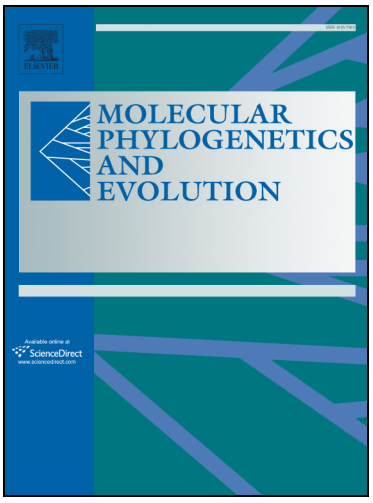


Multilocus coalescent species delimitation to evaluate traditionally defined morphotypes in *Hydrangea* sect. *Asperae* (Hydrangeaceae)

Yannick De Smet, Olivier De Clerck, Tatsuya Uemachi, Carolina Granados Mendoza, Stefan Wanke, Paul Goetghebeur, Marie-Stéphanie Samain

PII: S1055-7903(16)30394-3
DOI: http://dx.doi.org/10.1016/j.ympev.2017.05.021
Reference: YMPEV 5834

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 2 December 2016
Revised Date: 23 May 2017
Accepted Date: 23 May 2017

Please cite this article as: De Smet, Y., De Clerck, O., Uemachi, T., Granados Mendoza, C., Wanke, S., Goetghebeur, P., Samain, M-S., Multilocus coalescent species delimitation to evaluate traditionally defined morphotypes in *Hydrangea* sect. *Asperae* (Hydrangeaceae), *Molecular Phylogenetics and Evolution* (2017), doi: http://dx.doi.org/10.1016/j.ympev.2017.05.021

Multilocus coalescent species delimitation to evaluate traditionally defined morphotypes in *Hydrangea* sect. *Asperae* (Hydrangeaceae).

Yannick De Smet[a], Olivier De Clerck[b], Tatsuya Uemachi[c], Carolina Granados Mendoza[a,d,e], Stefan Wanke[f], Paul Goetghebeur[a] & Marie-Stéphanie Samain[a,g]

a *Ghent University, Department of Biology, Research Group Spermatophytes, K.L. Ledeganckstraat 35, 9000 Gent, Belgium*

b *Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281 (S8), 9000 Gent, Belgium*

c *School of Environmental Science, The University of Shiga Prefecture, Hikone 522-8533, Japan*

d *Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-367, 04510 Coyoacán, Distrito Federal, Mexico*

e *CONACyT División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica AC, Camino a la Presa de San José 2055, Lomas 4a. sección, C.P. 78216 San Luis Potosí, San Luis Potosí, Mexico*

f *Technische Universität Dresden, Institut für Botanik, Zellescher Weg 20b, 01062 Dresden, Germany*

g *Instituto de Ecología, A.C., Centro Regional del Bajío, Avenida Lázaro Cárdenas 253, 61600 Pátzcuaro, Michoacán, Mexico (current address)*

Author for correspondence: *Yannick De Smet, yvp.desmet@gmail.com*

Abstract:

The number of species recognized in section *Asperae* of the flowering plant genus *Hydrangea* differs widely between subsequent revisions. This variation is largely centered around the *H. aspera* species complex, with numbers of recognized species varying from one to nearly a dozen. Despite indications of molecular variation in this complex, no sequence-based species delimitation methods have been employed to evaluate the primarily morphology-based species boundaries. In the present study, a multi-locus coalescent-based approach to species delimitation is employed in order to identify separate evolutionary lines within *H.* sect. *Asperae*, using four chloroplast and four nuclear molecular markers. Eight lineages were recovered within the focal group, of which five correspond with named morphotypes. The other three lineages illustrate types of conflict between molecular species delimitation and traditional morphology-based taxonomy. One molecular lineage comprises two named morphotypes, which possibly diverged recently enough to not have developed sufficient molecular divergence. A second conflict is found in *H. strigosa*.

This morphotype is recovered as a separate lineage when occurring in geographic isolation, but when occurring in sympatry with two other morphotypes (*H. aspera* and *H. robusta*), the coalescent species delimitation lumps these taxa into a single putative species.

Key words: *Hydrangea*, *H*. sect. *Asperae*, species delimitation, coalescent, species tree

1. Introduction

Species are held to be fundamental biological units, on par in importance with fundamental units at lower levels of organization such as cells and organisms (Mayr, 1982). Despite the importance of the species category, the second half of the 20th century has seen widespread controversy concerning its definition. However, since the publication of Darwin's "On the Origin of Species", all species concepts formulated within an evolutionary worldview have shared a common central idea. This core idea can be traced back with a few minor modifications to Darwin's own vision of species as branches in the lines of descent (de Queiroz, 2011). The proliferation of species concepts, however, originated from the idea that these lines of descent need to develop a specific property in order to be recognized as species ("species criterion", e.g. reproductive isolation, reciprocal monophyly, etc.). In an attempt to create a unified species concept, de Queiroz (1998, 1999, 2007) proposed to eliminate these species criteria, effectively reducing the alternative species concept to their common denominator: the evolutionary component first proposed by Darwin. Under this unified species concept, species are independently evolving metapopulation lineages. These lineages may or may not develop the properties used to delimit species in previous species concepts (e.g. reproductive isolation, distinct ecological niche, etc.) in the early stages of divergence. Moreover, these properties underlying the differences between alternative species concepts remain important in this unified species concept in at least three ways (de Queiroz, 2011). First, all of these properties represent different lines of evidence to recognize certain entities as separately evolving lineages. Secondly, explicitly mentioning the properties that differ between a set of recognized species can offer insights into the processes that cause or maintain lineages separation. Finally, these secondary properties can be used to distinguish subcategories of the species category based on the species criteria they satisfy, resulting in more objective and informative subcategories.

Despite the conceptual elegance of this unified species concept, contrasting different types of data can be challenging. Most, if not all, operational criteria for species delimitation are prone to misinterpret species diversity in certain circumstances. The often-used operational criterion of reciprocal monophyly, for example, is prone to misinterpretation of evolutionary lines due to incomplete lineage sorting (Maddison, 1997) or introgressive hybridization (Nosil et al., 2009). Because of these difficulties associated with molecular data, many species-delimitation studies have turned to methods for analyzing DNA sequence data in a coalescent-based framework, capable of accounting for confounding processes such as incomplete lineage sorting (ILS; Bagley et al., 2015). The algorithm for species validation implemented in the Bayesian Phylogenetics and Phylogeography program (BP&P; Yang & Rannala, 2010), for example, tests different species hypotheses based on a species tree. The latter is generated from a sample of multiple, unlinked molecular markers, allowing for gene tree incongruence caused by ILS. Generation of gene trees or guide trees, however, generally requires an a priori assignment of individuals to species (but see: Bryant et al., 2012). The majority of studies employing Bayesian algorithms for species delimitation seem to focus on morphologically cryptic radiations, validating molecularly divergent, but morphologically similar lineages as separate species. In this study, however, we aim to utilize a coalescent approach to species delimitation in a species complex consisting of several morphotypes of uncertain species status. This approach, i.e. comparing traditional morphological species delimitations with a molecular-based species hypothesis has the advantage of potentially validating morphological characters useful for identifying molecularly diverged lineages. Such diagnostic characters are highly valuable, for instance, in the identification of threatened or commercially valuable independent lineages.

Species circumscription and identification is notoriously difficult in the genus *Hydrangea* L., with widely varying numbers of species recognized by different authors (e.g. McClintock, 1957: 24 worldwide; Wei & Bartholomew, 2001: 33 only in China). The previously paraphyletic *Hydrangea* (Samain et al., 2010; Granados Mendoza et al., 2013) was recently rendered monophyletic by expanding its circumscription to include eight closely related

genera (De Smet et al., 2015a). Furthermore, a new infrageneric classification, supported by morphological and molecular data was proposed, consisting of 16 monophyletic sections. The focal group of this study, *Hydrangea* section *Asperae* (Rehder) Y.De Smet & Samain (hereafter named sect. *Asperae*), is distributed throughout eastern and southeastern Asia, with the highest diversity in central China. Most revisions addressing the genus *Hydrangea* agree on the recognition of the Japanese and Taiwanese representatives of sect. *Asperae* as separate species, owing to their distinct morphology (fig. 1: A, B & C). The remaining nominal taxa constitute the *H. aspera* Buch.-Ham. ex D.Don species complex, within which species boundaries have been unclear. According to McClintock (1957), this complex represents a single, wide-spread species, *H. aspera*. Moreover, she proposed four subspecies, based on the pubescence of the abaxial leaf surface, and the shape of petioles and leaves: *H. aspera* subsp. *aspera*, *H. aspera* subsp. *strigosa*, *H. aspera* subsp. *robusta* and *H. aspera* subsp. *sargentiana*. In contrast, other classifications (e.g. Wei & Bartholomew, 2001) recognize these subspecies and several other nominal taxa as distinct species, splitting the *H. aspera* complex into eight (Wei in: Wei & Bartholomew, 2001) or nine (Bartholomew in: Wei & Bartholomew, 2001) species. These nominal taxa (morphospecies) differ greatly in their ecology and geographic distribution. Some can be found across a wide geographic area (*H. strigosa* Rehder) while others are only known from a single location (*H. sargentiana* Rehder). Furthermore, several members of sect. *Asperae* occur in sympatry while others remain strongly geographically isolated, such as *H. kawakamii* Hayata, endemic to the island of Taiwan. As is often the case for purely morphology-based classifications, the difference in number of recognized species in the *H. aspera* complex hinges on differential emphasis on certain morphological characters for species identification. This uncertainty regarding species boundaries is exacerbated by the lack of knowledge regarding molecular variation and therefore evolutionary relationships within sect. *Asperae*. However, two cytogenetic studies (Cerbah et al., 2001; Mortreau et al., 2010) have demonstrated variation in the genomic organization among members of sect. *Asperae*. While *Hydrangea* species typically present a chromosome number of 2n=36, most members of sect. *Asperae* have 2n=34, with the exception of *H. involucrata* Siebold (2n=30). Furthermore, studying the chromosomal organization of the subspecies recognized by McClintock, Mortreau et al. (2010) found that a subset of specimens in *H. aspera* subsp.

aspera to which they refer as the "kawakamii-group" shows a chromosome number 2n=36. The authors therefore suggest that *H. aspera* subsp. *aspera* can be split into two taxa, coinciding with the described species *H. villosa* Rehder and *H. kawakamii*, based on differing chromosome organization.

The unclear taxonomic status of distinct morphotypes, showing different geographic distributions and genomic organization render sect. *Asperae* an ideal candidate to evaluate the capability of coalescent-based species delimitation to stabilize taxonomy in difficult groups. To this end, this study compares the evolutionary lineages proposed by a multilocus coalescent-based species delimitation algorithm (Yang & Rannala, 2010) with species boundaries proposed by strict monophyly and the most recent morphological species delimitation in sect. *Asperae* (Wei & Bartholomew, 2001). Furthermore, the potential of leaf pubescence to discriminate between evolutionary lineages in this section will be evaluated, as this is one of the main morphological characters both traditionally and recently employed to distinguish between sect. *Asperae* morphotypes.

2. Material and Methods

2.1 Taxon sampling and initial morphological identification

This study included 29 specimens identified as representatives of sect. *Asperae* and one species from its sister clade *Hydrangea* sect. *Cornidia* as outgroup. Most of these specimens were collected in China (provinces of Sichuan and Hubei) and Japan in 2011 and 2012. Other samples were obtained from herbarium material (Table S1). Initial identification of specimens followed the identification key in the Flora of China (Wei and Bartholomew, 2001), using Wei's more restrictive species boundaries. However, this key excludes the two Japanese species *H. involucrata* and *H. sikokiana* Maxim., which were identified using their original description and comparing to type specimens. Furthermore, during field work in Hubei, specimens closely resembling the type of *H. villosa* were found. This taxon is not included in Wei's key, as this author considers this taxon to be synonymous with *H. aspera*. However, its distinct morphology and indications of aberrant genomic organization (Mortreau et al., 2010) warrant the inclusion of these specimens under the name *H. villosa*.

This resulted in the recognition of ten putative species as a starting point for the coalescent based species delimitation. This approach is beneficial, since the algorithm applied here is unable to split taxa containing two or more related species. Furthermore, each identified specimen was morphologically compared to type material and original descriptions. All published taxa belonging to sect. *Asperae* are included in this study, with the exception of *H. coacta* C.F. Wei which is morphologically indistinguishable from *H. aspera*, as described in the Flora of China (Wei and Bartholomew, 2001).

2.2 Extraction, amplification and sequencing

A modified CTAB method (Doyle & Doyle, 1987) was used to extract total genomic DNA from silica gel dried leaf tissue or herbarium material. Two chloroplast intergenic sequences (IGS) and one chloroplast intron sequence were obtained for each specimen (*trnV-ndhC* IGS, *rpl32-ndhF* IGS, *trnL-rpl32* IGS and *ndhA* intron), apart from sequencing four nuclear regions (*TIF3H1*, *SMC1*-44, *SMC1*-22 and ITS). Primers and PCR amplification conditions for the chloroplast regions followed Granados Mendoza et al. (2013), except for the *ndhA* intron, for which primers published by De Smet et al. (2015a) were used. The ITS region was amplified using primers ITS1 and ITS4, following PCR conditions as described by White et al. (1990). Primers for amplifying both regions of the *SMC1* gene and the *TIF3H1* gene were designed based on the sequences of *Cornus wisoniana*, *C. officinalis*, and *Philadelphus incanus* generated by Zhang et al. (2012), and are specific for sect. *Asperae*. For a list of primer sequences see Table S2. Loci *SMC1*-44 and *SMC1*-22 are two regions of the same *SMC1* gene, but as the connecting region could not be amplified, both regions are analyzed separately to avoid creating chimeric sequences by combining PCR fragments from different alleles. For the chloroplast as well as the ITS regions, PCR products were cleaned using EXO-FASTAP (Thermo scientific, Pittsburgh, PA, USA). PCR products for *TIF3H1*, *SMC1*-44 and *SMC1*-22 were cloned using the Pgem T-easy Cloning Kit (Promega, Fitchburg, WI, USA). A minimum of 5 clones per accession were PCR-amplified directly from plated cultures according to manufacturer's instructions. Sequencing used the SP6 and T7 primers for cloned copies, and the primers applied in the PCR cycles for other regions. All sequencing was performed at Macrogen Europe. Raw sequences were edited and combined into contigs with Sequencher v5.0.1 (Gene Codes Corporation , Ann Arbor, MI, USA).

Alignments were generated with Prank v120712 (Löytnyoja & Goldman, 2005). All newly generated sequences were deposited in the European Nucleotide Archive (ENA, Table S1).

2.3 Single gene trees and concatenated analysis

Phylogenetic analyses were conducted on each locus individually, using Bayesian methods. Models of sequence evolution were selected using the Akaike information criterion implemented in jModeltest v2.3.1 (Darriba et al., 2012). When models unavailable in MrBayes 3.2.1 (Huelsenbeck & Ronquist, 2001) were selected, the next most parameterized model available was used. Each analysis was run for 20 million generations, using four chains in each of four independent runs with a sample frequency of 1000. Convergence of the Markov chains was assessed using the standard deviation of split frequencies, assuming convergence when this parameter drops below 0.01. Furthermore, convergence for each run was assessed in Tracer v1.6 (Rambaut & Drummond, 2013), as were effective sample sizes for all parameters.

2.4 Species tree estimation

All single gene alignments were used in Bayesian species tree estimation with *BEAST (Heled & Drummond, 2010). Best substitution models recovered by jModeltest or the next most general model were used. We ran *BEAST with five independent runs of 200 million generations each, sampling every 10000 generations, using uncorrelated relaxed clock models. LogCombiner v1.6.2 (Drummond & Rambaut, 2007) was used to combine the logs for the five independent runs, checking the resulting log in Tracer to verify if the effective sample size for all parameters exceeded 200. Tree files were combined using LogCombiner, discarding the first 5000 sampled trees as burn-in for each separate run. TreeAnnotator v1.6.2 (Drummond & Rambaut, 2007) was applied to calculate the Maximum clade credibility (MCC) tree from the combined dataset of trees.

Since *BEAST requires the taxa to be a priori assigned to species, taxa were identified as mentioned above. Furthermore, since single gene trees showed diversification between two groups of *Hydrangea strigosa*, these two clusters were assigned to different taxa. As the species tree generated by *BEAST would be used for species delimitation with BP&P v.3.0 (Yang & Rannala, 2010), it is better to erroneously split a true species than to lump two

non-sister taxa (Reid et al., 2012), since this method can lump taxa in the input tree, but not split them.

2.5 Bayesian species delimitation

Bayesian species delimitation was conducted using BP&P for all eight sequenced loci. This method requires an a priori defined species tree, and thus an initial allocation of all specimens to potential species. We used the species tree resulting from the *BEAST analysis as guide tree for the BP&P runs, but since the position of *Hydrangea villosa* was only weakly supported in this phylogram, we ran independent analyses for each possible resolution for the position of *H. villosa* as suggested by Leaché & Fujita (2010). Furthermore, BP&P runs can use one of two possible algorithms (1 or 0), and different combinations for prior distribution on the ancestral population size (θ) and root age (τ_0). Since these priors have been shown (Zhang et al., 2011) to influence the outcome of species delimitation, we ran BP&P for three different combinations of priors as suggested by Leaché and Fujita (2010). Both priors are assigned a gamma distribution: G(α,β), with a prior mean α/β and variance α/β^2. The first combination of priors assumed small population sizes and relatively shallow divergences: $\theta \sim$ G(2,2000) and $\tau_0 \sim$ G(2,2000). The second set of priors assumed large population sizes and deep divergences: $\theta \sim$ G(1,10) and $\tau_0 \sim$ G(1,10). The final combination of priors is a mixture of priors that assumes large ancestral population sizes and relatively shallow divergence among species: $\theta \sim$ G(1,10) and $\tau_0 \sim$ G(2,2000), which is a conservative combination of priors favoring models containing fewer species. Each of these three prior combinations were run with both possible algorithms (1 and 0), and for each of three possible species trees, for a total of 18 combinations of parameters. Each BP&P run consisted of 100000 generations, sampling every second generation, with a burn-in of 4000 generations. Each combination of parameters was first run for a limited amount of generations to select the fine tuning parameters for the MCMC moves which resulted in acceptance proportions between 0.15 and 0.7. Furthermore, each analysis was run twice to ensure proper mixing of the transmodel algorithm.

2.6 Scanning electron microscopy

Pubescence of the abaxial leaf surface was documented with scanning electron microscopy for each sampled morphotype. Dried leaves of similar age were sampled. The area documented was the same for all leaves, being the location where the main vein meets a secondary vein close to the middle of the leaf blade. Microscopic examination was performed with a Supra 40 VP SEM (Carl Zeiss, Germany) equipped with a cryopreparation unit (Emitech K1250X, Quorum Technologies Ltd, Ashford, Kent, UK) to obtain high-resolution images of abaxial leaf surfaces. Samples were glued to metal holders using TissueTek® O.C.T.™ conducting fluid (Sakura Finetek Europe B.V., Alphen aan den Rijn, The Netherlands), frozen in liquid nitrogen, and transferred into the cryochamber (-130°C). After sublimation at -70°C for 25 minutes, samples were sputter-coated with approximately 10 nm of gold-palladium prior to examination in the SEM at an accelerating voltage of 5 kV while kept at -100 °C. At least three images were taken per leaf including close-up images of the surface and trichomes structures.

3. Results

3.1 Single gene trees

Our data matrix of 240 sequences shows 8 missing sequences. Despite several attempts, we were unable to generate sequences for these combinations of markers and specimens. Single gene trees for chloroplast and nuclear markers agree on topology of the deeper branches. Specimens identified as *Hydrangea longifolia* Hayata and *H. involucrata* form a well-supported clade in all gene trees (figs. S1-8), which is sister to a larger clade containing all other representatives of sect. *Asperae*. In the latter clade, *H. sikokiana* is recovered as monophyletic and sister to the *H. aspera* species complex. However, this sister position is not always strongly supported, and even absent in the gene trees recovered from *trnV-ndhC* IGS and ITS, where *H. sikokiana* is recovered as a sister clade to the *H. longipes* Franch. – *H. involucrata* clade, or *H. involucrata* respectively. Within the *H. aspera* complex, gene trees reveal widespread topological discordance and varying resolution. However, some well-supported clades are shared among gene trees. Specimens identified as *H. villosa* are consistently recovered in a supported monophyletic clade (with the exception of *SMC1*-44).

A clade consisting of specimens identified as *H. longipes* and *H. sargentiana* is recovered in all regions with very high support. Specimens ascribed to *H. kawakamii* are recovered in a supported clade, or are part of an unresolved polytomy. The taxon designated as *H. strigosa* is recovered as polyphyletic in all gene trees. In the nuclear gene trees, representatives of this species are distributed across two well-supported clades, coinciding with their geographic distribution; one clade contains specimens collected in Hubei (China), while the other specimens originated from Sichuan (China). Chloroplast gene trees recover a similar split for specimens ascribed to *H. strigosa*, but lack the resolution to support each clade as monophyletic. The remaining taxa *H. robusta* and *H. aspera* are not recovered as monophyletic groups, but specimens identified as these taxa cluster together in all chloroplast gene trees. However, although most specimens identified in the field as *H. aspera*, *H. robusta* or *H. strigosa* (Sichuan collections) are recovered as a highly supported clade in plastid gene trees, two specimens are repeatedly recovered outside this clade. These specimens (*H. aspera* 1349 and *H. robusta* 1351) were collected in Nepal and India respectively, at locations near the type locality for these taxa. Specimens of these three nominal taxa are not consistently grouped together in the nuclear gene trees.

3.2 Species tree

The MCC tree obtained from the five independent *BEAST analyses provides better resolution for the evolutionary relationships within sect. *Asperae* compared to the single gene trees (fig. 2). The topological placement for the Japanese species (*H. involucrata*, *H. sikoniana*) and *H. longifolia* concurs with that found in the single gene trees. Within the *H. aspera* complex, the *BEAST analysis provides improved resolution and nodal support over the single gene analyses. A split between *H. sargentiana* and *H. longipes* is well supported, and these two morphospecies form a clade sister to the rest of the complex, which is split into two clades. A first clade consists of *H. robusta*, *H. aspera* and *H. strigosa* (Sichuan population). This clade is recovered with posterior probability (PP) of 1; however, relationships within this clade remain unsupported (PP: 0.84). The second clade contains *H. villosa*, *H. kawakamii* and *H. strigos*a. The sister relationship between *H. strigosa* (Hubei population) and *H. kawakamii* received high support (PP: 1), whereas the position of *H. villosa* as sister to these two putative species remains unsupported (PP: 0.54).

3.3 Bayesian species delimitation

Bayesian species delimitation results for sect. *Asperae* are summarized in fig. 3. Only three nodes in the guide tree received speciation probabilities below 1 for all analyses: the node splitting *H. sargentiana* and *H. longipes*, and the two nodes separating *H. aspera*, *H. strigosa* (Sichuan population) and *H. robusta*. Placement of *H. villosa* in the guide tree and the choice of algorithm 0 or 1 did not affect the number of species recognized, only resulting in minor changes in the posterior probabilities for the three unsupported nodes. Prior distribution for τ and θ had a minor impact on the speciation probabilities for the nodes splitting *H. sargentiana* from *H. longipes* and *H. aspera* from *H. strigosa* (Sichuan population). However, speciation probability associated with the node splitting *H. robusta* from the *H. aspera – H. strigosa* clade varies strongly in response to changes in the prior distribution for τ and θ. Despite this variation, PP for this node never exceeds 0.95; consequently *H. robusta* is not supported as a separate species by BP&P. Remarkably, in only one of the 18 possible parameter combinations, the node splitting *H. sargentiana* and *H. longipes* receives a PP of 1 (fig. 3), while other combinations of parameters never result in a PP higher than 0.22. This PP remains constant after re-running BP&P for this combination of parameters.

3.4 Abaxial leaf surface pubescence

The different nominal taxa included in this study were morphologically heterogeneous with respect to the pubescence of their abaxial leaf surface (fig. 4). Most observed trichome types coincide with the types described in protologues and previous revisions of sect. *Asperae*. Besides variation in the morphology of the trichomes, differences in the ornamentation of the leaf surface were observed, more specifically, in the presence or absence of white papillae.

For nominal taxa *H. sikokiana* (fig. 4: A & B), *H. involucrata* (fig. 4: C & D), *H. kawakamii* (fig. 4: G & H) and *H. aspera* (fig.4: K & L), trichomes on the lower leaf surface can be described as long and erect, with conspicuous tubercles on their surface. Differences in appearance between these taxa is mainly due to variation in the density of the pubescence, and length of trichomes. A similar type of trichome is found in specimens morphologically ascribed to *H. villosa* (fig. 4: O & P), where they are supplemented with longer, stiff hairs on the larger

veins of the leaves. A similar situation occurs in *H. longipes* (fig. 4: M & N), but here dense groups of these hairs can be found in the axils formed by the main and secondary veins, visible as white tufts to the naked eye.

Two nominal taxa, *H. strigosa* (fig. 4: Q & R) and *H. robusta* (fig. 4: I & J) exhibit small appressed hairs on their lower leaf surface. The surface of these trichomes is adorned with small tubercles. Both taxa differ in the girth of these hairs, with those present in *H. strigosa* being much narrower than those of *H. robusta*.

Of the putative species examined in this study, two exhibited an autapomorphous type of trichomes. In *H. longifolia* (fig. 4: E & F), the lower leaf surface shows appressed hairs similar to those of *H. strigosa* and *H. robusta* interspersed with two-branched appressed hairs (fig. 4: F), which are especially dominant on larger veins and petioles. Petioles, flowering stems and main veins of the abaxial leaf surface show trichomes exhibiting a conspicuous fleshy base (fig. 4: S & T) in *H. sargentiana*, which are not observed in any other species of *Hydrangea*. These fleshy trichomes lend the petioles and inflorescences of this putative species its distinctive habit (fig. 1: D).

Apart from variation in pubescence type, two examined taxa differ from the others in the presence of papillae on the abaxial leaf surface. These are white and very prominent in *H. strigosa* (both Sichuan and Hubei populations) (fig. 4: Q & R), but less conspicuous in *H. kawakamii* (fig. 4: G & H).

4. Discussion

4.1 Reciprocal monophyly versus coalescent-based species delimitation

Our analyses support the recognition of several independent evolutionary lines within *H.* sect. *Asperae*, and is the first study to offer molecular evidence for the presence of separate lineages within the *H. aspera* complex. Furthermore, our results highlight an advantage of employing multilocus, coalescent-based species delimitation over reciprocal monophyly in single gene trees. Utilizing these coalescent-based methods provided better resolution for both evolutionary relationships and species boundaries within the focal section. Nevertheless, the operational criterion of reciprocal monophyly in gene trees is a valid way

of discerning independent evolutionary lineages, albeit a very strict one. Indeed, a substantial amount of generations can be required for two lineages to reach reciprocal monophyly (Hudson & Coyne, 2002; Knowles & Carstens, 2007). This criterion will therefore be unable to identify recently diverged lineages, as these have a high chance of harboring ancestral polymorphisms, rendering them polyphyletic for certain loci. In contrast, species delimitation methods based on coalescent theory represent a probabilistic approach to recognizing separate evolutionary lineages, not requiring reciprocal monophyly or fixed differences. Rather, these methods utilize information from multiple molecular markers to test alternative hypotheses of species delimitation, while allowing for gene tree discordance caused by genetic drift (ILS in the case of BP&P) (Rannala & Yang, 2003; Knowles & Carstens, 2007; Yang & Rannala, 2010). Although these coalescent-based methods are more sensitive in recognizing recently diverged lineages, most contemporary methods fail to discern lineages in the face of strong gene flow. Although the BP&P algorithm has been shown to be robust against a limited amount of gene flow (Zhang et al., 2011), this might limit its utility in sympatric species, where hybridization and introgression are more likely. Furthermore, the analysis has been shown to be sensitive to choice of the priors on ancestral population size and species divergence times (Leaché & Fujita, 2010; Zhang et al. 2011).

4.2 Species delimitation in *Hydrangea* section *Asperae*

Application of multi-locus coalescent-based species delimitation to our dataset of ten nominal taxa currently recognized in sect. *Asperae* resulted in the recognition of eight separate lineages. A number of these correspond to a single nominal taxon, whereas others show less straightforward correspondence to named morphotypes. These lineages include: 1) *H. involucrata* from Japan, 2) *H. longifolia* endemic to Taiwan, 3) the Japanese *H. sikokiana*, 4) specimens identified *as H. sargentiana* and *H. longipes*, 5) *H. kawakamii* endemic to Taiwan, 6) specimens identified as *H. strigosa* collected in Hubei, China, 7) *H. villosa* from China, and 8) specimens morphologically ascribed to the nominal taxa *H. robusta*, *H. aspera* and *H. strigosa* collected in Sichuan, China. A subset of these lineages correspond to highly supported monophyletic groups in all (1,2,3,4) or a substantial subset

(5,7) of the gene trees. Furthermore, they are morphologically clearly identifiable based on clear-cut diagnostic characters, such as abaxial leaf pubescence (e.g. fig 4).

Results from the coalescent analyses and gene trees suggest *H. involucrata*, *H. longifolia* and *H. sikokiana* to be separate evolutionary lineages. All gene trees recovered these lineages as monophyletic, which combined with their distinct morphology advocates their recognition as clearly diverged species. Geographic isolation from the other members of sect. *Asperae* is possibly the driving factor behind this pronounced divergence.

Within the *H. aspera* complex, two lineages identified in the coalescent analyses coincide with named morphospecies (*H. kawakamii*, *H. villosa*). Although they are only recovered as monophyletic in a subset of the gene trees, high speciation probabilities in all coalescent analyses and a distinctive morphology provide ample evidence to support these nominal taxa as separate evolutionary lineages. The lack of support for monophyly of these taxa in some, but not all, gene trees illustrates the shortcomings of using strict monophyly as the sole criterion for species recognition. Both taxa can represent separate evolutionary lineages, but some loci might experience ILS, or low sequence divergence, obscuring the evolutionary relationships of specimens belonging to *H. villosa* and *H. kawakamii*. The lack of resolution in most gene trees concerning the placement of these two species could represent an indication of the presence of these confounding factors.

The remaining three lineages recognized by the coalescent analyses present two opposing conflicts between nominal (morphology-based) taxonomy and sequence-based species delimitation. In a first case, two morphologically very distinct taxa are strongly supported to constitute a single species based on molecular data. In the second case, a morphologically homogenous group of specimens is split up into two evolutionary distinct lineages.

The operational criteria of strict monophyly and Bayesian species delimitation suggest morphospecies *H. sargentiana* and *H. longipes* to constitute a single species. Moreover, sequences recovered for all eight loci are nearly identical across specimens identified as these taxa. Morphologically however, both putative species are distinct. Petioles and stems

of *H. sargentiana* are covered with conspicuous fleshy trichomes (fig. 1D, 4T) while this type of indument is completely absent from *H. longipes*. Both putative species differ greatly in general appearance: *H. sargentiana* forming large leaves and inflorescences with purple central flowers, while *H. longipes* develops white central flowers and smaller leaves with distinct long and slender petioles. Furthermore, *H. sargentiana* is unique within section *Asperae* in being known from a single wild population in Hubei, China (De Smet et al., 2015b). While *H. longipes* does occur in the same region, its geographic distribution is far wider, covering the Chinese provinces of Hubei and Sichuan. Phenotypic divergence preceding molecular divergence can indicate a recent speciation event, caused by variation in a limited subset of loci. Such speciation would be difficult to detect using a limited subset of neutral markers, as these might not carry any record of the speciation event (Fujita et al., 2012). An alternative explanation for the lack of molecular divergence is strong and ongoing gene flow between both morphospecies. The lack of specimens with intermediate morphology, and the perseverance of the typical *H. sargentiana* morphology amidst a larger population of *H. longipes* morphotypes argue against strong intermixing of both forms. Since *H. sargentiana* can maintain its distinct morphology within the larger geographic distribution of *H. longipes*, we suggest that both morphotypes represent separate evolutionary lineages. Discordance between genetic and morphological divergence between *H. sargentiana* and *H. longipes* could suggest a recent divergence of *H. sargentiana* from the geographically more widespread morphotype. In this case sequence divergence between the two morphotypes would be expected to remain low, insufficient variation having accumulated, and ancestral polymorphisms not having sorted.

Hydrangea strigosa is reported to be a widespread species, distributed from Western Sichuan to Eastern Hubei. Our molecular data suggests two different lineages within this morphospecies; one situated in Hubei, and one from Sichuan. The Hubei lineage is supported as distinct by all coalescent-based analyses, as well as the monophyly criterion (for a subset of the sampled loci). The Sichuan lineage is supported as monophyletic by the ITS gene tree, whereas the remaining sequenced regions and all coalescent analyses failed to support this lineage as distinct. Instead, this Sichuan lineage of *H. strigosa* is closely related to *H. robusta* and *H. aspera*, which also occur in Sichuan. All coalescent-based

analyses support the recognition of these three morphotypes as a single evolutionary lineage. Our data therefore suggest that *H. strigosa* forms a distinct evolutionary lineage only when occurring in allopatry from the closely related nominal taxa *H. aspera* and *H. robusta*. Indeed, in Sichuan, where these putative species co-occur, a gradual transition can be found between populations of these species, along an altitudinal gradient (McClintock, personal observation on mt. Emei), strongly suggesting gene flow between these entities. In Hubei, on the other hand, no specimens morphologically identifiable as *H. aspera* or *H. robusta* were found in sympatry with the sampled *H. strigosa* specimens (personal observation). Similar patterns have been observed in fucoid brown algae, with species constituting separate evolutionary lines in allopatry, but exhibiting extensive gene flow in sympatry with closely related taxa (Zardi et al., 2011). Therefore, with the current knowledge, we consider the Hubei lineage of *H. strigosa* strongly supported as an independent evolutionary lineage. This lineage furthermore contains specimens collected at the type location of *H. strigosa*, ensuring the connection of this evolutionary lineage to the nominal taxon. Species boundaries between *H. strigosa*, *H. aspera* and *H. robusta* in the Chinese province of Sichuan are less straightforward. With the sampling of specimens and markers achieved in this study, it is unclear whether these named taxa represent a single evolutionary lineage, or if their lumping in our analyses is caused by the sensitivity of the utilized methods to gene flow. Future studies should explore the population level diversity of these taxa in Sichuan, addressing the possibility of extensive gene flow along altitudinal gradients.

Conclusions

Our analyses were able to unravel part of the difficult *H. aspera* species complex. Following our coalescent based species delimitation and the operational criterion of reciprocal monophyly, at least three morphotypes warrant recognition as species. These morphotypes are: *H. villosa*, *H. kawakamii* and *H. strigosa* (Hubei lineage). Despite the lack of molecular divergence, we propose the recognition of *H. sargentiana* and *H. longipes* as separate species, owing to their differing morphology and geographical isolation. Finally, this study was unable to provide evidence for the divergence of *H. strigosa* (Sichuan), *H. aspera* and *H. robusta*, suggesting them to represent a single, morphologically variable species, or a

species complex experiencing heavy gene flow. However, since these morphotypes were not sampled at their type location, the connection to these published names is uncertain. A similar study including specimens with a clear connection to these published names could provide further insight into their species status.

Acknowledgements:

The authors would like to thank Pieter Asselman for technical assistance, Prof. Dr. Chen Fangqing (China Three Gorges University) and Prof. Dr. Jer-Ming Hu (National Taiwan University) for providing wild-collected specimens, as well as Prof. Dr. Hong Ma for providing Cornales *SMC1* gene sequences for primer design. This work was supported by the Research Foundation Flanders (FWO Vlaanderen; FWO fellowship 1.1.518.11N), the Fondation Franklinia (Ghent University project number E/01394/01) and the Bundesministerium für Bildung und Forschung (BMBF) via the KMU-innovativ 9: Biotechnologie – BioChance project to the TU Dresden. We are grateful to all herbaria which sent us material for this study (CAS, WU, P, K, US). Furthermore, the authors thank two anonymous reviewers for valuable comments on earlier versions of the manuscript.

References:

Bagley, J.C., Alda, F., Breitman, M.F., Bermingham, E., van den Berghe, E.P., Johnson, J.B., 2015. Assessing Species Boundaries Using Multilocus Species Delimitation in a Morphologically Conserved Group of Neotropical Freshwater Fishes, the Poecilia sphenops Species Complex (Poeciliidae). PLoS One 10, e0121139. doi:10.1371/journal.pone.0121139

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., Roychoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. doi:10.1093/molbev/mss086

Cerbah, M., Mortreau, E., Brown, S., Siljak-Yakovlev, S., Bertrand, H., Lambert, C., 2001. Genome size variation and species relationships in the genus Hydrangea. Theor. Appl. Genet. 103, 45–51. doi:10.1007/s001220000529

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. doi:10.1038/nmeth.2109

De Smet, Y., Granados Mendoza, C., Wanke, S., Goetghebeur, P., Samain, M.S., 2015a. Molecular phylogenetics and new (infra)generic classification to alleviate polyphyly in tribe *Hydrangeeae* (Cornales: Hydrangeaceae). Taxon 64 (4). doi:10.12705/644.6

De Smet, Y., Larridon, I., Bauters, K., Goetghebeur., P, Wanke, S., Samain, M.S. 2015b. Rediscovering *Hydrangea sargentiana*, a taxon in need of conservation action.. Acta horticulturae 1087: 221-224. doi:10.17660/ActaHortic.2015.1087.26

de Queiroz, K., 1998. Chapter 5. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations., in: Endless Forms: Species and Speciation. pp. 57–75. doi:10.1080/10635150701701083

de Queiroz, K., 1999. The General Lineage Concept of species and the defining properties of the species category, in: Species: New Interdisciplinary Essays. pp. 49–89.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. doi:10.1080/10635150701701083

de Queiroz, K., 2011. Branches in the lines of descent: Charles Darwin and the evolution of the species concept. Biol. J. Linn. Soc. 103, 19–35. doi:10.1111/j.1095-8312.2011.01634.x

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15. doi:10.2307/4119796

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214. doi: 10.1186/1471-2148-7-214

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. doi:10.1016/j.tree.2012.04.012

Granados Mendoza, C., Wanke, S., Salomo, K., Goetghebeur, P., Samain, M.S., 2013. Application of the phylogenetic informativeness method to chloroplast markers: A test case of closely related species in tribe Hydrangeeae (Hydrangeaceae). Mol. Phylogenet. Evol. 66, 233–242. doi:10.1016/j.ympev.2012.09.029

Granados Mendoza, G., Naumann, J., Samain, M.-S., Goetghebeur, P., De Smet, Y., Wanke, S. 2015. A genome-scale mining strategy for recovering novel rapidly-evolving nuclear single-copy genes for addressing shallow-scale phylogenetics in *Hydrangea*. BMC Evolutionary Biology 15 (1). doi:10.1186/s12862-015-0416-z

Heled, J., Drummond, A.J., 2010. Bayesian Inference of Species Trees from Multilocus Data. Mol. Biol. Evol. 27, 570–580. doi:10.1093/molbev/msp274

Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. Evolution 56, 1557–1565.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755. doi:10.1093/bioinformatics/17.8.754

Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56, 887–895. doi:10.1080/10635150701701091

Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). Proc. Biol. Sci. 277, 3071–3077. doi:10.1098/rspb.2010.0662

Löytynoja, A., Goldman, N., 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. U. S. A. 102, 10557–10562. doi:10.1073/pnas.0409137102

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536. doi:10.2307/2413694

Mayr, E. 1982 The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Belknap Press of Harvard Univ. Press, Cambridge, MA

McClintock, E., 1957. A monograph of the genus *Hydrangea*. Proc.Calif. Acad. Sci. 29, 14–256.

Mortreau, E., Siljak-Yakovlev, S., Cerbah, M., Brown, S.C., Bertrand, H., Lambert., C., 2010. Cytogenetic characterization of *Hydrangea involucrata* Sieb. and *H. aspera* D. Don complex (Hydrangeaceae): genetic, evolutional, and taxonomic implications. Tree Genetics & Genomes 6: 137-148

Nosil, P., Harmon, L.J., Seehausen, O., 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. doi:10.1016/j.tree.2008.10.011

Rambaut, A., Drummond, A.J., 2013. Tracer V1.6. Available from http//beast.bio.ed.ac.uk/Tracer.

Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.

Reid, N., Demboski, J.R., Sullivan, J., 2012. Phylogeny Estimation of the Radiation of Western North American Chipmunks (Tamias) in the Face of Introgression Using Reproductive Protein Genes. Syst. Biol. 61(1) :44‑62.

Samain, M.S., Wanke, S., Goetghebeur, P., 2010. Unraveling Extensive Paraphyly in the Genus *Hydrangea* s. l. with Implications for the Systematics of Tribe Hydrangeeae. Syst. Bot. 35(3): pp. 593–600. doi:10.1600/036364410792495827

Wei, C., Bartholomew, B., 2001. *Hydrangea*. Flora of China. Beijing and Missouri Botanical Garden Press 8:145-157.

White, T.J., Bruns, S., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in: PCR Protocols: A Guide to Methods and Applications. pp. 315–322. doi:citeulike-article-id:671166

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. U. S. A. 107, 9264–9269. doi:10.1073/pnas.0913022107

Zardi, G.I., Nicastro, K.R., Canovas, F., Costa, J.F., Serrão, E.A., Pearson, G.A., 2011. Adaptive traits are maintained on steep selective gradients despite gene flow and hybridization in the intertidal zone. PLoS One 6. doi:10.1371/journal.pone.0019402

Zhang, C., Zhang, D.-X., Zhu, T., Yang, Z., 2011. Evaluation of a Bayesian Coalescent Method of Species Delimitation. Syst Biol 60 (6): 747-761. doi: 10.1093/sysbio/syr071


A


B


C


D


E

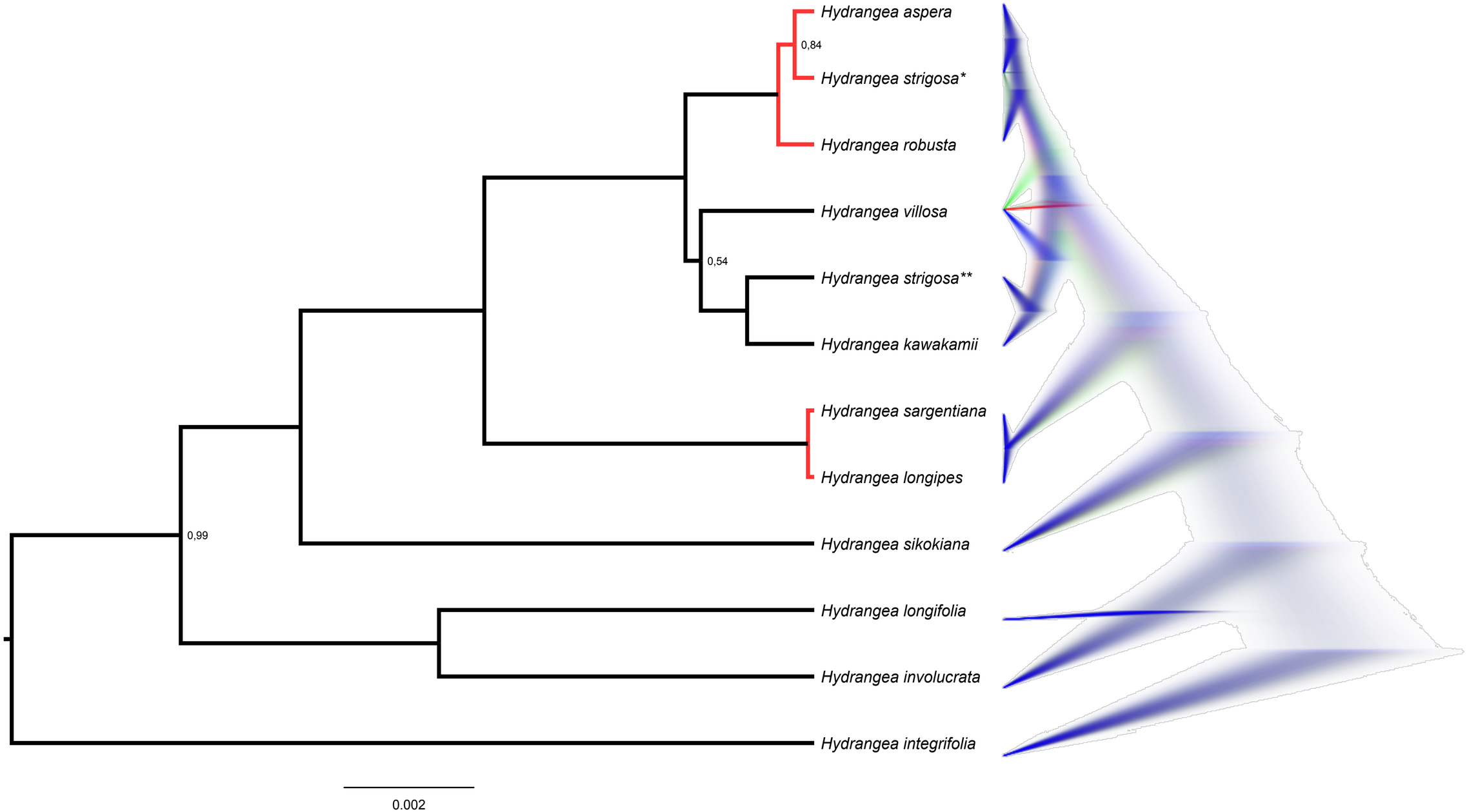

F


G

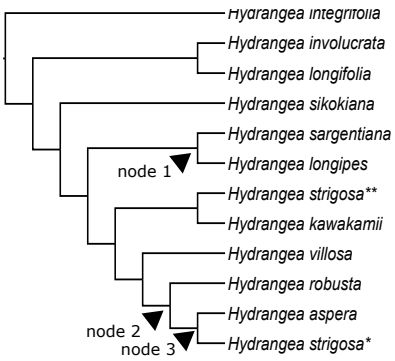

H

I

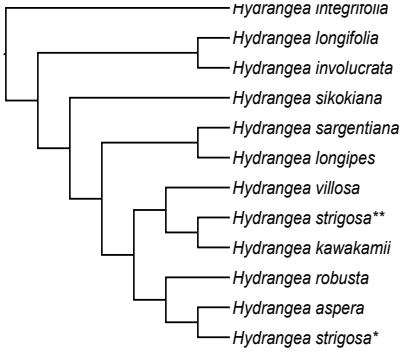
J


Hydrangea aspera
0,84
Hydrangea strigosa*
Hydrangea robusta
Hydrangea villosa
0,54
Hydrangea strigosa**
Hydrangea kawakamii
Hydrangea sargentiana
Hydrangea longipes
0,99
Hydrangea sikokiana
Hydrangea longifolia
Hydrangea involucrata
Hydrangea integrifolia
0.002

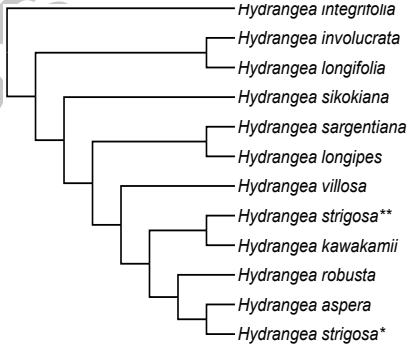
Algorithm	guide tree	theta	tau	node 1	node 2	node 3
0	topology 1	2 2000	2 2000	0,08	0,01	0,00
0	topology 1	1 10	1 10	0,15	0,79	0,18
0	topology 1	1 10	2 2000	0,22	0,92	0,32
1	topology 1	2 2000	2 2000	0,08	0,02	0,00
1	topology 1	1 10	1 10	0,00	0,79	0,20
1	topology 1	1 10	2 2000	0,00	0,84	0,26
0	topology 2	2 2000	2 2000	0,08	0,01	0,00
0	topology 2	1 10	1 10	0,05	0,49	0,06
0	topology 2	1 10	2 2000	0,00	0,61	0,10
1	topology 2	2 2000	2 2000	0,09	0,01	0,00
1	topology 2	1 10	1 10	0,06	0,47	0,05
1	topology 2	1 10	2 2000	0,07	0,72	0,16
0	topology 3	2 2000	2 2000	0,09	0,01	0,00
0	topology 3	1 10	1 10	0,00	0,82	0,20
0	topology 3	1 10	2 2000	0,00	0,81	0,22
1	topology 3	2 2000	2 2000	0,07	0,01	0,00
1	topology 3	1 10	1 10	0,02	0,79	0,21
1	topology 3	1 10	2 2000	1,00	0,90	0,29



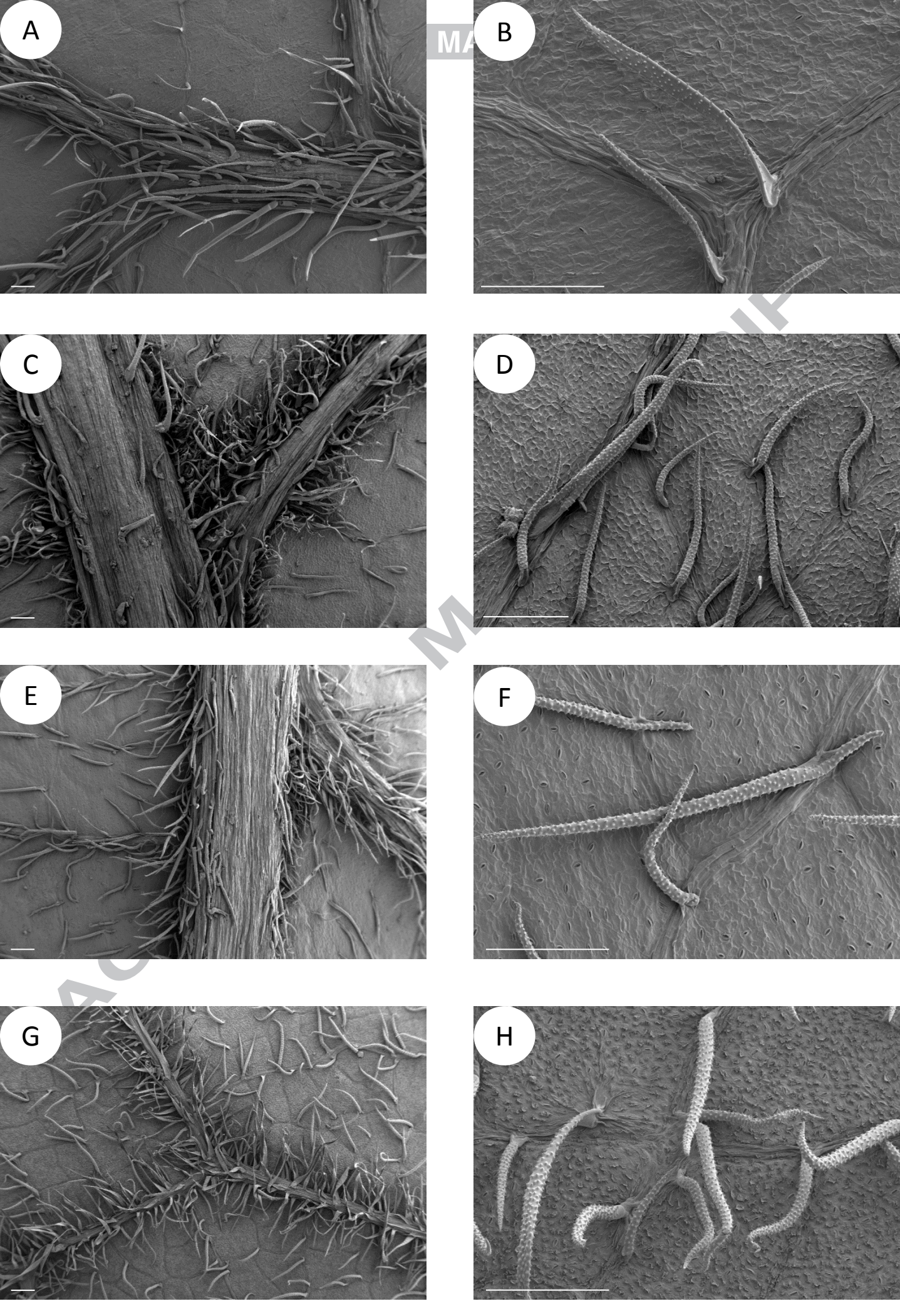
Topology 1

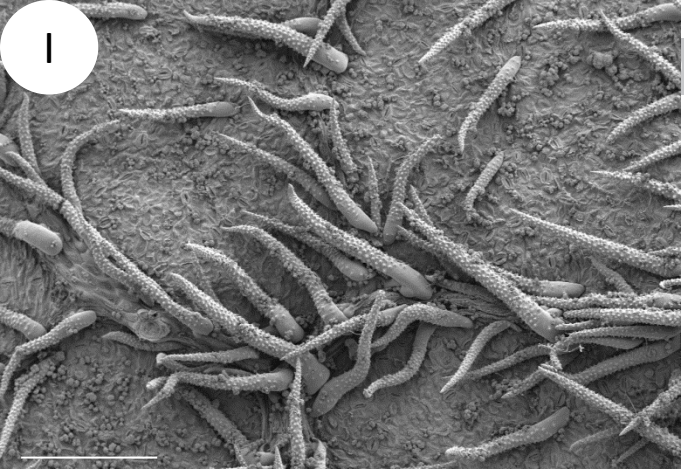


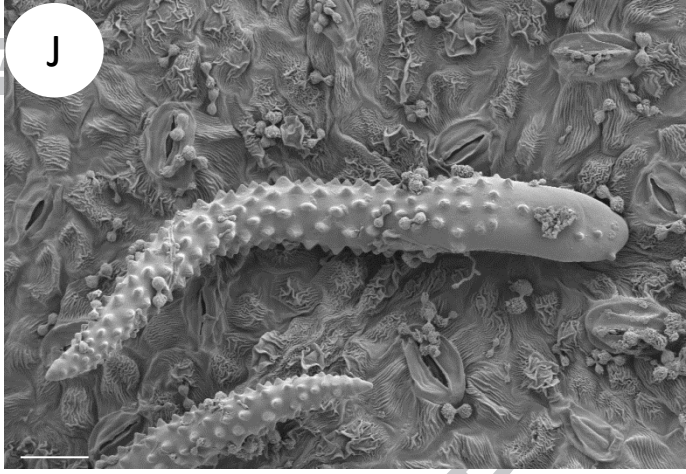
Topology 2

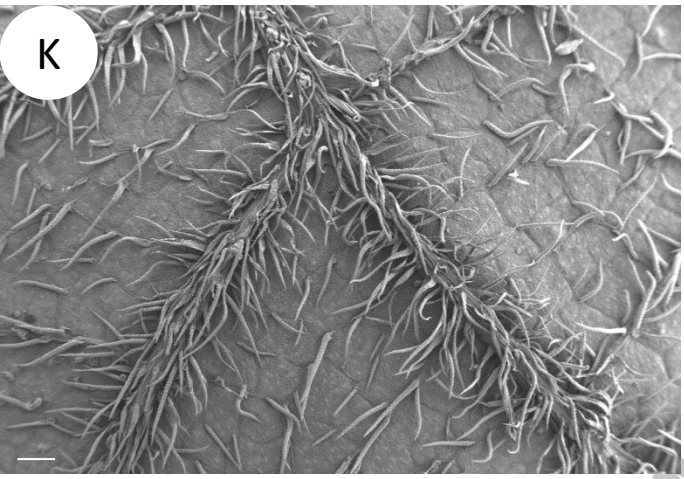


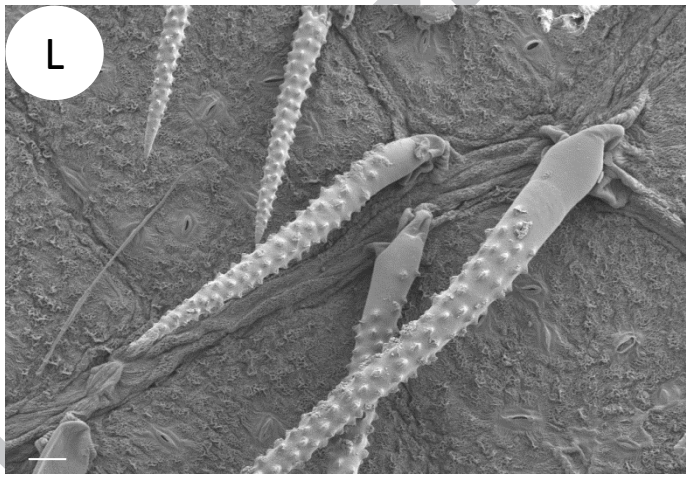
Topology 3

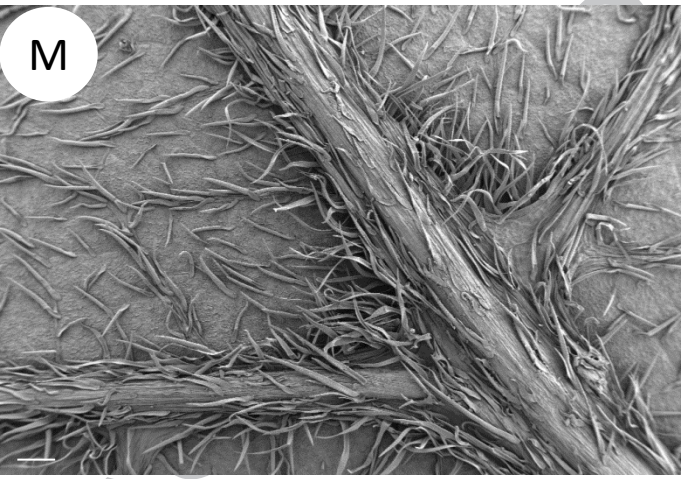

A
B
C
D
E
F
G
H

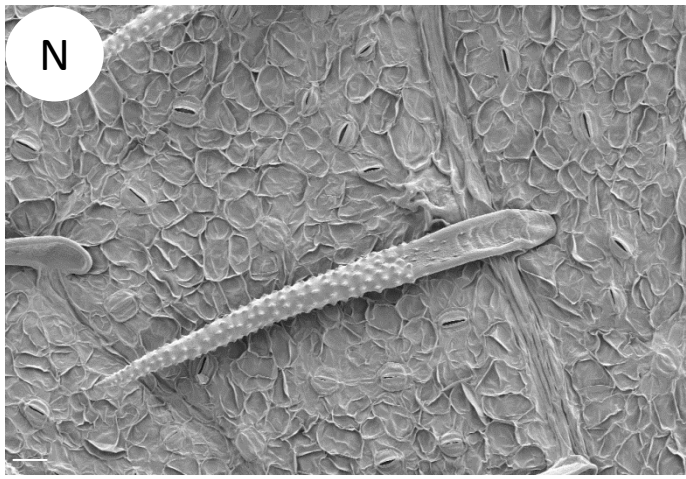

I


J

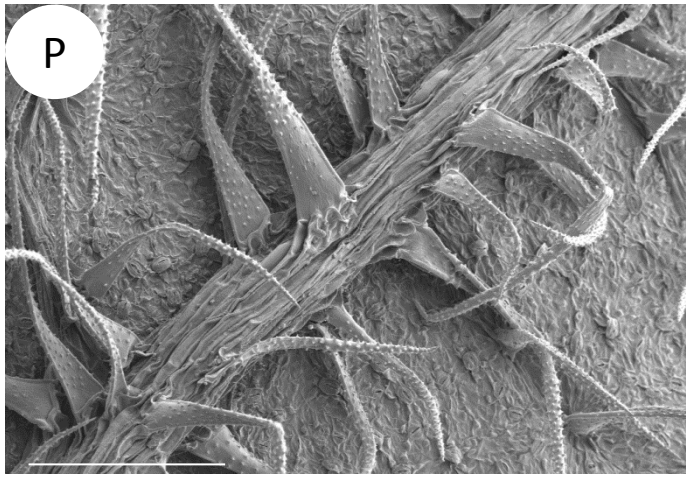

K

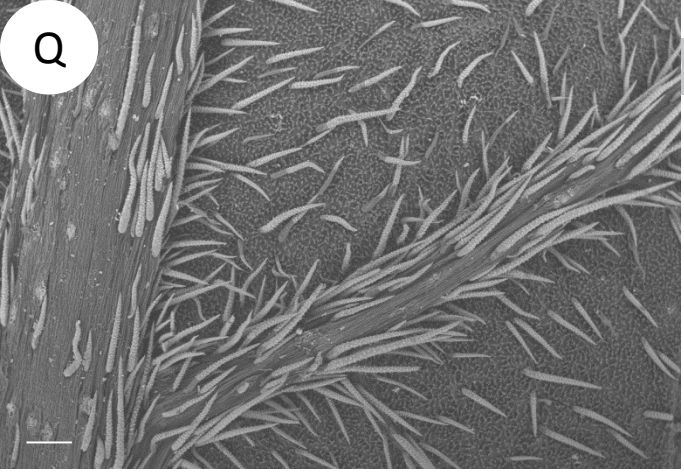

L

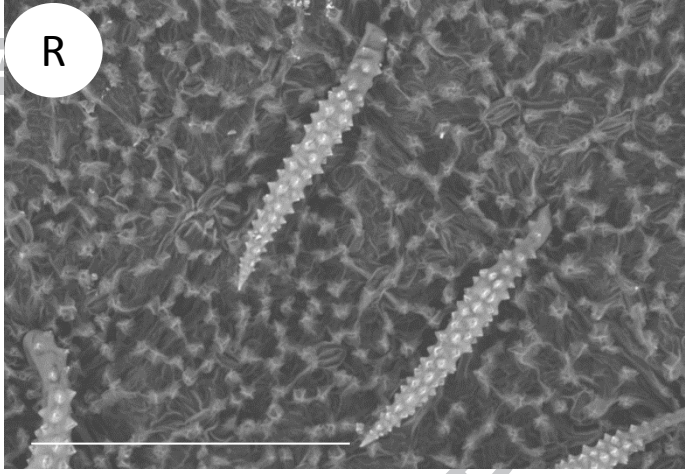

M

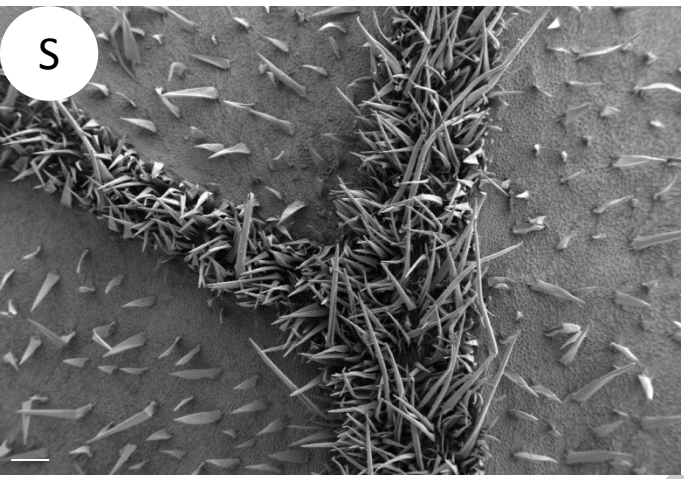

N

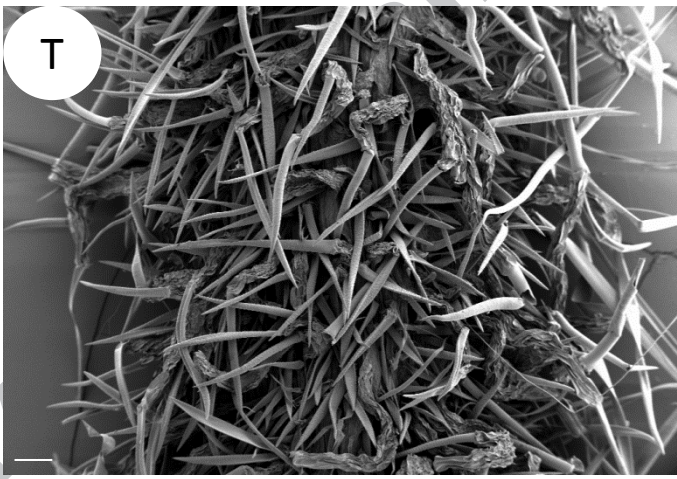

O


P









Hydrangea aspera
0,84
*Hydrangea strigosa**
Hydrangea robusta
Hydrangea villosa
0,54
*Hydrangea strigosa***
Hydrangea kawakamii
Hydrangea sargentiana
Hydrangea longipes
0,99
Hydrangea sikokiana
Hydrangea longifolia
Hydrangea involucrata
Hydrangea integrifolia
0.002

- We present the first molecular evaluation of species boundaries in *Hydrangea*.
- Coalescent-based species delimitation identifies eight separate lineages.
- Five well-supported species hypotheses can be proposed.
- Traditionally recognized *H. strigosa* is split into geographically isolated lineages.
- *H. strigosa*, *H. robusta* and *H. aspera* are recovered as a single lineage.
- *H. sargentiana* and *H. longipes* are recovered as a single lineage.