, density dependent	-
Linear Discriminant Analysis*	LDA	Parametric, density dependent	MASS (Venables and Ripley, 2002)
Mixture Discriminant Analysis*	MDA	Parametric, density dependent	mda (Hastie *et al.*, 2020)
Penalised Discriminant Analysis*	PDA	Parametric, density dependent	mda (Hastie *et al.*, 2020)
Random Ferns	rFerns	Probabilistic	rFerns (Kursa, 2012)
Random Forest	RF	Estimating boundaries	randomForest (Liaw and Wiener, 2002)
Regularised Random Forest	RRF	Estimating boundaries	RRF (Deng, 2013)
Support Vector Machine	SVM	Estimating boundaries	e1071 (Meyer *et al.*, 2019)

Note: * a classifier used in geometric data analysis only.

2.3.4. Combined data techniques

Two ways of combining techniques were used in this section. The first one is to create a combined dataset from linear and geometric datasets (direct combined datasets). The second technique is to combine the strength of the two best classifiers from both datasets (manually combine performance).

2.3.4.5. Direct combined datasets

The linear and geometric datasets were combined and partitioned into the training and testing sets as in 2.3.3.1. Then the training sets were used to train on nine classifiers shown in Table 2.3. The best fit model was selected and applied to the testing sets.

2.3.4.6. Manually combined performance

Combining predictions of multiple classifiers is a way to improve performance in machine learning. The combined performance of classifiers from different types is an ensemble method known as stacking (Witten *et al.*, 2017). In R, the "h2o" package (LeDell *et al.*, 2020) includes the method to do a multiclass stacking ensemble, however the fundamental assumption on datatype for each classifier is still needed to undertake multiclass stacking. In this study, the best classifier from the geometric dataset is PDA (see in result 2.4.2.4) which requires data to be normally distributed. A normal distribution cannot be obtained from qualitative data in linear morphometrics once a combined dataset includes linear and geometric data. Therefore, the performance of the two best classifiers from linear and geometric datasets were manually assembled in the same way as Christodoulou *et al.* (2018). This was based on an accuracy value of the class (so called 'balanced accuracy') and posterior probability estimate of a test sample from the prediction step of each classifier. Christodoulou *et al.* (2018) proposed four criteria for estimating the final prediction of the test sample:

1. If both classifiers agree on a prediction, the sample was classified as the agreed taxon.
2. If they disagreed and one classifier gained a balanced accuracy of more than 0.8 and accuracy of the other lower than 0.8. Then, the sample was classified as a taxon under the classifier with highest accuracy.

3. If they disagreed and both classifiers gained a high balanced accuracy (>0.8), then the sample was classified to a taxon under the classifier with the higher posterior probability estimate.
4. If both classifiers had a low balanced accuracy (<0.8), then posterior probability estimates were compared. The prediction from the classifier with the higher posterior probability estimate was selected.

2.4. Results

Vegetative characters of pistillate and staminate plants of *U. glabrum* (GL) and *U. longifolium* f.C (LC) were compared using PCA and LDA, shown in Figure 2.3–2.4, respectively. From Figure 2.3, samples of different sexes of both species were mixed, with no separation. The results are supported by the LDA analysis where pistillate and staminate samples of each species were mixed (Figure 2.4). Therefore, vegetative characters cannot be used successfully for sex identification, thus further analyses in this chapter were performed using all the samples of each species regardless of the sex of the sample.

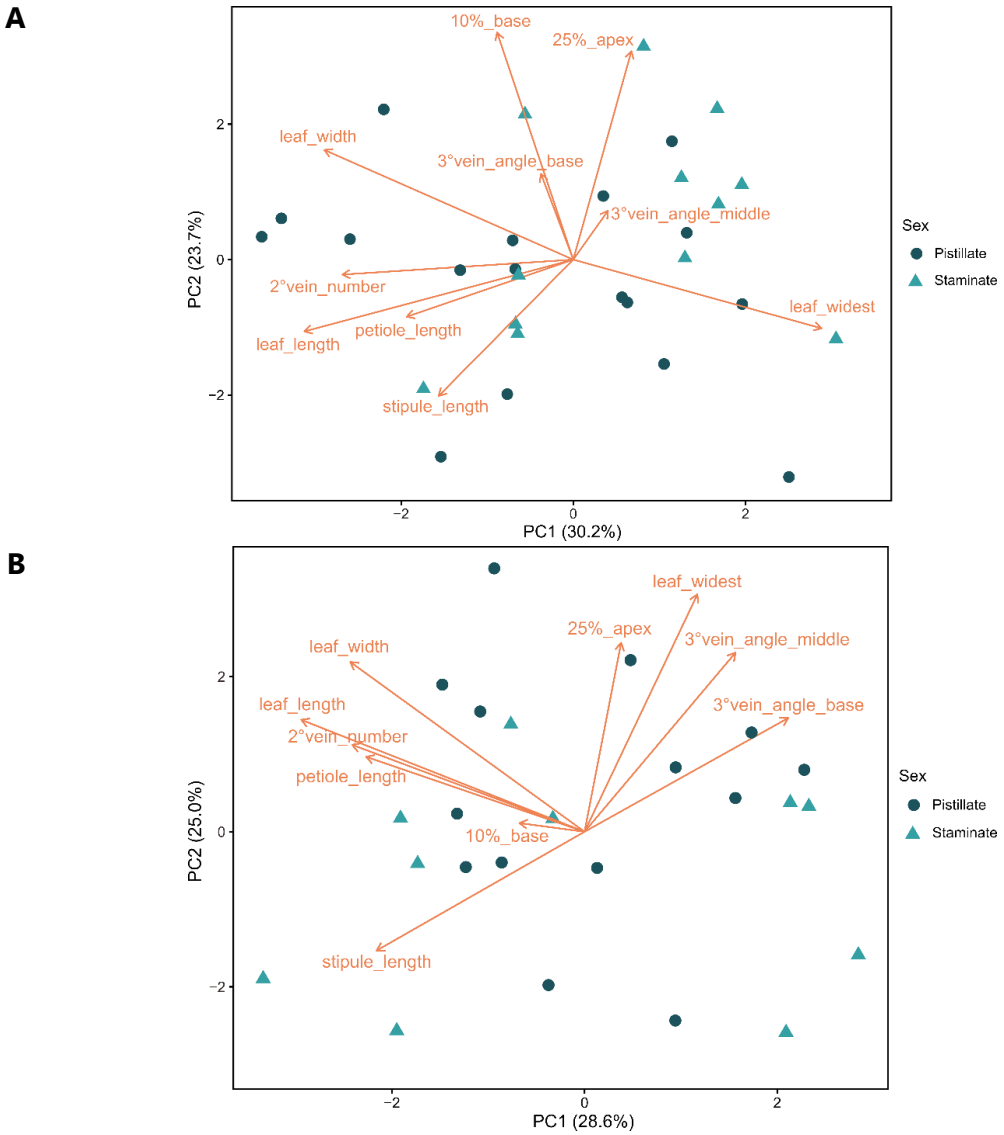


Figure 2.3 The first two principal components (PC1 and PC2) for ten vegetative characters of *U. glabrum* (A) and *U. longifolium* f.C (B). Dark green circles and light green triangles indicate pistillate and staminate samples, respectively. The orange arrows represent the loadings for the first two principal components.

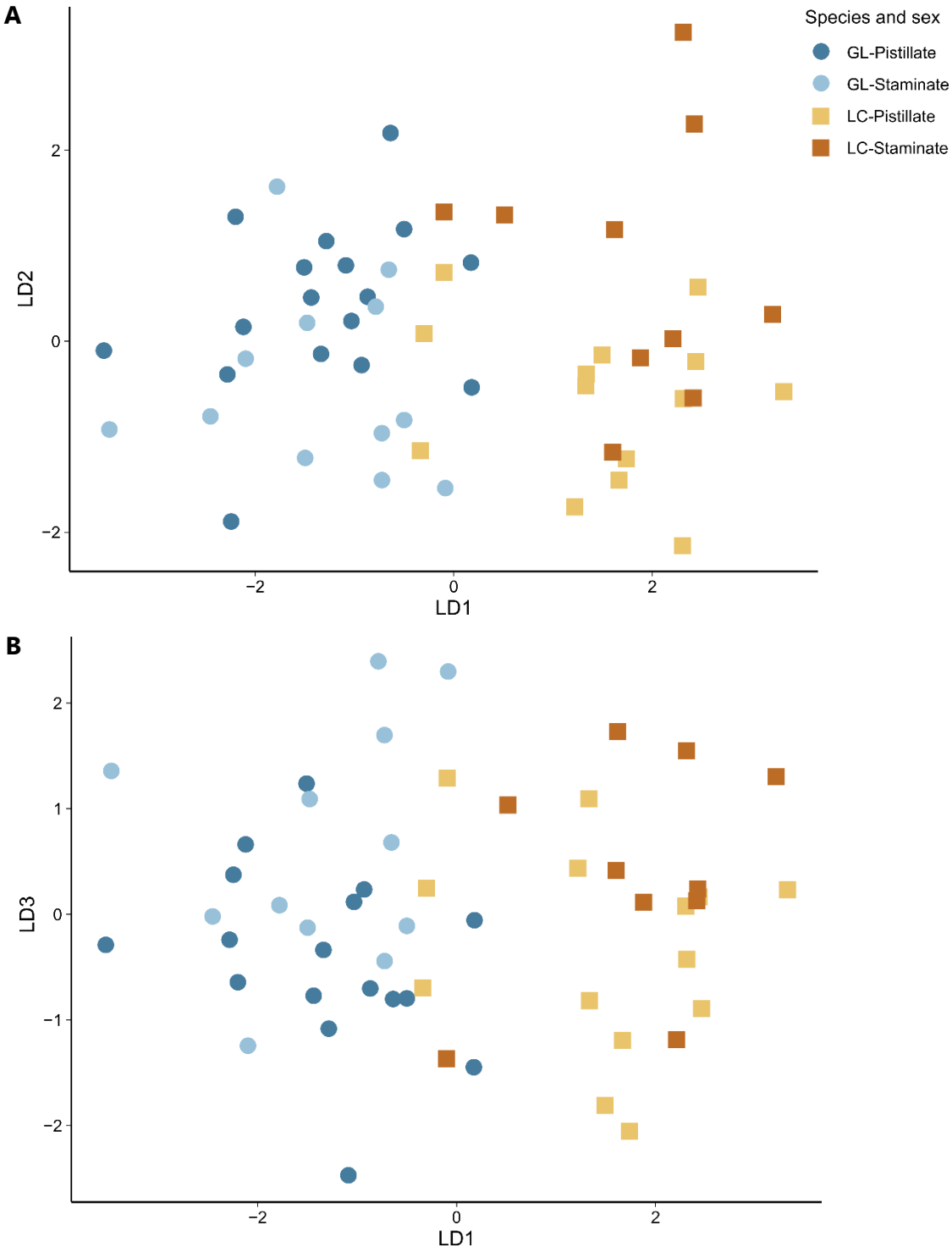


Figure 2.4 The three linear discriminants (LD1–LD3) for vegetative characters data. Colour shades represent species: light and dark blue are *U. glabrum*; light and dark brown are *U. longifolium* f.C.

2.4.1. Linear morphometrics

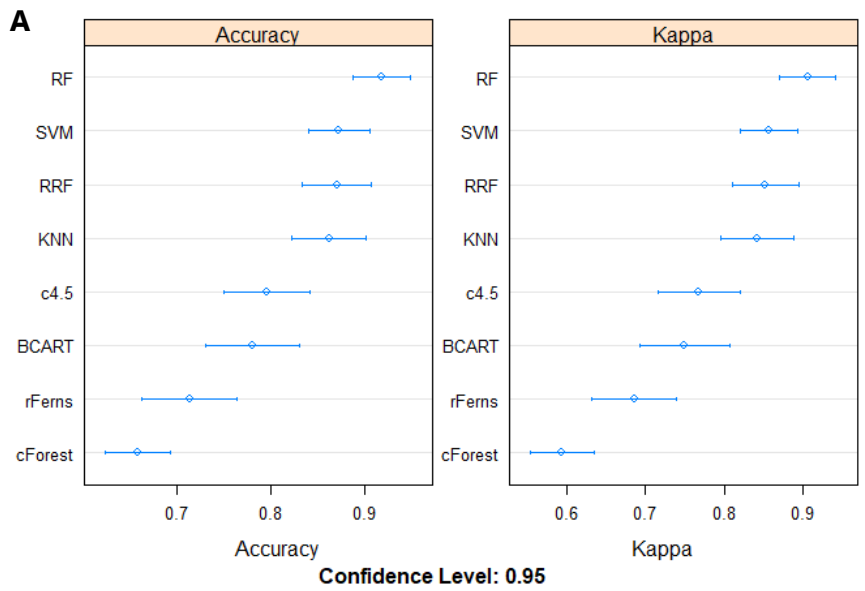
A correlation test revealed that covariances between characters are lower than 0.90 (Table S2.6). Therefore, all 21 variables were used in the analyses.

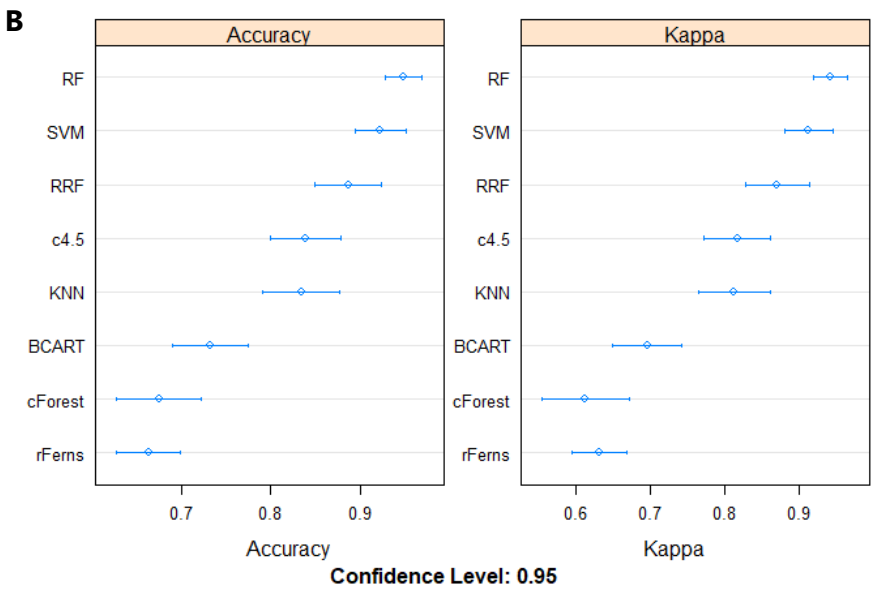
After data partitioning, the training sets were used in the validating step to train the nine classifiers with 5-fold cross-validation in three repetitions. The accuracy and kappa values from the Classification and Regression Tree (CART) were the lowest with accuracy ranging from 0.366–0.451 and kappa ranging from 0.277–0.371. These low values make the summary figure skewed, as the accuracy and kappa for the remaining classifiers are higher than 0.5. For this reason, CART was omitted from Figure 2.5.

The top three classifiers with highest accuracy and kappa values include Random Forest (RF), Support Vector Machine (SVM) and Regularised Random Forest (RRF) (Figure 2.5 and Table S2.10). These classifiers were therefore used to calculate AUC to find only one best model. The AUC values are summarised in Table 2.4. RF and SVM classifiers have similar AUC values in the three test sets. However, RF gained the highest AUC values for two out of three times. Therefore, RF was selected to be the best fit model for all three train sets and used as a classification model on the test sets (number of each taxon in test dataset are presented in Table 2.1). The RF model predictions are shown in a heat map confusion matrix in Figure 2.6.

Table 2.4 Area Under ROC Curve (AUC) for the top three highest classifiers from linear dataset (three times partitioning).

Rounds	Classifiers	Area under ROC curve (AUC)
1	SVM	0.994
	RF	0.992
	RRF	0.988
2	RF	0.994
	SVM	0.990
	RRF	0.988
3	RF	0.993
	SVM	0.987
	RRF	0.986





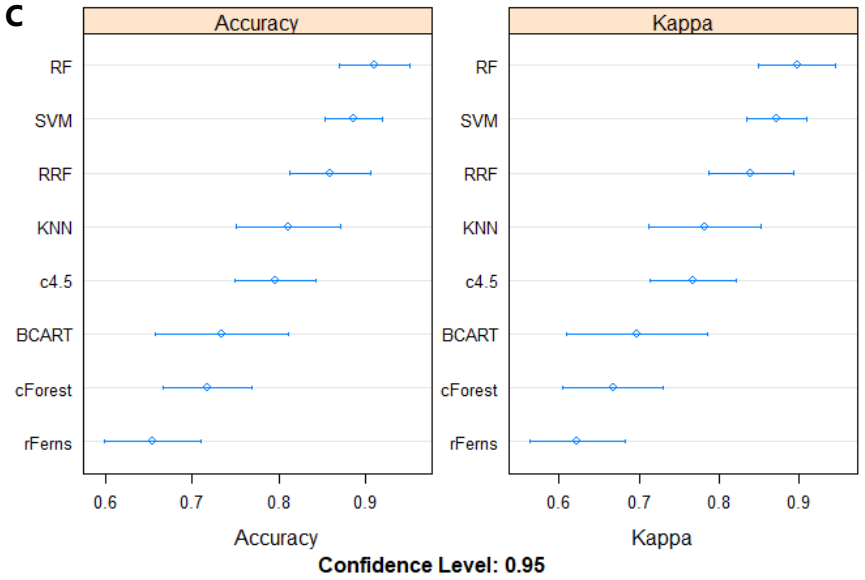
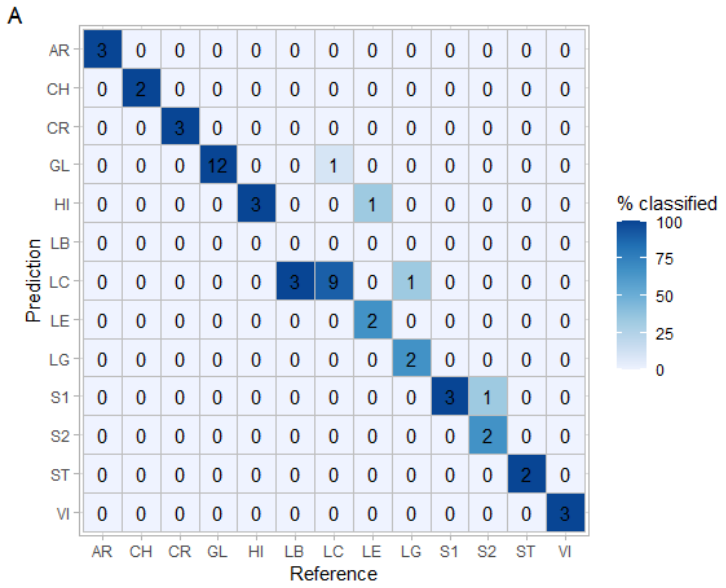
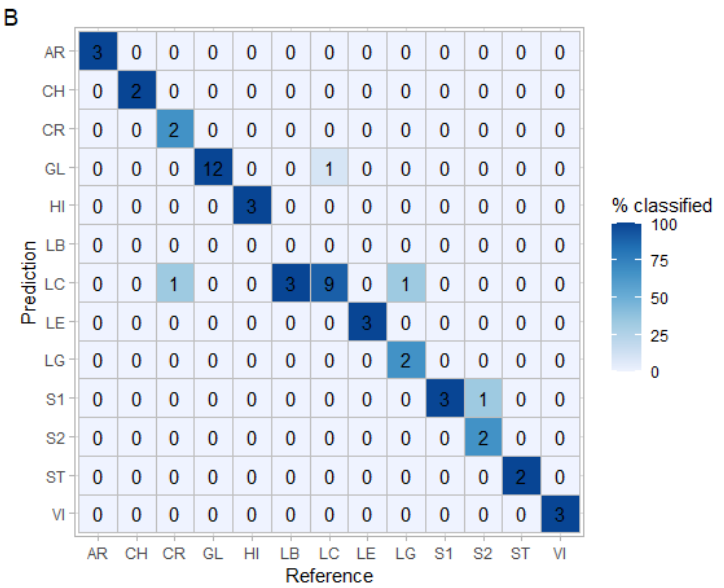


Figure 2.5 Average accuracy and kappa values of nine classifiers in three-time partitions (A, B and C) from training linear data set with 5-fold, 3 repeat cross-validation.





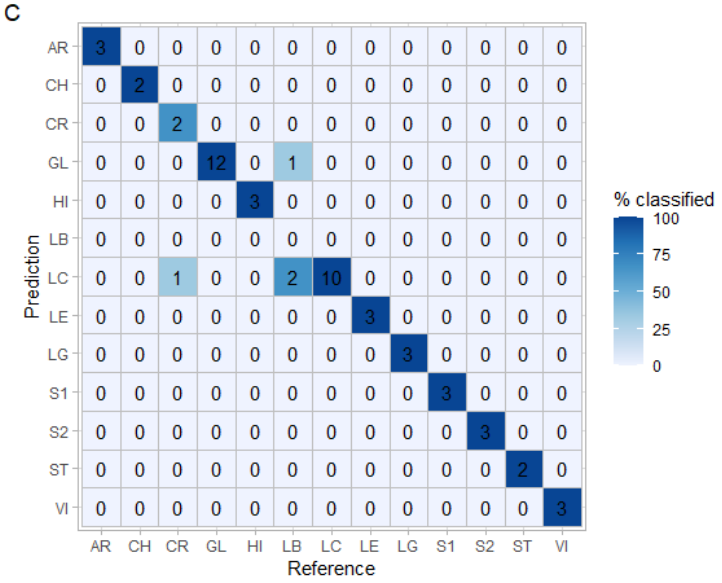


Figure 2.6 Confusion matrices from the Random Forest (RF) classification using the test linear datasets. Three replicates of partitioning are shown (A, B and C). Colours correspond to the percentage of classification in each category with numbers indicating the number of samples predicted for each taxon. The sample codes for each taxon follow Table 2.1.

From the confusion matrices, seven out of 13 taxa had a perfect match between predicted and reference classification (Figure 2.6). These are indicated by dark blue in the diagonal cells across the heatmap. One sample of *Urophyllum lecomtei* (LE) was misclassified in a testing round (Figure 2.6A). For four taxa (*U. crassum* (CR), *U. longifolium* f.C (LC), *U. longipes* (LG) and *U.* sp.2 (S2)), the classifier misclassified one sample of each in two out of three testing sets. In the case of *U. longifolium* f.B (LB), none of the samples were successfully identified in any testing sets.

To assess how misclassifications happened, the posterior probability provided from the classifier was recorded for all six taxa with at least one misidentified specimen. This value represents a probability that a test specimen can be assigned to a particular class. The posterior probability of misclassified taxa is summarised in Table 2.5. All specimens in a taxon were included (except for LC). If the specimen is correctly classified, only the posterior probability of the corrected class is presented (for LC, only specimens with the lowest and highest posterior probability are shown). If there was a misclassification, the posterior probability will be presented and ranked until the corrected class was assigned.

Table 2.5 Posterior probabilities for misclassifications. All taxa with at least one misclassification are included. If the sample was classified correctly, only posterior probability of corrected classification is shown.

Sampling round	Taxon	Sample 1	Sample 2	Sample 3
1	LB	1: LC (59.50%) 2: LB (21.70%)	1: LC (49.44%) 2: LB (22.50%)	1: LC (55.46%) 2: LB (20.70%)
	LC*	1: GL (39.98%) 2: LC (26.26%)	Sample 2–10 1: LC (38.38%–73.26%)	-
	LG	1: LC (24.30%) 2: LG (23.32%)	1: LG (51.40%)	1: LG (47.44%)
	LE	1: HI (29.96%) 2: LE (24.72%)	1: LE (39.66%)	1: LE (42.74%)
	S2	1: S1 (27.90%) 2: S2 (18.68%)	1: S2 (47.42%)	1: S2 (56.06%)
2	CR	1: LC (41.26%) 2: CR (24.42%)	1: CR (57.00%)	1: CR (32.72%)
	LB	1: LC (67.14%) 2: LB (18.52%)	1: LC (56.38%) 2: LB (25.66%)	1: LC (47.02%) 2: LB (32.38%)
	LC*	1: GL (40.92%) 2: LC (26.24%)	Sample 2–10 1: LC (39.00%–77.50%)	-
	LG	1: LC (37.96%) 2: LG (19.18%)	1: LG (55.78%)	1: LG (48.64%)
	S2	1: S1 (29.74%) 2: S2 (15.40%)	1: S2 (55.10%)	1: S2 (51.48%)
3	CR	1: LC (42.86%) 2: CR (22.00%)	1: CR (55.36)	1: CR (33.08%)
	LB	1: GL (54.08%) 2: LC (22.94%) 3: LB (7.32%)	1: LC (24.06%) 2: GL (19.64%) 3: LG (17.44%) 4: LB (9.24%)	1: LC (44.06%) 2: LB (33.84%)

Note: * taxon with 10 total test samples.

From the posterior probability, LB always had misclassified specimens. The misidentification, within all rounds, mostly failed to separated LB from LC, followed by GL in the last round. Misidentification between LC and GL were recognised in the first two rounds with only one specimen that failed to classify to the right taxon. These misidentifications represent the difficulty of separating these three taxa (LB, LC and GL) using the classifier.

Other cases of misclassification include one sample of CR and LG, both failed to be separated from LC. Similarly, S2 and LE failed to be separated from a specimen of S1 and HI, respectively. In total, there were seven misclassifications out of 53 specimens tested on the first and second rounds, and four misclassifications in the last round. In the first and second rounds, all seven misidentifications have the corrected class as the second choice. Within the first round, only one misclassified sample had a posterior probability of the corrected class lower than 20%. In the second round up to three samples had a posterior probability lower than 20%. In the third round, two out of four samples were misclassified with a posterior probability lower than 10% for the corrected class.

2.4.2. Geometric morphometrics

2.4.2.1. Digitisation error and variation between specimens and taxa

After landmark digitisation, the Procrustes ANOVA was performed on both the centroid size and shape data of the two replicates. From both data, the mean square calculated from the digitisation error accounted for 0.3% and 4.1% of the total sum of squares which was smaller than the error calculated among individual leaves ($P < 0.0001$; Table 2.6a and 2.6b). This suggests that digitisation error is negligible.

Another point on Procrustes ANOVA is the variation of individual leaves within and among specimens. From the Table 2.6, the mean square estimate from leaves was smaller than that from specimens, meaning that the difference between leaves was smaller than between specimens. Therefore, the leaves collected from the same specimen were averaged. The averaged leaf shape and centroid size were used to perform taxon classification in the later steps.

Table 2.6a Procrustes ANOVAs for measurement error for size (centroid size). SS, sum of squares; MS, mean square; df, degrees of freedom; F, F-value; P, P-value.

Effect	SS		MS	df	F	P
Taxa	63898.131	56.7%	5324.844	12	17.409	<.0001
Specimens	39457.293	35.0%	305.870	129	1.716	0.0156
Leaves (individual)	8914.456	7.9%	178.289	50	89.261	<.0001
Residual (digitisation error)	383.501	0.3%	1.997	192		
Total	112653.381	100.0%				

Table 2.6b Procrustes ANOVAs for measurement error for shape. SS, sum of squares; MS, mean square; df, degrees of freedom; F, F-value; P, P-value.

Effect	SS		MS	df	F	P
Taxa	2.622	42.2%	0.0156	168	10.726	<.0001
Specimens	2.628	42.3%	0.0015	1806	1.431	<.0001
Leaves (individual)	0.712	11.4%	0.0010	700	10.691	<.0001
Residual (digitisation error)	0.256	4.1%	0.0001	2688		
Total	6.217	100.0%				

2.4.2.2. Size differences (centroid size)

A one-way ANOVA was performed to test the size difference of the secondary vein loops by using the log transformed centroid size data. Arithmetic means can be found in Table S2.7. The result from the ANOVA suggests that there were statistical differences between taxa mean size with $P < 0.0001$ (See full ANOVA table in Table S2.8). The Tukey's pairwise test was performed to evaluate the difference between data pairs showing that *U. villosum* (VI) is significantly different in size from other taxa except *U. crassum* (CR) (Table S2.9). This result is consistent with what is presented in the box plot which shows VI had a larger median log centroid size than other taxa (Figure 2.7).

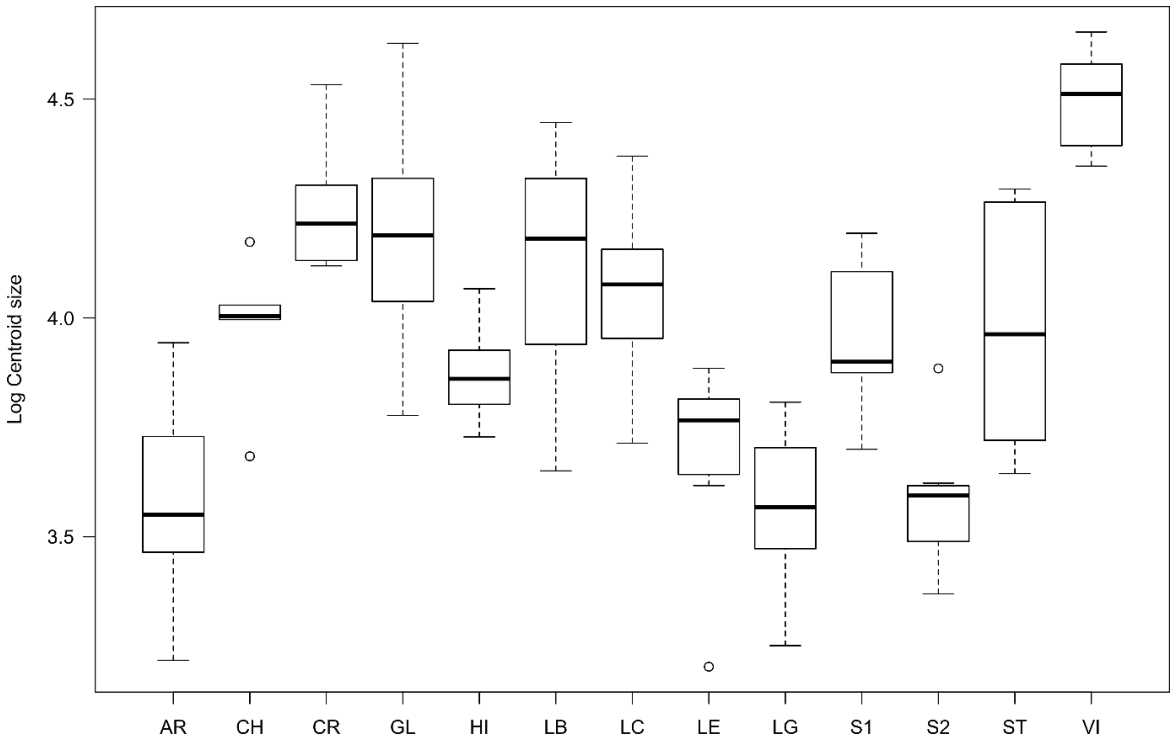


Figure 2.7 Box plot of secondary vein loop log centroid size in *Urophyllum* taxa. Black line represents median, box represents interquartile range, and open circles indicate outlier data points further than 1.5 times interquartile range from the box. The plot was drawn in R. In the case of CH, there is no line connecting the highest and lowest data points due to the small number of samples (5 in total) with two samples being outliers.

2.4.2.3. Shape differences

Shape differences were studied using Principal Component Analysis (PCA) and Canonical Variates Analysis (CVA). From the first three PCs and CVs which represent 92% and 81%, respectively (Figure 2.8 and 2.9), the shape of *U. longipes* (LG) was clearly separated from other groups. *U. hirsutum* (HI) and *U.* sp.2 (S2) shapes were separated from other taxa in CVA but not in PCA. The results from these analyses suggest that some taxa can be separated using the shape of the secondary vein loops. To compare the shape difference between each taxon pair, a permutation test and Mahalanobis distance results were recorded from CVA.

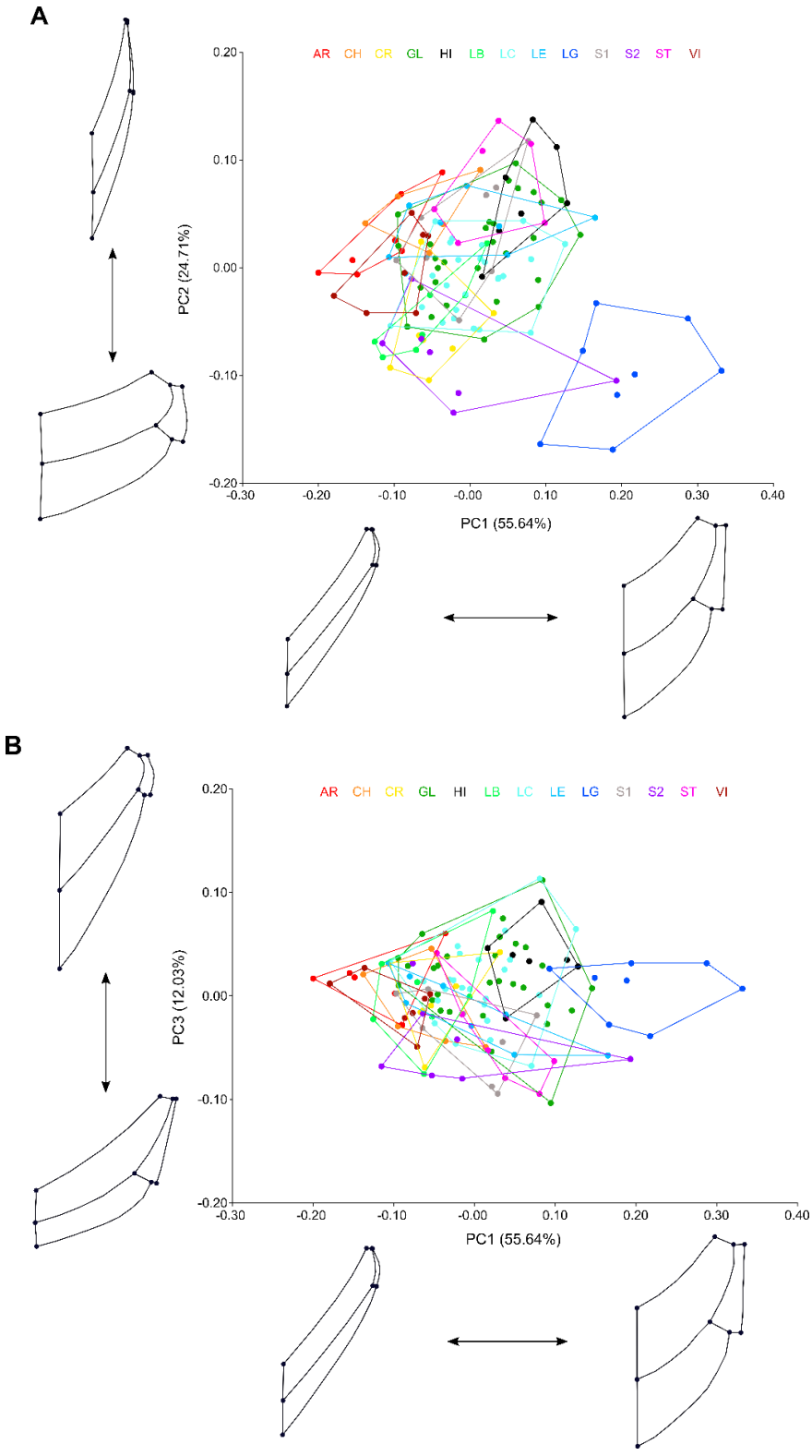


Figure 2.8 Principal Components Analysis of *Urophyllum* secondary vein loop shape. (A) PC1 vs PC2; (B) PC1 vs PC3. Convex polygons drawn around individuals from each taxon. Outlines illustrate shape changes along each principal component from −0.2 to 0.2 for PC2 and PC3 and -0.2 to 0.3 for PC1. Taxon codes follow Table 2.1.

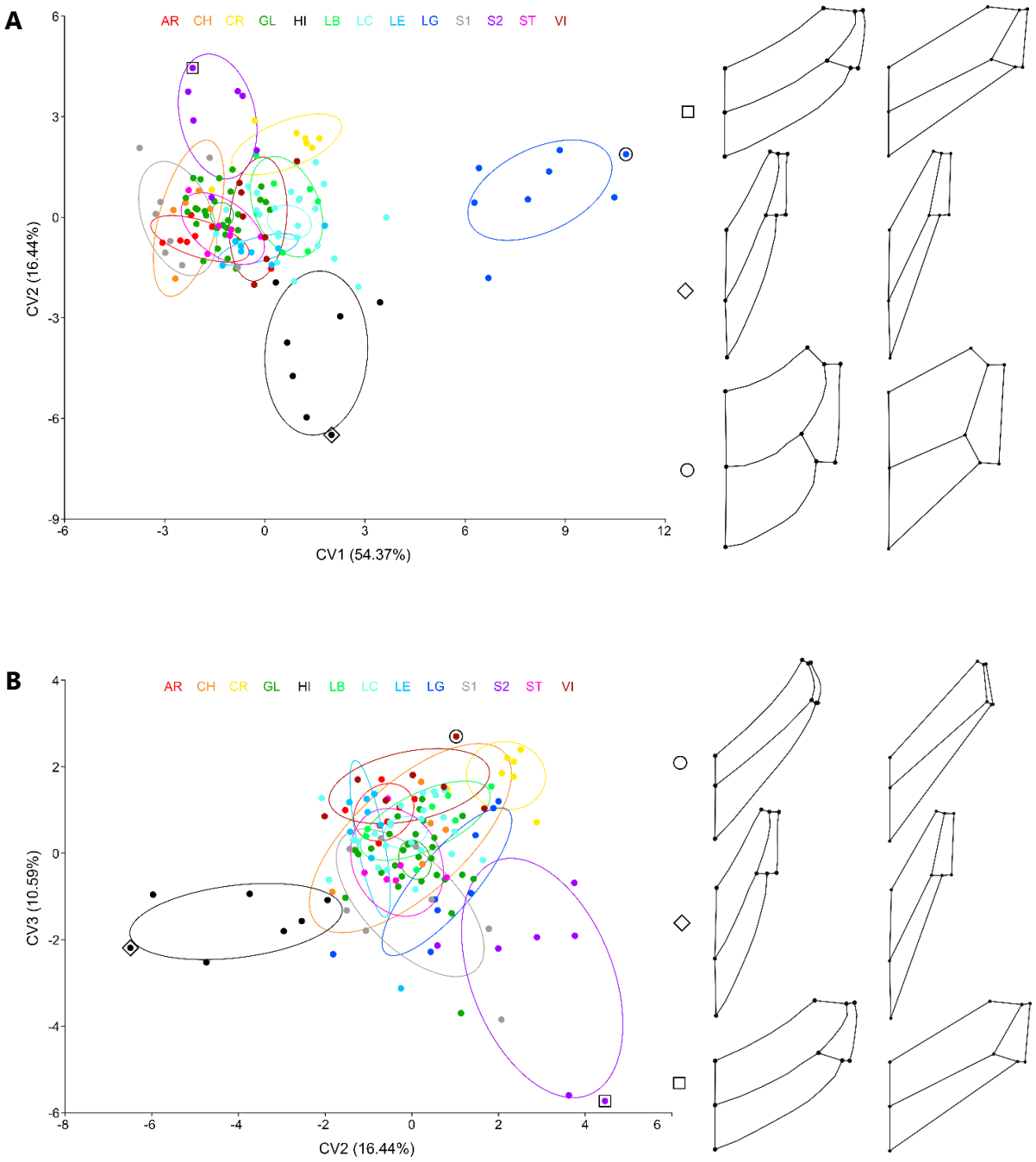


Figure 2.9 Canonical Variates Analysis of *Urophyllum* secondary vein loop shape. (A) CV1 vs CV2; (B) CV2 vs CV3. Confidence ellipse shows 95% probability. Outlines and wireframes illustrate shape changes at the points indicated with different symbols in the plot. Taxon codes follow Table 2.1.

Table 2.7 p-values from permutation tests (10000 permutation rounds) for Mahalanobis distances among taxa. Upper-right triangle exhibits p-values. Lower-left triangle indicates significance levels. Different significance levels are shown with asterisks: * = 95%; ** = 99%; *** = 99.9%. NS means not significant at 95% confidence intervals. Taxon codes follow Table 2.1.

	AR	CH	CR	GL	HI	LB	LC	LE	LG	S1	S2	ST	VI
AR		0.2437	0.0004	<.0001	<.0001	0.0008	<.0001	0.0039	<.0001	0.0015	<.0001	0.0013	0.0065
CH	NS		0.0024	0.0006	0.0011	0.0013	<.0001	0.0275	0.0003	0.0973	0.0026	0.0666	0.0023
CR	***	**		<.0001	0.0001	0.0136	<.0001	0.0001	<.0001	0.0002	0.0002	0.0013	0.0067
GL	***	***	***		<.0001	<.0001	<.0001	<.0001	<.0001	0.0004	<.0001	<.0001	<.0001
HI	***	**	***	***		<.0001	<.0001	0.0007	0.0001	0.0002	0.0004	0.0011	<.0001
LB	***	**	*	***	***		0.3758	0.0099	<.0001	<.0001	0.0005	<.0001	0.1703
LC	***	***	***	***	***	NS		0.0001	<.0001	<.0001	<.0001	<.0001	0.0051
LE	**	*	***	***	***	**	***		<.0001	0.0005	0.0001	0.039	0.1223
LG	***	***	***	***	***	***	***	***		0.0001	<.0001	<.0001	<.0001
S1	**	NS	***	***	***	***	***	***	***		0.0015	0.0051	<.0001
S2	***	**	***	***	***	***	***	***	***	**		<.0001	0.0002
ST	**	NS	**	***	**	***	***	*	***	**	***		0.001
VI	**	**	**	***	***	NS	**	NS	***	***	***	***	

The results of the permutation test and Mahalanobis distances between taxa after performing the CVA are summarised in Table 2.7. The table shows that many taxa are significantly different in shape from one another. Four taxa significantly different from others include *U. crassum* (CR), *U. glabrum* (GL), *U. hirsutum* (HI), and *U. longipes* (LG). In these cases, we expect that the classifier in the machine learning section could be highly accurate in classifying these four taxa using only shape data alone. On the other hand, *U. chinense* (CH) might be harder to distinguish from *U. argenteum* (AR), *U. streptopodium* (ST) and *U.* sp.1 (S1) because there is no significant difference between them.

2.4.2.4. Classification using machine learning

The training sets including log centroid size and the first 10 PC scores (explaining 99.9% of shape variation) were used to train the 12 classifiers shown in Table 2.3. The average accuracy and kappa values from 5-fold, three repeats cross-validation are reported in Figure 2.10. The six classifiers are LDA, PDA, MDA, SVM, RRF and RF, which gaining the highest accuracy and kappa values. Therefore, these classifiers were used to calculate AUC to select the most appropriate classifier for geometric data. The AUC values are presented in Table 2.8.

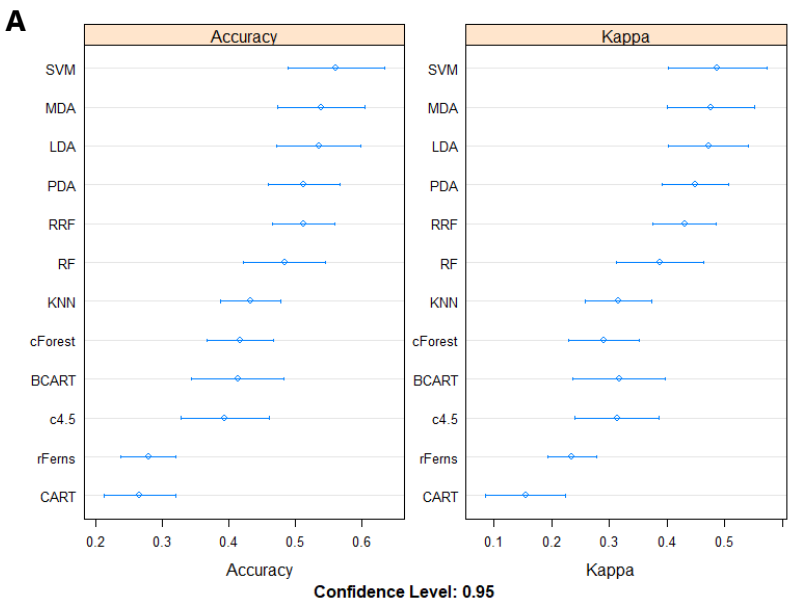
There were two classifiers showing a similar AUC value: Linear Discriminant Analysis (LDA) and Penalised Discriminant Analysis (PDA). They were selected to present the prediction for all partitions in test datasets. After testing, PDA and LDA classifiers presented very similar predictions with the accuracy and kappa values slightly higher in PDA (Table 2.9). Therefore, the confusion matrix heat map presented here is based upon PDA results only (Figure 2.11). For LDA, the heat map result is illustrated in Figure S2.1.

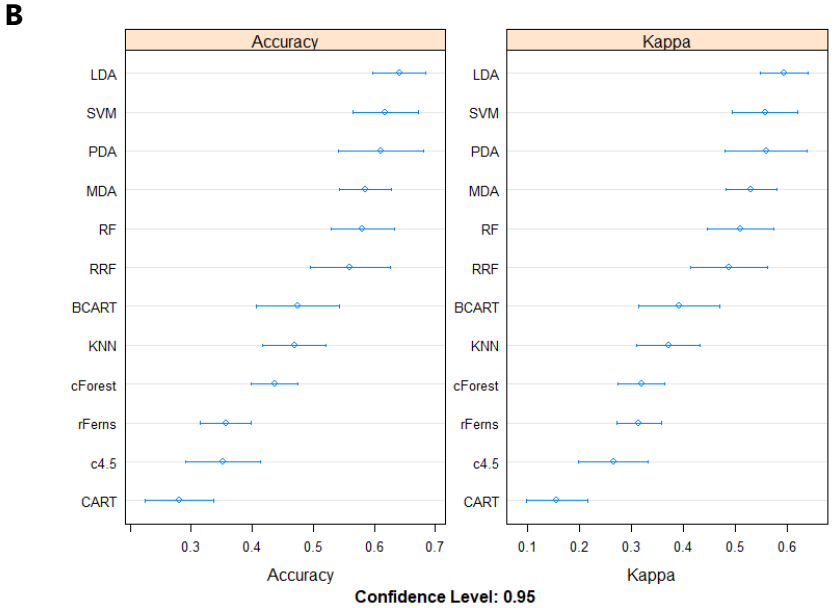
Table 2.8 Area Under ROC Curve (AUC) of the top six highest classifiers from geometric dataset (three times partitioning).

Round	Classifier	Area Under ROC Curve (AUC)
1	LDA	0.941
	PDA	0.938
	MDA	0.888
	SVM	0.874
	RRF	0.851
	RF	0.873
2	LDA	0.900
	PDA	0.899
	MDA	0.895
	SVM	0.843
	RRF	0.818
	RF	0.810
3	LDA	0.921
	PDA	0.920
	MDA	0.879
	SVM	0.868
	RRF	0.867
	RF	0.855

Table 2.9 Accuracy and kappa value after prediction on test set using LDA and PDA as classifiers.

Round	Classifier	Accuracy	Kappa
1	LDA	0.566	0.507
	PDA	0.585	0.530
2	LDA	0.604	0.548
	PDA	0.623	0.573
3	LDA	0.604	0.543
	PDA	0.623	0.566





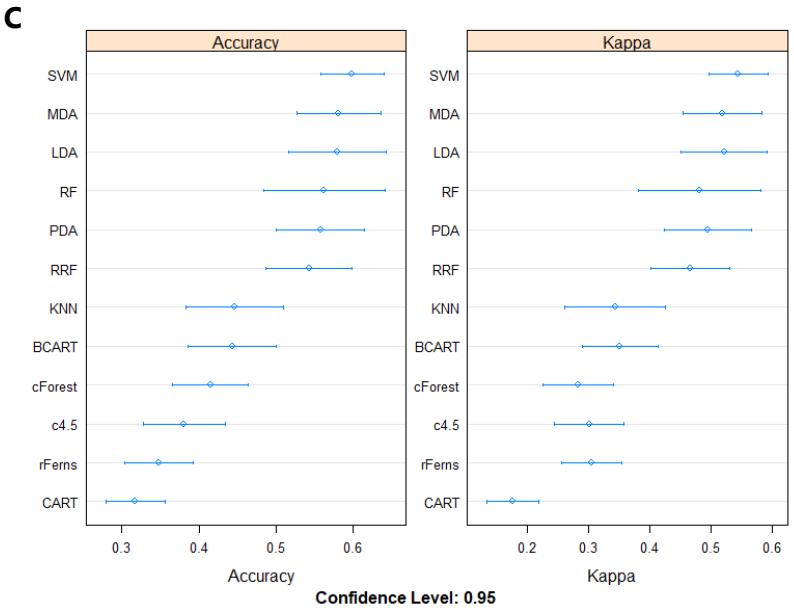
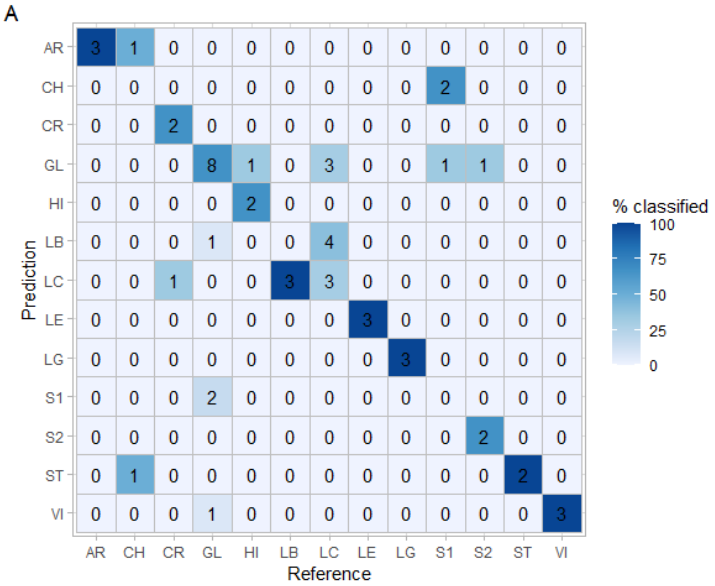
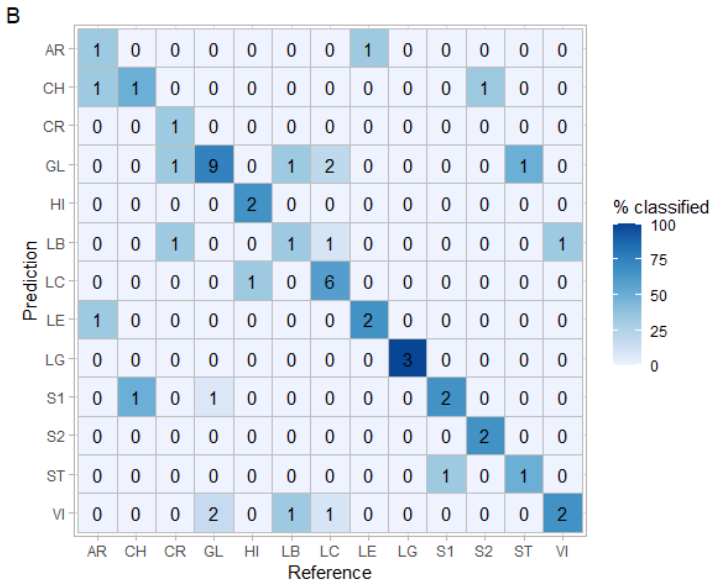


Figure 2.10 Average accuracy and kappa values of nine classifiers from training geometric dataset with 5-fold, 3 repeats cross-validation.





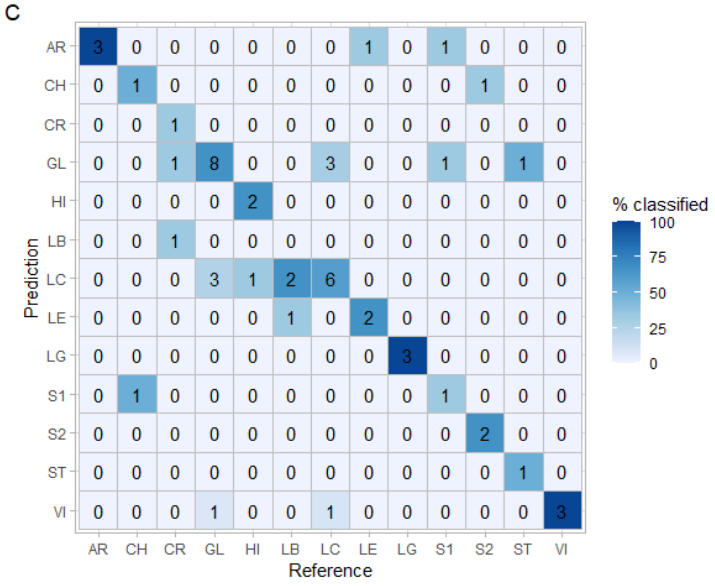


Figure 2.11 Confusion matrices from the Penalised Discriminant Analysis (PDA) classification using the test geometric datasets. Three replicates of partitioning are shown (A, B and C). Colours correspond to the percentage of classification in each category with numbers indicating the number of samples predicted for each taxon. The sample codes for each taxon follow Table 2.1.

The test results of PDA show that 100% of the *U. longipes* (LG) specimens were classified correctly, with no misclassification in any round. For *U. argenteum* (AR) and *U. villosum* (VI), there were two rounds of perfect classification and one round with misclassification. For *U. lecomtei* (LE), only one round predicted the right classification for all samples. Misclassifications were common in the geometric datasets; this suggests that the shape selected in this study is not sufficiently different when comparing all of the taxa.

2.4.3. Combined datasets

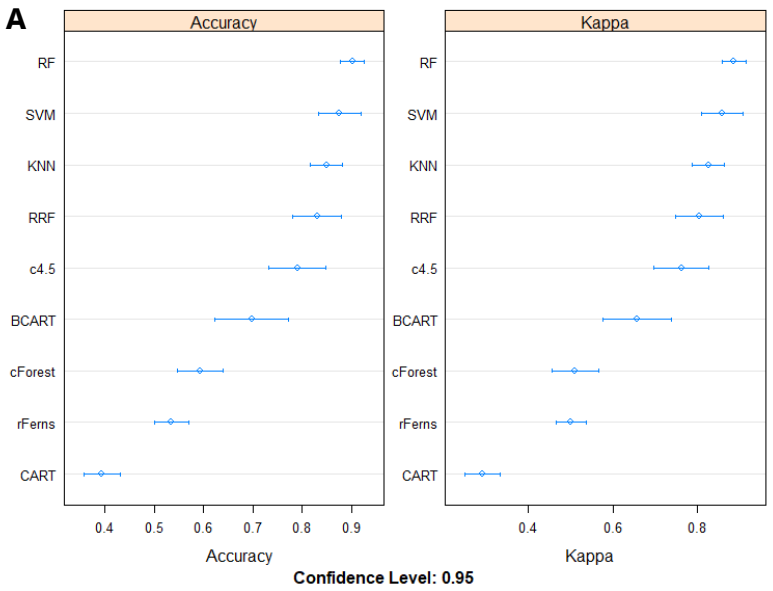
2.4.3.1. Direct combined datasets

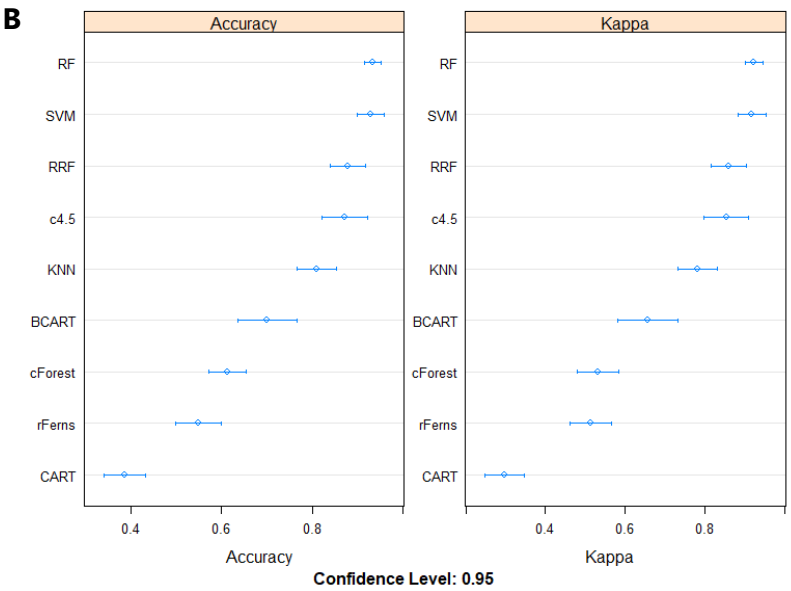
After training with 10 classifiers, the accuracy and kappa values were recorded as shown in Figure 2.12. The top five with highest values were RF, SVM, RRF, KNN and c4.5. Like previous datasets, the AUC was calculated for these classifiers (Table 2.10), then the classifier with highest AUC was selected to predict classification of the test dataset.

Based on AUC value, Random Forest (RF) is the most suitable classifier for the combined dataset. The confusion matrix of the test set prediction is presented in Figure 2.13.

Table 2.10 Area Under ROC Curve (AUC) of the top five highest classifiers from the combined dataset (three times partitioning).

Round	Classifier	Area under ROC curve (AUC)
1	RF	0.996
	SVM	0.994
	RRF	0.992
	KNN	0.990
	c4.5	0.956
2	RF	0.991
	SVM	0.987
	RRF	0.983
	KNN	0.985
	c4.5	0.885
3	RF	0.990
	SVM	0.972
	RRF	0.982
	KNN	0.978
	c4.5	0.867





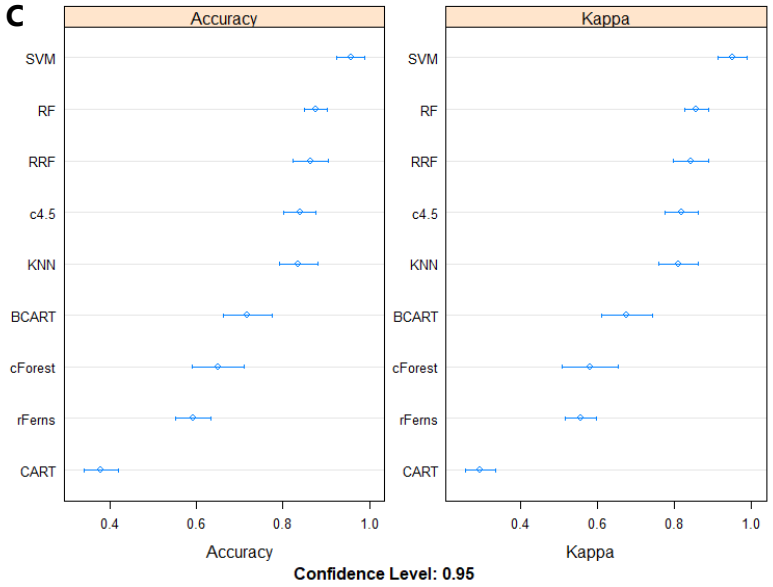


Figure 2.12 Average accuracy and kappa values of nine classifiers from training combined dataset with 5-fold, 3 repeats cross-validation.

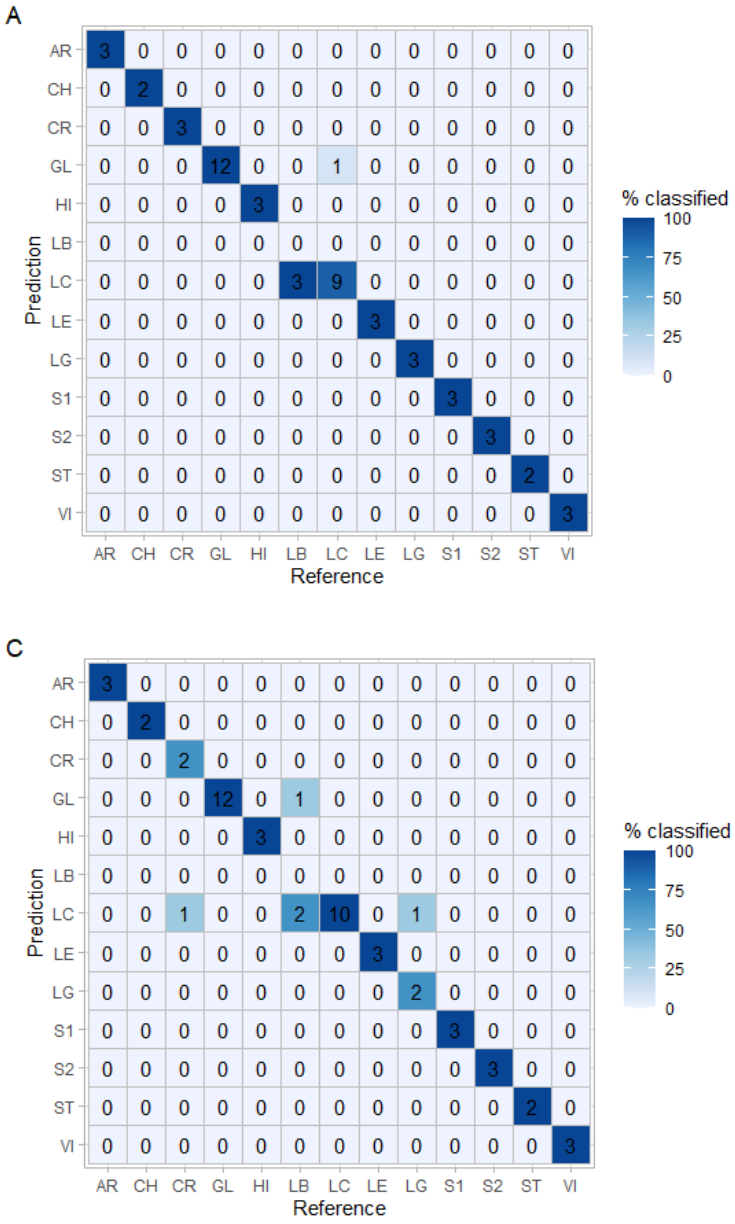
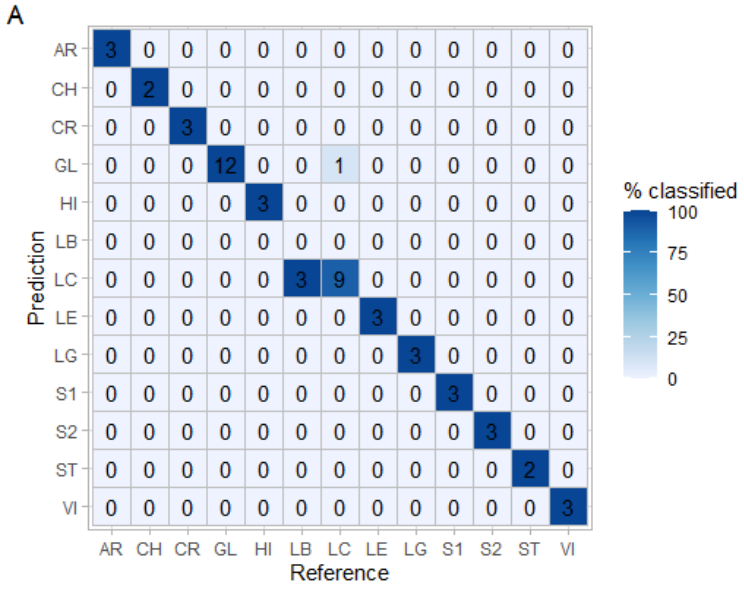


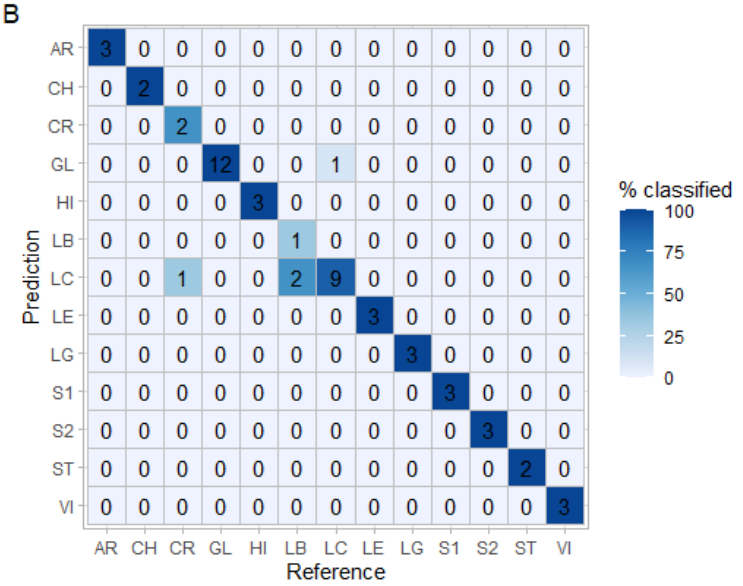
Figure 2.13 Confusion matrices from the Random Forest (RF) classification using the test combined datasets. Three replicates of partitioning are shown (A, B and C). Colours correspond to the percentage of classification in each category with numbers indicating the number of samples predicted for each taxon. The sample codes for each taxon follow Table 2.1.

2.4.3.2. Manually combined performance

The results of the manual prediction using a combination of two classifiers is shown in Figure 2.14. Samples from ten taxa were successfully identified to the right class. Two taxa (*U. crassum* (CR) and *U. longifolium* f.C (LC)) had one sample misclassified in two out of the three times they were tested. In CR, one out of the three samples was misclassified to LC. For LC, a sample was misclassified to *U. glabrum* (GL). *U. longifolium* f.B (LB) was always misclassified to LC, which is similar to the results in the previous three datasets.

Table 2.11 summarises the prediction rates of classifiers trained from different datasets. Overall, the manually combined performance outperformed the other classifiers with 91% averaged prediction, followed by RF in the direct combined datasets 87%, RF (linear) 86%, and PDA (geometric) 60%. In manual ensemble and both RFs, the classifiers can identify 12 taxa with a correct prediction rate higher than 78%. One exception was *U. longifolium* f.B (LB), it is not recognised in linear and combined analyses and has a correct prediction of just 11% in the ensemble method. Misclassifications of LB were usually misplaced to *U. longifolium* f.C (LC). This suggests that these two taxa are highly similar.





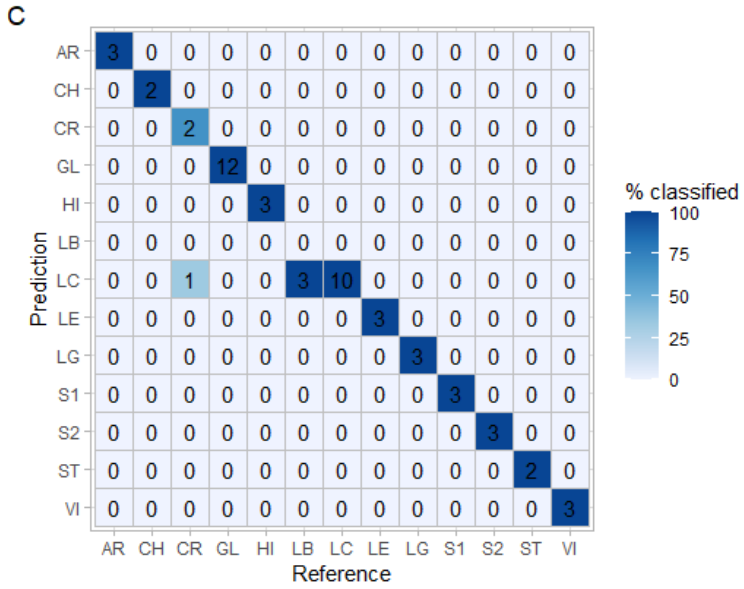


Figure 2.14 Confusion matrices from the Random Forest (RF) classification using the test manually combined datasets. Three replicates of partitioning are shown (A, B and C). Colours correspond to the percentage of classification in each category with numbers indicating the number of samples predicted for each taxon. The sample codes for each taxon follow Table 2.1.

Table 2.11 Summary of prediction rates from the best classifiers of the test datasets. The percentages were calculated by averaging the total samples from three rounds of partitioning of the test set.

Taxon code	RF: linear morphometrics	PDA: geometric morphometrics	RF: combined linear and geometric morphometrics	Manually combined performance
AR	100%	78%	100%	100%
CH	100%	33%	83%	100%
CR	78%	44%	78%	78%
GL	100%	69%	100%	100%
HI	100%	67%	100%	100%
LB	0%	11%	0%	11%
LC	93%	50%	93%	93%
LE	89%	78%	100%	100%
LG	78%	100%	89%	100%
S1	100%	33%	100%	100%
S2	78%	67%	89%	100%
ST	100%	67%	100%	100%
VI	100%	89%	100%	100%
Averaged prediction rates	**86%**	**60%**	**87%**	**91%**

2.5. Discussion

2.5.1. No perfect classifier for all datasets

According to the results, the best classifier for linear and geometric datasets are not the same. The best performing classifier for the linear dataset was Random Forest (RF) and for the geometric datasets was Penalised Discriminant Analysis (PDA). These patterns were similar to Christodoulou *et al.* (2018), and provides another example of the "No free lunch" theorem (Wolpert and Macready, 1997). The theory describes that the average performance of any two classifiers is equal when all the problems are accounted for. This means that the best classifier will not be the same for all problems. In this study, the two datasets each represent a classification problem that was best solved by different classifiers.

2.5.2. Automated learning for *Urophyllum* identification

The main aim of this chapter is to identify *Urophyllum* taxa in Thailand and to some extent Vietnam based upon external morphological characters and shape variation. This included two datasets, linear (both quantitative and qualitative data) and geometric data. For the linear dataset, the RF classifier can predict unknown specimens with a high averaged success rate (86%) and 100% correct prediction in seven taxa based upon 21 characters used. In contrast, the geometric data which represents the shape of the secondary vein closed loop, had a 60% successful prediction rate, but had an accurate prediction rate of 100% for *U. longipes* using PDA. This prediction of *U. longipes* using geometric data implies that the taxon can be distinguished using the shape of the secondary vein loop, this character would not be gathered in linear data. The different prediction rate for each taxon using different classifiers suggests combining datasets and an ensemble classifiers' performance would be more appropriate.

Combining methods presented a higher accuracy of prediction than individual datasets. The prediction success of Random Forest from the direct combined dataset was better than the prediction on linear data for *U. lecomtei*, *U. longipes,* and *U.* sp.2. However, it was worse in successfully predicting *U. chinense* specimens. It seems that merging the two datasets resulted in more noise in the classification process which is similar to the findings

in apple cultivar identification studied by Christodoulou *et al.* (2015). However, the ensemble method shows a different perspective. It not only provides an improved prediction for three taxa but the prediction is also 100% accurate for 10 of the 13 taxa sampled. Moreover, the prediction success is not lower for any taxa in the combined dataset (average prediction rate 91%), compared to the linear dataset (average prediction rate 86%). The success of the classification may be caused by the ensemble method utilising the strength of each classifier for accurately predicting a particular taxon. The ensemble method also worked better than other methods in apple cultivar identification (Christodoulou, Battey and Culham, 2018). This suggests the usefulness of the ensemble method in plant identification which should be considered when supervised machine learning is performed.

Persistent misclassification of *U. longifolium* f.B to *U. longifolium* f.C for all classifiers was not surprising. Morphological differences between these two taxa are hair density and the angle of hairs on the stipule (Figure 2.15). These characters were not included in the linear dataset as it is challenging to quantify under a stereomicroscope. The characters also show a gradient of variation between specimens (Figure 2.15). For example, specimens of *U. longifolium* f.C can be found with dense to scattered appressed hairs whereas specimens of *U. longifolium* f.B can be found with dense to scattered, less appressed to erect hairs. It is therefore challenging where the division in hair angle can be drawn, and this suggests that it is not a good character for identification. The difficulty to classify the taxa by machine learning suggests that the taxa were not distinguishable using any other characters than those hair characters stated above. Leaf hair density and angle could be the response to the growing environment, an effect previously reported in four arctic *Festuca* Tourn. ex L. species (Ramesar-Fortner, Dengler and Aiken, 1995). Leaf hair density is also associated with leaf size in *Sinapis arvensis* L. (Roy, Stanton and Eppley, 1999). The hair characters for *U. longifolium* taxa could therefore be due to morphological plasticity rather than species-specific adaptation. The relationships in the *U. longifolium* complex is included in the molecular analyses in Chapter 3.

Figure 2.15 Stipules of *U. longifolium* f.B (A and B) and *U. longifolium* f.C (C and D). Scale bar = 5 mm.

Another example of misclassification was *U. crassum* specimens being classified to *U. longifolium* f.C, in both the linear and combined methods. Although, the misplacement using machine learning shows that the characters and the shape of the secondary vein loop used in the analysis cannot be used to separate the taxa, the inflorescence structure and flower colour are different. The inflorescence of *U. crassum* is umbel-like with green flowers, compared to the compound cyme inflorescence and white flowers found in *U. longifolium* f.C (Figure 2.16). This means the inflorescence of *U. longifolium* f.C mostly has a rachis and secondary peduncles which are not present in *U. crassum*. The data for both the rachis and secondary peduncle have been recorded in the characters' list (Table 2.2: character 13 and 14), however they were omitted prior to the analysis due to the large minimum sample size (>25 individuals). The colour of flower could also not be recorded due to the uncertainty of colour preservation in herbarium specimens. Therefore, flower colour could only be retrieved from new collections during fieldwork (Appendix A). Another diagnostic character between these species (and generally for identifying other taxa in the genus) is stipule shape. However, aside from stipule length measurements it is challenging to describe and quantify stipule shape accurately, particularly using herbarium specimens. Since *Urophyllum* taxa tend to have either a folded or flat stipule, with or without a ridge forming at the mid-area (Figure 2.17), upon pressing it is difficult to measure the width of the stipule accurately. The pressing of specimens also means that landmark selection is limited for stipule characters, despite this being a good diagnostic character generally for *Urophyllum* species identification in the field.

Figure 2.16 Inflorescence and flower of *U. crassum* (A), *U. longifolium* f.C (pistillate) (B) and *U. glabrum* (pistillate) (C). Scale bar = 5 mm.

The prediction of *U. longifolium* f.C also misclassified to *U. glabrum*. The differences among these two species provide similar issues to the example of *U. crassum* and *U. longifolium* f.C above, involving characters of flower colour and stipule characters. *U. glabrum* has green flowers with a flat stipule especially at a mature stage, whereas the *U. longifolium* complex have white flowers with folded stipules (Figure 2.16 and 2.17). The most obvious character for identification of *U. glabrum* is the presence of hairy pocket domatia at every axil of branching between secondary veins to the midrib (Figure 2.18). However, the domatium structure was separated into two characters (23–presence of pocket domatia; & 24–presence of dense hairs) in Table 2.2 due to the domatia can be found as glabrous or hairy, and with or without pocket. This has likely led to confusion in machine learning.

Figure 2.17 Stipules of *Urophyllum* taxa. Sample codes correspond to Table 2.1. Scale bar = 5 mm.

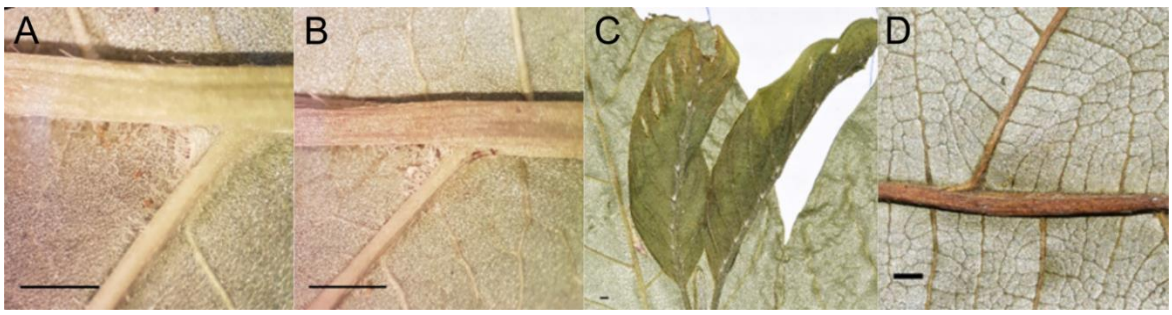


Figure 2.18 The domatia at the angle between the secondary vein and midrib of *U. glabrum* (A, B and C) compared to the angle in *U. longifolium* f.C (D). Scale bar = 1 mm.

2.6. Conclusion

The primary aim of this chapter was to use machine learning approaches to discover if morphological data could be used to assign specimens of *Urophyllum* to categories, categories in this instance being taxa, in doing so this work identifies how accurately taxa are identifiable. In conclusion 10 out of the 13 *Urophyllum* taxa sampled here could be routinely identified using machine learning approaches. This study demonstrates the power of classification is improved by incorporating multiple datasets and the use of ML for identification of *Urophyllum* taxa. Misclassifications, in some instances, occurred due to poor character preservation in herbarium materials, for example flower colour or stipule shape. However, the misclassification of specimens also reveals morphologically similar groups, for example the *U. longifolium* complex and *U. glabrum*. The relationship amongst these species is investigated further in Chapter 3 using molecular data.

The impact of this work is the use and application of a method. By using machine learning upon geometric and linear morphometric datasets, it is possible to classify groups of specimens, in this case *Urophyllum* taxa; however, there are broader applications in the use for identification and classification of other organisms. Moreover, the use of automated identification does not require expertise in a particular plant group for accurate identification. Although experts would play a crucial role in the development of such methods, particularly in identifying diagnostic characters. This would therefore allow non-experts to collect valuable records of species occurrence and variation, particularly of relevance to *Urophyllum* which is under recorded throughout SE Asia.

Chapter 3 Phylogenetic relationships of the genus *Urophyllum* in Thailand and Vietnam based on plastid genome and nuclear ribosomal DNA

3.1. Introduction

The genus *Urophyllum* comprises around 120 species of shade-tolerant dioecious shrubs and treelets (Metcalfe, Grubb and Turner, 1998) distributed throughout sub-tropical and tropical Asia (Smedmark, Eriksson and Bremer, 2010; Govaerts *et al.*, 2020). Approximately 20 species are found throughout mainland Southeast Asia, excluding Peninsular Malaysia (Govaerts *et al.*, 2020) with a limited number of studies of the genus in the area. The most comprehensive molecular study of the genus to date used both plastid (*rps16* intron, *trnT-F*) and nuclear ribosomal (ITS, ETS) DNA regions to elucidate the phylogenetic relationship within tribe Urophylleae, including 20 species of *Urophyllum* (Smedmark and Bremer, 2011). Of these, only four species are distributed in Thailand, Myanmar and Cambodia, and none of the sampled species occur in Laos or Vietnam.

Previous species-level studies of *Urophyllum* from mainland Southeast Asia were based upon morphological classifications and conducted nearly a century ago (Pitard, 1923; Craib, 1931; Ridley, 1932). More recently only regional treatments of *Urophyllum* have been published, that includes three species in Flora of China Vol. 19 (Taylor *et al.*, 2011) and seven species in Flora of Singapore Vol. 13 (Wong *et al.*, 2019). This covers a small proportion of the total number of *Urophyllum* species, and even fewer that extend into mainland Southeast Asia. Confusion, therefore, on the identification and extent of distribution for many *Urophyllum* species in this region still remains.

Genome skimming (shallow sequencing) is becoming increasingly more popular for molecular studies and can be used to recover high-copy number sequences of the genome, e.g., plastomes, mitochondrial genes, and ribosomal DNA repeats (Straub *et al.*, 2012). The benefits of this technology are that high-quality DNA is not a necessity, therefore the often degraded DNA collected from herbarium specimens can be used (Dodsworth, 2015).

Angiosperm plastid genomes typically have a conserved quadripartite circular structure including a large single copy (LSC), a pair of inverted repeats (IRs) and a small single copy (SSC) region (Green, 2011). The majority of plant plastomes are uniparentally inherited (Birky, 1995), most angiosperm plastomes are maternally inherited, whilst gymnosperm plastomes (except *Ginkgo* L., cycads and gnetophytes), are usually paternally inherited. However biparental inheritance of some seed plants including some conifers, and some members of Campanulaceae, Fabaceae, Geraniaeae, Hypericaceae, Lamiaceae, Onagraceae, Plumbaginaceae, Poaceae and Polygonaceae has been revealed (Harris and Ingram, 1991; Reboud and Zeyl, 1994; Wicke *et al.*, 2011).

The plastid structure of angiosperms is generally reported to be conserved (Downie and Jansen, 2015), although variation in gene organisation and losses have been identified in several plant groups. For example, in *Passiflora* L. *ycf1* and *ycf2* gene losses have been reported (Rabah *et al.*, 2019). In parasitic (*Diphelypaea* Nicolson in Orobanchaceae) and carnivorous plants (*Drosera* L. in Droseraceae), genes related to photosynthesis have been lost and plastome rearrangements have been reported (Gruzdev *et al.*, 2019; Nevill *et al.*, 2019). IR extension was also reported in Campanulaceae where seven genes from the SSC moved into the IR region (Li, Wang and Li, 2020) as well as in *Acacia* Mill. and *Inga* Mill. (Fabaceae) where nine genes from SSC moved into IR (Dugas *et al.*, 2015). The presence of two IRs in plastid genome is generally conserved in most angiosperm lineages. It is thought one function of the IR regions is to help stabilise the structure of the plastome (Palmer, 1983). However, despite having two IRs, the plastomes of *Pelargonium* L'Hér. ex Aiton and *Trachelium* Tourn. ex L. are highly rearranged in gene order (Chumley *et al.*, 2006; Haberle *et al.*, 2008). This indicates that plastome stability may not be based only upon the presence of the IRs (Zhang, Zhang and Xiang, 2019). Blazier *et al.* (2016) suggested that the number of repeats correlates with plastome instability, as repeat regions are in positions where nonhomologous recombination occurs. Furthermore, repeat regions are also linked to gene conversion that can lead to smaller scale IR extensions (Goulding *et al.*, 1996).

In Rubiaceae, a study of plastid genomes in subfamily Ixoroideae conducted by Ly *et al.*, (2020) reported *trnH-GUG* loss in *Tarenna grevei* (Drake) Homolle and the possible loss of one IR copy in some species (*Mussaenda pubescens* Dryand., *Feretia aeruginescens* Stapf

and *Pavetta schumanniana* F.Hoffm. ex K.Schum.) using short-read sequencing. In subfamily Rubioideae, the plastomes of nine species are currently available on GenBank (accessed 18 June 2020) including; *Hedyotis ovata* Thunb. ex Maxim. (MK203877), *Paralasianthus hainanensis* (Merr.) H.Zhu (synonym *Saprosma merrillii* H.S.Lo) (MK203879), *Galium mollugo* L. (KY562588), *Galium aparine* L. (KY562587), *Morinda citrifolia* L. (MN699649), *Gynochthodes officinalis* (F.C.How) Razafim. & B.Bremer (synonym *Morinda officinalis* F.C.How) (KR869730), *Gynochthodes nanlingensis* (Y.Z.Ruan) Razafim. & B.Bremer (KT852576), *Rubia cordifolia* L. (MN736957) and *Dunnia sinensis* Tutcher (MN883829). All of the available plastomes of Rubioideae have a canonical structure and gene organisation of angiosperms (Ruhlman *et al.*, 2017).

To date there have been no molecular studies focusing on regional interspecific relationships of *Urophyllum* in mainland Southeast Asia. Previous studies for tribal level classification have therefore limited sampling of species in the genus. The aims of this study are to understand plastid characteristics and structure, as well as provide the most comprehensive phylogeny of *Urophyllum* to date to understand the regional evolutionary relationships. In order to achieve this: 1) plastid genomes and nuclear ribosomal cistrons of 39 *Urophyllum* samples distributed in Thailand and Vietnam were sequenced; 2) a comparative plastome and repeat analyses were conducted; 3) phylogenetic trees from both data sets were constructed; 4) molecular variation is compared with morphological traits.

The results of the morphometric classification (Chapter 2) showed three groups of taxa that did not have 100% accurate identification based on morphology alone using supervised machine learning. These morphologically similar taxa are shown in Table 3.1. Phylogenetic analyses in this chapter provide further evidence for both inter- and intraspecific relationships not only within these taxa but also the genus *Urophyllum* in general.

Table 3.1 Taxa with highly similar morphological characters and <100% accurate identification using supervised machine learning in Chapter 2. Character differences shown here were from personal observations and not included in data analyses as discussed in Chapter 2.

Reference taxa	**Predict taxa**	**Character differences**
U. longifolium f.B	*U. longifolium* f.C	Hair density and pattern on the stipules. - *U. longifolium* f.B has scattered, long erect hairs. - *U. longifolium* f.C has dense, short appressed hairs.
U. longifolium f.C	*U. glabrum*	Stipule style, presence of pocket domatia on leaf, and flower colour. - *U. longifolium* f.C has folded stipule, domatium absent and white flower. - *U. glabrum* has flat stipule, domatium present, and green flower.
U. crassum	*U. longifolium* f.C	Inflorescence type and flower colour. - *U. crassum* has umbel-like inflorescence with green flower. - *U. longifolium* f.C has compound cyme inflorescence with white flower.

3.2. Materials and methods

3.2.1. Plant materials

Thirty-nine leaf samples of 18 *Urophyllum* taxa were collected during field expeditions in Thailand and Vietnam between 2017–2019 (Appendix A), and from the herbarium collections of FU, K and KKU (herbarium acronyms follow Index Herbariorum (Thiers, 2020). Figure 3.1 and Table 3.2 show the location and collection details of the sampled populations. As the genus comprises dioecious plants, where possible both pistillate and staminate samples were collected. Samples that could not be identified to species level, potentially new taxa, were coded by numbers (1 to 4).

Due to the difficulty in the identification of *U. longifolium* var. *longifolium*, *U. longifolium* var. *pilosum* and *U. talangense*, in this study, samples identified as *U. longifolium* or *U. talangense* were designated using the following taxa code: 1) *U. longifolium* var. *longifolium* = *U. longifolium* f.A; 2) *U. longifolium* var. *pilosum* = *U. longifolium* f.B; and 3) *U. talangense* = *U. longifolium* f.C (Table 3.2). The difficulty with identifications for these taxa is discussed further in Chapter 4.

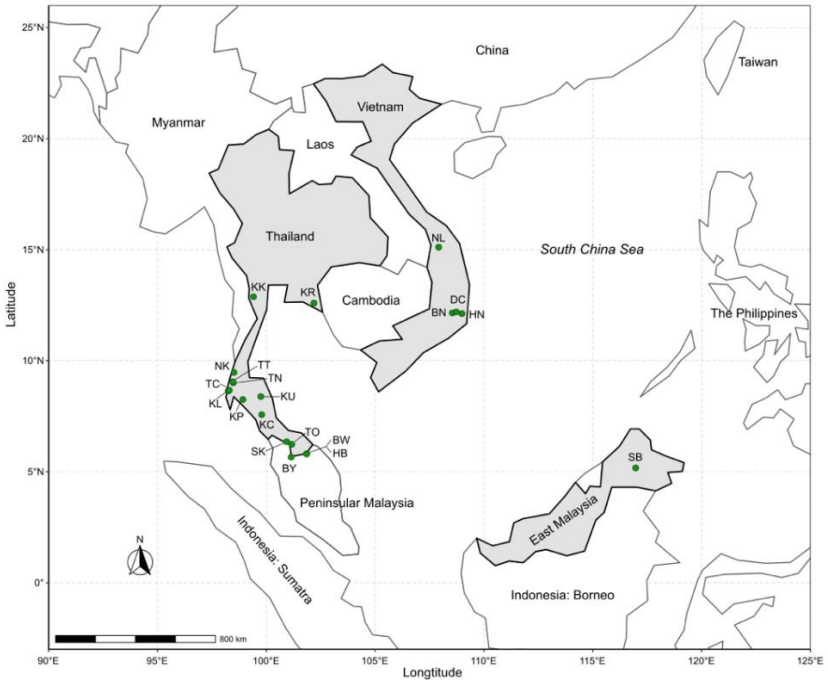


Figure 3.1 Map showing the localities of the sampled *Urophyllum* species. Locality codes correspond to those in Table 3.2.

Table 3.2 *Urophyllum* taxa, locations where all taxa used in this study were collected, and corresponding voucher specimens. Taxa codes show species followed by (forma 'f' if applicable), locality code, and sex (if known), respectively. Locality codes are indicated with two capital letters. F and M represent pistillate and staminate plants, respectively. Herbarium acronyms in column Voucher follow Index Herbariorum (Thiers, 2020).

Species (Gender)	**locality (code)**	**Taxon code**	**Voucher**
U. argenteum (♀ & ♂)	Hon Ba Nature Reserve, Khanh Hoa, Vietnam (HN)	*U. argenteum*_HN_F	S. Yooprasert *et al.* VN73-1 (BKF, HN)
		*U. argenteum*_HN_M	S. Yooprasert *et al.* VN74-1 (BKF, HN)
U. blumeanum	Ban Bawae, Waeng district, Narathiwat, Thailand (BW)	*U. blumeanum*_BW	S. Yooprasert *et al.* 199 (BKF, K)
U. chinense	Hon Ba Nature Reserve, Khanh Hoa, Vietnam (HN)	*U. chinense*_HN	S. Yooprasert *et al.* VN67-1 (BKF, HN)
U. crassum (♀ & ♂)	Krung Ching Falls, Khao Luang National Park, Nakhon Si Thammarat, Thailand (KU)	*U. crassum*_KU_F	S. Yooprasert *et al.* 140 (BKF, K)
		*U. crassum*_KU_M	S. Yooprasert *et al.* 146 (BKF, K)
	Than To Falls Forest Park, Yala, Thailand (TO)	*U. crassum*_TO	M. Poopath MP1999 (BKF)
U. glabrum (♀ & ♂)	Ton Chong Fa Falls, Khao Lak-Lamru National Park, Phangnga, Thailand (TC)	*U. glabrum*_TC_F	S. Yooprasert *et al.* 120 (BKF, K)
		*U. glabrum*_TC_M	S. Yooprasert *et al.* 171 (BKF, K)
	Sankala Khiri National Park, Songkhla, Thailand (SK)	*U. glabrum*_SK	S. Yooprasert *et al.* 181 (BKF, K)

Table 3.2 (continued) *Urophyllum* and outgroup taxa, locations where all taxa used in this study were collected, and corresponding voucher specimens. Taxa codes show species followed by (forma 'f' if applicable), locality code, and sex (if known), respectively. Locality codes are indicated with two capital letters. F and M represent pistillate and staminate plants, respectively. Herbarium acronyms in column Voucher follow Index Herbariorum (Thiers, 2020).

Species (Gender)	**locality (code)**	**Taxon code**	**Voucher**
U. glabrum (♀ & ♂)	Southern Botany and Forestry Development and Study Center, Trang, Thailand (KC)	*U. glabrum*_KC_F	S. Yooprasert *et al.* 92 (BKF, K)
		*U. glabrum*_KC_M	S. Yooprasert *et al.* 84 (BKF, K)
U. hirsutum	Ban Bawae, Waeng district, Narathiwat, Thailand (BW)	*U. hirsutum*_BW	S. Yooprasert *et al.* 198 (BKF, K)
	Hala-Bala Wildlife Sanctuary, Narathiwat, Thailand (HB)	*U. hirsutum*_HB	S. Yooprasert *et al.* 185 (BKF, K)
U. lecomtei	Da Chais Commune, Lac Duong district, Lam Dong, Vietnam (DC)	*U. lecomtei*_DC	S. Yooprasert *et al.* VN100-4 (BKF, HN)
	Ngoc Linh Nature Reserve, Dak Glei district, Kon Tum, Vietnam (NL)	*U. lecomtei*_NL	S. Tagane *et al.* V6085 (FU)
U. longifolium var. *annamense* Pit.	Hon Ba Nature Reserve, Khanh Hoa, Vietnam (HN)	*U. longifolium* var. *annamense*_HN	S. Yooprasert *et al.* VN73-3 (BKF, HN)
U. longifolium	Kaeng Krachan National Park, Phetchaburi, Thailand (KK)	*U. longifolium*_f.A_KK	S. Yooprasert *et al.* 160 (BKF, K)
U. longifolium var. *pilosum*	Southern Botany and Forestry Development and Study Center, Trang, Thailand (KC)	*U. longifolium*_f.B_KC	S. Yooprasert *et al.* 90 (BKF, K)
	Krung Ching Falls, Khao Luang National Park, Nakhon Si Thammarat, Thailand (KU)	*U. longifolium*_f.B_KU	S. Yooprasert *et al.* 144 (BKF, K)

Table 3.2 (continued) *Urophyllum* and outgroup taxa, locations where all taxa used in this study were collected, and corresponding voucher specimens. Taxa codes show species followed by (forma 'f' if applicable), locality code, and sex (if known), respectively. Locality codes are indicated with two capital letters. F and M represent pistillate and staminate plants, respectively. Herbarium acronyms in column Voucher follow Index Herbariorum (Thiers, 2020).

Species (Gender)	**locality (code)**	**Taxon code**	**Voucher**
U. longifolium var. *pilosum*	Khlong Nakha Wildlife Sanctuary, Ranong, Thailand (NK)	*U. longifolium*_f.B_NK	S. Yooprasert *et al.* 134 (BKF, K)
U. talangense	Trail by the visitor centre of Khao Lak-Lamru National Park, Phangnga, Thailand (KL)	*U. longifolium*_f.C_KL	S. Yooprasert *et al.* 105 (BKF, K)
	Khao Panom Bencha National Park, Krabi, Thailand (KP)	*U. longifolium*_f.C_KP	S. Yooprasert *et al.* 76 (BKF, K)
	Ton Chong Fa Falls, Khao Lak-Lamru National Park, Phangnga, Thailand (TC)	*U. longifolium*_f.C_TC	S. Yooprasert *et al.* 121 (BKF, K)
	Tam Nang Falls, Si PhangNga National Park, Phangnga, Thailand (TN)	*U. longifolium*_f.C_TN	S. Yooprasert *et al.* 123 (BKF, K)
	Ton Toei Falls, Si PhangNga National Park, Phangnga, Thailand (TT)	*U. longifolium*_f.C_TT	S. Yooprasert *et al.* 126 (BKF, K)
U. longipes	Betong District, Yala, Thailand (BY)	*U. longipes*_BY	J. Wai 2654 (BKF)
U. macrophyllum (Blume) Korth.	Hala-Bala Wildlife Sanctuary, Narathiwat, Thailand (HB)	*U. macrophyllum*_HB	S. Yooprasert *et al.* 188 (BKF, K)
U. memecyloides (C.Presl) S.Vidal	Imbak Canyon (Imbak Canyon Conservation Area), Sabah, Malaysia (SB)	*U. memecyloides*_SB	J. Gregson 27 (K)
U. schmidtii	Khlong Narai Falls, Chanthaburi, Thailand (KR)	*U. schmidtii*_KR	T. Srisuk Sri953 (KKU)
U. streptopodium	Betong District, Yala, Thailand (BY)	*U. streptopodium*_BY	S. Yooprasert *et al.* 212 (BKF, K)

Table 3.2 (continued) *Urophyllum* and outgroup taxa, locations where all taxa used in this study were collected, and corresponding voucher specimens. Taxa codes show species followed by (forma 'f' if applicable), locality code, and sex (if known), respectively. Locality codes are indicated with two capital letters. F and M represent pistillate and staminate plants, respectively. Herbarium acronyms in column Voucher follow Index Herbariorum (Thiers, 2020).

Species (Gender)	**locality (code)**	**Taxon code**	**Voucher**
U. villosum (♀ & ♂)	Than To Falls Forest Park, Yala, Thailand (TO)	*U. villosum*_TO	M. Poopath MP2016 (BKF)
	Betong District, Yala, Thailand (BY)	*U. villosum*_BY_F	J. Wai 2656 (BKF)
		*U. villosum*_BY_M	J. Wai 2663 (BKF)
U. sp.1 (*U. chinense* subsp. *latistipulum* sp. nov.)	Da Chais Commune, Lac Duong District, Lam Dong, Vietnam (DC)	*U.* sp.1_DC	S. Yooprasert *et al.* VN96-1 (BKF, HN)
U. sp.2 (*U. bidoupense* sp. nov.) (♀ & ♂)	Da Chais Commune, Lac Duong District, Lam Dong, Vietnam (DC)	*U.* sp.2_DC_F	S. Yooprasert *et al.* VN85-3 (BKF, HN)
		U. sp.2_DC_M	S. Yooprasert *et al.* VN87-3 (BKF, HN)
U. sp.3 (*U. pseudoschmidtii* sp. nov.)	Bidoup-NuiBa National Park. Lam Dong, Vietnam (BN)	*U.* sp.3_BN	S. Yooprasert *et al.* VN112-1b (BKF, HN)
U. sp.4 (*U. brochidodromum* sp. nov.)	Ngoc Linh Nature Reserve, Dak Glei District, Kon Tum, Vietnam (NL)	*U.* sp.4_NL	S. Tagane *et al.* V6648 (FU)

3.2.2. DNA extraction, genome sequencing and plastome assembly

Dried leaf material (20–25 mg) was ground using a TissueLyser II (QIAGEN, Manchester, UK) for two cycles of 35 Hz for one minute. Total genomic DNA was then extracted using a CTAB protocol (Doyle and Doyle, 1987) modified using Tel-Zur *et al.* (1999) and Krapp (2013) with the addition of 3 rinses in sorbitol buffer (100 mM Tris-HCl pH 8, 5 mM EDTA pH 8, 0.35M Sorbitol) on ice before incubation in CTAB buffer, and then adding 3M potassium acetate pH 5.5 prior to wash with chloroform and isoamyl alcohol solution (see full protocol in Appendix B).

DNA quality and quantity were checked using NanoDrop Lite™ (ThermoFisher, Paisley, UK), 0.7% agarose gel electrophoresis with Hyperladder™ 1kb (bioline, London, UK) as a reference, and Qubit™ Fluorometer (ThermoFisher, Paisley, UK). DNA aliquots of 1 µg were submitted for library preparation and 150 bp PE sequencing was completed by Novogene Co. Ltd. (Beijing, China).

Plastomes were assembled using two different assembly softwares; Fast-Plast v1.2.6 (McKain and Wilson, 2017) and NOVOplasty v3.7.0 (Dierckxsens, Mardulyn and Smits, 2016). For the Fast-Plast assemblies, the Bowtie reference index was set to Gentianales. For the NOVOplasty assemblies, a *rbcL* sequence of *U. glabrum* (GenBank accession: KJ594924) was used as the initial seed. Where NOVOplasty assembled more than two plastome options, further assemblies were performed using alternative seed sequences of *U. longifolium* (*rps16*-AM900616 and *trnL–F*-HM042602). In the cases where both pipelines produced unfinished plastomes, assemblies were performed in GetOrganelle v1.6.4 (Jin *et al.*, 2018) using the default settings. The up to three assembled draft plastomes for each taxon were then aligned using MAFFT v7.450 (Katoh and Standley, 2013) in Geneious Prime 2020.1.1 (https://www.geneious.com, Kearse *et al.*, 2012). Where assemblies did not produce consistent results, the draft plastomes were corrected manually by mapping raw paired-end reads to the conflict regions in Geneious Prime. IR regions were detected using the 'Repeat Finder v1.0.1' plugin in Geneious Prime.

The junctions between LSC-IRb, IRb-SSC, SSC-IRa, and IRa-LSC were confirmed using two different methods; PCR amplification and sequencing; and mapping raw paired-end reads to the regions. First, PCR amplification was conducted using published junction primers

(Costion *et al.*, 2011; Choi, Son and Park, 2015) and primers designed using draft assemblies in this study (see Table S1a for details). PCR reactions were performed in 20 µl volumes containing a final concentration of 1x BioMix (Bioline, London, UK), 0.75µM each of forward and reverse primers, and 10 ng/µl of DNA template. Cycling conditions for each of the primer combinations are provided in Table S1b. PCR products were visualised using 1% agarose gel electrophoresis stained with GelRed® (Biotum; Freemont, USA) and sized using Hyperladder™ 1kb as a reference. Amplicons were compared with the expected size based on the assembled draft plastomes. The PCR amplicons were purified and sequenced in both directions using dideoxynucleotide sequencing by Eurofins Genomics (Ebersberg, Germany). Consensus sequences were assembled in SeqMan Pro v13.0.2.422 (DNAStar, Madison, US) and then mapped to the draft plastomes in order to confirm the accuracy of the junction endpoints. The second method to check junction regions in the draft plastomes was to map raw paired-end reads to the coding regions nearest to the junction for both IR and single copy (SC) regions. The consensus threshold was then set to 75% in the resulting consensus contig produced from mapping the raw-reads. The IR endpoints were detected from the consensus contig where the identity was lower than 50%. The consensus contigs produced by mapping the raw-reads to both the LSC-IR and IR-SSC junctions were then extracted and mapped back to the draft plastomes to confirm accuracy. This method was also used to confirm other areas of conflict identified between different assembly methods. Then, the conflict free areas within LSC and SSC, and the corrected IR was extracted from the draft plastomes; the IR extraction was copied and converted to a reverse complement IR. These extracted counterparts from the draft plastomes together with the assembled consensus sequences of the PCR sequences, and the contigs built from mapping reads were used to assemble complete plastomes by performing the function 'De Novo Assemble…' in Geneious Prime. Finally, coverage analysis of the finished plastome assemblies was completed using Fast-Plast.

3.2.3. Plastome annotation

The *Urophyllum villosum* genome (TO) was initially annotated using both Geneious Prime and GeSeq (Tillich *et al.*, 2017). In Geneious Prime, annotations were transferred at 70% similarity to the *U. villosum*_TO from *Gynochthodes officinalis* (KR869730) and *Coffea arabica* L. (EF044213). Annotations were corrected by comparing with the output from

GeSeq and identifying the start and stop codons of protein coding genes. tRNA genes were confirmed using tRNAscan-SE v2.0.5 (Chan *et al.*, 2019) and ARAGORN v1.2.38 (Laslett and Canback, 2004). This annotated *U. villosum* plastome was then used as a reference, chosen as it is the type species of the genus. The annotations of *U. villosum*_TO were transferred to the other *Urophyllum* taxa, manually correcting the annotation when needed. A complete plastome map of *U. villosum*_TO was created in ORGDRAW v1.3.1 (Greiner, Lehwark and Bock, 2019). A schematic to visualise junction borders was also plotted in IRscope (Amiryousefi, Hyvönen and Poczai, 2018) and edited for clarity using INKSCAPE v1.0 (https://inkscape.org/).

3.2.4. Plastome comparison and repeat sequences analyses

One IR copy was excluded from all *Urophyllum* plastome sequences in the analyses. Prior to further analysis, the SSC region of all plastomes were reverse complimented, to unite the *ycf1* gene that is partially located in both SSC and IR regions. The plastome sequences were aligned in MAFFT v7.450 with default settings in Geneious Prime. To calculate the nucleotide variability (Pi) and number of variable sites, the alignment was used to conduct sliding window analysis in DnaSP v6.12 (Rozas *et al.*, 2017) with 200 bp step size and 600 bp window length. MEGA v10.1.7 (Kumar *et al.*, 2018) was used to analyse pairwise-distance. The plastome variation were compared and visualised using mVISTA (Frazer *et al.*, 2004) in Shuffle-LAGAN mode (Brudno *et al.*, 2003), using the annotated *U. villosum*_TO as a reference. Simple Sequence Repeats (SSRs) were identified using MISA v2.1 (Beier *et al.*, 2017) with the following number of repetitive units: ten units for mononucleotide, five units for dinucleotide, four units for trinucleotide and three units for tetra-, penta- and hexanucleotide SSRs. REputer (Kurtz *et al.*, 2001) was used to identify forward, reverse, compliment, and palindromic repeats with minimum repeat length set at 30 bp, sequence identity ≥90% and Hamming distance of three.

3.2.5. Nuclear ribosomal DNA cistron assembly and annotation

Nuclear ribosomal DNA (nrDNA) was extracted from the raw paired-end reads using GetOrganelle v1.6.4 with default settings for nrDNA assembly using a starting seed of *U. chinense* ITS2 sequence (KR532702), and *U. longifolium* ETS sequence (HM042526) downloaded from GenBank. Cistrons were annotated in Geneious Prime by transferring

annotation from a published nrDNA cistrons of *Amphidasya ambigua* (Standl.) Standl. (MK607892), *Coffea arabica* (RHJU01000188; NW_020849278), and *Asclepias syriaca* L. (JF312046). Only annotations matched with similarity higher than 80% were kept. The final nuclear ribosomal sequences used in the analyses consisted of the complete ribosomal cistron; small subunit rRNA (18S), ITS1, 5.8S rRNA, ITS2, and large subunit rRNA (26S).

3.2.6. Phylogenetic analyses

3.2.6.1. Plastome data

Plastomes assembled in this study were combined with four outgroup species downloaded from GenBank: *Amphidasya ambigua* (KY378703) (tribe Urophylleae), *Colletoecema dewevrei* (De Wild.) E.M.A.Petit (KY378707) (tribe Colletoecemateae), *Lasianthus* sp. (KY378708) (tribe Lasiantheae) and *Ophiorrhiza mungos* L. (KY378702) (tribe Ophiorrhizeae).

Phylogenetic analyses were performed on whole plastome sequences containing only one IR. Plastome sequences were aligned using MAFFT. Best substitution models were tested using jModelTest2 v2.1.6 (Darriba *et al.*, 2012) based on Akaike information criterion (AIC). The best-fit model was GTR+G. Subsequently phylogenetic trees were constructed using both maximum likelihood (ML) and Bayesian inference (BI) methods. ML estimation of phylogeny analyses were performed using RAxML v8.2.11 (Stamatakis, 2014) in Geneious Prime with 1,000 bootstrap replicates. BI analyses were performed using MrBayes v3.2.7a (Ronquist *et al.*, 2012) on XSEDE in CIPRES Portal v3.3, with Markov Chain Monte Carlo (MCMC) algorithm run for 1,500,000 generations, sampling trees every 1,000 generations and defaults runs and chains (two and four, respectively). The amount of generations exhibited efficient convergence diagnostics as described in Ronquist *et al.* (2020) (see Table S4). Burn-in was adjusted by examining the results in Tracer v1.7.1 (Rambaut *et al.*, 2018), and set as the first 25% of all trees. A consensus tree was constructed from the 75% remaining trees. The trees were visualised in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/, Rambaut, 2018) and *Lasianthus* sp. was chosen to root the phylogeny, this selection was based upon the results of a Rubiaceae phylogeny in Wikström *et al.* (2020), branches were coloured for clarity in INKSCAPE v1.0.

3.2.6.2. nrDNA data

The nuclear ribosomal DNA data was aligned using MAFFT plugin on Geneious Prime. Best-fit nucleotide substitution model was identified as GTR+I+G using jModetTest2 v2.1.6 on XSEDE in CIPRES Portal v3.3. For easy comparison, the same outgroup species as in phylogenetic analysis of plastome data were selected: *Amphidasya ambigua* (MK607892), *Colletoecema dewevrei* (MK607899), *Lasianthus* sp. (MK607919) and *Ophiorrhiza mungos* (MK607926); and *Lasianthus* sp. was chosen to root the trees. Phylogenetic trees were constructed using both maximum likelihood (ML) and Bayesian inference (BI). ML parameters were the same as the plastid datasets. Whereas BI were run for 2,000,000 generations of MCMC algorithm, burn-in was adjusted using Tracer and set as the first 25% of all trees. All remaining settings were the same as in plastid dataset.

3.2.7. Character states reconstruction

Eleven morphological characters and states used in this study are described in Table 3.3. The morphological matrix is shown in Table S2 (detailed characters are shown in Chapter 4).

The character states for *Urophyllum* taxa were obtained from voucher specimens summarised in Table 3.2 as well as herbarium specimens from AAU, ABD, BKF, BK, BM, C, E, FU, K, SING and QBG. For outgroup species, the characters were recorded from digitised type specimens on JSTOR Global Plant websites (https://plants.jstor.org/); except for *Lasianthus* sp. as the voucher specimen was not available online (collector: Kainulainen *et al.* 17 (S)). Leaf vein terminology follows Hickey (1979) and the Leaf Architecture Working Group (1999), other terminologies follow Beentje (2016). Taxa with polymorphic states were recorded in brackets (Table S2).

Ancestral state reconstructions were performed in Mesquite v3.61 (Maddison and Maddison, 2019) using function 'trace character over trees' under maximum parsimony method. Reconstructions were completed upon raw trees obtained from previous Bayesian inference analyses which were 2,250 and 3,000 trees in plastid and nrDNA datasets, respectively (excluding 25% of the burn-in trees); then the results were plotted on 50% majority rule consensus trees for each dataset.

Table 3.3 Characters and character states.

Characters	Character states
(1) Stipule division	0 entire; 1 divided
(2) Stipule fold morphology	0 appressed to the stem but not folded; 1 folded toward adaxial side
(3) Domatium morphology at abaxially secondary vein axil	0 absent; 1 pocket domatia, hairy and always present; 2 pocket domatia, hairy and sometimes present; 3 pocket domatia, glabrous and sometimes present; 4 hairs domatia, no pocket
(4) Secondary venation and closing loop	0 conspicuously brochidodromous; 1 weakly brochidodromous; 2 festooned brochidodromous; 3 eucamptodromous
(5) Petiole hair distribution	0 glabrous; 1 densely to sparsely hairy all over, similar density; 2 denser hairs at canalicular ridge, sparse/subglabrous other areas
(6) Inflorescence type	0 single flower; 1 simple cymose; 2 compound cymose; 3 helicoid cymose ;4 pedunculate umbellate; 5 two-tiers umbellate; 6 sessile umbellate
(7) Calyx lobe morphology	0 entire; 1 toothed; 2 lobed
(8) Corolla colour	0 white; 1 green; 2 white/pale yellow then green toward apex
(9) Abaxial corolla hair distribution	0 glabrous/scaly around apex; 1 hairy all over; 2 scaly all over
(10) Adaxial corolla throat, hair density	0 numerous densely hairy; 1 sparsely hairy, countable
(11) Adaxial corolla throat, membrane at lobed base	0 absent; 1 triangular, attached to each lobe; 2 tubular, connected to all lobes

3.3. Results

3.3.1. Plastome assembly and characteristics

Plastid genomes of 39 *Urophyllum* samples had a typical quadripartite structure consisting of a LSC, two IRs separated by an SSC (Figure 3.2). The plastid contents are summarised in Table 3.4 and Table 3.5 (content details by taxon can be found in Table S3 and S5).

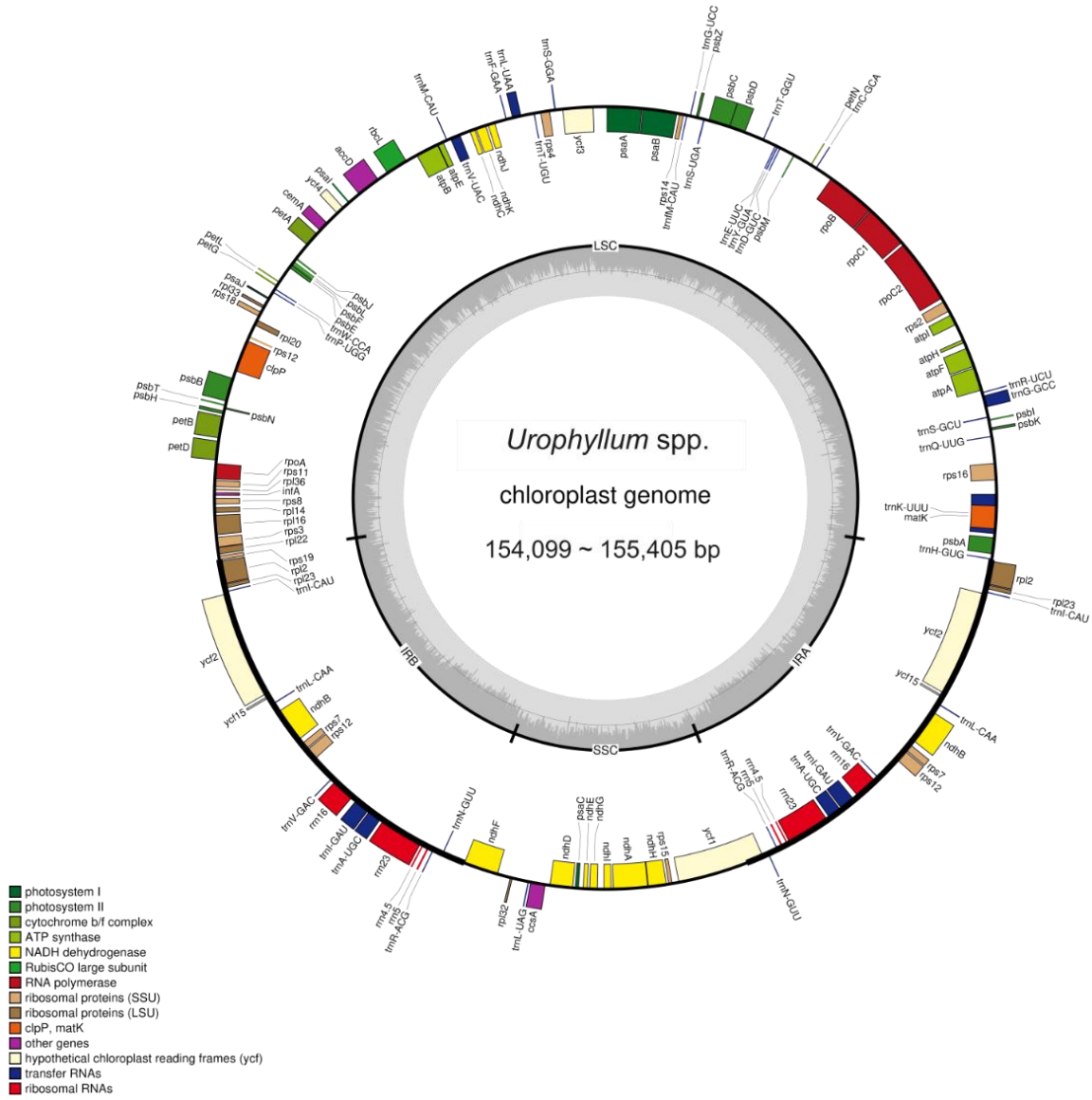


Figure 3.2 Gene map of *Urophyllum* plastid genomes. Genes located outside the circle are transcribed clockwise, those inside the circle are transcribed anticlockwise. Colour bars indicate groups of functional genes. Inner circle represents nucleotide content—dark grey corresponds to GC content and light grey corresponds to AT content.

Table 3.4 Summary of plastid genome content of 39 *Urophyllum* samples.

Plastid contents	**Number and unit of each content**
Paired-end reads (for the assemblies)	28,293,622–69,959,554 reads
Average coverage at 25 kmer	72–1,362×
Whole plastome sequences length - LSC - SSC - IR	154,099–155,405 bp 85,535–84,547 bp 18,177–18,342 bp 25,640–25,795 bp
GC contents	37.7% or 37.8%
Unique genes - Protein coding genes - tRNA genes - rRNA genes	113 genes 80 genes 29 genes 4 genes

Table 3.5 Gene content in 39 *Urophyllum* plastid genomes.

Category for gene	Group of genes	Name of genes
Self-replication	Large subunit of ribosome	*rpl2*[ab], *rpl14, rpl16, rpl20, rpl22, rpl23*[ab], *rpl32, rpl33, rpl36*
	Small subunit of ribosome	*rps2, rps3, rps4, rps7*[a], *rps8, rps11, rps12*[ab], *rps14, rps15, rps16*[b], *rps18, rps19*
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*[b], *rpoC2*
	rRNA gene	*rrn16, rrn23, rrn4.5, rrn5*
	tRNA gene	*trnA*-UGC[ab], *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnG*-GCC[b], *trnG*-UCC, *trnH*-GUG, *trnI*-CAU[a], *trnI*-GAU[ab], *trnK*-UUU[b], *trnL*-CAA[a], *trnL*-UAA[b], *trnL*-UAG, *trnM*-CAU, *trnN*-GUU[a], *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG[a], *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnT*-UGU, *trnV*-GAC[a], *trnV*-UAC[b], *trnW*-CCA, *trnY*-GUA
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	NADH-dehydrogenase	*ndhA*[b], *ndhB*[ab], *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Cytochrome b/f complex	*petA, petB*[b], *petD*[b], *petG, petL, petN*
	ATP synthase	*atpA, atpB, atpE, atpF*[b], *atpH, atpI*
	Rubisco	*rbcL*
Other genes	Translational initiation factor	*infA*
	Maturase	*matK*
	Protease	*clpP*[b]
	Envelope membrane protein	*cemA*
	Subunit of Acetyl-CoA-carboxylase	*accD*
	C-type cytochrome synthesis	*ccsA*
	Hypothetical chloroplast open reading frame	*ycf1, ycf2*[a], *ycf3*[b], *ycf4, ycf15*[a]

Notes. [a] = genes located in IR regions; [b] = intron containing genes

The variation at the junctions of the assembled plastomes can be classified into three patterns as shown in Figure 3.3. The variation relates to four genes; *rps19*, *ndhF*, *ycf1* and *trnH*-GUG. Pattern A was the most common found in *Urophyllum* plastomes (31 samples, 15 species) in which the LSC/IRb border was inside *rps19* gene; IRb/SSC border ranged from 13–34 bp downstream of *ndhF*; the SSC/IRa border was within *ycf1* gene; and IRa/LSC was inside *trnH*-GUG gene (Figure 3.3A). Patterns B and C were found in two taxa (four samples). Pattern B included all three *U. villosum* samples within this study and one unidentified species (*U.* sp.4). The boundaries of pattern B were similar to pattern A except for the IRb/SSC border that was inside *ndhF*. Pattern C was different by the LSC/IRb border was an intergenic spacer between *rps19* and *rpl2* genes. The taxa found with pattern C junction include two samples of *U. glabrum* collected from Ton Chong Fa Falls, PhangNga (TC) and two samples of one unidentified species (*U.* sp.2). This variation shows that the LSC/IRb endpoint of *U. glabrum* taxa shows two patterns either A or C in Figure 3.3 and Table 3.6. IR endpoints in other *Urophyllum* species show a consistent pattern within a species.

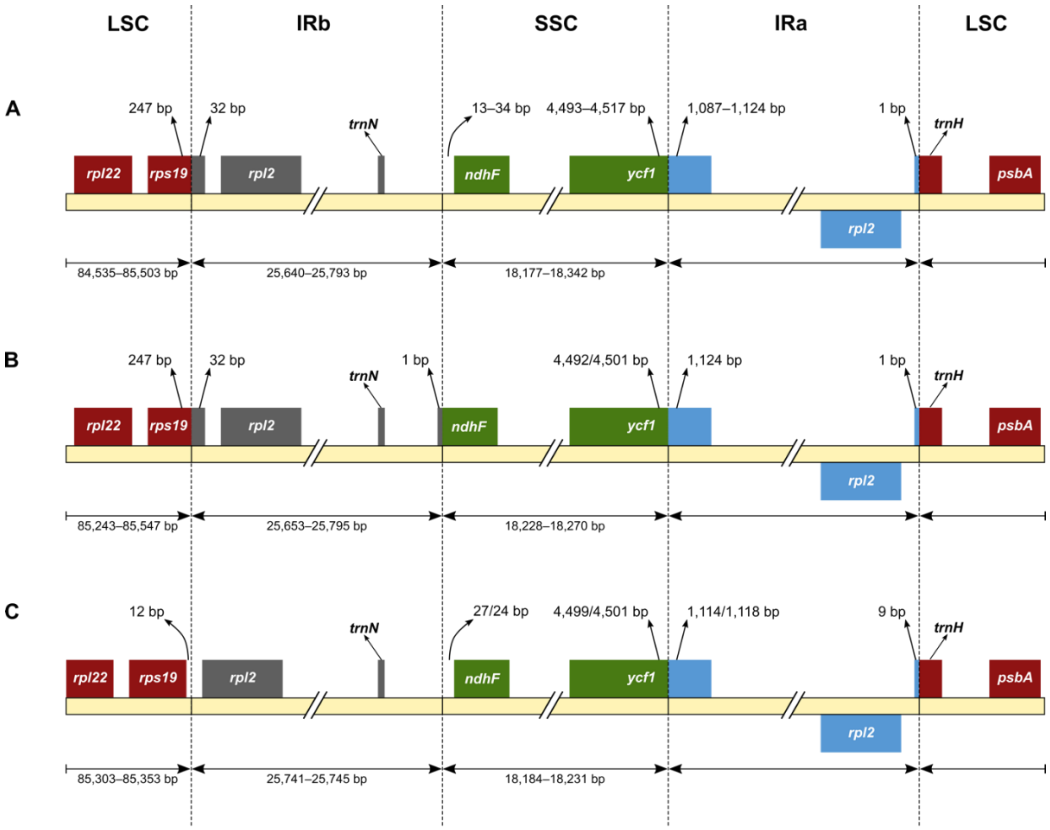


Figure 3.3 Comparison of LSC, SSC, and IR border regions among *Urophyllum* plastid genomes. Coloured boxes represent the genes in each region. Lists of *Urophyllum* taxa found with each boundary pattern (A, B and C) represent in Table 3.6.

Table 3.6 Lists of *Urophyllum* samples within each boundary pattern shown in Figure 3.3.

Boundary pattern	***Urophyllum* taxa**			**Total number of taxa**
A	*U. argenteum*_HN_F, *U. argenteum*_HN_M, *U. blumeanum*_BY, *U. chinense*_HN, *U. crassum*_KU_F, *U. crassum*_KU_M, *U. crassum*_TO, *U. glabrum*_TC_F, *U. glabrum*_TC_M, *U. glabrum*_SK, *U. hirsutum*_BW,	*U. hirsutum*_HB, *U. lecomtei*_DC, *U. lecomtei*_NL, *U. longifolium* var. *annamense*_HN, *U. longifolium*_f.A_KK, *U. longifolium*_f.B_KC, *U. longifolium*_f.B_KU, *U. longifolium*_f.B_NK, *U. longifolium*_f.C_KL, *U. longifolium*_f.C_KP,	*U. longifolium*_f.C_TC, *U. longifolium*_f.C_TN, *U. longifolium*_f.C_TT, *U. longipes*_BY, *U. macrophyllum*_HB, *U. memecyloides*_SB, *U. schmidtii*_KR, *U.* sp.1_DC, *U.* sp.3_BN, *U. streptopodium*_BY	31
B	*U.* sp.4_NL, *U. villosum*_TO, *U. villosum*_BY_F, *U. villosum*_BY_M			4
C	*U. glabrum*_KC_M, *U. glabrum*_KC_F, *U.* sp.2_DC_F, *U.* sp.2_DC_M			4

3.3.2. Repeat sequences analyses in *Urophyllum* plastid genomes

According to repeat analyses in MISA, the number of SSRs identified in 39 *Urophyllum* samples ranged from 32–44 repeats. Four types of SSR were detected in all taxa with the most abundant being mononucleotide repeats (17–26) followed by tetra- (5–9), tri- (4–7) and dinucleotide (1–6). In contrast, there was only one penta- and one hexanucleotide repeat in nine and five species, respectively (Figure S2A and Table S6). The majority of mononucleotide repeats for most species (except *U. macrophyllum*) were A and T motifs which accounted for 12.1–29.7% and 24.3–36.8% of all repeats, respectively, a C motif on the other hand, was rare (0–6.1%) and no G motif repeats were detected (Table S7). In *U. macrophyllum*, a T motif was the most frequently occurring type of repeat (30.8%), and A motif was less common than a C motif (7.7% and 10.3%, respectively). The repeat number of five types of plastid SSRs (A, T, AT, TTA and TAAA) found in all 39 *Urophyllum* samples varied in number both among and within species (Table S9). A further analysis of microsatellite locations revealed that highest amount of SSRs were found in LSC ranging from 25–35 SSRs (71.8–83.3%), followed by SSC which ranged from 4–10 SSRs (11.1–25.6%) and the IR region ranged from 0–2 SSRs (0–5.6%), respectively (Figure S2B and Table S8).

Repeat sequences, longer than 30 bp, were analysed using REputer. The total number of repeats in all 39 *Urophyllum* samples ranged from 21–29 repeats with the most common repeats between 30–39 bp in length (Figure S3A). The plastomes of *Urophyllum* contained 12–16 palindromic, 6–11 forward and 2–5 reverse repeats with no compliment repeat (Figure S3B). These repeats were mainly distributed in non-protein-coding regions (non-CDS) (Figure S4). Repeats found in CDS regions were in seven genes: *psaB*, *psaA*, *accD*, *rps18*, *ycf2*, *ycf1* and *ndhF* with the maximum repeat length at 60 bp (*ycf1*). All repeat locations can be found in Table S10 and S11. Repeat sequence locations varied among *Urophyllum* taxa. The regions found in all taxa include IGS regions: *petN–psbM*, *petN–psbM*, *rrn4.5-rrn5*, *psaC-ndhD* and *ndhD-ccsA*. Other regions that are unique to a particular species are IGS regions: *trnN–ndhF* in *U. crassum* plastomes, *atpI-rps2* in *U. argenteum* and *rpoB-trnC*(-*GCA*) in *U. villosum* (Table S10 and S11).

3.3.3. Identification of variable regions in *Urophyllum* plastid genomes

Divergent regions among 39 *Urophyllum* plastomes are shown in the mVISTA plot with *U. villosum*_TO as a reference annotation (Figure 3.4). Generally, the alignment revealed a high similarity between plastid sequences with some divergent regions particularly in intergenic spacers and introns (e.g., *rps16–trnQ*, *rpoB–trnC*, *psbM–trnD*, *ndhC–trnV*, *psbE–petL*, *ndhF–rpl32–trnL* and intron of *petD*). The most variable protein-coding gene was *ycf1*. Further analysis of nucleotide diversity (Pi) displayed similar trends where the diversity level was high in intergenic spacers and the intron of some genes including *petD* and *rpl16* (Figure 3.5). Additionally, all highly divergent sites (nucleotide diversity ≥0.01) were found in LSC and SSC regions. The IR region, on the contrary, showed very low divergence levels. The highest nucleotide diversity was detected between *ndhF–rpl32–trnL* genes (Pi=0.01886) in SSC, whereas the lowest (Pi=0) was found within ribosomal RNA genes in the IR (figure 3.5). The SSC was the most variable region with 3.85% (693 sites) of variable sites and average divergence levels equal to 0.00728 (Figure 3.5 and Table 3.4). LSC was the second most variable with 2.77% (2,279 sites) of variable sites and 0.0051 average nucleotide diversity. The lowest variability was the IR region with 0.54% and an average Pi equal to 0.0008.

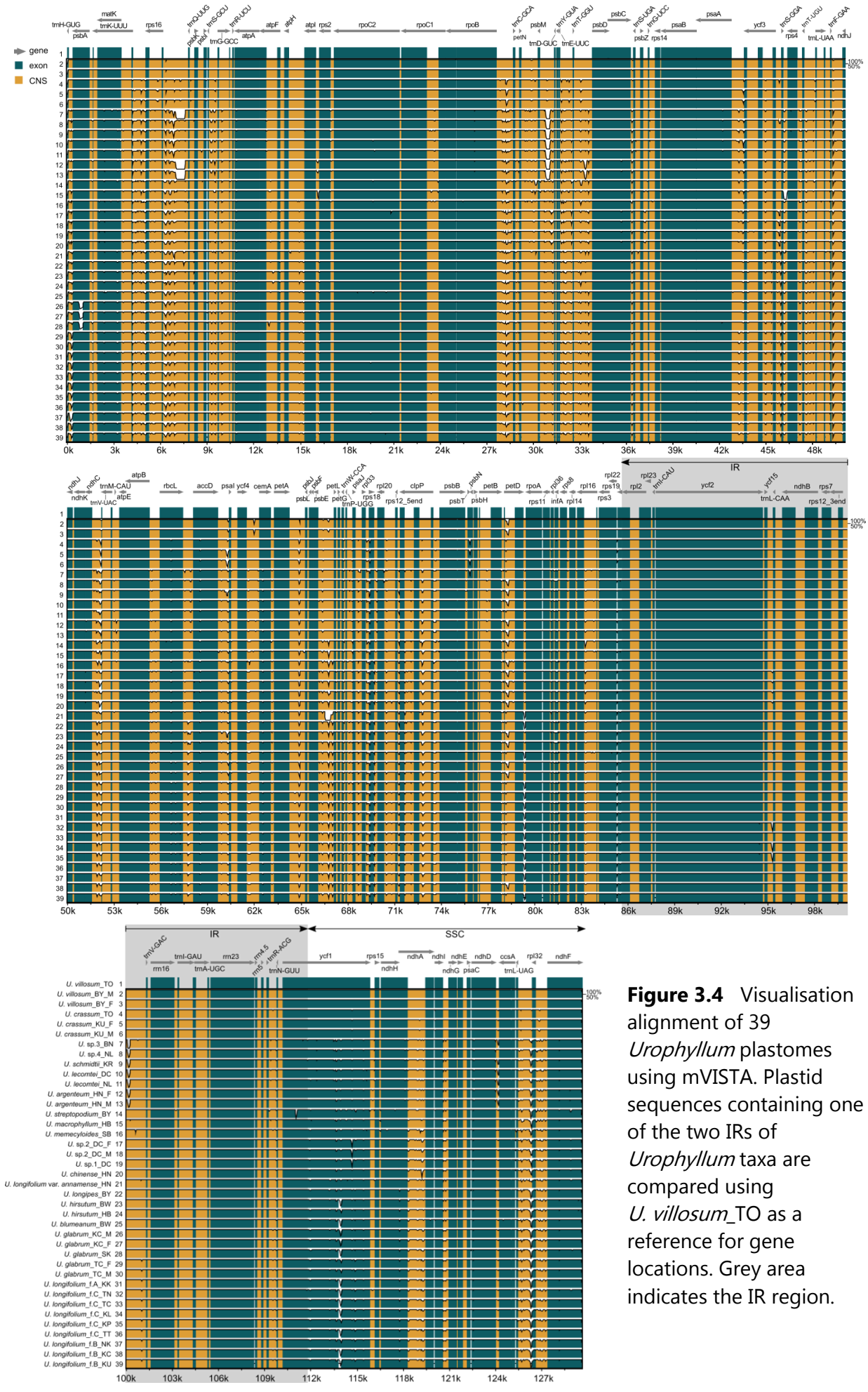


Figure 3.4 Visualisation alignment of 39 *Urophyllum* plastomes using mVISTA. Plastid sequences containing one of the two IRs of *Urophyllum* taxa are compared using *U. villosum*_TO as a reference for gene locations. Grey area indicates the IR region.

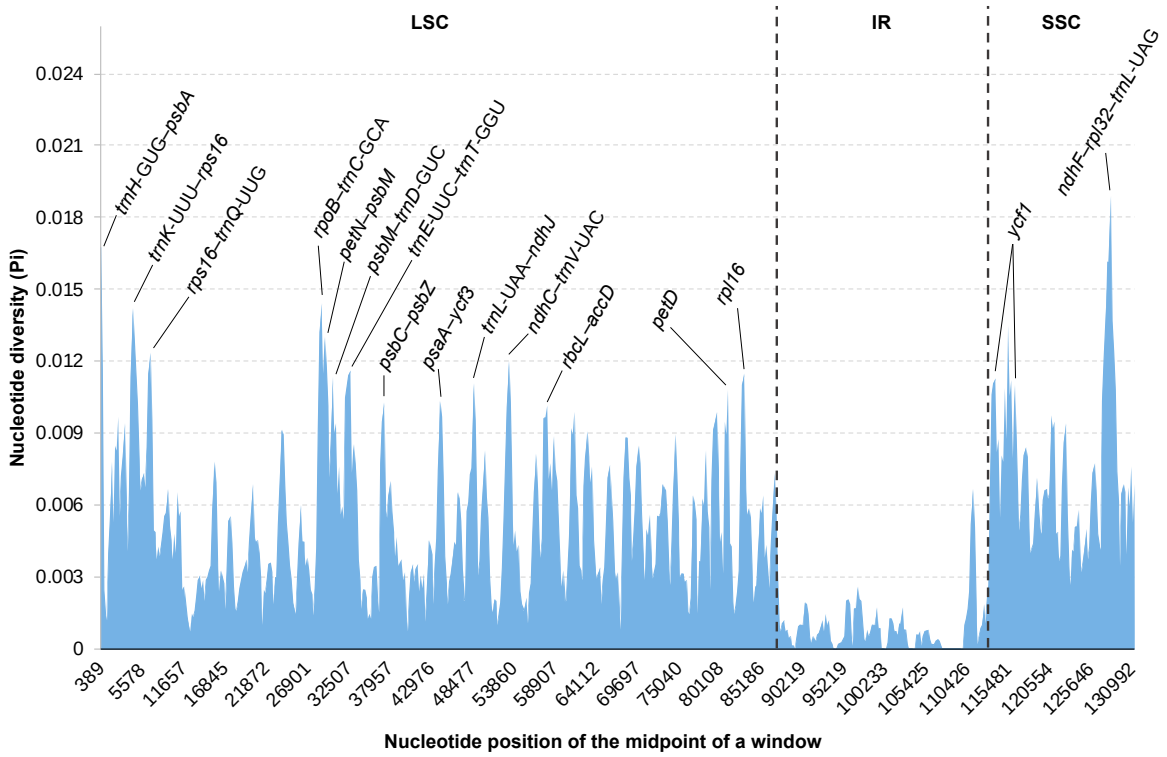


Figure 3.5 Sliding window analysis of 39 *Urophyllum* plastid genomes (window length: 600 bp, step size: 200 bp). Only one IR was included in each plastid genome. Regions with nucleotide diversity (Pi) more than 0.01 are labelled.

Table 3.7 Variable sites in 39 *Urophyllum* plastid genomes analysed by sliding window method. *= plastomes containing only one IR region.

Regions	Number of sites	Number of variable sites	Number of parsimony informative sites	Percentage of variable sites per total sites	Average nucleotide diversity (Pi)
Whole plastomes*	125,809	3,109	2,009	2.47%	0.00456
LSC	82,246	2,279	1,491	2.77%	0.00513
IR	25,578	137	79	0.54%	0.00080
SSC	17,985	693	439	3.85%	0.00728
CDS	68,829	1,253	777	1.82%	0.00328

3.3.4. Phylogenetic analyses

Both plastome and nrDNA phylogenetic trees reveal that all 18 *Urophylllum* spp. form a clade, *U. crassum* and *U. villosum* are sister to the rest of the species in the genus (Figure 3.6A and 3.7A).

3.3.4.1. Plastid genome data

Maximum likelihood (ML) and Bayesian inference (BI) analyses were performed with whole plastomes containing one IR dataset. The tree topologies of 39 *Urophyllum* samples based on the dataset were identical for both analyses. Therefore, the phylogenetic result reported here use BI tree with support values of both analyses shown at nodes (Figure 3.6).

All branches had both PP and BS support higher than 0.95 or 90% respectively, except for four branches; inside the *U. longifolium* and *U. glabrum* group, the branch leading to *U. hirsutum*, and a branch in the *U. chinense* group (Figure 3.6B). Overall phylogenies of plastome containing one IR are resolved to species and related species clades which were morphologically similar including *U. chinense* and *U. lecomtei* groups. (Figure 3.6B). One exception was found in *U. longifolium* clade where two samples of *U. glabrum* are nested within (Figure 3.6B). Both the pairwise distance and nucleotide substitution analyses support the close relationship of these two species (Table S11). Additionally, different sexes of the same taxon (indicated as F and M at the end of the name) consistently appeared together in the trees.

3.3.4.2. Nuclear ribosomal DNA tree

The BI and ML analyses of the nrDNA dataset resulted in identical tree topologies. Overall, species with more than one sample were resolved into monophyletic groups. Almost all branches had high support values (PP >0.95 and/or BS >90%) except from three branches: 1) *U. macrophyllum* and *U. streptopodium* with the support values 0.48(PP) and 44% (BS); 2) *U. longipes* with 0.58 (PP) and 44% (BS); and 3) the *U. glabrum* group with high support value of PP (0.99) but the BS support was 84%. The remaining 11 branches with low support were within species group clades.

3.3.4.3. Comparing nrDNA and plastome BI trees

When comparing the whole plastid to the nrDNA trees (Figure 3.7 and 3.8), *U. glabrum* and *U. longifolium* are resolved to form their own clades in nrDNA tree however in the plastid trees these two species form a group. The position of *U. longifolium* var. *annamense* in nrDNA is different, it is grouped with other taxa of *U. longifolium* as a sister taxon to the rest of the species. While in the plastid tree, this variety was at the node between *U. chinense* group and *U. longipes*. *U. memecyloides* was on its own in the plastid tree, however it is sister to *U. blumeanum* in nrDNA tree (Figure 3.7 and 3.8). Moreover, the placement of *U. schmidtii* changed from within the *U. lecomtei* group to be as a sister taxon to *U. longifolium* group.

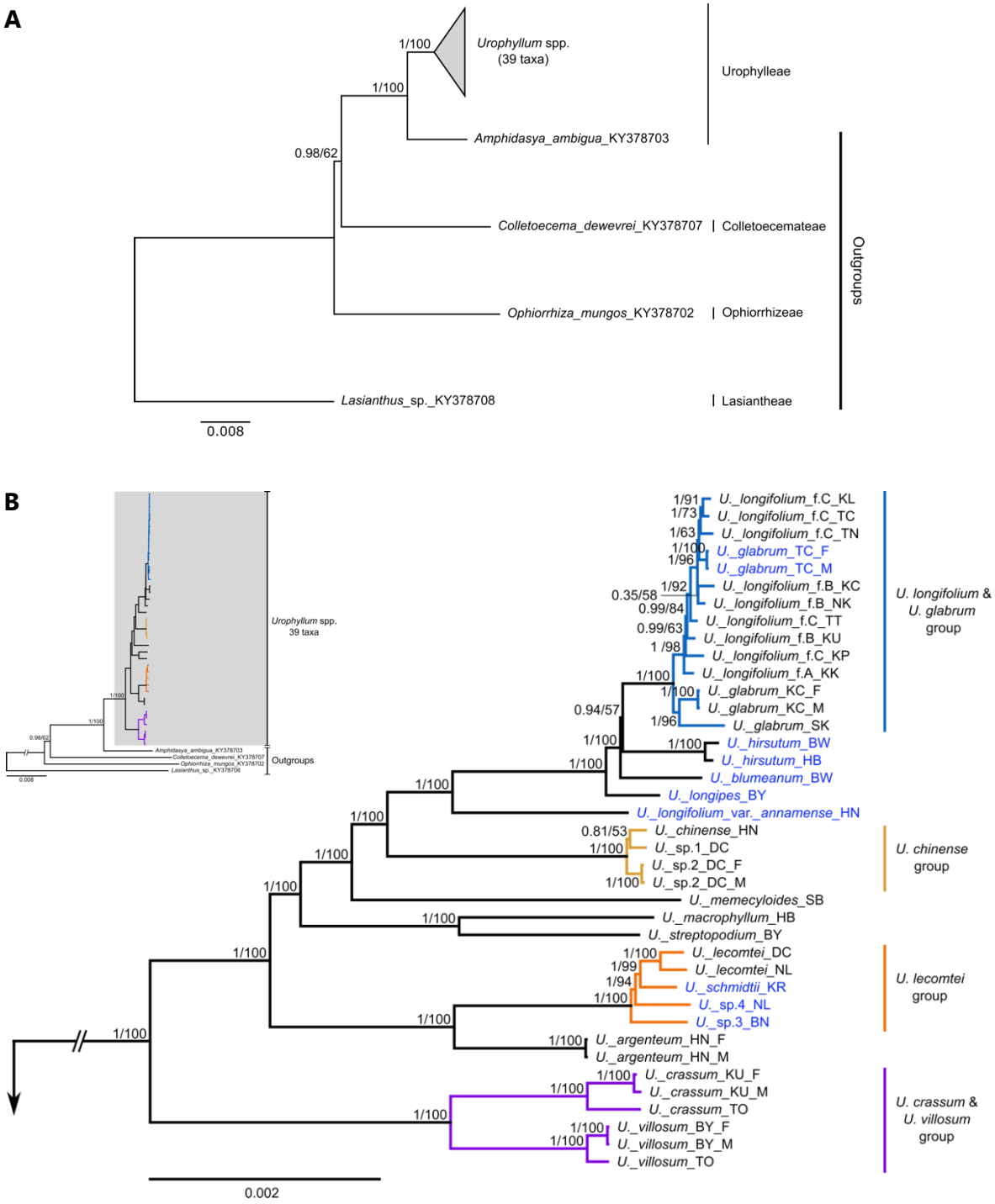


Figure 3.6 Bayesian Inference (BI) trees based upon different whole plastid sequences containing one IR. A) phylogenetic tree with branches collapsed; B) partial phylogenetic tree focusing upon *Urophyllum* samples only; inset - the whole phylogenetic tree including outgroup spp., *Urophyllum* taxa highlighted in grey. Number at each node represent Bayesian posterior probabilities (PP) and bootstrap support (BS), respectively. Taxa coloured blue indicate position changes among the plastid and nrDNA trees.

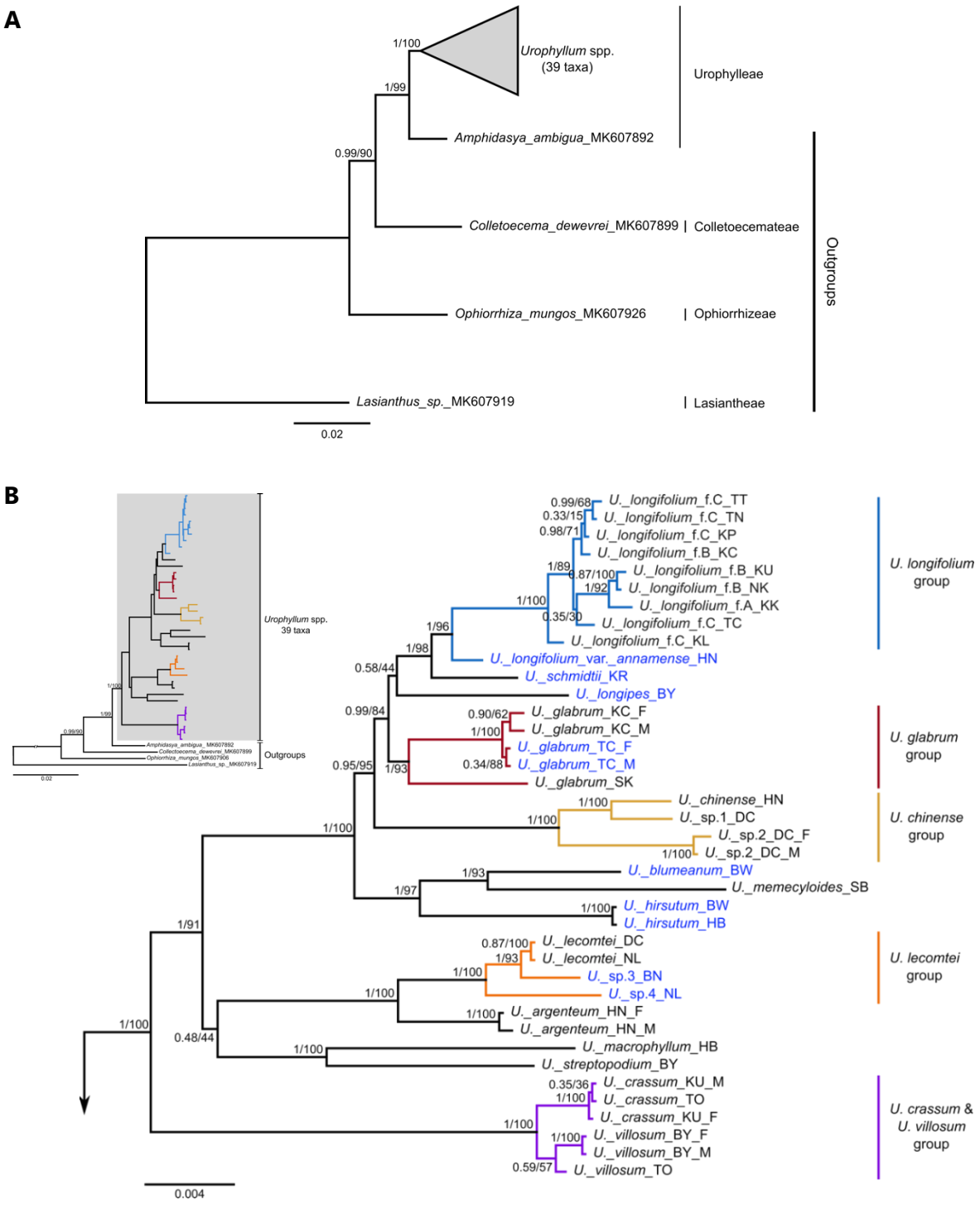


Figure 3.7 Bayesian Inference (BI) trees based upon nrDNA sequences. A) phylogenetic tree with branches collapsed. B) partial phylogenetic tree focusing upon the *Urophyllum* samples only; inset - the whole phylogenetic tree including outgroup spp., *Urophyllum* taxa highlighted in grey. Number at each node represent Bayesian posterior probabilities (PP) and bootstrap support (BS), respectively. Taxa coloured blue indicate position changes among the plastid and nrDNA trees.

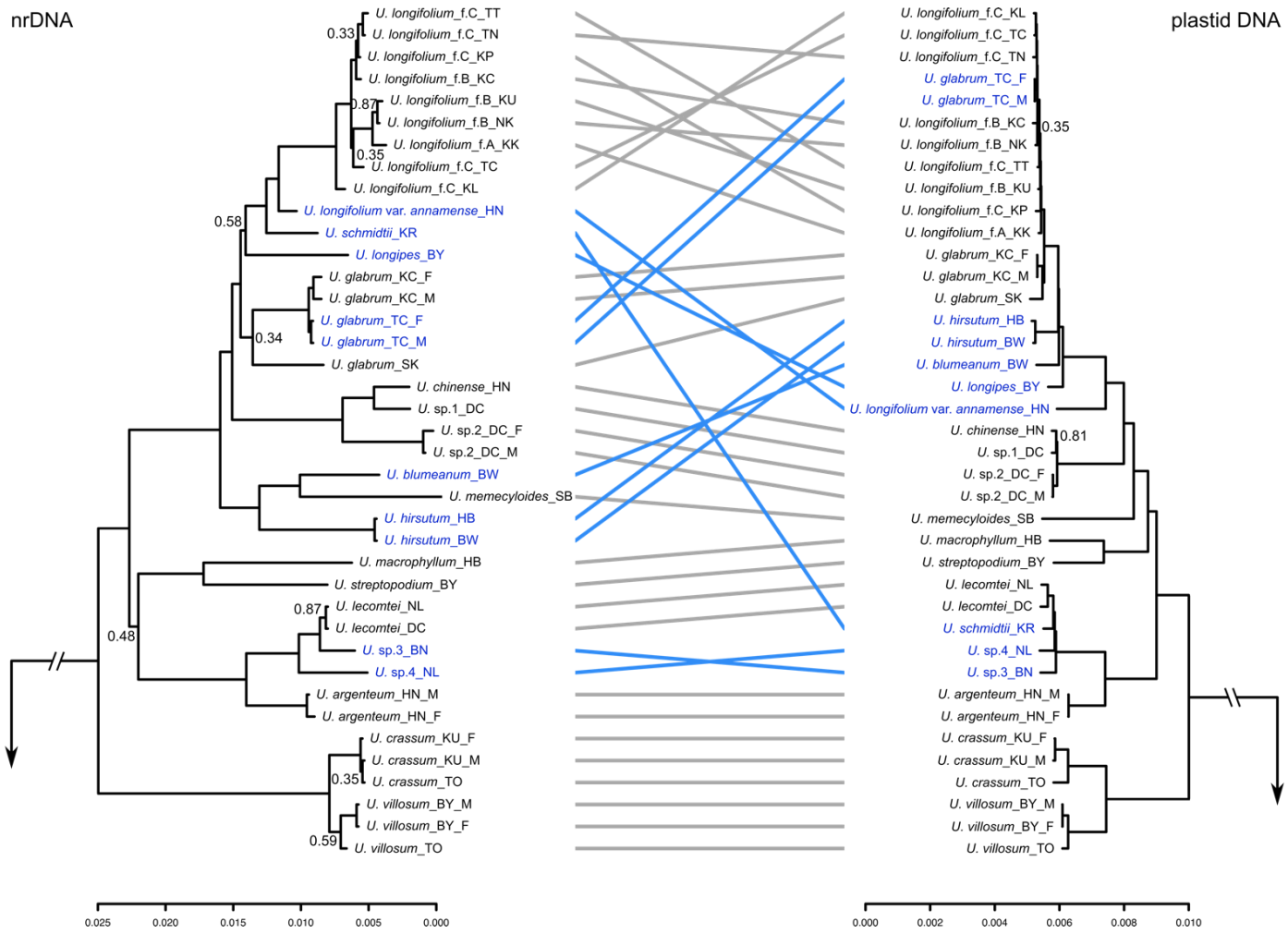


Figure 3.8 Comparison of nrDNA and plastid trees showing different species positions between the two datasets. Number at node indicates BI posterior probability (PP) when the value is lower than 1. Nodes with no support value mean PP > 0.9. Blue letters and lines indicate the position differences between two trees. Different positions within *U. longifolium* taxa were not hightlighted except *U. longifolium* var. *annamense*.

3.3.5. Character state reconstruction

Character state trees are shown in Figure 3.9–3.14 for both plastid and nrDNA sequences. The characters 1–4 presented in Figure 3.9 and 3.12 show that all 18 *Urophyllum* species have an entire stipule (Character 1). Most species (14 species) have stipules appressed to the stem but not folded (Character 2), only four species have a folded stipule: *U. argenteum*, *U. crassum, U. longifolium* taxa and *U. longipes*. The hairy pocket domatium at the secondary vein axil to the midrib (Character 3) is commonly found in every angle of *U. glabrum*, *U.* sp.4 and *U. villosum*. Whereas, it is glabrous and present in some angles in *U. schmidtii* and some samples of *U. longifolium* taxa (except *U. longifolium* var. *annamense* where the pocket domatium is absent). For *U. hirsutum*, the domatia has dense hairs at the angles. Many *Urophyllum* species (10 species) have festooned brochidodromous secondary vein loop patterns (Character 4), however *U.* sp.4 is the only species with conspicuously brochidodromous venation patterns.

Characters 5–8 are shown in Figure 3.10 and 3.13 for plastid and nrDNA trees, respectively. Many *Urophyllum* species in the study area have hairy petioles (Character 5), only *U.* sp.1. and *U.* sp.2 have glabrous petioles; in *U. schmidtii* and *U.* sp.3, the petioles are hairy with a higher density at the canalicular ridge. The inflorescence can be found as three main types (Character 6): 1) single flower, which is only found in some pistillate plants of *U. chinense* and *U. argenteum*; 2) cymose inflorescence which can be divided into two subcategories of simple or compound cymose, only *U. longipes* and *U. argenteum* (pistillate plant) have simple cymose inflorescence. Whereas, compound cymose inflorescences are found in *U. longifolium* taxa, a pistillate plant of *U. glabrum*, and staminate plant of *U. villosum*; 3) umbellate inflorescences can be divided into three subcategories as polymorphic characters within some species, for example, both pedunculate and sessile umbellate inflorescences are usually found within *U. crassum*, *U. hirsutum* and *U. streptopodium*, as well as pedunculate and two-tiered umbellate inflorescences are found in *U.* sp.3 and *U. villosum*. However, in *U. blumeanum* and *U. memecyloides*, the inflorescences are found as two-tiered umbellate, to date. For character 7, calyx lobe morphology, many *Urophyllum* either have a toothed or lobed calyx, only *U. memecyloides* has an entire calyx. The corolla colour (Character 8) can be found in three different patterns: 1) white corolla found in the *U. chinense* group, *U. lecomtei*, *U. longifolium* group (except

U. longifolium var. *annamense*), *U. longipes*, *U. schmidtii* and *U.* sp.3; 2) green corolla found in *U. argenteum*, *U. crassum*, *U. glabrum*, *U. longifolium* var. *annamense* and *U. villosum*; 3) mix between white/pale yellow, and green at apex, this calyx colour pattern is found in *U. hirsutum* and *U. streptopodium*. There are four species where the corolla colour is not yet known due to the lack of knowledge on living specimens: *U. blumeanum*, *U. macrophyllum*, *U. memecyloides* and *U.* sp.4.

The characters 9–11 are shown in Figure 3.11 and 3.14. Many *Urophyllum* species have glabrous corolla abaxially (Character 9). However, there are four out of 18 species in this study with hairy corolla abaxially: *U. argenteum*, *U. hirsutum*, *U. lecomtei* and *U. villosum*, while it is scaly all over in *U. crassum*. Hair density at the adaxial corolla lobes opening (character 10) of *U. blumeanum* was found to be unique with countable sparse hairs, whereas other species have many dense hairs. For character 11, the membrane at the corolla lobes opening where the hairs are attached, can be used to distinguish *U. crassum* and *U. villosum* with a triangular membrane attached to each corolla lobe, and *U. glabrum* with tubular membrane connecting to all corolla lobes.

Focusing on matching character states to the tree topologies identified from molecular datasets, a population of *U. glabrum* is nested within *U. longifolium* group in plastid tree where they form a clade in nrDNA, characters 1 (stipules), 2 (pocket membrane), 4 (secondary vein loop patterns), 8 (corolla colour) and 11 (membrane at corolla lobed base opening) are matched to the position of the *U. glabrum* group following nrDNA tree (Figure 3.12 to 3.14). Whereas, character 3 and one sample in character 6 are partially matched to the position in plastid tree (Figure 3.9 and 3.10). The incongruence of the positions among plastid and nrDNA trees are also found in *U. longifolium* var. *annamense* where the character states mostly match with the nrDNA trees except for corolla colour (character 8) where the variety has green colour instead of white like other *U. longifolium* taxa. In the cases of *U. blumeanum*, *U. hirsutum*, *U. longipes* and *U. schmidtii*, there is no conclusive match between plastid and nrDNA trees. For *U.* sp.3 and *U.* sp.4, the incongruence positions are limited to within a clade, therefore character states were inconclusive.

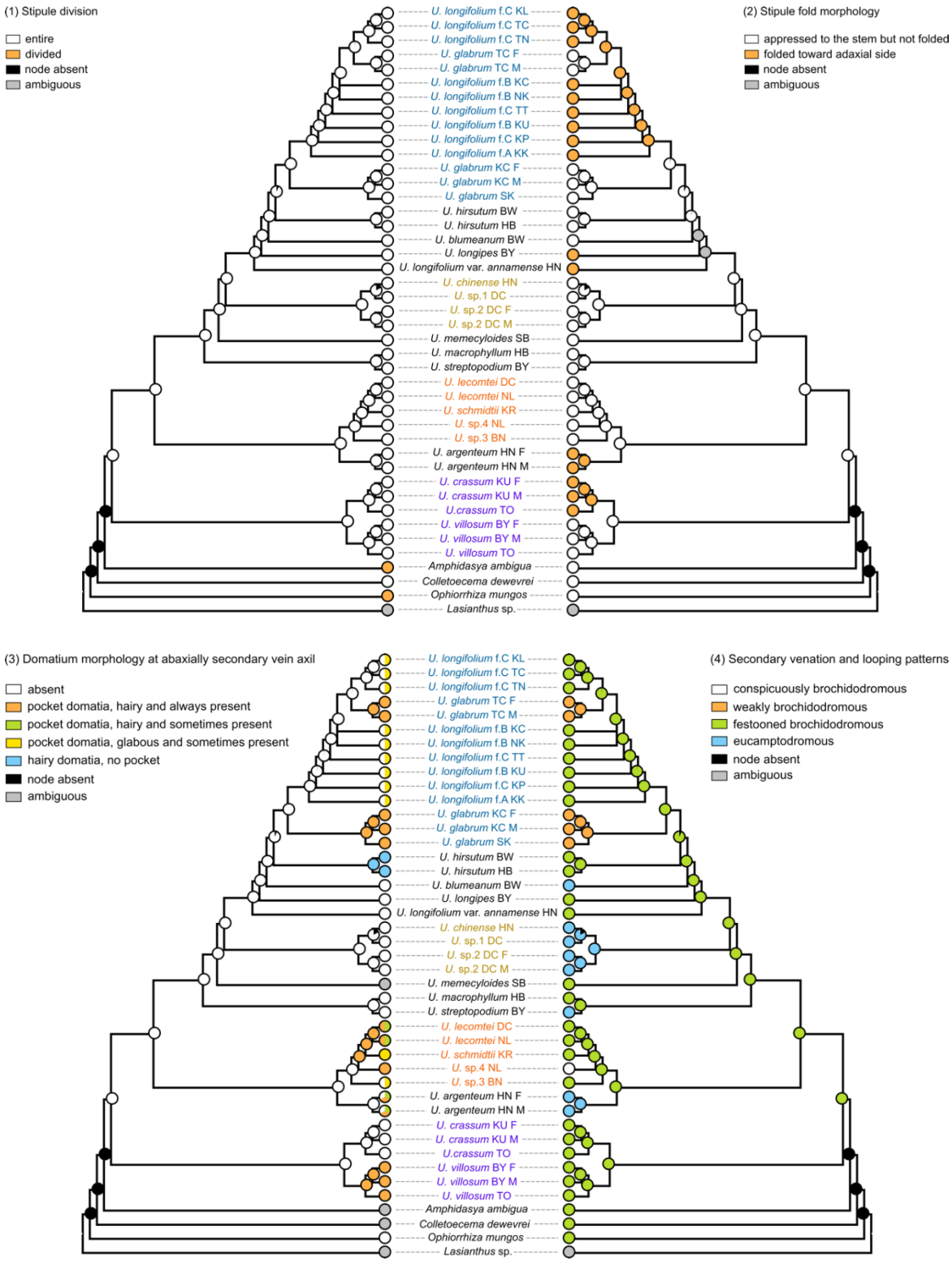


Figure 3.9 Ancestral character states of characters 1–4 reconstructed on a majority rule consensus tree (BI) of plastid sequences containing one IR. Taxon colour indicates group following Figure 3.6B.

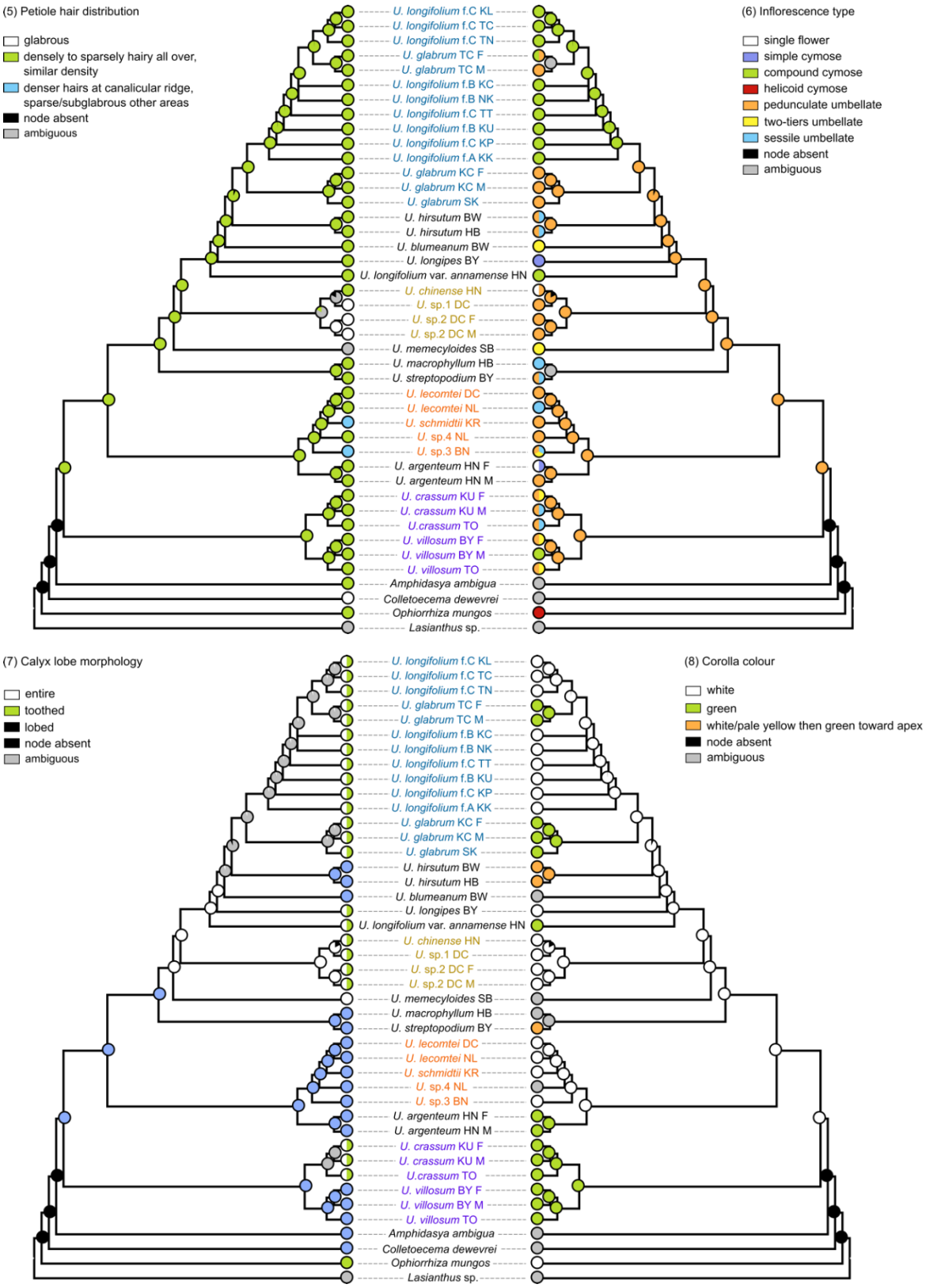


Figure 3.10 Ancestral character states of characters 5–8 reconstructed on a majority rule consensus tree (BI) of plastid sequences containing one IR dataset. Taxon colour indicates group following Figure 3.6B.

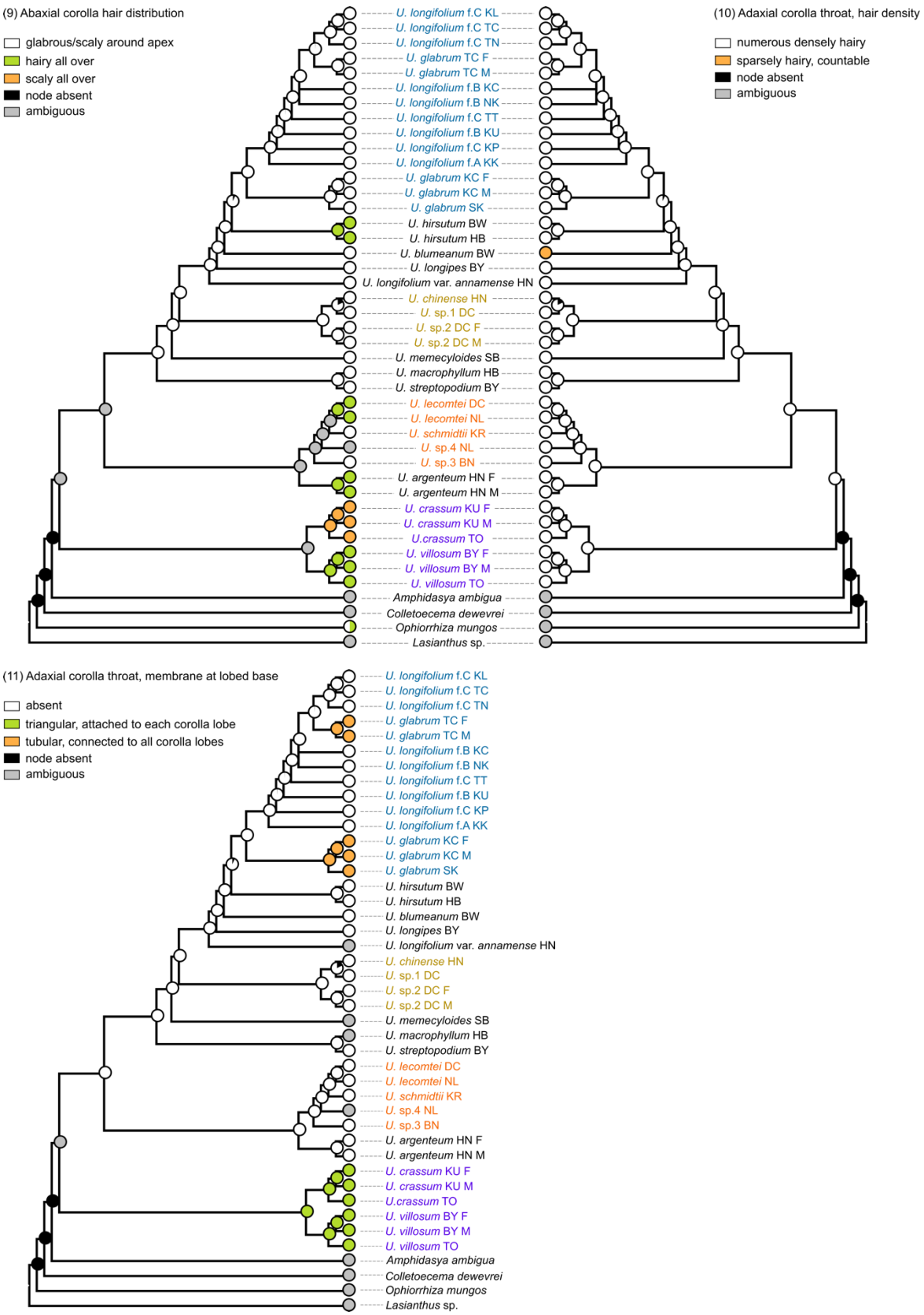


Figure 3.11 Ancestral character states of characters 9–11 reconstructed on a majority rule consensus tree (BI) of plastid sequences containing one IR dataset. Taxon colour indicates group following Figure 3.6B.

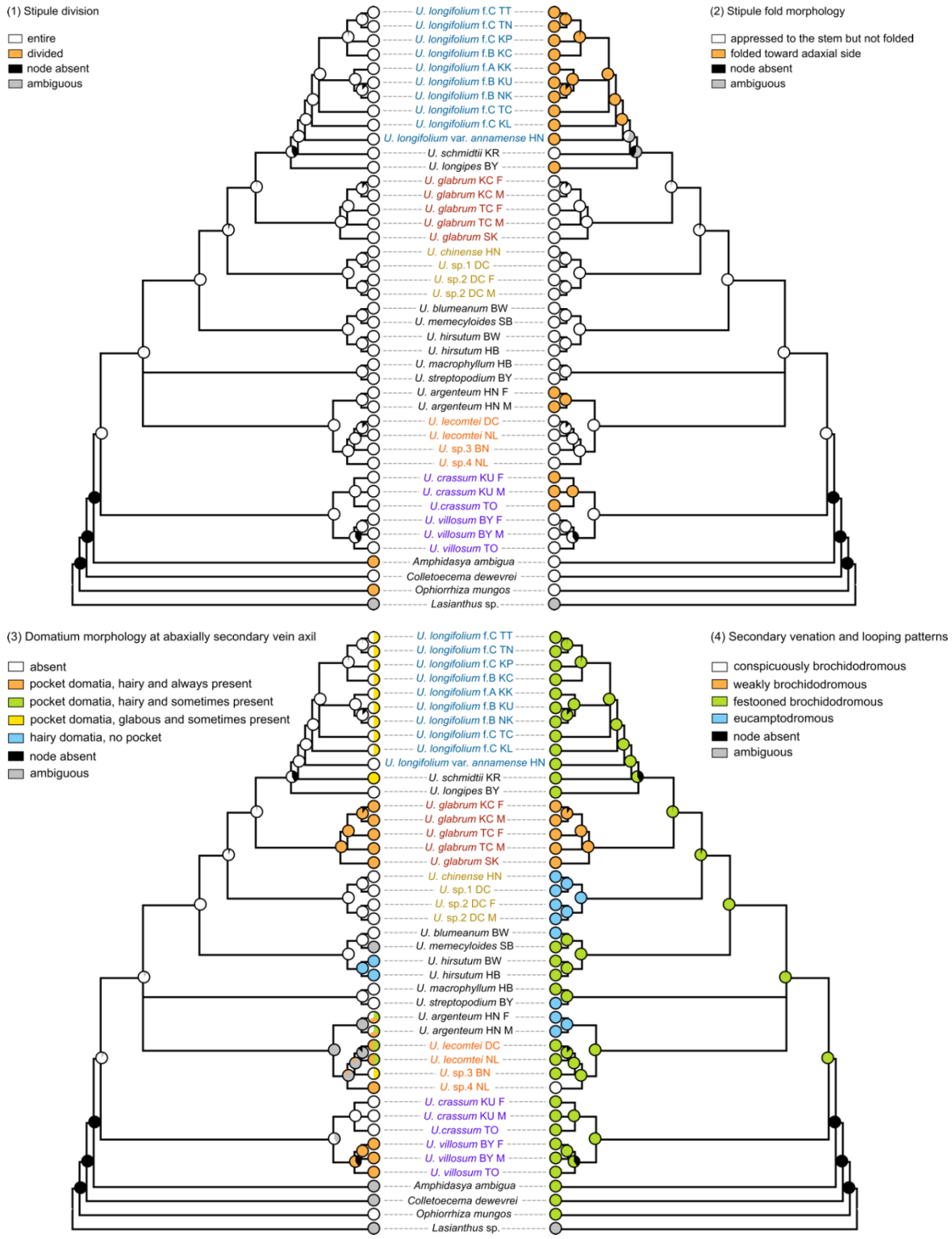


Figure 3.12 Ancestral character states of characters 1–4 reconstructed on a majority rule consensus tree (BI) of nrDNA sequences. Taxon colour indicates group following Figure 3.7B.

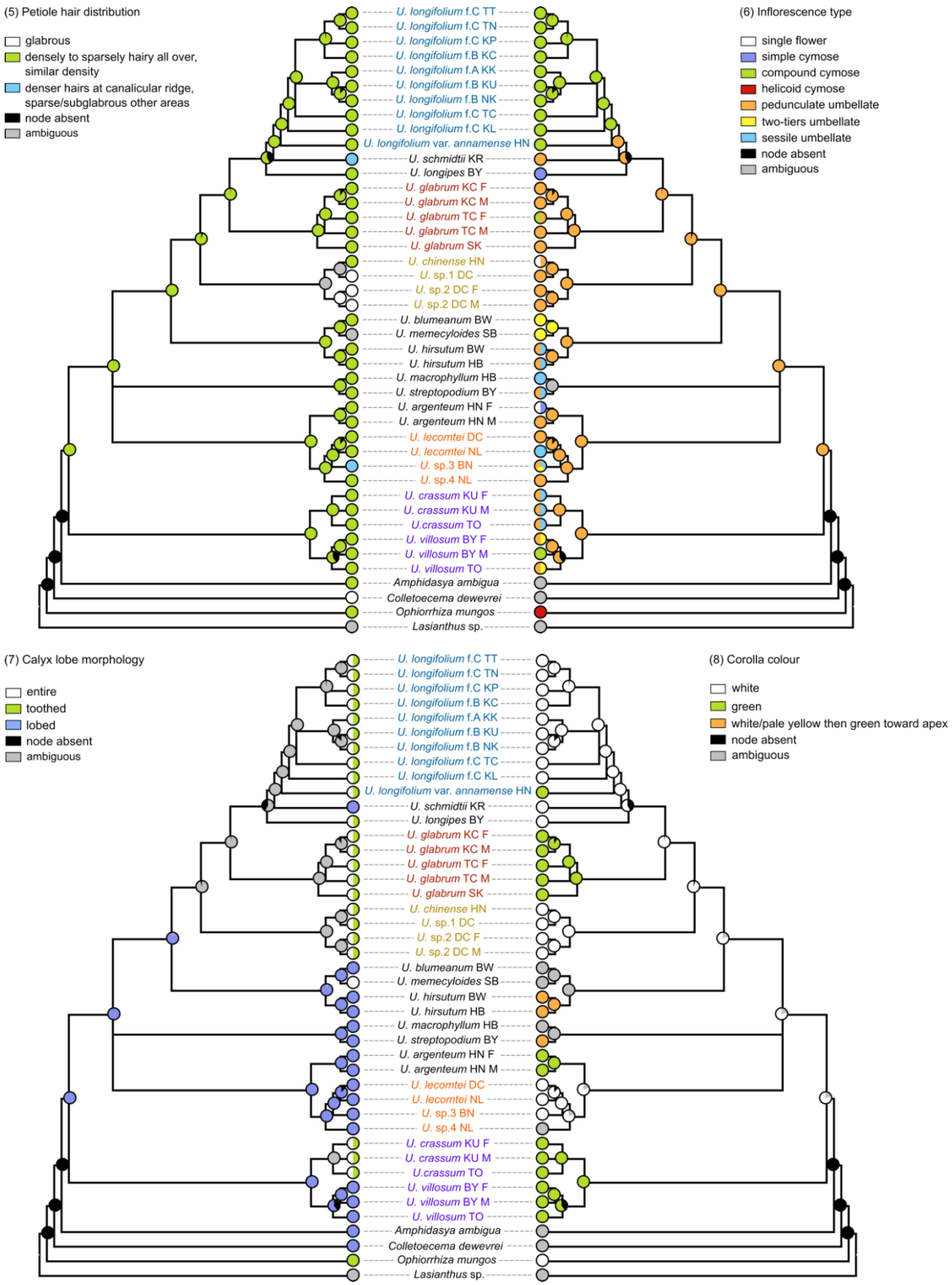


Figure 3.13 Ancestral character states of characters 5–8 reconstructed on a majority rule consensus tree (BI) of nrDNA sequences. Taxon colour indicates group following Figure 3.7B.

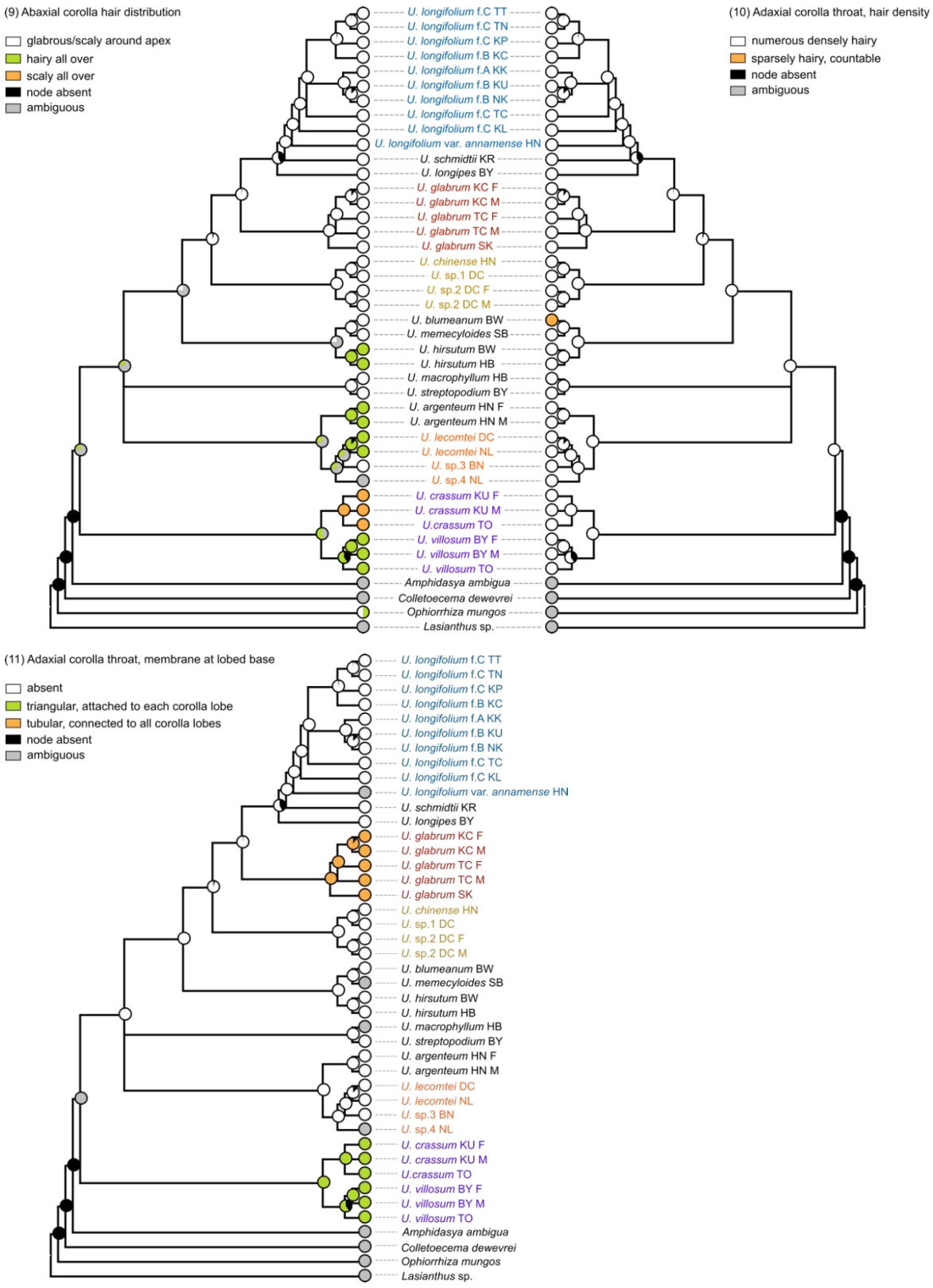


Figure 3.14 Ancestral character states of characters 9–11 reconstructed on a majority rule consensus tree (BI) of nrDNA sequences. Taxon colour indicates group following Figure 3.7B.

3.4. Discussion

3.4.1. Plastome features

All 39 plastomes of 18 *Urophyllum* species were successfully assembled in this study. They show a highly conserved quadripartite structure that is typical of angiosperm plastid genomes (LSC, SSC and two IRs) with 113 matching genes arranged in the same order. The plastomes differed in length by 1,306 bp with a range from 154,099 to 155,405 bp. There was minor variation in the position of the IR boundaries (Figure 3.3). Plastomes of *Urophyllum* species in this study had a similar gene composition and arrangement to those found in previous studies of Rubiaceae plastomes (Zhang *et al.*, 2016; X. F. Zhang *et al.*, 2019; Y. Zhang *et al.*, 2019; Ly *et al.*, 2020). However, there is variation in the IR endpoint that usually either contracts or extends to the surrounding genes or intergenic spacers (IGS). These small changes (<100 bp) of IR boundaries are common in angiosperm plastomes (Downie and Jansen, 2015). The LSC/IRb border for example, was found either within the *rps19* gene or the IGS between the *rps19* and *rpl2* genes in *Urophyllum* species, similar patterns have also been reported in subfamily Ixoroideae and *Gynochthodes officinalis* (syn. *Morinda officinalis*) (Zhang *et al.*, 2016; Ly *et al.*, 2020). However, the LSC/IRb boundary for two species in subfamily Rubioideae, *Hedyotis ovata* (X. F. Zhang *et al.*, 2019) and *Paralasianthus hainanensis* (syn. *Saprosma merrillii*) (Zhu *et al.*, 2019), extended to include the entire *rps19* gene in the IR. Changes in the IR border are suggested to result from either gene conversion (Goulding *et al.*, 1996) or double strand breakage and the subsequent repair process, followed by recombination at polyA tracts (Wang *et al.*, 2008). In the case of *Urophyllum*, IR border changes are likely due to gene conversion rather than double strand breakage as the latter process is extended to include one to many genes into the IR.

3.4.2. Variable and repeat regions in *Urophyllum* plastomes

Nucleotide variability within *Urophyllum* plastomes were found mainly in the LSC and SSC regions, and very low variation in the IR region. It is due to the fact that there are two identical copies of IR in the plastid genomes. If mutations occur in one IR, error correction can take place from the other, to minimise deleterious mutations (Weng *et al.*, 2016). The decrease in divergent hotspots in the IR compared to the SC regions has also been

observed in many other plastomes (Chen, Wu and Zhang, 2019; Thode and Lohmann, 2019; Valencia-D *et al.*, 2020). Nucleotide diversity revealed 18 variable regions (Pi≥0.01) that include 15 intergenic spacers and three genes (Figure 3.5). Of these variable regions some have been used as DNA markers for different taxonomic level studies in Rubiaceae including *accD-psaI*, *petD*, *rpl16*, *rpl32-trnL*, *trnH-psbA* (Maurin *et al.*, 2007; Tosh *et al.*, 2009; Cristians, Bye and Nieto-Sotelo, 2018; De Block *et al.*, 2018). Although, this suggests that these regions are variable within Rubiaceae and therefore suitable as DNA barcodes for identifying many species within the family, there are two species that are nested together in plastid tree of *Urophyllum*. In contrast, all species are resolved in the nrDNA tree and this is therefore better suited for DNA barcoding. DNA barcoding may be particularly important for accurate species identification such as from pollen on bees (Kamo *et al.*, 2018) or pollen on nectar feeding bats (Lim *et al.*, 2018) as well as forensic wood identification (Jiao *et al.*, 2018). In order to gain a better understanding of the phylogenetic relationship of species, this study demonstrates that the analyses of both plastid and nrDNA regions are required.

Single sequence repeats (SSRs) have been used as genetic markers for population genetics and genome polymorphism studies (Ebert and Peakall, 2009; Hendre and Aggarwal, 2014; Qi *et al.*, 2016). The number of SSRs in *Urophyllum* ranged from 32–44 repeats mostly located in LSC and rarely in IR. Like other angiosperms, mononucleotide of A or T motifs were the most common repeats found in the plastomes (Chen, Wu and Zhang, 2019; Thode and Lohmann, 2019). In this study, the number of repeat regions in *Urophyllum* range from 21 to 29 which are dispersed mainly in non-coding regions. The most abundant repeat sequences were 30–39 bp long, with palindromic repeats being the most common. Plastid SSRs provide insight into seed flow within the population (Provan, Powell and Hollingsworth, 2001). It gives useful information on the phylogeographical pattern of a plant group however, using them alone may not be sufficient enough to delimit species boundaries (Delplancke *et al.*, 2012). The study here found that there was variation of the repeat number for several SSR motifs (especially A and T repeats) within species, where multiple populations had been sampled (*U. longifolium* and *U. glabrum*) (Table S9). Therefore, the SSR motifs identified in this study could be useful for population level study.

3.4.3. Phylogenetic relationship

The phylogenetic analyses were performed on two datasets including plastid genomes and nuclear ribosomal DNA which represent different inheritance lineages of 18 *Urophyllum* species (Figure 3.6 and 3.7). There is broad congruence between the two data sets and many species form a clade including *U. crassum* and *U. villosum* as the basal clade of *Urophyllum* in this study. *Urophyllum argenteum* is sister to the *U. lecomtei* clade. In many cases, closely related species found in the plastid and nrDNA trees are congruent with morphological characters (Figure 3.9 to 3.14). This includes the *U. chinense* group where two unidentified species (*U.* sp.1 and *U.* sp.2) share most morphological characters with *U. chinense*. The diagnostic character for the two unidentified species is petiole hair distribution which is glabrous instead of densely hairy all over in *U. chinense*. Results from the phylogeny in this study indicate that *U.* sp.1 is sister to *U. chinense*, therefore it could be morphological variation within species, whereas *U.* sp.2 is monophyletic. Therefore, this group will require further taxonomic work to ascertain the relationship between *U. chinense* and the unknown taxa (Chapter 5).

The position of *U. blumeanum* and *U. hirsutum* were incongruent between the plastid and nrDNA phylogenies. Within the plastid tree, *U. hirsutum* is sister to the *U. longifolium* and the *U. glabrum* group with *U. blumeanum* as a sister to all these taxa. While in the nrDNA tree, *U. blumeanum* is placed as a sister to *U. memecyloides* with *U. hirsutum* as a sister to these two species. There are morphological characters shared between *U. blumeanum* and *U. memecyloides* such as a two-tiered umbellate inflorescence and glabrous corolla that support the nrDNA results (Figure 3.12 and 3.14). To ascertain the relationship within this group, *U. arboreum* (morphologically similar to both *U. blumeanum* and *U. memecyloides*) needs to be sampled. Unfortunately, it could not be sampled in this study, however it is being sourced from herbarium material for future work. Similarly, conflict has been found in a study of tribe Urophylleae (Smedmark and Bremer, 2011) which suggests uncertain species-level relationships within genus *Urophyllum*. Smedmark and Bremer (2011) sampled four species of *Urophyllum* that are included in this study, and found an incongruent relationship between plastid DNA (*rps16* intron and *trnTF*) and nrDNA (ITS and ETS) trees in the placement of three species: *U. schmidtii, U. blumeanum* and *U. longifolium.* Based upon nrDNA,

U. longifolium is sister to *U. schmidtii* and not *U. blumeanum* as with plastid DNA (Smedmark and Bremer, 2011). The results here show a similar incongruent pattern for these taxa based upon increased data. The fourth species reported in Smedmark and Bremer (2011) is *U. streptopodium* which was located in a different clade from the three species stated above. Their results show that the placement of *U. streptopodium* is incongruent between the plastid and nrDNA trees, either sister to *U. congestiflorum* Ridl. (nrDNA) or within a polytomy with eight other species (plastid) (Smedmark and Bremer, 2011). While in the study here, *U. streptpodium* is a sister to *U. macrophyllum*. However, Smedmark and Bremer (2011) did not sample *U. macrophyllum* and *U. congestiflorum* was not sampled in this study, therefore the placement of *U. streptopodium* and how these three species are related remains uncertain. Wider sampling of *Urophyllum* taxa and the use of other genome datasets (mitochondrial and nuclear) may help to understand the relationship between these closely related species.

Urophyllum longifolium and *U. glabrum* form a clade in the plastid trees. Samples from a population of *U. glabrum* collected in Ton Chong Fa Falls (TC) are nested inside *U. longifolium* clade, that were collected from the same waterfall (TC) and a nearby national park headquarters (Khao Lak-Lamru (KL)) (c. 6 km away) with one sample of *U. longifolium* collected from c. 30 km further away in Si Phangnga National Park (TT). However, the two species are morphologically distinct; *U. glabrum* has flat stipules (character 1), the presence of domatium-like structure (presence of a pocket membrane at every angle of secondary vein to midrib(character 2)), weakly brochidodromous secondary vein loop (character 4), green corolla (character 8), and the presence of a tubular membrane at the corolla lobe opening (character 11), in contrast to *U. longifolium,* has folded stipules, the pocket membrane absent or if present, only at some angles, festooned brochidodromous secondary vein loop, white corolla, and lacking the membrane at the corolla opening (Figure 3.12 to 3.14). In the nrDNA tree, *U. glabrum* and *U. longifolium* are both monophyletic with high posterior probability support (0.99 and 1, respectively). In spite of no evidence on sexual incompatibility between these two species, the contrasting pattern may be a result of hybridization. Natural hybridisation in dioecious plants has been reported in closely related species of several plant groups include *Ficus* Tourn. ex L. and *Nepenthes* L. (Parrish *et al*., 2003; Peng and Clarke, 2015). In *Nepenthes* species,

natural hybridisation was occasionally observed within disturbed habitats which caused a disruption on flowering seasons of co-occurring *Nepenthes* species, resulting in the cross-pollination between these species (Peng and Clarke, 2015). In *Ficus* species, hybridisation were thought to be a result of pollinator-specificity breakdown especially on islands or in harsh environmental habitats (Janzen, 1979). The study on three *Ficus* species (*F. septica* Burm.f., *F. fistulosa* Reinw. ex Blume and *F. hispida* L.f.) on Krakatau islands, Indonesia by Parrish *et al.* (2003), found natural fertile hybrids between the three *Ficus* species using AFLP and plastid haplotypes from nine intergenic spacers. Parrish *et al.* (2003) suggested that hybridisation may be caused by the lower numbers or absence of mutualistic wasp pollinators on one species of the *Ficus* which led the isolated individual plant to accept the pollinator wasps that are abundant in number instead. This might have happened in the case of *U. longifolium* and *U. glabrum*. Since two individuals of *U. glabrum* are nested inside the *U. longifolium* group (plastid tree), were collected at the northern distribution limit of the species which could suggest a hybrid established in this region. Therefore, further study on pollinators and reproductive system might help to understand the relationship between these two species. The phylogenetic pattern could also be explained by the early stage of speciation and therefore incomplete lineage sorting of these taxa (Zhou *et al.*, 2017; del Valle *et al.*, 2019). This is reflected in the short branches on the phylogenies. Another possible explanation of the incongruence may be caused by plastome capture in the form of horizontal gene transfer (HGT) between sympatric species. The evidence was tested under a man-made setting via a grafting method among *Nicotiana* L. species (Stegemann *et al.*, 2012). Even though grafting has been reported to happen naturally among trees especially root graft (Graham and Bormann, 1966; Lev-Yadun and Sprugel, 2011), there is no record of the event occurring in shrubs and small trees. Therefore, this is unlikely to be the case for the *U. longifolium* and *U. glabrum* relationship. Further population level study into the variation within *U. longifolium* and the relationship with *U. glabrum* is required using low-copy nuclear markers which represent biparental lineages (Zhang *et al.*, 2012) as shown to be a useful tool to resolve low-level taxonomic relationships (Meseguer *et al.*, 2014).

For the *U. lecomtei* group, there is incongruence between plastid and nrDNA trees in the placement of *U. schmidtii,* which is found either within *U. lecomtei* group (plastid) or as

sister to *U. longifolium* group (nrDNA). *U. schmidtii* is found in eastern Thailand (Chanthaburi and Trat Provinces) and Cambodia. The distribution of *U. lecomtei* and two unidentified species (*U.* sp.3 and *U.* sp.4) are in central to southern Vietnam. At this point, it seems that the placement of *U. schmidtii* as sister to *U. lecomtei* in the plastid data might be a result of hybridisation and incomplete lineage sorting that is similar to the case of *U. longifolium* and *U. glabrum* stated above. However, there are limited number of plants recorded from Cambodia so this cannot be certain.

Morphologically different taxa within the *U. longifolium* group were not recovered using the molecular data (Figure 3.6 and 3.7). Phylogenetic trees from both plastid and nrDNA data show all taxa: *U. longifolium* var. *longifolium* (A), *U. longifolium* var. *pilosum* (B) and *U. talangense* (C), are mixed except *U. longifolium* var. *annamense*, particularly considering plastid data where it is placed in a different branch from the rest of the species. Differences in morphological characteristics could therefore, reflect localised adaptation to habitat that has led to the taxonomic inflation. This is especially true for leaf size, and hair density and angle that can be different due to environmental effects, this has been reported in many species, such as *Cynoglossum officinale* L., *Centaurea* L. spp. *Festuca* spp., *Quercus* spp. *Sinapis arvensis* (Upadhyaya and Furness, 1994; Ramesar-Fortner, Dengler and Aiken, 1995; Roy, Stanton and Eppley, 1999; Mediavilla *et al.*, 2019). The morphological characters and distribution range of *U. longifolium* var. *annamense* are different from other *U. longifolium* varieties and *U. talangense*. At present it is only known to occur in Vietnam. It has a green corolla instead of a white corolla like other varieties of *U. longifolium* (Figure 3.10 and 3.13). *U. longifolium* var. *annamense* is placed in a different clade in plastid tree, which tended to group species based upon geographical distribution, this was also reported in *Ostrya* Scop. species using whole plastid data of multi-samples per species (Jiang *et al.*, 2019). Furthermore, it is sister to the other varieties in *U. longifolium* in nrDNA tree indicating it is closely related to the group, which is supported by morphological data. This evidence might suggest the change in taxonomic level. Although, more samples of *U. longifolium* var. *annamense* would be required to resolve the phylogenetic relationship with *U. longifolium*.

3.5. Conclusion

This work has provided the first complete plastid genomes of the genus *Urophyllum* and an estimate of the phylogenetic relationship between 18 *Urophyllum* species distributed in Thailand and Vietnam. Plastomes had canonical quadripartite structure and were highly conserved among *Urophyllum* species. Phylogenetic analyses performed on two datasets, both plastid and nuclear ribosomal DNA reveal incongruence for complex species groups such as *U. longifolium* that will require further study to ascertain species boundaries. Plastid trees revealed the possibility of hybridisation and incomplete lineage sorting between *U. glabrum* and *U. longifolium* distributed within the same geographical range. Almost all species formed a clade in the nrDNA tree, and these data provide support for morphological boundaries, whereas position of species on plastid tree seem to be associated with geographic distribution. This study represents the most comprehensive molecular study of *Urophyllum* to date, revealing levels of variation and importantly reveals species complexes within the genus that require further investigation.

Chapter 4 Synopsis of *Urophyllum* species in Thailand and their occurrence in Peninsular Malaysia and Singapore

4.1. Introduction

4.1.1. Genus *Urophyllum*

Urophyllum was first published by Nathaniel Wallich in Roxburgh (1824: 184) which enumerated two species in the genus: *U. villosum*, and *U. glabrum*. The authority for the genus, Wallich, was doubted by Griffith (1844) (footnote of page 17), who instead proposed that the authority should be William Jack as both the description of the genus and the two species were prepared by him. Merrill (1952) disagreed with Griffith (1844), on the basis that Jack proposed to name the genus either *Patisna* or *Wallichia* (Gage and Burkill, 1916), but ultimately left Wallich to decide. However, the two names suggested by Jack could be either misleading or illegitimate. Gage and Burkill (1916) compiled Jack's works based upon his letters, and may have mistaken 'Patisna' for 'Patima' from the original version (pages 196–198). This was supported by a label written by Jack on a specimen of *U. glabrum* (barcode E00130812) in the Herbarium of Royal Botanic Garden Edinburgh which indicated that the unknown genus was similar to *Patima* Aubl. (the same text is found in Wallich's protologue of *U. glabrum*). This justification was also communicated in the work of Cowan (1954). Furthermore, Wallich may have known that the genus name 'Wallichia', as proposed by Jack, was already used by Roxburgh (1820) in Arecaceae. Therefore, the name 'Urophyllum' was selected by Wallich based upon Jack's description of the acuminate leaves. According to Art. 46.2 of the ICN Shenzhen Code (2018), *Urophyllum* Wall. is used here.

The typification also proved controversial with regards to which species should be used as a type species for *Urophyllum* Wall.; Bremekamp (1940) suggested that *U. arboreum* (Reinw. ex Blume) Korth. (1851: 194) was more suitable than *U. villosum* (the first species in Wallich's protologue) based upon two reasons: 1) *Urophyllum* was divided into several genera in Bremekamp (1940) and *U. villosum* was transferred to the genus *Maschalocorymbus* Bremek.; and 2) *U. arboreum* (syn. *Wallichia arborea* Reinw. ex Blume (1823: 11)) was the earliest known published species of *Urophyllum*, even though the

genus name 'Wallichia' was illegitimate. However, phylogenetic analyses by Smedmark & Bremer (2011) on plastid and nrDNA sequences (*rps*16, *trn*T–F, ITS and ETS) did not support the several genera recognised by Bremekamp (1940), and therefore *Maschalocorymbus* was combined to *Urophyllum sensu lato*. This meant that there was no conflict in designating *U. villosum* as the type species of the genus (Wong *et al.*, 2019).

The taxonomic revisions of *Urophyllum* in the 1800s synonymised the genus *Axanthes* Blume and the species within this genus described both by Blume (1826) and Wight (1847) to *Urophyllum* (Korthals, 1851; Hooker, 1880). Revisions at the regional level started from the 1900s that mostly contributed to the Flora of Malay Peninsula, Flora of Malaya and Flora of Singapore (Ridley, 1923, 1932; Wong, 1989; Wong *et al.*, 2019). Only three publications relating to the areas in Thailand have been published (Schmidt, 1902; Craib, 1931, 1932), with the latest revision by Craib (1932). Species delimitation of *Urophyllum* found in Thailand is challenging as no taxonomic revision includes key characters or identification keys. Therefore, the genus is in need of taxonomic revision.

The geographical areas in Thailand where *Urophyllum* species are found can be divided into two main regions: 1) Eastern Thailand, on the Cardamom mountain ranges which lie from the Khao Soi Dao Mountains in Khao Soi Dao Wildlife Sanctuary, Chanthaburi Province (Thailand) to the Elephant Mountains in Bokor National Park, Kampot Province (Cambodia) (Grismer *et al.*, 2011); and 2) Peninsular Thailand where there are four mountain ranges from Myanmar to Peninsular Malaysia (Figure 4.1): 1) the Tenassarim Mountains (border between Thailand and Myanmar); 2) the Nakhon Sri Thammarat Mountains (Peninsular Thailand); 3) the Sankalakhiri Mountains (between Satun–Songkhla Provinces (Thailand) and Perlis State (Peninsular Malaysia); and 4) the Titiwangsa Mountains (along Thai–Malaysian border and into Peninsular Malaysia) (Grismer *et al.*, 2011). Given the overlap of the distribution of *Urophyllum* in Thailand and neighbouring countries: Cambodia, Myanmar, and Peninsular Malaysia, the present taxonomic study of *Urophyllum* species in Thailand does not only consider specimen records within a political border but extends to the surrounding regions.

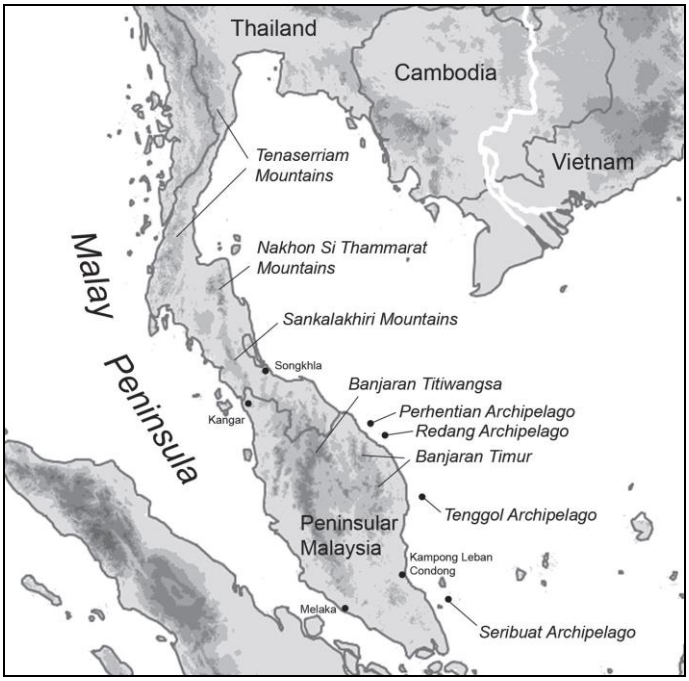


Figure 4.1 Map showing the mountain ranges in Peninsular Thailand and Malaysia. Source: Grismer *et al.* (2011).

4.1.2. Morphologically similar species

Morphological variation within *U. longifolium* has led to three varieties being described based upon hair density and hair length on both the stem and leaves (Hooker, 1880; Pitard, 1923; Craib, 1932). These include *U. longifolium* var. *longifolium* (Wight) Hook.f., *U. longifolium* var. *pilosum* Craib and *U. longifolium* var. *annamense* Pierre ex Pitard. After examining type specimens, differences of stipule hair density and hair pattern between *U. longifolium* var. *longifolium* and *U. longifolium* var. *pilosum* were found. *U. longifolium* var. *longifolium* has densely appressed hairs rather than the scattered erect hairs found in *U. longifolium* var. *pilosum*. The case for *U. longifolium* var. *annamense* is more complex, as the first publication includes five specimens that can be classified into three categories based upon morphological variation. The first category is comprised of one specimen collected by Pierre number 1840 (in Bien Hoa, south-eastern Vietnam) which contains similar morphological characters to the *U. longifolium* group. However, it differs from the other two *U. longifolium* varieties by the pattern of hairs on the stipule being shorter soft hairs (observed under magnification). Thus, the first category likely represents *U. longifolium* var. *annamense*. The second and third categories can be identified as *U. schmidtii* C.B.Clarke (collector — Pierre 1251 collected from Thpong, Cambodia) and

U. chinense Merr. & Chun (collector — Eberhardt 3867 (collected from Thai-Nguyen, north-eastern Vietnam), Chevalier 38709 and Lecomte & Finet 713 (both collected from Hon Ba, south-central coast of Vietnam)). The ecoregions of Vietnam used above follow Phuong *et al.* (2012). Furthermore, there has been a species named *U. talangense* Craib that is morphologically similar to *U. longifolium,* published in Contributions to the Flora of Siam (Craib, 1931). The diagnostic character provided for *U. talangense* is based upon the lower number of secondary leaf veins compared with *U. longifolium* (Craib, 1931). However, after observing specimens, there is overlap of the secondary leaf vein number. The only character that could be used to separate *U. talangese* and *U. longifolium* is hair density and pattern on stipules, where *U. talangense* has shorter appressed hairs compared to *U. longifolium* varieties (pers. comm.).

4.2. Aims

This chapter provides a taxonomic revision of *Urophyllum* in Thailand, as a brief summary of the genus (synopsis). The chapter aims to:

1. Provide the key morphological characters that are important to identify *Urophyllum* species.
2. Produce an identification key to the species found in the study area.
3. Discuss how the species differ from morphologically similar species or species within the same distribution range, and how the molecular data (Chapter 3) support the hypotheses.

4.3. Materials and methods

4.3.1. Plant materials

This taxonomic study of *Urophyllum* in Thailand was based primarily on the examination and measurements of morphological characters from herbarium specimens in AAU, ABD, BKF, C, FU, K, K-W, SING and newly collected specimens from fieldwork expeditions in Thailand during 2017–2019 (Appendix A). Herbarium records from BK, BM, CMUB, E, KKU, PSU and QBG, and online digitised records from P, Naturalis Biodiversity Centre BioPortal (specimens of AMD, L, U and WAG), and JSTOR Global Plants (https://plants.jstor.org/) were also examined. The acronym K-W refers to the Herbarium of East India Company (also known as Wallich Herbarium) where some of the plant collections of Nathaniel Wallich are kept separate from the main Kew collection. Other herbarium acronyms follow Index Herbariorum (Thiers, 2020). Leaf vein morphological terms follow Hickey (1979) and the Leaf Architecture Work Group (1999). Other morphological terms are from Beentje (2016). The descriptions provided here are based upon both literature and personal observations expanding to include specimens distributed outside the study area (Cambodia, Myanmar, Peninsular Malaysia and Singapore). The description of additional characters is based upon personal observation unless stated otherwise.

4.3.2. Conservation status assessment

Locality data was gathered from specimen labels, and coordinates used when available. Specimens lacking coordinates, were georeferenced using a point-radius method (Wieczorek, Guo and Hijmans, 2004) by obtaining coordinates of the main areas from GeoNames (https://www.geonames.org/). If offset data are provided on the label, the data was used to adjust coordinates on Google Earth Pro v7.3.3.7786. The conservation assessments were estimated following the IUCN Red list categories and criteria v3.1 (IUCN, 2012b) and IUCN guidelines (IUCN Standards and Petitions Committee, 2019). For species that have a distribution range that includes areas in Borneo and Indonesia (e.g., *Urophyllum blumeanum* (Wight) Hook.f., *U. hirsutum* (Wight) Hook.f., *U. macrophyllum* (Blume) Korth. and *U. streptopodium* Wall. ex Hook.f.), the regional conservation assessments were undertaken (IUCN, 2012a) for the study area and the neighbouring regions: Peninsular Malaysia and Singapore. The Extent Of Occurrence (EOO) and Area Of

Occupancy (AOO) were calculated using the GeoCAT software (Bachman *et al.*, 2011) with the cell size set to 2×2 km^2. Google Earth Pro v7.3.3.7786 was used to view the quality of habitat at different time periods (from 2000 to 2020), such as the changes in the amount of existing forest compared to expansion of urban and cultivation areas. Protected areas and IUCN Management Categories were referenced using the World Database on Protected Areas (UNEP-WCMC & IUCN, 2020). As the occurrence of species used in this study were based upon herbarium specimens and the fieldwork from 2017–2019, they could not be used to estimate population size and trends as required for criteria A and C, as well as quantitative analysis of population viability (criteria E); thus, only the criteria B and D have been applied. Point occurrence maps were generated on R (R Core Team, 2019).

4.3.3. Species concepts and groupings

The findings in Chapter 2, using supervised machine learning with the manual ensemble method revealed that morphological characters can be used to distinguish 10 *Urophyllum* species with 100% accuracy. Within these, five species were found in Thailand: *U. glabrum*, *U. hirsutum*, *U. longipes* Craib, *U. streptopodium* and *U. villosum*. From the findings in Chapter 3, these species were resolved to their own clade in phylogenetic analyses either in one dataset plastid (e.g., *U. longipes*) or nrDNA (e.g., *U. glabrum*) or both datasets for some species (e.g., *U. hirsutum*, *U. streptopodium* and *U. villosum*). The remaining five taxa with 100% accuracy are found in Vietnam and China (*U. argenteum* Pit., *U. chinense* Merr. & Chun, *U. lecomtei* Pit., *U.* sp.1 (*U. chinense* subsp. *latistipulum* sp. nov.), *U.* sp.2 (*U. bidoupense* sp. nov.) and *U.* sp.3 (*U. pseudoschmidtii* sp. nov.), the publication of these taxa is included in Chapter 5.

There were three taxa where the accuracy from the machine learning with the manual ensemble method was less than 100% including *U. crassum* Craib, *U. longifolium* var. *pilosum* (*U. longifolium* f.B in Chapters 2&3) and *U. talangense* (or *U. longifolium* f.C in Chapters 2&3). Only *U. longifolium* var. *pilosum* and *U. talangense* were unresolved in the plastid and nrDNA trees (Chapter 3). In the case of *U. crassum*, all three individuals formed a species clade for both data trees, and the species is sister to the *U. villosum* clade (Chapter 3). The taxonomic revision of *Urophyllum* in Thailand is based upon both the

phylogenetic species concept (*sensu* Mishler and Theriot (2000) with the broader definition of monophyly following Baum and Smith (2012)) and morphological species concept for species and lower ranks such as subspecies, using criteria as follows: 1) if taxa of the same group form a clade in either plastid or nrDNA trees or both, they were assigned to a species; 2) where taxa of different groups formed a clade in the trees they were treated in the following three ways: 2.1) If their recognised characters are influenced by habitat (phenotypic plasticity) such as leaf size, hair density, angle of hairs and hair length, despite other characters being identical, these taxa were assigned to the same species; 2.2) If they are morphologically distinct, they are assigned to different species; 2.3) If they are morphologically similar and: their distribution (data known to date) do not overlap, each taxon was resolved to a subspecies. This provided a robust framework for a working species concept for *Urophyllum* using the data available.

4.3.4. Key morphological characters

Stipule folding morphology

Stipule features are very useful characters for species identification in *Urophyllum* and can be used to group *Urophyllum* species as an early step. There are two patterns of stipule morphology - either folded and not appressed to the stem or appressed to the stem and not folded. A folded stipule is folded longitudinally toward the adaxial side, and they usually are not appressed to the stem as found in *U. blumeanum*, *U. crassum*, *U. longifolium* and *U. longipes* (Figure 4.2A–C). The appressed but not folded longitudinally stipules can be found in the other seven species found in Thailand with the examples shown in Figure 4.2D–F.

Figure 4.2 Stipules of *Urophyllum* taxa. **A–C** folded toward adaxial side. **A** *U. crassum*. **B** *U. longifolium*. **C** *U. longipes*. **D–F** appressed to the stem but not folded. **D** *U. glabrum*. **E** *U. streptopodium*. **F** *U. villosum*. Scale bar = 5 mm.

Leaf venation and domatia

Secondary venation and looping patterns, tertiary vein spacing, the divergence angle relative to the midrib, and the presence of domatia are also key characters to recognise *Urophyllum* species. There are two patterns of secondary vein looping found in the genus in Thailand: 1) festooned brochidodromous, where the secondary veins connect to one another in a series of small loops (Figure 4.3A–B), this type is commonly found in many *Urophyllum* spp. including *U. crassum*, *U. schmidtii* and *U. trifurcum*; and 2) weakly brochidodromous, where the secondary veins are inconspicuously connected to one another in a series of small loops, the connecting veins are similar in size to tertiary veins (Figure 4.3C–D), this type can be found in *U. blumeanum*, *U. glabrum* and *U. streptopodium*.

Tertiary vein spacing and divergence angle relative to the midrib can be divided into two groups: 1) closely spaced between veins, usually <2 mm apart and the veins perpendicular or subperpendicular to the midrib (angle c. 90°) (Figure 4.4B), this is found in

U. blumeanum and *U. streptopodium*; and 2) well spaced tertiary veins (>2 mm) and obtuse to the midrib (angle >90°) (Figure 4.4D), which is commonly found in the remaining species.

Domatia are found in the axils at the branching point between the midrib and secondary veins on the abaxial leaf surface. They can be found in form of either pocket shaped domatia which are glabrous (e.g., *U. longifolium*, *U. schmidtii*) or hairy (*U. glabrum* and *U. villosum*) (Figure 4.4E), or as a group of dense hairs only at the axil (*U. hirsutum*). Domatia are inconspicuous when the midrib is densely hairy, as found in *U. hirsutum* and *U. villosum*.

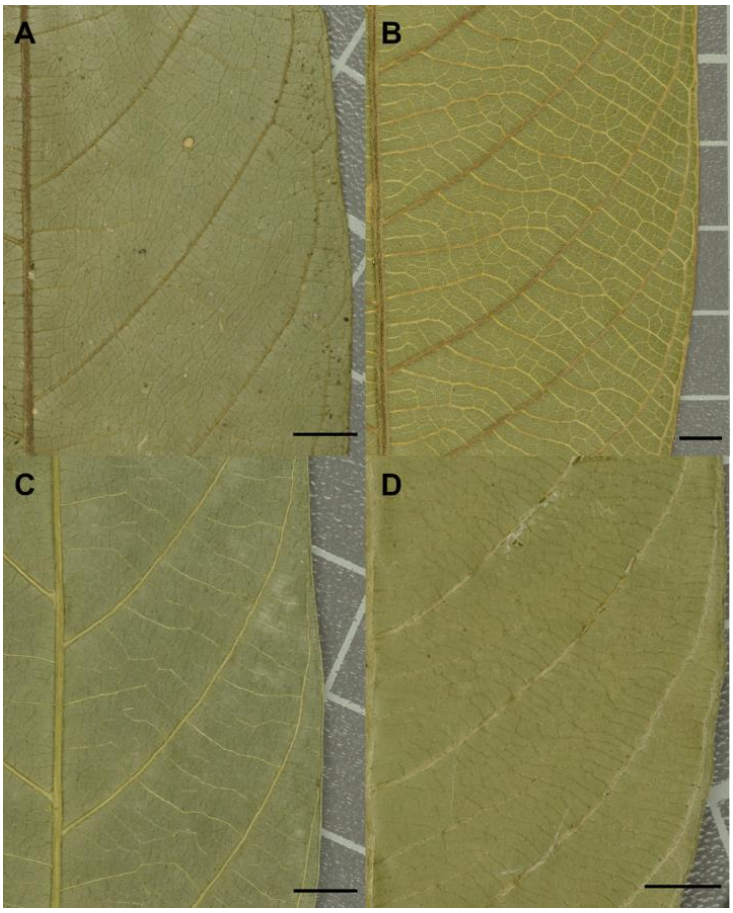


Figure 4.3 Secondary venation and loop patterns. **A–B** festooned brochidodromous in *U. longifolium* (**A**) and *U. villosum* (**B**). **C–D** weakly brochidodromous in *U. glabrum* (**C**) and *U. streptopodium* (**D**). Scale bar = 5 mm.

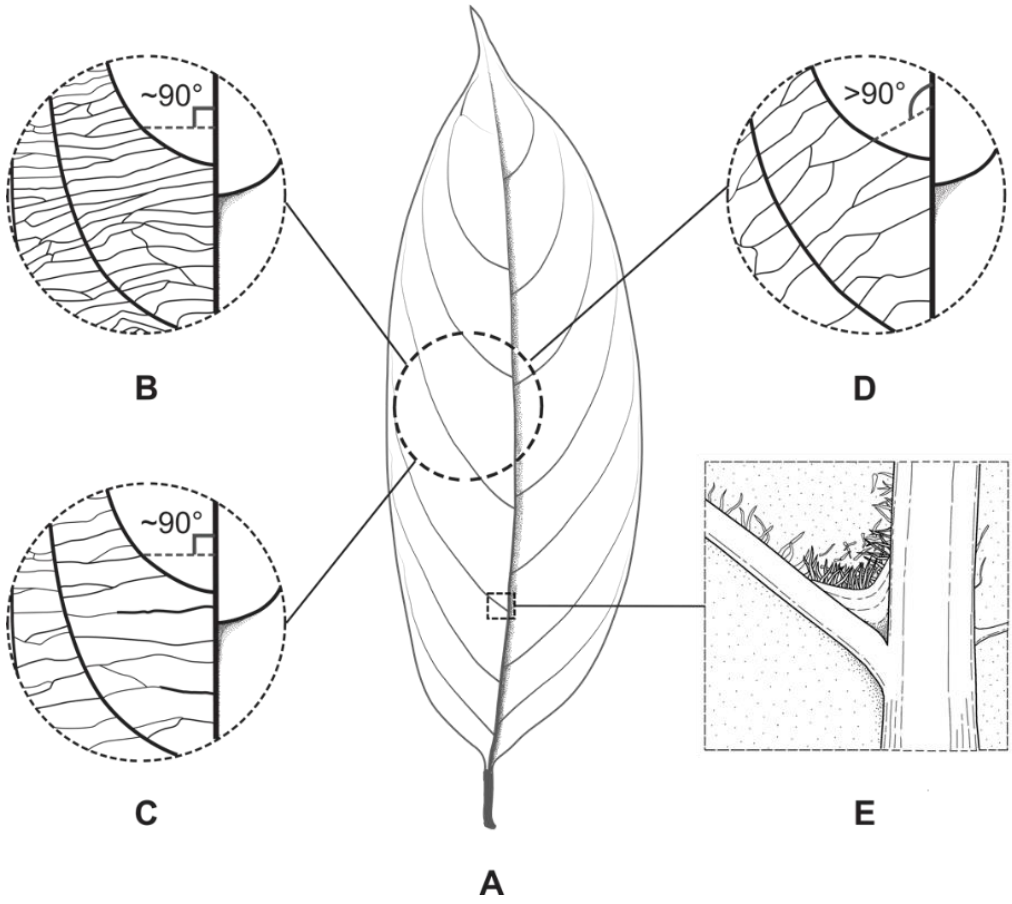


Figure 4.4 Divergence angle of tertiary veins relative to the midrib at the central portion of a *Urophyllum* leaf (**A**) and domatia (**E**). **B** = angle perpendicular with closely spaced tertiary veins (<2 mm space); **C** = perpendicular angle with well spaced tertiary veins (>2 mm space); **D** angle obtuse with well spaced. A densely hairy pocket shaped domatium at the secondary vein axil to the midrib in *U. glabrum* (**E**). Illustration is own work.

Inflorescences

Urophyllum inflorescences are mostly axillary cymose or umbellate (rarely solitary flower). The main category is consistent within a species; however, its subcategory is very variable especially the length of peduncle. For example, *U. longifolium* can be constantly found with compound cymose inflorescences (Figure 4.5D–E) that is either sessile or pedunculate, as well as *U. streptopodium* and *U. hirsutum* where the inflorescences can be both pedunculate and sessile umbellate (Figure 4.5F&H). Only *U. blumeanum* has an inflorescence subcategory that is consistent being two-tiered umbellate (Figure 4.5G). The simple cymose inflorescence (Figure 4.5B) is usually found in *U. longipes* with the two-tiered cymose arrangement (Figure 4.5C) found rarely in some populations.

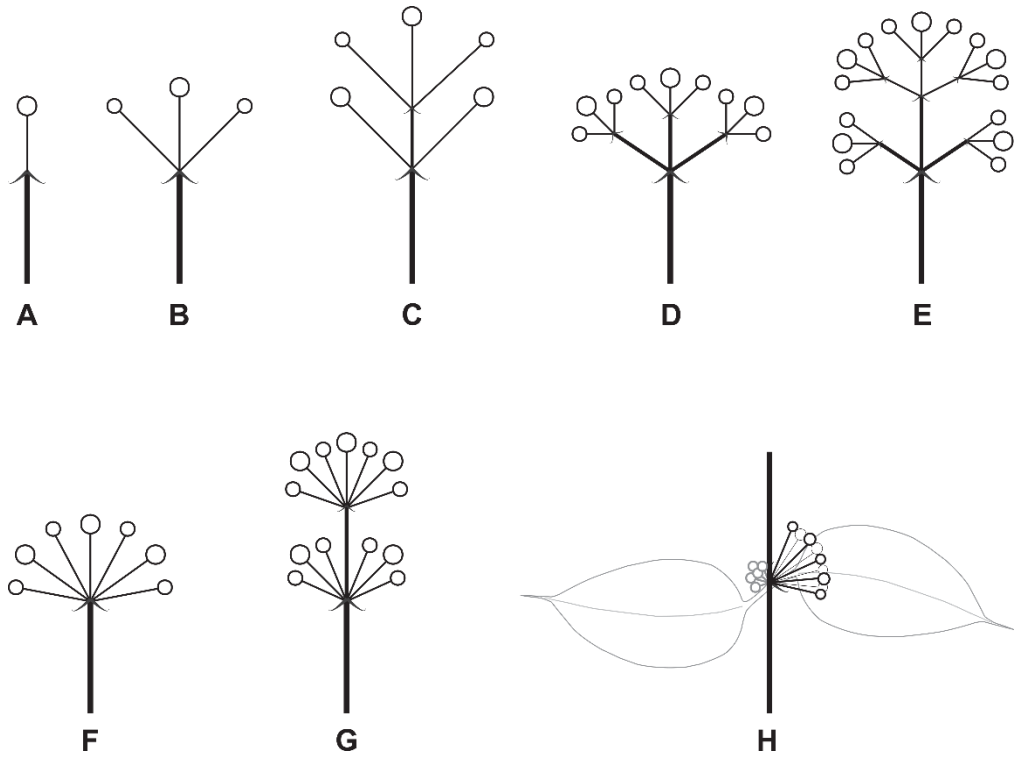


Figure 4.5 Inflorescences of *Urophyllum*. **A** solitary flower. **B–E** cymose: **B** simple cymose; **C** two-tiered cymose. **D–E** compound cymose: **D** trichotomous cymose; E two-tiered trichotomous cymose. **F–H** Umbellate: **F** pedunculate umbellate; **G** two-tiered umbellate. **H** sessile umbellate. Illustration is own work.

Hairs and membrane at the base of the corolla lobes

Some species of *Urophyllum* in Thailand can be recognised using hair density and the membrane shape at the base of the corolla lobes. Hairs at the base of the corolla lobes are sparse and countable in *U. blumeanum*, whereas they are dense and numerous in other species. The presence of membranes at the base of the corolla lobes and its shape can be used to identify *U. glabrum* where it is tubular in shape connected to all the lobes (Figure 4.6A). In *U. crassum*, *U. trifurcum* and *U. villosum*, the membranes are triangular and attached to each corolla lobe (Figure 4.6B). The membranes are absent in the other species found in Thailand.

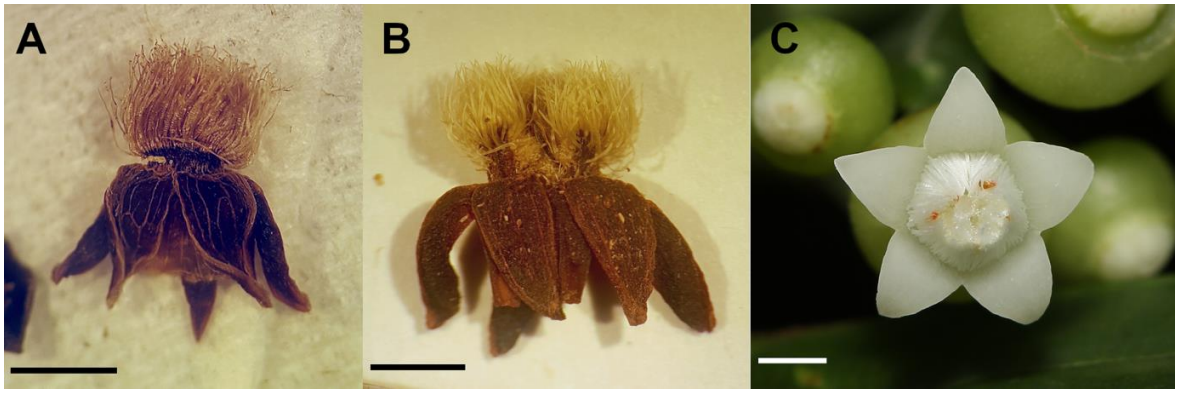


Figure 4.6 Position of hairs at the base of the corolla lobes. **A**) Hairs on a tubular membrane (*U. glabrum*); **B**) Hairs on triangular membranes (*U. trifurcum*); **C**) Hairs directly on corolla lobes (*U. longifolium*). Scale bar 2 mm.

4.4. Taxonomic treatment

Urophyllum Wall. in Roxburgh (1824: 184). Type species: *Urophyllum villosum* Wall. in Roxburgh (1824: 185) (lectotype selected by Wong *et al.* (2019)).

Shrubs or treelets to 5 m, young branches ridged, tetragonous, branching usually opposite. Indumentum of simple, short to long hairs where present, plants glabrous to densely appressed hairy especially on young branches. Stipules interpetiolar, narrowly to broadly lanceolate, caducous to persistent for 3–4 nodes, subglabrous to densely appressed hairy, colleters present from base to the middle of adaxial side. Leaves simple, opposite-distichous, chartaceous to coriaceous, ovate to oblong to elliptic, sometimes obovate, apex attenuate to acuminate to caudate, base cuneate, margin entire; secondary veins pinnate, conspicuously brochidodromous or weakly to festooned brochidodromous, primary vein impressed adaxially and prominent abaxially, adaxial lamina usually glabrous to sparsely hairy on midrib and secondary veins, abaxial veins subglabrous to densely hairy. Petiole canaliculate, always present, subglabrous to hairy. Inflorescence axillary (rarely cauliflorous), distichous, trichotomous simple to compound cymose or umbellate, peduncle and pedicel usually hairy. Flowers unisexual (plants dioecious), usually 5-merous (rarely 4 or 6-merous); calyx entire to subtruncate, finely toothed or lobed; corolla fused at base, corolla lobes 4, 5 or 6, valvate, glabrous or sparsely to densely hairy abaxially, base

of the corolla lobes sparsely to densely hairy adaxially, hairs white to yellow. Staminate flowers: stamens 4, 5 or 6, alternate with petals, filaments short attached to petal tube, anthers dorsifixed, introrse, 2-celled; stigma reduced borne directly from disk or on very short style; nectariferous disk present, circular to conical; ovary vestigial, ovules numerous or absent. Pistillate flowers: staminodes present or absent, when present with the same arrangement, shape and size as staminate flowers but lacking pollen; ovary inferior, globose or subglobose, locules usually 5, axile placentation; disk present, same shape as in staminate flowers, smooth, rugose when dried; stigma linear or subglobular, usually 5-lobed, subglabrous or sparsely hairy. Fruits globose or subglobose, baccate, unripe fruits light green to yellowish green, mature fruits orange ripening black, with persistent calyx and disk; seed numerous, ovate to elliptic, surface alveolate and orange-brown to dark brown *in sicco*, enclosed in a fleshy orange aril *in vivo*, disappearing when dried.

DISTRIBUTION. In tropical Asia including Sri Lanka, India (Kolkata, Andaman and Nicobar Islands), southern China, through to Malesia and reaching Papua New Guinea.

Key to *Urophyllum* species in Thailand.

1. Secondary vein number 5–7 pairs, rarely 8–10; divergence angle of tertiary veins relative to the midrib at lamina mid-point perpendicular or sub-perpendicular (angle c. 90°) with tertiary veins closely spaced (distance between veins ≤2 mm) (Figure 4.4B); 2

   Secondary vein number >8 pairs; divergence angle of tertiary veins relative to primary vein at lamina mid-point, obtuse (angle >90°) with tertiary veins relatively well spaced apart (distance between veins >2 mm) (Figure 4.4D) or mixed between closely and well spaced.....3

2. (1) Stipules not constricted, hair density similar all over, hairs light brown *in sicco*; petiole densely hairy on canalicular ridge, glabrous to subglabrous elsewhere; inflorescence two-tiered umbellate (Figure 4.5G); adaxial corolla lobes with sparse hairs at the base of the lobes (<10) ..... **1. *Urophyllum blumeanum***

   Stipules constricted above the base, densely hairy at base, sparsely toward apex revealing a contrast pale yellow at the base then dark brown toward apex *in sicco*; petiole densely hairy all over; inflorescence sessile to pedunculated umbellate with very short peduncle (<3 mm) (Figure 4.4F or H); adaxial corolla lobes with numerous hairs at the base of the lobes (>12 or difficult to count)..... .....**9. *Urophyllum streptopodium***

3. (1) Stipules folded longitudinally toward the adaxial side (Figure 4.2A–C).....4

   Stipules appressed to the stem but not folded or only top half folded (Figure 4.1 A–B).....7

4. (3) Inflorescence compound cymose, pedunculate or sessile umbellate (Figure 4.4D, E, F or H) 5

   Inflorescence pedunculate, simple cymose, rarely two-tiered cymose (Figure 4.4B or C) .....**6. *Urophyllum longipes***

5. (4) Inflorescence compound cymose; corolla white.....**5. *Urophyllum longifolium***

   Inflorescence pedunculate or sessile umbellate; corolla green or white/yellow at base then green toward apex.....6

6. (5) Abaxial leaf subglabrous or sparsely hairy on midrib to tertiary veins; flower with short pedicel (<4 mm) or inconspicuous and flowers appearing subsessile; calyx

usually subtruncate or toothed; abaxial corolla surface scaly ..... **2. *Urophyllum crassum***

Abaxial leaf densely hairy all over; flower with conspicuous pedicel (>5 mm); calyx lobed; abaxial corolla surface glabrous..... **7. *Urophyllum macrophyllum***

7. (3) Corolla glabrous all over abaxially..... 8

Corolla hairy all over abaxially 9

8. (7) Leaf secondary veins weakly brochidodromous; densely hairy pocket domatia present at every axil at the branching point between the midrib and secondary veins on the abaxial leaf surface (Figure 4.3E); calyx toothed; corolla green; southern Thailand..... **3. *Urophyllum glabrum***

Leaf secondary veins festooned brochidodromous; glabrous pocket domatia present, but not in every axil at the branching point between the midrib and secondary veins on the abaxial leaf surface; calyx lobed; corolla white; eastern Thailand..... **8. *Urophyllum schmidtii***

9. (8) Adaxial corolla with triangular membranes at the base of the lobes (Figure 4.6B).. 10

Adaxial corolla lacking membranes at the base of the lobes (Figure 4.6C)..... .....**4. *Urophyllum hirsutum***

10. (9) Abaxial leaf glabrous or hairy only on the midrib to tertiary veins, veins reddish brown *in sicco*; calyx entire to toothed..... **10. *Urophyllum trifurcum***

Abaxial leaf hairy, veins yellowish brown *in sicco*; calyx lobed..... ..... **11. *Urophyllum villosum***

1. ***Urophyllum blumeanum*** (Wight) Hook.f. (Hooker 1880: 99). *Axanthes blumeana* Wight (1847: 145). Type: [Malaysia, Peninsular Malaysia], Malacca, s.a., *Griffith* s.n. (♂) (lectotype, designated by Wong *et al.*, 2019: K! [K000740827]).

DISTRIBUTION. Southern Thailand (from Trang Province southward), Peninsular Malaysia (except Perlis and Kedah states), Singapore, Borneo, and Sumatra (Figure 4.7).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: JOHOR, Kota Tinggi, Bandar Tenggara, Linggiu Forest Reserve [1°54'26''N, 103°40'39''E], 65 m, 23 Jul 1991, *Lesmy FRI 35919* (♀) (K!); Segamat, Eastern boundary of Segamat Wildlife Reserve [2°35'44.36"N, 102°56'6.45"E], 6 Feb 1970, *Loh FRI 17142* (♀) (K!); Mersing, 29 Jul 1992, *Thomas & Teo KL 4142* (♀) (P-photo!); **KELANTAN**, Gua Musang District, near Terengganu border, Ulu Lebir Kechil, steep hillside [5°00'30''N, 102°28'51''E], 800 f [c. 240 m], 18 Sep 1967, *Whitmore FRI 4421* (♀) (K!); ibid., Kuala Mersing, Sg. Brok, Ulu Kelantan [4°38'N, 101°36'E], 450 f [c. 140 m], 6 Dec 1967, *Ng FRI 5397* (♂) (K!); ibid., Sungai Chalil, Sungai Lebir [5°09'51''N, 102°19'15''E], 5 Jul 1935, *Henderson SFN 29543* (♂) (BKF!, K!); ibid., 2 miles E. Kuala Aring [5°01'N, 102°23'E], 13 Sep 1967, *Cockburn FRI 7079* (♀) (K!); ibid., Ulu Kelantan, Relai FR, ridge top [5°08'N, 102°12'E], 18 Oct 1967, *Cockburn FRI 7246* (♀) (K!); Jeli District, Sungai Yong [Long] off Sungai Pergau, Jeli [5°42'55.53"N, 101°47'1.68"E], 22 Sep 1986, *Latiff et al. ALM 1612* (♀) (PSU!); Kuala Krai District, Machang, Sg. Durian FR [5°35'N, 102°20'E], 7 Jul 1987, *Damanhuri & Khairuddin FRI 35980* (♀) (K!); **KUALA LUMPUR**, Wild Hill Reserve, 10 Oct 1922, *Guard 8528* (♀) (K!); s.loc., 3 May 1915, *Ridley s.n.* (♀) (K!); Damansara Road, Jan 1921, *Ridley s.n.* (♀) (K!); Ampang Intake Catchment Reserve [3°11'N, 101°47'E], 14 May 1981, *Wong FRI 32223* (♀) (K!); **MALACCA**, s.loc., 18 Jan 1905, *Griffith s.n.* (♂) (K! [K000740831]); s.loc., 21 Jan 1908, *Griffith s.n.* (♂) (K! [K000740828]); s.loc., s.a., *Griffith s.n.* (♂) (K! [K000740830]); **PAHANG**, Jerantut District, Sg. Tekam Fr, Jengka [4°01'N, 102°33'E], 8 Jul 1979, *Chan FRI 23935* (♀) (K!); ibid., Ulu Sungai Sat near Kuala Kelepah [Kelapah] [4°32'54.31"N, 102°35'16.43"E], 9 Jul 1970, *Shah 1745* (♀) (L-photo!); Raub District [3°53'N, 101°50'E], 20 Aug 1930, *s.coll. 22529* (♀) (K!); ibid., 54th mile gap Rd., hillside, meranti Bukit plot [3°42'3.87"N, 101°45'16.53"E], 3000 f [c. 910 m], 7 Oct 1980, *Kochummen FRI 29109* (♀) (K!); **PENANG**, Timur Laut District, Penang hill [5°26'N, 100°16'E], Apr 1882, *Hullett 165* (♀) (K!); ibid., Penang Hill toward the submit [5°26'N, 100°16'E], 2000 f [c. 610 m], 5 Mar 1966, *Stone 6330* (♀) (K!); s.loc., Jun 1885,

Luetis 261 (♂) (K!); Gunung Hill, 13 Mar 1905, *Ridley 10251* (♂) (K!); s.loc., 26 Dec 1904, *Wallich 6218* (sterile) (K-W!); s.loc., 24 Feb 1905, *Wallich 9067 [specimen on the left]* (sterile) (K-W!); **PERAK**, Taman Negara, Ulu Sat near K. Kelapah Ranger post. [4°32'54.31"N, 102°35'16.43"E], 350 f [c. 110 m], 7 Sep 1970, *Whitmore FRI 15213* (♂) (K!); Batang Padang District, Slim Hill FR [4°00'N, 101°27'E], 9 Mar 1966, *Whitmore FRI 0775* (♀) (K!); Kinta District, Goping [Gopeng], Oct 1880, *Dr. King's collector 770* (♀) (K!); ibid., Kledang Saiong FR, Batu Gajah [4°38'N, 101°01'E], 600 f [c. 180 m], 4 Apr 1968, *Ng FRI 6036* (♀) (K!); Kuala Kangsar District, Near logging road running NW/SE from Kg. Ayer into G. Bubu Massif [4°37'37.60"N, 100°45'33.05"E], 1300 f [c. 400 m], 20 Feb 1970, *Everett FRI 13917* (♀) (K!); Larut Matang District, Taiping, 2.5 miles up road, road verge, 1500 f [c. 460 m], 30 Oct 1969, *Everett FRI 13587* (♀) (K!); ibid., 4th mile, Maxwell's Hill [4°51'N, 100°47'E], 12 May 1965, *Shah & Sidek MS 1112* (♀) (K!, L-photo!); ibid., Larut, Buki Maxwell [Bukit Larut] [4°51'N, 100°47'E], 400 m, 8 Jun 1983, *Stone 15521* (♂) (AAU!); ibid., by Tea garden [4°52'2.12"N, 100°46'38.58"E], 2100 f [c. 640 m], 18 Nov 1969, *Whitmore FRI 12882* (♀) (K!); s.loc., Jun 1886, *Dr. King's collector 10103* (♀) (K!, P-photo!); Ulu Bubong, 400–600 f [120–180 m], Jan 1884, *Dr. King's collector 10311* (♂) (K!); **SELANGOR**, Gombak District, Uluh Gombak [03°18'00"N, 101°47'00"E], Nov 1986, *David 273* (♀) (P-photo!); ibid., Ulu Gombak FR, roadside [3°17'1.03"N, 101°43'55.44"E], 18 Jun 1988, *Kamarudin FRI 34505* (♀) (K!); ibid. [3°17'N, 101°45'E], 10 Apr 1968, *Kochummen FRI 2553* (♀) (K!); ibid., Genting Bidai [3°18'N, 101°49'E], May 1896, *Ridley 7443* (♀) (K!); ibid., Ampang FR [3°11'N, 101°47'E], 11 Jun 1978, *Suppiah FRI 28186* (♂) (K!); Hulu Langat District, Bukit Tangkol, K. Pansom [3°13'N, 101°53'E], 2 Jan 1959, *anak Umbai KL 1387* (♀) (K!); ibid., Forest near dam at K. Pansom [3°13'N, 101°53'E], 15 May 1959, *anak Umbai KL 1489* (♀) (L-photo!); Bukit Kuyu Kapun, 13 Mar 1905, *Ridley 10587* (♀) (K!); **TERENGGANU**, Dungun District, Eastern face of G. Mandi Angin, Shale, valley bottom [4°41'N, 102°51'E], 2000 f [c. 610 m], 13 Jul 1968, *Whitmore FRI 12072* (♂) (K!); **SINGAPORE.** Bukit Timah [1°23'13"N, 103°48'3"E], 19 Jun 1938, *Corner SING 34997* (♂) (K!); ibid., 3 Dec 1964, *Hardial 118* (♂) (K!, L-photo!); s.loc., 19 Jan 1905, *Lobb 279* (♀) (K!); Lawn V., Botanical Garden Singapore [1°18'38.53"N, 103°48'54.29"E], 12 May 1929, *Nai s.n.* (♂) (K!); Bukit Junat, 4 Mar 1905, *Ridley 4910* (♀) (K!); Hollam rd., 13 Mar 1905, *Ridley 10371* (♂) (K!); Main Road, Bukit Timah Nature Reserve [1°23'13"N, 103°48'3"E], 21 Jul 1970, *Hamzah H 6* (♀) (K!, L-photo!); s.loc., 3 Jan 1905, *Walker 268* (♀) (K!); s.loc., s.a., *Wallich 8318* (♂) (E!); s.loc., s.a., *Wallich 8318 [specimen on the right and*

middle] (♂) (P-photo!); **THAILAND. NARATHIWAT**, Waeng, Bawae, Beside the road [5°48'5.4"N, 101°50'28.5"E], 94 m, 11 May 2019, *Yooprasert et al. 199* (sterile) (BKF!, K!); **PHATTHALUNG**, Kao Soi dao [7°20'N, 99°54'E], 100 m, 29 Apr 1930, *Kerr 19232* (♂) (BK!, BM!, E!, K!); **SONGKHLA**, Hat Yai, Ton Nga Chang WS [6°56'47"N, 100°13'45"E], 760 m, 13 Apr 1998, *Puangpen N 526* (♂) (QBG!); **TRANG**, Khao Chong [Kachong] [7°31'05"N, 99°48'14"E], 7 Mar 1976, *Chermsirivatthana 2175* (♂) (BK!); ibid., Khao Pappa [7°25'N, 99°35'E], 150 m, 13 Mar 1974, *Larsen & Larsen 33265* (♂) (AAU!, BKF!, K!).

HABITAT. Lowland to montane evergreen forest, sometimes mixed with Bamboo. Usually found around riverbank, by streams or on steep hills or slopes; elev. 60–950 m.

REGIONAL CONSERVATION STATUS. Near Threatened (NT). The estimated AOO (108 km^2) is rated as Endangered (EN); however, the estimated EOO is more than 95,000 km^2 and the species has been collected from 19 localities (10 localities within protected areas in Thailand and Peninsular Malaysia), this means the species does not fulfil the criteria for the threatened categories. Specimens of *Urophyllum blumeanum* from Peninsular Malaysia collected before 1990 (all but two specimens) were within forest areas that have since been cleared for agricultural land (especially oil palm and durian plantations). Forest clearing has occurred within the past few years in some areas e.g., in the Gua Musang and Mersing districts (Google Earth imagery). Therefore, it is rated as Near Threatened at the regional level with concerns on quality of habitat and the recommendation that new collections in Peninsular Malaysia are required.

PHENOLOGY. Collected in flower and fruit from February to November.

NOTES. *Urophyllum blumeanum* can be recognised by its longitudinally folded stipules, the abaxial leaf surface is densely hairy with hairs appressed to the surface, tertiary veins are closely spaced (space <2 mm apart), and a two-tiered umbellate inflorescence.

This species is morphologically similar to *U. arboreum,* highlighted by *U. blumeanum* being synonymised to *U. arboreum* sensu Wong *et al.* (2019). They share a similar number of lateral veins (5–)6–8(–10) pairs, neatly arranged tertiary venation with the divergence angle relative to midrib ~90° and a two-tiered umbellate inflorescence (in some populations of *U. arboreum* e.g., in Java and Sumatra). However, they differ by the

hairiness of the abaxial leaf surface: glabrous in *U. arboreum* but densely minute hairs in *U. blumeanum*. Therefore, the species is morphologically distinctive and can be recognised as its own species, *U. blumeanum*. Although, Wong *et al.* (2019) discussed in the Notes to *U. arboreum* that the specimen of *Hamzah H 6*, which can be identified to *U. blumeanum* here, bears both male and female flowers on the same or different inflorescences, this is not the case for duplicates in K and L where the specimens only have female flowers and young fruits. Therefore, *U. blumeanum* is dioecious in this study.

The species is also similar to *U. streptopodium* in having leaves with dense, appressed hairs abaxially, and subperpendicular divergence angle of tertiary veins relative to the midrib, but it differs by the stipule morphology and inflorescence type. *Urophyllum streptopodium* has stipules that are densely hairy at base and sparsely hairy toward the apex, revealing a contrast pale yellow at the base then dark brown toward apex when dried (Figure 4.3E), and it has a sessile to pedunculated umbellate inflorescence, whereas *U. blumeanum* has stipules with one colour from light to dark brown, and a two-tiered umbellate inflorescence.

The phylogenetic trees from plastid and nrDNA datasets were incongruent in the position of *U. blumeanum* either as sister to the *U. hirsutum* and *U. longifolium*–*U. glabrum* group or to *U. memecyloides* S.Vidal, respectively (Chapter 3). However, morphologically *U. blumeanum* is more similar to *U. memecyloides* than to *U. hirsutum* in leaf characters and inflorescence type (divergence angle of tertiary veins relative to the midrib perpendicular or subperpendicular and two-tiered umbellate inflorescence; rather than obtuse angle of tertiary veins, and sessile or shortly pedunculate umbellate) (Chapter 3). They are unlikely to be confused as *U. memecyloides* has leaves submembranaceous, abaxial leaf surface subglabrous, tertiary veins on mature leaves dry wavy abaxially, and calyx subtruncate, whereas *U blumeanum* has leaves chartaceous, abaxial leaf surface densely hairy, tertiary veins dry smooth, and calyx lobed.

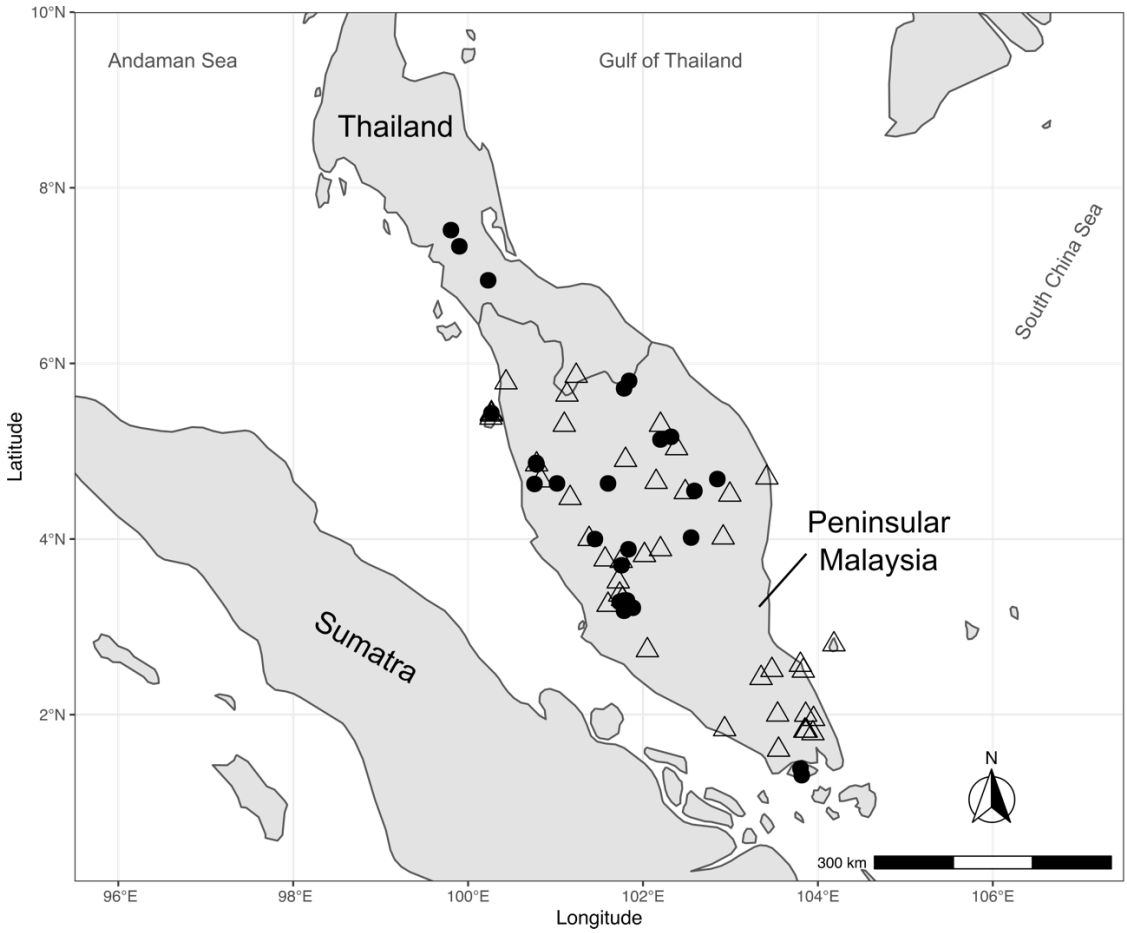


Figure 4.7 Occurrence map (partial distribution) of *Urophyllum blumeanum* (●) and *U. streptopodium* (△).

2. ***Urophyllum crassum*** Craib (1931: 445). Type: Thailand, Yala, Banang Sta (Bannang Sata) [6°17'N], 14 June 1930, *Kiah 24326* (♀) (lectotype, designated here K!; isolectotypes ABD!, BK!).

DISTRIBUTION. Peninsular Malaysia (Kelantan and Perak states) and southern Thailand (Nakhon Sri Thammarat, Narathiwat and Yala Provinces) (Figure 4.8).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: KELANTAN, Gua Musang District, Ulu Sungai Aring, near Kuala Tapah [4°51'N, 102°20'E], 21 Sep 1967, *Cockburn FRI 7152* (♀) (K!); ibid., Sungai Tapah [4°51'N, 102°20'E], 22 Sep 1967, *Cockburn FRI 7169* (♀) (K!); ibid., Ulu Kelantan, Relai FR, compartment 33, ridge top [5°08'N, 102°12'E], 17 Oct 1967, *Cockburn FRI 7222* (♀) (K!); Jeli District, Sungai Yong [Long] off Sungai Pergau, Jeli [5°42'55.53"N, 101°47'1.68"E], 22 Sep 1986, *Latiff et al. ALM 1608* (♀) (PSU!); **PERAK**, Hulu Perak District, K. Temengor, State Land [5°24'N, 101°18'E], 1100 f [c. 340 m], 20 Jul 1966, *Chelliah KEP 98681* (♀) (K!); ibid., Belum FR, Sg. Semiliang [5°43'N, 101°25'E], 290 m, 02 Jun 1998, *Chua et al. FRI 40687* (♂) (K!); ibid., Grik, Temenggor, Sg. Singor, Sg. Remei trail [5°31'20''N, 101°28'45''E], 260 m, 24 Sep 1993, *Noorsiha et al. FRI 39438* (♀) (K!); **THAILAND. NAKHON SRI THAMMARAT**, Nopphitam District, Khrung Ching Falls, trail to waterfall [8°43'14"N, 99°40'28"E], 270 m, 18 Jun 2017, *Yooprasert et al. NSK 137* (sterile) (BKF!, K!); ibid., [8°43'12"N, 99°40'31"E], 300 m, 18 Jun 2017, *Yooprasert et al. NSK 138* (♀) (BKF!, K!); ibid., [8°43'12"N, 99°40'31"E], 300 m, 18 Jun 2017, *Yooprasert et al. NSK 139* (♀) (BKF!, K!); ibid., *Yooprasert et al. NSK 140* (♀) (BKF!, K!); ibid., [8°42'40"N, 99°41'01"E], 277 m, 18 Jun 2017, *Yooprasert et al. NSK 145* (sterile) (BKF!, K!); ibid., [8°42'38"N, 99°41'18"E], 279 m, 18 Jun 2017, *Yooprasert et al. NSK 146* (♂) (BKF!, K!); ibid., *Yooprasert et al. NSK 147* (♂) (BKF!, K!); ibid., [8°42'38"N, 99°41'26"E], 275 m, 18 Jun 2017, *Yooprasert et al. NSK 148* (♂) (BKF!, K!); ibid., [8°42'38"N, 99°41' 32"E], 277 m, 18 Jun 2017, *Yooprasert et al. NSK 149* (♀) (BKF!, K!); ibid., [8°42'38"N, 99°41'33"E], 278 m, 18 Jun 2017, *Yooprasert et al. NSK 151* (♀) (BKF!, K!); **NARATHIWAT**, Bukit [6°10'N, 101°50'E], 300 m, 07 Jul 1923, *Kerr 7093* (♂) (BK!, BM!, K!); Sungai Padi District, Chatwarin fall, Sungai Padi [6°4' N, 101°52'E], 19 Oct 1970, *Charoenphol et al. 3985* (♀) (BKF!, K!); ibid., [6°6'09.8"N, 101°50'6"E], 100–150 m, 08 Oct 1991, *Larsen et al. 42210* (♀) (AAU!, BKF!); Waeng District, 200 m, 13 Jun 1970, *Smitinand 10927* (♂) (K!); ibid., Hala-Bala WS, rubber plantation in Hala-Bala [5°28'32"N, 101°30'0.9"E], 120 m, 22 Aug 2006, *Poopath et al. 8* (♀) (E!); ibid.,

Ban Bala, Hala-Bala WS, Research trail [5°47'55"N, 101°50'0.7"E], 115 m, 10 May 2019, *Yooprasert et al. 183* (sterile) (BKF!, K!); ibid., Bawae, Beside the road [5°48'6.9"N, 101°50'26.8"E], 96 m, 11 May 2019, *Yooprasert et al. 195* (♂) (BKF!, K!); **PATTANI** [Narathiwat, Pattani and Yala Provinces], s.loc., s.a., *Kerr 7993* (♂) (K!); **RANONG**, Nature trail, Huai Kraminj near headquarter, Khlong Naka, 150 m, 18 Jul 2000, *Chamchumroon vc 882* (♂) (BKF!); **YALA**, North side of Banglang Resevoir, Toh moh [6°4'N, 101°23'E], 200–250 m, 17 Jun 1992, *Larsen et al. 42955* (♂) (AAU!, BKF!, PSU!); Than To District, along Tomo river in Chulaphon Phatthana 7 area. [6°8'90"N, 101°38'E], 160 m, 20 May 2005, *Middleton et al. 3486* (♂) (BKF!, E!); ibid., Ban chulaphon Phattana 7, Bang Lang Dam [6°5'18"N, 101°22'50"E], 200 m, 31 Oct 2001, *Pooma et al. 3173* (♀) (BKF!); ibid., Maewat [Mae Wat], the route to nature trail, the Ecology permanent plot [6°00'15"N, 101°16'32"E], 260 m, 16 Oct 2017, *Poopath 1999* (♀) (BKF!); Betong District, relic forest along Pattani river [5°47'45.3"N, 101°10'14.8"E], 275 m, 01 Aug 2016, *Poopath et al. 1569* (♂) (BKF!); ibid., Maewat, nature trail to waterfall, along Than Roi Jai stream [6°5'19"N, 101°22'49"E], 180 m, 06 Aug 2016, *Poopath et al. 1674* (♂) (BKF!).

HABITAT. Tropical evergreen forest with dipterocarps, usually on slopes along streams; elev. 90–350 m.

CONSERVATION STATUS. Least Concern (LC). The species has been collected from nine localities in southern Thailand and northern Peninsular Malaysia, with seven populations within protected areas around the Thailand–Malaysia border and one collection from Khao Luang National Park, another protected area in Nakhon Sri Thammarat Province, southern Thailand. The remaining specimen was collected from Kuala Tapah in Gua Musang District, Kelantan State, Peninsular Malaysia. The risk of habitat loss in the last few years, due to the expansion of agricultural land, in Kuala Tapah, has raised concern to the species being threatened. Satellite images have shown evidence of forest clearing for the purpose of palm plantations, in the region where *U. crassum* has been previously recorded (Google Earth imagery). Shevade and Loboda (2019) highlighted that over 99% of forest loss in Malaysia from 1988-2012 was within 1 km of existing plantations. The proximity of oil palm plantations therefore demonstrates a high risk to the forest in Kuala Tapah. The loss or encroachment upon forests from agriculture has been widely reported (Wilson *et al.*, 2016; Corlett, 2019), and in recent decades Malaysia has one of the highest deforestation

rates in SE Asia (Hughes, 2017). Even with this threat to the habitat, the number of localities being less than 10 and the estimated AOO (68 km^2) under 2,000 km^2, the seven populations of *U. crassum* are found within protected areas in Thailand where the threat is low. The species is therefore rated as Least Concern; however, there is risk of habitat loss in Kelantan State, Peninsular Malaysia.

PHENOLOGY. Collected in flower from May to August, fruit from July to October.

NOTES. The diagnostic characters of *U. crassum* are its subglabrous appearance, folded stipules, leaves abaxially subglabrous to the naked eye, the veins on drying leaves are reddish brown, inflorescences sessile or subsessile umbellate, flowers with short pedicel (0–4 mm), calyx truncate to toothed, and green corolla scaly abaxially with triangular membranes at the adaxial base of the corolla lobes.

Urophyllum crasssum is similar to *U. glabrum* and *U. longifolium* in their subglabrous appearance and truncate to toothed calyx; they are also found in the same areas (Krung Ching Falls in Nakhon Sri Thammarat Province and Bang Lang Dam in Yala Province). However, *U. crassum* differs from the other two species in its longitudinally folded stipules, inflorescence type, abaxial corolla hair distribution and the shape of membrane at base of the corolla lobes as described in the previous paragraph. *Urophyllum glabrum* has hairy appressed stipules that are not folded, pedunculate umbellate to dichotomous cymose inflorescence, abaxial corolla glabrous and a tubular membrane at the base of corolla lobes. The species also has hairy pocket domatia at the points of braching between midrib and secondary veins which are absent in *U. crassum*. In the case of *U. longifolium*, it differs by having a compound cymose inflorescence, white corolla and glabrous abaxially, and it lacks a membrane at the base of the corolla lobes.

The phylogenetic trees from plastid and nrDNA datasets revealed that *U. crassum* is the sister group to the *U. villosum* clade (Chapter 3). The only character shared by both species is the triangular membrane at the base of the corolla lobes, otherwise they are morphologically distinctive. *Urophyllum villosum* is densely hairy in appearance, its leaves dried yellowish brown and are densely hairy abaxially, calyx lobed, and corolla hairy abaxially.

The lectotype designated here, *Kiah 24326* deposited at K, as Craib's biography in Taxonomic Literature II (Stafleu and Mennega, 1997) show that the type material is deposited in K and WRSL, and the Bulletin of Miscellaneous Information, Kew (Kerr, 1933) details Craib frequently working at K.

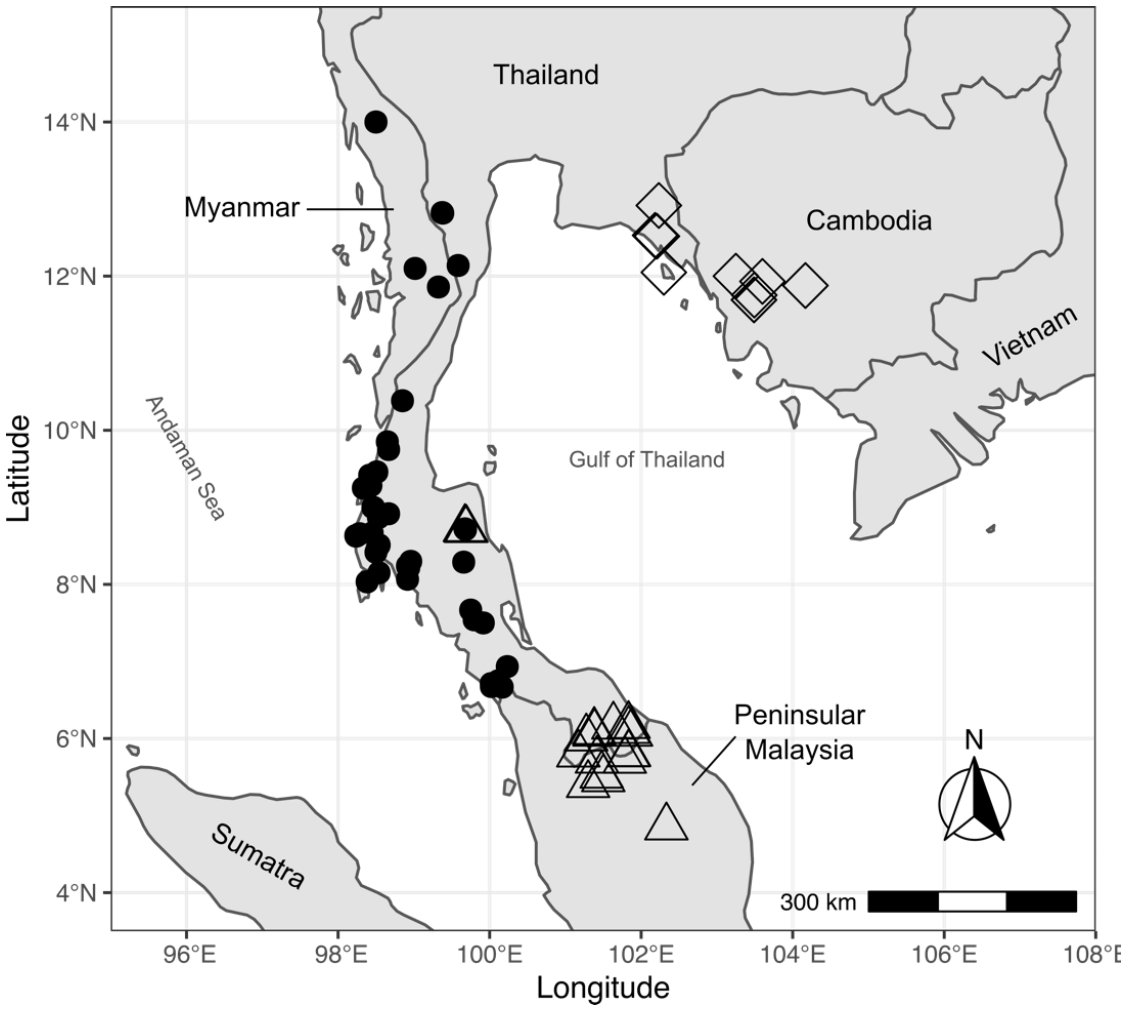


Figure 4.8 Occurrence map of *Urophyllum crassum* (△), *U. longifolium* (●) and *U. schmidtii* (◇).

3. ***Urophyllum glabrum*** Wall. in Roxburgh (1824: 185). Type: Penang, *Wallich 8316* (lectotype, designated by Wong *et al.* (2019) K-W! (♀) [K001125235]; isolectotypes BR-photo! [BR0000005620937], K-W! [K001125234], [K001125237])

Urophyllum aequale Craib (Bull. Misc. Inform. Kew 1931: 444), **synon. nov.** Type: Siam [Thailand], Satul [Satun], Klawng Ton, 300 m, 10 Mar 1928, *Kerr 14448* (♂) (lectotype, designated here K!; isolectotypes ABD!, BM!, BK!, BKF!, C-photo!, E-photo!, TCD-photo!).

Urophyllum fuscum Craib (Bull. Misc. Inform. Kew 1931: 446), **synon. nov.** Type: Siam [Thailand], Pang Nga [PhangNga], Kao Kata Kwam [8°30'N, 98°38'E], 400 m, 7 Mar 1930, *Kerr 18411* (♂) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!, K!, TCD-photo!).

Urophyllum oblongum Craib (Fl. Siam. ii. 1932: 84), **synon. nov.** Type: Siam [Thailand], Takuapa [district in PhangNga Province], Kapong, c. 100 m, 17 Feb 1929, *Kerr 17131* (♂) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!).

DISTRIBUTION. Throughout Peninsular Malaysia, Singapore, and southern Thailand (Surat Thani Province towards the south). Possibly found in Sumatra as mentioned in William Jack's letter to Nathaniel Wallich (Gage and Burkill, 1916) and the label of the specimen deposited at E [E00130812] (Figure 4.9).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: s.loc., s.a., *Griffith 2945* (♀) (K!); s.loc., s.a., *s.coll. 279* (♀) (K!); **JOHOR**, Johor Bahru District, Sungai Tukong estate [1°27'N, 103°56'E], 1 Aug 1932, *Corner 996* (sterile) (K!); ibid., *Sparz 996* (sterile) (K!); Kluang District, at 7 mile towards Keesing, 16 Nov 1922, *Horttum 9255* (♀) (K!); Kota Tinggi District, Kuala Sedili new road [1°56'N, 104°08'E], 25 Jun 1959, *Kadim & Noor 174* (♀) (K!); ibid., Pengerang estate, Pengerang [1°31'N, 104°03'E], 17 Oct 1934, *Teruya 2562* (sterile) (K!); ibid., Sedili, 9 Jan 1929, *Teruya 903* (♂) (K!); Mersing District, Kg. Hubang [Hubong] Development area, 100 ms. Endau Rd., 100 f [c. 30 m], 14 Jul 1959, *Burkill HMB 1896* (♀) (BKF!, K!, L-photo!); Bukit Alor, 22 Nov 1966, *Hardial 534* (♀) (K!); Bukit Kuing, 23 Jun 1934, *Corner SFN 28651* (♀) (E!, L-photo!); **KEDAH**, Langkawi Island District, Gunong Raya FR, Compt. 11 [6°22'N, 99°49'E], >1000 f [c. 300 m], 13 Mar 1969, *Chelliah FRI 6905* (♀) (K!, L-photo!); ibid., 100 f [c. 30 m], 17 May 1957, *Chew CWL 153* (♂) (K!); ibid., West coast

Malay Peninsula, 1 Nov 1916, *Robinson s.n.* (♂) (K!); Padang Terap District, Koh Mai [Moi] Forest Reserve [6°26'N, 100°37'E], 2 Apr 1938, *Kiah SFN 35125* (♀) (K!, L-photo!); ibid., south facing slope of Bukit Perak [FR] [5°58'N, 100°38'E], 1500 f [c. 460 m], 27 Nov 1969, *Everett FRI 13707* (♀) (K!); Sik District, Enggang Forest Reserve, central Kedah [5°49'N, 100°41'E], 13 Jul 1956, *Kochummen KEP 78874* (♀) (L-photo!, K!); Chubar FR, compt. 4, ridge side, 800 f [c. 240 m], 21 Jan 1969, *Bray FRI 11799* (♀) (K!); Pulau Tiifn advecy, 1 Apr 1911, *s.coll. 15870* (sterile) (K!); **KELANTAN**, Gua Musang District, Bertam, Ulu Kelantan [4°48'N, 101°27'E], 28 Jul 1962, *s.coll. 94* (♀) (L-photo!); ibid., S. Nenggiri near K. Jenera, ridge crest [5°06'N, 101°46'E], 18 Jul 1967, *Whitmore FRI 4084* (♀) (K!); Kota Bahru District, s.loc., s.a., *Gwynne-Vaughan 558* (♂) (L-photo!); ibid., s.loc. [6°07'N, 102°15'E], s.a., *s.coll. 558* (♂) (K!); Kuala Krai District, Sg. Durian Vjr., Machang [5°32'N, 102°17'E], 1000 f [c. 300 m], 7 Jul 1987, *Khairuddin & Damanhuri FRI 31952* (♀) (K!); Machang District, Kampong Gobek, Kerilla Estates, Tamangan [5°40'N, 102°11'E], 2 Mar 1959, *Shah & Kadim 514* (♂) (K!, L-photo!); ibid., Ulu Sat FR, hillside [5°42'N, 102°20'E], 18 Jun 1968, *Suppiah KEP 108852* (♂) (K!); G. stong, east face leading ridge, broad ridge top on granite, 2600 f [c. 790 m], 13 Aug 1969, *Whitmore FRI 12411* (♀) (K!); **KUALA LUMPUR**, Damansara Road, 10 Oct 1920, *Ridley s.n.* (♀) (K!); **MALACCA**, s.loc., s.a., *Griffith s.n. [AAU 2690, specimen on the right]* (♀) (AAU!); ibid., s.loc., s.a., *Griffith s.n.* (♀) (K!); ibid., s.loc., s.a., *Maingay 2616* (♀) (K!); ibid., s.loc. s.a., *Maingay 883* (♀) (K!); **NEGERI SEMBILAN**, Jelebu District, Pasoh FR, IBP study [3°02'N, 102°20'E], 23 Jul 1975, *Chan FRI18187* (♀) (K!); Tampin District, Tebong FR [2°31'N, 102°20'E], 1000 f [c. 300 m], 12 Nov 1959, *Poore 135* (♂) (K!); **PAHANG**, Cameron highlands District, Sungei Bertam valley [4°24'N, 101°27'E], 10 Oct 1963, *Chew CWL 905* (♀) (K!, L-photo!); ibid., Sungei Boh Valley [4°27'N, 101°27E], 9 Oct 1963, *Chew CWL 888* (♀) (K!, L-photo!); Jerantut District, along Sungai Tembelling, trail from Kuala Tahan H.Q., Taman Negara [4°23'N, 102°24'E], 80 m, 20 Apr 1975, *Balgooy 2472* (♀) (AAU!); ibid., Jalan Bukit Tersek [Terusek], T. Negera [Taman Negara] [4°29'N, 102°25'E], 30 Apr 1975, *Teo & Pachiappan 554* (♀) (K!, L-photo!); Lipis District, Taman Negara, K. Relau-K. Juram Rd. [4°39'N, 102°07'E], 440 m, 8 Aug 1996, *Saw FRI 44817* (♀) (K!); Raub District, Sungei Belut, 11 Aug 1929, *Kalong 20458* (♀) (K!); Segamat District, E. boundary of Segamat Wildlife reserve, North east of Segamat [2°35'44.36"N, 102°56'6.45"E], 6 Mar 1970, *Everett FRI 14307* (♂) (K!); Temerloh District, Tasek Bera [Tasik Bera], along the path leading to the lakes, 300 m, 15 May 1968, *Soepadmo 437* (♀) (K!, L-photo!); Taman Negara, 300 f [c. 90

m], 19 Apr 1975, *Cheng FRI 23323* (♀) (K!); **PENANG**, Penang Hill, May 1887, *Luetis 1189* (♂) (K!); ibid., 1 Aug 1940, *Nauen SING 37666* (♀) (K!); Poulo-Pinang [Penang], 27 Dec 1904, *s.coll. s.n.* (♀) (P-photo!); ibid., 28 Dec 1904, *s.coll. s.n.* (sterile) (P-photo!); s.loc., 500–700 f [c. 150–210 m], Apr 1881, *Dr. King's collector 1651* (♂) (K!, P-photo!); s.loc., Jun 1885, *Luetis 260 [specimen on the left]* (♂) (K!); s.loc., Jun 1885, *Luetis 260 [specimen on the right]* (♀) (K!); s.loc., 28 Dec 1904, *Phillips s.n.* (♀) (K!); s.loc., s.a., *Wallich 8319* (♂) (K-W!); **PERAK**, Batang Padang District, high forest, broad ridge, Slim Hill FR, SE Perak [4°00'N, 101°27'E], 9 Sep 1966, *Whitemore FRI 0848* (♀) (AAU!, K!); Hulu Perak District, Gunong Batu Puteh [Gunung Batu Putih] [5°40'N, 101°30'E], s.a., *Wray Jr 262* (♂) (L-photo!); ibid., Temengor FR, Compt. 44, block 5 (3) [5°31'52"N, 101°35'39"E], 18 May 2010, *Kamarul Hisham et al. FRI 67195* (♀) (L-photo!); ibid., 20 May 2010, *Kamarul Hisham et al. FRI 67207* (♂) (L-photo!); Kinta District, Goping [Gopeng], Oct 1880, *Dr. King's Collector 825* (♂) (P-photo!); Kuala Kangsar District, Sg. Plus, Chior FR, Sg. Siput [4°59'N, 101°09'E], 600 f [c. 180 m], 10 Jun 1967, *Ng FRI 5748* (♀) (K!); Larut Matang District, above Speedy's house, Maxwell's hill [4°51'N, 100°47'E], 3 Dec 1965, *Shah & Sidek MS 1068* (♀) (K!, L-photo!); ibid., Taiping hill [Bukit Larut], 3500 f [c. 1070 m], 16 Feb 1907, *Harilt & Jun 2357* (♀) (K!); ibid., Jalan Pokok Asam, 4 Oct 1968, *bin Kiah S 316* (sterile) (K!); Hill Garden Larut, s.a., *Wray Jr. 66* (♂) (K!); Jambong Rabok, s.a., *Scortechini 177* (♂) (K!, P-photo!); ibid., s.a., *Scortechini 177* (♀) (K!); Jaupriq Hautu, Mar 1896, *Ridley 7189* (♂) (K!); Kalatall hill, May 1888, *Wray Jr. 2079* (♀) (K!); Larut, Sep 1881, *Dr. King's collector 2409* (♂) (L-photo!); s.loc., 2000–3000 f [c. 610–910 m], Oct 1883, *Dr. King's collector 5034* (♂) (K!); s.loc., s.a., *Wray Jr. 2940* (♀) (P-photo!); **SELANGOR**, Gombak District, Hutan Lipur Sungai Sendat [3°24'2"N, 101°40'97"E], 2 Apr 2007, *Syahida Emiza & Angan FRI 55041* (♂) (K!, L-photo!); ibid., Rawang FR [3°19'N, 101°35'E], 8 Jul 1914, *Kloss s.n.* (♀) (K!); ibid., 8 Jun 1914, *Kloss s.n.* (♂) (K!); Hulu Selangor District, Rantau Panjang [3°25'N, 101°32'E], 30 Jul 1914, *Kloss 65* (♀) (K!); Petaling District, Bkt. Cheraka Res. [Bukit Cherakah FR] [3°10'N, 101°28'E], 7 May 1918, *Foxvorlky 2383* (♀) (K!); **TERENGGANU**, Sunga Ryah, Oct 1880, *Dr. King's collector 797* (♀) (K!); Dungun District, B. Bauk FR, 5th ml. [4°45'N, 103°21'E], 23 Jul 1962, *bin Zainuddin KEP 94974* (♀) (K!, L-photo!); ibid., Jerangau FR [4°55'42"N, 103°5'71"E], 51 m, 23 Jul 2009, *Julius FRI 56193* (♀) (K!); ibid., compt. 35 area [4°55'8"N, 103°6'35"E], 26 m, 24 Jul 2009, *Mohd. Hairul et al. FRI 69812* (♀) (K!); **PENANG AND SINGAPORE**, s.loc., s.a., *Wallich 8316* (♀) (E!, K-W!, P-photo!); ibid., s.loc., s.a., *Wallich 8316* (♂) (K-W!); **SINGAPORE**, Changi, Mar

1889, *Ridley 147* (♀) (K!); ibid., 6 Mar 1905, *Ridley 3601a* (♂) (K!); Chua Chu Kay, s.a., *Ridley 4907* (♀) (K!); s.loc., s.a., *Lobb 279* (♀) (E!); s.loc., s.a., *Lobb 331* (♀) (E!); s.loc., s.a., *Lobb s.n.* (♀) (K!); **THAILAND. NAKHON SRI THAMMARAT**, Phrom Khiri District, Phrom Lok Falls, Khao Luang NP, 200–300 m, 13 Jul 1993, *Puff 930712-1/11* (♀) (PSU!); **NARATHIWAT**, Bacho District, s.loc. [6°35'N, 101°4E], 150–350 m, 18 Jun 1992, *Larsen 42985* (♀) (AAU!); ibid., s.loc., 5 Jun 1961, *Sangkhachand 187* (♂) (BKF!, C!, K!); ibid., s.loc., 22 Dec 1968, *Sangkhachand 1603* (♀) (BK!); Sukhirin District, Hala-Bala WS, Nature trail from research station, 150 m, 22 Jul 2004, *Pooma et al. 4526* (♀) (BKF!); ibid., Khao Nakarat, 500–600 m, 20 Oct 1996, *Niyomdham 4849* (♀) (BKF!); Sungai Padi District, Ban Yuan Yahng, Group 3, 50 m, 6 Jun 1987, *Maxwell 87-500* (♀) (BKF!, PSU!); ibid., Chatwarin Falls, Budo-Sungai Padi NP [6°4'°N, 101°52'E], 25 Dec 1999, *Wongprasert 9912-55* (♀) (BKF!); ibid., [6°5'N, 101°50'E], 150–300 m, 15 Aug 1995, *Larsen et al. 45570* (♀) (AAU!, BKF!); Waeng District, Bala-Hala, 3 Aug 1999, *Puudjaa & Cholkulchana 621* (♀) (BKF!); ibid., Ban Bala, Hala-Bala WS, Research trail [5°47'55"N, 101°50'0.7"E], 115 m, 10 May 2019, *Yooprasert et al. 182* (♂) (BKF!, K!); ibid., *Yooprasert et al. 186a* (♀) (BKF!, K!); ibid., Bawae, Beside the road [5°47'55"N, 101°50'0.7"E], 118 m, 11 May 2019, *Yooprasert et al. 192* (♂) (BKF!, K!); ibid., [5°48'5.4"N, 101°50'28.5"E], 94 m, 11 May 2019, *Yooprasert et al. 197* (♀) (BKF!, K!); ibid., *Yooprasert et al. 200* (♀) (BKF!, K!); ibid., Hala-Bala, 15 Aug 2003, *Puudjaa 1194* (♀) (BKF!); ibid., Nature trail of Ornamental wild plant in southern project, 40 m, 30 Apr 2004, *Poopath 2* (♂) (BKF!); ibid., rubber plantation in Hala-Bala [5°28'32"N, 101°30'9"E], 120 m, 22 Aug 2006, *Poopath et al. 6* (♀) (E!); Budo [Budo Sungai Padi] NP, 50–350 m, 18 Jul 1993, *Puff 930718-1/2* (♀) (PSU!); Kao Re chau, Toh moh [5°48'10.6", 101°42'40.5"], 1800 f [c. 550 m], 20 Apr 1931, *Lakshnakara 730* (♀) (BK!, E!, K!); Khao Sana, Ban Krabulae, Kaburotai, 14 Jul 1987, *Niyomdham 715* (♂) (BKF!); **PHANGNGA**, Ko Yao District, Ban Chong Lad, Trail up hill, Scrub forest near rubber plantation [8°04'04"N, 98°35'4"E], 1 May 2007, *Suddee et al. 3156* (♂) (BKF!); Ta Kua Pa District, s.loc., 22 Mar 2008, *Chamchumroon et al. vc 2483* (♀) (BKF!); ibid., Ton Chong Fa waterfall, KhaoLak-Lamru NP [8°39'31"N, 98°17'2"E], 60 m, 11 Jun 2017, *Yooprasert et al. PHT 106* (sterile) (BKF!, K!); ibid., *Yooprasert et al. PHT 107* (♀) (BKF!, K!); ibid., [8°39'26"N, 98°17'1"E], 75 m, 11 Jun 2017, *Yooprasert et al. PHT 109* (♀) (BKF!, K!); ibid., [8°39'24"N, 98°17'3"E], 95 m, 11 Jun 2017, *Yooprasert et al. PHT 111* (sterile) (BKF!, K!); ibid., [8°39'18"N, 98°17'2"E], 113 m, 11 Jun 2017, *Yooprasert et al. PHT 116* (♀) (BKF!, K!); ibid., *Yooprasert et al. PHT 117* (sterile) (BKF!, K!); ibid., *Yooprasert et al.*

PHT 119 (♀) (BKF!, K!); ibid., *Yooprasert et al. PHT 120* (♀) (BKF!, K!); ibid., [8°39'27.3"N, 98°17'1.2"E], 75 m, 7 Apr 2018, *Yooprasert et al. PHT 171* (♂) (BKF!, K!); ibid., [8°39'20.4"N, 98°17'6.5"E], 97 m, 7 Apr 2018, *Yooprasert et al. PHT 173* (♀) (BKF!, K!); ibid., *Yooprasert et al. PHT 174* (♀) (BKF!, K!); ibid., [8°39'13"N, 98°17'4.8"E], 132 m, 7 Apr 2018, *Yooprasert et al. PHT 175* (♀) (BKF!, K!); ibid., [8°37'1.7"N, 98°14'48.2"E], 94 m, 7 Apr 2018, *Yooprasert et al. PHT 176* (♀) (BKF!, K!); Thung Rha Suung, 700 m, 27 Mar 2000, *Suksathan 2568* (♂) (QBG!); **PHATTHALUNG**, Tamot District, Mom Tui [Mom Jui] falls, Khao Ban tad [Banthat] WS, 100–200 m, 24 Jul 1993, *Puff 930724-1/2* (♀) (PSU!); ibid., NW of Hat Yai [7°2'N, 100°05'E], 150 m, 20 Aug 1995, *Larsen 45801* (♀) (AAU!, QBG!); **PHUKET**, Thalang District, Khao Phra Tao Non-hunting area, trail between Bang Pae and Ton Sai falls [8°02'N, 98°23'E], 120 m, 21 Apr 2006, *Gardner ST 2604* (♀) (K!); **SATUN**, Klawng Ton [6°40'N, 100°10'E], 200 m, 14 Mar 1928, *Kerr 14579* (♂) (BK!, BM!, E!, K!, TCD-photo!); Terutao [Tarutao] [6°35'N, 99°40'E], 5 m, 20 Jan 1928, *Kerr 14201* (sterile) (BK!, BM!, K!); **SONGKHLA**, Hat Yai District, Haad Yai [Hat Yai], Dton Nga Chang Reserve [Ton Nga Chang Waterfall], 450 m, 28 Sep 1985, *Maxwell 85-919* (♂) (L-photo!, PSU!); ibid., Ko hong hill, 300 m, 3 Nov 1984, *Maxwell 84-393* (♂) (PSU!); ibid., east side, 50 m, 20 Apr 1986, *Maxwell 86-246* (♂) (BKF!, L-photo!, PSU!); ibid., Ton Nga Chang WS [Falls], a few 100 m north of waterfall level 5 [6°57'N, 100°13'E], 300 m, 15 May 2004, *Gardner ST 0533* (♂) (BKF!, K!); ibid., Thung Tamsao, Ton Nga Chang waterfall, few hundred metres from 5th level falls on ridge side forest [6°56'44"N, 100°13'39"E], 366 m, 20 Jun 2017, *Yooprasert et al. SKT 157* (♀) (BKF!, K!); ibid., 400 m, 20 Jun 2017, *Yooprasert et al. SKT 158* (♀) (BKF!, K!); Nathawi District, Khao Nam Khang NP, trail to waterfall behind Headquarters [6°35'N, 100°35'E], 130 m, 18 May 2004, *Gardner & Sidisunthorn ST 0555* (♀) (BKF!, K!); ibid., 6 Oct 2004, *Tippayasri & Sidisunthorn ST 1062* (♀) (BKF!); ibid., 9 Apr 2018, *Yooprasert et al. SKN 178* (sterile) (BKF!, K!); ibid., 10 Apr 2018, *Yooprasert et al. SKN 179* (sterile) (BKF!, K!); ibid., south of Chana [6°45'N, 100°43'E], 100–150 m, 13 Jun 1992, *Larsen et al. 42866* (♀) (AAU!, BKF!, PSU!); ibid., [6°35'N, 100°34'E], 180 m, 28 Aug 1995, *Larsen et al. 46090* (♀) (AAU!, BKF!); Rattaphum District, Boriphat falls, 5 Aug 1993, *Puff 930805-1/5* (♀) (PSU!); ibid., 24 Apr 1987, *Sirirugsa 1066* (♀) (PSU!); Saba Yoi District, San Kalakhiri NP, trail by the stream, 10 Apr 2018, *Yooprasert et al. SKS 180* (sterile) (BKF!, K!); ibid., *Yooprasert et al. SKS 181* (♀) (BKF!, K!); Ban Prakawp [6°34'N, 100°40'E], 100 m, 18 Jul 1928, *Kerr 15844* (♀) (BM!, E!, K!); **SURAT THANI**, Phanom District, Khlong Panom [Phanom] NP, Rafflesia nature trail

[8°52'°N, 98°32'E], 300 m, 18 Jun 2004, *Gardner ST 0533* (♂) (K!); **TRANG**, Chong District, Kao chong [Khao Chong], 13 Mar 1969, *Sangkhachand 1782* (♀) (BK!); ibid., [7°40'N, 99°45'E], 150 m, 14 Jun 1974, *Geesink et al. 7223* (♀) (C!, K! [sheet 1]); ibid., Khao Chong Botanical Garden, Nayong [7°32'28"N, 99°47'42"E], 160 m, 9 Jun 2017, *Yooprasert et al. TRC 84* (♂) (BKF!, K!); ibid., 157 m, 9 Jun 2017, *Yooprasert et al. TRC 85* (♂) (BKF!, K!); ibid., *Yooprasert et al. TRC 86* (sterile) (BKF!, K!); ibid., [7°32'23"N, 99°47'43"E], 209 m, 9 Jun 2017, *Yooprasert et al. TRC 87* (sterile) (BKF!, K!); ibid., [7°32'22"N, 99°47'45"E], 205 m, 9 Jun 2017, *Yooprasert et al. TRC 88* (♂) (BKF!, K!); ibid., [7°32'25"N, 99°47'49"E], 158 m, 9 Jun 2017, *Yooprasert et al. TRC 89* (♀) (BKF!, K!); ibid., [7°32'30"N, 99°47'44"E], 132 m, 9 Jun 2017, *Yooprasert et al. TRC 92* (♀) (BKF!, K!); ibid., 350 m, 13 Jul 1985, *Maxwell 85-713* (♀) (BKF!, L!, PSU!); ibid., Trail from Bot garden Headquarter toTon Yai [7°33'N, 99°47'E], 80 m, 9 Jul 2000, *Middleton et al. 358* (♀) (BKF!, E!); Palian District, Ton Tae [Ton Te] Falls, 250 m, 24 Feb 2002, *Chamchumroon et al. vc 1320* (♂) (BKF!); Yan Ta Khao District, s.loc., 150–200 m, 29 Jul 1993, *Puff 930729-1/3* (♀) (PSU!); ibid., Sai Roong Falls, 350 m, 26 Apr 1987, *Maxwell 87-429* (♀) (BKF!, L-photo!, PSU!); **YALA**, Banang Sata District, Hill beside the road [6°19'8.7"N, 101°22'59.6"E], 627 m, 14 May 2019, *Yooprasert et al. 223* (♂) (BKF!, K!); ibid., [6°19'9.9"N, 101°22'39"E], 499 m, 14 May 2019, *Yooprasert et al. 224* (♂) (BKF!, K!); Than To District, Mae-Wat, Bang Lang NP, Ban Wang Sai, Trail to summit at 800 m alt., 300 m, 18 Jul 2004, *Pooma et al. 4358* (♀) (BKF!); Betong District, Rubber tree trail, Malaysia border road No.4266 Km [5°38'37.5"N, 101°7'50.1"E], 750 m, 13 May 2019, *Yooprasert et al. 211* (sterile) (BKF!, K!); ibid., Than Nam Thip, Rd to Ban Cha-ro Susu near Malay border [5°43'66"N, 101°7'77"E], 241 m, 22 Apr 2005, *Pooma et al. 5133* (♂) (AAU!, BKF!).

HABITAT. Most specimens were collected from primary or secondary tropical evergreen forest or swamp forest; sometimes mixed with dipterocarps and bamboos; scrub forest or sandy beach forest; usually found in shaded areas; elev. 0–1,070 m.

CONSERVATION STATUS. Least Concern (LC). *Urophyllum glabrum* has a widespread distribution in many protected areas in southern Thailand, Peninsular Malaysia and Singapore with recent collections from more than 15 locations since the year 2000. The estimated AOO (244 km^2) is rated as Endangered, however there is a large EOO (179,943 km^2) and the species is commonly found in many protected areas, this species is therefore less likely to be threatened.

PHENOLOGY. Collected in flower from January to November; collected in fruit from March to December.

NOTES. *Urophyllum glabrum* is recognisable by the presence of hairy pocket domatia on almost every axil between the midrib and secondary veins on the abaxial leaf surface. This character is shared by *U. villosum* in southern Thailand, however the abaxial leaf surface of *U. villosum* is densely hairy, whereas it is subglabrous in *U. glabrum* resulting in the inconspicuous presence of domatia in *U. glabrum*. The species also has stipules appressed to the stem but not folded, calyx truncate to toothed, glabrous green corolla with a tubular membrane connected to all lobes at the base. *Urophyllum longifolium* is morphologically similar to *U. glabrum* but it differs by having folded stipules, white corolla and lacking a membrane at the base of the corolla lobes. Other species found in the same distribution area in Narathiwat Province are *U. macrophyllum* and *U. trifurcum*. The first species, *U. macrophyllum*, differs by having stipules folded, secondary veins festooned brochidodromous, and a sessile umbellate inflorescence. *Urophyllum trifurcum*, like *U. villosum*, has an abaxially hairy corolla and triangular membrane at the base of the corolla lobes.

After examination of type specimens of *U. aequale* (*Kerr 14448*), *U. fuscum* (*Kerr 18411*) and *U. oblongum* (*Kerr 17131*), their characters are within the morphological range of *U. glabrum*. All specimens have hairy pocket domatia, umbellate inflorescences and a tubular membrane at the base of the corolla lobes. The characters differ between these species are leaf shape and peduncle length. The leaves of the type of *U. aequale* are elliptic to oblong-elliptic; they are oblong in *U. oblongum*; and obovate in *U. fuscum*. However, after examined many herbarium specimens of *U. glabrum*, leaf shape and peduncle length can be very variable within this species. To this extent, all three species are included to *U. glabrum*. The designated lectotypes for these synonymous species were based upon Craib's biography as in *U. crassum*.

The specimens deposited in K were selected as the lectotype of *U. aequale*, *U. fuscum* and *U. oblongum* based upon Craib's biography on the Taxonomic Literature II (Stafleu and Mennega, 1997) and Bulletin of Miscellaneous Information, Kew (Kerr, 1933) as discussed previously for *U. crassum*.

Phylogenetic study on plastid and nrDNA datsets were incongruent on the position of a population of *U. glabrum* collected from Ton Chong Fa Falls in PhangNga Province (Chapter 3). This population was nested within *U. longifolium* clade in plastid data with the other two populations forming a species clade and being sister to the *U. longifolium* clade. While, in nrDNA, the relationship is resolved, and all populations form a species clade. This evidence supports that *U. glabrum* and *U. longifolium* are closely related species and are morphologically similar.

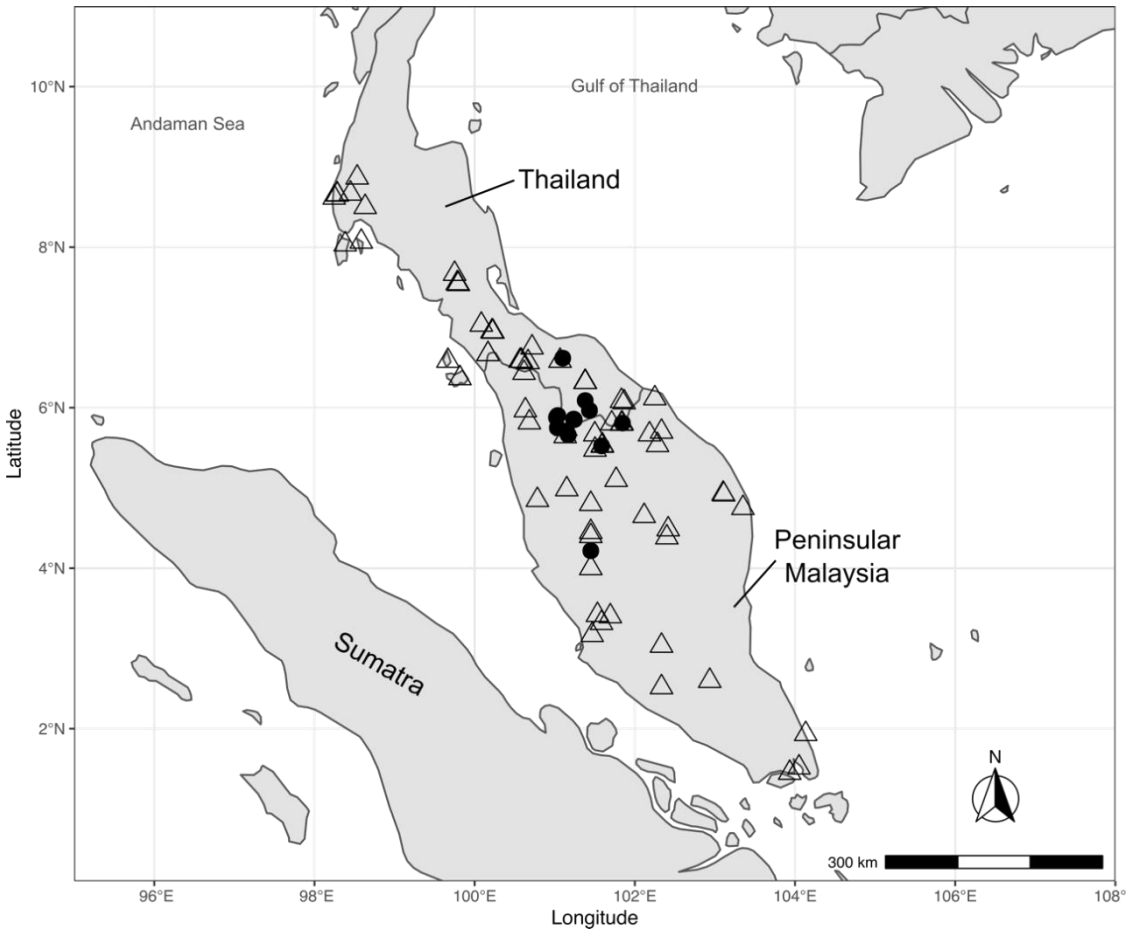


Figure 4.9 Occurrence map of *Urophyllum glabrum* (partial distribution) (△) and *U. longipes* (●).

4. ***Urophyllum hirsutum*** (Wight) Hook.f. (Hooker 1880: 99). *Axanthes hirsuta* Wight (1847: 148). Type: [Malaysia, Peninsular Malaysia], Malacca, 1845, *Griffith s.n.* (♂) (lectotype, designated by Wong *et al.* (2019) K! [K000740842]).

DISTRIBUTION. Throughout Peninsular Malaysia, Singapore, and southernmost Thailand (Yala and Narathiwat provinces). Possibly distributed in Sumatra and Borneo (Figure 4.10).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: Malaya [3°14'N, 101°45'E], 6 Apr 1905, *Maingay 873* (♀) (L-photo!); 20 m. Genting Sempah, 24 Sep 1926, *Strugnell 12137* (♂) (E!); s.loc., s.a., *Griffith 2939* (♂) (P-photo!); **JOHOR**, Kluang District, Gunung Lambak recreational area [2°02'N, 103°22'E], 120 m, 22 Jul 1991, *Kamarudin FRI 31481* (♀) (K!); ibid., Kluang FR, G. Belumut, trail to summit [2°03'40"N, 103°32'11"E], 145 m, 11 Aug 2009, *Kamarul Hisham & Syahida Emiza FRI 67152* (♀) (L-photo!); ibid., *Kamarul Hisham & Syahida Emiza FRI 67154* (♀) (L-photo!); ibid., 2 Feb 1966, *Ng KEP 97997* (♀) (K!); ibid., Rengam FR [2°01'N, 103°21'E], 750 f [c. 230 m], 17 Nov 1975, *Ang FRI 23437* (♀) (K!); Kota Tinggi District, Panti FR, compartment 16 [1°49'N, 103°51'E], 27 Oct 1988, *Zainudin AZ 2611* (♂) (K!); Mersing District, Ulu Sungei Sedili Besar, east of G. Sumalayang [2°04'16''N, 103°52'00''E], 29 Jan 1970, *Everett FRI 13879* (♀) (K!); Segamat District, Labis FR, near Pahang [2°27'N, 103°06'E], 19 Feb 1971, *Suppiah FRI 14794* (♀) (K!); ibid., Sungei Juasseh [2°32'10''N, 102°55'12''E], 28 Jun 1970, *Ahmad S 300* (♀) (L-photo!); Bahru, Ulu Sibul FR, 300–400 f [c. 90–120 m], 2 Jun 1980, *Mat FRI 25581* (♂) (K!); **KEDAH**, Pendang District, Bukit Perak FR [5°58'N, 100°38'E], 1500 f [c. 460 m], 27 Nov 1969, *Chan FRI 13201* (♀) (K!); Sik District, North facing slope Bukit Enggang [5°49'N, 100°41'E], 12 Apr 1969, *Everett FRI 13776* (♀) (K!); **KELANTAN**, Gua Musang District, Ulu S. Aring, near K. Telong [4°39'19''N, 102°20'47''E], 21 Sep 1967, *Whitmore FRI 4438* (♀) (K!); **NEGERI SEMBILAN**, Jelebu District, Jelebu FR [2°59'N, 102°04'E], 7 May 1969, *Suppiah FRI 11315* (♀) (K!); Juuong Augsi, Dec 1898, *Ridley 10100* (♂) (K!); **PAHANG**, Bentong District, Bukit Lenting [Lentang] [3°24'N, 102°00'E], 14 Jun 1935, *Lymington 40508* (♀) (K!); ibid., Sabai Estate, near Bentong [3°20'N, 102°05'E], 400 f [c. 120 m], 27 Jan 1958, *Shah 163* (♀) (L-photo!); Jerantut District, near Kuala Teku [4°33'N, 102°19'E], 20 Jul 1936, *Kiah SING 31745* (♀) (BKF!); ibid., Taman Negara, foot of Gua Peningat [4°22'N, 102°33'E], 500 f [c. 150 m], 15 Jul 1970, *Loh FRI 17245* (♀) (K!); **PAHANG**, Jerantut District, Taman Negara, Kuala Juram, Sg. Tanum [4°39'N, 102°8'E], 30 m, 5 May 1997, *Chua et al. FRI 38845* (♀) (BKF!); ibid., Ulu

Sat near Tg. Petir [4°37'N, 102°36'E], 1000 f [c. 300 m], 7 Nov 1970, *Whitmore FRI 15237* (♀) (K!); ibid., near Kuala Kelapah [4°33'N, 102°34'E], 100 f [c. 30 m], 10 Jul 1970, *Shah & Noor MS 1784* (♀) (L-photo!); ibid., Ulu Cheka, Benom forest (Quadrat 4, tree 44) [3°53'N, 102°12'E], 13 Jun 1968, *Teo & Pachiappan 109* (♂) (L-photo!); ibid., Ulu Sungai Sepia near Kuala Aur [4°30'N, 102°42'E], 100–200 f [c. 30–60 m], 16 Jul 1970, *Shah & Noor MS 1900* (♀) (L-photo!); ibid., Ulu Tekam [3°58'N, 102°35'E], 550 f [c. 170 m], 27 Jun 1972, *Ng & Beltran FRI 6442* (♀) (K!); Lipis District, Sungai Yu [4°30'03''N, 102°00'10''E], 14 Aug 1964, *Hardial & Noor 55* (♀) (L-photo!); Raub District, G. Benom Game reserve, Ulu Krau [3°49'N, 102°01'E], 2300 f [c. 700 m], 19 Apr 1967, *Ismail KEP 97815* (♂) (K!); ibid., State land [3°45'N, 101°45'E], 1500 f [c. 460 m], 25 Mar 1971, *Sohadi FRI 14723* (♀) (K!); ibid., Ulu Tranum FR, at 14 miles [3°40'N, 101°47'E], 3 Sep 1976, *Mat FRI 21677* (♀) (K!); Rompin District, Lesong FR [2°45'N, 103°08'E], 21 Feb 1980, *Maxwell 80-94* (♀) (L-photo!); ibid., 6 Oct 1979, *Chan FRI 25198* (♀) (K!); ibid., 200 f [c. 60 m], 29 Jun 1972, *Chan FRI 16904* (♀) (K!); Temeloh District, Kuala Lompat, Krau FR [3°42'N, 102°08'E], 70 m, 29 Jun 1988, *Kamarudin FRI 34517* (♀) (K!); ibid., S. boundary Krau Game reserve NR, Sungai Rangit [3°45'N, 102°19'E], 200 f [c. 60 m], 11 Oct 1969, *Everett FRI 13622* (♀) (K!); Taman Negara, 500 f [c. 150 m], 20 Apr 1975, *Ang FRI 23336* (♀) (K!); **PENANG**, Timur Laut District, Penara Bukit [5°23'N, 100°16'E], Mar 1890, *Luetis 1759* (♂) (K!); Batu Etam Bass, 2000 f [c. 600 m], Apr 1886, *Lurtis 785* (♂) (K!); Pulo-Pinang [Penang Island], s.a., *Wallich 8320* (♂) (K-W!); ibid., s.a., *Wallich 8320* (♂) (E!, K-W!); **PERAK**, Larut Matang District, Sg. Wang Bubu FR [4°38'27''N, 100°42'51''E], 1000 f [c. 300 m], 27 Apr 1968, *Ng FRI 6084* (♂) (K!); Manjung District, Dindings, South Pangkor FR [4°19'N, 100°39'E], 350 f [c. 110 m], 6 Jul 1955, *Burkill HMB 180* (♀) (L-photo!); Larut, Aug 1882, *Dr. King's Collector 3236* (♂) (P-photo!); **SELANGOR**, Gombak District, FRI, Kepong, field 4 B [3°14'N, 101°37'E], 7 Mar 1980, *Vethevelu FRI 25423* (♀) (K!); ibid., Klang Gates [3°14'N, 101°45'E], 11 Dec 1953, *Sinclair 7908* (♂) (E!, L-photo!); Hulu Langat District, Dusun Tua [3°08'N, 101°50'E], May 1896, *Ridley 7436* (♀) (K!); Hulu Selangor District, Gading FR, Ulu Selangor [3°40'N, 101°38'E], 1400 f [c. 430 m], 19 Jul 1969, *Loh FRI 13370* (♀) (K!); **TERENGGANU**, Hulu Terengganu District, Tasik Kenyir, Hutan Simpan Tembat, Sg. Jalang, compt. 108 [5°13'21"N, 102°34'46"E], 345 m, 4 Jan 2009, *Kamarul Hisham et al. FRI 67069* (♂) (K!); Kuala Kangsar District, Gunong Bubu reserve, compt. 47, line 1 of hill forest regeneration reserve [4°40'N, 100°48'E], 17 Jun 1969, *Selvaraj FRI 11125* (♀) (K!); **SINGAPORE.** Bukit Timah Forest

Reserve [1°18'N, 103°48'E], 16 Oct 1948, *Sinclair s.n.* (♂) (P-photo!); ibid., *Sinclair 5243* (♂) (E!); ibid., 23 Oct 1967, *Hardial 622* (♀) (BKF!, L!); ibid., 25 Apr 1984, *Leeuwenberg 13351* (♂) (WAG-photo!, L-photo!); ibid., 50 m, 17 Nov 1978, *Maxwell 78-388* (♀) (L-photo!); ibid., Fern valley, 14 Nov 1982, *Axelius 176* (♀) (L-photo!); ibid., *Axelius 177* (♀) (L-photo!); ibid., 50 m, 4 Dec 1980, *Maxwell 80-216* (♀) (L-photo!); ibid., 26 Mar 1981, *Maxwell 81-50* (♀) (L-photo!); Jamping, Karong, s.a., *s.coll. 4913* (sterile) (L-photo!); Sungei Morai, 19 Dec 1953, *Sinclair SING 40180* (♀) (E!, L-photo!); **THAILAND. NARATHIWAT**, Waeng District, Bala-Hala, 3 Aug 1999, *Puudjaa & Cholkulchana 613* (♀) (BKF!); ibid., Hala-Bala WS, Research trail [5°47'55"N, 101°50'0.7"E], 115 m, 10 May 2019, *Yooprasert et al. 185* (sterile) (BKF!, K!); ibid., *Yooprasert et al. 186* (sterile) (BKF!, K!); ibid., *Yooprasert et al. 190* (♀) (BKF!, K!); ibid., *Yooprasert et al. 198* (sterile) (BKF!, K!); ibid., Bawae, Beside the road [5°47'54.6"N, 101°45'30.4"E], 241 m, 11 May 2019, *Yooprasert et al. 202* (♂) (BKF!, K!); **YALA**, Betong District, Hala-bala WS, unnamed trail 1490 mt reach from shore of Bang Lang resevoir, 600 m, 22 May 2005, *Middleton et al. 3566* (♂) (BKF!, E!); ibid., Hill beside the road [6°19'9.9"N, 101°22'39"E], 499 m, 14 May 2019, *Yooprasert et al. 225* (♀) (BKF!, K!); Than To District, Ban Chulaphon Phattana 7, Bang Lang Dam, in partly opened by stream [6°5'18"N, 101°22'50"E], 200 m, 31 Oct 2001, *Pooma et al. 3184* (♀) (BKF!, P!); ibid., Maewat, trail behind the Prince Chulaphon (Princess Chulabhorn) resting place [6°5'0.2"N, 101°22'30.8"E], 270 m, 6 Aug 2016, *Poopath et al. 1696* (♀) (BKF!); ibid., *Poopath et al. 1698* (♀) (BKF!).

HABITAT. Lowland evergreen dipterocarp forest; sometimes on clay; by streams or slopes; open to shaded area; elev. 30–700 m.

REGIONAL CONSERVATION STATUS. Least Concern (LC). Despite the large estimated EOO (95,363 km^2), suitable habitats of *Urophyllum hirsutum* are likely declining due to land development for both urban and agricultural use, especially in the Bentong, Ruab, Temeloh and Lipis districts in Pahang State, Peninsular Malaysia. However, with many localities found, the large EOO and recent collections from five locations in the 2000s, the species does not currently fulfil the criteria of the threatened categories and, therefore, has been rated as Least Concern at the regional level.

Although, *Urophyllum hirsutum* was reported to be endemic to Peninsular Malaysia, Singapore and Thailand in Wong *et al.* (2019), collections recorded from Indonesia and Borneo deposited in E, K, L and P have similar characters to *U. hirsutum* found in Peninsular Thailand and Malay Peninsula. Further study is required to investigate the identification of the specimens with similar characters from Indonesia and Borneo to assess the overall distribution of *U. hirsutum*, as this could affect the global conservation status.

PHENOLOGY. Collected in flower and fruit throughout the year.

NOTES. *Urophyllum hirsutum* can be recognised by its hairy appearance, secondary veins festooned brochidodromous, sessile to subsessile umbellate inflorescence, calyx lobed, corolla densely hairy abaxially, whitish to pale yellow at base then dark green toward apex. *Urophyllum hirsutum* is similar to *U. blumeanum* (vegetatively) and *U. streptopodium* (reproductively). For *U. blumeanum*, the shared characters are chartaceous leaves, oblong to elliptic with 10–17.5 cm long, 2–6 cm wide, apex attenuate and densely hairy abaxially. However, *U. blumeanum* differs by its folded stipules, secondary veins weakly brochidodromous and two-tiered umbellate inflorescences, whereas *U. hirsutum* has appressed stipules but not folded with other characters previously stated. *Urophyllum streptopodium* is similar to *U. hirsutum* by its sessile to subsessile umbellate inflorescences, abaxial corolla hairy and corolla pale white/yellow at base then green toward apex, but they differ in vegetative characters as *U. streptopodium* has stipules constricted above the base, leaves coriaceous and secondary veins weakly brochidodromous, whereas in *U. hirsutum* stipules not constricted, leaves chartaceous and secondary veins festooned brochidodromous. Other species found in the same areas (Yala Province (Thailand) and northern half of Peninsular Malaysia southward to Kuantan City, Pahang State) are *U. longipes* and *U. villosum*. These two species differ to *U. hirsutum* by their inflorescences; simple cymose in *U. longipes* and pedunculate to two-tiered umbellate in *U. villosum*, whereas inflorescences of *U. hirsutum* are sessile to subsessile umbellate. *Urophyllum longipes* also differs by being subglabrous in all plant parts, and abaxial corolla glabrous; whereas *U. hirsutum* is densely hairy especially on the stem at shoot, stipules and abaxial leaf surface, and abaxial corolla hairy. For *U. villosum*, it has

coriaceous leaves, apex distinctly acuminate with long tail (usually 2–3 cm long), while in *U. hirsutum*, leaves are chartaceous with apex attenuate.

The position of *U. hirsutum* is incongruent among plastid and nrDNA trees, it is either sister to the *U. longifolium* and *U. glabrum* clade (plastid data) or to *U. blumeanum* and *U. memecyloides* (nrDNA data). However, morphologically, *U. hirsutum* is more similar to *U. blumeanum* and *U. memecyloides* than to *U. longifolium* and *U. glabrum*. The latter two species tend to have larger leaves (10.7–28.3 cm long, 2.7–10.2 cm, wide), the plants are usually subglabrous (except *U. longifolium* collected from Kaeng Krachan National Park) and calyx subtruncate or toothed, whereas *U. hirsutum* leaves are smaller (as previously discussed), densely hairy in appearance and the flowers have a lobed calyx. The characters shared between *U. memecyloides* and *U. hirsutum* are similar to those with *U. blumeanum* as previously discussed except the abaxial leaf surface of *U. memecyloides* is subglabrous, there is a two-tiered umbellate inflorescence, and the abaxial corolla is glabrous.

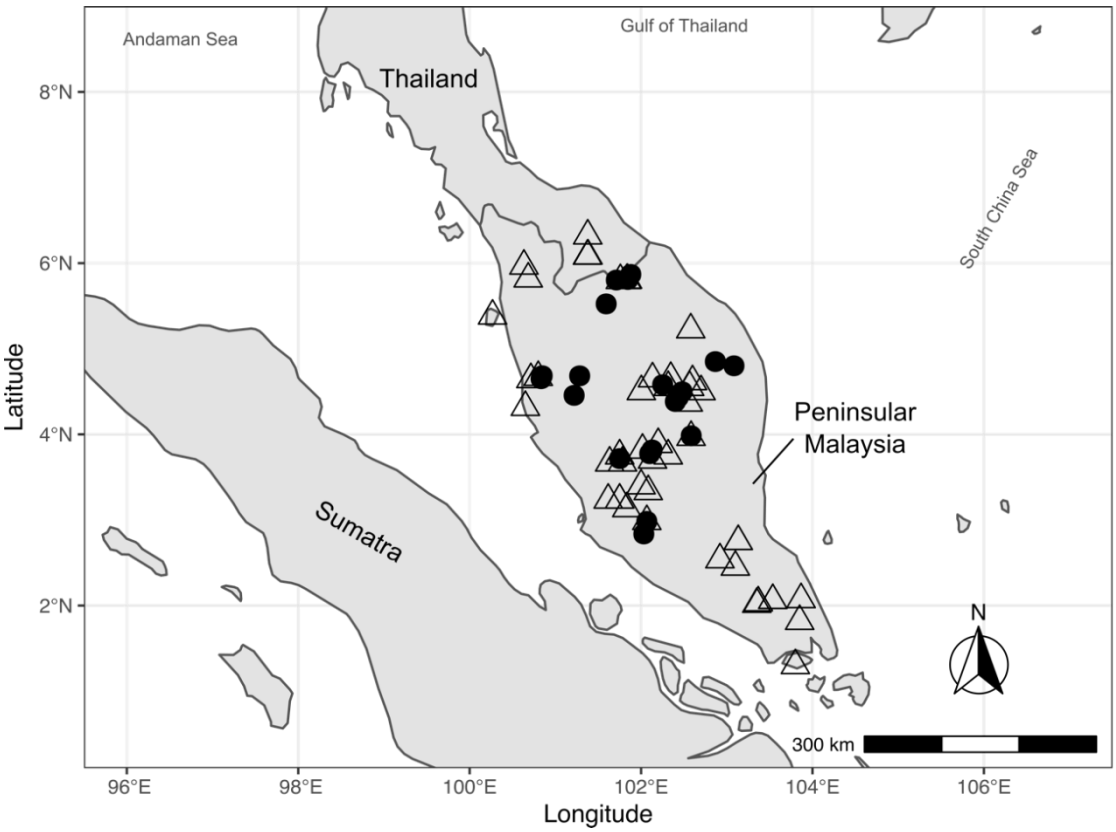


Figure 4.10 Occurrence map (partial distribution) of *Urophyllum hirsutum* (△) and *U. macrophyllum* (●).

5. ***Urophyllum longifolium*** (Wight) Hook.f. (Hooker 1880: 99). *Axanthes longifolia* Wight (1847: 145). Type: [Myanmar], Mergui, *Griffith 9* (♂) (second-step lectotype, designated here K! [K000031290]; isolectotypes E! [E00847781], K! [K000031291, K000031292]).

Urophyllum longifolium var. *pilosum* Craib (1932: 84), **synon. nov.** Type: Siam [Thailand], Trang, Ampo Kao Kao, Kuan Pra, 30 June 1929, *Rabil 251* (♀) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!).

Urophyllum talangense Craib (in Bull. Misc. Inform. Kew 1931: 447), **synon. nov.** Type: Siam [Thailand], Puket [Phuket], Talang [Thalang], c. 50 m, 11 Mar 1929, *Kerr 17437* (♀) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!).

DISTRIBUTION. India (West Bengal State), Myanmar (Tanintharyi Region) and Thailand (Phetchaburi to Songkhla Provinces) (Figure 4.8).

SPECIMENS EXAMINED. BIRMA&MALAY PENINSULA. S.loc., s.a., *Herb Griffith 2945* (sterile) (L-photo!); **INDIA.** Bengalia circa Calcuttam, 11 Jan 1905, *Helfer 97* (♂) (K!); ibid., s.a. 1838, *Helfer 509* (♂) (K!); **MYANMAR. TANINTHARYI**, Tenasserim [12°6′N, 99°1′E], 50 m, 5 Jun 1932, *Kerr 21669* (♀) (BK!, K!); ibid., Along the roadside en route from Mawtaung to Tanintharyi, ca. 43 km SE of Tanintharyi town [11°51′32.63″N, 99°19′28.02″E], 172 m, 6 Jun 2016, *Tagane et al. MY 339* (♂) (FU!); ibid., Tavoy District, hill west of Paungdaw Power station [14°N, 98°30′E], Aug 1961, *Keenan et al. 865* (♀) (E!, K!); ibid., *Keenan et al. 1174* (♀) (E!, K!); ibid., *Keenan et al. 1182* (♀) (E!, K!); **TENASSERIM AND ANDAMANS.** s.loc., s.a., *Helfer 2940* (♂) (K!, P-photo!); ibid., s.a., *s.coll. 2944* (♂) (K!); **THAILAND. CHUMPHON**, Phato Watershed Conservation and management Unit, 4 Nov 2008, *Wessumritt 130* (♂) (QBG!); **KRABI**, Huai Tai Falls, Khao Phanom Benja NP, Mueang [8°14′N, 98°55′E], 50 m, 9 May 2002, *Pooma et al. 3640* (♂) (BKF!); ibid., Khao Panom [Phanom] Bencha NP, trail from Ban San to top of Panom Bencha, along trail in forest [8°17′56″N, 98°57′32″E], 400 m, 18 Jun 2006, *Williams et al. 1856* (♀) (BKF!, E!); ibid., foothill of S range [8°13′N, 98°56′E], 100–360 m, 11 Jul 1992, *Larsen et al. 43305* (♀) (AAU!, BKF!); ibid., trail near headquarters [8°14′N, 98°55′E], 100 m, 16 Jul 2000, *Middleton et al. 489* (♀) (BKF!, E!, K!); ibid., [8°14′25″N, 98°55′E], 195 m, 8 June 2017, *Yooprasert et al. KRP 77* (♀) (BKF!, K!); ibid., [8°14′25″N, 98°55′4″E], 261 m, 8 June 2017, *Yooprasert et al. KRP 80* (♀) (BKF!, K!); ibid.,

trail to viewpoint [8°14′25″N, 98°54′56″E], 166 m, 8 June 2017, *Yooprasert et al. KRP 76* (♂) (BKF!, K!); ibid., [8°14′25″N, 98°55′4″E], 261 m, 8 June 2017, *Yooprasert et al. KRP 79* (sterile) (BKF!, K!); Nai Chawng, 24 May 1960, *Chirayupin 67* (♀) (BK!); Nong Khon, 15 Aug 1964, *Sangkhachand 1027* (♀) (BKF!, C!, K!); Tambon Kao Panom [8°4′N, 98°55′E], 100 m, 30 Mar 1930, *Kerr 18749* (♀) (BM!); ibid., *Kerr 18769* (♂) (BK!, E!, K!); **NAKHON SRI THAMMARAT**, Gahrome Falls, Langsagah, Khao Luang NP, shaded, rocky area, primary evergreen forest, 300 m, 19 May 1985, *Maxwell 85-511* (♀) (BKF!, L-photo!, PSU!); Khao Luang Khiriwong, 21 Oct 1951, *Smitinand 980* (♂) (BKF!); Khao Luang NP, Khrung Ching Falls, Nopphitam [8°43′14″N, 99°40′28″E], 270 m, 18 Jun 2017, *Yooprasert et al. NSK 136* (sterile) (BKF!, K!); ibid., [8°43′11″N, 99°40′33″E], 292 m, 18 Jun 2017, *Yooprasert et al. NSK 141* (♂) (BKF!, K!); ibid., [8°42′44″N, 99°40′55″E], 278 m, 18 Jun 2017, *Yooprasert et al. NSK 144* (sterile) (BKF!, K!); ibid., 200 m, 27 Feb 2002, *Chamchumroon et al. vc 1340* (♀) (BKF!); Khao Soop, 600 m, 10 Apr 1955, *Snan 93* (sterile) (BKF!); ibid., Lansageh, Gahrome Falls, Khao Luang NP, 200 m, 14 Apr 1985, *Maxwell 85-399* (♂) (BKF!, E!, L-photo!, PSU!); Thap Chang, 10 May 1954, *Phloenchit 798* (♂) (BKF!); Yong Falls NP, Khao Maen, Primary evergreen forest, Na Bon [8°17′20″N, 99°39′29″E], 600 m, 8 Feb 2005, *Williams et al. 1278* (♂) (BKF!, E!); **PHETCHABURI**, Kaeng Krachan NP [12°49′26″N, 99°22′24″E], 789 m, 26 Jun 2017, *Yooprasert et al. PCK 160* (♀) (BKF!, K!); ibid., [12°49′29″N, 99°22′28″E], 765 m, 26 Jun 2017, *Yooprasert et al. PCK 161* (sterile) (BKF!, K!); ibid., *Yooprasert et al. PCK 164* (sterile) (BKF!, K!); ibid., Trail to Tan Thip Wfall [Waterfall] [12°8′17″N, 99°35′5″E], 700 m, 12 May 2005, *Middleton 3409* (♂) (BKF!, E!); ibid., transect line 4, [12°49′3.5″N, 99°22′53.5″E], 860 m, 28 Oct 2013, *Tagane T 2393* (sterile) (BKF!); **PHANGNGA**, Kuraburi District, Bangwan stream, 16 Apr 2007, *Muadsub 273* (♂) (BKF!, PSU!); ibid., 14 May 2007, *Muadsub 280* (♀) (PSU!); ibid., Si Phang Nga NP, Khuraburi, Bangwan, Tamnang falls [8°59′48″N, 98°28′10″E], 84 m, 12 Jun 2017, *Yooprasert et al. PHS 128* (♂) (BKF!, K!); ibid., Ton Tei falls [8°59′53″N, 98°27′36″E], 48 m, 12 Jun 2017, *Yooprasert et al. PHS 122* (sterile) (BKF!, K!); ibid., *Yooprasert et al. PHS 123* (♂) (BKF!, K!); ibid., [8°59′54″N, 98°27′40″E], 36 m, 12 Jun 2017, *Yooprasert et al. PHS 125* (♀) (BKF!, K!); ibid., [8°59′59″N, 98°27′41″E], 56 m, 11 Jun 2017, *Yooprasert et al. PHS 126* (♀) (BKF!, K!); ibid., [8°59′48″N, 98°28′10″E], 84 m, 12 Jun 2017, *Yooprasert et al. PHS 129* (♂) (BKF!, K!); Mueang District, Nop-pring, Sa Nang Manora Forest Park [8°30′39″N, 98°32′29″E], 87 m, 10 Jun 2017, *Yooprasert et al. PHM 100* (♂) (BKF!, K!); ibid., 85 m, 10 Jun 2017, *Yooprasert et al. PHM 101* (♀) (BKF!, K!); ibid.,

Yooprasert et al. PHM 102 (♀) (BKF!, K!); ibid., *Yooprasert et al. PHM 103* (♀) (BKF!, K!); ibid., [8°30′41″N, 98°32′26″E], 70 m, 10 Jun 2017, *Yooprasert et al. PHM 93* (♀) (BKF!, K!); ibid., *Yooprasert et al. PHM 94* (♀) (BKF!, K!); ibid., [8°9′4″N, 98°32′28″E], 73 m, 10 Jun 2017, *Yooprasert et al. PHM 95* (♀) (BKF!, K!); ibid., *Yooprasert et al. PHM 96* (♂) (BKF!, K!); ibid., *Yooprasert et al. PHM 97* (sterile) (BKF!, K!); ibid., [8°30′38″N, 98°32′29″E], 78 m, 10 Jun 2017, *Yooprasert et al. PHM 98* (sterile) (BKF!, K!); ibid., *Yooprasert et al. PHM 98a* (♂) (BKF!, K!); ibid., *Yooprasert et al. PHM 99* (♀) (BKF!, K!); ibid., evergreen forest over limestone along stream [8°28′N, 98°31′E], 50 m, 22 Aug 1999, *Puff 990822-1/1* (♀) (BKF!); Ta Kua Pa District, Hill along new road c. 30 km east from Takua Pa, 100 m, 11 May 1968, *van Beusekom & Phengkhlai 701* (♂) (BKF!, K!); Kao Pub pa, 7 Aug 1975, *Satheesorn 3382* (♀) (BK!); ibid., Khao Lak-Lam Ru NP, KL. Section (6 km trail from HQ to ton chong waterfall) [8°38′N, 98°15′E], 50–150 m, 21 Aug 1999, *Puff 990821-1/4* (♀) (BKF!); ibid., common in shaded lowland evergreen dipterocarp forest by seashore [8°38′N, 98°14′E], 20 m, 9 May 2002, *Pooma et al. 3669* (♀) (BKF!); ibid., [8°3734″N, 98°14′16″E], 20 m, 10 Jun 2017, *Yooprasert et al. PHL 104* (♀) (BKF!, K!); ibid., Ton Chong Fa falls [8°39′18″N, 98°17′01″E], 75 m, 11 Jun 2017, *Yooprasert et al. PHT 108* (♂) (BKF!, K!); ibid., [8°39′26″N, 98°17′02″E], 113 m, 11 Jun 2017, *Yooprasert et al. PHT 105* (♂) (BKF!, K!); ibid., *Yooprasert et al. PHT 115* (♂) (BKF!, K!); ibid., [8°39′26″N, 98°17′01″E], 75 m, 11 Jun 2017, *Yooprasert et al. PHT 110* (♀) (BKF!, K!); ibid., [8°39′18″N, 98°17′02″E], 113 m, 11 Jun 2017, *Yooprasert et al. PHT 121* (♀) (BKF!, K!); Khlong Nang Yon [9°15′N, 98°20′E], 100 m, 30 Apr 1973, *Geesink & Santisuk 5076* (♀) (BKF!, C!, E!, K!); Limestone Hill, east of PhangNga, 30–150 m, 23 Aug 1967, *Shimizu et al. T 7841* (♀) (BKF!); Nai Chong [8°25′N, 98°30′E], <100 m, 11 May 1973, *Geesink & Santisuk 5343* (♀) (BKF!, C!, K!); **PHATTHALUNG**, Chawng, Wang [7°30′N, 99°55′E], 200 m, 14 Apr 1928, *Kerr 15208* (♂) (BK!, BM!, E!, K!); **PHUKET**, Khao Phra tao Non-Hunting area, Gibbon rehabilitation side, Bang Pae and Ton Sai waterfalls [8°02′N, 98°23′E], 50 m, 6 Jun 2004, *Gardner & Sidisunthorn ST 0673* (♂) (BKF!, K!); ibid., 120 m, 21 Apr 2006, *Gardner ST 2604* (♂) (BKF!, K!); **RANONG**, Suksamran District, Aow Tey, (on slope, along sea shore), 0–20 m, 27 Apr 2005, *Phengkhlai 15009* (♀) (BKF!); ibid., Khlong Naka (Nakha) [9°45′N, 98°40′E], 50 m, 22 Jun 1974, *Geesink et al. 7400* (♀) (BKF!, C!, K!); ibid., [9°25′N, 98°25′E], 100 m, 24 Apr 1974, *Larsen 33340* (♂) (K!); ibid., nature trail [9°27′36″N, 98°30′38″E], 51 m, 13 Jun 2017, *Yooprasert et al. RNK 131* (sterile) (BKF!, K!); ibid., *Yooprasert et al. RNK 132* (sterile) (BKF!, K!); ibid., [9°27′34″N, 98°30′40″E], 37 m, 13

Jun 2017, *Yooprasert et al. RNK 133* (sterile) (BKF!, K!); ibid., [9°27′30″N, 98°30′41″E], 53 m, 13 Jun 2017, *Yooprasert et al. RNK 134* (sterile) (BKF!, K!); Boonyapal waterfall, 8 Sep 1984, *Fukuoka et al. T 36007* (♀) (BKF!); Kapoe [Kaper District], 30–150 m, 7 Jul 1993, *Puff and Sridith 930707-1/3* (♀) (PSU!); ibid., Khlong Bang Man, 100–250 m, 15 Jul 1979, *Niyomdham 317* (♀) (BKF!, C!, E!, K!); Khao Saideng, 400 m, 5 Apr 1968, *van Beusekom & Phengkhlai 558* (♂) (C!, E!, K!); Khlong Kam Puang [9°15′N, 98°20′E], 26 Apr 1973, *Geesink 4937* (♂) (AAU!, E!); Kraburi District, Hua Sieng, 18 Apr 1967, *Satheesorn 2280* (♀) (BK!); ibid., Kao pak kua, 20 Apr 1967, *Satheesorn 2318* (♂) (BK!); ibid., Nikhom Pakchan, 26 Apr 1974, *SP. & SS. 525* (♀) (BKF!); ibid., Thungraya Nasak WS, along Bok Krai river from WS headquarter [10°23′N, 98°51′E], 200 m, 28 Aug 2002, *Middleton et al. 1410* (♀) (BKF!, E!, P-photo!); Ngao falls, south of Ranong [9°51′N, 98°39′E], 100–200 m, 7 Aug 1992, *Larsen 43233* (♀) (AAU!); s.loc., 18 Sep 1968, *Phengkhlai 1302* (♀) (BKF!, C!, K!); s.loc., 12 Aug 1973, *Pochanart 422* (♀) (BKF!, C!, K!); **SATUN**, Klawng Ton [6°40′N, 100°10′E], 400 m, 11 Mar 1928, *Kerr 14466* (♂) (BK!, BM!, K!); Kuan Kalong, 5 May 1967, *s.coll. 363* (sterile) (BKF!); ibid., near Nam Rah village, Toong Ngui, 200 m, 26 Aug 1984, *Maxwell 84-144* (♀) (PSU!); Wang Prachan District, Thale Ban NP [6°40′30″N, 100°10′22″E], 70 m, 3 Jun 2001, *Pooma et al. 1961* (♂) (BKF!); ibid., 100–250 m, 13 Sep 1990, *Puff et al. 900913-1/2* (♀) (BKF!); ibid., nature trail [6°42′46″N, 100°10′15″E], 150 m, 19 Jun 2017, *Yooprasert et al. STT 152* (♂) (BKF!, K!); Yar Roy waterfall, c 25 km NE of Satun [6°45′N, 100°7′E], 100–200 m, 6 Nov 1990, *Larsen et al. 41198* (♀) (BKF!); **SONGKHLA**, Had Yai, Ko Hong Hill, west slope, 200 m, 23 Jun 1985, *Maxwell 85-627* (♀) (BKF!, L-photo!); Ton Nga Chang WS, Nature trail west of headquarters [6°56′N, 100°14′E], 140 m, 2 Oct 2004, *Gardner & Sidisunthorn ST 0927* (♀) (BKF!, K!); **SURAT THANI**, 30 km east of Takua Pa, Khao Sok NP [8°55′N, 98°40′E], 100–200 m, 9 Jun 1992, *Larsen 42766* (♀) (AAU!, PSU!); ibid., Along Khlong Sok, Khao Sok NP, 150 m, 17 Jul 2000, *Chamchumroon vc 872* (♀) (BKF!); ibid., 100–200 m, 5 Jul 1993, *Puff and Sridith 930704-1/1* (♀) (PSU!); ibid., Trail to 11 Chan [levels] falls, Phanom District [8°52′N, 101°57′E], 100 m, 28 Mar 1993, *Chantaranothai 1505* (♀) (K!); Khlong Phanom NP, Rafflesia nature trail [8°52′N, 98°32′E], 300 m, 18 Jun 2004, *Gardner & Sidisunthorn ST 0789* (♀) (BKF!); Khlong Sok, 17 Aug 1975, *Prapat 37* (♀) (BKF!, C!, K!); Nasan, 27 Apr 1989, *Chiratanakon s.n.* (♂) (BKF!); **TRANG**, Chaung, 120 m, 14 Sep 1933, *Collins 2369* (♀) (BM!, K!, P-photo!); Foothill of Khao Phra Mi [9°17′N, 98°26′E], 60 m, 7 Aug 1972, *Larsen et al. 30766* (♀) (BKF!, E!, K!, P-photo!); Kachong District, Khao Chong [Khao Kachong] [7°40′N,

99°45′E], 150 m, 14 Jun 1974, *Geesink et al. 7223* (♂) (BKF!, K! [sheet 2]); ibid., 200 m, 11 Aug 1975, *Maxwell 75-744* (♀) (BK!, L-photo!); ibid., 2 Apr 1969, *Sangkhachand 1831* (♂) (BK!); ibid., Forest behind Kachong Bot. Garden, 30–120 m, 11 Sep 1990, *Puff et al. 900911-1/8* (♀) (BKF!); ibid., Khao Chong Botanical Garden, Nayong [7°32′27″N, 99°47′48″E], 149 m, 9 June 2017, *Yooprasert et al. TRC 90* (♂) (BKF!, K!); ibid., *Yooprasert et al. TRC 91* (sterile) (BKF!, K!); s.loc., s.a., *Geesink 5343* (♂) (E!).

HABITAT. Lowland evergreen forest, sometimes mixed with dipterocarps and bamboo; shaded to semi-shaded areas, usually by streams or on steep slopes; elev. 20–800 m.

CONSERVATION STATUS. Least Concern (LC). *Urophyllum longifolium* is widely distributed across India and Myanmar and western to southern Thailand. It is found in many National Parks and Wildlife Sanctuaries and is therefore unlikely to be threatened.

PHENOLOGY. Specimens with flowers have been collected from January to November; fruits collected from February to November.

NOTES. The useful characters to identify *U. longifolium* are the longitudinally folded stipules toward the adaxial side, leaves with secondary veins festooned brochidodromous, compound cymose inflorescences, toothed calyx and glabrous white corolla. Hairs of *U. longifolium* can be variable in length, density and the angle of hairs especially on stipules, leaves and inflorescence bracts. The type specimen, *Griffith 9* (Myiek, Myanmar), and the populations found in Kaeng Krachan National Park (Petchaburi Province) and Krung Ching Falls (Khao Luang National Park, Nakhon Sri Thammarat Province) have long and dense, appressed hairs compared with other populations. Whilst populations found in PhangNga and Phuket Provinces have short and sparse, appressed hairs. The other populations have long, erect hairs varying in their density. This highly variable character has led to three taxa being described, *U. longifolium* var. *longifolium* (dense and long, appressed hairs), *U. longifolium* var. *pilosum* (dense and long, erect hairs) and *U. talangense* (sparse and short, appressed hairs) as discussed in the introduction. However, these differences were not supported by phylogenetic studies using both plastid and nrDNA data (Chapter 3), and all the taxa are synonymised to *U. longifolium* in this study. The morphologically similar species (*U. glabrum*) and sympatric species (*U. crassum*) differ in their characters to *U. longifolium* as discussed in the Notes to those species.

The species was first published by Wight under the genus *Axanthes* as *A. longifolia* and was combined with the genus *Urophyllum* by Hooker in 1880 with specimens mentioned in the protologue to the collection by Griffith from Mergui [Myeik] without a collection number. Furthermore, there is more than a single specimen in that collection. Noltie (2005) considered *Griffith 9*, with Wight's handwritten note '*A. longifolius* RW, 9 Mergui Griffith', at K as the holotype and the other two specimens with a 'Herb. Hookerianum' stamp as isotypes. Noltie (2005) can be classed as typification, but there are misused terms (holotype and isotype) that cannot be treated as 'an error to be corrected' (Art. 9.10 of the ICN Shenzhen Code (2018) as the publication did not include the words 'designated here' or equivalent, as required in Art. 7.11 of the ICN Shenzhen Code (2018). Therefore, a second-step designation is proposed in this study, lectotype and isolectotypes are selected based upon the holotype and isotypes listed in Noltie (2015). Lectotypes of *U. longifolium* var. *pilosum* and *U. talangense* were selected based upon Craib's biography as previously discussed (cf *U. crassum*) that his type specimens were deposited in K and WRSL.

6. ***Urophyllum longipes*** Craib (in Bull. Misc. Inform. Kew 1931: 446). Type: Siam [Thailand], Pattani [Yala], Betong, c. 300 m, 13 Aug 1923, *Kerr 7607* (♀) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!, K!, TCD-photo!).

DISTRIBUTION. Peninsular Malaysia (Kelantan and Perak states) and southern most Thailand (Narathiwat and Yala Provinces) (Figure 4.9).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: KELANTAN, Jeli District, Upper Sungai Pergau, 25 Sep 1986, *Latiff et al. ALM 1797* (♂) (PSU!); ibid., Sungai Renyok junction off Sungai Pergau, 27 Sep 1986, *Latiff et al. ALM 1871* (♀) (PSU!); **PERAK**, Hulu Perak District, Temengor FR, Compt.44, block 5 (3) [5°31'24"N, 101°35'3.5"E], 653 m, 20 May 2010, *Kamarui Hisham et al. FRI 67211* (♀) (K!); ibid., Gunung Batu Puteb [Puteh] [4°13'N, 101°27'E], s.a., *Wray Jr. 223* (♂) (P-photo!); **THAILAND. NARATHIWAT**, Waeng District, Hala-Bala, Khao Sam Sip [5°48'28.7"N, 101°50'42.2"E], 300–500 m, 25 Sep 1996, *Niyomdham 4781* (♀) (BKF!); **YALA**, Betong District, [5°45'N, 101°2'E], 400 m, 1 Aug 1923, *Kerr 7455* (♂) (BK!, BM!, K!); ibid., Hala-Bala Wildlife Sanctuary. Trail to the summit of unnamed '1490' mountain reached from the shores of Bang Lang Reservoir. [5°58N, 101°26'E], 1400 m, 24 May 2005, *Middleton 3670* (♀) (E!); ibid., Than Num Thip, the route start from Sa Ho check dam [5°39 33.9N, 101°10'8"E], 500 m, 21 Jul 2015, *Poopath et al. 1274* (♀) (BKF!); ibid., Ta Noh Mae Roh, the route to Phuk Num plots and to Bukit Lata Papa Lang peak [5°54 9.6N, 101°2'15.4"E], 830 m, 23 Jul 2015, *Poopath et al. 1311* (♀) (BKF!); ibid., Than Num Thip, Nature trail to Than Num Thip waterfall [5°41 19N, 101°9'15.7"E], 300 m, 3 Aug 2016, *Poopath et al. 1604* (♀) (BKF!); ibid., Ta Noh Mae Roh, the road No. Yala-3004, ro to Mai Muang Nao Garden [5°53 35.8N, 101°1'46.3"E], 950 m, 3 Aug 2016, *Poopath et al. 1613* (♂) (BKF!); ibid., Maewat, nature trail to waterfall, along Than Roi Jai stream [6°5 19N, 101°22'49"E], 180 m, 6 Aug 2016, *Poopath et al. 1683* (♀) (BKF!); ibid., Khlong Ka pa, near Ban Chulabhorn Phathana 10 [5°51N, 101°14'38"E], 500 m, 21 Oct 2017, *Wai 2654* (♀) (BKF!); ibid., trail large tree [5°51'28.8"N, 101°14'11"E], 560 m, 13 May 2019, *Yooprasert et al. 214* (♀) (BKF!, K!); ibid., trail large tree [5°51'28.8"N, 101°14'11"E], 560 m, 13 May 2019, *Yooprasert et al. 215* (sterile) (BKF!, K!); ibid., trail large tree [5°51'28.8"N, 101°14'11"E], 560 m, 13 May 2019, *Yooprasert et al. 216* (sterile) (BKF!, K!); ibid., Beside the road entry to Piyamitr 2 village [5°52'40.3"N, 101°1'18"E], 841 m, 14 May 2019, *Yooprasert et al. 220* (♂) (BKF!, K!); Kabang District, Kao Kalakiri, Pattani [Yala]

[6°37'N, 101°6'E], 800 m, 11 Sep 1923, *Kerr 7806* (♀) (BK!, BM!, K!); ibid., 500 m, 31 Mar 1928, *Kerr 14915* (♀) (BK!, BM!, E!, K!).

HABITAT. Montane evergreen forest, sometimes mixed dipterocarp forest; slightly shaded on slopes or by streams; elev. 300–1,400 m.

CONSERVATION STATUS. Near Threatened (NT). *Urophyllum longipes* is found on mountain ranges around the Thailand-Malaysia border including protected (IUCN category Ia and II) and non-protected areas where there is no sign of agricultural expansion reaching the area to date (according to Google Earth imagery); the small number of locations (9 locations), the estimated EOO (5,878 km^2) and AOO (48 km^2) of *Urophyllum longipes* indicate the species should be rated as Vulnerable (VU). As the species is not at risk however, it should be assessed as Near Threatened.

PHENOLOGY. Collected in flower from May to August; collected in fruit from August to March.

NOTES. In addition to the characters discussed previously (cf *U. hirsutum*), *U. longipes* differs from other species in Thailand by having a folded stipule, obovate to elliptic coriaceous leaves, abaxially shiny, secondary veins festooned brochidodromous, tertiary veins and veinlets inconspicous, calyx truncate to toothed, and corolla glabrous, white in colour. A morphologically similar species is *U. griffithianum* (Wight) Hook.f. which resembles *U. longipes* in appearance except *U. griffithianum* is not found in Thailand and differs by its wide stipules, appressed to the stem but not folded, and pedunculate umbellate inflorescence. Other species found in the same area in Thailand to *U. longipes* are *U. streptopodium* and *U. villosum*. These two species are morphologically different to *U. longipes* as both have densely hairy shoots, stipules and abaxial leaf surface and stipules are appressed to the stem but not folded. More detailed characters for these species can be found in the Notes to those species. The lectotype of *U. longipes* was selected for the reasons discussed previously (cf *U. crassum, U. glabrum* and *U. longifolium*).

7. *Urophyllum macrophyllum* (Blume) Korth. (Korthals 1851: 194). *Axanthes macrophylla* Blume (1826-1827: 1002). Type: [Indonesia], Java, s.a., *Blume 1452* (♀) (lectotype, designated here L(photo seen) [L 0820395]).

DISTRIBUTION. Thailand (Narathiwat Province), Peninsular Malaysia (Negeri sembilan, Pahang, Perak, Selangor and Terengganu states), and Indonesia (Java and Sumatra) (Figure 4.10).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: Kelam, Tujin, Jun 1888, *Wray Jr. 2901* (♀) (K!); s.loc., Mar 1884, *Father Scortechini 267* (♀) (K!); s.loc., May 1884, *Father Scortechini 715* (♀) (K!); **NEGERI SEMBILAN**, Jelebu District, Berembun FR, Bukit Lantai, about 4.4 km on the road to summit Gunung Telapak Buruk [2°50.2'N, 102°1.89'E], 840 m, 8 Apr 2008, *Siti Mastura et al. FRI 66505* (♀) (K!); ibid., Jelebu FR [2°59'N, 102°4'E], 2000 f [c. 610 m], 7 Mar 1969, *Suppiah FRI 11284* (♀) (K!); **PAHANG**, Jerantut District, confluence of Sg. Tekam and Sg. Balol [Balul] [3°58'58.48"N, 102°35'5.13"E], 200 f [c. 60 m], 25 Jun 1972, *Ng & Beltran FRI 6385* (♀) (K!); ibid., S. [Sg.] Tembeling NR, K. Keniyum [Keniyam], S. [Sg.] Redab [4°29'50''N, 102°28'44''E], 3 Jun 1968, *Whitmore FRI 8564* (sterile) (K!); ibid., Taman Negara, plot 1, along Sungai Tahan trail c. 1.5 km from Kuala Tahan H.Q. [4°23'N, 102°24'E], 18 Apr 1975, *Balgooy 2444* (♀) (AAU!); ibid., Tembeling Valley at K. Trenggan [Terengan] [4°26'N, 102°26'E], 17 Aug 1982, *Wong FRI 32608* (♀) (K!); ibid., Teku river, Gunung Tahan [4°34'46.20"N, 102°15'6.29"E], 2 Jun 1922, *Harilt 8072* (♀) (K!); Raub District, Fraser's Hill, Bishop's trail [3°43'N, 101°45'E], 2000 m, 13 Sep 1992, *Chua FRI 39001* (♀) (K!); ibid., near Richmond [Bungalow] [3°43'N, 101°45'E], 25 Apr 1955, *Purseglove P 4325* (♂) (K!); ibid., Selangor residency Bangalow [3°43'N, 101°45'E], 4000 f [c. 1220 m], 28 Sep 1959, *Shah & Kadim MS 706* (♀) (BKF!, E!, K!); Temerloh District, Krau GR [Wildlife Reserve], Sg. Lompat, Ulu Sg. Lompat [3°46'26''N, 102°6'E], 6 Feb 2000, *Damanhuri & Ayau FRI 45356* (♀) (K!); ibid., Kuala Lompat, Lata Tujuh, Trail to Batu Begambar [3°49'2"N, 102°7'43"E], 12 Jul 2007, *Mohd. Hairul et al. FRI 58916* (♂) (L-photo!); Telua Ruie, 1 Jun 1901, *Ridley 13907* (♀) (K!); **PERAK**, Hulu Perak District, Temengor FR, Compt. 44, block 5 (3) [5°31'26"N, 101°35'34"E], 20 May 2010, *Kamarul Hisham et al. FRI 67209* (♀) (L-photo!); Kinta District, Gunong Kerbau [Korbu] [4°41'N, 101°17'E], 20 Mar 1913, *Robinson 4200* (♂) (K!); ibid., Sg. Groh [Geroh], hill east of Gopeng [4°27'19''N, 101°13'7''E], 1500 f [c. 460 m], 10 Mar 1966, *Ng FRI 1587* (♀) (K!); Kuala Kangsar District, Bubu FR, S. Gading, first base

camp. [4°41'4"N, 100°50'47"E], 7 Jul 2009, *Imin et al. FRI 68228* (♀) (L-photo!); ibid., Sungei Guar [4°39'N, 100°50'E], 2200 f [c. 670 m], 18 Aug 1966, *Whitmore FRI 0679* (♀) (K!); Larut Matang District, Larut, 400 f [c. 120 m], Apr 1882, *Hustler 2948* (♀) (K!, P-photo!); s.loc., Jan 1886, *Dr. King's collector 10733* (♀) (K!); s.loc., 300 f [c. 90 m], May 1889, *Wray Jr. 3511* (♀) (K!); **SELANGOR**, Gombak District, Guiting Bidai Huk pil, 23 May 1896, *Ridley 1440* (♂) (K!); Hulu Langat District, Sungei Lalang Kajang S, s.a., *Symingtan 22741* (♂) (K!); **TERENGGANU**, Dungun District, Jerangau FR, Compt 95 [4°48'3"N, 103°5'1"E], 27 Jan 2010, *Kamarul Hisham et al. FRI 67179* (♂) (L-photo!); Hulu Terengganu District, Gunong Padang expedition, Ulu Brang, Camp. 1 near K. Lallang [4°51'N, 102°52'E], 300 f [c. 90 m], 15 Sep 1969, *Whitmore FRI 12532* (♀) (K!); **THAILAND. NARATHIWAT**, Sukhirin District, Hala-Bala WS, trail to To Mo mine [5°48'9"N, 101°42'40"E], 200 m, 21 Jul 2004, *Pooma et al. 4491* (♀) (E!); Waeng District, Hala-Bala [5°48'31.7"N, 101°50'41.8"E], 300 m, 22 Mar 2000, *Niyomdham et al. 6105* (♀) (BKF!); ibid., Khao Bo Lue Sa, 5 Aug 1999, *Puudjaa & Cholkulchana 643* (♀) (BKF!); ibid., NikhomWang [Nikhom Waeng] [5°52'6.5"N, 101°52'47.5"E], 3 Apr 1968, *Sangkhachand 1274* (♀) (BK!); ibid., s.loc., 23 Mar 1968, *Phusomsaeng 403* (♀) (BKF!, K!); ibid., s.loc., 1 Apr 1968, *Phusomsaeng 439* (♂) (BKF!).

HABITAT. Evergreen forest and lowland dipterocarp forest, sometimes mixed with bamboo; by streams, rocky, rich soil, or on slopes; elev. 90–1,220 m.

REGIONAL CONSERVATION STATUS. Least Concern (LC). The estimated EOO for *Urophyllum macrophyllum* is greater than 40,000 km^2 and it is found from 14 locations (nine localities in National Parks, State Parks and Wildlife Reserves (IUCN categories II and Ia) and seven locations with collections after the year 2000), thus the species does not currently qualify as threatened and is rated as Least Concern at regional level. It should be noted that its lowland habitat in Forest Reserves (e.g., Bubu and Kledang Saiong Water Catchment Forests (IUCN category VI)) in Peninsular Malaysia means there is a potential risk of habitat loss from agriculture expansion in the region, as observed using Google Earth imagery from the last five years.

PHENOLOGY. Collected in flower from January to July; in fruit from February to September.

NOTES. *Urophyllum macrophyllum* is recognisable by its longitudinally folded stipules, chartaceous obovate leaves, very delicate when dried and usually found torn and damaged in herbarium specimens; (11–)13–18 secondary veins pairs, loops festooned brochidodromous, tertiary veins well-spaced (>2 mm apart), obtuse divergence angle of tertiary veins relative to the midrib; sessile umbellate inflorescence, lobed calyx, and glabrous corolla. It is morphologically similar to *U. streptopodium* with phylogenetic analysis showing they are sister species (Chapter 3). The two species share several characters including being appressed densely hairy in most parts, sessile umbellate inflorescences and flowers with a lobed calyx. Differing characters of *U. streptopodium* are appressed to the stem but not folded stipules, coriaceous leaves, usually elliptic; 5–10 secondary vein pairs, loops weakly brochidodromous, and tertiary veins closely spaced (<2 mm) with angle to the midrib perpendicular to subperpendicular.

The lectotype designation is based upon Blume's biography on Taxonomic Literature II (Stafleu and Mennega, 1993) detailing his original materials are deposited at L and the specimen, Blume 1452 [L 0820395], has the determinavit slip written by Blume to *Axanthes macrophylla* and the type label to this species on it.

8. ***Urophyllum schmidtii*** C.B.Clarke (1902: 334). Type: Koh Chang [Trat Province, Thailand], Klong Son, 1000 ft [c. 300 m], *Schmidt 664* (♀) (lectotype, designated here C-photo!; isolectotype M-photo!)

Urophyllum olivaceum Craib (1931: 447), **synon. nov.** Type: Siam [Thailand], Chantabun [Chanthaburi], Kao Sabap, 5 Jul 1927, *Put 903* (♀) (lectotype, designated here K!; isolectotypes ABD!, BK!, BM!)

Urophyllum longifolium var. *annamense* Pierre ex Pit. (Pitard in Lecomte *et al.* 1923: 202), pro parte, only specimen of *Pierre 1251* (♀) (BKF!, C!, E!, K!, L-photo!, P-photo!).

DISTRIBUTION. Cambodia (Kampong Speu and Koh Kong Provinces) and eastern Thailand (Chanthaburi and Trat Provinces) (Figure 4.8).

SPECIMENS EXAMINED. CAMBODIA. KAMPONG SPEU, Aural [Aoral District] Phnom Aural, trail above camp 1 [11°52'40"N, 104°10'05"E], 12 Feb 2001, *Boyce MoE 454* (♀) (K!); ibid., trail above camp 2 [11°52'40"N, 104°10'05"E], 17 Feb 2001, *Boyce MoE 530* (♀) (K!); Tpong [Thpong] District, in Mt Knang Repeu, May 1870, *Pierre 1251* (♀) (BKF!, C!, E!, K!, L-photo!, P-photo!); **KOH KONG**, Cardamom mountain [11°55'53.7"N, 103°36'02.3"E], 1250 m, 25 Mar 2000, *Eanghourt 69* (♂) (K!); ibid., Phnom Koh Khchang, Koh Khchang village [11°41'25.4"N, 103°29'21.3"E], 485 m, 27 Mar 2000, *Eanghourt 75* (♂) (K!); ibid., pres de Stung Tauck (Cardamones centrales) [12°00'00"N, 103°15'00"E], 1 Feb 1970, *Mane 1711* (♂) (P-photo!); ibid., Phnom Rodam Muoy Daeum, 23 Feb 1966, *Martin 347* (♀) (P-photo!); ibid., Thma Baing District, Russei Chrum Commune, Trapeang Chheu Trao village [11°41'34.1"N, 103°29'20"E], 532 m, 17 May 2010, *Newman et al. 2328* (♀) (E!); ibid., Central Cardamon [11°45'12.94"N, 103°29'36.11"E], 495 m, 21 Apr 2011, *Toyama et al. 812* (♀) (FU!); **THAILAND. CHANTHABURI**, Pong Nam Ron District, Khao Soi dao WS, Khao Soi dao Tai [12°55'N, 102°14'E], 590 m, 13 Jul 2008, *Phonsena et al. 6112* (♀) (BKF!); ibid., Khao Sabap [12°31'04"N, 102°12'30"E], 5 Jul 1827, *Put 903* (♀) (BK!, BM!, K!); ibid., Ang-hong falls, Kao Sra-bap, Pliew experiment station [12°31'46"N, 102°11'02"E], 18 Jun 1976, *Vacharee 126* (♀) (BK!); **TRAT**, Koh Chang, Aw Ong Kang [10°20'N, 102°40'E], 30 m, 7 May 1974, *Geesink et al. 6592* (♀) (BKF!, C!, K!).

HABITAT. Evergreen rain forest, sometimes on granitic hills; in the Cardamom Mountain Ranges; elev. 30–1,250 m.

CONSERVATION STATUS. Least Concern (LC). The status is based upon *Urophyllum schmidtii* being known from seven localities (13 specimens) with the estimated EOO as 13,060 km^2 and AOO as 36 km^2. Google Earth imagery identify that the areas surrounding one location (Phnom Aural, Kampong Speu Province, Cambodia) has changed to agricultural areas within the last decade. This will affect the habitat quality and loss. However, the other six localities are within protected areas, therefore the species is rated as Least Concern.

PHENOLOGY. Collected in flower from February to May; in fruit from May to July.

NOTES. *Urophyllum schmidtii* is the only species found in eastern Thailand. It has stipules appressed to the stem but not folded, sometimes glabrous pocket domatia are present at

the axils of branching between secondary veins and the midrib abaxially, divergence angle of tertiary veins relative to the midrib obtuse, calyx lobed and corolla white, glabrous abaxially. It is similar to *U. pseudoschmidtii* but *U. schmidtii* has pedunculate umbellate inflorescence and pistillate flower with the presence of staminodes, whereas the inflorescence of *U. pseudoschmidtii* is variable from sessile to pedunculate umbellate, sometimes two-tiered cymose (only in pistillate plants), and lacks staminodes in pistillate flowers.

Craib (1931) described *Urophyllum olivaceum* based on a specimen collected from Khao Sabap, Chanthaburi Province being subglabrous throughout, having longer leaves, and leaves usually olive green when dried. However, without more available specimens, these characters fall within the variation of *U. schmidtii*. The leaves of the specimens collected from Kampong Speu are occasionally shorter and have denser hairs than specimens from Koh Kong and Chanthaburi. The colour of leaves when dried can often be a result of the method used in the drying process (e.g., specimens preserved with 70% alcohol before drying usually have a darker colour than specimens dried in paper). Therefore, these morphological characters could not be used to distinguish the two species based upon herbarium material, it is therefore recognised as *U. schmidtii* in this study.

9. ***Urophyllum streptopodium*** Wall. ex Hook.f. (Hooker 1880: 99). Type: [Malaysia, Peninsular Malaysia], Penang, s.a., *Wallich s.n.* [EIC 8317] (♀) (lectotype, designated by Wong *et al.* (2019) K! [K000740838]; isolectotypes BR-photo! [BR0000005620913], CAS-photo! [CAS0005344], K-W! [K001125241(element on the leftmost)])

DISTRIBUTION. Indonesia, Peninsular Malaysia, Sabah and Sarawak, Singapore and southernmost Thailand (Yala and Narathiwat Provinces) (Figure 4.7).

SPECIMENS EXAMINED. BIRMA AND MALAY PENINSULA. s.loc., s.a., *Griffith 2941* (♂) (P-photo!); ibid., s.loc., s.a., *Griffith 2941a* (♂) (K!); ibid., s.loc., s.a., *Griffith 2941b* (♀) (K!); **MALAYSIA. PENINSULAR MALAYSIA:** Pasoh FR, 320 f [c. 100 m], 25 Jul 1980, *Wong FRI 28928* (♀) (K!); s.loc., s.a., *Wallich 8317 [specimen on the left]* (♀) (K-W!); **JOHOR**, Batu Pahat District, s.loc. [1°50'N, 102°56'E], s.a., *Hullett 505* (♀) (K!); ibid., Gunong Pulai [1°36'N, 103°33'E], 2000 f [c. 600 m], 1 Oct 1956, *Purseglove P 5510* (♀) (BKF!, K!, L-photo!); ibid., 1100 f [c. 340 m], 9 Mar 1971, *Chan FRI 17640* (♀) (K!); Kluang District, Hutan Lipur, Gunong Belumut, trail to summit [2°00'N, 103°31'82"E], 91 m, 4 May 2011, *Imin FRI 74678* (♀) (K!); Kota Tinggi District, 5.5 miles Kota Tinggi-Mawai Rd [1°47'14.8"N, 103°56'35.6"E], 2 Feb 1935, *Corner 28699* (sterile) (K!); ibid., Bukit Tinjau Laut [1°57'N, 103°57'E], 8 Apr 1939, *Ngadiman SING 36943* (♀) (K!); ibid., Gunong Panti FR [1°49'N, 103°51'E], 7 May 1970, *Samsuri S 326* (♀) (K!, L-photo!); ibid., 14 Jun 1981, *Maxwell 81-141* (♀) (L-photo!); ibid., 350 f [c. 110 m], 2 Mar 1980, *Vethevelu FRI 25278* (♂) (K!); ibid., Gunong Panti West [1°49'N, 103°52'E], 8 Apr 1977, *Maxwell 77-184* (♀) (L-photo!); ibid., S. Kayu Ara, Mawai-Jemaluang Rd. [1°59'53''N, 103°51'36''E], 23 Jun 1935, *Corner 29480* (♀) (K!); Mersing District, Compt. 90, Arong FR [2°34'N, 103°48'E], 25 Apr 1967, *Ng FRI 5208* (♂) (K!); ibid., Telok Ayer Papan [2°30'N, 103°50'E], 17 Apr 1977, *Maxwell 77-193* (♂) (L-photo!); Segamat District, Endau river, Labis FR [2°30'29.3"N, 103°28'28.1"E], 23 Jul 1977, *Maxwell 77-352* (♀) (AAU!); ibid., Labis FR [2°25'N, 103°21'E], 700 f [c. 210 m], 4 May 1972, *Chan FRI 19968* (♂) (K!); ibid., Compt 81, 500 f [c. 150 m], 14 Apr 1967, *Suppiah KEP 104970* (♀) (K!, L-photo!); **KEDAH**, Kuala Muda District, Kedah peak [Gunung Jerai] [5°47'N, 100°26'E], 2800–4000 f [c. 850–1220 m], Dec 1915, *Robinson & Kloss s.n.* (sterile) (K!); **KELANTAN**, Gua Musang District, Kuala Betis track [4°54'N, 101°48'E], 18 Jul 1935, *Henderson SING 29723* (♀) (K!); ibid., Sungai Lebir, 2 miles E. Kuala Aring [5°02'N, 102°23'E], 13 Sep 1967, *Cockburn FRI 7113* (♀) (K!); ibid., Sg. Brok [Berok], Ulu Kelantan, 800 f [c. 240 m], 6 Oct 1967, *Ng FRI*

5348 (♀) (K!); Kuala Krai District, Relai VJR., Machang [5°18'N, 102°12'E], 2000 f [c. 600 m], 7 Sep 1987, *Khairuddin & Damanhuri FRI 31962* (♀) (K!); **MALACCA**, s.loc., s.a., *Griffith s.n. [AAU 2690, specimen on the left]* (♀) (AAU!); s.loc., s.a., *Griffith 2942* (♀) (K!); s.loc., s.a., *Griffith s.n.* (♀) (K!); s.loc., Aug 1886, *Hervey s.n.* (♂) (K!); **NEGERI SEMBILAN**, Seremban District, Gunong Angsi FR, Pedas [2°44'N, 102°03'E], 1500 f [c. 460 m], 18 Feb 1971, *bin Sohadi FRI 14616* (♂) (K!); **PAHANG**, Jerantut District, Gunung Aais Forest Reserve, Compt. 141, along the vicinity of base camp and Sungai Jeram Perahu trail, 160 m, 8 Jul 2004, *Chung & Angan RC 153* (♀) (BKF!, L-photo!); ibid.,Taman Negara, Sg. Tanam [4°39'N, 102°9'E], 30 m, 8 May 1997, *Chua et al. FRI 40647* (♀) (K!, L-photo!); ibid., Ulu Cheka, Benom forest, quadrat 4, tree 15 [3°53'N, 102°12'E], 13 Jun 1968, *T. & P. 106* (♀) (K!); ibid., Bukit Terom, Ulu Keniyam [4°32'N, 102°29'E], 1000–2000 f, 3 May 1968, *Shah MS 1561* (♂) (K!, L-photo!); Kuantan District, Gunung Tapis [4°01'N, 102°55'E], 13 Jun 1934, *Symingtan & Kiah SING 28810* (♂) (K!); Lipis District, Sg. Telom rigde N. of Sg. Kadjau, 500 f [c. 150 m], 27 May 1971, *bin Sohadi FRI 14745* (♀) (K!); Raub District, G. Benom Game Reserve, Ulu Krau [3°49'N, 102°01'E], 2300 f [c. 700 m], 19 Apr 1967, *bin Ismail KEP 97814* (♀) (K!, L-photo!); ibid., 1900 f [c. 580 m], 20 Apr 1967, *bin Yusoff KEP 99113* (♂) (K!, L-photo!); ibid., 13 Jul 1967, *Chelliah KEP 104406* (♀) (K!); ibid., Main NE ridge, boundary Krau Game reserve [3°49'N, 102°01'E], 4000 f [c. 1220 m], 16 Mar 1967, *Whitmore FRI 3222* (♀) (K!); ibid., State land [3°45'N, 101°45'E], 1200 f [c. 370 m], 22 Mar 1971, *bin Sohadi FRI 14679* (♂) (K!); Rompin District, Tanjay Dratak, Pulau Tinman [Pulau Tioman] [2°48'N, 104°11'E], 28 Jun 1915, *Sulleie 1121* (♀) (K!); **PENANG**, Timur Laut District, Bukit Penara Forest Reserve, trail to Telekom Satation [5°22'51"N, 100°15'42"E], 18 May 2006, *Imin et al. FRI 50711* (♀) (K!, L-photo!); ibid., Government Hill [5°25'N, 100°16'E], Feb 1867, *Maingay 2212* (♀) (K!); ibid., Penang Hill [5°26'N, 100°16'E], 31 Jan 1921, *Ridley s.n.* (♀) (K!); Pulo-Pinang [Penang Island], 20 Jan 1905, *Wallich 8320* (♂) (P-photo!); Richmond Pool, Mar 1915, *Ridley s.n.* (♂) (K!); s.loc., s.a., *Maingay 882 [specimen on the left]* (♂) (K!); s.loc., s.a., *Maingay 882 [specimen on the right]* (♀) (K!); s.loc., 24 Feb 1905, *Wallich 9067 [specimen on the right]* (♂) (K-W!); **PERAK**, Batang Padang District, Behrang FR, main range [3°46'N, 101°34'E], 1500 f [c. 460 m], 30 Nov 1966, *Ng FRI 1783* (♀) (K!); ibid., Trolak FR [3°60'N, 101°23'E], 18 Mar 1967, *Chelliah KEP 104605* (♀) (K!); Hulu Perak District, Compt. 57, Papulut FR, Grik [5°18'N, 101°06'E], 500 f [c. 150 m], 7 Jul 1966, *Chelliah KEP 98604* (♀) (K!); Kinta District, Goping [Gopeng], Larut [4°28'N, 101°10'E], 300–500 f [c. 90–150 m], Apr

1884, *Dr. King's collector 5784* (♂) (K!, P-photo!); Kuala Kangsar District, Gunong Bubu [4°40'N, 100°50'E], 2000 f [c. 600 m], 14 Aug 1966, *Chew CWL 1188* (♀) (K!, L-photo!); Larut Matang District, Maxwell's Hill [Bukit Larut] [4°51'N, 100°47'E], 4 Mar 1965, *Hardial & bin Samsuri 289* (♀) (K!, L-photo!); s.loc., May 1884, *Scortechini 671* (♀) (K!); **SELANGOR**, Gombak District, Bukit Lagong FR, K.L. [3°15'N, 101°36'E], 8 Aug 1961, *Yong KEP 98282* (♀) (K!, L-photo!); ibid., Kepong [3°15'N, 101°36'E], 7500 f [c. 2286 m], 6 Mar 1962, *bin Rajab 347* (♀) (K!); ibid., Gunong Bunga Buah [3°22'N, 101°44'E], 2800 f [c. 850 m], 28 May 1966, *Whitmore FRI 0325* (♀) (K!); ibid., Ulu Gombak, UMFSC [3°19'N, 101°46'E], 900 f [c. 270 m], May 1971, *3rd year student s.n. [AAU2689]* (♀) (AAU!); ibid., Virgin Jungle FR [3°19'N, 101°46'E], 450 m, 24 Aug 1966, *Hou 680* (♀) (K!); Hulu Selangor District, Bukit Kutu [3°31'N, 101°43'E], 20 Jun 1896, *Ridley 7439* (♂) (K!); **TERENGGANU**, Dungun District, Bt. Bauk FR [4°42'N, 103°25'E], 15 May 1976, *Chan FRI 25059* (♀) (K!, L-photo!); ibid., Jengai FR, cpt. 77 [4°30'17"N, 102°59'35"E], 80 m, 13 Jul 2010, *Julius & Mohd.-Nazri FRI 57756* (♀) (K!); **SINGAPORE.** Bukit Timah FR, 27 May 1948, *Sinclair 4782* (♀) (E!); ibid., 23 Oct 1967, *Hardial 629* (♀) (L-photo!); ibid., 9 Jul 1959, *Shah MS 750* (♀) (E!); Pulo Obris, Mar 1885, *Hullett 865* (♀) (K!); Upper MacRitchie Resevoir area, Island Club evegreen forest., 4 Jun 1981, *Maxwell 81-116* (♀) (L-photo!); **THAILAND. NARATHIWAT**, Sukhirin District, Bala-Hala, 4 Dec 1997, *Niyomdham 4992* (♀) (AAU!); **YALA**, Betong District, Rubber tree trail, Malaysia border road No.4266 Km [5°38'37.5"N, 101°7'50.1"E], 750 m, 13 May 2019, *Yooprasert et al. 210* (sterile) (BKF!, K!); ibid., trail large tree [5°51'28.8"N, 101°14'11"E], 560 m, 13 May 2019, *Yooprasert et al. 212* (♂) (BKF!, K!).

HABITAT. Lowland dipterocarp forest to montane forest; slightly shaded areas; on clay; by streams, on slopes, or along ridge top; elev. 30–2,300 m.

REGIONAL CONSERVATION STATUS. Least Concern (LC). In Thailand *Urophyllum streptopodium* is found in only one location (Betong District, Yala Province); however, the species is relatively widespread from Peninsular Malaysia to Singapore with the large estimated EOO (>100,000 km^2). Furthermore, it is found in 26 locations with several recent collections from five locations, it has a wide range of ecological requirements as it is found from lowland to montane forest with elevations ranging from 30–2,300 m., and habitat loss is therefore likely to be reduced. The only location that currently shows a risk of habitat loss is Gua Musang District, Kelantan State but the vegetation seems undisturbed

at other localities according to Google Earth imagery. The species is therefore not identified a threatened within this study and rated as Least Concern at the regional level.

PHENOLOGY. Collected in flower from January to November; in fruit whole year round.

NOTES. In addition to the characters discussed previously (cf *U. blumeanum* and *U. macrophyllum*), *U. streptopodium* has a superposed axillary inflorescence, lobed calyx, glabrous corolla, pale yellow at base then green toward apex and lacks a membrane at adaxial lobes base. The species is easy recognised by its coriaceous leaves, yellowish colour when dried, tertiary veins closely spaced with perpendicular divergence angle relative to the midrib.

The isolectotypes designated by Wong *et al.* (2019), *Wallich s.n.* [EIC 8317] (K-W!), with barcode K001125241 (element on the right and centre) was not included here in the type list as the elements being a staminate plant implying that it was not collected from the same plant as the lectotype (pistillate plant). The specimens with barcode K001125239 and K001125240 were also not included, as none of the elements on the sheet match the characters to be identified as *U. streptopodium*.

10. *Urophyllum trifurcum* H.Pearson ex King & Gamble (1904: 194) Type: [Peninsular Malaysia], Pahang, Pekan, 4 May 1890, *Ridley 1180* (♂) (lectotype, cited by Wong (2018) SING-photo!; isolectotype K!).

DISTRIBUTION. Throughout Peninsular Malaysia and Narathiwat Province in Thailand (Figure 4.11).

SPECIMENS EXAMINED. MALAYSIA. PENINSULAR MALAYSIA: JOHOR, Kuala Sembrong, 6 Mar 1905, *Luke & Kelsall 4084* (♂) (K!); **KELANTAN**, Bukit Batu Papan, Sugnai Lebir, 500 f [c. 150 m], 7 Apr 1935, *Henderson SING 29506* (♀) (K!); Machang District, Ulu Sat FR, hillside [5°42'N, 102°20'E], 17 Jun 1968, *Suppiah KEP 104578* (♀) (K!); **NEGERI SEMBILAN**, Seremban District, Gn. Angsi, Compt. 8 [2°44'N, 102°03'E], 1500 f [c. 460 m], 16 Feb 1971, *Loh FRI 17301* (♀) (K!); Jelebu District, 9th ml. Seremban to K. Klawang [2°50'56.3"N, 102°00'02.9"E], 21 Jan 1968, *Ismail KEP 109425* (♂) (K!); Seremban District, Gunong Telapak Burok [2°44'N, 102°03'E], 27 Mar 1977, *Maxwell 77-150* (♀) (L-photo!); **PAHANG**, Jalan Bukit Tersek towards Simpon FR, Taman Negara, 29 Apr 1975, *T. & P. 557*

(♀) (K!); Taman Negara [4°35'N, 102°25'E], 500 f [c. 150 m], 15 Apr 1975, *Ang FRI 23303* (♂) (K!); ibid., 600 f [c. 180 m], 18 Apr 1975, *Ang FRI 23304* (♂) (K!); ibid., 1 Mar 1983, *Weber 168* (♀) (L-photo!); ibid., Kuala Kenyam [Keniyam], low undulating forest along Sg. Kenyam [Keniyam] [4°31'N, 102°28'E], 6 Dec 1971, *Whitmore FRI 20161* (♀) (K!); Teluu, 30 Mar 1909, *s.coll. 13647* (♀) (K!); Bentong District, 20 km from Karak towards Maneis Seremban [3°24'00"N, 102°02'00"E], 2 Sep 1982, *T. & P. 704* (sterile) (K!); ibid., Sabai Estate, near Bontong [Bentong] [3°20'N, 102°05'E], 28 Jan 1958, *Shah 183* (♂) (BKF!); Cameron Highlands District, Cameron Highlands [4°27'N, 101°28'E], 4000 f [c. 1220 m], 11 Apr 1937, *Nur SING 32612* (♂) (K!, L-photo!, P-photo!); ibid., Sungei Boh valley [4°27'N, 101°27'E], 3500 f [c. 1070 m], 10 Sep 1963, *Chew 887* (♀) (K!, L-photo!); Jerantut District, Sungai Belar, Ulu Sungai Tembeling [4°18'02''N, 102°55'20''E], 7 Mar 1968, *Shah MS 1620* (♀) (K!, L-photo!); ibid., Taman Negara Expedition, 1.5 miles West of Kuala Yong on Southern side of Sungai Yong [4°22'N, 102°23'E], 120 m, 7 Sep 1970, *Everett FRI 14358* (♂) (K!); ibid., Sungei Riul, 4 miles directly west of S. Tembiling, 2 miles south of Bukit [4°22'07''N, 102°20'37''E], 60 m, 7 Dec 1970, *Everett FRI 14427* (♀) (K!); ibid., Kuala Tahan [4°23'N, 102°24'E], 6 Jul 1982, *T. & P. 766* (♀) (K!); ibid., riverine forest, trail along Sungai Tahan [4°30'13.9"N, 102°20'33.2"E], 80 m, 16 Apr 1975, *Balgooy 2418* (♀) (AAU!); ibid., Tembeling Valley at K. Trenggan [Terengan] [4°26'N, 102°26'E], 17 Aug 1982, *Wong FRI 32610* (♀) (K!); ibid., Ulu Sungai Sat [4°45'50''N, 102°38'10''E], 13 Jul 1970, *Shah & Noor 1861* (♀) (L-photo!); Lipis District, NE Kuala Mesong, Sg. Telom, NW Pahang [4°29'N, 101°39'E], 700 f [c. 210 m], 30 May 1971, *bin Sohadi FRI 17868* (♂) (K!, L-photo!); Raub District, Raub State land [3°45'N, 101°45'E], 1200 f [c. 370 m], 22 Mar 1971, *bin Sohadi FRI 14668* (♀) (K!, L-photo!); Rompin District, Labong Endau [2°36'N, 103°36'E], 1 Aug 1917, *Evan s.n.* (♂) (K!); Temerloh District, 6 miles N-W of Temerloh, 30 Oct 1961, *Chew & Noor CWL 283* (♂) (K!); ibid., Krau Game reserve [Krau WR], Kuala Lompat [3°45'N, 102°19'E], 70 m, 29 Jun 1988, *Saw FRI 36288* (♀) (K!); ibid., 15 Apr 1967, *Whitmore FRI 3519* (♀) (K!); ibid., 16 Apr 1967, *Whitmore FRI 3561* (♀) (K!); ibid., 28 Jul 1994, *Zainudin et al. AZ 5184* (♀) (K!, L-photo!); ibid., Sungai Mai estate, Jenerak Halt., railway track [3°47'49.68"N, 102°21'37.77"E], 28 Mar 1959, *Kadim & Mahmud K 45* (♂) (K!, L-photo!); ibid., Ulu Sungai Krau, NE Gunong Benom [3°50'49''N, 102°08'53''E], 800 f [c. 240 m], 3 Jan 1967, *Whitmore FRI 3126* (♂) (K!); **PERAK**, Batang Padang District, Tapah Hills FR [4°14'N, 101°22'E], 4 Apr 1971, *Loh FRI 17385* (♂) (K!); s.loc., s.a., *Scortechini s.n.* (♂) (K!); **SELANGOR**, Kepong Belukar, 25 Nov 1930,

Foxworthy & Whitty 16915 (♂) (P-photo!); lower side of Kluang VJR, 21 Oct 1969, *Kochummen FRI 2847* (♂) (K!); Gombak District, Bt. Lagong FR, Kepong [3°15'N, 101°36'E], 500 f [c. 150 m], 1 Apr 1967, *Pagi KEP 99275* (♀) (K!); ibid., 800 f [c. 240 m], 21 Nov 1959, *Kochummen KEP 93490* (♀) (BKF!, K!, L-photo!); ibid., Guiting Bidai [3°18'N, 101°49'E], 8 May 1896, *Ridley 7441* (♂) (K!); ibid., Ulu Gombak [3°19'N, 101°46'E], 2000 f [c. 610 m], 4 Nov 1965, *Whitmore KEP 115657* (♂) (K!); ibid., Ulu Gombak Rd. [3°20'52''N, 101°46'47''E], 1 Mar 1915, *Ridley s.n.* (♀) (K!); ibid., Ulu Gombak, 20 miles [3°19'N, 101°46'E], 24 Jun 1971, *T. & P. 366* (♂) (K!); Hulu Langat District, Ulu Langat [3°05'N, 101°49'E], 5 Jul 1969, *Suppiah FRI 11265* (♀) (K!); ibid., 1 Mar 1966, *Whitmore FRI 0102* (♀) (K!); ibid., Gading FR [3°40'N, 101°38'E], 19 Jul 1969, *Loh FRI 13367* (♀) (K!); Hulu Selangor District, Semangkok pass [3°41'N, 101°45'E], 21 Jan 1921, *Ridley s.n.* (♂) (K!); Kuala Lumpur District, FRI Kepong, about 200 yds. North of swimming pool [3°14'07.2"N, 101°38'03.4"E], 24 Jun 1969, *Selvaraj FRI 11176* (♀) (K!); **TERENGGANU**, Besut District, Gunong Tebu [5°34'N, 102°36'E], 2000 f [c. 610 m], 7 Jun 1969, *Selvaraj FRI 11200* (♀) (K!); Setiu District, Bukit Kesing [5°18'N, 102°52'E], 28 Jul 1983, *Penomat & Teo KL 4271* (sterile) (P-photo!); **THAILAND. NARATHIWAT**, Waeng District, Way to Khlong Saphan 2, 400 m, 24 Apr 2012, *Puudjaa & Hemrat 1790* (♀) (BKF!); Kao Re chau, Toh moh [5°48'10.6"N, 101°42'40.5"E], 1800 f [c. 550 m], 20 Apr 1931, *Lakshnakara 731* (♂) (BK!, BM!, E!, K!);ibid., 2000 f [c. 610 m], 21 Apr 1931, *Lakshnakara 738* (♀) (BK!, BM!, K!).

HABITAT. Lowland evergreen forest to montane forest, sometimes mixed with bamboos; on slope, by rivers or streams, sometimes on sandstones; elev. 60–600 m. with two collections in Cameron Highlands with elevation ranging from 1,000–1,300 m.

CONSERVATION STATUS. Near Threatened (NT). In Peninsular Malaysia, *Urophyllum trifurcum* is seemingly widespread (EOO >59,000 km^2) and found in more than 16 localities from Narathiwat Province, Thailand to Sembrong, Kluang District, Johor State; many specimens recorded are collected before 1990s and many areas have changed to agricultural land (e.g., in Bentong, Raub, as well as Cameron Highlands districts) or to urban areas (Ulu Langat, Selangor) (according to Google Earth imagery). The expansion of agricutltural land makes the habitat loss and the presence of the species uncertain especially considering the small number of new collections. Therefore, a Near Threatened

(NT) rating is proposed to this species, it is recommended that new collection data are gathered to accurately provide a conservation status assessment.

PHENOLOGY. Collected in flower almost every month except December; in fruit from February to December.

NOTES. *Urophyllum trifurcum* can be easily recognised by its softly short densely hairy appearance, sometimes subglabrous; stipules appressed to the stem but not folded; leaves usually chartaceous, elliptic to ovate, the midrib hairy adaxially, secondary veins festooned brochidodromous, major veins and veinlets usually dried reddish to dark brown abaxially; inflorescence pedunculate umbellate, two-tiered umbellate and two-tiered trichotomous cymose, calyx subtruncate to toothed, corolla densely hairy abaxially, and triangular membrane presence at the base of corolla lobes. *Urophyllum villosum* and *U. trifurcum* are morphologically similar in sharing linear triangular densely hairy stipules, leaves ovate to elliptic or oblong, apex acuminate with long tail, corolla and the membrane at lobes base morphologies. However, *U. villosum* has coriaceous leaves, abaxially always densely hairy, tertiary veins and veinlets conspicuously prominent, and calyx lobed.

The specimens of *Urophyllum trifurcum* show large variation in inflorescence type and peduncle length (personal observation). Specimens with pedunculate to two-tiered umbellate inflorescences are usually rather densely hairy appearance (especially on the petiole and abaxial leaf surface) with shorter peduncle (0.5–1.8 cm long) than those with two-tiered trichotomous cymose ((0.7–)1.3–6.1 cm long). However, variation of hair density and inflorescence type in *Urophyllum* species is large, these characters are therefore not sufficient to propose a new taxon.

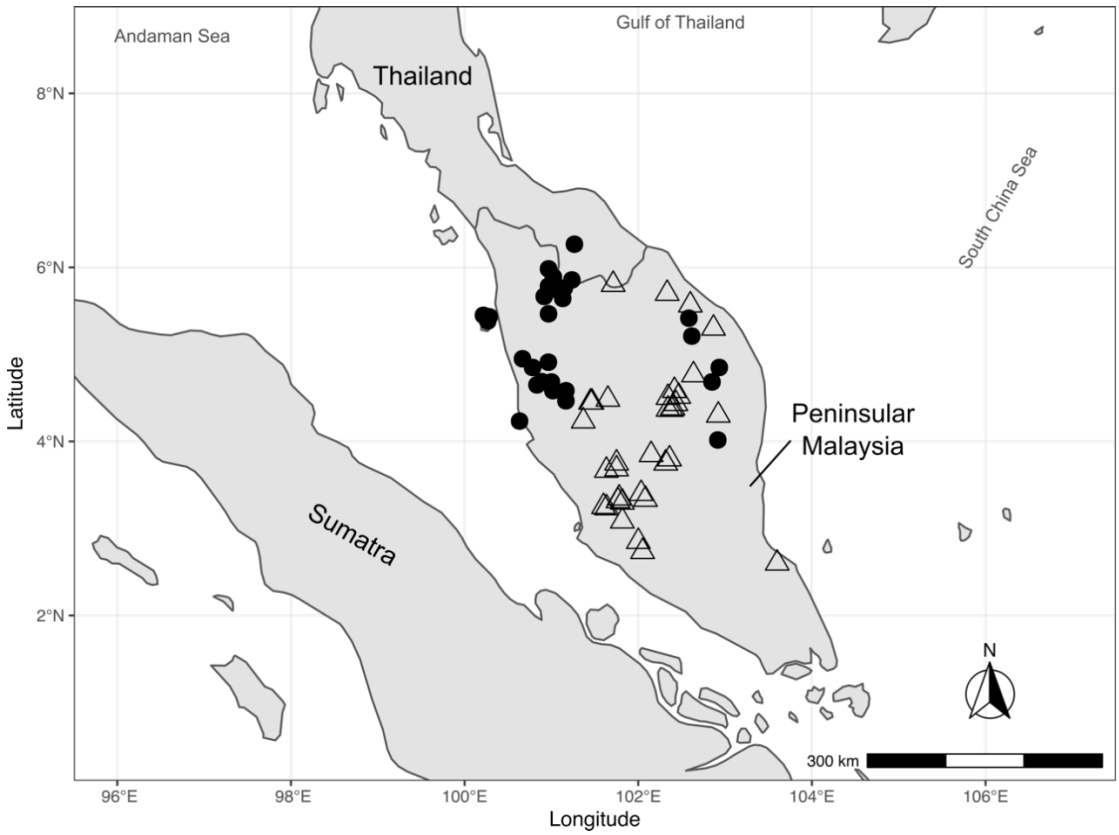


Figure 4.11 Occurrence map of *Urophyllum trifurcum* (△) and *U. villosum* (●).

11. ***Urophyllum villosum*** Wall. in Roxburgh (1824: 185). Type: [Malaysia, Peninsular Malaysia], Penang, *Wallich s.n.* [EIC 8314] (♀) (lectotype, designated by Wong *et al.* (2019) K-W! [K001125227]; isolectotypes K-W! [K001125226, K001125228 (element on the leftmost)]).

DISTRIBUTION. Almost throughout Peninsular Malaysia (to Kluang District, Johor State) and the southernmost Thailand (Yala Province) (Figure 4.11).

SPECIMENS EXAMINED. BIRMA AND MALAY PENINSULA, s.loc., s.a., *Griffith 2938/1* (♀) (P-photo!); **MALAYSIA. PENINSULAR MALAYSIA:** s.loc., s.a., *Kiah SING 35342* (♀) (K!); s.loc., s.a., *Maingay 884* (♂) (L-photo!); **JOHOR**, Kluang District, Mersing-Kluang road, simplot plot No.6, 2 Aug 1984, *Deverre 220* (♀) (P-photo!); **KEDAH**, Baling District, s.loc. [5°40'N, 100°55'E], 5 Nov 1938, *Kiah SING 35395* (♀) (K!); ibid., Gunong Lang [5°47'N, 100°58'E], 26 Mar 1938, *Kiah SING 35073* (♂) (BKF!, K!); ibid., Ulu Muda FR [5°59'N, 100°58'E], 1000 f [c. 300 m], 20 Jan 1969, *Chan FRI 6763* (♂) (K!); ibid., S. Section compt. 90 [5°59'N, 100°58'E], 600 f [c. 180 m], 15 Jan 1969, *Bray FRI 11510* (♂) (K!); **KELANTAN**, Kuala Krai District, Ulu Kelantan, Sungai Jenal, 5 miles N. Sungai Galas, 27 Oct 1967, *Cockburn FRI 7450* (♀) (K!); **PAHANG**, Kuantan District, Gunung Tapis Expedition, ridge top [4°01'N, 102°55'E], 2100 f [c. 640 m], 29 Sep 1971, *Chan FRI 19889* (♀) (K!); **PENANG**, Bukit Lassnanu, Mar 1885, *Lurtis 178* (♂) (K!); Mounts road, top end, 29 Jul 1917, *Burkill 2641* (♀) (K!); s.loc., 20 Jan 1905, *Wallich s.n.* (♀) (P-photo!); Barat Daya District, Tullek [Teluk] Bahang [5°27'N, 100°13'E], Aug 1980, *Lewis s.n.* (♂) (P-photo!); Timur Laut District, Government Hill [5°25'00"N, 100°16'00"E], Feb 1867, *s.coll. 2269* (♀) (K!); ibid., Penang Botanical Garden [5°26'N, 100°17'E], 40 m, 13 Sep 1966, *Hou 827* (♀) (K!); ibid., Penang Hill [5°26'N, 100°16'E], 4 Feb 1966, *Selvaraj KEP 99654* (♀) (K!); ibid., Penara Bukit [5°23'N, 100°16'E], 6 Mar 1938, *Yahaya SING 21445* (♀) (K!); Poelu Pinang [Penang Island], s.a., *s.coll. s.n.* (sterile) (U-photo!); ibid., s.a., *Blume s.n.* (♀) (L-photo!); ibid., Aug 1823, *s.coll. s.n.* (♀) (P-photo!); ibid., s.a., *Wallich 8314a* (♀) (E!, P-photo!); **PERAK**, Khler, Jun 1889, *Wray Jr. 3649* (♂) (K!, P-photo!); s.loc., 9 Nov 1920, *Harilt & Jun 6944* (♂) (K!); s.loc., Jun 1889, *Wray Jr. 3675* (♂) (L-photo!, P-photo!); Hulu Perak District, Kg. Tera, Grik State Land [5°28'N, 100°58'E], 15 Nov 1966, *bin Ismail KEP 95030* (♀) (L-photo!); Kerian District, Gunong Semanggol [4°57'N, 100°40'E], 200 m, 14 Jun 1938, *Sparz SING 34577* (♀) (K!); Kinta District, Goping [Gopeng] [4°28'N, 101°10'E], Aug 1880, *Dr. King's collector 526* (♀) (K!);

ibid., Goping [Gopeng], Apr 1888, *Scortechini 1983* (♂) (K!); ibid., Gunong Keledang [4°35'N, 101°1'E], Sep 1898, *Ridley 9711* (♀) (K!); ibid., Kinta Hill FR, Ipoh [4°35'N, 101°10'E], 1500 f [c. 460 m], 10 Dec 1966, *Zainudin KEP 99753* (♀) (K!); Kuala Kangsar District, Chior FR, 7 Nov 1967, *Ismail KEP 99820* (♀) (K!); ibid., Gunong Bubu FR, entrance at B. Rubber, Seedling nursery at the Kuala Kangsar/Dindings boundary about 18.5 miles south of Kuala Kangsar, 1 Nov 1958, *Sinclair 9913* (♀) (L-photo!); ibid., [4°39'N, 100°50'E], 710 m, 14 Aug 1966, *Hou 622* (♀) (K!); ibid., Saiong FR, Keledang [Kledang] [4°41'N, 101°00'E], 26 Sep 1989, *Damahuri FRI 36681* (♀) (K!); ibid., Sg. Legap, Chior FR, Sg. Siput [4°54'38''N, 100°57'53''E], 600 f [c. 180 m], 10 Sep 1967, *Ng FRI 5814* (♀) (K!); ibid., Sg. Plus, Chior FR, Sg. Siput [4°54'38''N, 100°57'53''E], 600 f [c. 180 m], 10 Jun 1967, *Ng FRI 5761* (♀) (K!); ibid., Ulu Kenas, Ulu Kenas recreational forest, G. Bubu FR [4°41'15"N, 100°53'24"E], 22 Jul 2009, *Imin et al. FRI 68141* (♀) (L-photo!); Larut Matang District, Maxwell Hill [Bukit Larut] base, Taiping [4°51'N, 100°47'E], 9 Apr 1968, *bin Kiah S. 307* (♀) (L-photo!); ibid., between 8th and 9th mile [4°51'N, 100°47'E], 636–1400 f [c. 190–430 m], 19 Sep 1949, *Sinclair & Kiah SING 38824* (♀) (E!); Manjung District, Lumut [4°14'N, 100°38'E], 10 Nov 1992, *Thomas & Teo 4174* (♀) (P-photo!); **TERENGGANU**, Dungun District, Mandi Angin Expedition, Southern watershed of S. Loh [4°41'N, 102°51'E], 2600 f [c. 790 m], 7 Sep 1968, *Whitmore FRI 12032* (♀) (K!); ibid., Ulu Sungai Loh [4°41'05''N, 102°50'56''E], 2500 f [c. 760 m], 13 Jul 1968, *Cockburn FRI 10800* (♀) (K!); Hulu Terengganu District, Gunung Lawit via Kampong Buloh [5°25'N, 102°35'E], 400 m, 12 Mar 1975, *Shah et al. 3487* (♀) (L-photo!); ibid., Tasik Kenyir, Hutan Simpan Tembat, Kaki Gunung Tembat, compt. 96 [5°12'37"N, 102°37'E], 792 m, 4 Feb 2009, *Kamarul Hisham et al. FRI 67076* (♂) (K!, L-photo!); ibid., Ulu Brang [Berang] [4°51'N, 102°56'E], 2500 f [c. 760 m], Jul 1937, *Moysey & Kiah SING 33648* (♀) (K!); **SINGAPORE.** s.loc., s.a., *Jack s.n.* (♀) (L-photo!); **THAILAND. YALA**, Bannang Sata District, route to Lee Pae peak [06°15'59"N, 101°15'53"E], 680 m, 17 Oct 2017, *Poopath 2016* (♀) (BKF!); Betong District, Ban Chulabhorn Phattana 10, Trail to giant Tetrameles nudiflora tree [5°51'21.7"N, 101°14'07.9"E], 480 m, 21 Apr 2005, *Pooma et al. 5102* (♀) (AAU!, BKF!); ibid., Behind 10,000 Buppha Garden [5°53'19.6"N, 101°1'19.9"E], 878 m, 14 May 2019, *Yooprasert et al. 218* (♂) (BKF!, K!); ibid., Pattani [5°45'N, 101°2'E], 500 m, 11 Mar 1925, *Kerr 10066* (♂) (BK!, BM!, K!); ibid., Rubber tree trail, Malaysia border road No.4266 Km [5°38'35.1"N, 101°7'47.9"E], 705 m, 13 May 2019, *Yooprasert et al. 206* (♂) (BKF!, K!); ibid., Ta Noh Mae Roh, trail to Thailand-Malaysia border ridge, behind Mai

Muang Nao Garden [5°53'19.6"N, 101°1'19.9"E], 2 Feb 2015, *Poopath et al. 913* (♀) (BKF!); ibid., Yarome, Ban Lu Bo Bue Day [5°45'46"N, 101°8'53"E], 300 m, 22 Oct 2017, *Wai 2656* (♀) (BKF!); ibid., *Wai 2663* (♂) (BKF!); Sukhirin District, Hala-Bala WS, Betong, 5 Aug 1996, *Puudjaa 227* (♀) (BKF!).

HABITAT. Lowland evergreen forest mixed with dipterocarps; on slopes, ridges; by streams, waterfalls, or rivers; usually shaded areas; elev. 40–900 m.

CONSERVATION STATUS. Least Concern (LC). Many collections of *Urophyllum villosum* (47 out of 58 specimens) were collected before the 1990s, however the areas where it is found are seemingly undisturbed (according to Google Earth imagery). It is also found in many protected areas, e.g., Bang Lang National Park, Hala-Bala Wildlife Sanctuary, Bukit Larut Water Catchment Forest, Bubu Water Catchment Forest, Taman Negara (Terengganu), Remen Chereh Soil Protection Forest with the estimated EOO >79,000 km^2 and several newly collections from the 2000s. Therefore, Least Concern is proposed for this species for the conservation assessment.

PHENOLOGY. Collected in flower from January to November; in fruit from February to December.

NOTES. Morphological characters of *Urophyllum villosum* are distinctive and the species is easily identified by its coriaceous leaves, usually ovate (sometimes elliptic), apex acuminate with long tail (hence to be type of the genus), abaxially densely hairy, yellowish to light brown when dried, secondary veins festooned brochidodromous; pedunculate to two-tiered umbellate inflorescence, calyx lobed, corolla hairy abaxially, and triangular membrane present at lobes base. Differences between the morphologically similar species (*U. trifurcum*), distribution overlapped species in Thailand (*U. glabrum*, *U. hirsutum* and *U. longipes*) and closely related species by molecular data (*U. crassum*), are previously discussed in those species.

The specimens with barcode K000740725 and K001125228 (except the leftmost element) are designated as isolectotypes in Wong *et al.* (2019), however neither specimen could be assigned here, as both bear staminate inflorescences and the species is dioecious, therefore staminate and pistillate material could not be found on the same individual.

Therefore, they are omitted from the type list in this study. On JSTOR Global Plants, there is a specimen of Wallich (EIC 8314) in BR [BR0000005621262], but without inflorescences, the specimen could not be assigned as isolectotype of this species. Other specimens with EIC 8314a not included in the type list are: K [K000740726] and NY [NY00133474].

4.5. Dubious specimens

The following specimens resemble *Urophyllum hirsutum* by their very densely hairy appearance, leaves with secondary veins festooned brochidodromous; calyx lobed and corolla hairy abaxially, but their pedunculate umbellate inflorescence has a longer peduncle (2.6–4.6 mm) whereas the length in *U. hirsutum* is 0–1.8(–2.5) mm. The character might be variation within *U. hirsutum*, thus further evidence on the molecular study is recommended.

MALAYSIA. PENINSULAR MALAYSIA: PAHANG, Lipis district, Sg. Telom NE of Kuala Mesong, 700 f [c. 210 m], 30 May 1971, *Sohadi FRI 17866* (♀) (K!); **TERENGGANU**, [Hulu Terengganu District], Ulu S. Trengan, 1/2 mile upstream from K. Petang, 300 f [c. 90 m], 4 Jun 1968, *Cockburn FRI 8474* (♀) (K!); **THAILAND. NARATHIWAT**, Sungai Kolok District, Nikhom Waeng [5°5′N, 101°5′E], 27 Feb 1974, *Larsen 32673* (♀) (K!).

Chapter 5 New species and new status of *Urophyllum* Wall. (Rubiaceae) from Cambodia and Vietnam

5.1. Overview

This chapter has been submitted as a manuscript to the journal *Adansonia*, and is awaiting review at the time of submission of this thesis. The aim of this paper is to publish five new species of *Urophyllum* found in Cambodia and Vietnam, and to change the taxonomic rank of *U. longifolium* var. *annamense*. Both morphological data collected in Chapter 2 together with molecular data gathered in Chapter 3 (excluding for *U. pulchristipulum* sp. nov.) are used to provide the evidence for the publication of these new species names. The table below summarises the new species names with the sample codes that are included within this thesis. Moreover, point occurrence maps are provided for each of the taxa with data gathered from herbarium specimens and field collections, these data have also been used to undertake initial conservation assessments. Crucially, an identification key is provided for *Urophyllum* within the region, providing diagnostic characters for the identification of the new taxa. This publication therefore, contributes towards the Rubiaceae account for the Flora of Cambodia, Laos, and Vietnam, which is still to be published. The results published in this paper also highlight the need for collections of *Urophyllum* in this region.

Table 5.1 Overview of the species names published in Chapter 5, with the sample codes used in the analyses throughout this thesis. Symbols denote chapters where samples of each taxon are included; *=both Chapters 2 & 3; +=Chapter 3 only.

Sample Code	Species name	Authority	Distribution
U. longifolium var. *annamense*[+]	*U. annamense*	(Pierre ex Pit.) Jooprasert, Culham & Utteridge	Vietnam
Sp.2[*]	*U. bidoupense*	Jooprasert, Culham & Utteridge	Vietnam
Sp.4[+]	*U. brochidodromum*	Jooprasert, Culham & Utteridge	Vietnam
Sp.1[*]	*U. chinense* subsp. *latistipulum*	Jooprasert, Culham, Yahara & Utteridge	Vietnam
Observation of specimens only	*U. pulchristipulum*	Jooprasert, Culham & Utteridge	Cambodia
Sp.3[+]	*U. pseudoschmidtii*	Jooprasert, Culham, Yahara, Tagane & Utteridge	Vietnam

Author Contributions: SY conceived the paper with guidance from TU and AC. SY conducted fieldwork with VDN and KSN. ST and TY provided additional field collection data and notes for *U. chinense* subsp. *latistipulum* and *U. pseudoschmidtii*. SY collected both morphological and molecular data for all the taxa included in this study. SY wrote the initial draft of the manuscript followed by TU and AC reviewing drafts of the manuscript. TU, AC, KSN, and ST provided feedback upon a later draft of the manuscript and SY made necessary changes before submission of the manuscript.

New species and new status of *Urophyllum* Wall. (Rubiaceae) from Cambodia and Vietnam

Sawita YOOPRASERT

The Forest Herbarium, Department of National Parks, Wildlife and Plant Conservation, 61 Phaholyothin Rd., Ladyao, Chatuchak, Bangkok, 10900 Thailand

and School of Biological Sciences, Health and Life Sciences Building, University of Reading, Whiteknights, Reading, RG6 6EX, U.K.

saveta.yo@gmail.com (corresponding author)

Alastair CULHAM

School of Biological Sciences, Health and Life Sciences Building, University of Reading, Whiteknights, Reading, RG6 6EX, U.K.

a.culham@reading.ac.uk

Shuichiro TAGANE

Kagoshima University Museum, Kagoshima University, 1-21-30 Korimoto, Kagoshima 890-0065, Japan.

stagane29@gmail.com

Tetsukazu YAHARA

Kyushu Open University, Fukuoka 819-0395, Japan

tet.yahara@gmail.com

Van Du NGUYEN

Institute of Ecology and Biological Resources (IEBR) and Graduate University of Science and Technology–Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Street,

Cau Giay, Hanoi, Vietnam

vandu178@gmail.com

Khang Sinh NGUYEN

Institute of Ecology and Biological Resources (IEBR)–Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Street, Cau Giay, Hanoi, Vietnam

nskhang@gmail.com

Timothy M.A. UTTERIDGE

Royal Botanic Gardens, Kew, Herbarium, Kew, Richmond, Surrey, TW9 3AE, U.K.

t.utteridge@kew.org

5.2. Abstract

Five new species of *Urophyllum* Wall. endemic to Cambodia and Vietnam are herein described and illustrated, and *U. annamense* (Pierre ex Pit.) Yooprasert, Culham & Utteridge is raised to species status (previously *U. longifolium* var. *annamense*). The new species are unique due to the morphological combination of indumentum, secondary venation, stipule shape and inflorescence structure as follows *U. bidoupense* Yooprasert, Culham & Utteridge: plant almost glabrous, secondary veins weakly brochidodromous, stipule glabrous, oblong-lanceolate; *U. chinense* subsp. *latistipulum* Yooprasert, Culham & Utteridge: similar appearance to the *U. bidoupense* but differs with stipule ovate to elliptic, hairy along the midline; *U. brochidodromum* Yooprasert, Culham & Utteridge: stem and branches hairy, adaxial leaf surface hairy on midrib and secondary veins, secondary veins conspicuously brochidodromous, stipule hairy, lanceolate; *U. pulchristipulum* Yooprasert, Culham & Utteridge: plant glabrous, secondary veins festooned brochidodromous, stipule glabrous and subcordate, inflorescence sessile; *U. pseudoschmidtii* Yooprasert, Culham & Utteridge: stem and branches hairy to subglabrous, secondary veins festooned brochidodromous, stipule sparsely to densely hairy, linear-lanceolate to oblong-lanceolate, pistillate flower with no staminodes. An identification key to *Urophyllum* species of Cambodia, Laos and Vietnam is provided. Point occurrence maps are presented for each species, as well as provisional conservation assessments based on IUCN guidelines.

Key words: Indochina, key identification, taxonomy, Urophylleae, understorey plants, endemic, Gentianales

Mots clés: Indochine, Clés d' Identification, taxonomie, Urophylleae, plantes de sous-bois, endémique, Gentianales

5.3. Introduction

Urophyllum Wall. is a genus in coffee family (Rubiaceae), comprised of c. 120 species mostly of understorey treelets and shrubs in subtropical and tropical Asia from India to Papua New Guinea (Metcalfe, Grubb and Turner, 1998; Govaerts *et al.*, 2020). The genus was first published in Flora of India by Nathaniel Wallich in Roxburgh (1824) with two species: *Urophyllum villosum* Wall. and *U. glabrum* Wall. As *U. villosum* is the first species mentioned in this publication, it has been selected as the type specimen of the genus by Wong *et al.* (2019). Members of *Urophyllum* can be recognised by their usually long linear-lanceolate to oblong stipules, acuminate leaf apex, secondary veins usually festooned brochidodromous or weakly brochidodromous, inflorescence axillary, flowers urn shaped, stigma with 4–5 lobes, and baccate fruits bearing numerous seeds with an alveolate surface (Yooprasert, 2021).

The genus in mainland Indochina has been revised at a regional scale and published in local Floras including Pitard (1923) for Indochina; Pham (2000) for Vietnam; Schumann in Schmidt (1902) for Koh Chang and Puff *et al.* (2005) for Thailand; Chen & Taylor (2011) for China; and Nagahama *et al.* (2019) for Bidoup-Nui Ba National Park, Vietnam.

As a result of taxonomic study from both fieldwork and herbarium specimen observations of *Urophyllum* in mainland Southeast Asia, five new species of *Urophyllum* and a new status for *U. annamense* (Pierre ex Pit.) Yooprasert, Culham & Utteridge (based on *U. longifolium* (Wight) Hook.f. var. *annamense* Pierre ex Pit.) were identified. Within these, four species are endemic to Vietnam: *U. bidoupense* Yooprasert, Culham & Utteridge, *U. brochidodromum* Yooprasert, Culham & Utteridge, *U. chinense* Merr. & Chun subsp. *latistipulum* Yooprasert, Culham, Yahara & Utteridge, *U. pseudoschmidtii* Yooprasert, Culham, Yahara, Tagane & Utteridge; and a single species endemic to Cambodia - *U. pulchristipulum* Yooprasert, Culham & Utteridge. All these species, except the new species from Cambodia, have been included in a comprehensive molecular analysis by the first author (Yooprasert, 2021), which supports recognition of the new taxa at species or

subspecies level as presented here. The new species bring the total number of *Urophyllum* species to 12 in Cambodia, Laos and Vietnam. Due to the difficulty of species identification within the genus, and the resulting number of misidentified specimens of *Urophyllum* in Indochina, a regional key is provided for the species recorded from Cambodia, Laos and Vietnam aligning with the scope of the regional Flora project Flore du Cambodge, du Laos et du Viêtnam.

During the taxonomic study of *Urophyllum* in Thailand and Indochina, there are several characters which are especially useful to distinguish species: shape and indumentum of stipules (Figure 5.1A–D), secondary venation and looping patterns (Figure 5.1E–G), inflorescence type (e.g. pedunculate vs sessile umbellate–see Figure 5.4D; 11D), calyx lobe morphology (Figure 5.1H–I) and abaxial corolla hair distribution (Figure 5.1J–K) (Tan, Chua and Turner, 1995; Yooprasert, 2021); these were used as diagnostic characters in this study.

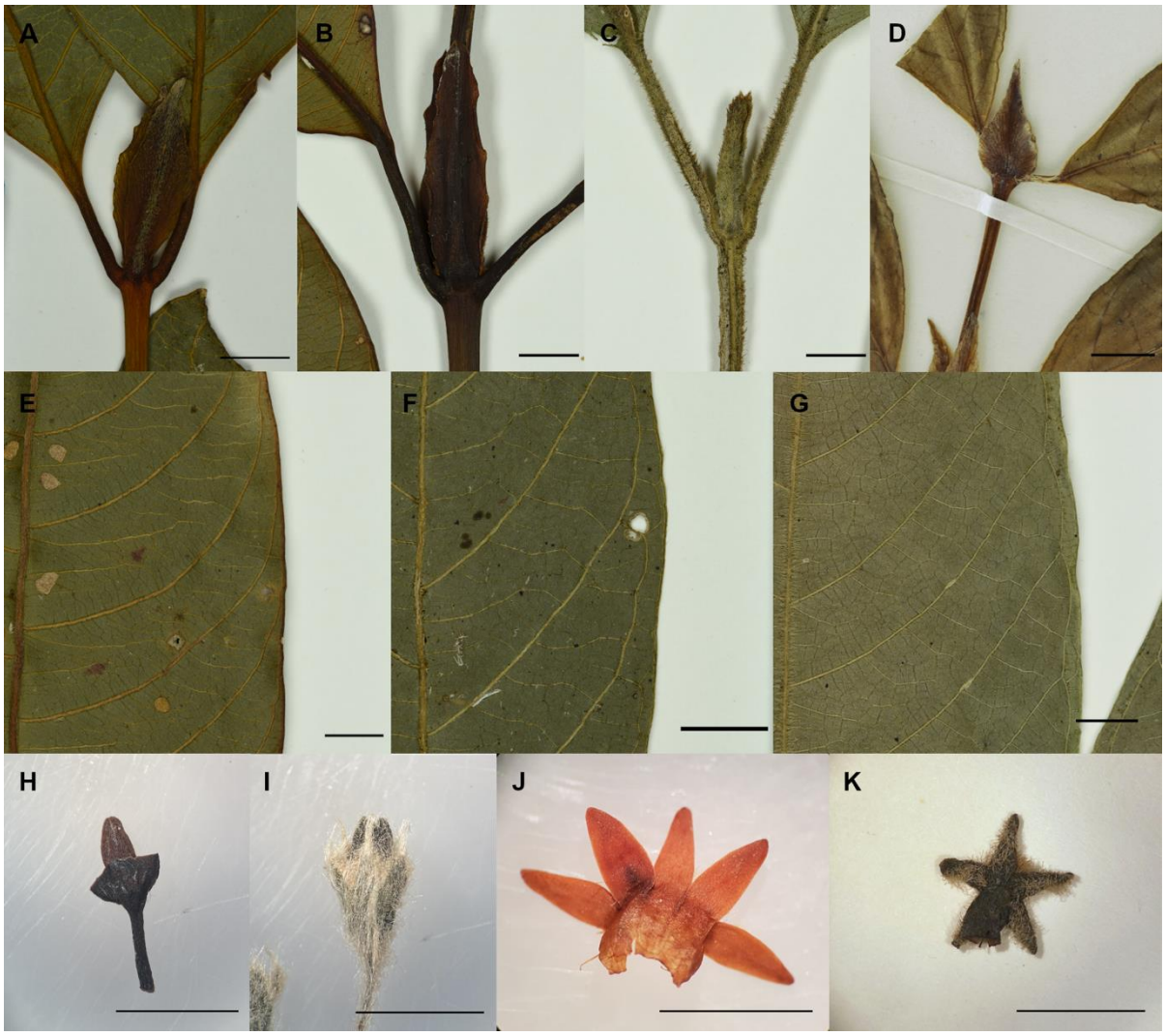


Figure 5.1 Stipule shape and indumentum (A–D). **A** ovate and hairy along midline (*Urophyllum chinense* subsp. *latistipulum* Yooprasert, Culham, Yahara & Utteridge). **B** oblong-lanceolate and subglabrous, hairy only at apex (*U. bidoupense* Yooprasert, Culham & Utteridge). **C** linear-lanceolate and hairy all over (*U. brochidodromum* Yooprasert, Culham & Utteridge). **D** subcordate and glabrous (*U. pulchristipulum* Yooprasert, Culham & Utteridge). Secondary venation and looping patterns (E–G). **E** weakly brochidodromous (*U. bidoupense*). **F** festooned brochidodromous (*U. pseudoschmidtii* Yooprasert, Culham, Yahara, Tagane & Utteridge). **G** conspicuously brochidodromous (*U. brochidodromum*). Calyx lobe morphology (H–I). **H** toothed (*U. bidoupense*). **I** lobed (*U. argenteum* Pit.). Abaxial corolla hair distribution (J–K). **J** glabrous (*U. bidoupense*). **K** hairy (*U. lecomtei* Pit.). Scale bar 5 mm.

5.4. Materials and methods

This taxonomic study of *Urophyllum* Wall. in Vietnam and Cambodia was based primarily on herbarium specimens from AAU, FU, K and newly collected specimens from fieldwork in Vietnam in May–June 2019, where they were used for morphological character measurement and examination; digitised records from BM, C, CMUB, E, HN, HNU, VNM, NY, P and the Naturalis Biodiversity Centre BioPortal (specimens of L, U and WAG available at http://bioportal.naturalis.nl/) were also examined. The locations visited in the fieldwork were selected based upon type locality information, herbarium specimens examined and in-country knowledge from Vietnamese botanists. All specimens cited have been seen, unless indicated by "n.v.". Leaf vein morphological terms follow Hickey (1979) and Leaf Architecture Working Group (1999), other morphological terms are from Beentje (2016). The conservation assessments were estimated following the IUCN Red list categories and criteria v3.1 (IUCN, 2012b) and IUCN guidelines (IUCN Standards and Petitions Committee, 2019). The Extent Of Occurrence (EOO) and Area Of Occupancy (AOO) were calculated using the GeoCAT software (Bachman *et al.*, 2011) with the cell size set to 2 × 2 km^2. Google Earth Pro v7.3.3.7786 was used to view the quality of habitat at different time periods (from 2008 to 2020), such as the changes in the amount of existing forest compared to expansion of urban and cultivation areas. Protected areas and IUCN Management Categories were referenced using the World Database on Protected Areas (UNEP-WCMC and IUCN, 2020). If the category was not assigned in this database, the reports in Biodiversity and Protected areas–Cambodia, Laos, and Vietnam (Clarke, 2000a, 2000b, 2000c), were used. The descriptions of species used in this study were based upon herbarium specimens and the fieldwork in 2019, they could not be used to estimate population size and trends as required for criteria A and C, as well as quantitative analysis of population viability (criterion E); thus, only the criteria B and D have been applied. The point occurrence maps were generated on R (R Core Team, 2019).

5.5. Taxonomic treatment

1. *Urophyllum annamense* (Pierre ex Pit.) Yooprasert, Culham & Utteridge **stat. nov.** (Figure 5.2)

Diagnosis. *Urophyllum annamense* is similar to *U. longifolium* Hook.f. in most of the appearance but differs in its corolla colour, which is green instead of white. It also differs from *U. glabrum* and *U. schmidtii* C.B.Clarke in having stipules which are folded toward the adaxial side rather than the stipules appressed to the stem but not folded. The species also differs from *U. schmidtii* in its compound cymose inflorescence and calyx with small teeth or nearly entire (not conspicuously lobed), instead of pedunculate umbellate inflorescences and calyx lobed, that are found in *U. schmidtii*.

Urophyllum longifolium (Wight) Hook.f. var. *annamense* Pierre ex Pit. (Pitard in Lecomte 1923: 202). **Type.** VIETNAM. Bao Chiang [in Bien Hoa, Dong Nai Province], July 1877, *Pierre 1840* (♀) (lectotype selected here P-photo! [P03922480]; isolectotypes BM!, C!, K!, NY-photo!, P-photo! [P03922444, P03922481]).

Specimens examined. VIETNAM. Lam Dong Province: Bao Loc, piste menant a la montagne [track leading to the mountain], 16 July 1984, *Tirvengadum 1637* (♀) (AAU!); **Khanh Hoa Province:** Khanh Vinh, Khanh Phu, Hon Ba NR, Trail by the road [12°6'48.4"N, 108°58'8"E], 884 m elev., 29 May 2019, *Yooprasert et al. VN 73-3* (sterile) (BKF!, HN!); ibid. [12°6'43.1"N, 108°58'25.9"E], 874 m elev., 29 May 2019, *Yooprasert et al. VN 74-5* (sterile) (K!).

Distribution. Vietnam: Khanh Hoa Province (Hon Ba Nature Reserve), Lam Dong Province (Bao Loc) and Dong Nai Province (Bien Hoa) (Figure 5.3).

Habitat. Primary subtropical evergreen forest; elev. ca. 800 m.

Phenology. Collected in flower and young fruit from June to July.

Conservation status. Vulnerable (VU) D2. To date, *Urophyllum annamense* is found in southern Vietnam from Khanh Hoa to Dong Nai Provinces recorded from four specimens. The type specimen of the species, *Pierre 1840*, was collected in Bao Chiang, Cochinchine where it can be traced to Bien Hoa Province from the specimen of *Dialium cochinchinense* Pierre, *Pierre 1814* (P-photo! [P00330626]). The type specimen was not included for

evaluating the conservation status due to historic date of collection (1877) and declining forest habitat in the locality (Bien Hoa Province) visible on the Google Earth imagery, leaving three specimens collected from two locations in Bao Loc city and Hon Ba Nature Reserve (IUCN category not reported) to be reviewed. The visit to the nature reserve in 2019 showed it was in a good condition and well managed by the reserve staff. Although agriculture reached to the border of the reserve, the population of *U. annamense* is further away from the edge. With these small number of locations found and restricted AOO (8 km^2), the species is rated as Vulnerable (VU). Further collections are recommended to provide a full assess of species existence.

Remarks. *Urophyllum annamense* was first published as *U. longifolium* var. *annamense* by Pitard using the manuscript of Pierre in Flore Générale de l'Indo-Chine with no taxon diagnosis but five specimens were listed without indication of a type (Pitard in Lecomte, 1923). These specimens are *Eberhardt 3867* (P-photo!), *Chevalier 38709* (P-photo!), *Lecomte & Finet 713* (P-photo!), *Pierre 1251* (BKF!, C!, E!, K!, L-photo!, P-photo!) and *Pierre 1840* (BM!, C!, K!, P-photo!). After observing the specimens, they can be classified into two known taxa: *U. chinense* subsp. *chinense* Merr. & Chun (*Eberhardt 3867*, *Chevalier 38709*, and *Lecomte & Finet 713*) and *U. schmidtii* (*Pierre 1251*). The specimen of *Pierre 1840* is different from those two taxa and shows morphological characters similar to *U. longifolium*. Despite the lack of diagnosis in the original description of the taxon, and that most parts of the taxon description were not specific to any one of the specimens, there were two unique elements of the description that identify the specimen of *Pierre 1840* which were the axillary, cymose inflorescences, and the illustration of a flower in the publication the original of which is attached to the specimen. Therefore, the specimen of *Pierre 1840* is selected here to represent as a type specimen of the variety. The herbarium specimen deposited in P was chosen as the lectotype based upon Pierre's biography in Taxonomic Literature Second edition (Stafleu and Cowan, 1983) showing that he went back to Paris in 1877 and worked on the manuscript there.

Populations of *U. longifolium* var. *longifolium* are recorded from Tenasserim, Myanmar through Peninsular Thailand with the most southern in Songkhla Province (Yooprasert, 2021). However, specimens assigned to *U. longifolium* var. *annamense* are found only in Vietnam, further away and disjunct from other populations of *U. longifolium*. These specimens can also be distinguished by its corolla colour (green instead of white).

Together, these data support the change in taxonomic status from the variety to recognition at species rank as *Urophyllum annamense*.

Figure 5.2 Lectotype of *Urophyllum annamense* (Pierre ex. Pit.) Yooprasert, Culham & Utteridge. Digitised image from Muséum national d'Histoire naturelle, Paris (France), Collection: Vascular plants (P), Specimen P03922480 (http://coldb.mnhn.fr/catalognumber/mnhn/p/p03922480).

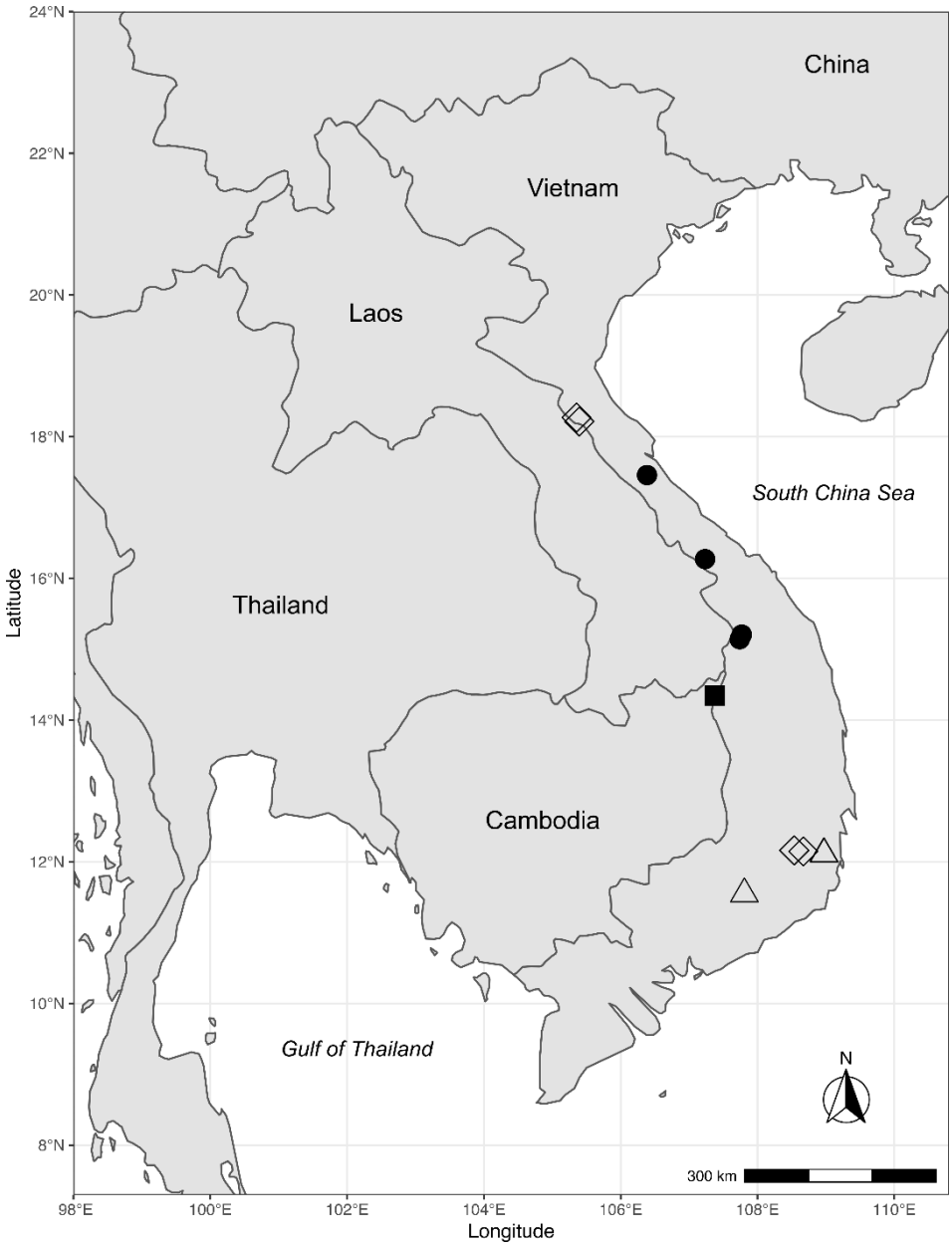


Figure 5.3 Occurrence of *Urophyllum annamense* (△), *U. brochidodromum* (●), *U. pulchristipulum* (■) and *U. pseudoschmidtii* (◇).

2. *Urophyllum bidoupense* Yooprasert, Culham & Utteridge **sp. nov.** (Figure 5.4 and 5.5)

Diagnosis. *Urophyllum bidoupense* is similar to *U. chinense* subsp. *chinense*, but differs in: the stipules which are subglabrous abaxially, the hairs (if present) are only around the apex, and the leaves are glabrous to the naked eye and sparsely hairy with magnification. While *U. chinense* subsp. *chinense* has uniformly densely hairy stipules, and hairy leaves to the naked eye. It differs from *U. chinense* subsp. *latistipulum*, as the stipules have hairs along the midline from base to apex.

Type. VIETNAM. Lam Dong Province: Lac Duong District, Bidoup-Nui Ba National Park, Trail by Ranger station Hon Giao, border to Son Thai, Khanh Vinh district [12°11'14.3" N, 108°42'51.7" E], 1,633 m altitude, 1 June 2019, *Yooprasert et al. VN 91-1* (♀) (holotype HN!; isotype BKF!).

Additional specimens examined. VIETNAM. Khanh Hoa Province: Khanh Son District, 41 km to NE from Dalat city, E macroslope of Hon Giao mt. ridge [12°11'N, 108°43'E], 1,500–1,600 m elev., 22 April 1997, *Averyanov et al. VH 4154* (♂) (AAU!, HN!); Khanh Vinh District, Son Thai [12°11'47.7"N, 108°43'33.2"E], 1518 m elev., 28 June 2018, *Tagane et al. V 9201* (♂) (DLU n.v., FU!, KAG!); **Lam Dong Province:** Lac Duong District, Bidoup-Nui Ba NP, Trail by Ranger station Hon Giao [12°11'11.8"N, 108°42'53.3"E], 1,575 m elev., 1 June 2019, *Yooprasert et al. VN 85-3* (♂) (BKF!, HN!); ibid., 1,651 m elev., 1 June 2019, *Yooprasert et al. VN 87-1* (♂) (BKF!, HN!); ibid., *Yooprasert et al. VN 87-3* (♂) (BKF!, HN!); ibid., *Yooprasert et al. VN 87-4* (♂) (BKF!, HN!); ibid., municipality Da chay [Da Chais Commune], 35 km to NE from Dalat city, W macroslope of Gia Rinh mt. ridge [12°9'N, 108°41'E], 1,700–1,800 m elev., 18 April 1997, *Averyanov et al. VH 4059* (♀) (AAU!, HN!).

Etymology. The specific epithet refers to its type locality, Bidoup-Nui Ba National Park.

Distribution. Vietnam: Lam Dong (Bidoup-Nui Ba National Park) and Khanh Hoa (Son Thai) provinces (Figure 5.6).

Habitat. Primary broadleaved evergreen montane forest mixed with conifers in the Hon Giao Mountain range; elev. 1,500–1,800 m.

Phenology. Collected in flower and fruit from April to June.

Conservation status. Least Concern (LC). *Urophyllum bidoupense* is known from only two previous collected specimens within the area of Bidoup-Nui Ba National Park (IUCN

category Ia) in 1997. The species was still found in the same locality in 2018 and 2019, close to the Hon Giao ranger station, where the habitat was observed to be undisturbed, and the species is common: the plot survey in Hon Giao by third and fourth authors recorded 108 individuals within 0.1 ha. Despite a small AOO (8 km^2), the species is not at risk, therefore it can be rated as Least Concern (LC).

Description

Shrub or *Treelet* to 10 m high, 10 cm in DBH, almost glabrous. *Branches* terete; young shoots flattened, ridged. *Stipules* caducous, 1–2 nodes at shoot, oblong-lanceolate to elliptic, 1–2 cm long, 0.3–0.6 cm wide, appressed to the stem but not folded, abaxially glabrous to subglabrous, hairs present at around apex; conspicuous to inconspicuous netted veins. *Leaves* coriaceous, elliptic to obovate, 7–15 cm long, 3–5 cm wide, apex acuminate, base cuneate, adaxially glabrous, abaxially glabrous to the naked eye, sparsely hairy with short hairs <0.5 mm long under magnification, drying dull yellowish green to brown adaxially, usually shiny yellowish green abaxially with reddish-brown midrib; secondary veins [10–]12–16 pairs, weakly brochidodromous; tertiary veins mixed percurrent, angle increasing basally, middle of the leaves subperpendicular to obtuse, veins diverging at angle of c. 95°–110° to the midrib; petioles [0.8–]1.1–1.7 cm long, canaliculate. *Inflorescence* axillary, pedunculated umbellate, peduncle 0.8–4.0 mm long, bract triangular with dense hairs at base and margin, 1–2 mm long, bearing 3–6 flowers in staminate plant, 3–4 flowers in pistillate plant; nectary disk annular, presence both sexes. *Staminate flowers* calyx fused, toothed 5, 1.1–1.6 mm long; corolla fused forming a tube at base towards 1/3 their length, apex lobed 5, valvate, 2–4 mm long, densely hairs at throat inside; stamens 5, filament fused to the corolla tube, anthers 2-celled, dorsifixed, introrse; pistillode present, ovary vestigial. *Pistillate flowers* calyx fused, toothed to shallow lobed 5, 1–2 mm long; corolla as staminate flowers; staminodes 5, anthers lacking pollen; ovary cupuliform, locules 5; style ca. 2 mm long, sparsely scaly; stigma lobed 5 ca. 0.5 mm long, covered with scales. *Fruits* baccate, subglobose, ca. 3 mm diam. *in sicco*; calyx and disk persistent. *Seed* numerous, subelliptic, surface alveolate, orange brown *in sicco*.

Remarks. *Urophyllum bidoupense* is recognised by its abaxially glabrous to subglabrous stipules. Additional characters useful for identifying the species are coriaceous leaves, abaxially usually a shiny olive colour, glabrous to the naked eye but sparsely hairy under magnification, fruits glabrous, calyx persistent with small teeth to shallow lobes.

Urophyllum bidoupense is similar to *U. chinense* subsp. *latistipulum* and their distribution overlaps within Bidoup-Nui Ba National Park. While *U. chinense* subsp. *latistipulum* is widely distributed in the national park, *U. bidoupense* is restricted to Hon Giao and its adjacent area. The main morphological character that differs between these two taxa are the stipule and leave morphologies. In *U. bidoupense*, the stipules are subglabrous abaxially, hairs present only at apex with coriaceous leaves drying shiny abaxially, whereas the stipules of *U. chinense* subsp. *latistipulum* are abaxially hairy along midline with leaves usually chartaceous drying dull abaxially. Other species found in the National Park and nearby area (Giang Ly and Iar Giang) are *U. pseudoschmidtii* and *U. lecomtei* Pit. The distinguish character of *U. bidoupense* from these two species is stipules which are subglabrous abaxially, whereas the latter two have conspicuously hairy stipules.

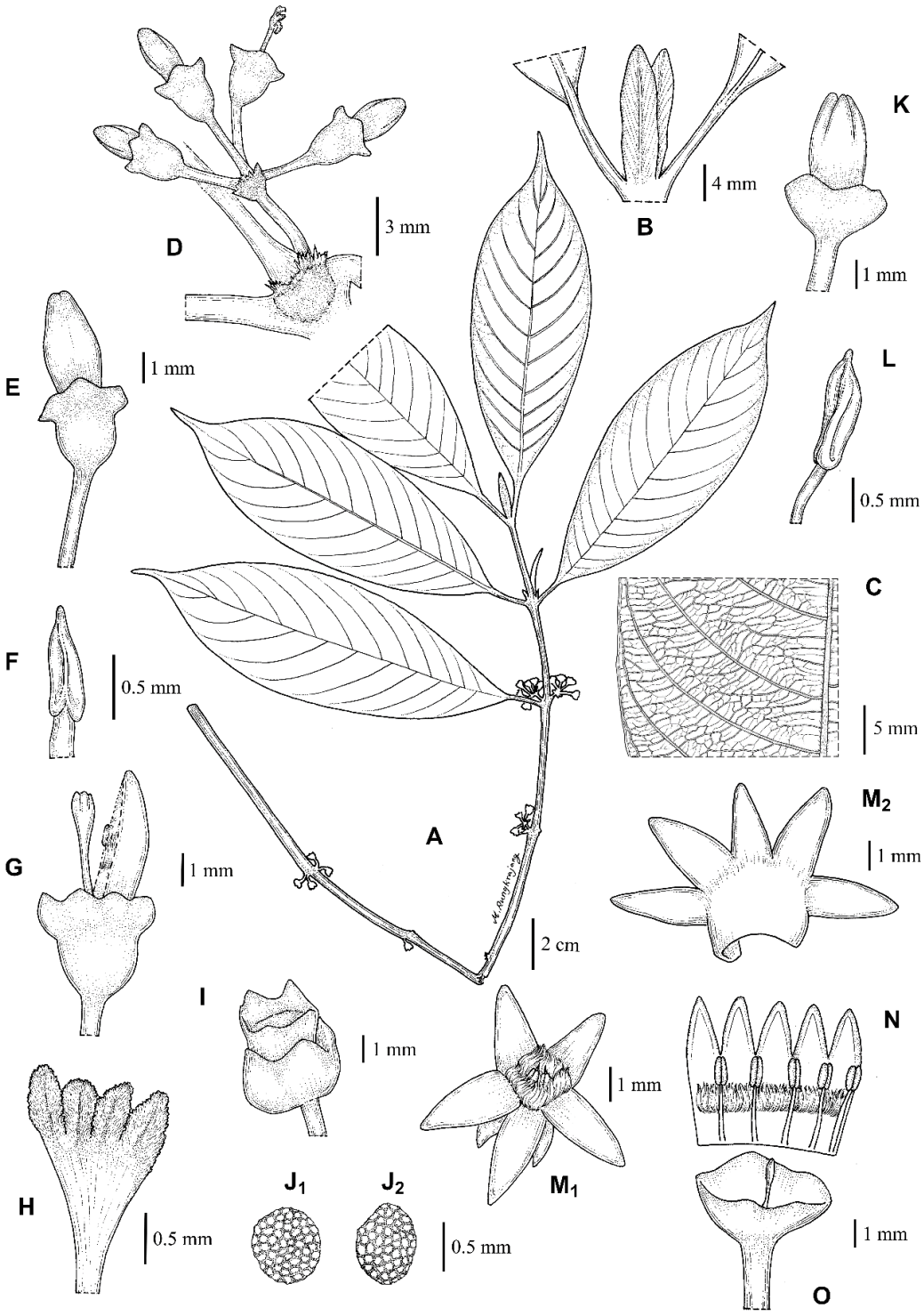


Figure 5.4 *Urophyllum bidoupense* Yooprasert, Culham & Utteridge. A Flowering branch. B Stipule. C Abaxial leaf lamina. D–J Pistillate plant: D Inflorescence; E Flower bud; F Staminode; G Immature stigma positioning in a flower; H Stigma; I Fruit; J Seed, top view (J_1) and side view (J_2). K–O Staminate plant: K Flower bud; L Stamen; M Corolla, adaxial side (M_1) and abaxial side (M_2); N Corolla opened out to show stamen and throat hair arrangement; O Calyx and pistillode. Drawn by Mahsarahka Rungkrajang. Drawn from: A *Averyanov et al. VH 4154*; B–G, O *Yooprasert et al. VN 91-1*; H–I *Averyanov et al. VH 4154*; J–N *Yooprasert et al. VN 87-1*.

Figure 5.5 Holotype of *Urophyllum bidoupense* Yooprasert, Culham & Utteridge.

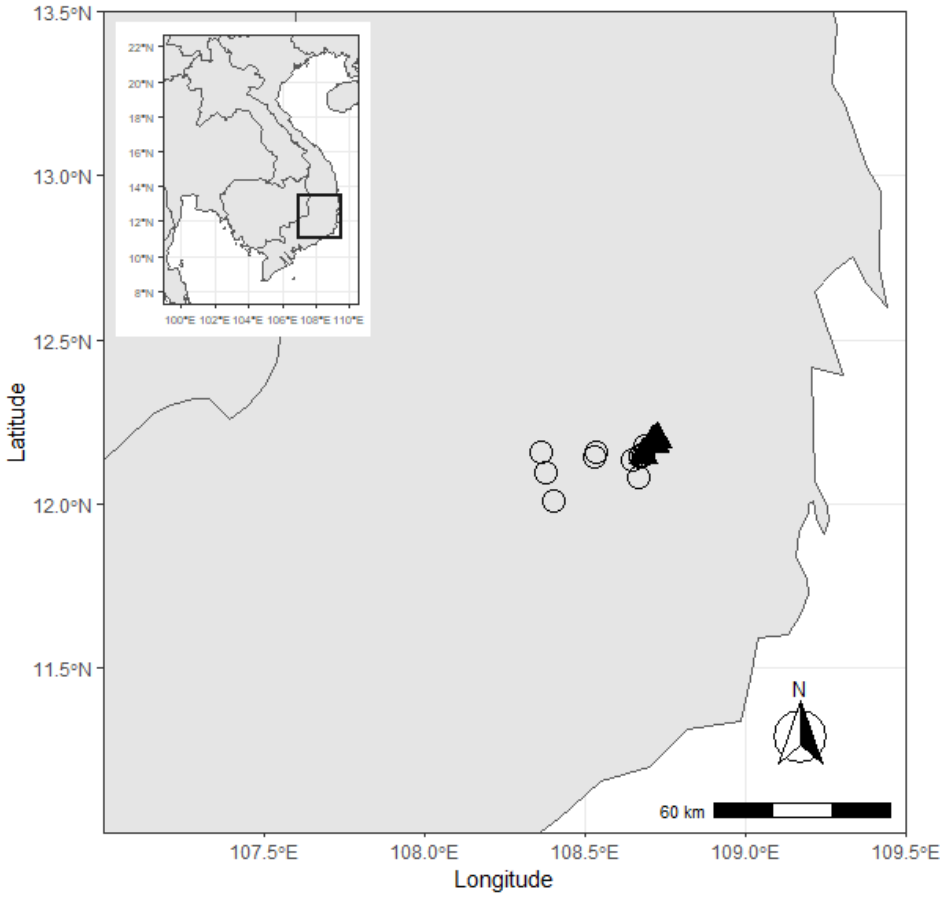


Figure 5.6 Occurrence of *Urophyllum bidoupense* (O) and *U. chinense* subsp. *latistipulum* (▲).

3. *Urophyllum brochidodromum* Yooprasert, Culham & Utteridge **sp. nov.** (Figure 5.7 and 5.8)

Diagnosis. *Urophyllum brochidodromum* is similar to *U. lecomtei*. but differs in the secondary veins strongly joining together to form closed loops close to the leaf margin with no additional sets of loops (conspicuously brochidodromous), while in *U. lecomtei*, the veins are joined with a series of small loops (festooned brochidodromous). The species also differs in its calyx lobes, which are recurved compared to the erect to incurved lobes found in *U. lecomtei*.

Type. VIETNAM. Kon Tum Province: Dak Gley [Dak Glei District], about 12 km to N of Dak Gley town (24 km by road), near Mang Khen village, 1,100–1,200 m altitude, 14 November 1995, *Averyanov et al. VH 1648* (♀) (holotype AAU!).

Additional specimens examined. VIETNAM. Bin Tri Thien [Quang Binh, Quang Tri and Thua Thien Hue provinces], 9 August 1980, *Hoang 67* (♀) (HN!); **Kon Tum Province:** Ngoc Linh Nature Reserve [15°12'24.2"N, 107°46'10.4"E], 1,365 m elev., 14 February 2017, *Tagane et al. V 6648* (♀) (DLU n.v., FU!); **Quang Binh Province:** Bo Trach District, Hung Trach municipality, Phong Nha-Ke Bang NP, between km 51 and 56 of west branch of Ho Chi Minh road. [17°27'30"N, 106°23'6"E], ca. 650 m elev., 4 July 2004, *Wu et al. WP 928* (♀) (HN!); **Thua Thien Hue Province:** A Luoi District, 25 May 2005, *Binh & Cuong VN 1506* (♂) (HN!).

Etymology. The specific epithet refers to the conspicuously brochidodromous lateral veins.

Distribution. Vietnam: Kon Tum Province (Ngoc Linh Nature Reserve), Quang Binh Province (Phong Nha-Ke Bang National Park) and Thua Thien Hue Province (A Luoi District) (Figure 5.3).

Habitat. Evergreen forest to evergreen montane forest, by stream; elev. 650–1,400 m.

Phenology. Flowers collected in May and fruits from August to February.

Conservation status. Vulnerable (VU) B1ab(iii)+2ab(iii)+D2. The species has an estimated EOO of 2,470 km^2 and AOO of 16 km^2 both below the Endangered (EN) threshold, and is only known from three locations; however, two locations are within the protected areas of Ngoc Linh Nature Reserve (IUCN category IV) and Phong Nha-Ke Bang National Park (IUCN category II) where the risk of deforestation is unlikely to increase dramatically. One

specimen was collected in A Luoi District which is not a protected area, with concerns for the expansion of agriculture land use growing *Acacia* species and *Cassava* in the future (Pham, 2019). Therefore, the species is rated as Vulnerable. However, there is Phong Dien Nature Reserve (IUCN category IV) which has continuous forest to the northeast of A Luoi District viewing on the Google Earth in 2021. If further collections reveal the presence of the species in this area, the rating could be reviewed as Near Threatened.

Description

Shrub or *treelet* to 2 m high. *Indumentum* of appressed to erect hairs, 0.5–1.5 mm long, scattered on old stems, dense on other parts except adaxial leaf lamina. *Branches* terete; young shoots flattened, ridged. *Stipules* caducous usually left only at shoot, lanceolate 1.1–1.4 cm long, 0.15–0.33 cm wide, appressed to the stem but not folded, abaxially densely hairy. *Leaves* chartaceous, elliptic to oblong, 13.0–15.1 cm long, 3.9–4.8 cm wide, apex attenuate to acuminate, base cuneate; adaxially glabrous, abaxially densely hairy, drying dull dark yellowish green adaxially, lighter abaxially with yellowish brown veins, adaxially midrib hairy; secondary veins 13–17 pairs, adaxially densely hairy, conspicuously brochidodromous, usually without additional sets of small loops; tertiary veins alternate percurrent, angle increasing exmedially, obtuse to midrib; petioles 1.0–1.5 cm long, canaliculate. *Inflorescence* axillary, pedunculated umbellate, peduncle 2.2–2.9 mm long, bracts lanceolate, 3–4 mm long, only fruit bearing pistillate plant known. *Fruits* baccate, subglobose, ca. 3 mm diam. *in sicco*; petiole ca. 5 mm; calyx fused, lobed 5, persistent, 1–2 mm long, recurved. *Seed* numerous, subelliptic, surface alveolate, orange brown *in sicco*.

Remarks. *Urophyllum brochidodromum* is recognised by having strongly brochidodromous and dense hairs on both adaxial midrib and secondary venation. The persistent recurved calyx is an additional diagnostic character. The species is similar to *U. lecomtei* in its erect dense hairs over all parts but differs by the characters previously stated in the diagnosis: *U. lecomtei* has hairs only on the adaxial midrib, secondary veins festooned brochidodromous (main loops with sets of smaller loops) and calyx erect or sometimes curved toward adaxial side in fruit. There are records of *U. argenteum* Pit. found in Bach Ma National Park, Thua Thien Hue Province. This national park is further south ca. 70 km away from A Luoi District where one population of *U. brochidodromum* was found. *Urophyllum argenteum* is distinguished from *U. brochidodromum* by its coriaceous leaves, adaxial leaf side shiny and glabrous, secondary veins

eucamptodromous to weakly brochidodromous with inconspicuous loops and erect calyx lobes.

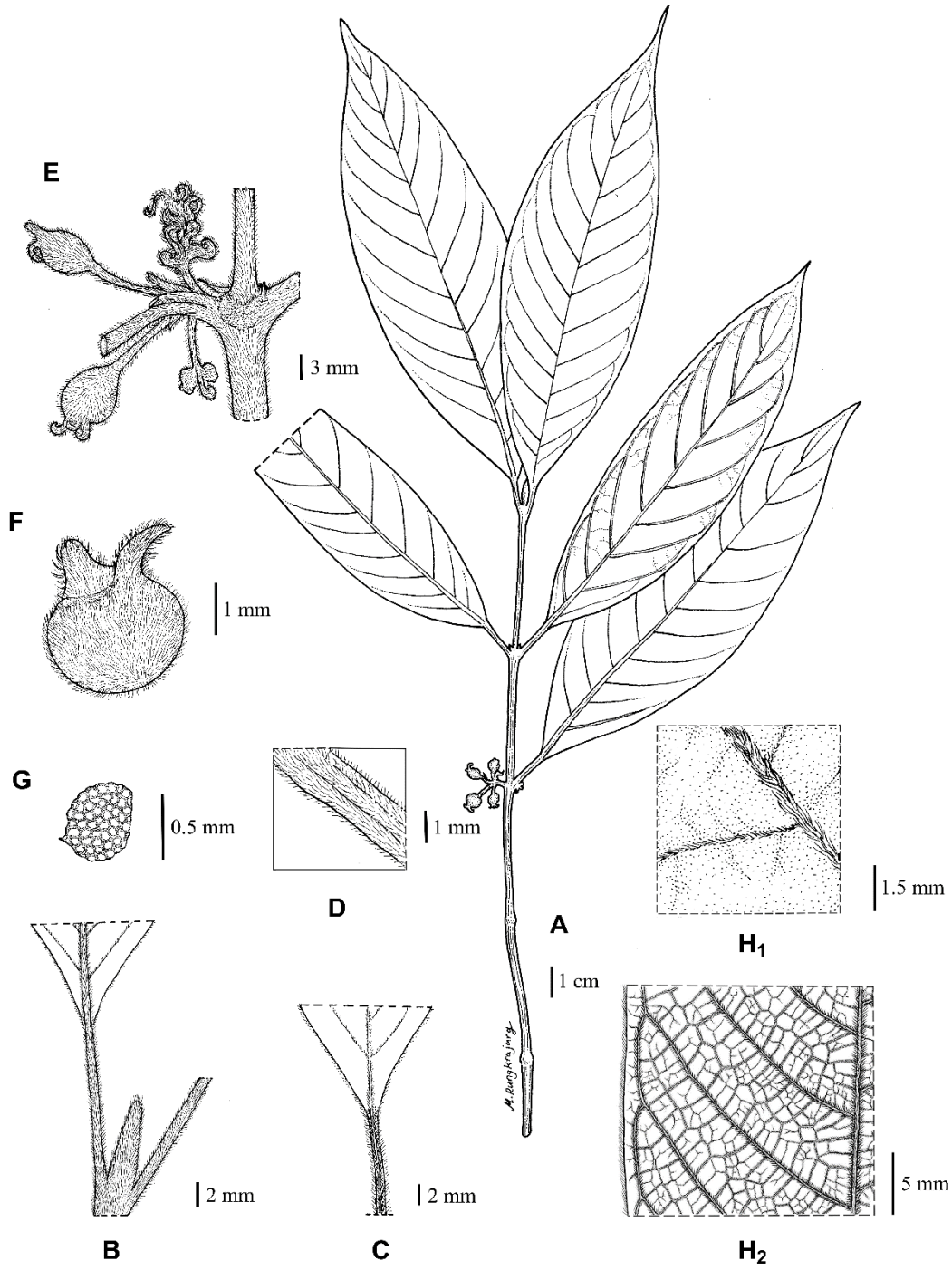


Figure 5.7 *Urophyllum brochidodromum* Yooprasert, Culham & Utteridge. Illustration of pistillate plant. A Fruiting branch. B Stipule and abaxial petiole. C Adaxial petiole. D Young stem. E Infructescence; F Fruit; G Seed. H Leaf lamina, adaxial side showing midrib and secondary veins (H_1) and abaxial side showing midrib and veins (H_2). Drawn by Mahsarahka Rungkrajang. Drawn from *Averyanov et al. VH 1648*.

Figure 5.8 Holotype of *Urophyllum brochidodromum* Yooprasert, Culham & Utteridge.

4. *Urophyllum chinense* Merr. & Chun subsp. ***latistipulum*** Yooprasert, Culham, Yahara & Utteridge **subsp. nov.** (Figure 5.9 and 5.10)

Diagnosis. *Urophyllum chinense* subsp. *latistipulum* is similar to *U. chinense* subsp. *chinense*, but differs in its stipules, which are broadly ovate to elliptic, 4.4–7.6 mm wide and with hairs present only along the midline; whereas in *U. chinense* subsp. *chinense*, the stipules are linear-oblong to ligulate, 1.7–2.9 mm wide with hairs present all over. The species also differs in its leaves which are abaxially glabrous to the naked eye and sparsely hairy under magnification, instead of conspicuously densely hairy, visible to the naked eye in *U. chinense* subsp. *chinense*.

Type. VIETNAM. Lam Dong Province: Lac Duong District, Da Nhim Commune, Bidoup-Nui Ba NP, Trail next to Ranger station Dung Iar Gieng, 12°09'35.9" N, 108°32'10.1" E, 1,621 m altitude, 3 June 2019, *Yooprasert et al. VN 112-2* (♀) (holotype: HN!; isotype: BKF!).

Additional specimens examined. VIETNAM. Lam Dong Province: 15 kil. [kilometre] au nord de Dankia, Langbiang, 1,500–1,800 m elev., 27 October 1930, *Poilane 18661* (♀) (P-photo! [P04950968]); Da Lat City, Rong Phon Forest, Suoivang, 1,500 m elev., 1 January 1998, *Wongprasert 981-15* (♀) (BKF!); ibid., Da Tong, Bidoup Nui Ba NP, Beside the road No. DT722, across a stream [12°5'50.4"N, 108°22'36"E], 1,711 m elev., 4 June 2019, *Yooprasert et al. VN 116-5* (♀) (BKF!, HN!); Dam Rong District, Da Long, Bidoup Nui Ba NP, Beside the road to Ranger station Da Long [12°9'21.5"N, 108°21'57.6"E], 1,495 m elev., 2 June 2019, *Yooprasert et al. VN 105-1* (sterile) (BKF!, K!); Lac Duong District, Da Chay [Da Chais Commune], 29 km to NE from Dalat City, on main peak of Bi Dup mt. system [12°5'N, 108°40'E], 2,260 m elev., 21 March 1997, *Averyanov et al. VH 2981* (♀) (AAU!); ibid., 31 km NE of Dalat City, vinicities Klong Lanh village, NW macroslope of Bi Dup mt system [12°8'N, 108°39'E], 1,850–1,950 m elev., 15 March 1997, *Averyanov et al. VH 2651* (♀) (AAU!); ibid., 20 March 1997, *Averyanov et al. VH 2941* (♀) (AAU!); ibid., 35 km to NE from Dalat City, W macroslope of Gia Rinh mt. ridge [12°9'N, 108°41'E], 1,800 m elev., 18 April 1997, *Averyanov et al. VH 4075* (♀) (AAU!); ibid., Bidoup Nui Ba NP [12°2'48.13"N, 108°26'6.67"E], 1,918 m elev., 24 June 2018, *Tagane et al. V 8982* (sterile) (DLU n.v., FU!); ibid., Mt. Langbian [12°2'46.3"N, 108°26'1.5"E], 1,905 m elev., 25 March 2018, *Yahara et al. V 7930* (sterile) (FU!); ibid., Trail by Ranger station Giang Ly [12°8'44.6"N, 108°40'17.2"E],

1,510 m elev., 1 June 2019, *Yooprasert et al. VN 102-7* (♀) (BKF!, HN!); ibid., [12°8'48.3"N, 108°40'16.6"E], 1,487 m elev., 1 June 2019, *Yooprasert et al. VN 100-1* (♀) (BKF!, HN!); ibid., Trail opposite Ranger station Giang Ly, crossing a stream [12°10'47.5"N, 108°41'6.9"E], 1,490 m elev., 1 June 2019, *Yooprasert et al. VN 96-1* (sterile) (BKF!, HN!); ibid., *Yooprasert et al. VN 96-2* (sterile) (BKF!, K!); ibid. [12°10'34.92"N, 108°41'8.35"E], 1,543 m elev., 22 June 2018, *Yahara et al.* (fl. bud) (DLU n.v., FU!); ibid., Da Nhim Commune, Bidoup Nui Ba NP, Trail next to Ranger station Dung Iar Gieng [12°7'59.9"N, 108°39'1.9"E], 1,439 m elev., 3 June 2019, *Yooprasert et al. VN 107-2* (sterile) (BKF!); ibid., Trail to Thac Thien Thai waterfall [12°8'41.4"N, 108°31'47.5"E], 1,510 m elev., 31 May 2019, *Yooprasert et al. VN 82-1* (sterile) (BKF!, K!); ibid., *Yooprasert et al. VN 82-2* (sterile) (BKF!, K!).

Etymology. The specific epithet derived from stipule's broader shape.

Distribution. Vietnam. Endemic, to date only known from Lam Dong Province (Figure 5.6).

Habitat. Primary broadleaved evergreen montane forest and mixed evergreen deciduous forest and bamboo thickets, on slope; elev.1,400–2,300 m.

Phenology. Collected in flower in June and fruit all year round.

Conservation status. Vulnerable (VU) B1ab(iii)+2ab(iii). The taxon is endemic and known from 18 collections within six locations in Lam Dong Province (ca. 15 km radius). The estimated EOO (441 km^2) and AOO (44 km^2) are within the threshold for the Endangered (EN) category, however five locations are in Bidoup-Nui Ba National Park (IUCN category Ia) where the risk of extinction is low. Only one population in Rung Thong Da Lat Cultural and Historical Site has shown the slowly expanding urban and cultivation areas in the past 12 years on the Google Earth imagery, however the area is listed in IUCN category V where the land use can be balanced between human and nature. For these reasons, this endemic subspecies is rated as Vulnerable.

Description

Shrub or *treelet* to 7 m, almost glabrous. *Branches* terete; young shoots flattened, ridged. *Stipules* caducous usually left only at apex, ovate to elliptic 1.2–2.1 cm long, 0.44–0.76 cm wide, appressed to the stem but not folded, abaxially densely hairy along midline, conspicuous netted veins. *Leaves* chartaceous, rarely coriaceous, usually elliptic to oblong, rarely ovate to obovate, 10.0–16.2 cm long, 3.3–6.4 cm wide, apex acuminate, base cuneate, adaxially glabrous, abaxially glabrous to the naked eye, sparsely hairy with short hairs <0.5 mm long under magnification, drying dull yellowish green to brown adaxially,

dull (rarely shiny) yellowish green abaxially with yellow to brown veins; secondary veins 7–11 pairs, weakly to festooned brochidodromous; tertiary veins mixed percurrent, angle increasing basally, middle of the leaves subperpendicular to obtuse, veins diverging at angle of c. 89°–113° from the midrib; petioles [0.5–]0.7–1.4[–2.1] cm long, canaliculate. *Inflorescence* axillary, pedunculate to subsessile umbellate, peduncle [0.5–]2.2–7.6[–16.9] mm long, bracts triangular with densely hairy at base, ca. 1 mm long, bearing 4–5 flowers in pistillate plant, nectary disk annular, only pistillate plant known. *Pistillate flowers* calyx fused, 1.0–2.2 mm long, toothed to shallow lobed 5, 0.3–0.8 mm deep; corolla fused forming a tube at base towards 1/3 their length, lobed 5, valvate, immature lobes ca. 1.5 mm long, densely hairs at throat inside; staminodes 5, anthers no pollen; ovary cupuliform, locules 5; style ca. 0.8 mm long in flower bud, sparsely scaly; stigma lobed 5, ca. 0.8 mm long, covered with scales. *Fruits* baccate, subglobose, 5.1–5.8 mm diam. *in sicco*; calyx and disk persistent. *Seed* numerous, subelliptic, surface alveolate, orange brown *in sicco*.

Remarks. *Urophyllum chinense* subsp. *latistipulum* is sympatric with *U. bidoupense* in Bidoup-Nui Ba National Park. They are morphologically similar being superficially glabrous in appearance, only sparsely hairy on the abaxial leaf surface under magnification. However, the taxon differs to *U. bidoupense* by its stipules being hairy along the midline abaxially and usually chartaceous leaves drying dull abaxially instead of subglabrous stipules with hairs found only at apex and coriaceous leaves drying shiny abaxially found in *U. bidoupense*. Other taxa found in the area are as discussed in the Remarks to *U. bidoupense*, which are *U. pseudoschmidtii* and *U. lecomtei*. The taxon differs from these taxa by its stipule hairy in the midline, where in the others stipule hairy all around.

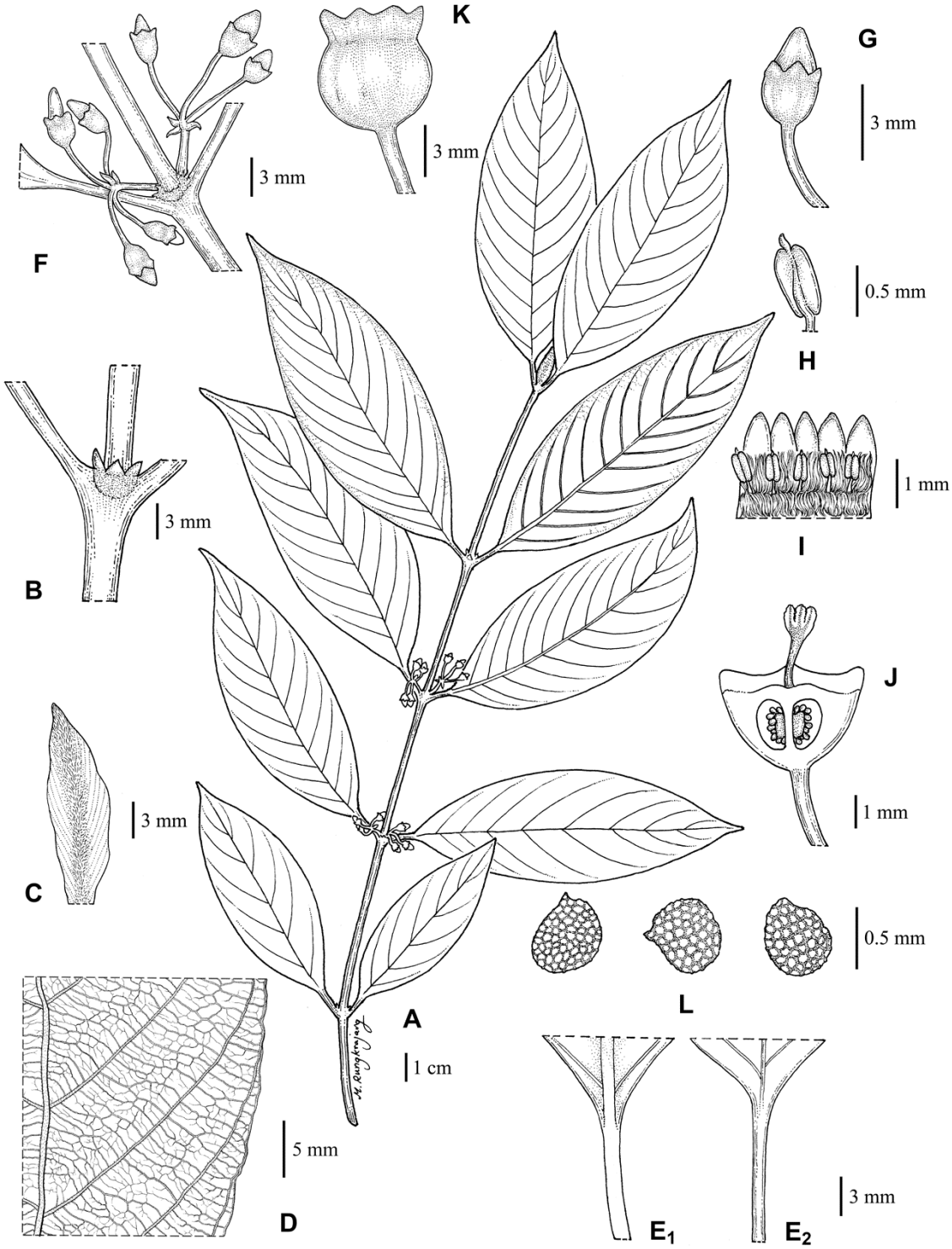


Figure 5.9 *Urophyllum chinense* subsp. *latistipulum* Yooprasert, Culham, Yahara & Utteridge. Illustration from pistillate plant. A Flowering branch. B Stem and stipule scar. C Stipule. D Abaxial leaf lamina. E Petiole, abaxial side (E_1) and adaxial side (E_2). F Inflorescence. G Flower bud; H Staminode; I Adaxial corolla showing staminodes and throat hairs arrangement. J Immature stigma positioning in a flower; K Fruit; L Seed. Drawn by Mahsarahka Rungkrajang. Drawn from *Yooprasert et al. VN 112-2*.

Figure 5.10 Holotype of *Urophyllum chinense* subsp. *latistipulum* Yooprasert, Culham, Yahara & Utteridge.

5. *Urophyllum pulchristipulum* Yooprasert, Culham & Utteridge **sp. nov.** (Figure 5.11 and 5.12)

Diagnosis. *Urophyllum pulchristipulum* is similar to *U. schmidtii* and *U. pseudoschmidtii*, but differs in its stipules, which are broadly subcordate, with conspicuous netted veins and abaxially glabrous, whereas in *U. schmidtii*, they are linear-oblong, have only inconspicuous netted veins and abaxially densely hairy. Its inflorescences are sessile umbellate rather than pedunculate umbellate as always found in *U. schimidtii*.

Type. CAMBODIA. Ratanikiri [Ratanakiri] Province: Virachay [Virachey] National Park, Western slopes and ridge leading to Mt. Yak Kham, East of Ho Chi Minh trail, 48P. 756282: 1587299 [14°20′45.24″N, 107°22′34.25″E], 800–1,100 m altitude, 13 Dec 2005, *Thomas et al. 26* (♀) (holotype E! [E00220453]; isotype: L-photo! [L.2972067]).

Etymology. The specific epithet refers to its stipules which have a beautiful subcordate shape and a neatly arranged network venation.

Distribution. Cambodia: Ratanakiri Province. Endemic, to date only known from the type specimen from Virachey National Park (Figure 5.3).

Habitat. Mixed evergreen forest and bamboo thickets, on slope; elev. 800–1,100 m.

Phenology. Flower season is unknown. A specimen with fruits collected in December.

Conservation status. Data deficient (DD). Because only one specimen of *Urophyllum pulchristipulum* has been collected from Virachey National Park (IUCN category II) in Cambodia, it does not provide enough data to estimate EOO and AOO for this species. The species has a potential to be found within nearby National Parks includes Chu Mom Ray (in Vietnam, IUCN category II), Dong Ampham and Xe Pian National Protected Areas (In Laos PDR, both in IUCN category VI) where the habitat is similar. Further collections in the areas should be carried out to assess existence of the species.

Description

Shrub. *Indumentum* almost glabrous, hairy only on petiole ridges, stipule base and margin. *Branches* quadrate to terete; young shoots flattened, ridged. *Stipules* caducous, usually left only at shoot, subcordate, 9–11 mm long, 3–5 mm wide, appressed to the stem but not folded, abaxially glabrous, conspicuous netted veins *in sicco*. *Leaves* chartaceous, elliptic to oblong, 5.4–8.3 cm long, 1.3–2.2 cm wide, apex acuminate, base cuneate to

obtuse; adaxially glabrous, abaxially sparsely hairy on veins, drying dull brown to dark green adaxially, dull light yellowish brown abaxially; secondary veins 7–8 pairs, festooned brochidodromous, angle decreasing basally; tertiary veins mixed percurrent, angle decreasing exmedially and increasing basally, obtuse angle to the midrib; petioles 3.7–5.3 mm long, canaliculate, adaxially densely hairy on ridges, abaxially glabrous. *Inflorescence* axillary, sessile umbellate, bracts lanceolate leafy-like, ca. 1.9 mm long, only fruits bearing pistillate plant known. *Fruits* baccate, subglobose, 1.6–2.5 mm diam. *in sicco*; pedicel 1.4–1.9 mm; calyx lobed 5, persistent, 0.7–0.9 mm long.

Remarks. Despite lacking flowers and staminate plants for *Urophyllum pulchristipulum*, it is unique by its subcordate stipules with the abaxial side glabrous, sessile umbellate inflorescence in pistillate plants and is only species recorded from north-eastern Cambodia (Ratanakiri Province) to date. The other *Urophyllum* species found in Cambodia is *U. schmidtii* recorded from eastern Thailand (Chanthaburi and Trat provinces) through south-western Cambodia (Koh Kong and Kampong Speu provinces), which has linear-lanceolate stipules, abaxially densely hairy and a pedunculate umbellate inflorescence. An additional morphologically similar species is *U. pseudoschmidtii*, however the stipules also differ from *U. pulchristipulum* in the same way as *U. schmidtii*.

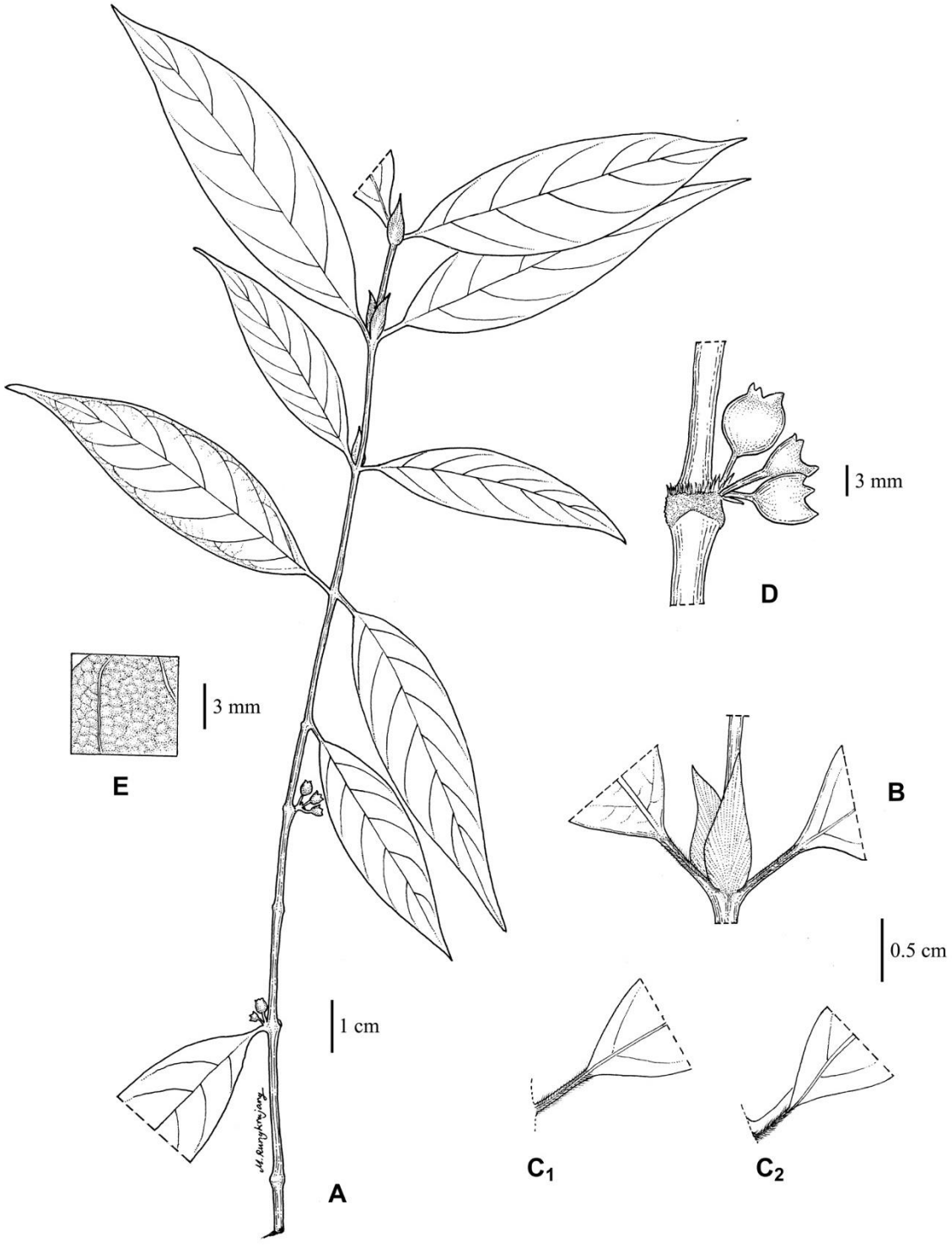


Figure 5.11 *Urophyllum pulchristipulum* Yooprasert, Culham & Utteridge. A Fruiting branch. B Stipule. C Petiole, adaxial view (C1), side view (C2). D Infructescence. E Abaxial leaf lamina. Drawn by Mahsarahka Rungkrajang. Drawn from *Thomas et al. 26*.

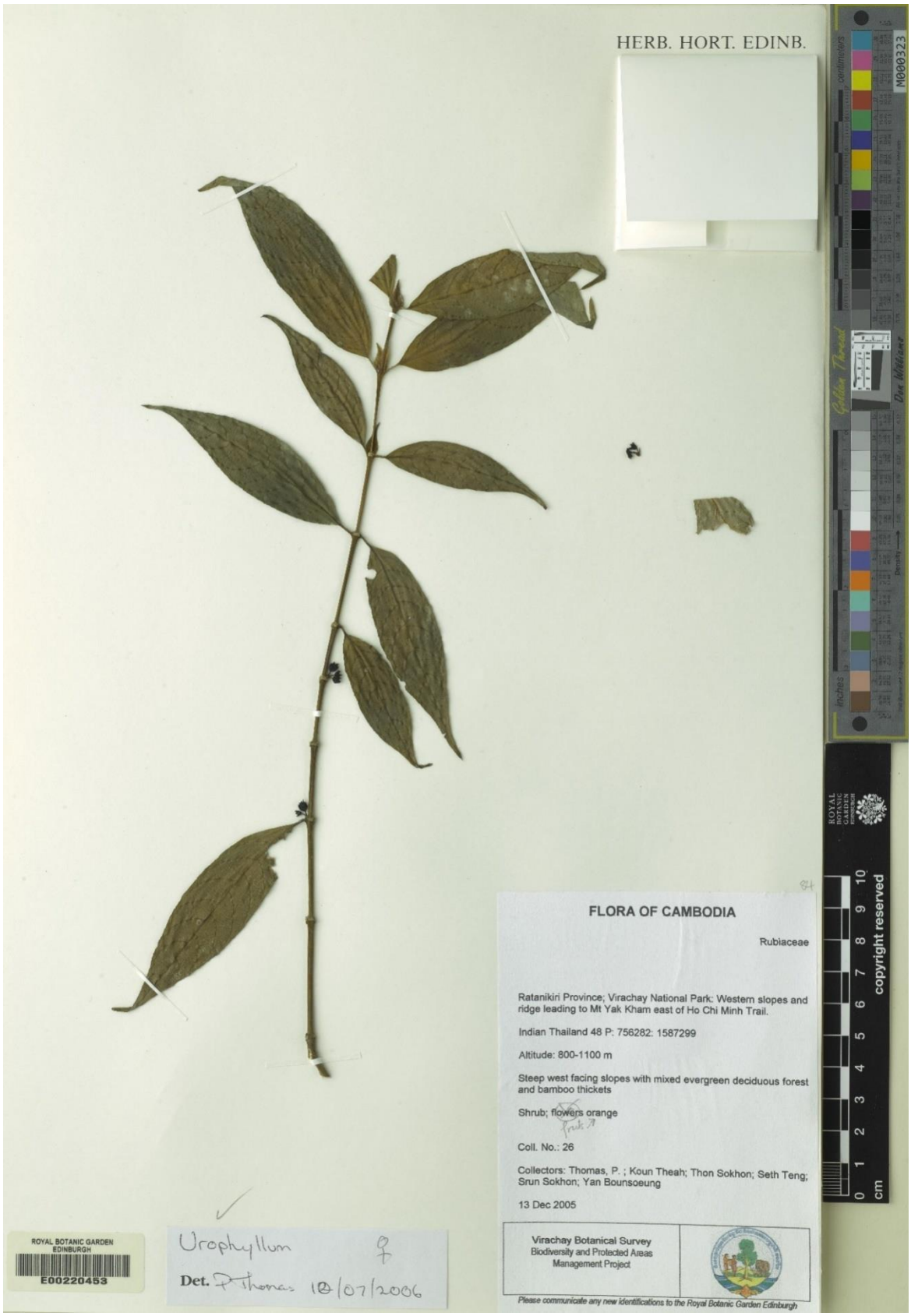


Figure 5.12 Holotype of *Urophyllum pulchristipulum* Yooprasert, Culham & Utteridge. Digitised image from Royal Botanic Garden Edinburgh (E), Specimen E00220453 (http://data.rbge.org.uk/herb/E00220453).

6. *Urophyllum pseudoschmidtii* Yooprasert, Culham, Yahara, Tagane & Utteridge **sp. nov.** (Figure 5.13 and 5.14)

Diagnosis. *Urophyllum pseudoschmidtii* is similar to *U. schmidtii*, but differs in its pistillate flowers, which lack staminodes which are always found in pistillate flowers of *U. schmidtii*. The species also differs in its staminate inflorescences, which are sessile umbellate and the flowers are seemingly sessile having very short pedicels; whereas *U. schmidtii* has pedunculate umbellate inflorescences and the flowers with conspicuous pedicels.

Type. VIETNAM. Ha Tinh Province: Vu Quang National Park, along trail to the summit of Mt. Rao CO [18°12′49.7″N, 105°23′42.8″E], 904 m altitude, 21 June 2016, *Yahara et al. V 5618a* (♀) (holotype K!; isotypes DLU n.v., FU!).

Additional specimens examined. VIETNAM. s.loc., 7 June 1921, *Hayata 205* (♀) (P-photo! [P03922484]); **Ha Tinh Province:** Vu Quang NP, in transect line 2 [18°16'9.1"N, 105°21'17.5"E], 649 m elev., 24 July 2015, *Yahara et al. V 3614* (sterile) (DLU n.v., FU!); ibid., along trail to the summit of Mt. Rao CO [18°12′49.7″N, 105°23′42.8″E], 904 m altitude, 21 June 2016, *Yahara et al. V 5618b* (♂) (DLV n.v., K!, FU!); **Lam Dong Province:** Grand Piton Lang-bian [Lang Biang], pres du village de Beneur [near the village Beneur], 1,500–2,000 m elev., 15 February 1914, *Chevalier 30836* (♀) (P-photo! [P03922421]); Lac Duong District, Da Chais Commune, Bidoup Nui Ba NP, Trail by Ranger station Giang Ly [12°8'49.1"N, 108°40'16.2"E], 1,469 m elev., 1 June 2019, *Yooprasert et al. VN 99-1* (♂) (BKF!, HN!); ibid., Da Nhim Commune, Bidoup Nui Ba NP, Trail next to Ranger station Dung Iar Gieng [12°9'35.9"N, 108°32'10.1"E], 1,621 m elev., 3 June 2019, *Yooprasert et al. VN 112-1a* (♂) (BKF!, HN!); ibid., *Yooprasert et al. VN 112-1b* (♂) (HN!, K!).

Etymology. The specific epithet refers to its similar morphological characters to *U. schmidtii*.

Distribution. Vietnam: Ha Tinh Province (Vu Quang National Park) and Lam Dong Province (Bidoup-Nui Ba National Park) (Figure 5.3).

Habitat. Evergreen hill forest, lower montane forest, by stream; elev. ca. 600–2,000 m.

Phenology. Collected in flower in June. Fruit is only known from two herbarium specimens collected in February and June.

Conservation status. Vulnerable (VU) D2. The conservation status of *Urophyllum pseudoschmidtii* is assessed using six specimens collected from four locations in Bidoup-Nui Ba National Park (IUCN category Ia) and Vu Quang National Park (IUCN category II). The other two specimens, *Chevalier 30836* and *Hayata 205*, were not included due to lacking precise localities and old ages (in 1914 and 1921, respectively). The estimated AOO (16 km^2) of these six specimens is within the Endangered threshold, however the species is found in protected areas where the deforestation has limited chance. Even though, the estimated EOO (5,135 km^2) is within the Vulnerable threshold, there is a big gap between these two National Parks that causing habitat fragmentation and meaning estimated EOO is not reflected the real species occurrence. Therefore, *U. pseudoschmidtii* is rated as Vulnerable with restricted AOO (<20 km^2) and small number of locations found. This assessment could be reviewed to Near Threatened or Least Concern if there is an evidence of species existing in nearby area such as Nakai-Nam Theun National Park (Laos, IUCN category II) or Phong Nha-Ke Bang National Park (IUCN category II).

Description

Shrub to 1 m high. *Indumentum* sparsely to densely hairy except leaves and corolla. *Branches* terete; young shoots flattened, ridged. *Stipules* caducous, present 1–2 nodes at shoot, linear-lanceolate to oblong-lanceolate, 1.0–1.1 cm long, 0.8–1.4 mm wide, appressed to the stem but not folded, abaxially sparsely to densely hairy. *Leaves* chartaceous, ovate or elliptic, 10–13 cm long, 3.5–4.5 cm wide, apex acuminate, base cuneate to obtuse, adaxially glabrous, abaxially subglabrous, sparsely hairy with short hairs <1 mm long on midrib to tertiary veins, drying dark greenish brown adaxially, yellowish to dark green abaxially with brown midrib and yellow veins; secondary veins [6–]8–10 pairs, festooned brochidodromous, angle decreasing basally; tertiary veins mixed percurrent, angle decreasing exmedially and increasing basally, obtuse angle to the midrib; petioles 4–6 mm long, canaliculate, adaxially densely hairy on ridges, abaxially sparsely hairy. *Inflorescence* axillary, staminate plant sessile umbellate, bracts minute triangular ca. 0.8 mm long, adaxially and abaxially sparsely hairy, pedicels very short ca. 0.5 mm long, nectary disk dome shaped; pistillate plant sessile, pedunculated to two-tier umbellate, peduncle 0–1.5 mm long, rachis 1.5–2 mm long, bracts lanceolate leafy-like, 2.3–5.2 mm long, adaxially glabrous, abaxially densely hairy, pedicels 2.8–7.9 mm long, nectary disk annular. *Staminate flowers* calyx fused, 5-lobed, 0.5–0.8 mm deep; corolla 5-lobed, valvate,

fused to form a tube at base towards 1/3–1/2 their length, lobed ca. 1.5 mm long, densely hairy at throat inside; stamens 5, filament fused to the corolla tube, anthers 2-celled, dorsifixed, introrse; pistillode present, ovary vestigial. *Pistillate flowers* calyx and corolla as staminate flowers; staminodes absent; style ca. 1.9 mm long, sparsely scaly; stigma lobed 5, ca. 1.4 mm long, covered with scales.

Remarks. *Urophyllum pseudoschmidtii* is recognised by its pistillate flowers without staminodes. The population in Bidoup-Nui Ba National Park has a distribution overlapping with *U. lecomtei*. However, the species differs from *U. lecomtei* by its petioles with densely hairy on canalicular ridges, abaxially corolla glabrous and pistillate flowers lacking staminodes, instead of petioles with similar hairs density all over, abaxially corolla hairy and pistillate flower with staminodes as found in *U. lecomtei*. Other taxa found in the same area are *U. bidoupense* and *U. chinense* subsp. *latistipulum*. The differences have been discussed earlier in both taxa.

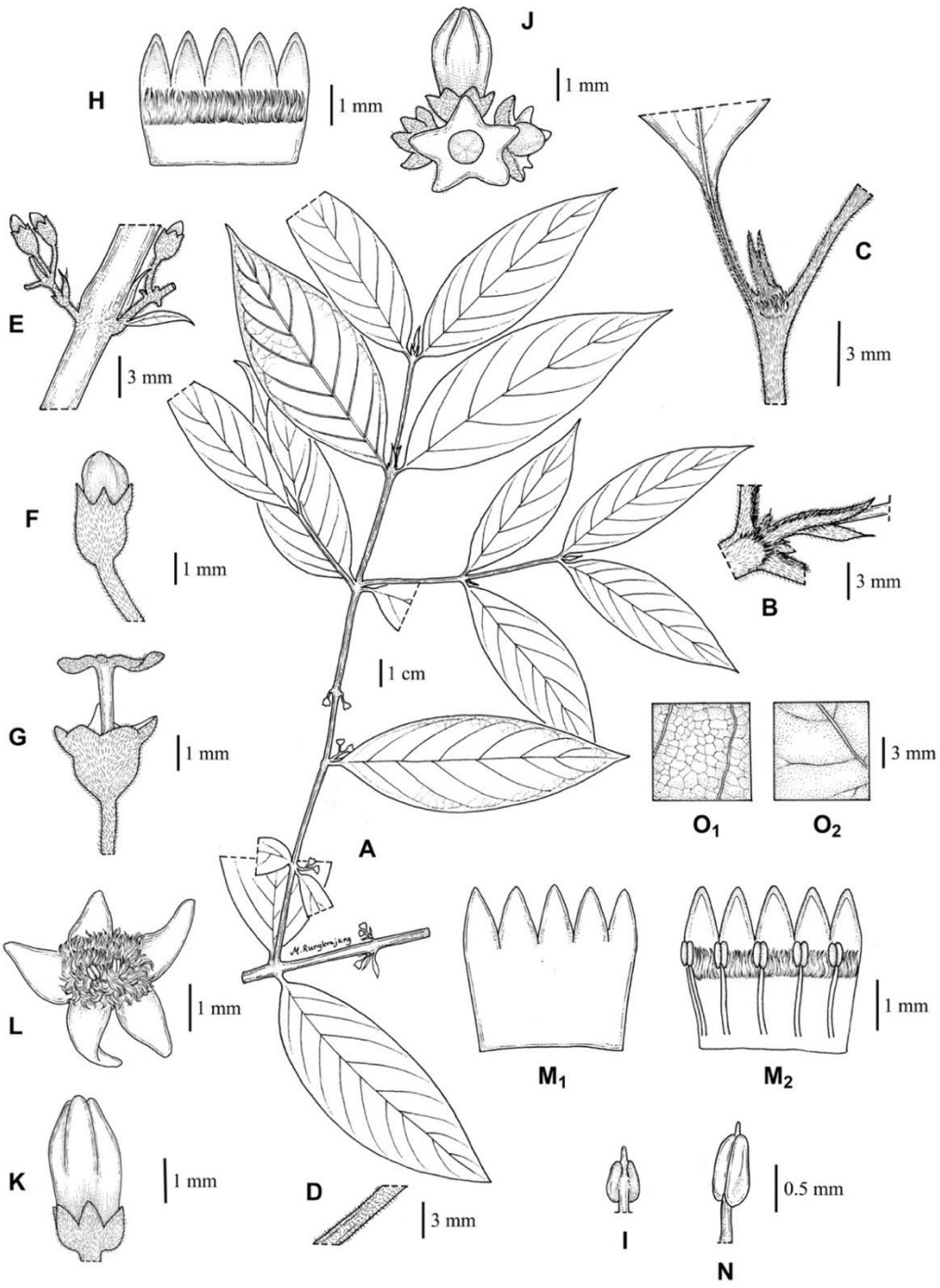


Figure 5.13 *Urophyllum pseudoschmidtii* Yooprasert, Culham, Yahara, Tagane & Utteridge. A Flowering branch. B Stipule. C Petiole. D Young stem. E–I Pistillate plant: E Inflorescence; F Flower bud; G Stigma positioning in a flower; H Adaxial corolla showing throat hairs arrangement; I Staminode. J–N Staminate plant: J Inflorescence; K Flower bud; L Opening flower (top view); M Corolla, abaxial side (M_1); adaxial side showing stamens and throat hairs arrangement (M_2); N Stamen. O Leaf lamina, abaxial side (O_1) and adaxial side (O_2). Drawn by Mahsarahka Rungkrajang. Drawn from: A-I, O *Yahara et al. V 5618a*; J–M *Yahara et al. V 5618b*.

Figure 5.14 Holotype of *Urophyllum pseudoschmidtii* Yooprasert, Culham, Yahara, Tagane & Utteridge.

5.6. Key to *Urophyllum* species in Cambodia, Laos and Vietnam

1. Stipule glabrous to subglabrous abaxially, sparse hairs present at margin and dense at base adaxially or only at apex abaxially, sometimes hairy only along midline; netted veins conspicuous *in sicco*.....2

   Stipule hairy abaxially throughout; netted veins inconspicuous *in sicco*.....4

2.(1) Petiole glabrous to the naked eye; inflorescence pedunculate umbellate; Vietnam..3

   Petiole hairy to the naked eye on canalicular ridges; inflorescence sessile umbellate; Cambodia ..... ***Urophyllum pulchristipulum***

3.(2) Stipule glabrous to subglabrous with hairs only at the apex ***Urophyllum bidoupense***

   Stipule hairy along the midline from base to apex.....

   ..... ***Urophylllum chinense* subsp. *latistipulum***

4.(1) Petiole hairs conspicuously denser on canalicular ridges, sometimes hairy only on the ridges.....5

   Petiole hairs similar density all over.....7

5.(4) Stipule margin and petiole canalicular ridges with pale cream hairs denser than other areas ..... ***Urophyllum parviflorum***

   Stipule hair density similar throughout; petiole canalicular ridges with hairs not different in density and colour 6

6.(5) Staminate inflorescence sessile umbellate; staminodes absent in pistillate flowers; Vietnam ..... ***Urophyllum pseudoschmidtii***

   Staminate inflorescence pedunculate umbellate; staminodes present in pistillate flowers; Thailand and Cambodia ..... ***Urophyllum schmidtii***

7.(4) Secondary veins weakly brochidodromous (loop inconspicuous) or eucamptodromous.....8

   Secondary veins conspicuously brochidodromous or festooned brochidodromous10

8.(7) All parts covered with long sericeous hairs; corolla with sparse sericeous hairs abaxially..... ***Urophyllum argenteum***

   All parts subglabrous or covered with hairs but not sericeous; corolla glabrous abaxially.....9

9.(8) Leaves coriaceous, abaxial lamina densely appressed hairy; stipules appressed to the stem, sometimes recurved at margins but not folded; inflorescence single

flower or pedunculate umbellate, small number of flowers usually <5; corolla white ..... ***Urophyllum chinense* subsp. *chinense***

Leaves chartaceous, abaxial lamina glabrous; stipule folded toward adaxial side; inflorescence compound cymose, many flower >7; corolla green ***Urophyllum annamense***

10.(7) Adaxial leaf veins glabrous; leaf reddish brown *in sicco*; tertiary veins not prominent, inconspicuous abaxially; inflorescence sparsely hairy; abaxial corolla surface glabrous..... ***Urophyllum tsaianum***

Adaxial leaf veins hairy, sometimes only on midrib; leaf yellowish or greenish brown to dark grey *in sicco*; tertiary veins prominent, conspicuous abaxially; inflorescence densely hairy; abaxial corolla surface hairy..... 11

11.(10) Mature leaves with secondary veins hairy adaxially, brochidodromous without a series of small loops outside the main loop or inconspicuous; flower and fruit calyx lobes recurved ..... ***Urophylllum brochidodromum***

Mature leaves with secondary veins glabrous, brochidodromous with a series of small loops (festooned); flower and fruit calyx lobes erect or folded toward adaxial side..... 12

12.(11) Pistillate plant with pedunculate umbellate inflorescence..... ***Urophylllum lecomtei***

Pistillate plant with sessile umbellate inflorescence..... ***Urophyllum tonkinense***

Acknowledgements

We thank the curators and staff of AAU, C and FU herbaria that sent materials on loan to be studied; BM, CMUB, E, HN, HNU, VNM and K for giving accesses to the collections in the herbaria; Naturalis Biodiversity Centre (L, U and WAG), NY, Muséum national d'Histoire naturelle (MNHN)–P for digitised image online database. The Royal Thai government and Department of National Parks, Wildlife and Plant Conservation for financial support and an opportunity of Sawita Yooprasert's PhD research. Rachun Pooma and Nannapat Pattharahirantricin (BKF) for advice and support, BKF staff for processing voucher collections, Mahsarahka Rungkrajang for the illustrations, Lesley Scott and Natalie Zarte (E) for a high-resolution digitised image.

Chapter 6 General discussion

6.1. Overview of the thesis

Rubiaceae is a major family of woody and herbaceous plants that includes many economically important plants such as *Coffea* L., *Cinchona* L., and ornamental plants such as *Gardenia* J.Ellis and *Ixora* L.. The family is in need of taxonomic treatment for the Flora of Thailand project, and this thesis has provided a method to test species concepts and therefore produce robust taxonomic accounts that can be used more widely in other taxa. A novel combination of morphological and molecular data and techniques have been applied to produce a taxonomic account for *Urophyllum* in Thailand and mainland Indochina to evaluate the species delimitations. Both linear and geometric morphometric data were gathered to conduct analyses combined with supervised machine learning (Chapter 2). The results from Chapter 2 demonstrate that this is an effective method for the classification of *Urophyllum* taxa, as 10 out of the 13 (91% success rate) *Urophyllum* taxa sampled were successfully identified. The misclassification of three species, *U. longifolium*, *U. glabrum*, and *U. crassum* and the phylogenetic relationships of taxa in the genus were investigated further in the most comprehensive molecular analysis of the genus to date (Chapter 3). Both whole plastid and nrDNA sequences were used and compared with morphological data. The results of Chapter 2 and 3 informed the taxonomic revision of *Urophyllum* in Thailand (Chapter 4) and identified four new taxa in Vietnam (Chapter 5). A new species from Cambodia was also described based only upon morphological characters (Chapter 5). The publication of new taxa and crucially the development of an identification key provides a basis for the Flora of Cambodia, Laos, and Vietnam. Moreover, this thesis provides valuable novel data for a poorly known plant group such as *Urophyllum* with recommendations of how to study the group in the future.

6.2. From description-based taxonomy to an integrated discipline

Morphological characters are the most readily available source of data for taxonomists, as they are convenient and easy to collect (Christodoulou, Clark and Culham, 2020). A routine approach many taxonomists have employed for separating species is to observe specimens and select characters for identification, specimens can then be sorted into groups based upon the possession of similar characters. Whilst this is appropriate initially,

an explicit and robust methodology (for example, detailed quantitative analyses) is often lacking, and this causes species delimitations to be disputed (Henderson, 2005). Taxonomic revision, like any other branch of science, is a hypothesis driven and tests species boundaries (Wheeler, 2004). To make taxonomic studies more robust, supporting evidence, the species concept employed, and quantitative methodology need to be included.

There are two types of morphometric data that are commonly applied: linear (distance measurement) and geometric (shape analysis). Linear morphometrics have been applied more widely to address questions of plant classification, largely due to the simplicity in gathering data (Christodoulou, Clark and Culham, 2020). Da Costa *et al.* (2009) studied the variation of both vegetative and reproductive parts in the *Vriesea paraibica* Wawra complex (Bromeliaceae), identifying four species. Similarly, Nagahama *et al.* (2014) used linear morphometrics to separate two species within the *Andropogon lateralis* Nees complex and their hybrids (Poaceae). As well as being a more modern approach, geometric morphometric data can detect subtle differences in shape more easily compared to linear morphometrics. Geometric data of the lip shape in *Dactylorhiza* Neck. orchids has been used for species identification and to confirm hybrid taxa (Shipunov and Bateman, 2005). Liu (2018), also gathered geometric morphometric data of *Quercus* L. leaf shapes among sympatric species and was able to distinguish *Q. dentata* Thunb. and *Q. aliena* Blume and their hybrids. A common approach is to combine the two types of data to investigate species boundaries (Shipunov and Bateman, 2005; Menini Neto, Van den Berg and Forzza, 2019). However, this project is different from these earlier studies as it combines the analyses of morphometric data with supervised machine learning (ML), to provide the quantitative morphological evidence of 13 pre-grouped *Urophyllum* taxa found in Thailand and Vietnam. The results reveal that ML trained on linear data alone can successfully classify seven of the *Urophyllum* taxa (100% success rate), however, the combination of linear and geometric data increased the successful classification to 10 of the 13 *Urophyllum* taxa (100% success rate). Similar results were also found in apple cultivar identification by using ML on both linear and geometric data gained higher prediction rate than those on each dataset individually (Christodoulou, Battey and Culham, 2018). This thesis therefore provides strong evidence for the use of ML approaches that

have been trained upon a combination of linear and geometric data for taxonomic classification. Although *Urophyllum* has been used as a study group to test the classification of groups (or taxa) using ML methods in this thesis, the methods have a much wider application not only to provide the basis for taxonomic accounts, but also outside of botany (Guisande *et al.*, 2010; Santana *et al.*, 2014; da Silva *et al.*, 2015).

Supervised machine learning is a powerful tool that can be trained on a dataset to create a classifier that can be used on new data (Kotsiantis, 2007), this process is fundamentally similar to grouping specimens in plant classification. Machine learning has the additional advantage of using statistical analyses and no human interference in the classification step. However, a researcher decides which characters to gather the data from. Therefore, the process of applying ML in taxonomy can provide a valuable quantitative basis for pre-defined groups of species that are often compiled in traditional taxonomic studies in an intuitive manner. Machine learning approaches can also be used with data from external morphological characters, which prevents the destruction of specimens (Christodoulou, Battey and Culham, 2018), a similar method that is usually required when using herbarium specimens, and therefore this method can harness the vast amount of data available in herbaria. Furthermore, it also allows for digitised specimens to be used for plant classification. Machine learning is becoming a popular approach in several other fields (Christodoulou, Clark and Culham, 2020) but is not yet a common approach in taxonomic studies. The performance of combining machine learning and morphometrics in apple cultivar identification (Christodoulou, Battey and Culham, 2018) and *Urophyllum* species classification in this study, demonstrates that it can be applied to taxonomic studies. The success rate of classifying *Urophyllum* specimens using ML was 91% (Chapter 2); however, there were three taxa (*U. longifolium*, *U. glabrum*, and *U. crassum*) that were sometimes misclassified. The misclassifications were due to the morphological characters used which were not informative enough to classify them. Key diagnostic characters such as flower colour (for *U. glabrum* misclassifications to *U. longifolium*) could not be used as this character is poorly preserved in herbaria collections. It is therefore recommended that characters that do not preserve well when drying herbarium specimens are recorded in new collections, such as recording flower colour with a standard colour chart (e.g. RHS colour chart). It should not be ruled out that misclassifications can occur due to either the

taxa being closely related, or they are not identifiable as distinct groups. In this case, further analysis, such as molecular studies, should be conducted to compliment morphological data, this combination of techniques was used in this thesis to study the phylogenetic relationship of *Urophyllum* and the misclassified taxa in Chapter 2.

With the advances in next generation sequencing (NGS) over the past decade it is now possible to gather molecular data from high copy regions of the genome easily and with high accuracy. In this thesis, the plastomes and nrDNA cistrons of 39 *Urophyllum* samples (18 species) were assembled. A comparative analysis has identified variable nucleotide regions that can be developed as DNA markers to provide a molecular tool for species identification in *Urophyllum*.

Phylogenetic analysis in *Urophyllum* was previously based on four regions in the plastome and nrDNA, to study the relationship of *Urophyllum* and its closely related genera (Smedmark and Bremer, 2011). Smedmark & Bremer (2011) included only four species of *Urophyllum* found in Thailand. However, this project, sampled a greater number of species of *Urophyllum* found in Thailand and Vietnam. Moreover, NGS sequencing was used to assemble whole plastomes and mine nrDNA regions to construct phylogenetic trees. Therefore, this thesis provides the most comprehensive phylogenetic study of *Urophyllum* to date. Incongruence between tree topologies of plastome and nrDNA data was identified in Chapter 3; conflicting patterns of tree topology between plastid and nrDNA trees has also been reported in many plant groups (Fehrer *et al.*, 2007; French, Brown and Bayly, 2016; Ji *et al.*, 2019; Wikström, Bremer and Rydin, 2020). Conflict is thought to be due to hybridisation, incomplete lineage sorting, and horizontal gene transfer (Philippe *et al.*, 2005; Folk, Mandel and Freudenstein, 2017). The nrDNA tree is largely congruent to the classification of taxa using machine learning (Chapter 2). The three misclassified taxa identified in Chapter 2 were also resolved using nrDNA. This project reveals that morphological and phylogenetic approaches are a good combination to provide robust evidence to delimit species. Where incongruence between tree topology occurs between different molecular datasets, morphological data can be used to provide additional evidence to resolve species boundaries. An integrated approach of applying both phylogenomic and morphometric data for species delimitation has also been successfully used in other plant groups. Frajman *et al.* (2019) used a combination of phylogenomics,

phylogenetics and linear morphometrics to delimit the Eurasian *Euphorbia seguieriana* Neck. *s.l.*, Karbstein *et al.* (2020) used target enrichment sequencing together with geometric morphometrics to delimit the Eurasian populations of the *Ranunculus auricomus* L. complex (Ranunculaceae). In the study of *Urtica dioica* L. complex Rejlová *et al.* (2021) used target enrichment sequencing together with linear and geometric morphometrics to study the differences between two ploidy levels. To date, the combination of morphometric and molecular data, seem to be used primarily in resolving species complexes. However, the study of *Urophyllum* here demonstrates that a combination of these datasets can be used in the taxonomic revision of a genus.

6.3. Plants do not respect political boundaries

The initial aim of this study was to provide a taxonomic revision of *Urophyllum* to contribute towards the Rubiaceae account for the Flora of Thailand project. However, Thailand is located between four biogeographical regions, thus the flora is heavily influenced by the Indochinese, Indo-Burmese and Malesian regions (Van Welzen *et al.*, 2011). This highlights that many plant species found in Thailand are likely to be found in the neighbouring countries and not limited by a political border. This was also found in the synopsis of *Urophyllum* in this study (Chapter 4), as all the species recorded in Thailand can also be found in Cambodia, Myanmar, and Peninsular Malaysia. Furthermore, the Flora of China lists two endemic species (*U. parviflorum* F.C.How ex H.S.Lo and *U. tsaianum* F.C.How ex H.S.Lo), however, both species can be found in either Vietnam (*U. tsaianum*, specimens in HN and FU) or Laos (*U. parviflorum*, specimens in CMUB, L and P). These species have been included in the identification key of *Urophyllum* in Cambodia, Laos and Vietnam (Chapter 5). It is therefore essential to investigate specimens from the entire region when undertaking a taxonomic account of a particular plant group. Wood *et al.* (2020) discusses the consequences of taxonomic revisions limited to local regions in the monograph of the large sweet potato genus, *Ipomoea* L. A study that covers the whole distribution of a species can also provide the best account of species records which will help to provide the most accurate conservation assessment. This highlights the need for a more collaborative approach between researchers in different countries of SE Asia. The collaboration of researchers would also help to speed up the taxonomic work at regional levels (Cámara-Leret *et al.*, 2020).

What are the factors, therefore, that currently limit local taxonomists expanding their work outside of their own country? Several factors are thought to provide difficulty for local taxonomists (pers. comm. Rachun Pooma (Director of BKF)): 1) there is a lack of collaboration among herbaria and taxonomists within the same biogeographical region as newly collected specimens are exchanged with well-known and established herbaria (such as E, K, L, P, and NY) rather than local herbaria in neighbouring countries. For example, *Urophyllum* specimens deposited in BKF are mainly local collections within Thailand, with duplicates typically found in AAU, C, E and K. More consideration should be given to specimen exchange with herbaria in neighbouring countries; 2) the difficulty of getting permits to collect samples - this can be challenging due to the complex and highly variable system for granting permits in different countries. In Thailand, researchers are required to apply for permits from both the Forest and Plant Conservation Research office and each national park. Although, more recently foreign researchers, are required to collaborate with a Thai institution prior applying for a permit (http://park.dnp.go.th/dnp/media/media_110209_54214.pdf, access 4 May 2021), this system is not in place in neighbouring countries; 3) funding usually limits researchers to research within their country with limited opportunity to fund plant collections in neighbouring countries.

The first issue above can be largely overcome by providing online open access to herbaria, as is the aim of the WFO (http://www.worldfloraonline.org, accessed 4 May 2021). Cámara-Leret *et al.* (2020) suggest that access to specimens and literature online would speed up the compilation of species checklists in hyperdiverse regions. Online access to specimens removes the need to travel to visit herbaria; however, fieldwork remains a vital component of any study, to observe plants in their habitat, and therefore access to permits is a necessity. Issues two and three are more challenging but could be mitigated by collaboration between research groups working in the region. This can be achieved by networking and collaboration at conferences such as the Flora of Thailand and Flora Malesiana.

6.4. Future work

6.4.1. The future plan for the study of *Urophyllum*

This thesis provides a detailed study of the genus *Urophyllum* in Thailand and mainland Indo China, with the most comprehensive sampling to date (Smedmark and Bremer, 2011). The study can be viewed as a start of the development of a monograph for the genus. Therefore, there are several directions for further study into the genus. These include:

1) Phylogenetic construction based on mitochondrial genes and further study on nuclear single copy genes.

Intracellular gene transfer between genomes within land plants is a common process including the transfer of plastid DNA to mitochondrion and nuclear genomes or plastid and mitochondrial DNA to nuclear and vice-versa (Alverson *et al.*, 2010; Smith, 2011; Raman *et al.*, 2019). The phylogenetic relationships of *Urophyllum* were studied using whole plastid genomes and nrDNA sequences and revealed the incongruence between these two data. Mitochondrial genomes have been used successfully to resolve the phylogenetic relationship of Mediterranean olive species where the plastid genomes could not (Van de Paer, Bouchez and Besnard, 2018). Therefore, gathering genetic data from mitochondrial DNA may provide a better understanding of this incongruence between the two datasets in this thesis. The genome skim data gathered in this thesis could be used to retrieve mitochondrial genes (Ripma, Simpson and Hasenstab-Lehman, 2014), however the genes were not mined during this project due to constraints on time. Moreover, as DNA sequencing technology is developing at a fast pace, there is opportunity to retrieve single copy nuclearDNA regions, for example using Angiosperm353 (Johnson *et al.*, 2019), this will help understand the relationship of *Urophyllum* from both parental lineages. Together, these data can be contributed towards the monograph of the genus.

2) The taxonomic revision of *Urophyllum* in Cambodia, Laos and Vietnam, and other regions to improve conservation status assessments.

There are a small number of *Urophyllum* specimens recorded from Cambodia and Laos (<20 specimens) to date. This was the most important point that limited the taxonomic work for mainland Indochina. The availability of more specimens would help to provide

valuable information on species occurrences to accurately evaluate the conservation status of the species in the whole region. Moreover, occurrence data from specimens in Vietnam is poorly databased. Databasing herbarium collections would help to progress taxonomic works for *Urophyllum* in particular and many species occur there.

There are approximately 120 species of *Urophyllum* (Puff, Chayamarit and Chamchumroon, 2005; Smedmark and Bremer, 2011). The number of species studied in this project accounts for approximately 10% of the total species in the genus. For *Urophyllum,* the most diverse regions in terms of species richness are Indonesia and Malaysia (including Borneo, Java and Sumatra), according to the protologues (pers. obs.). With the risk of habitat loss from deforestation in these areas, it is an urgent to revise the taxonomy of the whole genus to provide an accurate assessment of each species conservation status.

A monograph of *Urophyllum,* therefore, remains a large undertaking, although revisions in species rich groups, such as *Ipomoea* (~425 species) demonstrate it is achievable. Moreover, the development of methods in this thesis provides a framework for testing species delimitation in *Urophyllum* that can be used across the distribution range of the genus. For example, a combination of machine learning and morphometrics could be used to build an identification platform to investigate the morphological similarity between species; and informative plastid DNA regions identified in Chapter 3 can be used for species delimitation (*matK*, *rpoB–trnC* and *petN–psbM*).

3) Generic delimitation of *Urophyllum s.l.*

According to Bremekamp (1940) and Smedmark and Bremer (2011), *Urophyllum s.l.* is paraphyletic. There remain eight closely related genera that have not been included in a phylogenetic study of Urophylleae to date. The diagnostic characters of some closely related genera proposed by Bremekamp (1940) can be found in *Urophyllum* species as shown in the synopsis in this study, these included the presence of leaf domatia (*Antherostele* p.p.), hairs in corolla throat inserted on a scale at the base of corolla lobes (*Lepidostoma*), hairs in corolla throat forming a ring and stipules glabrous inside (*Leucolophus*), and corolla throat densely covered with stiff, white hairs (*Praravinia* p.p.). This suggests that many closely related genera share morphological characters with

Urophyllum s.l., the extent of this close affinity should therefore be researched using an integrative approach of morphological and molecular data. As stated in Smedmark and Bremer (2011), obtaining plant samples of these genera was challenging, although herbarium specimens could be sampled to overcome this (Nevill *et al.*, 2020). It is important to include samples from these genera in a future phylogenetic study to understand the relationship of *Urophyllum s.l.* and its closely related genera.

6.4.2. Impact of the work for the flora of Thailand

The application of linear and geometric morphometrics combined with supervised machine learning can be a useful tool for taxonomic revision in the Flora of Thailand project moving forwards. Data can be easily gathered directly from herbarium specimens deposited in herbaria and digitised specimens online. This provides an affordable method for examining specimens that relies on free software (MorphoJ and R). It can be developed to use in routine plant identification for the staff both in the herbarium and in protected areas, which would help to increase the number of occurrence records of species in poorly recorded groups such as *Urophyllum*. The use of NGS is becoming increasingly widespread, although there is still an obstacle to use these methods in some regions, as access to facilities can be limited within organisations. Phylogenomic studies supporting morphometric investigations could therefore be undertaken in a more collaborative nature. For example, herbaria can provide access to specimens for sampling, and perform morphometric analyses with machine learning, collaborating with a university with the facilities to undertake molecular work. This provides a pragmatic approach to undertaking integrative taxonomy to ensure taxonomic treatments are thorough and robust.

6.5. Conclusion

This study has demonstrated the use of a novel morphological based statistical technique to classify *Urophyllum* species in Thailand and some species in mainland Indochina. The technique includes the combination of linear and geometric morphometrics data with supervised machine learning to perform the species classification. This study demonstrates the use of this method for classification of *Urophyllum* species, but also highlights the wider use for taxonomic research, and for the Flora of Thailand project. Whole plastid genomes of *Urophyllum* were also assembled, providing a comparative analysis to identify

new variable gene regions that can be used for marker development for species identification. The comprehensive phylogenetic relationship of *Urophyllum* species were provided based on the assembled whole plastid genomes and nrDNA sequences. Incongruent tree topologies between these two datasets were identified, nrDNA tree was largely congruent with morphological data. The morphological and phylogenetic results were used to produce a robust taxonomic revision of *Urophyllum* in Thailand and identify new species in Cambodia and Vietnam.

References

Adams, D. C., Rohlf, F. J. and Slice, D. E. (2004) 'Geometric morphometrics: ten years of progress following the "revolution"', *Italian Journal of Zoology*, 71(1), pp. 5–16. doi: 10.1080/11250000409356545.

Adams, D. C., Rohlf, F. J. and Slice, D. E. (2013) 'A field comes of age: geometric morphometrics in the 21st century', *Hystrix, the Italian Journal of Mammalogy*, 24(1), pp. 7–14. doi: 10.4404/hystrix-24.1-6283.

Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K. and Palmer, J. D. (2010) 'Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae)', *Molecular Biology and Evolution*, 27(6), pp. 1436–1448. doi: 10.1093/molbev/msq029.

Amiryousefi, A., Hyvönen, J. and Poczai, P. (2018) 'IRscope: an online program to visualize the junction sites of chloroplast genomes', *Bioinformatics*, 34(17), pp. 3030–3031. doi: 10.1093/bioinformatics/bty220.

Angermueller, C., Pärnamaa, T., Parts, L. and Stegle, O. (2016) 'Deep learning for computational biology', *Molecular Systems Biology*, 12(7), p. 878. doi: https://doi.org/10.15252/msb.20156651.

Bachman, S., Moat, J., Hill, A. W., de Torre, J. and Scott, B. (2011) 'Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool', *ZooKeys*. 2011/11/28, (150), pp. 117–126. doi: 10.3897/zookeys.150.2109.

Baum, D. A. and Smith, S. D. (2012) *Tree thinking: an introduction to phylogenetic biology*. New York: W.H.Freeman and Company.

Beentje, H. (2016) *The Kew Plant Glossary: an illustration dictionary of plant term*. Second. Surrey: Kew publishing, Royal Botanic Gardens, Kew.

Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M. (2017) 'MISA-web: a web server for microsatellite prediction', *Bioinformatics*, 33(16), pp. 2583–2585. doi: 10.1093/bioinformatics/btx198.

Birky, C. W. (1995) 'Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution', *Proceedings of the National Academy of Sciences*, 92(25), p. 11331 LP-11338. doi: 10.1073/pnas.92.25.11331.

Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M.-L. and Ruhlman, T. A. (2016) 'Variable presence of the inverted repeat and plastome stability in *Erodium*', *Annals of Botany*, 117(7), pp. 1209–1220. doi: 10.1093/aob/mcw065.

De Block, P., Rakotonasolo, F., Ntore, S., G. Razafimandimbison, S. and Janssens, S. (2018) 'Four new endemic genera of Rubiaceae (Pavetteae) from Madagascar represent multiple radiations into drylands', *PhytoKeys*, 99, pp. 1–66. Available at: https://doi.org/10.3897/phytokeys.99.23713.

Blume, C. (1823) *Catalogus van eenige der Merkwaardigste Zoo in- als Uitheemse*

Gewassen te Vinden in 's Lands Plantentuin te Buitenzorg. Batavia (Jakarta).

Blume, C. L. (1826) *Bijdragen tot de flora van Nederlandsch Indië*. Batavia (Jakarta): Ter Lands Drukkerij,. Available at: https://www.biodiversitylibrary.org/item/9225.

Bookstein, F., Schäfer, K., Prossinger, H., Seidler, H., Fieder, M., Stringer, C., Weber, G. W., Arsuaga, J.-L., Slice, D. E., Rohlf, F. J., Recheis, W., Mariam, A. J. and Marcus, L. F. (1999) 'Comparing frontal cranial profiles in archaic and modern *Homo* by morphometric analysis', *The Anatomical Record*, 257(6), pp. 217–224. doi: https://doi.org/10.1002/(SICI)1097-0185(19991215)257:6<217::AID-AR7>3.0.CO;2-W.

Bremekamp, C. E. B. (1940) 'On *Urophyllum* Wall., and its nearest allies', *Recueil des Travaux Botaniques Néerlandais*, pp. 171–197.

Bremer, B. and Eriksson, T. (2009) 'Time tree of Rubiaceae: phylogeny and dating the family, subfamilies, and tribes', *International Journal of Plant Sciences*, 170(6), pp. 766–793. doi: 10.1086/599077.

Bremer, B. and Manen, J.-F. (2000) 'Phylogeny and classification of the subfamily Rubioideae (Rubiaceae)', *Plant Systematics and Evolution*, 225(1), pp. 43–72. doi: 10.1007/BF00985458.

Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I. and Batzoglou, S. (2003) 'Glocal alignment: finding rearrangements during alignment', *Bioinformatics*, 19(Suppl 1), pp. i54–i62. doi: 10.1093/bioinformatics/btg1005.

Cámara-Leret, R. *et al.* (2020) 'New Guinea has the world's richest island flora', *Nature*, 584(7822), pp. 579–583. doi: 10.1038/s41586-020-2549-5.

Candolle, A. P. de (1830) *Prodromus systematis naturalis regni vegetabilis, sive, Enumeratio contracta ordinum generum specierumque plantarum huc usque cognitarium, juxta methodi naturalis, normas digesta*. Parisii: Sumptibus Sociorum Treuttel et Wurtz. Available at: https://www.biodiversitylibrary.org/item/7153.

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R. and De Rosario, H. (2020) *pwr: basic functions for power analysis*. Available at: https://cran.r-project.org/package=pwr (Accessed: 11 December 2020).

Chan, P. P., Lin, B. Y., Mak, A. J. and Lowe, T. M. (2019) 'tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes', *bioRxiv*, p. 614032. doi: 10.1101/614032.

Chen, Q., Wu, X. and Zhang, D. (2019) 'Phylogenetic analysis of *Fritillaria cirrhosa* D.Don and its closely related species based on complete chloroplast genomes', *PeerJ*, 7, p. e7480. doi: 10.7717/peerj.7480.

Choi, K. S., Son, Og. and Park, S. (2015) 'The chloroplast genome of *Elaeagnus macrophylla* and *trnH* duplication event in Elaeagnaceae', *PLOS ONE*. Edited by Z.-H. Chen, 10(9), p. e0138727. doi: 10.1371/journal.pone.0138727.

Christodoulou, M. D. (2015) *Quantification of fruit shape in apple: development of methodologies and assessment of their potential use in cultivar identification*. University of Reading.

Christodoulou, M. D., Battey, N. H. and Culham, A. (2018) 'Can you make morphometrics work when you know the right answer? Pick and mix approaches for apple identification', *PLOS ONE*, 13(10), p. e0205357. Available at: https://doi.org/10.1371/journal.pone.0205357.

Christodoulou, M. D., Clark, J. Y. and Culham, A. (2020) 'The Cinderella discipline: morphometrics and their use in botanical classification', *Botanical Journal of the Linnean Society*, 194(4), pp. 385–396. doi: 10.1093/botlinnean/boaa055.

Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L. and Jansen, R. K. (2006) 'The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants', *Molecular Biology and Evolution*, 23(11), pp. 2175–2190. doi: 10.1093/molbev/msl089.

Clarke, J. E. (2000a) *Biodiversity and Protected Areas - Cambodia*. Available at: http://www.mekonginfo.org/document/0003675-environment-biodiversity-and-protected-areas-cambodia.

Clarke, J. E. (2000b) *Biodiversity and Protected Areas - Laos P.D.R.* Available at: http://www.mekonginfo.org/document/0002547-environment-biodiversity-and-protected-areas-lao-p-d-r.

Clarke, J. E. (2000c) *Biodiversity and Protected Areas - Viet Nam*. Available at: http://www.mekonginfo.org/document/0002519-environment-biodiversity-and-protected-areas-vietnam.

Corlett, R. T. (2019) 'Chapter 7: threats to terrestrial biodiversity', in *The Ecology of Tropical East Asia*. Oxford University Press. doi: 10.1093/oso/9780198817017.001.0001.

Da Costa, A. F., Rodrigues, P. J. F. P. and Wanderley, M. D. G. L. (2009) 'Morphometric analysis and taxonomic revision of the *Vriesea paraibica* complex (Bromeliaceae)', *Botanical Journal of the Linnean Society*, 159(1), pp. 163–181. doi: 10.1111/j.1095-8339.2008.00919.x.

Costion, C., Ford, A., Cross, H., Crayn, D., Harrington, M. and Lowe, A. (2011) 'Plant DNA barcodes can accurately estimate species richness in poorly known floras', *PLOS ONE*. Edited by J. A. Gilbert, 6(11), p. e26841. doi: 10.1371/journal.pone.0026841.

Cowan, J. M. (1954) 'Some information on the Menzies and Jack collections in the Herbarium, Royal Botanic Garden, Edinburgh', *Notes From The Royal Botanic Garden Edinburgh*, 21, pp. 219–227.

Cox, P. G., Morris, P. J. R., Hennekam, J. J. and Kitchener, A. C. (2020) 'Morphological and functional variation between isolated populations of British red squirrels (*Sciurus vulgaris*)', *Journal of Zoology*, n/a(n/a). doi: https://doi.org/10.1111/jzo.12829.

Craib, W. G. (1931) 'Contributions to the Flora of Siam. Additamentum XXXII', *Bulletin of Miscellaneous Information (Royal Botanic Gardens, Kew)*, 1931(9), pp. 441–448. Available at: http://www.jstor.org/stable/4102478.

Craib, W. G. (1932) *Florae Siamensis Enumeratio Vol. 2*. Bangkok: W.H. Mundie, The Bangkok Times Press.

Cristians, S., Bye, R. and Nieto-Sotelo, J. (2018) 'Molecular markers associated with chemical analysis: a powerful tool for quality control assessment of copalchi medicinal plant complex', *Frontiers in Pharmacology*, 9, p. 666. doi: 10.3389/fphar.2018.00666.

Darriba, D., Taboada, G. L., Doallo, R. and Posada, D. (2012) 'jModelTest 2: more models, new heuristics and parallel computing', *Nature Methods*, p. 772. doi: 10.1038/nmeth.2109.

Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E. and Arrigo, N. (2012) 'Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers', *Evolutionary Applications*, 5(4), pp. 317–329. doi: https://doi.org/10.1111/j.1752-4571.2011.00223.x.

Deng, H. (2013) 'Guided random forest in the RRF package'. arXiv:1306.0237. Available at: https://cran.r-project.org/package=RRF (Accessed: 11 December 2020).

Dierckxsens, N., Mardulyn, P. and Smits, G. (2016) 'NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data', *Nucleic Acids Research*, 45(4). doi: 10.1093/nar/gkw955.

Dodsworth, S. (2015) 'Genome skimming for next-generation biodiversity analysis', *Trends in Plant Science*, 20(9), pp. 525–527. doi: https://doi.org/10.1016/j.tplants.2015.06.012.

Downie, S. R. and Jansen, R. K. (2015) 'A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions', *Systematic Botany*, 40(1), pp. 336–351. doi: 10.1600/036364415X686620.

Doyle, J. J. and Doyle, J. L. (1987) 'A rapid DNA isolation procedure from small quantities of fresh leaf tissue', *Phytochemical Bulletin*, 19(1), pp. 11–15.

Dugas, D. V, Hernandez, D., Koenen, E. J. M., Schwarz, E., Straub, S., Hughes, C. E., Jansen, R. K., Nageswara-Rao, M., Staats, M., Trujillo, J. T., Hajrah, N. H., Alharbi, N. S., Al-Malki, A. L., Sabir, J. S. M. and Bailey, C. D. (2015) 'Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*', *Scientific Reports*, 5(1), p. 16958. doi: 10.1038/srep16958.

Ebert, D. and Peakall, R. (2009) 'Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species', *Molecular Ecology Resources*, 9(3), pp. 673–690. doi: 10.1111/j.1755-0998.2008.02319.x.

Fehrer, J., Gemeinholzer, B., Chrtek, J. and Bräutigam, S. (2007) 'Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in Pilosella hawkweeds (Hieracium, Cichorieae, Asteraceae)', *Molecular Phylogenetics and Evolution*, 42(2), pp. 347–361. doi: https://doi.org/10.1016/j.ympev.2006.07.004.

Fogel, G. B. (2008) 'Computational intelligence approaches for pattern discovery in biological systems', *Briefings in Bioinformatics*, 9(4), pp. 307–316. doi: 10.1093/bib/bbn021.

Folk, R. A., Mandel, J. R. and Freudenstein, J. V (2017) 'Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms', *Systematic Biology*, 66(3), pp. 320–337. doi: 10.1093/sysbio/syw083.

Frajman, B., Záveská, E., Gamisch, A., Moser, T. and Schönswetter, P. (2019) 'Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana s.l.* (Euphorbiaceae)', *Molecular Phylogenetics and Evolution*, 134, pp. 238–252. doi: 10.1016/j.ympev.2018.10.046.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004) 'VISTA: computational tools for comparative genomics', *Nucleic Acids Research*, 32(suppl_2), pp. W273–W279. doi: 10.1093/nar/gkh458.

French, P. A., Brown, G. K. and Bayly, M. J. (2016) 'Incongruent patterns of nuclear and chloroplast variation in Correa (Rutaceae): introgression and biogeography in South-Eastern Australia', *Plant Systematics and Evolution*, 302(4), pp. 447–468. doi: 10.1007/s00606-016-1277-7.

Gage, A. T. and Burkill, I. H. (1916) 'William Jack's letters to Nathaniel Wallich, 1819-1821', *Journal of the Straits Branch of the Royal Asiatic Society.*, 73, pp. 147–268. Available at: https://www.biodiversitylibrary.org/part/184560.

GenBank (2020) *GenBank*, *Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1982]*. Available at: https://www.ncbi.nlm.nih.gov/nucleotide/ (Accessed: 18 June 2020).

Goulding, S. E., Wolfe, K. H., Olmstead, R. G. and Morden, C. W. (1996) 'Ebb and flow of the chloroplast inverted repeat', *Molecular and General Genetics MGG*, 252(1), pp. 195–206. doi: 10.1007/BF02173220.

Govaerts, R., Ruhsam, M., Andersson, L., Robbrecht, E., Bridson, D., Davis, A., Schanzer, I. and Sonk, B. (2020) *World checklist of Rubiaceae*, *Royal Botanic Gardens, Kew (Published on the Internet)*. Available at: http://wcsp.science.kew.org/ (Accessed: 9 June 2020).

Graham, B. F. and Bormann, F. H. (1966) 'Natural root grafts', *The Botanical Review*, 32(3), pp. 255–292. doi: 10.1007/BF02858662.

Green, B. R. (2011) 'Chloroplast genomes of photosynthetic eukaryotes', *The Plant Journal*, 66(1), pp. 34–44. doi: 10.1111/j.1365-313X.2011.04541.x.

Greiner, S., Lehwark, P. and Bock, R. (2019) 'OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes', *Nucleic Acids Research*, 47(W1), pp. W59–W64. doi: 10.1093/nar/gkz238.

Griffith, W. (1844) *Calcutta journal of natural history, and miscellany of the arts and sciences in India*. W. Ridsdale: Bishop's College Press. Available at: https://www.biodiversitylibrary.org/item/176237.

Grismer, L., Grismer, J., Wood Jr, P., Ngo, V. T., Thy, N. and Onn, C. (2011) 'Herpetology on the fringes of the sunda shelf: a discussion of discovery, taxonomy, and biogeography', *Bonner Zoologische Monographien*, 57, pp. 57–97.

Gruzdev, E. V, Kadnikov, V. V, Beletsky, A. V, Mardanov, A. V and Ravin, N. V (2019) 'Extensive plastome reduction and loss of photosynthesis genes in *Diphelypaea coccinea*, a holoparasitic plant of the family Orobanchaceae', *PeerJ*, 7, pp. e7830–e7830. doi: 10.7717/peerj.7830.

Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J. M., Patti, C., Patti, B., Mazzola, S., Jiménez, S., Duque, V. and Salmerón, F. (2010) 'IPez: an expert system for the taxonomic identification of fishes based on machine learning techniques', *Fisheries Research*, 102(3), pp. 240–247. doi: https://doi.org/10.1016/j.fishres.2009.12.003.

Haberle, R. C., Fourcade, H. M., Boore, J. L. and Jansen, R. K. (2008) 'Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes', *Journal of Molecular Evolution*, 66(4), pp. 350–361. doi: 10.1007/s00239-008-9086-4.

Hammer, Ø., Harper, D. A. T. and Ryan, P. D. (2001) 'PAST: paleontological statistics software package for education and data analysis', *Palaeontologia Electronica*, 4(1).

Hansen, K. T., Elven, R. and Brochmann, C. (2000) 'Molecules and morphology in concert: tests of some hypotheses in Arctic *Potentilla* (Rosaceae)', *American Journal of Botany*, 87(10), pp. 1466–1479. doi: https://doi.org/10.2307/2656873.

Harris, S. A. and Ingram, R. (1991) 'Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission', *Taxon*, 40(3), pp. 393–412. doi: 10.2307/1223218.

Hastie, T., Tibshirani, R., Leisch, F., Hornik, K., Ripley, B. D. and Narasimhan, B. (2020) *mda: mixture and flexible discriminant analysis*. Available at: https://cran.r-project.org/package=mda (Accessed: 11 December 2020).

Hedrick, B. P. and Dumont, E. R. (2018) 'Putting the leaf-nosed bats in context: a geometric morphometric analysis of three of the largest families of bats', *Journal of Mammalogy*, 99(5), pp. 1042–1054. doi: 10.1093/jmammal/gyy101.

Henderson, A. (2005) 'The methods of herbarium taxonomy', *Systematic Botany*, 30(2), pp. 456–459. doi: 10.1600/0363644054223701.

Henderson, A. (2006) 'Traditional morphometrics in plant systematics and its role in palm systematics', *Botanical Journal of the Linnean Society*, 151(1), pp. 103–111. doi: https://doi.org/10.1111/j.1095-8339.2006.00526.x.

Hendre, P. S. and Aggarwal, R. K. (2014) 'Development of genic and genomic SSR markers of robusta coffee (*Coffea canephora* Pierre ex A.Froehner)', *PLOS ONE*, 9(12), p. e113661. Available at: https://doi.org/10.1371/journal.pone.0113661.

Hickey, L. J. (1979) 'A revised classification of the architecture of dicotyledonous leaves', in Metcalfe, C. R. and Chalk, L. (eds) *Anatomy of the dicotyledons*. Ed.2 Vol.1. Oxford: Clarendon Press, pp. 25–39.

Hooker, J. D. (1880) *The flora of British India*. London: L. Reeve & Co. Available at: https://www.biodiversitylibrary.org/page/18705441.

Hooker, W. J. (1849) *Niger Flora*. London: Hippolyte Bailliere.

Hornik, K., Buchta, C. and Zeileis, A. (2009) 'Open-source machine learning: R meets Weka', *Computational Statistics*, 24(2), pp. 225–232. doi: 10.1007/s00180-008-0119-7.

Hothorn, T., Hornik, K. and Zeileis, A. (2008) 'Party: a laboratory for recursive part(y)itioning', *R package version 0.9-0, URL http://CRAN. R-project. org*.

Hua, Z. (2002) 'A revision of the genus *Lasianthus* (Rubiaceae) from China', *Systematics and Geography of Plants*, 72, pp. 63–110. doi: 10.2307/3668764.

Hughes, A. C. (2017) 'Understanding the drivers of Southeast Asian biodiversity loss', *Ecosphere*, 8(1), p. e01624. doi: https://doi.org/10.1002/ecs2.1624.

Ibañez, A. L., Cowx, I. G. and O'Higgins, P. (2007) 'Geometric morphometric analysis of fish scales for identifying genera, species, and local populations within the Mugilidae', *Canadian Journal of Fisheries and Aquatic Sciences*, 64(8), pp. 1091–1100. doi: 10.1139/f07-075.

IUCN (2012a) *Guidelines for application of IUCN Red List criteria at regional and national levels: version 4.0*. Gland, Switzerland and Cambridge, UK: IUCN.

IUCN (2012b) *IUCN Red List categories and criteria, version 3.1, Second edition*, *Gland and Cambridge*. Available at: https://www.iucnredlist.org/resources/categories-and-criteria (Accessed: 15 February 2021).

IUCN (2019) *Guidelines for Using the IUCN Red List Categories and Criteria. Version 14*. Prepared by the Standards and Petitions Committee. Available at: http://www.iucnredlist.org/documents/RedListGuidelines.pdf (Accessed: 9 March 2021).

Janzen, D. H. (1979) 'How to be a fig', *Annual Review of Ecology and Systematics*, 10(1), pp. 13–51. doi: 10.1146/annurev.es.10.110179.000305.

Ji, Y., Yang, L., Chase, M. W., Liu, C., Yang, Z., Yang, J., Yang, J.-B. and Yi, T.-S. (2019) 'Plastome phylogenomics, biogeography, and clade diversification of *Paris* (Melanthiaceae)', *BMC Plant Biology*, 19(1), p. 543. doi: 10.1186/s12870-019-2147-6.

Jiang, Y., Yang, Y., Lu, Z., Wan, D. and Ren, G. (2019) 'Interspecific delimitation and relationships among four *Ostrya* species based on plastomes', *BMC Genetics*, 20(1), p. 33. doi: 10.1186/s12863-019-0733-0.

Jiao, L., Yu, M., Wiedenhoeft, A. C., He, T., Li, J., Liu, B., Jiang, X. and Yin, Y. (2018) 'DNA barcode authentication and library development for the wood of six commercial *Pterocarpus* species: the critical role of xylarium specimens', *Scientific Reports*, 8(1), p. 1945. doi: 10.1038/s41598-018-20381-6.

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., DePamphilis, C. W., Yi, T.-S. and Li, D.-Z. (2018) 'GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes', *bioRxiv*, p. 256479. doi: https://doi.org/10.1101/256479.

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J. H., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K., Baker, W. J. and Wickett, N. J. (2019) 'A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering', *Systematic Biology*, 68(4), pp. 594–606. doi: 10.1093/sysbio/syy086.

Jolliffe, I. T. and Cadima, J. (2016) 'Principal component analysis: a review and recent developments', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), p. 20150202. doi: 10.1098/rsta.2015.0202.

Kamo, T., Kusumoto, Y., Tokuoka, Y., Okubo, S., Hayakawa, H., Yoshiyama, M., Kimura, K. and Konuma, A. (2018) 'A DNA barcoding method for identifying and quantifying the composition of pollen species collected by European honeybees, *Apis mellifera* (Hymenoptera: Apidae)', *Applied Entomology and Zoology*, 53(3), pp. 353–361. doi: 10.1007/s13355-018-0565-9.

Karbstein, K., Tomasello, S., Hodac, L., Dunkel, F. G., Daubert, M. and Hörandl, E. (2020) 'Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (Ranunculaceae)', *Taxon*, 69, pp. 1191–1220. doi: 10.1101/2020.01.07.896902.

Katoh, K. and Standley, D. M. (2013) 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', *Molecular Biology and Evolution*, 30(4), pp. 772–780. doi: 10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012) 'Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28(12), pp. 1647–1649. doi: 10.1093/bioinformatics/bts199.

Kerr, A. F. G. (1933) 'Miscellaneous Notes', *Bulletin of Miscellaneous Information (Royal Botanic Gardens, Kew)*, 1933(8), pp. 408–416. Available at: http://www.jstor.org/stable/4113435.

King, G. and Gamble, J. S. (1904) *The journal of the Asiatic Society of Bengal Part II*. Calcutta: Baptist Mission Press. Available at: https://www.biodiversitylibrary.org/item/110113.

Klingenberg, C. P. (2008) 'Novelty and "Homology-free" morphometrics: what's in a name?', *Evolutionary Biology*, 35(3), pp. 186–190. doi: 10.1007/s11692-008-9029-4.

Klingenberg, C. P. (2011) 'MorphoJ: an integrated software package for geometric morphometrics', *Molecular Ecology Resources*, 11(2), pp. 353–357. doi: https://doi.org/10.1111/j.1755-0998.2010.02924.x.

Klingenberg, C. P., Duttke, S., Whelan, S. and Kim, M. (2012) 'Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves', *Journal of Evolutionary Biology*, 25(1), pp. 115–129. doi: https://doi.org/10.1111/j.1420-9101.2011.02410.x.

Korthals, P. W. (1851) 'Overzigt der Rubiaceen van de Nederlandsch-Oostindische Kolonien', in De Veriese, W. H., Dozy, F., and Molkenboer, J. H. (eds) *Nederlandsch kruidkundig archief*. Leyden: Jacs. Hazenberg Corns. Zoon, pp. 145–269. Available at: https://www.biodiversitylibrary.org/item/107156.

Kotsiantis, S. B. (2007) 'Supervised machine learning: a review of classification techniques', in Maglogiannis, I., Karpouzis, K., Wallace, B. A., and Soldatos, J. (eds) *Emerging Artificial Intelligence Applications in Computer Engineering*. IOS Press.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V and Fotiadis, D. I. (2015) 'Machine learning applications in cancer prognosis and prediction', *Computational and Structural Biotechnology Journal*, 13, pp. 8–17. doi: https://doi.org/10.1016/j.csbj.2014.11.005.

Krapp, F. (2013) *Phylogenie und evolution der gattung Dyckia (Bromeliaceae)*. University of Kassel, Germany. Available at: https://kobra.bibliothek.uni-kassel.de/bitstream/urn:nbn:de:hebis:34-2013011142369/5/DissertationFlorianKrapp.pdf.

Kuhn, M. (2008) 'Building predictive models in R using the caret package', *Journal of Statistical Software; Vol 1, Issue 5 (2008)*. doi: 10.18637/jss.v028.i05.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) 'MEGA X: molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kursa, M. B. (2012) 'Random ferns method implementation for the general-purpose machine learning', *CoRR*, abs/1202.1. Available at: http://arxiv.org/abs/1202.1121.

Kurtz, S., Choudhuri, J. V, Ohlebusch, E., Schleiermacher, C., Stoye, J. and Giegerich, R. (2001) 'REPuter: the manifold applications of repeat analysis on a genomic scale', *Nucleic Acids Research*, 29(22), pp. 4633–4642. doi: 10.1093/nar/29.22.4633.

Laslett, D. and Canback, B. (2004) 'ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences', *Nucleic Acids Research*, 32(1), pp. 11–16. doi: 10.1093/nar/gkh152.

Leaf-Architecture-Working-Group (1999) *Manual of leaf architecture–morphological description and categorization of dicotyledonous and net-veined monocotyledonous angiosperms*. Leaf Architecture Working Group.

LeDell, E. *et al.* (2020) *h2o: R interface for the 'H2O' scalable machine learning platform*. Available at: https://cran.r-project.org/package=h2o (Accessed: 11 December 2020).

Lev-Yadun, S. and Sprugel, D. (2011) 'Why should trees have natural root grafts?', *Tree Physiology*, 31(6), pp. 575–578. doi: 10.1093/treephys/tpr061.

Li, C.-J., Wang, R.-N. and Li, D.-Z. (2020) 'Comparative analysis of plastid genomes within the Campanulaceae and phylogenetic implications', *PLOS ONE*, 15(5), p. e0233167. Available at: https://doi.org/10.1371/journal.pone.0233167.

Liaw, A. and Wiener, M. (2002) 'Classification and regression by randomForest', *R News*, 2(3), pp. 18–22. Available at: https://cran.r-project.org/doc/Rnews/ (Accessed: 11 December 2020).

Lim, V.-C., Ramli, R., Bhassu, S. and Wilson, J.-J. (2018) 'Pollination implications of the diverse diet of tropical nectar-feeding bats roosting in an urban cave', *PeerJ*, 6, p. e4572. doi: 10.7717/peerj.4572.

Liu, Y., Li, Y., Song, J., Zhang, R., Yan, Y., Wang, Y. and Du, F. K. (2018) 'Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae)', *Annals of Forest Science*, 75(4), p. 90. doi: 10.1007/s13595-018-0770-2.

Ly, S. N., Garavito, A., De Block, P., Asselman, P., Guyeux, C., Charr, J.-C., Janssens, S., Mouly, A., Hamon, P. and Guyot, R. (2020) 'Chloroplast genomes of Rubiaceae: comparative genomics and molecular phylogeny in subfamily Ixoroideae', *PLOS ONE*, 15(4), pp. 1–21. doi: 10.1371/journal.pone.0232295.

Mace, G. M. (2004) 'The role of taxonomy in species conservation', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. Edited by H. C. J. Godfray and S. Knapp, 359(1444), pp. 711–719. doi: 10.1098/rstb.2003.1454.

Macfarlane, T. D., Sokoloff, D. D. and Remizowa, M. V. (2017) 'Filling a morphological gap between Australasian and Eurasian/African members of Althenia (Potamogetonaceae, Alismatales): A. hearnii sp. nov. from SW Western Australia', *Phytotaxa*, 317(1), p. 53. doi: 10.11646/phytotaxa.317.1.5.

MacLeod, N. (1999) 'Generalizing and extending the eigenshape method of shape space visualization and analysis', *Paleobiology*, 25(1), pp. 107–138. Available at: http://www.jstor.org/stable/2665995.

Maddison, W. P. and Maddison, D. R. (2019) 'Mesquite: a modular system for evolutionary analysis'. Available at: http://www.mesquiteproject.org (Accessed: 29 March 2021).

Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y. and Fay, M. F. (2007) 'Towards a phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences', *Annals of Botany*, 100(7), pp. 1565–1583. doi: 10.1093/aob/mcm257.

McKain, M. R. and Wilson, M. (2017) *Fast-Plast: rapid de novo assembly and finishing for whole chloroplast genomes*. doi: https://doi.org/10.5281/zenodo.973887.

Meade, C. and Parnell, J. (2003) 'Multivariate analysis of leaf shape patterns in Asian species of the *Uvaria* group (Annonaceae)', *Botanical Journal of the Linnean Society*, 143(3), pp. 231–242. doi: 10.1046/j.1095-8339.2003.00223.x.

Mediavilla, S., Martín, I., Babiano, J. and Escudero, A. (2019) 'Foliar plasticity related to gradients of heat and drought stress across crown orientations in three Mediterranean *Quercus* species', *PLOS ONE*, 14(10), p. e0224462. Available at: https://doi.org/10.1371/journal.pone.0224462.

Menini Neto, L., Van den Berg, C. and Forzza, R. C. (2019) 'Linear and geometric morphometrics as tools to resolve species circumscription in the *Pseudolaelia vellozicola* complex (Orchidaceae, Laeliinae)', *Plant Ecology and Evolution*, 152(1), pp. 53–67. doi: 10.5091/plecevo.2019.1531.

Merrill, E. (1952) 'William Jack's genera and species of Malaysian plants', *Journal of the Arnold Arboretum*, 33, pp. 199–251. Available at: https://www.biodiversitylibrary.org/part/29334.

Meseguer, A. S., Sanmartín, I., Marcussen, T. and Pfeil, B. E. (2014) 'Utility of low-copy nuclear markers in phylogenetic reconstruction of *Hypericum* L. (Hypericaceae)', *Plant Systematics and Evolution*, 300(6), pp. 1503–1514. Available at: http://www.jstor.org/stable/43498428.

Metcalfe, D. J., Grubb, P. J. and Turner, I. M. (1998) 'The ecology of very small-seeded shade-tolerant trees and shrubs in lowland rain forest in Singapore', *Plant Ecology*, 134(2), pp. 131–149. doi: 10.1023/A:1009789922595.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F. (2019) 'e1071: misc functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien'. Available at: https://cran.r-project.org/package=e1071.

Middleton, D. J. *et al.* (2019) 'Progress on Southeast Asia's Flora projects', *Gardens' Bulletin Singapore*, 71(2), pp. 267–319. doi: 10.26492/gbs71(2).2019-02.

Mishler, B. D. and Theriot, E. C. (2000) 'The phylogenetic species concept (*sensu* Mishler & Theriot): monophyly, apomorphy and phylogenetic species concepts', in Wheeler, Q. and Meier, R. (eds) *Species Concepts and Phylogenetic Theory*. New York: Columbia University Press, pp. 44–54.

Nagahama, A. *et al.* (2019) *A picture guide for the Flora of Bidoup-Nui Ba National Park I: Mt. Langbian*. Fukuoka: Center for Asian Conservation Ecology, Kyushu University. Available at: https://ffish.asia/pdfs/PictureGuide_Mt_Langbian.pdf.

Nagahama, N., Anton, A. M. and Norrmann, G. A. (2014) 'Taxon delimitation in the *Andropogon* lateralis complex (Poaceae) in Southern South America based on morphometrical analyses', *Systematic Botany*, 39(3), pp. 804–813. doi: 10.1600/036364414X681509.

Nevill, P. G., Howell, K. A., Cross, A. T., Williams, A. V, Zhong, X., Tonti-Filippini, J., Boykin, L. M., Dixon, K. W. and Small, I. (2019) 'Plastome-wide rearrangements and gene losses in carnivorous Droseraceae', *Genome Biology and Evolution*, 11(2), pp. 472–485. doi: 10.1093/gbe/evz005.

Nevill, P. G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., van Leeuwen, S., Boykin, L. M. and Small, I. (2020) 'Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics', *Plant Methods*, 16(1), p. 1. doi: 10.1186/s13007-019-0534-5.

Nobis, M., Klichowska, E., Nowak, A., Gudkova, P. D. and Rola, K. (2016) 'Multivariate morphometric analysis of the *Stipa turkestanica* group (Poaceae: *Stipa* sect. *Stipa*)', *Plant Systematics and Evolution*, 302(2), pp. 137–153. doi: 10.1007/s00606-015-1243-9.

Noltie, H. J. (2005) *The botany of Robert Wight*. Ruggell, Lichenstein: Gantner Verlag.

O'hanlon, J. C., Li, D. and Norma-Rashid, Y. (2013) 'Coloration and morphology of the orchid mantis *Hymenopus coronatus* (Mantodea: Hymenopodidae)', *Journal of Orthoptera Research*, 22(1), pp. 35–44. doi: 10.1665/034.022.0106.

Obico, J. J. A. and Alejandro, G. J. D. (2012) 'Molecular phylogeny of the Philippine endemic genus *Antherostele* Bremek. (Rubiaceae) inferred from ITS data (nrDNA) and its conservation status', *American Journal of Plant Sciences*, 03(10), pp. 1399–1407. doi: 10.4236/ajps.2012.310169.

Olden, J. D., Lawler, J. J. and Poff, N. L. (2008) 'Machine learning methods without tears: a primer for ecologists', *The Quarterly Review of Biology*, 83(2), pp. 171–193. doi: 10.1086/587826.

Van de Paer, C., Bouchez, O. and Besnard, G. (2018) 'Prospects on the evolutionary mitogenomics of plants: a case study on the olive family (Oleaceae)', *Molecular Ecology Resources*, 18(3), pp. 407–423. doi: https://doi.org/10.1111/1755-0998.12742.

Palmer, J. D. (1983) 'Chloroplast DNA exists in two orientations', *Nature*, 301(5895), pp. 92–93. doi: 10.1038/301092a0.

Parrish, T. L., Koelewijn, H. P., van Dijk, P. J. and Kruijt, M. (2003) 'Genetic evidence for natural hybridization between species of dioecious *Ficus* on island populations', *Biotropica*, 35(3), pp. 333–343. Available at: http://www.jstor.org/stable/30043049.

Peng, H. S. and Clarke, C. (2015) 'Prey capture patterns in Nepenthes species and natural hybrids–are the pitchers of hybrids as effective at trapping prey as those of their parents?', *Carnivorous Plant Newsletter*, 44(2), pp. 62–79. Available at: https://cpn.carnivorousplants.org/articles/CPNv44n2p62_79.pdf.

Pham, G. T. (2019) *Soil properties mapping and land evaluation for potential agricultural land use types in A Luoi district, Thua Thien Hue province, Central Vietnam*. Georg-August University of Göttingen. Available at: https://d-nb.info/118990473X/34.

Pham, H. H. (2000) *Cây cỏ Việt Nam: an illustrated flora of Vietnam, Vol. 3*. Ho Chi Minh City: Youth Publishing House.

Philippe, H., Delsuc, F., Brinkmann, H. and Lartillot, N. (2005) 'Phylogenomics', *Annual Review of Ecology, Evolution, and Systematics*, 36(1), pp. 541–562. doi: 10.1146/annurev.ecolsys.35.112202.130205.

Phuong, V. T., Inoguchi, A., Birigazzi, L., Henry, M. and Sola, G. (2012) 'Introduction and background of the study, Vietnam', in Inoguchi, A., Henry, M., Birigazzi, L., and Sola, G. (eds) *Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam (Part A)*. Hanoi, Viet Nam: UN-REDD Programme.

Di Pietro, R., Conte, A. L., Di Marzio, P., Gianguzzi, L., Spampinato, G., Caldarella, O. and Fortini, P. (2020) 'A multivariate morphometric analysis of diagnostic traits in southern Italy and Sicily pubescent oaks', *Folia Geobotanica*, 55(3), pp. 163–183. doi: 10.1007/s12224-020-09378-0.

Pitard, J. (1923) 'RUBIACÉES', in Lecomte, H., Humbert, H., and Gagnepain, F. (eds) *Flore générale de l'Indo-Chine*. Volume 3. Paris: Masson et Cie, 1912-1936, pp. 20–442.

Priyadi, H., Takao, G., Rahmawati, I., Supriyanto, B., Ikbal Nursal, W. and Rahman, I. (2010) *Five hundred plant species in Gunung Halimun Salak National Park, West Java: a checklist including Sundanese names, distribution and use*. Bogor, Indonesia: CIFOR. doi: 10.13140/2.1.2823.5845.

Provan, J., Powell, W. and Hollingsworth, P. M. (2001) 'Chloroplast microsatellites: new tools for studies in plant ecology and evolution', *Trends in Ecology & Evolution*, 16(3), pp. 142–147. doi: https://doi.org/10.1016/S0169-5347(00)02097-8.

Puff, C., Chayamarit, K. and Chamchumroon, V. (2005) *Rubiaceae of Thailand: a pictorial guide to indigenous and cultivated genera*. Bangkok: The Forest Herbarium, National Park, Wildlife and Plant Conservation Department.

Qi, W., Lin, F., Liu, Y., Huang, B., Cheng, J., Zhang, W. and Zhao, H. (2016) 'High-throughput development of simple sequence repeat markers for genetic diversity research in *Crambe abyssinica*', *BMC Plant Biology*, 16(1), p. 139. doi: 10.1186/s12870-016-0828-y.

R Core Team (2019) 'R Core Team', *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Available at: https://www.r-project.org/.

Rabah, S. O., Shrestha, B., Hajrah, N. H., Sabir, M. J., Alharby, H. F., Sabir, M. J., Alhebshi, A. M., Sabir, J. S. M., Gilbert, L. E., Ruhlman, T. A. and Jansen, R. K. (2019) '*Passiflora* plastome sequencing reveals widespread genomic rearrangements', *Journal of Systematics and Evolution*, 57(1), pp. 1–14. doi: https://doi.org/10.1111/jse.12425.

Raman, G., Park, S., Lee, E. M. and Park, S. (2019) 'Evidence of mitochondrial DNA in the chloroplast genome of *Convallaria keiskei* and its subsequent evolution in the Asparagales', *Scientific Reports*, 9(1), p. 5028. doi: 10.1038/s41598-019-41377-w.

Rambaut, A. (2018) *FigTree: tree figure drawing tool. Version 1.4.4*, *Institute of Evolution Biology, University of Edinburgh*. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (Accessed: 1 July 2020).

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. and Suchard, M. A. (2018) 'Posterior summarization in Bayesian phylogenetics using Tracer 1.7', *Systematic Biology*, 67(5), pp. 901–904. doi: 10.1093/sysbio/syy032.

Ramesar-Fortner, N. S., Dengler, N. G. and Aiken, S. G. (1995) 'Phenotypic plasticity in leaves of four species of arctic *Festuca* (Poaceae)', *Canadian Journal of Botany*, 73(11), pp. 1810–1823. doi: 10.1139/b95-192.

Reboud, X. and Zeyl, C. (1994) 'Organelle inheritance in plants', *Heredity*, 72(2), pp. 132–140. doi: 10.1038/hdy.1994.19.

Rejlová, L., Böhmová, A., Chumová, Z., Hořčicová, Š., Josefiová, J., Schmidt, P.-A., Trávníček, P., Urfus, T., Vít, P. and Chrtek, J. (2021) 'Disparity between morphology and genetics in *Urtica dioica* (Urticaceae)', *Botanical Journal of the Linnean Society*, 195(4), pp. 606–621. doi: 10.1093/botlinnean/boaa076.

Ridley, H. N. (1923) *The flora of the Malay Peninsula*. London,: L. Reeve & co., ltd.,. Available at: https://www.biodiversitylibrary.org/item/256567.

Ridley, H. N. (1932) 'The genus *Urophyllum* in Malaya', *Journal of Botany*. Edited by A. B. Rendle, pp. 191–197, 219–226.

Ripma, L. A., Simpson, M. G. and Hasenstab-Lehman, K. (2014) 'Geneious! simplified genome skimming methods for phylogenetic systematic studies: a case study in *Oreocarya* (Boraginaceae)', *Applications in Plant Sciences*, 2(12), p. 1400062. doi: 10.3732/apps.1400062.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C. and Müller, M. (2011) 'pROC: an open-source package for R and S+ to analyze and compare ROC curves', *BMC Bioinformatics*, 12(1), p. 77. doi: 10.1186/1471-2105-12-77.

Rohlf, F. J. (2017) *TPSDig2 v2.31*. Available at: http://www.sbmorphometrics.org (Accessed: 15 June 2019).

Rohlf, F. J. (2019) *TPSUtil v1.78*. Available at: http://www.sbmorphometrics.org (Accessed: 15 June 2019).

Ronquist, F., Huelsenbeck, J. P., Teslenko, M. and Nylander, J. A. A. (2020) *MrBayes version 3.2 manual: tutorials and model summaries*. Available at: https://github.com/NBISweden/MrBayes/blob/develop/doc/manual/Manual_MrBayes_v3.2.pdf (Accessed: 18 June 2020).

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012) 'MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space', *Systematic Biology*, 61(3), pp. 539–542. doi: 10.1093/sysbio/sys029.

Roxburgh, W. (1820) *Plants of the coast of Coromandel*. Vol. III. London: W. Bulmer and Co.

Roxburgh, W. (1824) '*Urophyllum* Wall.', in Carey, W. (ed.) *Flora Indica*. Volume 2. Serampore: The Mission Press, pp. 184–186.

Roy, Stanton and Eppley (1999) 'Effects of environmental stress on leaf hair density and consequences for selection', *Journal of Evolutionary Biology*, 12(6), pp. 1089–1103. doi: https://doi.org/10.1046/j.1420-9101.1999.00107.x.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. and Sánchez-Gracia, A. (2017) 'DnaSP 6: DNA sequence polymorphism analysis of large data sets', *Molecular Biology and Evolution*, 34(12), pp. 3299–3302. doi: 10.1093/molbev/msx248.

Ruhlman, T. A., Zhang, J., Blazier, J. C., Sabir, J. S. M. and Jansen, R. K. (2017) 'Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure', *American Journal of Botany*, 104(4), pp. 559–572. doi: https://doi.org/10.3732/ajb.1600453.

Santana, F., Costa, A., Truzzi, F., Silva, F., Leal, S., Francoy, T. and Saraiva, A. (2014) 'A reference process for automating bee species identification based on wing images and digital image processing', *Ecological Informatics*, 24. doi: 10.1016/j.ecoinf.2013.12.001.

Schmidt, J. (1902) 'Flora of Koh Chang. Contributions to the knowledge of the vegetation in the Gulf of Siam. Part VI: Rubiaceae', *Botanisk Tidsskrift*, 24, pp. 329–367.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*, 9(7), pp. 671–675. doi: 10.1038/nmeth.2089.

Shevade, V. S. and Loboda, T. V (2019) 'Oil palm plantations in Peninsular Malaysia: determinants and constraints on expansion', *PLOS ONE*, 14(2), p. e0210628. Available at: https://doi.org/10.1371/journal.pone.0210628.

Shipunov, A. B. and Bateman, R. M. (2005) 'Geometric morphometrics as a tool for understanding *Dactylorhiza* (Orchidaceae) diversity in European Russia', *Biological Journal of the Linnean Society*, 85(1), pp. 1–12. doi: https://doi.org/10.1111/j.1095-8312.2005.00468.x.

da Silva, F. L., Grassi Sella, M. L., Francoy, T. M. and Costa, A. H. R. (2015) 'Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images', *Computers and Electronics in Agriculture*, 114, pp. 68–77. doi: 10.1016/j.compag.2015.03.012.

Smedmark, J. E. E. and Bremer, B. (2011) 'Molecular systematics and incongruent gene trees of Urophylleae (Rubiaceae)', *Taxon*, 60(5), pp. 1397–1406.

Smedmark, J. E. E., Eriksson, T. and Bremer, B. (2010) 'Divergence time uncertainty and historical biogeography reconstruction–an example from Urophylleae (Rubiaceae)', *Journal of Biogeography*, 37(12), pp. 2260–2274. doi: 10.1111/j.1365-2699.2010.02366.x.

Smedmark, J. E. E., Rydin, C., Razafimandimbison, S. G., Khan, S. A., Liede-Schumann, S. and Bremer, B. (2008) 'A phylogeny of Urophylleae (Rubiaceae) based on *rps16* intron data', *Taxon*, 57(1), pp. 24–32. doi: 10.2307/25065945.

Smith, D. R. (2011) 'Extending the limited transfer window hypothesis to inter-organelle DNA migration', *Genome Biology and Evolution*, 3, pp. 743–748. doi: 10.1093/gbe/evr068.

Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical taxonomy: the principles and practice of numerical classification*. San Francisco: W. H. Freeman.

Stafleu, F. A. and Cowan, R. S. (1983) *Taxonomic literature, Volume IV: P-Sak*. Second Ed. Utrecht/Antwerpen; The Hague/Boston: Bohn, Scheltema & Holkema; dr. W. Junk b.v., Publishers.

Stafleu, F. A. and Mennega, E. A. (1993) *Taxonomic literature: a selective guide to botanical publications and collections with dates, commentaries and types*. 2d ed. Konigstein: Koeltz Scientific Books. Available at: https://www.biodiversitylibrary.org/item/103859.

Stafleu, F. A. and Mennega, E. A. (1997) *Taxonomic literature: a selective guide to botanical publications and collections with dates, commentaries and types*. 2d ed. Utrecht : Bohn, Scheltema & Holkema,. Available at: https://www.biodiversitylibrary.org/item/103860.

Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics*, 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Stamate, D., Smith, R., Tsygancov, R., Vorobev, R., Langham, J., Stahl, D. and Reeves, D. (2020) 'Applying deep learning to predicting dementia and mild cognitive impairment', *Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II*. Edited by I. Maglogiannis, L. Iliadis, and E. Pimenidis, 584, pp. 308–319. doi: 10.1007/978-3-030-49186-4_26.

Stegemann, S., Keuthe, M., Greiner, S. and Bock, R. (2012) 'Horizontal transfer of chloroplast genomes between plant species', *Proceedings of the National Academy of Sciences*, 109(7), p. 2434 LP-2438. doi: 10.1073/pnas.1114076109.

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. and Liston, A. (2012) 'Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics', *American journal of botany*, 99 2, pp. 349–364.

Tan, H. T. W., Chua, K. S. and Turner, I. M. (1995) 'Rubiaceae of Bukit Timah Nature Reserve', in Chin, S. C., Corlett, R. T., Wee, Y. C., and Gee, S. Y. (eds) *Garden's Bulletin Singapore Supplement No. 3*. National Parks Board Singapore, pp. 29–61.

Tan, P. N., Steinbach, M. and Kumar, V. (2014) *Introduction to data mining*. Pearson ne. Harlow, Essex, England: Pearson Education Limited.

Tarca, A. L., Carey, V. J., Chen, X., Romero, R. and Drăghici, S. (2007) 'Machine learning and its applications to biology', *PLOS Computational Biology*, 3(6), p. e116. Available at: https://doi.org/10.1371/journal.pcbi.0030116.

Taylor, C. M., Ehrendorfer, F., Lantz, H., Funston, A. M. and Puff, C. (2011) 'Rubiaceae', in Wu, Z. Y., Raven, P. H., and Hong, D. Y. (eds) *Flora of China Vol. 19 (Cucurbitaceae through Valerianaceae, with Annonaceae and Berberidaceae)*. Vol. 19. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis, pp. 57–368.

Tel-Zur, N., Abbo, S., Myslabodski, D. and Mizrahi, Y. (1999) 'Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae)', *Plant Molecular Biology Reporter*, 17(3), pp. 249–254. doi: 10.1023/A:1007656315275.

Thiers, B. (2020) *Index Herbariorum: a worldwide index of 3,100 herbaria and 12,000 associated staff where a total of 390 million botanical specimens are permanently housed*, *New York Botanical Garden's Virtual Herbarium*. Available at: http://sweetgum.nybg.org/ih/ (Accessed: 12 February 2020).

Thode, V. A. and Lohmann, L. G. (2019) 'Comparative chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignonieae, Bignoniaceae)', *Frontiers in Plant Science*, p. 796. Available at: https://www.frontiersin.org/article/10.3389/fpls.2019.00796.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R. and Greiner, S. (2017) 'GeSeq-versatile and accurate annotation of organelle genomes', *Nucleic Acids Research*, 45(W1), pp. W6–W11. doi: 10.1093/nar/gkx391.

Tosh, J., Davis, A. P., Dessein, S., Block, P. De, Huysmans, S., Fay, M. F., Smets, E. and Robbrecht, E. (2009) 'Phylogeny of *Tricalysia* (Rubiaceae) and its relationships with allied genera based on plastid DNA data: resurrection of the genus *Empogona*', *Annals of the Missouri Botanical Garden*, 96(1), pp. 194–213. doi: 10.3417/2006202.

Turland, N., Wiersema, J., Barrie, F., Greuter, W., Hawksworth, D., Herendeen, P., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T., McNeill, J., Monro, A., Prado, J., Price, M. and Smith, G. (eds) (2018) *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Glashütten: Koeltz Botanical Books (Regnum Vegetabile 159). doi: 10.12705/Code.2018.

UNEP-WCMC and IUCN (2020) *Protected planet: the world database on protected areas (WDPA)*. Cambridge, UK: UNEP-WCMC & IUCN. Available at: www.protectedplanet.net (Accessed: 9 March 2021).

Upadhyaya, M. K. and Furness, N. H. (1994) 'Influence of light intensity and water stress on leaf surface characteristics of *Cynoglossum officinale*, *Centaurea* spp., and *Tragopogon* spp.', *Canadian Journal of Botany*, 72(9), pp. 1379–1386. doi: 10.1139/b94-169.

Valencia-D, J., Murillo-A, J., Orozco, C. I., Parra-O, C. and Neubig, K. M. (2020) '-Complete plastid genome sequences of two species of the Neotropical genus *Brunellia* (Brunelliaceae)', *PeerJ*. Edited by T. Tatarinova, 8, p. e8392. doi: 10.7717/peerj.8392.

del Valle, J. C., Casimiro-Soriguer, I., Buide, M. L., Narbona, E. and Whittall, J. B. (2019) 'Whole plastome sequencing within *Silene* section *Psammophilae* reveals mainland hybridization and divergence with the Balearic Island populations', *Frontiers in Plant Science*, p. 1466. Available at: https://www.frontiersin.org/article/10.3389/fpls.2019.01466.

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M. and Zhao, S. (2019) 'Applications of machine learning in drug discovery and development', *Nature Reviews Drug Discovery*, 18(6), pp. 463–477. doi: 10.1038/s41573-019-0024-5.

Venables, W. N. and Ripley, B. D. (2002) *Modern applied statistics with S*. Fourth. New York: Springer. Available at: http://www.stats.ox.ac.uk/pub/MASS4/.

Verdcourt, B. (1958) 'Remarks on the classification of the Rubiaceae', *Bulletin du Jardin botanique de l'État a Bruxelles*, 28(3), pp. 209–290. doi: 10.2307/3667090.

Vigalondo, B., Fernández-Mazuecos, M., Vargas, P. and Sáez, L. (2015) 'Unmasking cryptic species: morphometric and phylogenetic analyses of the Ibero-North African *Linaria incarnata* complex', *Botanical Journal of the Linnean Society*, 177(3), pp. 395–417. doi: 10.1111/boj.12251.

Viscosi, V. and Cardini, A. (2011) 'Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners', *PLOS ONE*, 6(10), p. e25630. Available at: https://doi.org/10.1371/journal.pone.0025630.

Viscosi, V., Fortini, P., Slice, D. E., Loy, A. and Blasi, C. (2009) 'Geometric morphometric analyses of leaf variation in four oak species of the subgenus *Quercus* (Fagaceae)', *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, 143(3), pp. 575–587. doi: 10.1080/11263500902775277.

de Vriese, W. H., Dozy, F. and Molkenboer, J. H. (1851) *Nederlandsch kruidkundig archief*. Part II. Leyden: S. en J. Luchtmans.

Wang, R.-J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M. and Chaw, S.-M. (2008) 'Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots', *BMC Evolutionary Biology*, 8(1), p. 36. doi: 10.1186/1471-2148-8-36.

Webster, M. and Sheets, H. D. (2010) 'A practical introduction to landmark-based geometric morphometrics', *The Paleontological Society Papers*. 2017/07/21, 16, pp. 163–188. doi: DOI: 10.1017/S1089332600001868.

Wei, L., Li, Y.-F., Zhang, H. and Liao, W.-J. (2015) 'Variation in morphological traits in a recent hybrid zone between closely related *Quercus liaotungensis* and *Q. mongolica* (Fagaceae)', *Journal of Plant Ecology*, 8(2), pp. 224–229. doi: 10.1093/jpe/rtv023.

Van Welzen, P. C. *et al.* (2011) 'The current and future status of floristic provinces in Thailand', in Trisurat, Y., Shrestha, R. P., and Alkemade, R. (eds) *Land Use, Climate Change and Biodiversity Modeling: Perspectives and Applications*. Information Science Reference, pp. 219–247. doi: 10.4018/978-1-60960-619-0.ch011.

Wheeler, Q. D. (2004) 'Taxonomic triage and the poverty of phylogeny', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1444), pp. 571–583. doi: 10.1098/rstb.2003.1452.

Wicke, S., Schneeweiss, G. M., DePamphilis, C. W., Müller, K. F. and Quandt, D. (2011) 'The evolution of the plastid chromosome in land plants: gene content, gene order, gene function', *Plant Molecular Biology*, 76(3), pp. 273–297. doi: 10.1007/s11103-011-9762-4.

Wieczorek, J., Guo, Q. and Hijmans, R. (2004) 'The point-radius method for georeferencing locality descriptions and calculating associated uncertainty', *International Journal of Geographical Information Science*, 18(8), pp. 745–767. doi: 10.1080/13658810412331280211.

Wight, R. (1847) 'Notes on Indian Botany', in *Calcutta journal of natural history, and miscellany of the arts and sciences in India*. Calcutta: Bishop's College Press, pp. 143–189. Available at: https://www.biodiversitylibrary.org/item/124837.

Wikström, N., Bremer, B. and Rydin, C. (2020) 'Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae)', *Journal of Systematics and Evolution*, 58(4), pp. 440–460. doi: https://doi.org/10.1111/jse.12566.

Wilson, E. O. (2004) 'Taxonomy as a fundamental discipline', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. Edited by H. C. J. Godfray and S. Knapp, 359(1444), pp. 739–739. doi: 10.1098/rstb.2003.1440.

Wilson, M. C. *et al.* (2016) 'Habitat fragmentation and biodiversity conservation: key findings and future challenges', *Landscape Ecology*, 31(2), pp. 219–227. doi: 10.1007/s10980-015-0312-3.

Witten, I. H., Frank, E., Hall, M. A. and Pal, C. J. (2017) 'Chapter 12 - ensemble learning', in Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (eds) *Data Mining*. Fourth Edition. Morgan Kaufmann, pp. 479–501. doi: https://doi.org/10.1016/B978-0-12-804291-5.00012-X.

Wolpert, D. H. and Macready, W. G. (1997) 'No free lunch theorems for optimization', *IEEE Transactions on Evolutionary Computation*, 1(1), pp. 67–82. doi: 10.1109/4235.585893.

Wong, K. M. (1989) 'Rubiaceae', in Ng, F. S. P. (ed.) *Tree Flora of Malaya. Vol. 4*. Kuala Lumpur: Longman, pp. 324–425.

Wong, K. M. (2018) 'Flora of Singapore Precursors, 11. A new species of Urophyllum (Rubiaceae) from the Malay Peninsula', *Phytotaxa*, 373(3), p. 236. doi: 10.11646/phytotaxa.373.3.7.

Wong, K. M., Turner, I. M., Wang, R. J., Harwood, R., Seah, W. W., Ng, X. Y., Lim, R. C. J., Lua, H. K. and Mahyuni, R. (2019) 'Rubiaceae', in Middleton, D. J., Lindsay, S., and Leong-Škorničková, J. (eds) *Flora of Singapore Vol. 13 (Gentianales)*. National Parks Board Singapore, pp. 1–358.

Wood, J. R. I., Muñoz-Rodríguez, P., Williams, B. R. M. and Scotland, R. W. (2020) 'A foundation monograph of Ipomoea (Convolvulaceae) in the New World', *PhytoKeys*, 143, pp. 1–823. Available at: https://doi.org/10.3897/phytokeys.143.32821.

Yooprasert, S. (2021) *Taxonomy and Systematics of Urophyllum (Rubiaceae) in Thailand and Indochina*. University of Reading (unpublished thesis).

Zelditch, M. L., Swiderski, D. L. and Sheets, H. D. (2012) *Geometric morphometrics for biologists*. second. Amsterdam ; Boston: Elsevier Academic Press.

Zhang, H.-R., Zhang, X.-C. and Xiang, Q.-P. (2019) 'Directed repeats co-occur with few short-dispersed repeats in plastid genome of a spikemoss, *Selaginella vardei* (Selaginellaceae, Lycopodiopsida)', *BMC Genomics*, 20(1), p. 484. doi: 10.1186/s12864-019-5843-6.

Zhang, N., Zeng, L., Shan, H. and Ma, H. (2012) 'Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms', *New Phytologist*, 195(4), pp. 923–937. doi: https://doi.org/10.1111/j.1469-8137.2012.04212.x.

Zhang, R., Li, Q., Gao, J., Qu, M. and Ding, P. (2016) 'The complete chloroplast genome sequence of the medicinal plant *Morinda officinalis* (Rubiaceae), an endemic to China', *Mitochondrial DNA Part A*, 27(6), pp. 4324–4325. doi: 10.3109/19401736.2015.1089484.

Zhang, X. F., Wang, J. H., Zhao, K. K., Fan, W. W., Wang, H. X., Zhu, Z. X. and Wang, H. F. (2019) 'Complete plastome sequence of *Hedyotis ovata* Thunb. ex Maxim (Rubiaceae): an endemic shrub in Hainan, China', *Mitochondrial DNA Part B*, 4(1), pp. 675–676. doi: 10.1080/23802359.2019.1572467.

Zhang, Y., Zhang, J. W., Yang, Y. and Li, X. N. (2019) 'Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* Gaertn.f. and related Rubiaceae species', *Forests*, 10(11). doi: 10.3390/f10111000.

Zhou, Y., Duvaux, L., Ren, G., Zhang, L., Savolainen, O. and Liu, J. (2017) 'Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions', *Heredity*, 118(3), pp. 211–220. doi: 10.1038/hdy.2016.72.

Zhu, Z.-X., Wang, J.-H., Zhao, K.-K., Fan, W.-W., Wang, H.-X. and Wang, H.-F. (2019) 'Complete chloroplast genome of *Saprosma merrillii* Lo. (Rubiaceae): a Near Threaten (NT) shrub species endemic to Hainan Province, China', *Mitochondrial DNA Part B*, 4(1), pp. 742–743. doi: 10.1080/23802359.2019.1565931.

Appendix A Fieldwork collection

Three fieldwork expeditions were conducted in Thailand and Vietnam from 2017–2019. The total number of *Urophyllum* specimen collections are 141 specimens as listed below.

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
1	Yooprasert *et al.*	KRP 76	8 June 2017	Krabi	Thailand	8°14'25"	N	98°54'56"	E	*U. longifolium*
2	Yooprasert *et al.*	KRP 77	8 June 2017	Krabi	Thailand	8°14'25"	N	98°54'56"	E	*U. longifolium*
3	Yooprasert *et al.*	KRP 79	8 June 2017	Krabi	Thailand	8°14'25"	N	98°54'56"	E	*U. longifolium*
4	Yooprasert *et al.*	KRP 80	8 June 2017	Krabi	Thailand	8°14'25"	N	98°54'56"	E	*U. longifolium*
5	Yooprasert *et al.*	TRC 84	9 June 2017	Trang	Thailand	7°32'28"	N	99°	E	*U. glabrum*
6	Yooprasert *et al.*	TRC 85	9 June 2017	Trang	Thailand	7°32'28"	N	99°	E	*U. glabrum*
7	Yooprasert *et al.*	TRC 86	9 June 2017	Trang	Thailand	7°32'28"	N	99°	E	*U. glabrum*
8	Yooprasert *et al.*	TRC 87	9 June 2017	Trang	Thailand	7°32'23"	N	99°	E	*U. glabrum*
9	Yooprasert *et al.*	TRC 88	9 June 2017	Trang	Thailand	7°32'22"	N	99°	E	*U. glabrum*
10	Yooprasert *et al.*	TRC 89	9 June 2017	Trang	Thailand	7°32'25"	N	99°	E	*U. glabrum*
11	Yooprasert *et al.*	TRC 90	9 June 2017	Trang	Thailand	7°32'37"	N	99°47'48"	E	*U. longifolium*
12	Yooprasert *et al.*	TRC 91	9 June 2017	Trang	Thailand	7°32'37"	N	99°47'48"	E	*U. longifolium*
13	Yooprasert *et al.*	TRC 92	9 June 2017	Trang	Thailand	7°32'30"	N	99°	E	*U. glabrum*
14	Yooprasert *et al.*	PHL 104	10 June 2017	PhangNga	Thailand	8°37'34"	N	98°14'16"	E	*U. longifolium*
15	Yooprasert *et al.*	PHM 100	10 June 2017	PhangNga	Thailand	8°30'38"	N	98°32'29"	E	*U. longifolium*
16	Yooprasert *et al.*	PHM 101	10 June 2017	PhangNga	Thailand	8°30'38"	N	98°32'29"	E	*U. longifolium*
17	Yooprasert *et al.*	PHM 102	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'29"	E	*U. longifolium*
18	Yooprasert *et al.*	PHM 103	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'29"	E	*U. longifolium*
19	Yooprasert *et al.*	PHM 93	10 June 2017	PhangNga	Thailand	8°30'41"	N	98°32'26"	E	*U. longifolium*
20	Yooprasert *et al.*	PHM 94	10 June 2017	PhangNga	Thailand	8°30'41"	N	98°32'26"	E	*U. longifolium*
21	Yooprasert *et al.*	PHM 95	10 June 2017	PhangNga	Thailand	8°30'41"	N	98°32'28"	E	*U. longifolium*

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
22	Yooprasert *et al.*	PHM 96	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'28"	E	*U. longifolium*
23	Yooprasert *et al.*	PHM 97	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'28"	E	*U. longifolium*
24	Yooprasert *et al.*	PHM 98	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'29"	E	*U. longifolium*
25	Yooprasert *et al.*	PHM 98a	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'29"	E	*U. longifolium*
26	Yooprasert *et al.*	PHM 99	10 June 2017	PhangNga	Thailand	8°30'39"	N	98°32'29"	E	*U. longifolium*
27	Yooprasert *et al.*	PHS 126	11 June 2017	PhangNga	Thailand	8°59'59"	N	98°27'41"	E	*U. longifolium*
28	Yooprasert *et al.*	PHT 105	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. longifolium*
29	Yooprasert *et al.*	PHT 106	11 June 2017	PhangNga	Thailand	8°39'31"	N	98°17'2"	E	*U. glabrum*
30	Yooprasert *et al.*	PHT 107	11 June 2017	PhangNga	Thailand	8°39'31"	N	98°17'2"	E	*U. glabrum*
31	Yooprasert *et al.*	PHT 108	11 June 2017	PhangNga	Thailand	8°39'26"	N	98°17'1"	E	*U. longifolium*
32	Yooprasert *et al.*	PHT 109	11 June 2017	PhangNga	Thailand	8°39'26"	N	98°17'1"	E	*U. glabrum*
33	Yooprasert *et al.*	PHT 110	11 June 2017	PhangNga	Thailand	8°39'26"	N	98°17'1"	E	*U. longifolium*
34	Yooprasert *et al.*	PHT 111	11 June 2017	PhangNga	Thailand	8°39'24"	N	98°17'1"	E	*U. glabrum*
35	Yooprasert *et al.*	PHT 115	11 June 2017	PhangNga	Thailand	8°29'18"	N	98°17'2"	E	*U. longifolium*
36	Yooprasert *et al.*	PHT 116	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. glabrum*
37	Yooprasert *et al.*	PHT 117	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. glabrum*
38	Yooprasert *et al.*	PHT 119	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. glabrum*
39	Yooprasert *et al.*	PHT 120	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. glabrum*
40	Yooprasert *et al.*	PHT 121	11 June 2017	PhangNga	Thailand	8°39'18"	N	98°17'2"	E	*U. longifolium*
41	Yooprasert *et al.*	PHS 122	12 June 2017	PhangNga	Thailand	8°59'53"	N	98°27'36"	E	*U. longifolium*
42	Yooprasert *et al.*	PHS 123	12 June 2017	PhangNga	Thailand	8°59'53"	N	98°27'36"	E	*U. longifolium*
43	Yooprasert *et al.*	PHS 125	12 June 2017	PhangNga	Thailand	8°59'53"	N	98°27'40"	E	*U. longifolium*
44	Yooprasert *et al.*	PHS 128	12 June 2017	PhangNga	Thailand	8°59'48"	N	98°28'10"	E	*U. longifolium*
45	Yooprasert *et al.*	PHS 129	12 June 2017	PhangNga	Thailand	8°59'48"	N	98°28'10"	E	*U. longifolium*
46	Yooprasert *et al.*	RNK 131	13 June 2017	Ranong	Thailand	9°27'36"	N	98°30'38"	E	*U. longifolium*
47	Yooprasert *et al.*	RNK 132	13 June 2017	Ranong	Thailand	9°27'36"	N	98°30'38"	E	*U. longifolium*
48	Yooprasert *et al.*	RNK 133	13 June 2017	Ranong	Thailand	9°27'34"	N	98°30'40"	E	*U. longifolium*

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
49	Yooprasert *et al.*	RNK 134	13 June 2017	Ranong	Thailand	9°27'30"	N	98°30'41"	E	*U. longifolium*
50	Yooprasert *et al.*	NSK 136	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'14"	N	99°40'28"	E	*U. longifolium*
51	Yooprasert *et al.*	NSK 137	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'14"	N	99°40'28"	E	*U. crassum*
52	Yooprasert *et al.*	NSK 138	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'12"	N	99°40'28"	E	*U. crassum*
53	Yooprasert *et al.*	NSK 139	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'12"	N	99°40'28"	E	*U. crassum*
54	Yooprasert *et al.*	NSK 140	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'12"	N	99°40'28"	E	*U. crassum*
55	Yooprasert *et al.*	NSK 141	18 June 2017	Nakhon Sri Thammarat	Thailand	8°43'11"	N	99°40'33"	E	*U. longifolium*
56	Yooprasert *et al.*	NSK 144	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'44"	N	99°40'55"	E	*U. longifolium*
57	Yooprasert *et al.*	NSK 145	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'40"	N	99°40'55"	E	*U. crassum*
58	Yooprasert *et al.*	NSK 146	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'38"	N	99°40'55"	E	*U. crassum*
59	Yooprasert *et al.*	NSK 147	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'38"	N	99°40'55"	E	*U. crassum*
60	Yooprasert *et al.*	NSK 148	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'38"	N	99°40'55"	E	*U. crassum*
61	Yooprasert *et al.*	NSK 149	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'38"	N	99°40'55"	E	*U. crassum*
62	Yooprasert *et al.*	NSK 151	18 June 2017	Nakhon Sri Thammarat	Thailand	8°42'38"	N	99°40'55"	E	*U. crassum*
63	Yooprasert *et al.*	STT 152	19 June 2017	Satun	Thailand	6°42'46"	N	100°10'15"	E	*U. longifolium*
64	Yooprasert *et al.*	SKT 157	20 June 2017	Songkhla	Thailand	6°56'44"	N	100°10'15"	E	*U. glabrum*

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
65	Yooprasert *et al.*	SKT 158	20 June 2017	Songkhla	Thailand	6°56'45"	N	100°10'15"	E	*U. glabrum*
66	Yooprasert *et al.*	PCK 160	26 June 2017	Petchaburi	Thailand	12°49'26"	N	99°22'24"	E	*U. longifolium*
67	Yooprasert *et al.*	PCK 161	26 June 2017	Petchaburi	Thailand	12°49'29"	N	99°22'28"	E	*U. longifolium*
68	Yooprasert *et al.*	PCK 164	26 June 2017	Petchaburi	Thailand	12°49'29"	N	99°22'28"	E	*U. longifolium*
69	Yooprasert *et al.*	PHT 171	7 April 2018	PhangNga	Thailand	8°39'27.3"	N	98°17'1.2"	E	*U. glabrum*
70	Yooprasert *et al.*	PHT 173	7 April 2018	PhangNga	Thailand	8°39'20.4"	N	98°17'6.5"	E	*U. glabrum*
71	Yooprasert *et al.*	PHT 174	7 April 2018	PhangNga	Thailand	8°39'20.4"	N	98°17'6.5"	E	*U. glabrum*
72	Yooprasert *et al.*	PHT 175	7 April 2018	PhangNga	Thailand	8°39'13"	N	98°17'4.8"	E	*U. glabrum*
73	Yooprasert *et al.*	PHT 176	7 April 2018	PhangNga	Thailand	8°37'1.7"	N	98°14'48.2"	E	*U. glabrum*
74	Yooprasert *et al.*	SKN 178	9 April 2018	Songkhla	Thailand	6°33'53"	N	100°35'34"	E	*U. glabrum*
75	Yooprasert *et al.*	SKN 179	10 April 2018	Songkhla	Thailand	6°33'53"	N	100°35'34"	E	*U. glabrum*
76	Yooprasert *et al.*	SKS 180	10 April 2018	Songkhla	Thailand	6°31'47"	N	100°54'44"	E	*U. glabrum*
77	Yooprasert *et al.*	SKS 181	10 April 2018	Songkhla	Thailand	6°31'47"	N	100°54'44"	E	*U. glabrum*
78	Yooprasert *et al.*	182	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. glabrum*
79	Yooprasert *et al.*	183	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. crassum*
80	Yooprasert *et al.*	185	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. hirsutum*
81	Yooprasert *et al.*	186	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. hirsutum*
82	Yooprasert *et al.*	190	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. hirsutum*
83	Yooprasert *et al.*	186a	10 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. glabrum*
84	Yooprasert *et al.*	192	11 May 2019	Narathiwat	Thailand	5°47'55"	N	101°50'0.7"	E	*U. glabrum*
85	Yooprasert *et al.*	195	11 May 2019	Narathiwat	Thailand	5°48'6.9"	N	101°50'28.5"	E	*U. crassum*
86	Yooprasert *et al.*	197	11 May 2019	Narathiwat	Thailand	5°48'5.4"	N	101°50'28.5"	E	*U. glabrum*
87	Yooprasert *et al.*	198	11 May 2019	Narathiwat	Thailand	5°48'5.4"	N	101°50'28.5"	E	*U. hirsutum*
88	Yooprasert *et al.*	199	11 May 2019	Narathiwat	Thailand	5°48'5.4"	N	101°50'28.5"	E	*U. blumeanum*
89	Yooprasert *et al.*	200	11 May 2019	Narathiwat	Thailand	5°48'5.4"	N	101°50'28.5"	E	*U. glabrum*
90	Yooprasert *et al.*	202	11 May 2019	Narathiwat	Thailand	5°47'54.6"	N	101°45'30.4"	E	*U. hirsutum*
91	Yooprasert *et al.*	206	13 May 2019	Yala	Thailand	5°38'35.1"	N	101°7'47.9"	E	*U. villosum*

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
92	Yooprasert *et al.*	210	13 May 2019	Yala	Thailand	5°38'37.5"	N	101°7'50.1"	E	*U. streptopodium*
93	Yooprasert *et al.*	211	13 May 2019	Yala	Thailand	5°38'37.5"	N	101°7'50.1"	E	*U. glabrum*
94	Yooprasert *et al.*	212	13 May 2019	Yala	Thailand	5°51'28.8"	N	101°14'11"	E	*U. streptopodium*
95	Yooprasert *et al.*	214	13 May 2019	Yala	Thailand	5°51'28.8"	N	101°14'11"	E	*U. longipes*
96	Yooprasert *et al.*	215	13 May 2019	Yala	Thailand	5°51'28.8"	N	101°14'11"	E	*U. longipes*
97	Yooprasert *et al.*	216	13 May 2019	Yala	Thailand	5°51'28.8"	N	101°14'11"	E	*U. longipes*
98	Yooprasert *et al.*	218	14 May 2019	Yala	Thailand	5°53'19.6"	N	101°1′19.9″	E	*U. villosum*
99	Yooprasert *et al.*	220	14 May 2019	Yala	Thailand	5°52'40.3"	N	101°1'18"	E	*U. longipes*
100	Yooprasert *et al.*	223	14 May 2019	Yala	Thailand	6°19'8.7"	N	101°22'39"	E	*U. glabrum*
101	Yooprasert *et al.*	224	14 May 2019	Yala	Thailand	6°19'9.9"	N	101°22'39"	E	*U. glabrum*
102	Yooprasert *et al.*	225	14 May 2019	Yala	Thailand	6°19'9.9"	N	101°22'39"	E	*U. hirsutum*
103	Yooprasert *et al.*	VN 42-2	25 May 2019	Thua Thien-Hue	Vietnam	16°11'47.7"	N	107°51'42.5"	E	*U. argenteum*
104	Yooprasert *et al.*	VN 42-3	25 May 2019	Thua Thien-Hue	Vietnam	16°11'47.7"	N	107°51'42.5"	E	*U. argenteum*
105	Yooprasert *et al.*	VN 43-1	25 May 2019	Thua Thien-Hue	Vietnam	16°11'50.4"	N	107°51'39.1"	E	*U. argenteum*
106	Yooprasert *et al.*	VN 45-1	25 May 2019	Thua Thien-Hue	Vietnam	16°11'56.3"	N	107°51'25.9"	E	*U. argenteum*
107	Yooprasert *et al.*	VN 46-1	25 May 2019	Thua Thien-Hue	Vietnam	16°11'53.9"	N	107°51'23.4"	E	*U. argenteum*
108	Yooprasert *et al.*	VN 61-3	26 May 2019	Thua Thien-Hue	Vietnam	16°7'34.8"	N	107°49'18.1"	E	*U. argenteum*
109	Yooprasert *et al.*	VN 64-4	28 May 2019	Khanh Hoa	Vietnam	12°7'6.9"	N	108°56′47.8″	E	*U. chinense*
110	Yooprasert *et al.*	VN 65-2	28 May 2019	Khanh Hoa	Vietnam	12°7'7"	N	108°56′47.8″	E	*U. chinense*
111	Yooprasert *et al.*	VN 65-3	28 May 2019	Khanh Hoa	Vietnam	12°7'7"	N	108°56′47.8″	E	*U. chinense*
112	Yooprasert *et al.*	VN 67-1	28 May 2019	Khanh Hoa	Vietnam	12°6'55.9"	N	108°56′39.7″	E	*U. chinense*
113	Yooprasert *et al.*	VN 73-1	29 May 2019	Khanh Hoa	Vietnam	12°6'48.4"	N	108°58'8"	E	*U. argenteum*
114	Yooprasert *et al.*	VN 73-3	29 May 2019	Khanh Hoa	Vietnam	12°6'48.4"	N	108°58'8"	E	*U. annamense*
115	Yooprasert *et al.*	VN 74-1	29 May 2019	Khanh Hoa	Vietnam	12°6'43.1"	N	108°58'25.9"	E	*U. argenteum*
116	Yooprasert *et al.*	VN 74-2	29 May 2019	Khanh Hoa	Vietnam	12°6'43.1"	N	108°28'25.9"	E	*U. argenteum*
117	Yooprasert *et al.*	VN 74-5	29 May 2019	Khanh Hoa	Vietnam	12°6'43.1"	N	108°58'25.9"	E	*U. annamense*
118	Yooprasert *et al.*	VN 82-1	31 May 2019	Lam Dong	Vietnam	12°8'41.4"	N	108°31'47.5"	E	*U. chinense* subsp.

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
										latistipulum
119	Yooprasert *et al.*	VN 82-2	31 May 2019	Lam Dong	Vietnam	12°8'41.4"	N	108°31'47.5"	E	*U. chinense* subsp. *latistipulum*
120	Yooprasert *et al.*	VN 100-1	1 June 2019	Lam Dong	Vietnam	12°8'48.3"	N	108°40'16.6"	E	*U. chinense* subsp. *latistipulum*
121	Yooprasert *et al.*	VN 100-2	1 June 2019	Lam Dong	Vietnam	12°8'48.3"	N	108°40'16.6"	E	*U. lecomtei*
122	Yooprasert *et al.*	VN 100-3	1 June 2019	Lam Dong	Vietnam	12°8'48.3"	N	108°40'16.6"	E	*U. lecomtei*
123	Yooprasert *et al.*	VN 100-4	1 June 2019	Lam Dong	Vietnam	12°8'48.3"	N	108°40'16.6"	E	*U. lecomtei*
124	Yooprasert *et al.*	VN 102-7	1 June 2019	Lam Dong	Vietnam	12°8'44.6"	N	108°40'17.2"	E	*U. chinense* subsp. *latistipulum*
125	Yooprasert *et al.*	VN 85-3	1 June 2019	Khanh Hoa	Vietnam	12°11'11.8"	N	108°42'53.3"	E	*U. bidoupense*
126	Yooprasert *et al.*	VN 87-1	1 June 2019	Khanh Hoa	Vietnam	12°11'16.1"	N	108°42'50.9"	E	*U. bidoupense*
127	Yooprasert *et al.*	VN 87-3	1 June 2019	Khanh Hoa	Vietnam	12°11'16.1"	N	108°42'50.9"	E	*U. bidoupense*
128	Yooprasert *et al.*	VN 87-4	1 June 2019	Khanh Hoa	Vietnam	12°11'16.1"	N	108°42'50.9"	E	*U. bidoupense*
129	Yooprasert *et al.*	VN 91-1	1 June 2019	Khanh Hoa	Vietnam	12°11'14.3"	N	108°42'51.7"	E	*U. bidoupense*
130	Yooprasert *et al.*	VN 96-1	1 June 2019	Lam Dong	Vietnam	12°10'47.5"	N	108°41'6.9"	E	*U. chinense* subsp. *latistipulum*
131	Yooprasert *et al.*	VN 96-2	1 June 2019	Lam Dong	Vietnam	12°10'47.5"	N	108°41'6.9"	E	*U. chinense* subsp. *latistipulum*
132	Yooprasert *et al.*	VN 99-1	1 June 2019	Lam Dong	Vietnam	12°8'49.1"	N	108°40'16.2"	E	*U. pseudoschmidtii*
133	Yooprasert *et al.*	VN 99-2	1 June 2019	Lam Dong	Vietnam	12°8'49.1"	N	108°40'16.2"	E	*U. lecomtei*
134	Yooprasert *et al.*	VN 99-5	1 June 2019	Lam Dong	Vietnam	12°8'49.1"	N	108°40'16.2"	E	*U. lecomtei*
135	Yooprasert *et al.*	VN 105-1	2 June 2019	Lam Dong	Vietnam	12°9'21.5"	N	108°21'57.6"	E	*U. chinense* subsp. *latistipulum*
136	Yooprasert *et al.*	VN 107-2	3 June 2019	Lam Dong	Vietnam	12°7'59.9"	N	108°39'1.9"	E	*U. chinense* subsp. *latistipulum*
137	Yooprasert *et al.*	VN 112-1a	3 June 2019	Lam Dong	Vietnam	12°9'35.9"	N	108°32'10.1"	E	*U. pseudoschmidtii*
138	Yooprasert *et al.*	VN 112-1b	3 June 2019	Lam Dong	Vietnam	12°9'35.9"	N	108°32'10.1"	E	*U. pseudoschmidtii*

No.	Collector	Collector number	Collection Date	Province	Country	Latitude		Longitude		species
139	Yooprasert *et al.*	VN 112-2	3 June 2019	Lam Dong	Vietnam	12°9'35.9"	N	108°32'10.1"	E	*U. chinense* subsp. *latistipulum*
140	Yooprasert *et al.*	VN 116-5	4 June 2019	Lam Dong	Vietnam	12°5'50.4"	N	108°22'36"	E	*U. chinense* subsp. *latistipulum*
141	Yooprasert *et al.*	VN 116-6	5 June 2019	Lam Dong	Vietnam	12°5'50.4"	N	108°22'36"	E	*U. lecomtei*

Appendices B and C are supplied in electronic form.

Appendix B

Table S2.1 130 samples from 13 taxa of *Urophyllum* used in characters measurement. Herbarium acronyms (AAU, BKF and FU) followed Index Herbariorum (Thiers, 2020) ; SY indicates samples collecting by the researcher.

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
FU	Tagane, S. *et al.*	V1717	Vietnam	*argenteum*	AR	ar01
FU	Yahara, T. *et al.*	V2376	Vietnam	*argenteum*	AR	ar02
SY	Yooprasert, S. *et al.*	VN42-3	Vietnam	*argenteum*	AR	ar03
SY	Yooprasert, S. *et al.*	VN43-1	Vietnam	*argenteum*	AR	ar04
SY	Yooprasert, S. *et al.*	VN46-1	Vietnam	*argenteum*	AR	ar05
SY	Yooprasert, S. *et al.*	VN73-1	Vietnam	*argenteum*	AR	ar06
SY	Yooprasert, S. *et al.*	VN74-1	Vietnam	*argenteum*	AR	ar07
AAU	Huang	V110455	China	*chinense*	CH	ch01
FU	Tagane, S. *et al.*	V1725	Vietnam	*chinense*	CH	ch02
SY	Yooprasert, S. *et al.*	VN65-2	Vietnam	*chinense*	CH	ch03
SY	Yooprasert, S. *et al.*	VN65-3	Vietnam	*chinense*	CH	ch04
SY	Yooprasert, S. *et al.*	VN67-1	Vietnam	*chinense*	CH	ch05
AAU	Larsen, K.	42955	Thailand	*crassum*	CR	cr01
BKF	Chamchumroon, V.	vc882	Thailand	*crassum*	CR	cr02
BKF	Poopath, M.	MP1674	Thailand	*crassum*	CR	cr03
BKF	Poopath, M.	MP1999	Thailand	*crassum*	CR	cr04
SY	Yooprasert, S. *et al.*	NSK140	Thailand	*crassum*	CR	cr05
SY	Yooprasert, S. *et al.*	NSK149	Thailand	*crassum*	CR	cr06
SY	Yooprasert, S. *et al.*	NSK151	Thailand	*crassum*	CR	cr07
AAU	Larsen, K.	42866	Thailand	*glabrum*	GL	gl01
AAU	Larsen, K.	45570	Thailand	*glabrum*	GL	gl02
AAU	Pooma, R.	5133	Thailand	*glabrum*	GL	gl03
BKF	Gardner, S.	ST0555	Thailand	*glabrum*	GL	gl04

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
BKF	Larsen, K.	46090	Thailand	*glabrum*	GL	gl05
BKF	Pooma, R.	4358	Thailand	*glabrum*	GL	gl06
BKF	Poopath, M.	2	Thailand	*glabrum*	GL	gl07
BKF	Puudjaa, P.	621	Thailand	*glabrum*	GL	gl08
BKF	Suddee, S.	3156	Thailand	*glabrum*	GL	gl09
BKF	Tippayasri, P.	ST1062	Thailand	*glabrum*	GL	gl10
SY	Yooprasert, S. *et al.*	182	Thailand	*glabrum*	GL	gl11
SY	Yooprasert, S. *et al.*	192	Thailand	*glabrum*	GL	gl12
SY	Yooprasert, S. *et al.*	223	Thailand	*glabrum*	GL	gl13
SY	Yooprasert, S. *et al.*	224	Thailand	*glabrum*	GL	gl14
SY	Yooprasert, S. *et al.*	186a	Thailand	*glabrum*	GL	gl15
SY	Yooprasert, S. *et al.*	PHT107	Thailand	*glabrum*	GL	gl16
SY	Yooprasert, S. *et al.*	PHT109	Thailand	*glabrum*	GL	gl17
SY	Yooprasert, S. *et al.*	PHT116	Thailand	*glabrum*	GL	gl18
SY	Yooprasert, S. *et al.*	PHT119	Thailand	*glabrum*	GL	gl19
SY	Yooprasert, S. *et al.*	PHT120	Thailand	*glabrum*	GL	gl20
SY	Yooprasert, S. *et al.*	PHT171	Thailand	*glabrum*	GL	gl21
SY	Yooprasert, S. *et al.*	PHT174	Thailand	*glabrum*	GL	gl22
SY	Yooprasert, S. *et al.*	PHT176	Thailand	*glabrum*	GL	gl23
SY	Yooprasert, S. *et al.*	SKT158	Thailand	*glabrum*	GL	gl24
SY	Yooprasert, S. *et al.*	TRC84	Thailand	*glabrum*	GL	gl25
SY	Yooprasert, S. *et al.*	TRC85	Thailand	*glabrum*	GL	gl26
SY	Yooprasert, S. *et al.*	TRC88	Thailand	*glabrum*	GL	gl27
SY	Yooprasert, S. *et al.*	TRC89	Thailand	*glabrum*	GL	gl28
SY	Yooprasert, S. *et al.*	TRC92	Thailand	*glabrum*	GL	gl29
BKF	Kiah	SING31745	Peninsular Malaysia	*hirsutum*	HI	hi01

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
BKF	Pooma, R.	3184	Thailand	*hirsutum*	HI	hi02
BKF	Poopath, M.	MP1696	Thailand	*hirsutum*	HI	hi03
BKF	Poopath, M.	MP1698	Thailand	*hirsutum*	HI	hi04
SY	Yooprasert, S. *et al.*	190	Thailand	*hirsutum*	HI	hi05
SY	Yooprasert, S. *et al.*	202	Thailand	*hirsutum*	HI	hi06
SY	Yooprasert, S. *et al.*	225	Thailand	*hirsutum*	HI	hi07
AAU	Averyanov, L.	VH1649	Vietnam	*lecomtei*	LE	le01
AAU	Averyanov, L.	VH2685	Vietnam	*lecomtei*	LE	le02
AAU	Averyanov, L.	VH63	Vietnam	*lecomtei*	LE	le03
FU	Tagane, S. *et al.*	V6085	Vietnam	*lecomtei*	LE	le04
SY	Yooprasert, S. *et al.*	VN100-3	Vietnam	*lecomtei*	LE	le05
SY	Yooprasert, S. *et al.*	VN100-4	Vietnam	*lecomtei*	LE	le06
SY	Yooprasert, S. *et al.*	VN99-2	Vietnam	*lecomtei*	LE	le07
SY	Yooprasert, S. *et al.*	VN99-5	Vietnam	*lecomtei*	LE	le08
AAU	Larsen, K.	43305	Thailand	*longifolium* f.B	LB	lb01
BKF	Gardner, S.	ST0927	Thailand	*longifolium* f.B	LB	lb02
BKF	Geesink, R.	7223	Thailand	*longifolium* f.B	LB	lb03
BKF	Phengkhlai, C.	15009	Thailand	*longifolium* f.B	LB	lb04
SY	Yooprasert, S. *et al.*	NSK141	Thailand	*longifolium* f.B	LB	lb05
SY	Yooprasert, S. *et al.*	STT152	Thailand	*longifolium* f.B	LB	lb06
SY	Yooprasert, S. *et al.*	TRC90	Thailand	*longifolium* f.B	LB	lb07
BKF	Chamchumroon, V.	vc872	Thailand	*longifolium* f.C	LC	lc01
BKF	Fukuoka, N.	T36007	Thailand	*longifolium* f.C	LC	lc02
BKF	Gardner, S.	ST0673	Thailand	*longifolium* f.C	LC	lc03
BKF	Gardner, S.	ST2604	Thailand	*longifolium* f.C	LC	lc04
BKF	Phengkhlai, C.	1302	Thailand	*longifolium* f.C	LC	lc05
FU	Tagane, S. *et al.*	MY339	Myanmar	*longifolium* f.C	LC	lc06

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
SY	Yooprasert, S. *et al.*	KRP76	Thailand	*longifolium* f.C	LC	lc07
SY	Yooprasert, S. *et al.*	KRP77	Thailand	*longifolium* f.C	LC	lc08
SY	Yooprasert, S. *et al.*	PHM100	Thailand	*longifolium* f.C	LC	lc09
SY	Yooprasert, S. *et al.*	PHM101	Thailand	*longifolium* f.C	LC	lc10
SY	Yooprasert, S. *et al.*	PHM102	Thailand	*longifolium* f.C	LC	lc11
SY	Yooprasert, S. *et al.*	PHM103	Thailand	*longifolium* f.C	LC	lc12
SY	Yooprasert, S. *et al.*	PHM93	Thailand	*longifolium* f.C	LC	lc13
SY	Yooprasert, S. *et al.*	PHM94	Thailand	*longifolium* f.C	LC	lc14
SY	Yooprasert, S. *et al.*	PHM95	Thailand	*longifolium* f.C	LC	lc15
SY	Yooprasert, S. *et al.*	PHM96	Thailand	*longifolium* f.C	LC	lc16
SY	Yooprasert, S. *et al.*	PHM99	Thailand	*longifolium* f.C	LC	lc17
SY	Yooprasert, S. *et al.*	PHS123	Thailand	*longifolium* f.C	LC	lc18
SY	Yooprasert, S. *et al.*	PHS125	Thailand	*longifolium* f.C	LC	lc19
SY	Yooprasert, S. *et al.*	PHT105	Thailand	*longifolium* f.C	LC	lc20
SY	Yooprasert, S. *et al.*	PHT108	Thailand	*longifolium* f.C	LC	lc21
SY	Yooprasert, S. *et al.*	PHT110	Thailand	*longifolium* f.C	LC	lc22
SY	Yooprasert, S. *et al.*	PHT115	Thailand	*longifolium* f.C	LC	lc23
SY	Yooprasert, S. *et al.*	PHT121	Thailand	*longifolium* f.C	LC	lc24
BKF	Poopath, M.	MP1274	Thailand	*longipes*	LG	lg01
BKF	Poopath, M.	MP1311	Thailand	*longipes*	LG	lg02
BKF	Poopath, M.	MP1604	Thailand	*longipes*	LG	lg03
BKF	Poopath, M.	MP1613	Thailand	*longipes*	LG	lg04
BKF	Poopath, M.	MP1683	Thailand	*longipes*	LG	lg05
BKF	Wai, J.	2654	Thailand	*longipes*	LG	lg06
SY	Yooprasert, S. *et al.*	214	Thailand	*longipes*	LG	lg07
SY	Yooprasert, S. *et al.*	220	Thailand	*longipes*	LG	lg08

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
AAU	3rd year student, Uni. of Malaya	sn (AAU2689)	Peninsular Malaysia	*streptopodium*	ST	st01
AAU	Maxwell, J.F.	77-352	Peninsular Malaysia	*streptopodium*	ST	st02
AAU	Pereira, J. *et al.*	JTP453	Malaysia	*streptopodium*	ST	st03
BKF	Purseglove, J.W.	P5510	Peninsular Malaysia	*streptopodium*	ST	st04
BKF	Saerudin, D.	370	Indonesia	*streptopodium*	ST	st05
SY	Yooprasert, S. *et al.*	212	Thailand	*streptopodium*	ST	st06
BKF	Pooma, R.	5102	Thailand	*villosum*	VI	vi01
BKF	Poopath, M.	913	Thailand	*villosum*	VI	vi02
BKF	Poopath, M.	2016	Thailand	*villosum*	VI	vi03
BKF	Wai, J.	2656	Thailand	*villosum*	VI	vi04
BKF	Wai, J.	2663	Thailand	*villosum*	VI	vi05
SY	Yooprasert, S. *et al.*	218	Thailand	*villosum*	VI	vi06
SY	Yooprasert, S. *et al.*	206	Thailand	*villosum*	VI	vi07
AAU	Averyanov, L.	VH2651	Vietnam	species1	S1	s101
AAU	Averyanov, L.	VH2941	Vietnam	species1	S1	s102
AAU	Averyanov, L.	VH2981	Vietnam	species1	S1	s103
AAU	Averyanov, L.	VH4075	Vietnam	species1	S1	s104
SY	Yooprasert, S. *et al.*	VN100-1	Vietnam	species1	S1	s105
SY	Yooprasert, S. *et al.*	VN102-7	Vietnam	species1	S1	s106
SY	Yooprasert, S. *et al.*	VN112-2	Vietnam	species1	S1	s107
SY	Yooprasert, S. *et al.*	VN116-5	Vietnam	species1	S1	s108
AAU	Averyanov, L.	VH4059	Vietnam	species2	S2	s201
AAU	Averyanov, L.	VH4154	Vietnam	species2	S2	s202
SY	Yooprasert, S. *et al.*	VN85-3	Vietnam	species2	S2	s203
SY	Yooprasert, S. *et al.*	VN87-1	Vietnam	species2	S2	s204

Herbarium or collection*	Collectors	Collector no.	Collected country	Specific ephithet	Taxa acronym	Code
SY	Yooprasert, S. *et al.*	VN87-3	Vietnam	species2	S2	s205
SY	Yooprasert, S. *et al.*	VN87-4	Vietnam	species2	S2	s206
SY	Yooprasert, S. *et al.*	VN91-1	Vietnam	species2	S2	s207

Table S2.2 Effect size and minimum sample size number of quantitative characters in linear dataset. Highlights indicate characters that omitted from the study.

No.	Characters	Effect size (η^2)	Minimum sample size
1	Petiole length	0.53	8
2	Leaf width	0.70	5
3	Leaf length	0.78	4
4	The widest point of the leaf to total leaf length (calculated by length from leaf base to the widest part x 100 /total leaf length)	0.21	46
5	Lateral vein number	0.78	4
6	Angle at 10% leaf length from the base	0.19	56
7	Angle at 25% leaf length from the base	0.18	62
8	Angle at 25% leaf length from the apex	0.44	11
9	Angle of tertiary veins to midrib at the base of a leaf (number of tertiary veins within 1 cm diameter)	0.56	7
10	Angle of tertiary veins to midrib at the mid of a leaf (number of tertiary veins within 1 cm diameter)	0.59	7
11	Stipule length	0.50	9
12	Primary peduncle length	0.40	13
13	Rachis length	0.29	25
14	secondary peduncle length	0.18	62
15	Pedicel length (up to six flowers)	0.19	56

Table S2.3 Raw characters data collected from *Urophyllum* specimens. (can be found in excel file Supplement CH2)

Table S2.4 Raw co-ordinates data of each leaf after landmarks digitisation. (can be found in excel file Supplement CH2)

Table S2.5 Log centroid sizes and principle component scores of geometric morphometric used in the study. (can be found in excel file Supplement CH2)

Table S2.6 Covariance correlation of linear morphometric data. (can be found in excel file Supplement CH2)

Table S2.7 Mean log centroid size (±standard deviation) of secondary vein loop of *Urophyllum* taxa.

Taxa	log centroid size (mean±SD)
AR	3.59±0.241
CH	3.98±0.179
CR	4.25±0.156
GL	4.17±0.209
HI	3.87±0.115
LB	4.11±0.294
LC	4.04±0.162
LE	3.69±0.216
LG	3.57±0.183
S1	3.96±0.163
S2	3.58±0.162
ST	3.97±0.269
VI	4.49±0.125

Table S2.8 One-Way ANOVA table testing on log centroid size mean between *Urophyllum* taxa.

	SS	**df**	**MS**	**F**	**P**
Between groups:	1.50540	12	0.12545	17.72	4.12E-21
Within groups:	0.82824	117	0.00708		**Permutation P** (n=99999)
Total:	2.33364	129			1.00E-05

Table S2.9 P values from Tukey's pairwise post-hoc tests of secondary vein loop log centroid size between *Urophyllum* taxa. Significant comparisons highlighted.

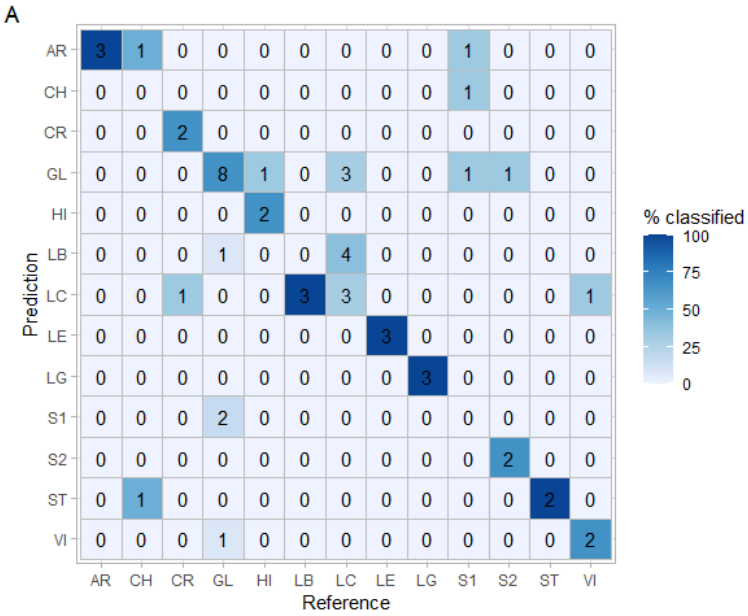
	AR	CH	CR	GL	HI	LB	LC	LE	LG	S1	S2	ST
AR		-	-	-	-	-	-	-	-	-	-	-
CH	0.0426		-	-	-	-	-	-	-	-	-	-
CR	3.08E-07	0.4720		-	-	-	-	-	-	-	-	-
GL	4.26E-09	0.6448	0.9997		-	-	-	-	-	-	-	-
HI	0.2446	0.9993	0.0245	0.0164		-	-	-	-	-	-	-
LB	0.0001	0.9927	0.9857	0.9999	0.5004		-	-	-	-	-	-
LC	0.00003	1	0.3814	0.3211	0.7304	0.9995		-	-	-	-	-
LE	0.998	0.3292	1.46E-05	4.70E-07	0.8488	0.0034	0.0018		-	-	-	-
LG	1	0.0176	3.92E-08	1.35E-10	0.1232	0.0001	0.0001	0.9852		-	-	-
S1	0.0214	1	0.1708	0.1851	0.9998	0.9341	0.9977	0.257	0.0067		-	-
S2	1	0.03302	1.96E-07	2.33E-09	0.2006	0.0001	0.0001	0.9953	1	0.01567		-
ST	0.0283	1	0.3595	0.4850	0.9992	0.9848	0.9999	0.2733	0.0103	1	0.0214	
VI	1.64E-12	9.72E-04	0.4721	0.0113	1.80E-06	0.0216	2.08E-05	8.61E-11	1.47E-13	3.21E-05	1.01E-12	3.15E-04

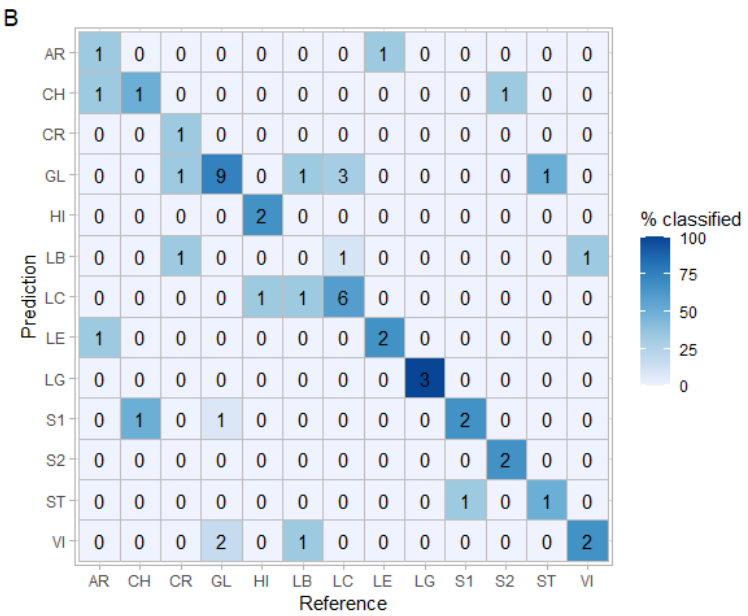
Table S2.10 Accuracy and kappa values from 3 repeats 5-fold cross-validation of top 3–6 classifiers from three datasets. Sub-column name under accuracy and kappa indicates rounds of training analyses.

Dataset	Classifier	Accuracy			Kappa		
		1	2	3	1	2	3
Linear	RF	0.910	0.938	0.916	0.896	0.928	0.904
	SVM	0.876	0.928	0.956	0.859	0.918	0.950
	RRF	0.863	0.885	0.831	0.844	0.869	0.808
Geometric	LDA	0.536	0.641	0.579	0.472	0.594	0.521
	PDA	0.513	0.611	0.557	0.449	0.560	0.495
	MDA	0.539	0.586	0.581	0.476	0.531	0.519
	SVM	0.562	0.618	0.598	0.487	0.557	0.545
	RRF	0.513	0.560	0.542	0.430	0.489	0.466
	RF	0.484	0.581	0.562	0.388	0.510	0.482
Mixed	RF	0.902	0.933	0.876	0.887	0.923	0.857
	SVM	0.876	0.928	0.956	0.859	0.918	0.950
	RRF	0.830	0.878	0.863	0.805	0.861	0.843
	KNN	0.849	0.809	0.836	0.826	0.782	0.811
	c4.5	0.790	0.871	0.839	0.762	0.854	0.819

Table S2.11 Accuracy and kappa values from prediction on test datasets.

Dataset	Classifier	Test round	Accuracy	Kappa
Linear	RF	1	0.868	0.848
		2	0.868	0.848
		3	0.925	0.913
		averaged	0.887	0.870
Geometric	PDA	1	0.585	0.530
		2	0.623	0.573
		3	0.627	0.566
		averaged	0.612	0.556
Mixed	RF	1	0.925	0.914
		2	0.868	0.848
		3	0.906	0.891
		averaged	0.900	0.884





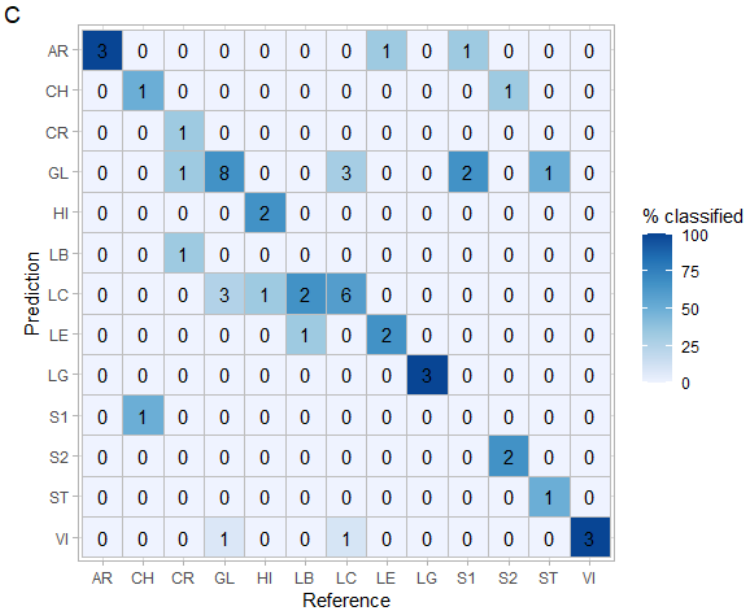


Figure S2.1 Confusion matrices from the linear discriminant analysis (LDA) classification using the Test manually combined dataset. Three replicates of partitioning are shown (A, B and C). Colours correspond to the percentage of classification in each category with numbers indicating the number of samples predicted for each taxon.

Appendix C

Supplement S1 Modified CTAB protocol.

Methods:

1. Weigh 20-30 mg of dried leaf material in a 2 ml Eppendorf tube, add 2 Qiagen tungsten carbide beads and a small amount of sand. Grind using TissueLyser at 30 hz for 45s, turn block and grind for a further 45s.
2. Remove beads and add 1.5 ml of ice-cold sorbitol wash buffer, mix well by inverting and flicking (shaking for slimy samples). Incubate for 10 mins on ice, inverting every 3 mins.
3. Centrifuge at 13,000 rpm for 5 mins (10,000 rpm for 3 mins for slimy samples), remove supernatant and discard, leaving only a pellet in the tube.
4. Repeat steps 2 & 3 twice.
5. Add 600µl of 2% CTAB buffer to the pellet and incubate at 60°C for 50 mins, mix by inverting/vortexing every 5-10 mins.
6. Add 3µl RNase A, mix well and incubate at 60°C for 10 mins.
7. Add 195µl of 3M potassium acetate, mix well and incubate on ice for 10 mins, mix by inverting after 5 mins.
8. Add 800µl of chloroform/isoamyl alcohol (24:1) and incubate on ice for 10 mins, inverting every 2-3 min.
9. Centrifuge at 13,000 rpm for 5 mins. From this centrifugation, there should be 3 separate phases, transfer the upper aqueous phase using a pipette into a new 2 ml Eppendorf tube.
10. Repeat step 8-9, inverting every 2-3 mins, centrifuge and transfer to a new 1.5 ml Eppendorf tube.
11. Add 0.08 volumes of 7.5M ammonium acetate, mix well by inverting.
12. Add 0.54 volume of Isopropanol (calculate total volume including ammonium acetate). Mix well by inverting then incubate in -20 ºC freezer for at least 60 mins or longer (overnight for samples with low yields of DNA). Longer times can yield more DNA but also more contaminants.
13. Centrifuge at 13,000 rpm for 15 mins. Pour or pipette off the supernatant, taking care not to lose the pellet.
14. Add 700µl of cold 70% ethanol, mix by flicking and leave it for 3 minutes or until the pellet becomes free.
15. Centrifuge at 13,000 rpm for 1 min.
16. Pipette off the liquid, taking care not to lose the pellet.
17. Repeat steps 14-16 twice.
18. Dry the pellet using CentriVap at 35°C (approximately 7-10 mins).
19. Resuspend pellet in 50µl of 10 mM Tris-Cl, pH 8 (EB buffer, Qiagen).
20. Leave the sample in the fridge (4°C) overnight for resuspension.

Solutions and Reagents

Stock solutions

- Sorbitol Buffer (100 mM Tris-HCl pH 8, 5 mM EDTA, 0.35M Sorbitol)
- 2% CTAB-Buffer, pH 8 (100 mM Tris-HCl, 20 mM EDTA, 1.4M NaCl, 2% w/v CTAB)
- Chloroform/Isoamylalcohol (24:1 v/v)
- 3M potassium acetate, pH 5.5
- 7.5M ammonium acetate
- Isopropanol (propan-2-ol)
- 70% Ethanol

- EB-Buffer, pH 8 (Qiagen)
- RNase A (100 mg/ml, Qiagen)

Working solutions

- 2% CTAB Buffer (2% CTAB-Buffer plus 0.2% v/v 2-mercaptoethanol and 4% w/v PVP-40)
- Sorbitol Wash Buffer (Sorbitol-Buffer plus 0.2% v/v 2-Mercaptoethanol, 1% w/v PVP-40)

Table S1a Primers used in the study.

No.	Junction	Primers	
		Forward	Reverse
1	IRa–LSC	ATGTTGGGGTGAACCAGAAA	GTTATGCATGAACGTAATGCTC
2	LSC–IRb	TGTCCGGCTATATACTCTGC	GTTATGCATGAACGTAATGCTC
3	IRb–SSC	GGTATTAGTCTGGATACAGC	CTATCTCTATGGGGTAAGGG
4	SSC–IRa	GGTATTAGTCTGGATACAGC	CAGTAAGAATACTATGAATCCG

Table S1b PCR cycling conditions.

Junction 1 and 2

No. of cycle	Temp. (℃)	Time (min)
1	94	2
35	94	1
	55	1
	72	2
1	72	8
1	4	10

Junction 3 and 4

No. of cycle	Temp. (℃)	Time (min)
1	94	2
35	94	1
	51	1
	72	2
1	72	8
1	4	10

Table S2 Morphological matrix of 39 *Urophyllum* taxa and outgroup species.

Species	Character states (12 characters)
Amphidasya ambigua	10?21?2?????
Colletoecema dewevrei	00?20?2?????
Lasianthus sp.	????????????
Ophiorrhiza mungos	10021310(0 1)???
*U. argenteum*_HN_F	01(0 1 2)31(0 1)211001
*U. argenteum*_HN_M	01(0 1 2)314211001
*U. blumeanum*_BW	0003152?0101
*U. chinense*_HN	00031(0 4)(0 1)00001
*U. crassum*_KU_F	01021(4 6)(0 1)12011
*U. crassum*_KU_M	01021(4 6)(0 1)12011
*U. crassum*_TO	01021(4 6)(0 1)12011
*U. glabrum*_KC_F	001114(0 1)10021
*U. glabrum*_KC_M	001114(0 1)10021
*U. glabrum*_SK	001114(0 1)10021
*U. glabrum*_TC_F	00111(2 4)(0 1)10021
*U. glabrum*_TC_M	001114(0 1)10021
*U. hirsutum*_BW	00421(4 6)221001
*U. hirsutum*_HB	00421(4 6)221001
*U. lecomtei*_DC	00(1 2)214201001
*U. lecomtei*_NL	00(1 2)216201001
*U. longifolium*_f.A_KK	01(0 3)212(0 1)00001
*U. longifolium*_f.B_KC	01(0 3)212(0 1)00001
*U. longifolium*_f.B_KU	01(0 3)212(0 1)00001
*U. longifolium*_f.B_NK	01(0 3)212(0 1)00001
*U. longifolium*_f.C_KL	01(0 3)212(0 1)00001
*U. longifolium*_f.C_KP	01(0 3)212(0 1)00001
*U. longifolium*_f.C_TC	01(0 3)212(0 1)00001
*U. longifolium*_f.C_TN	01(0 3)212(0 1)00001
*U. longifolium*_f.C_TT	01(0 3)212(0 1)00001
*U. longifolium*_var._*annamense*_HN	010212(0 1)100?1
*U. longipes*_BY	010211(0 1)00001
*U. macrophyllum*_HB	0002162?00?1
*U. memecyloides*_SB	00?2?50?00??
*U. schmidtii*_KR	3224200001
U. sp.1_DC	000304(0 1)00001
U. sp.2_DC_F	000304(0 1)00001
U. sp.2_DC_M	000304(0 1)00001
U. sp.3_BN	00(0 3)22(4 5 6)200000
U. sp.4_NL	0010142??0??
*U. streptopodium*_BY	00031(4 6)220001
*U. villosum*_BY_F	00121(4 5)211011

Species	Character states (12 characters)
*U. villosum*_BY_M	1212211011
*U. villosum*_TO	00121(4 5)211011

Table S3 Paired-end (PE) reads and average coverage of 39 *Urophyllum* plastome assemblies.

Species	No. of PE reads	average coverage
U. streptopodium_BY	38,450,090	1362
U. sp.2_DC_F	69,959,554	1254
U. chinense_HN	46,674,034	1074
U. argenteum_HN_F	43,863,582	1069
U. sp.2_DC_M	41,328,298	999
U. hirsutum_HB	45,737,898	875
U. sp.3_BN	46,893,758	768
U. longifolium_f.C_KL	40,278,920	669
U. argenteum_HN_M	50,748,624	660
U. sp.1_DC	56,803,372	609
U. longifolium_f.C_TC	37,559,570	577
U. longifolium_f.B_NK	49,086,242	564
U. longipes_BY	39,207,944	475
U. longifolium_f.B_KU	42,668,866	473
U. lecomtei_DC	45,421,612	444
U. glabrum_SK	34,840,328	432
U. longifolium_f.C_TN	35,124,928	412
U. blumeanum_BW	34,682,210	404
U. macrophyllum_HB	54,632,876	399
U. lecomtei_NL	49,049,798	374
U. longifolium var. annamense_HB	41,064,324	357
U. sp.4_NL	44,029,796	355
U. longifolium_f.B_KC	35,507,636	347
U. glabrum_KC_M	38,653,346	346
U. hirsutum_BW	37,324,216	343
U. glabrum_TC_M	29,500,626	332
U. schmidtii_KR	49,190,842	319
U. memecyloides_SB	43,644,492	311
U. glabrum_TC_F	38,267,050	310
U. longifolium_f.A_KK	34,919,330	274
U. longifolium_f.C_TT	32,752,204	257
U. glabrum_KC_F	34,771,876	226
U. longifolium_f.C_KP	29,615,176	197
U. crassum_TO	29,176,050	187
U. villosum_TO	35,120,048	145
U. villosum_BY_F	28,293,622	130
U. villosum_BY_M	28,952,394	109
U. crassum_KU_M	35,631,482	80
U. crassum_KU_F	32,033,868	72

Table S4 Convergence diagnostics of BI analyses from different datasets.

Datasets	Model	MCMC Generations	ASDSF*	95% HPD Interval			
				Parameters	min ESS	avg ESS	PSRF
plastome	GTR+G	1.5 millions	0.0024	TL	965.62	1016.55	1.000
				r(A<->C)	512.51	515.64	1.000
				r(A<->G)	445.10	505.52	1.000
				r(A<->T)	613.27	632.32	1.000
				r(C<->G)	530.27	590.70	1.000
				r(C<->T)	465.48	506.07	1.000
				r(G<->T)	396.62	487.71	1.001
				pi(A)	228.14	301.93	1.000
				pi(C)	313.58	320.74	1.004
				pi(G)	322.00	374.06	1.000
				pi(T)	245.29	282.76	1.000
				alpha	1126.00	1126.00	1.000

Table S4 Convergence diagnostics of BI analyses from different datasets (continued).

Datasets	Model	MCMC Generations	ASDSF*	95% HPD Interval			
				Parameters	min ESS	avg ESS	PSRF
nrDNA (39 taxa)	GTR+I+G	2 millions	0.0043	TL	1072.26	1134.86	1.000
				r(A<->C)	1218.00	1231.12	1.000
				r(A<->G)	804.67	904.28	1.000
				r(A<->T)	1020.20	1035.93	1.000
				r(C<->G)	1043.93	1167.70	1.000
				r(C<->T)	691.99	735.21	1.000
				r(G<->T)	904.56	1015.68	1.001
				pi(A)	1042.98	1102.60	1.000
				pi(C)	1249.05	1293.03	1.000
				pi(G)	1096.47	1150.02	1.000
				pi(T)	1159.70	1191.45	1.000
				alpha	664.23	668.89	1.000
				pinvar	606.29	623.07	1.000
* = Average standard deviation of split frequencies							

Table S5 Summary of *Urophyllum* plastomes sequenced including plastomes length, GC content and genes number.

Species	Plastome Length (bp)	LSC length (bp)	SSC length (bp)	IR length (bp)	GC content (%)	total genes	Coding genes (duplicate in IR)	tRNA (duplicate in IR)	rRNA (duplicate in IR)
*U. argenteum*_HB_F	154,099	84,544	18,275	25,640	37.8	113	80 (6)	29 (7)	4 (4)
*U. argenteum*_HB_M	154,099	84,544	18,275	25,640	37.8	113	80 (6)	29 (7)	4 (4)
U. sp.3_BN	154,166	84,535	18,313	25,659	37.8	113	80 (6)	29 (7)	4 (4)
*U. schmidtii*_KR	154,716	85,163	18,247	25,653	37.8	113	80 (6)	29 (7)	4 (4)
U. longifolium var. *annamense*_HN	154,738	84,868	18,290	25,790	37.8	113	80 (6)	29 (7)	4 (4)
*U. lecomtei*_DC	154,749	85,166	18,277	25,653	37.7	113	80 (6)	29 (7)	4 (4)
U. sp.4_NL	154,777	85,243	18,228	25,653	37.7	113	80 (6)	29 (7)	4 (4)
*U. lecomtei*_NL	154,779	85,194	18,279	25,653	37.7	113	80 (6)	29 (7)	4 (4)
*U. macrophyllum*_HB	154,842	85,021	18,235	25,793	37.7	113	80 (6)	29 (7)	4 (4)
*U. streptopodium*_HB	154,895	85,154	18,231	25,755	37.7	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.B_NK	154,899	85,171	18,178	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.C_KL	154,945	85,228	18,179	25,769	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.C_TN	154,953	85,236	18,179	25,769	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.C_KP	154,963	85,242	18,183	25,769	37.8	113	80 (6)	29 (7)	4 (4)
*U. glabrum*_TC_F	154,964	85,234	18,180	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. glabrum*_TC_M	154,964	85,234	18,180	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.B_KU	154,970	85,240	18,180	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.B_KC	154,971	85,243	18,178	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.A_KK	154,974	85,247	18,177	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. hirsutum*_BW	154,976	85,216	18,192	25,784	37.8	113	80 (6)	29 (7)	4 (4)

Table S5 Summary of *Urophyllum* plastomes sequenced including plastomes length, GC content and genes number (continued).

Species	Plastome Length (bp)	LSC length (bp)	SSC length (bp)	IR length (bp)	GC content (%)	total genes	Coding genes (duplicate in IR)	tRNA (duplicate in IR)	rRNA (duplicate in IR)
*U. longifolium*_f.C_TT	154,978	85,250	18,178	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. longifolium*_f.C_TC	154,987	85,259	18,178	25,775	37.8	113	80 (6)	29 (7)	4 (4)
*U. hirsutum*_HB	154,990	85,228	18,194	25,784	37.8	113	80 (6)	29 (7)	4 (4)
*U. blumeanum*_BY	154,991	85,224	18,187	25,790	37.8	113	80 (6)	29 (7)	4 (4)
*U. glabrum*_SK	155,003	85,249	18,186	25,784	37.8	113	80 (6)	29 (7)	4 (4)
*U. glabrum*_KC_F	155,013	85,347	18,184	25,741	37.8	113	80 (6)	29 (7)	4 (4)
*U. longipes*_BY	155,018	85,238	18,206	25,787	37.8	113	80 (6)	29 (7)	4 (4)
*U. glabrum*_KC_M	155,019	85,353	18,184	25,741	37.8	113	80 (6)	29 (7)	4 (4)
U. sp.2_DC_F	155,024	85,303	18,231	25,745	37.7	113	80 (6)	29 (7)	4 (4)
U. sp.2_DC_M	155,041	85,320	18,231	25,745	37.7	113	80 (6)	29 (7)	4 (4)
U. sp.1_DC	155,113	85,305	18,248	25,780	37.7	113	80 (6)	29 (7)	4 (4)
*U. chinense*_HB	155,117	85,305	18,236	25,788	37.7	113	80 (6)	29 (7)	4 (4)
*U. crassum*_TO	155,330	85,442	18,308	25,790	37.7	113	80 (6)	29 (7)	4 (4)
*U. crassum*_KU_F	155,352	85,464	18,308	25,790	37.7	113	80 (6)	29 (7)	4 (4)
*U. crassum*_KU_M	155,352	85,464	18,308	25,790	37.7	113	80 (6)	29 (7)	4 (4)
*U. villosum*_TO	155,367	85,507	18,270	25,795	37.7	113	80 (6)	29 (7)	4 (4)
*U. memecyloides*_SB	155,377	85,503	18,342	25,766	37.7	113	80 (6)	29 (7)	4 (4)
*U. villosum*_BY_F	155,405	85,547	18,268	25,795	37.7	113	80 (6)	29 (7)	4 (4)
*U. villosum*_BY_M	155,405	85,547	18,268	25,795	37.7	113	80 (6)	29 (7)	4 (4)

Table S6a Type and number of SSRs motifs.

Taxa	A	T	C	AT	TA	AAT	ATA	ATT	CAG	TAT	TTA	AAAT	AATA	AATT	ATAA	ATAG	ATTT	CATT	CCTT
*U. villosum*_TO	8	14	1	2	2	0	1	0	0	1	4	0	0	0	0	1	0	1	1
*U. villosum*_BY_F	7	13	1	2	2	0	1	1	0	1	4	0	0	0	0	1	0	1	1
*U. villosum*_BY_M	7	13	1	2	2	0	1	1	0	1	4	0	0	0	0	1	0	1	1
*U. crassum*_TO	8	14	1	2	4	0	1	0	0	1	4	0	1	0	0	1	0	1	1
*U. crassum*_KU_F	8	12	2	2	4	0	1	0	0	1	5	0	1	0	0	1	0	1	1
*U. crassum*_KU_M	8	12	2	2	4	0	1	0	0	1	5	0	1	0	0	1	0	1	1
U. sp.3_BN	9	13	1	2	1	0	0	0	0	1	3	0	1	1	0	0	0	0	0
U. sp.4_NL	11	12	1	2	1	0	0	0	0	1	4	0	1	1	0	1	0	0	0
*U. schmidtii*_KR	9	12	1	2	1	0	0	0	0	1	4	0	1	1	1	1	0	0	0
*U. lecomtei*_DC	11	9	1	2	1	0	0	0	0	1	4	0	1	1	0	1	0	0	0
*U. lecomtei*_NL	10	10	1	2	1	0	0	0	0	1	4	0	1	1	0	1	0	0	0
*U. argenteum*_HN_F	11	10	1	2	0	0	0	0	0	1	4	0	1	1	0	1	0	1	0
*U. argenteum*_HN_M	11	10	1	2	0	0	0	0	0	1	4	0	1	1	0	1	0	1	0
*U. streptopodium*_BY	4	11	2	1	0	0	0	0	1	0	4	0	1	0	0	1	0	1	0
*U. macrophyllum*_HB	3	12	4	2	1	0	1	0	1	1	4	0	1	0	0	1	0	1	0
*U. memecyloides*_SB	8	15	0	3	1	0	0	0	0	0	5	1	1	0	1	1	0	1	0
U. sp.2_DC_F	9	14	0	3	0	0	0	0	0	1	4	0	1	0	0	1	0	1	0
U. sp.2_DC_M	9	14	0	3	0	0	0	0	0	1	4	0	1	0	0	1	0	1	0
U. sp.1_DC	10	15	0	3	1	1	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. chinense*_HN	9	12	0	3	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
U. longifolium var. *annamense*_HN	9	15	2	2	1	0	1	0	0	1	4	0	1	0	1	1	0	1	0
*U. longipes*_BY	10	14	1	2	1	0	0	0	0	1	3	0	1	0	0	1	0	1	0
*U. hirsutum*_BW	6	12	1	2	1	0	0	0	0	1	4	0	1	0	0	1	1	1	0
*U. hirsutum*_HB	6	14	0	2	1	0	0	0	0	1	4	0	1	0	0	1	1	1	0
*U. blumeanum*_BW	6	12	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. glabrum*_KC_F	8	13	1	2	0	0	0	0	0	1	4	0	1	0	0	1	0	1	0

Taxa	A	T	C	AT	TA	AAT	ATA	ATT	CAG	TAT	TTA	AAAT	AATA	AATT	ATAA	ATAG	ATTT	CATT	CCTT
*U. glabrum*_KC_M	8	13	1	2	0	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. glabrum*_SK	11	14	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. glabrum*_TC_F	10	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. glabrum*_TC_M	10	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.A_KK	9	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.C_TN	8	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.C_TC	9	13	2	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.C_KL	9	14	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.C_KP	8	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.C_TT	9	13	2	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.B_NK	9	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.B_KC	8	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0
*U. longifolium*_f.B_KU	9	13	1	2	1	0	0	0	0	1	4	0	1	0	0	1	0	1	0

Table S6b Type and number of SSRs motifs (continued).

Taxa	CTAT	TAAA	TCTA	TCTT	TTAA	TTAT	TTTA	TTTC	TATAT	TATCC	TATTT	AAGACC	ATAGGT	ATATCA
*U. villosum*_TO	0	2	1	0	1	1	0	0	0	0	0	0	0	0
*U. villosum*_BY_F	0	2	1	0	1	1	0	0	0	0	0	0	0	0
*U. villosum*_BY_M	0	2	1	0	1	1	0	0	0	0	0	0	0	0
*U. crassum*_TO	0	1	1	0	1	1	0	0	0	0	0	0	0	0
*U. crassum*_KU_F	0	1	1	0	1	1	0	0	0	0	0	0	0	0
*U. crassum*_KU_M	0	1	1	0	1	1	0	0	0	0	0	0	0	0
U. sp.3_BN	1	1	0	0	0	1	0	0	0	0	1	0	0	0
U. sp.4_NL	1	1	1	0	0	1	0	0	0	0	1	0	0	0
*U. schmidtii*_KR	1	1	1	0	0	1	1	0	0	0	1	0	0	0
*U. lecomtei*_DC	1	1	1	0	0	1	0	0	0	0	1	0	0	0
*U. lecomtei*_NL	1	1	1	0	0	1	0	0	0	0	1	0	0	0
*U. argenteum*_HN_F	1	1	1	0	0	1	0	0	0	0	1	0	0	0
*U. argenteum*_HN_M	1	1	1	0	0	1	0	0	0	0	1	0	0	0
*U. streptopodium*_BY	1	1	1	0	1	1	0	0	0	0	0	1	0	0
*U. macrophyllum*_HB	1	1	1	0	1	1	0	0	0	0	0	0	1	0
*U. memecyloides*_SB	0	1	1	0	1	1	0	0	0	0	0	0	0	0
U. sp.2_DC_F	1	1	1	0	1	1	0	0	1	0	0	0	0	0
U. sp.2_DC_M	1	1	1	0	1	1	0	0	1	0	0	0	0	0
U. sp.1_DC	1	1	1	0	1	1	0	0	1	0	0	0	0	0
*U. chinense*_HN	1	1	1	0	1	1	0	0	1	0	0	0	0	0
U. longifolium var. *annamense*_HN	1	1	1	0	1	0	0	1	0	1	0	0	0	0
*U. longipes*_BY	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. hirsutum*_BW	2	1	1	0	1	0	0	1	0	0	0	0	0	0
*U. hirsutum*_HB	2	1	1	0	1	0	0	1	0	0	0	0	0	0
*U. blumeanum*_BW	1	1	1	1	1	1	0	0	0	0	0	0	0	1
*U. glabrum*_KC_F	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. glabrum*_KC_M	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. glabrum*_SK	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. glabrum*_TC_F	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. glabrum*_TC_M	1	1	1	0	1	1	0	1	0	0	0	0	0	0

Taxa	CTAT	TAAA	TCTA	TCTT	TTAA	TTAT	TTTA	TTTC	TATAT	TATCC	TATTT	AAGACC	ATAGGT	ATATCA
*U. longifolium*_f.A_KK	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.C_TN	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.C_TC	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.C_KL	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.C_KP	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.C_TT	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.B_NK	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.B_KC	1	1	1	0	1	1	0	1	0	0	0	0	0	0
*U. longifolium*_f.B_KU	1	1	1	0	1	1	0	1	0	0	0	0	0	0

Table S6c Type and number of SSRs motifs (continued).

Taxa	ATTTCC	TACCTA	TATACA	TATTGA	TGGTCT	Total	Percentage		
							A	T	C
*U. villosum*_TO	0	0	0	0	0	41	19.5	34.1	2.4
*U. villosum*_BY_F	0	0	0	0	0	40	17.5	32.5	2.5
*U. villosum*_BY_M	0	0	0	0	0	40	17.5	32.5	2.5
*U. crassum*_TO	0	0	0	0	0	43	18.6	32.6	2.3
*U. crassum*_KU_F	0	0	0	0	0	43	18.6	27.9	4.7
*U. crassum*_KU_M	0	0	0	0	0	43	18.6	27.9	4.7
U. sp.3_BN	0	0	0	0	0	36	25.0	36.1	2.8
U. sp.4_NL	0	0	1	0	0	41	26.8	29.3	2.4
*U. schmidtii*_KR	0	0	0	0	0	40	22.5	30.0	2.5
*U. lecomtei*_DC	0	0	0	0	0	37	29.7	24.3	2.7
*U. lecomtei*_NL	0	0	0	0	0	37	27.0	27.0	2.7
*U. argenteum*_HN_F	0	0	0	0	0	38	28.9	26.3	2.6
*U. argenteum*_HN_M	0	0	0	0	0	38	28.9	26.3	2.6
*U. streptopodium*_BY	0	0	0	0	1	33	12.1	33.3	6.1
*U. macrophyllum*_HB	0	1	0	0	0	39	7.7	30.8	10.3
*U. memecyloides*_SB	0	0	0	0	0	41	19.5	36.6	0.0
U. sp.2_DC_F	0	0	0	0	0	40	22.5	35.0	0.0
U. sp.2_DC_M	0	0	0	0	0	40	22.5	35.0	0.0
U. sp.1_DC	0	0	0	0	0	44	22.7	34.1	0.0
*U. chinense*_HN	0	0	0	0	0	39	23.1	30.8	0.0
U. longifolium var. *annamense*_HN	0	0	0	0	0	45	20.0	33.3	4.4
*U. longipes*_BY	0	0	0	0	0	41	24.4	34.1	2.4
*U. hirsutum*_BW	1	0	0	0	0	38	15.8	31.6	2.6
*U. hirsutum*_HB	0	0	0	0	0	38	15.8	36.8	0.0
*U. blumeanum*_BW	0	0	0	1	0	38	15.8	31.6	2.6
*U. glabrum*_KC_F	0	0	0	0	0	38	21.1	34.2	2.6

Taxa	ATTTCC	TACCTA	TATACA	TATTGA	TGGTCT	Total	Percentage		
							A	T	C
*U. glabrum*_KC_M	0	0	0	0	0	38	21.1	34.2	2.6
*U. glabrum*_SK	0	0	0	0	0	43	25.6	32.6	2.3
*U. glabrum*_TC_F	0	0	0	0	0	41	24.4	31.7	2.4
*U. glabrum*_TC_M	0	0	0	0	0	41	24.4	31.7	2.4
*U. longifolium*_f.A_KK	0	0	0	0	0	40	22.5	32.5	2.5
*U. longifolium*_f.C_TN	0	0	0	0	0	39	20.5	33.3	2.6
*U. longifolium*_f.C_TC	0	0	0	0	0	41	22.0	31.7	4.9
*U. longifolium*_f.C_KL	0	0	0	0	0	41	22.0	34.1	2.4
*U. longifolium*_f.C_KP	0	0	0	0	0	39	20.5	33.3	2.6
*U. longifolium*_f.C_TT	0	0	0	0	0	41	22.0	31.7	4.9
*U. longifolium*_f.B_NK	0	0	0	0	0	40	22.5	32.5	2.5
*U. longifolium*_f.B_KC	0	0	0	0	0	39	20.5	33.3	2.6
*U. longifolium*_f.B_KU	0	0	0	0	0	40	22.5	32.5	2.5

Table S7 Number of SSRs at each region in 39 *Urophyllum* plastomes.

Taxa	LSC	SSC	IR	total	% in LSC	% in SSC	% in IR
*U. villosum*_TO	31	6	2	39	79.5	15.4	5.1
*U. villosum*_BY_F	31	5	2	38	81.6	13.2	5.3
*U. villosum*_BY_M	31	5	2	38	81.6	13.2	5.3
*U. crassum*_TO	31	8	2	41	75.6	19.5	4.9
*U. crassum*_KU_F	33	6	2	41	80.5	14.6	4.9
*U. crassum*_KU_M	33	6	2	41	80.5	14.6	4.9
U. sp.3_BN	28	8	0	36	77.8	22.2	0.0
U. sp.4_NL	31	8	1	40	77.5	20.0	2.5
*U. schmidtii*_KR	28	10	1	39	71.8	25.6	2.6
*U. lecomtei*_DC	28	7	1	36	77.8	19.4	2.8
*U. lecomtei*_NL	29	6	1	36	80.6	16.7	2.8
*U. argenteum*_HN_F	29	7	1	37	78.4	18.9	2.7
*U. argenteum*_HN_M	29	7	1	37	78.4	18.9	2.7
*U. streptopodium*_BY	25	6	1	32	78.1	18.8	3.1
*U. macrophyllum*_HB	29	6	2	37	78.4	16.2	5.4
*U. memecyloides*_SB	31	8	1	40	77.5	20.0	2.5
U. sp.2_DC_F	31	7	1	39	79.5	17.9	2.6
U. sp.2_DC_M	31	7	1	39	79.5	17.9	2.6
U. sp.1_DC	34	8	1	43	79.1	18.6	2.3
*U. chinense*_HN	31	6	1	38	81.6	15.8	2.6
U. longifolium var. *annamense*_HN	35	8	1	44	79.5	18.2	2.3
*U. longipes*_BY	32	7	1	40	80.0	17.5	2.5
*U. hirsutum*_BW	30	6	1	37	81.1	16.2	2.7
*U. hirsutum*_HB	29	7	1	37	78.4	18.9	2.7
*U. blumeanum*_BW	30	4	2	36	83.3	11.1	5.6
*U. glabrum*_KC_F	28	8	1	37	75.7	21.6	2.7
*U. glabrum*_KC_M	28	8	1	37	75.7	21.6	2.7
*U. glabrum*_SK	33	8	1	42	78.6	19.0	2.4
*U. glabrum*_TC_F	32	7	1	40	80.0	17.5	2.5
*U. glabrum*_TC_M	32	7	1	40	80.0	17.5	2.5
*U. longifolium*_f.A_KK	31	7	1	39	79.5	17.9	2.6
*U. longifolium*_f.C_TN	32	7	1	40	80.0	17.5	2.5
*U. longifolium*_f.C_TC	30	7	1	38	78.9	18.4	2.6
*U. longifolium*_f.C_KL	31	8	1	40	77.5	20.0	2.5
*U. longifolium*_f.C_KP	30	7	1	38	78.9	18.4	2.6
*U. longifolium*_f.C_TT	32	7	1	40	80.0	17.5	2.5
*U. longifolium*_f.B_NK	31	7	1	39	79.5	17.9	2.6
*U. longifolium*_f.B_KC	30	7	1	38	78.9	18.4	2.6
*U. longifolium*_f.B_KU	31	7	1	39	79.5	17.9	2.6

Table S8 Number of SSRs types in different repeat lengths found in 39 *Urophyllum* plastomes. On the top row, letters indicate SSR type; numbers indicate repeat length.

Taxa	**A 10**	**A 11**	**A 12**	**A 13**	**A 14**	**A 15**	**T 10**	**T 11**	**T 12**	**T 13**	**T 14**	**T 15**	**AT 5**	**AT 6**	**AT 7**	**TTA 4**	**TTA 5**	**TTA 6**	**TAAA 3**	**TAAA 4**	**TAAA 5**
*U. villosum*_TO	7	0	1	0	0	0	6	5	1	0	1	1	1	0	1	4	0	0	2	0	0
*U. villosum*_BY_F	6	0	1	0	0	0	7	3	0	2	0	1	1	0	1	4	0	0	2	0	0
*U. villosum*_BY_M	6	0	1	0	0	0	7	3	0	2	0	1	1	0	1	4	0	0	2	0	0
*U. crassum*_TO	6	2	0	0	0	0	8	5	0	0	0	1	1	1	0	4	0	0	1	0	0
*U. crassum*_KU_F	6	0	1	1	0	0	7	4	0	0	0	1	2	0	0	5	0	0	1	0	0
*U. crassum*_KU_M	6	0	1	1	0	0	7	4	0	0	0	1	2	0	0	5	0	0	1	0	0
U. sp.3_BN	2	3	1	3	0	0	6	4	1	1	0	1	1	0	1	3	0	0	0	1	0
U. sp.4_NL	6	3	2	0	0	0	7	2	0	2	0	1	1	0	1	3	0	1	0	1	0
*U. schmidtii*_KR	5	1	2	1	0	0	6	4	1	0	0	1	1	0	1	4	0	0	0	1	0
*U. lecomtei*_DC	7	2	2	0	0	0	3	4	0	1	0	1	1	0	1	4	0	0	0	1	0
*U. lecomtei*_NL	3	3	3	0	1	0	4	4	1	0	0	1	1	0	1	4	0	0	0	1	0
*U. argenteum*_HN_F	4	3	3	0	1	0	4	4	0	1	0	1	1	1	0	1	0	0	0	0	1
*U. argenteum*_HN_M	4	3	3	0	1	0	4	4	0	1	0	1	1	1	0	1	0	0	0	0	1
*U. streptopodium*_BY	3	0	0	1	0	0	4	4	2	0	0	1	1	0	0	4	0	0	0	1	0
*U. macrophyllum*_HB	2	0	0	0	1	0	7	3	0	1	0	1	1	1	0	4	0	0	0	1	0
*U. memecyloides*_SB	3	3	0	2	0	0	6	4	2	1	1	1	2	1	0	4	1	0	0	1	0
U. sp.2_DC_F	2	4	1	2	0	0	7	4	1	0	1	1	3	0	0	4	0	0	1	0	0
U. sp.2_DC_M	2	4	1	2	0	0	7	4	1	0	1	1	3	0	0	4	0	0	1	0	0
U. sp.1_DC	3	4	2	1	0	0	7	5	1	0	1	1	3	0	0	4	0	0	1	0	0
*U. chinense*_HN	2	3	3	1	0	0	5	4	1	0	1	1	3	0	0	4	0	0	1	0	0
U. longifolium var. *annamense*_HN	4	2	2	1	0	0	10	3	1	0	0	1	1	0	1	4	0	0	0	1	0
*U. longipes*_BY	4	2	3	1	0	0	4	8	1	0	0	1	1	0	1	3	0	0	0	1	0
*U. hirsutum*_BW	1	3	2	0	0	0	6	3	1	1	0	1	1	0	1	4	0	0	0	1	0

Taxa	**A 10**	**A 11**	**A 12**	**A 13**	**A 14**	**A 15**	**T 10**	**T 11**	**T 12**	**T 13**	**T 14**	**T 15**	**AT 5**	**AT 6**	**AT 7**	**TTA 4**	**TTA 5**	**TTA 6**	**TAAA 3**	**TAAA 4**	**TAAA 5**
*U. hirsutum*_HB	1	2	3	0	0	0	7	4	1	1	0	1	1	0	1	4	0	0	0	1	0
*U. blumeanum*_BW	2	0	4	0	0	0	3	7	1	0	0	1	1	1	0	1	0	0	0	1	0
*U. glabrum*_KC_F	2	3	3	0	0	0	5	2	3	2	0	1	1	0	1	3	0	1	0	1	0
*U. glabrum*_KC_M	2	3	3	0	0	0	5	2	3	2	0	1	1	0	1	3	0	1	0	1	0
*U. glabrum*_SK	6	2	2	1	0	0	7	1	3	2	0	1	1	0	1	4	0	0	0	1	0
*U. glabrum*_TC_F	3	3	3	1	0	0	5	4	1	2	0	1	1	0	1	3	0	1	0	1	0
*U. glabrum*_TC_M	3	3	3	1	0	0	5	4	1	2	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.A_KK	3	3	3	0	0	0	6	2	3	0	1	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.C_TN	2	4	3	0	0	0	5	4	1	2	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.C_TC	0	3	5	0	0	0	6	2	4	0	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.C_KL	2	3	4	0	0	0	7	3	1	2	0	1	1	0	1	4	0	0	0	1	0
*U. longifolium*_f.C_KP	3	1	4	0	0	0	6	2	2	2	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.C_TT	3	2	4	0	0	0	6	3	0	3	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.B_NK	3	3	3	0	0	0	6	3	1	2	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.B_KC	1	3	3	1	0	0	6	3	1	2	0	1	1	0	1	3	0	1	0	1	0
*U. longifolium*_f.B_KU	2	2	4	1	0	0	6	3	1	2	0	1	1	0	1	3	0	1	0	1	0

Table S9 Number of long repeats (≥30 bp) in different locations within plastomes.

Species	CDS	Non-CDS	Total repeats	Percentage of CDS	Percentage of non-CDS
*U. villosum*_TO	4	20	24	16.7	83.3
*U. villosum*_BY_F	4	23	27	14.8	85.2
*U. villosum*_BY_M	4	23	27	14.8	85.2
*U. crassum*_TO	4	19	24	16.7	79.2
*U. crassum*_KU_F	4	20	26	15.4	76.9
*U. crassum*_KU_M	4	20	26	15.4	76.9
U. sp.3_BN	4	19	23	17.4	82.6
U. sp.4_NL	4	18	22	18.2	81.8
*U. schmidtii*_KR	5	16	21	23.8	76.2
*U. lecomtei*_DC	5	17	22	22.7	77.3
*U. lecomtei*_NL	5	16	21	23.8	76.2
*U. argenteum*_HN_F	5	20	25	20.0	80.0
*U. argenteum*_HN_M	5	20	25	20.0	80.0
*U. streptopodium*_BY	5	18	23	21.7	78.3
*U. macrophyllum*_HB	4	21	25	16.0	84.0
*U. memecyloides*_SB	4	23	27	14.8	85.2
U. sp.2_DC_F	5	18	23	21.7	78.3
U. sp.2_DC_M	5	18	23	21.7	78.3
U. sp.1_DC	5	19	24	20.8	79.2
*U. chinense*_HN	5	21	26	19.2	80.8
U. longifolium var. *annamense*_HN	5	20	25	20.0	80.0
*U. longipes*_BY	5	20	25	20.0	80.0
*U. hirsutum*_BW	6	21	27	22.2	77.8
*U. hirsutum*_HB	6	20	26	23.1	76.9
*U. blumeanum*_BW	6	20	26	23.1	76.9
*U. glabrum*_KC_F	6	20	26	23.1	76.9
*U. glabrum*_KC_M	6	20	26	23.1	76.9
*U. glabrum*_SK	6	22	28	21.4	78.6
*U. glabrum*_TC_F	6	20	26	23.1	76.9
*U. glabrum*_TC_M	6	20	26	23.1	76.9
*U. longifolium*_f.A_KK	6	21	27	22.2	77.8
*U. longifolium*_f.C_TN	6	22	28	21.4	78.6
*U. longifolium*_f.C_TC	6	23	29	20.7	79.3
*U. longifolium*_f.C_KL	6	20	26	23.1	76.9
*U. longifolium*_f.C_KP	6	20	26	23.1	76.9
*U. longifolium*_f.C_TT	6	21	27	22.2	77.8
*U. longifolium*_f.B_NK	6	20	26	23.1	76.9
*U. longifolium*_f.B_KC	6	19	25	24.0	76.0
*U. longifolium*_f.B_KU	6	20	26	23.1	76.9

Table S10a Locations (CDS and non-CDS) that long repeats found in plastomes of different *Urophyllum* species.

Taxa	CDS						Non-CDS							
	psaB-psaA	*accD*	*rps18*	*ycf2*	*ycf1*	*ndhF*	*ycf1(CDS)-ycf1–ndhF*	*trnS(-GCU)-trnS(-GGA)*	*trnS(-GCU)-trnS(-UGA)*	*trnS(-UGA)-trnS(-GGA)*	*trnG(-UCC)-trnG(-GCC)*	*ycf3-ndhA*	*rpl16*	*ndhA*
*U. villosum*_TO	-	1	1	1	1	-	-	1	1	1	1	1	-	2
*U. villosum*_BY_F	-	1	1	1	1	-	-	1	1	1	1	1	-	2
*U. villosum*_BY_M	-	1	1	1	1	-	-	1	1	1	1	1	-	2
*U. crassum*_TO	1	1	-	1	1	-	1	1	1	1	1	1	-	1
*U. crassum*_KU_F	1	1	-	1	1	-	1	1	1	1	1	1	-	1
*U. crassum*_KU_M	1	1	-	1	1	-	1	1	1	1	1	1	-	1
U. sp.3_BN	1	1	-	1	1	-	-	1	1	1	-	1	-	2
U. sp.4_NL	1	1	-	1	1	-	-	1	1	1	-	1	-	2
*U. schmidtii*_KR	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. lecomtei*_DC	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. lecomtei*_NL	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. argenteum*_HN_F	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. argenteum*_HN_M	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. streptopodium*_BY	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. macrophyllum*_HB	1	1	-	1	1	-	-	1	1	1	-	1	-	3
*U. memecyloides*_SB	1	1	-	1	1	-	-	1	1	1	-	1	-	-
U. sp.2_DC_F	1	1	-	1	1	1	-	1	1	1	-	1	1	-
U. sp.2_DC_M	1	1	-	1	1	1	-	1	1	1	-	1	1	-
U. sp.1_DC	1	1	-	1	1	1	-	1	1	1	-	1	1	-
*U. chinense*_HN	1	1	-	1	1	1	-	1	1	1	-	1	1	2
U. longifolium var. *annamense*_HN	1	1	-	1	1	1	-	1	1	1	-	1	-	2
*U. longipes*_BY	1	1	-	1	1	1	-	1	1	1	-	1	-	2

Taxa	CDS						Non-CDS							
	psaB-psaA	*accD*	*rps18*	*ycf2*	*ycf1*	*ndhF*	*ycf1(CDS)-ycf1–ndhF*	*trnS(-GCU)-trnS(-GGA)*	*trnS(-GCU)-trnS(-UGA)*	*trnS(-UGA)-trnS(-GGA)*	*trnG(-UCC)-trnG(-GCC)*	*ycf3-ndhA*	*rpl16*	*ndhA*
*U. hirsutum*_BW	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. hirsutum*_HB	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. blumeanum*_BW	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. glabrum*_KC_F	1	1	-	1	2	1	-	1	1	1	-	1	-	1
*U. glabrum*_KC_M	1	1	-	1	2	1	-	1	1	1	-	1	-	1
*U. glabrum*_SK	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. glabrum*_TC_F	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. glabrum*_TC_M	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.A_KK	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.C_TN	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.C_TC	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.C_KL	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.C_KP	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.C_TT	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.B_NK	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.B_KC	1	1	-	1	2	1	-	1	1	1	-	1	-	2
*U. longifolium*_f.B_KU	1	1	-	1	2	1	-	1	1	1	-	1	-	2

Table S10b Locations (12 non-CDS) that long repeats found in plastomes of different *Urophyllum* species.

Taxa	Non-CDS											
	1)trnH(-GUG)-psbA-2)trnT(-GGU)-psbD	*trnH(-GUG)-psbA*	*matK-rps16*	*rps16-trnQ(-UUG)*	*trnS(-GCU)-trnG(-UCC)*	*atpI-rps2*	*rpoB-trnC(-GCA)*	*rpoB-trnC(-GCA) & psaA-ycf3*	*petN-psbM*	*trnE(-UUC)-trnT(-GGU)*	*trnT(-GGU)-psbD*	*psaA-ycf3*
*U. villosum*_TO	-	-	-	1	-	-	1	1	1	-	-	1
*U. villosum*_BY_F	-	-	-	1	-	-	1	1	1	-	-	1
*U. villosum*_BY_M	-	-	-	1	-	-	1	1	1	-	-	1
*U. crassum*_TO	-	-	-	1	-	-	-	-	1	1	1	1
*U. crassum*_KU_F	-	-	1	2	-	-	-	-	1	1	1	1
*U. crassum*_KU_M	-	-	1	2	-	-	-	-	1	1	1	1
U. sp.3_BN	-	1	-	-	-	-	-	-	1	-	-	-
U. sp.4_NL	-	1	-	-	2	-	-	-	1	-	-	-
*U. schmidtii*_KR	-	1	-	-	-	-	-	-	1	-	-	-
*U. lecomtei*_DC	-	1	-	1	-	-	-	-	1	-	-	-
*U. lecomtei*_NL	-	1	-	-	-	-	-	-	1	-	-	-
*U. argenteum*_HN_F	-	1	-	-	1	1	-	-	1	-	-	-
*U. argenteum*_HN_M	-	1	-	-	1	1	-	-	1	-	-	-
*U. streptopodium*_BY	-	-	-	2	-	-	-	-	1	-	-	-
*U. macrophyllum*_HB	-	1	-	2	-	-	-	-	1	-	-	-
*U. memecyloides*_SB	-	1	-	3	-	-	-	-	1	-	-	-
U. sp.2_DC_F	-	1	-	1	-	-	-	-	1	-	-	-
U. sp.2_DC_M	-	1	-	1	-	-	-	-	1	-	-	-
U. sp.1_DC	-	1	-	1	-	-	-	-	1	-	-	-
*U. chinense*_HN	-	1	-	1	-	-	-	-	1	-	-	-
U. longifolium var. *annamense*_HN	-	1	-	1	-	-	-	-	n	-	-	-

Taxa	Non-CDS											
	1)trnH(-GUG)-psbA-2)trnT(-GGU)-psbD	*trnH(-GUG)-psbA*	*matK-rps16*	*rps16-trnQ(-UUG)*	*trnS(-GCU)-trnG(-UCC)*	*atpI-rps2*	*rpoB-trnC(-GCA)*	*rpoB-trnC(-GCA) & psaA-ycf3*	*petN-psbM*	*trnE(-UUC)-trnT(-GGU)*	*trnT(-GGU)-psbD*	*psaA-ycf3*
*U. longipes*_BY	-	1	-	2	-	-	-	-	1	-	-	-
*U. hirsutum*_BW	-	1	-	1	-	-	-	-	1	-	1	-
*U. hirsutum*_HB	-	1	-	1	-	-	-	-	1	-	-	-
*U. blumeanum*_BW	-	1	-	1	-	-	-	-	1	-	-	-
*U. glabrum*_KC_F	-	1	-	1	-	-	-	-	1	-	-	-
*U. glabrum*_KC_M	-	1	-	1	-	-	-	-	1	-	-	-
*U. glabrum*_SK	1	1	-	2	-	-	-	-	1	-	-	-
*U. glabrum*_TC_F	-	1	-	1	-	-	-	-	1	-	-	-
*U. glabrum*_TC_M	-	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.A_KK	1	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.C_TN	1	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.C_TC	1	1	-	1	2	-	-	-	1	-	-	-
*U. longifolium*_f.C_KL	-	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.C_KP	-	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.C_TT	-	1	-	2	-	-	-	-	1	-	-	-
*U. longifolium*_f.B_NK	-	1	-	1	-	-	-	-	1	-	-	-
*U. longifolium*_f.B_KC	-	1	-	-	-	-	-	-	1	-	-	-
*U. longifolium*_f.B_KU	-	1	-	1	-	-	-	-	1	-	-	-

Table S10c Locations (12 non-CDS) that long repeats found in plastomes of different *Urophyllum* species.

Taxa	**Non-CDS**											
	rps4-trnT(-UGU)	*ndhC-trnV(-UAC)*	*accD-psaI*	*ycf4-cemA*	*petA-psbJ*	*psbE-petL*	*clpP-psbB*	*psbT-psbN*	*petD-rpoA*	*rps12-trnV(-GAC)*	*rrn4.5-rrn5*	*psaC-ndhD*
*U. villosum*_TO	-	-	1	-	-	-	-	1	-	1	1	1
*U. villosum*_BY_F	-	-	2	-	1	1	-	1	-	1	1	1
*U. villosum*_BY_M	-	-	2	-	1	1	-	1	-	1	1	1
*U. crassum*_TO	-	-	-	-	-	-	-	1	-	1	1	1
*U. crassum*_KU_F	-	-	-	-	-	-	-	1	-	1	1	1
*U. crassum*_KU_M	-	-	-	-	-	-	-	1	-	1	1	1
U. sp.3_BN	-	1	1	-	-	-	-	1	1	-	1	1
U. sp.4_NL	-	n	1	-	-	-	-	1	1	-	1	1
*U. schmidtii*_KR	-	-	1	-	-	-	-	1	1	-	1	1
*U. lecomtei*_DC	-	-	1	-	-	-	-	1	1	-	1	1
*U. lecomtei*_NL	-	-	1	-	-	-	-	1	1	-	1	1
*U. argenteum*_HN_F	-	-	1	-	-	-	1	1	1	1	1	1
*U. argenteum*_HN_M	-	-	1	-	-	-	1	1	1	1	1	1
*U. streptopodium*_BY	-	-	1	-	-	-	-	1	-	1	1	1
*U. macrophyllum*_HB	-	-	1	-	-	-	-	1	1	1	1	1
*U. memecyloides*_SB	-	1	1	1	-	-	-	1	1	1	1	1
U. sp.2_DC_F	-	-	-	-	-	-	-	1	1	1	1	1
U. sp.2_DC_M	-	-	-	-	-	-	-	1	1	1	1	1
U. sp.1_DC	-	-	1	-	-	-	-	1	1	1	1	1
*U. chinense*_HN	-	-	1	-	-	-	-	1	1	1	1	1
U. longifolium var. *annamense*_HN	-	1	1	-	-	-	-	1	1	1	1	1
*U. longipes*_BY	-	-	1	-	-	-	-	1	1	1	1	1
*U. hirsutum*_BW	-	1	1	-	-	-	-	1	1	1	1	1

Taxa	**Non-CDS**											
	rps4-trnT(-UGU)	*ndhC-trnV(-UAC)*	*accD-psaI*	*ycf4-cemA*	*petA-psbJ*	*psbE-petL*	*clpP-psbB*	*psbT-psbN*	*petD-rpoA*	*rps12-trnV(-GAC)*	*rrn4.5-rrn5*	*psaC-ndhD*
*U. hirsutum*_HB	-	1	1	-	-	-	-	1	1	1	1	1
*U. blumeanum*_BW	-	1	1	-	-	-	-	1	1	2	1	1
*U. glabrum*_KC_F	-	1	2	-	-	-	-	1	1	1	1	1
*U. glabrum*_KC_M	-	1	2	-	-	-	-	1	1	1	1	1
*U. glabrum*_SK	-	1	1	-	-	-	-	1	1	1	1	1
*U. glabrum*_TC_F	-	1	1	-	-	-	-	1	1	1	1	1
*U. glabrum*_TC_M	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.A_KK	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.C_TN	1	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.C_TC	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.C_KL	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.C_KP	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.C_TT	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.B_NK	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.B_KC	-	1	1	-	-	-	-	1	1	1	1	1
*U. longifolium*_f.B_KU	-	1	1	-	-	-	-	1	1	1	1	1

Table S10d Locations (six non-CDS) that long repeats found in plastomes of different *Urophyllum* species.

Taxa	Non-CDS					
	ndhD-ccsA	ccsA-trnL(UAG)	trnL(UAG)-rpl32	ycf3(int.)-rps12-trnV(-GAC)	ndhC-trnV(-UAC)-ndhA(int.)	rps12-trnV(-GAC)-ndhA(int.)
*U. villosum*_TO	1	-	-	1	-	1
*U. villosum*_BY_F	1	-	-	1	-	1
*U. villosum*_BY_M	1	-	-	1	-	1
*U. crassum*_TO	1	-	-	1	1	1
*U. crassum*_KU_F	1	-	-	1	1	1
*U. crassum*_KU_M	1	-	-	1	1	1
U. sp.3_BN	1	-	2	1	-	1
U. sp.4_NL	1	-	-	1	-	1
*U. schmidtii*_KR	1	-	-	1	-	1
*U. lecomtei*_DC	1	-	-	1	-	1
*U. lecomtei*_NL	1	-	-	1	-	1
*U. argenteum*_HN_F	1	-	-	1	-	1
*U. argenteum*_HN_M	1	-	-	1	-	1
*U. streptopodium*_BY	1	-	-	2	-	1
*U. macrophyllum*_HB	1	-	-	2	-	1
*U. memecyloides*_SB	1	1	1	2	-	1
U. sp.2_DC_F	1	-	1	2	-	1
U. sp.2_DC_M	1	-	1	2	-	1
U. sp.1_DC	1	-	1	2	-	1
*U. chinense*_HN	1	-	1	2	-	1
U. longifolium var. *annamense*_HN	1	-	1	2	-	1
*U. longipes*_BY	1	-	-	2	-	1
*U. hirsutum*_BW	1	-	-	2	-	1

Taxa	Non-CDS					
	ndhD-ccsA	ccsA-trnL(UAG)	trnL(UAG)-rpl32	ycf3(int.)-rps12-trnV(-GAC)	ndhC-trnV(-UAC)-ndhA(int.)	rps12-trnV(-GAC)-ndhA(int.)
*U. hirsutum*_HB	1	-	-	2	-	1
*U. blumeanum*_BW	1	-	-	2	-	-
*U. glabrum*_KC_F	1	-	-	2	-	1
*U. glabrum*_KC_M	1	-	-	2	-	1
*U. glabrum*_SK	1	-	-	2	-	1
*U. glabrum*_TC_F	1	-	-	2	-	1
*U. glabrum*_TC_M	1	-	-	2	-	1
*U. longifolium*_f.A_KK	1	-	-	2	-	1
*U. longifolium*_f.C_TN	1	-	-	2	-	1
*U. longifolium*_f.C_TC	1	-	-	2	-	1
*U. longifolium*_f.C_KL	1	-	-	2	-	1
*U. longifolium*_f.C_KP	1	-	-	2	-	1
*U. longifolium*_f.C_TT	1	-	-	2	-	1
*U. longifolium*_f.B_NK	1	-	-	2	-	1
*U. longifolium*_f.B_KC	1	-	-	2	-	1
*U. longifolium*_f.B_KU	1	-	-	2	-	1

Table S11 Number of nucleotide substitutions (top-right triangle) and pairwise distance (lower-left triangle) in plastid genomes of four closely related *Urophyllum* taxa. Species code refer to: 1) first two letters: GL = *U. glabrum*, LA = *U. longifolium*_f.A, LB = *U. longifolium*_f.B, LC = *U. longifolium*_f.C; 2) next two letters indicate locality as refers in Table 3.1; 3) last one letter indicates sex (if presence).

Species code	GL_KC_F	GL_KC_M	GL_SK	GL_TC_F	GL_TC_M	LB_KU	LB_KC	LB_NK	LA_KK	LC_TT	LC_TC	LC_KL	LC_KP	LC_TN
GL_KC_F		6	205	211	211	197	212	269	197	197	220	228	211	239
GL_KC_M	0.0000		199	205	205	191	206	263	191	191	214	222	205	233
GL_SK	0.0005	0.0005		216	216	216	224	278	220	218	229	225	236	246
GL_TC_F	0.0004	0.0004	0.0005		0	40	42	100	50	60	55	57	65	70
GL_TC_M	0.0004	0.0004	0.0005	0.0000		40	42	100	50	60	55	57	65	70
LB_KU	0.0003	0.0003	0.0006	0.0002	0.0002		40	100	32	46	53	57	45	70
LB_KC	0.0005	0.0005	0.0006	0.0002	0.0002	0.0002		98	50	58	51	55	61	70
LB_NK	0.0004	0.0004	0.0005	0.0001	0.0001	0.0002	0.0002		110	118	113	115	123	128
LA_KK	0.0004	0.0004	0.0006	0.0002	0.0002	0.0001	0.0003	0.0002		54	59	65	50	80
LC_TT	0.0004	0.0004	0.0006	0.0002	0.0002	0.0001	0.0002	0.0001	0.0002		73	75	68	90
LC_TC	0.0004	0.0004	0.0005	0.0001	0.0001	0.0002	0.0002	0.0001	0.0002	0.0002		60	64	79
LC_KL	0.0004	0.0004	0.0006	0.0001	0.0001	0.0002	0.0002	0.0001	0.0002	0.0002	0.0001		60	69
LC_KP	0.0004	0.0004	0.0006	0.0003	0.0003	0.0002	0.0003	0.0003	0.0002	0.0002	0.0002	0.0003		77
LC_TN	0.0005	0.0005	0.0006	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0002	0.0001	0.0002	0.0003	

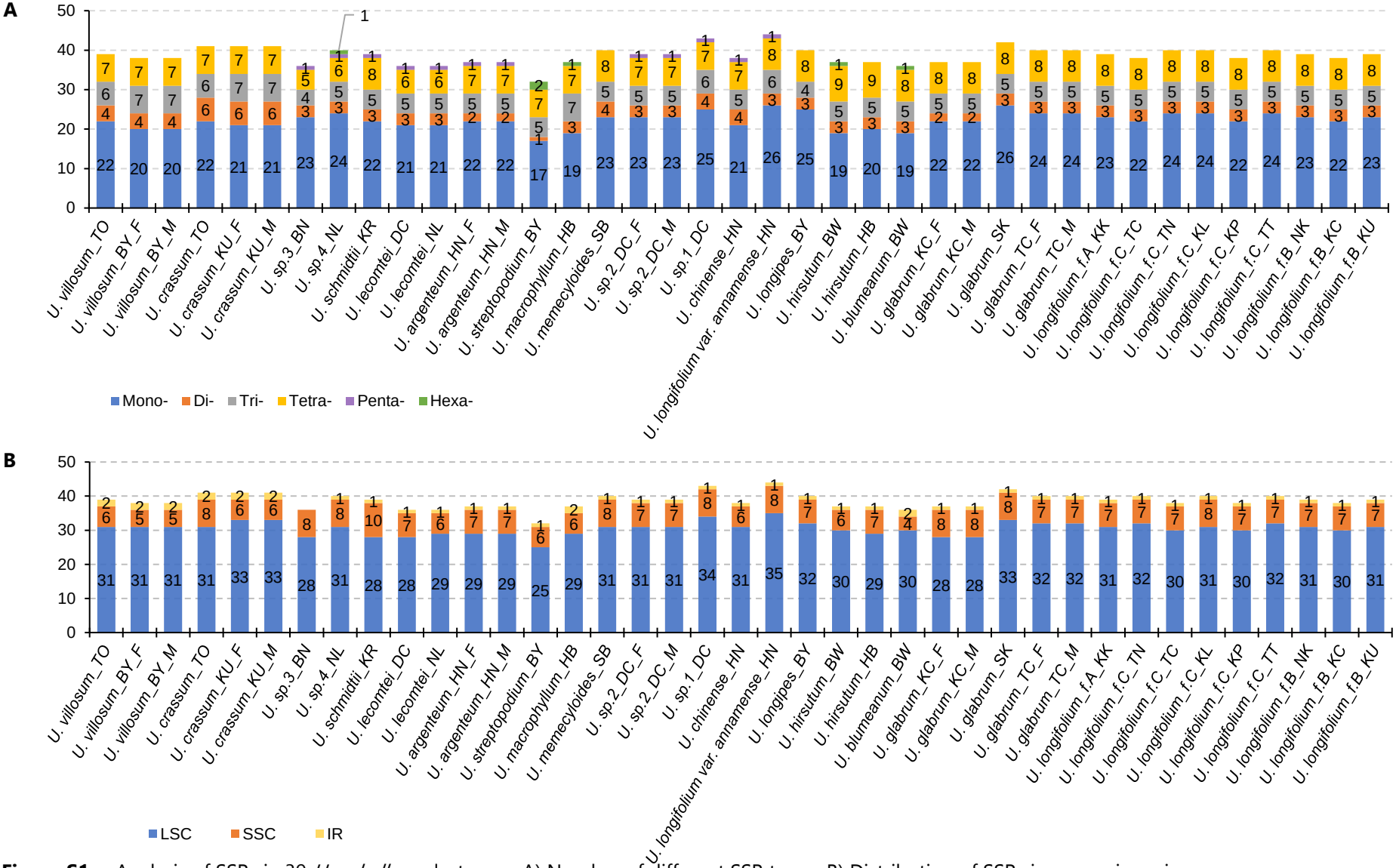


Figure S1 Analysis of SSRs in 39 *Urophyllum* plastomes. A) Number of different SSR types. B) Distribution of SSRs in genomic regions.

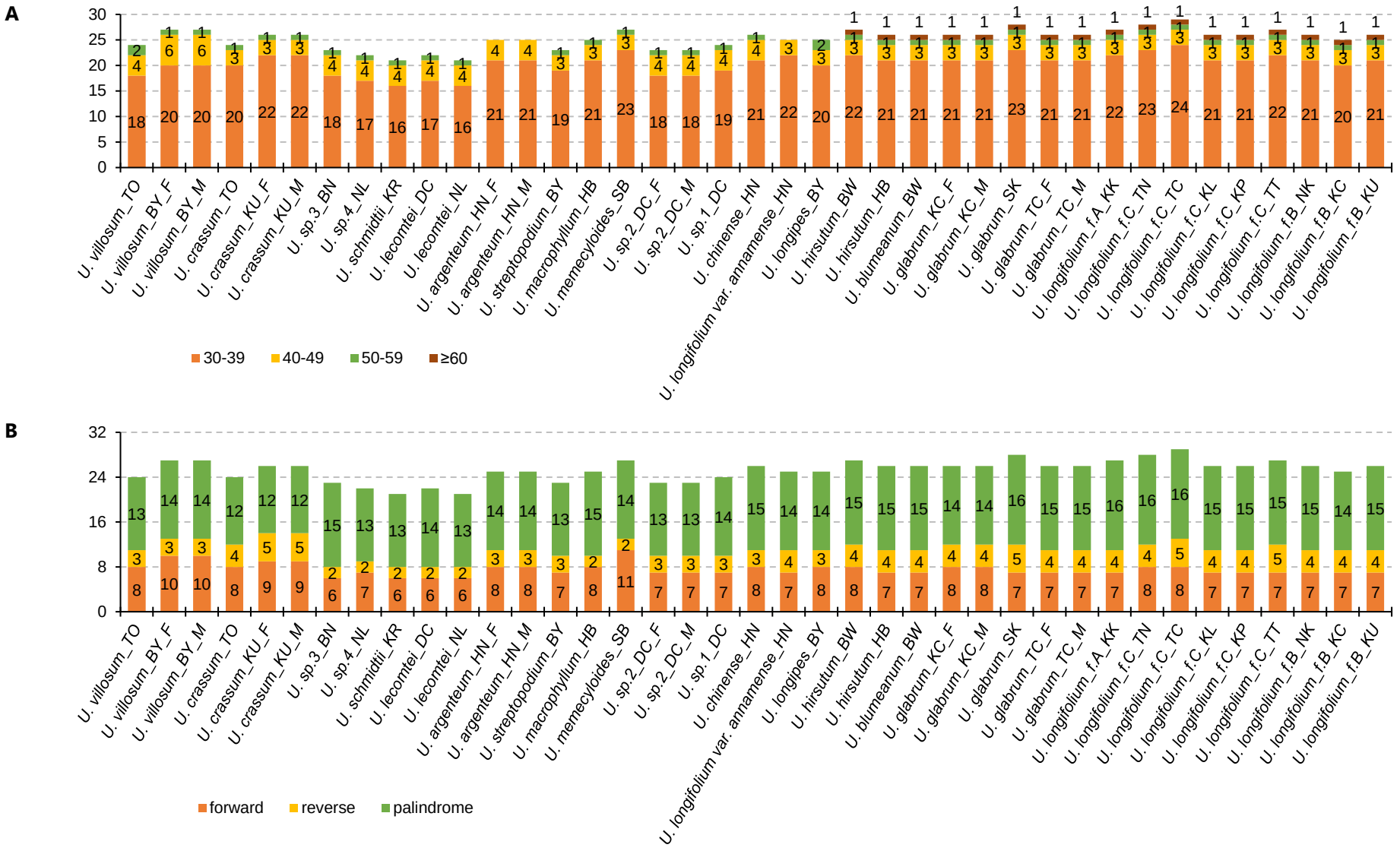


Figure S2 Number of long repeats (≥30 bp) found in 39 *Urophyllum* plastomes. A) Different length. B) Different repeat types.

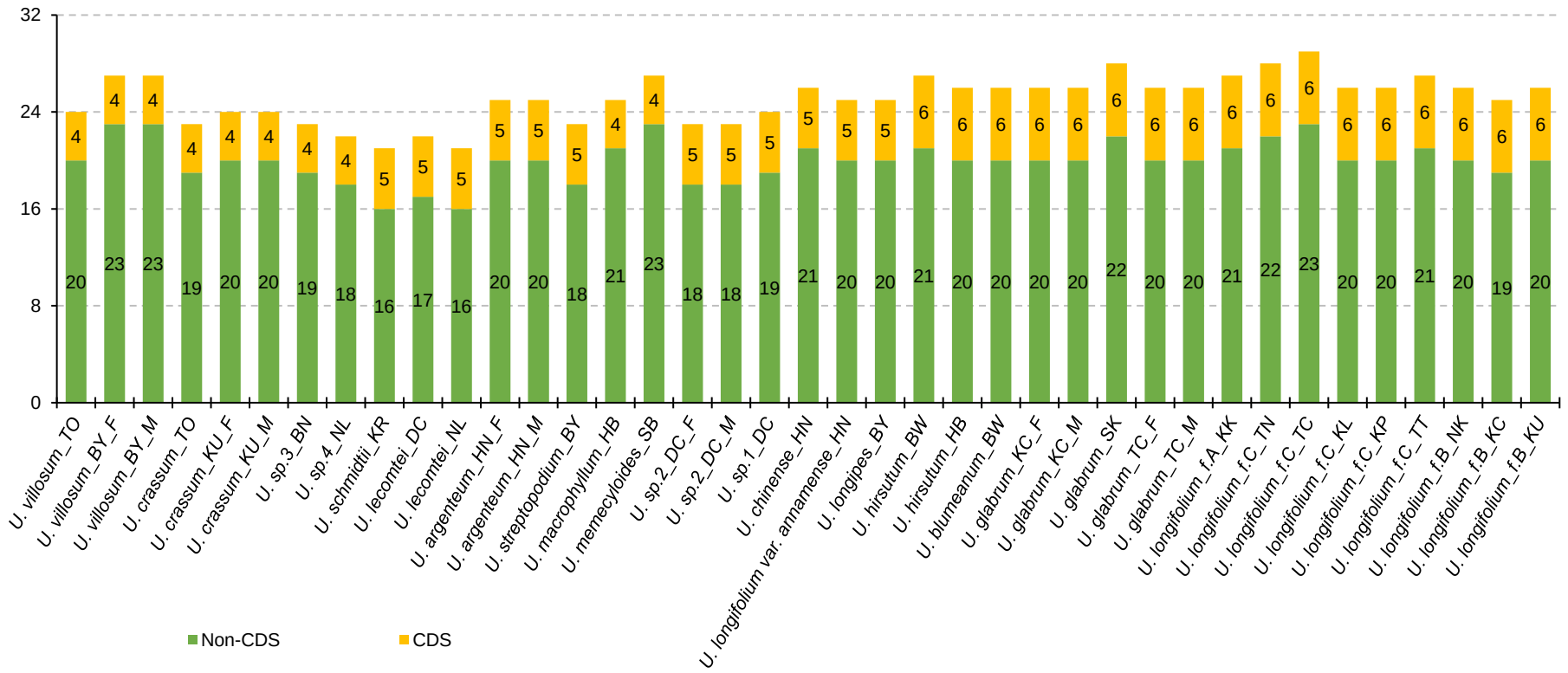


Figure S3 Locations of repeats found in 39 *Urophyllum* plastomes. CDS indicates protein-coding region. IGS indicates intergenic spacer. Note: non-CDS includes RNA, introns, IGS and IGS-introns.