

Cryptic species, gene recombination and hybridization in the genus Spiraeanthemum (Cunoniaceae) from New Caledonia

Yohan Pillon, Helen Hopkins, Jérôme Munzinger, Hamid Amir, Mark Chase

▶ **To cite this version:**

Yohan Pillon, Helen Hopkins, Jérôme Munzinger, Hamid Amir, Mark Chase. Cryptic species, gene recombination and hybridization in the genus Spiraeanthemum (Cunoniaceae) from New Caledonia. Botanical Journal of the Linnean Society, 2009, Botanical Journal of the Linnean Society, 161 (2), pp.137-152. 10.1111/j.1095-8339.2009.00997.x . ird-03635321

HAL Id: ird-03635321
https://ird.hal.science/ird-03635321

Submitted on 8 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cryptic species, gene recombination and hybridisation in the genus *Spiraeanthemum* (Cunoniaceae) from New Caledonia

Yohan Pillon[1,2], Helen C.F. Hopkins[3], Jérôme Munzinger[1], Hamid Amir[2] & Mark W. Chase[4]

[1] Laboratoire de Botanique, Institut de Recherche pour le Développement, BP A5, 98848 Nouméa Cedex, New Caledonia

[2] Laboratoire Insulaire du Vivant et de l'Environnement, Université de la Nouvelle-Calédonie, BP R4, 98851 Nouméa Cedex, New Caledonia.

[3] c/o The Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK

[4] Jodrell laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK

Running title: Cryptic species in New Caledonian *Spiraeanthemum*

ABSTRACT

The Oceanian plant genus *Spiraeanthemum* (Cunoniaceae) has a centre of diversity in New Caledonia, where it is represented by seven species. Its diversification was investigated using two low-copy nuclear genes, *ncpGS* and *GapC*, and phylogenetic analyses were based on maximum parsimony, maximum likelihood, and recombination networks. We detected several cases of gene recombination in both datasets, and these have obscured history within the genus. For *S. ellipticum* and *S. pubescens*, accessions from southern populations on ultramafic soils were genetically distinct from accessions from northern populations on non-ultramafic soils. Given that no obvious morphological characters distinguish northern and southern populations in either taxon, both may be considered examples of cryptic species. Incongruence between gene trees and species delimitation may be explained by the parallel evolution of similar morphology, differential lineage sorting leading to differential fixation of alleles, or different introgression patterns in the north and south leading to allele displacement. In New Caledonia, some species with broad ecological preferences may thus be artificial concepts. This suggests that they should be treated more critically in monographs and that the species richness of the New Caledonian flora may be underestimated. Problems associated with the typification of *S. ellipticum* and evidence of hybridisation events in the history of *Spiraeanthemum* are also discussed.

Keywords: Biodiversity hotspots, conservation of island endemics, hybridisation, island biogeography, single- nuclear copy gene, Pacific islands, serpentine, taxonomic inflation

INTRODUCTION

Cryptic species are defined as "two or more distinct species that are erroneously classified (and hidden) under one species name" (Bickford *et al.*, 2007). The importance of cryptic species probably varies among ecosystems, geographical areas and taxonomic groups, and few studies are available in the tropics. A few cases are known in higher plants (Bickford *et al.*, 2007), including the *Caesalpinia hintonii* complex (Fabaceae, Sotuyo *et al.*, 2007) in Mexican dry forests, which contains two sibling species (sensu Bickford *et al.*, 2007) within *C. oyamae* S. Sotuyo & G. P. Lewis that are geographically distinct. Other recently reported examples belong to the genera *Allium* (Alliaceae, Guruschidze, Fritsch & Blattner, 2008) and *Cedrela* (Meliaceae, Muellner, Pennington & Chase, 2009) in which sibling species and cryptic species that are sometimes only distantly related have been identified in molecular phylogenetic analyses. Further occurrences have also been reported in *Potamogeton* (Potamogetonaceae, Whittall *et al.*, 2004), *Veronica* (Plantaginaceae, Martínez-Ortega *et al.*, 2004) and possibly *Liparis* (Orchidaceae, Pillon *et al.*, 2007), although in these cases the "cryptic species" revealed by molecular data were already described taxa and so were not really hidden, but rather poorly characterized. New examples of cryptic species are revealed every year with the use of molecular tools in a wide range of organisms (Rieseberg, Vines & Kane, 2009), especially animals, but reports in flowering plants are still few. This is probably linked to the low variability observed in most markers used for molecular phylogenetics, which has also made DNA barcoding more difficult in plants than in animals (Chase *et al.*, 2007). Nevertheless, methodologies used in population genetics would be expected to have revealed more cases of cryptic species in plants if they were common.

New Caledonia is an archipelago in the southwestern Pacific well known for its rich, unique and endangered flora, hence its recognition as a biodiversity hotspot for conservation priorities (Myers *et al.*, 2000). It is often considered that the richness of this flora is linked to

the diversity of edaphic conditions found on the main island (Jaffré, 1993). About one-third of the main island is covered with ultramafic soils (Pelletier, 2006), which have several characteristics that are limiting for plant growth: low fertility (low N, P, K), high Mg/Ca ratio, high concentrations of heavy metals (Co, Cr, Ni etc.), and low water-holding capacity; in addition, fire is more frequent in vegetation on such substrates (Proctor, 2003). These soils, including those that are derived from "true" serpentine, usually present a challenging environment for plant development, but they are nevertheless covered by a species-rich vegetation in New Caledonia, with a high level of endemism, exceeding 90% (Jaffré *et al.*, 2001). Other soil types in New Caledonia are often referred to as non-ultramafic, and they develop from rocks that include limestone, basalt, shale, etc. This basic dichotomy between ultramafic and non-ultramafic soils is a simple way to define the ecology of plant species because the majority are found on one or the other type, even though this may imply a disjunct or fragmented distribution. For instance, Jaffré *et al.* (1987) calculated that 1176 species of seed plants were restricted to ultramafic substrates (see also Morat, 1993).

Nevertheless, species naturally found on both types of soils are not uncommon. Within Cunoniaceae, a family that has radiated in New Caledonia, some species with broad soil tolerance are monomorphic, e.g. *Geissois hirsuta* Brongn. & Gris, whereas others are complex and present considerable taxonomic challenges, e.g. *Pancheria ternata* Brongn. & Gris and *Geissois pruinosa* Brongn. & Gris (Hopkins & Pillon, unpublished manuscript). While investigating the molecular systematics of the genus *Spiraeanthemum*, also in the Cunoniaceae, we wanted to assess the validity of the circumscription of some species with wide substrate preferences. This genus encompasses 18 species of trees and shrubs distributed from the Moluccas to Samoa, including New Guinea, Queensland, the Solomon Islands, Vanuatu, Fiji, and New Caledonia, the last having the largest number of species (Hoogland, 1979; Hoogland, 1987; Pillon *et al.*, 2009a). The genus can be divided into three groups

(Pillon *et al.*, 2009a), two of which occur in New Caledonia. The first (the *S. densiflorum* group) has a single New Caledonian species, *S. densiflorum* Brongn. & Gris, which is characterised by hairy apical buds, inflorescences branched near the base and several ovules per carpel; other species in this group are found in Australia and New Guinea. The *S. brongniartianum* group, characterised by a single ovule per carpel and pedunculate inflorescences, comprises six species in New Caledonia plus one, *S. vitiense* A. Gray, in Fiji. The third group is *Spiraeanthemum sensu stricto* (*S. samoense* group), which occurs on several Pacific islands but not New Caledonia; its species have hairy apical buds, opposite leaves, pedunculate inflorescences and two ovules per carpel.

Of the six species in the *Spiraeanthemum brongniartianum* group in New Caledonia (Fig. 1), *S. meridionale* (Hoogland) Pillon and *S. pedunculatum* Schltr. are restricted to ultramafic soils in the south; *S. collinum* (Hoogland) Pillon has a scattered distribution on a few ultramafic massifs in the centre; *S. brongniartianum* Schltr. is widespread on non-ultramafic soils; and *S. pubescens* Pamp. is also widespread but occurs on both soil types. Finally, *S. ellipticum* Pamp. is known from two mountains in the north on non-ultramafic soils, whereas in the south it occurs on several ultramafic ones, and it is unknown in between. The disjunction in the range and ecology of *S. ellipticum* raises doubts about its circumscription.

In this study, we aimed to assess the circumscription of species in *Spiraeanthemum* using molecular markers. We selected for this purpose two nuclear low-copy genes, *ncpGS* and *gapC*, which have already been used successfully in Cunoniaceae (Pillon *et al.*, 2009a; Pillon *et al.*, 2009b). Low-copy nuclear genes are generally more informative than the most commonly used markers such as plastid or nuclear ribosomal loci because of their higher variability, biparental inheritance and general absence of concerted evolution (Mort & Crawford, 2004; Small, Cronn & Wendel, 2004). Previous works have shown that plastid

genes display low levels of variation within genera of Cunoniaceae (Bradford, 2002, Pillon, unpublished), including within the *S. brongniartianum* group (Pillon *et al.*, 2009a) and that multiple paralogues of ITS are recovered in several genera (Bradford, 2002, Pillon, unpublished).

MATERIALS & METHODS

All currently accepted species in the genus *Spiraeanthemum* (see Pillon *et al.*, 2009a) in New Caledonia were sampled. The main focus of this study was the *S. brongniartianum* group in which multiple accessions per species were used whenever possible. Accessions of *S. densiflorum* (*S. densiflorum* group, New Caledonia) and *S. macgillivrayi* Seem. (*Spiraeanthemum* s.s., Vanuatu) were used as outgroups. A previous study showed that the *S. brongniartianum* group was well-supported and that *S. macgillivrayi* and *S. densiflorum* were its successive sister groups, a set of relationships that was strongly supported (Pillon *et al.*, 2009a). No material of *S. vitiense* (Fiji), the only species of the *S. brongniartianum* group occurring outside New Caledonia, was available. Leaf material was dried in the field in silica gel (Chase & Hills, 1991), and DNA was extracted with a CTAB protocol modified from Doyle and Doyle (1987) by Csiba and Powell (2006); the list of accessions is given in the appendix.

A portion of chloroplast-expressed glutamine synthetase (*ncpGS*), a low-copy nuclear gene, was amplified using the primers 687F and 994R of Emshwiller & Doyle (1999). The amplification mix included 45μL of 2.5mM $MgCl_2$ Reddy PCR Master Mix 1.1× (ABgene Ltd), 1μL of 0.4% bovine serum albumin (BSA), 0.5μL of each primer (100ng/μL) and 3μL of template DNAs of unknown concentrations. The following amplification program was used: 2 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 50°C, 1 min 30 s at 72°C and a final extension of 5 min at 72°C. A portion of the glyceraldehyde 3-phosphate dehydrogenase

(*GapC*, also known as *G3pdh*), another low-copy nuclear gene was amplified using the primers GPDX7F and GPDX9R from Strand *et al.* (1997) with the same amplification mix as for *ncpGS*. The following program was used: 2 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 53°C, 1 min at 72°C and a final extension of 5 min at 72°C.

No chromosome counts are available for *Spiraeanthemum*, but diploid chromosome numbers in other genera of Cunoniaceae vary between 24 and 32 (Hopkins & Hoogland, 2002). In *Populus trichocarpa* Torr. & A. Gray (Salicaceae), the most closely related model group to Cunoniaceae for which the entire genome has been sequenced (Tuskan *et al.*, 2006), one homeologue of *GapC* is found in linkage group I and homeologs of *ncpGS* are found in linkage groups VIII and X. Thus, the two genes are probably not linked in *Spiraeanthemum*.

If the electropherorams of directly sequenced amplification products of *ncpGS* indicated the presence of more than one allele in an accession and if these alleles differed in length or in more than one base, the PCR product was cloned. An average of six clones was routinely sequenced. Further clones were sequenced if all the base diversity observed in the electropherograms from the direct sequencing was not recovered. Satisfactory results from direct sequencing of *GapC* were not obtained, and thus amplification products from all accessions were cloned and sequenced (up to six clones per accession). For both genes, a consensus sequence was built for each cluster of clones on the phylogenetic tree in order to remove substitutions produced during PCR; these were compared with electropherograms from the direct sequencing when readable. PCR-induced recombinant sequences were detected by eye (with the help of the phylogenetic tree) and excluded. Sequences were submitted to Genbank (accession numbers GQ332607 to GQ332639), and full details are provided in the appendix.

Alignment was carried out easily by eye in a PAUP matrix, and parsimony analysis were carried out with PAUP*4.01b10 (Swofford, 2002) with DELTRAN character state

optimisation. For *GapC*, the most parsimonious trees were recovered using branch-and-bound searches, and node support was measured with 1000 replicates of bootstrap, also with branch-and-bound searches. For *ncpGS*, we used heuristic searches due to the larger size of the matrix, with TBR branch-swapping and MULTREES option on; node support was measured with 100 replicates using the same settings. Maximum likelihood analyses were carried out using PHYML (Guindon & Gascuel, 2003; Guindon *et al.*, 2005) with a HKY model. This model best fitted all datasets according to Modeltest analyses (Posada & Crandall, 1998), except for the *GapC* dataset with recombinant sequences excluded, for which the K81uf model had a slightly lower AIC. Recombination networks (Huson & Kloepper, 2005) were reconstructed using Splitstree4 (Huson & Bryant, 2006) with default settings. Outgroups were not included in the networks since their inclusion leads to graphical distortion because they are on a long branch and we wanted to emphasize the ingroup (i.e. the *S. brongniartianum* group).

To test for the occurrence of recombination in our datasets, we used GARD software (Kosakovsky Pond *et al.*, 2006) with a HKY85 substitution model, assuming no rate variation among sites. The latter setting proves the most sensitive considering the relatively low level of polymorphism observed in our datasets. Recombinant sequences were then identified by examination of the alignments. Further putative recombinant sequences were identified by progressive exclusion of sequences from the dataset until no crossovers were observed in the recombination networks. We then ran the phylogenetic analysis again excluding all putative recombinant alleles.

RESULTS

Detailed information on the *ncpGS* and *GapC* datasets is given in table 1. No sequences of *GapC* were obtained for *Spiraeanthemum meridionale*, but otherwise both genes

were successfully amplified and sequenced for at least one accession of each New Caledonian species of *Spiraeanthemum*. Direct sequencing revealed that the accession *Pillon et al. 20* of *S. ellipticum* had at least two alleles, but only one was successfully recovered by cloning. Fewer accessions were sequenced for *GapC* because direct sequencing was not possible, and cloning was necessary in each case. We detected six likely PCR-induced recombinant sequences out of 59 clones (10.3%) in the *ncpGS* datset, and four out of 31 (12.9%) in the *GapC* dataset.

With the GARD analyses, several cases of recombination were detected in both *ncpGS* and *GapC* datasets. Identification of recombinant sequences required a detailed examination of the alignment, and by eye we detected six recombinant sequences in the *ncpGS* dataset and one in the *GapC* dataset. After exclusion of these recombinant sequences, GARD analyses no longer detected recombination in the *ncpGS* dataset, although they still did for *GapC* . The recombination networks (see below) indicate two further recombinant sequences in the *ncpGS* dataset and one in *GapC*. After exclusion of these recombinant sequences from the *GapC* dataset, recombination was still detected, but examination of the alignment indicates that recombination is not likely and that this may be a false positive result. Table 2 provides the base composition of the different accessions at the parsimony-informative sites for *ncpGS*.

Phylogenetic analyses of *GapC* sequences produce the same topology with both maximum parsimony and maximum likelihood, so only the first is shown (figure 2). In this phylogenetic tree, two groups can be distinguished. Group A consists of the northern accessions of both *Spiraeanthemum pubescens* and *S. ellipticum*, plus one of two alleles of *S. collinum* (allele 1). Group B contains *S. brongniartianum*, the southern accession of *S. ellipticum*, and allele 2 of *S. collinum*, which may be a recombinant sequence. For *S. pedunculatum*, allele 2 also belonged to group B, but allele 1 showed evidence of

recombination around position 430; the first part of the alignment (positions 1-429) clustered with group A and the second part (positions 430-910) with group B.

In the *ncpGS* analysis, a well-supported group A was recovered in both maximum parsimony and maximum likelihood analyses (figure 3), containing the two northern accessions of both *Spiraeanthemum ellipticum* and *S. pubescens.* Group B formed a polytomy with group 1 nested in it and with little internal resolution; it contained the two accessions of *S. brongniartianum,* the single accession of *S. collinum*, the two southern accessions of *S. ellipticum,* both accessions of *S. meridionale,* both of *S. pedunculatum* and the single accession of southern *S. pubescens*. Several nodes were not present in the strict consensus of the maximum parsimony analysis, and these nodes were supported in the maximum likelihood analysis. We tentatively recognised five subgroups, B1 to B5, excluding the putative recombinant sequences. B1 is recovered in both analyses as part of the sister group to group A. Groups B2 and B3 were each supported in the maximum likelihood analysis with 85 and 76 bootstrap percentages, respectively. Putative recombinant sequences either formed clusters with subgroups B1 to B5 or independent clades.

When all putative recombinant sequences were excluded from the *ncpGS* analysis, a better resolved phylogenetic tree was obtained using maximum likelihood (figure 4), with higher bootstrap percentages supporting group A and of subgroups B2 and B3. Group B was recovered (85 bootstrap percentage) but on a short branch. The topology obtained with the maximum parsimony analysis was the same except that group B was not monophyletic; subgroup B5 was recovered as sister to group A, but with low bootstrap support (<50%).

Recombination networks based on *ncpGS* and *GapC* sequence data are shown in figure 5. Groups A and B are present in both cases. In the *GapC* analysis, two crossovers were detected involving groups A and B, whereas in the *ncpGS* analysis, all crossovers detected were within group B, and almost no structure was observed within the latter, apart from

subgroups B2 and B3. When outgroups were included, the rooting was imprecise because of the recombination events within the ingroup, but the outgroup tended to attach between groups A and B.

DISCUSSION

RECOMBINATION

Recombination events seem to have been common in New Caledonian *Spiraeanthemum*, especially in *ncpGS*, but also in *GapC*, which was less well sampled. Because of the frequency of the phenomenon, it is difficult to tease apart clearly recombinant from non-recombinant sequences. Similarly high numbers of recombinant sequences were detected in *Nicotiana* (Solanaceae; Kelly *et al.*, submitted). Recombination networks also detected putative cases of recombination not found by visual inspection of the alignments. As many as eight cases of recombination for *ncpGS* are suggested here, two of which possibly involved two breakpoints each. Some recombination may be associated with PCR and revealed only because of cloning (these recombinations would be unlikely to be observed in direct sequencing because they would be a minority signal); however, most alleles were cloned more than once. Examination of the direct sequences of *S. meridionale* shows that all differences between the two alleles occur in the last 163 base pairs, so in this case at least recombination is not an artefact; if differences are evident in direct sequencing, they are probably present in roughly 50% of the PCR copies, which is too frequent for them to be the result of PCR alone. Furthermore, no evidence of recombination was found in a phylogenetic study of *Codia* (Cunoniaceae) using the same set of markers and carried out at the same time (Pillon *et al.*, 2009b), although several species of putative hybrid origins displayed multiple alleles.

It is not clear why recombination in *ncpGS* is so common in *Spiraeanthemum*. Many more cases were observed for this gene than for the less well sampled *GapC* (only two instances), and furthermore *ncpGS* is longer, so recombination events are more likely to occur. With a comparable level of variability, *ncpGS* offers more variable sites, which makes recombination easier to detect. We are not aware of cases of recombination for either gene reported previously in the literature. A possible explanation for the high numbers of recombination events is frequent hybridisation during the history of the genus (see below).

A first consequence of recombination is that it affects reconstruction of phylogenetic trees by increasing homoplasy. Recombination is expected to reduce the accuracy of phylogenetic analyses (Posada & Crandall, 2002), and indeed a large number of equally parsimonious trees were recovered, making the consensus tree highly unresolved. The exclusion of the recombinant sequences gives a much better resolved phylogenetic tree with good to strong support for most nodes, although some questions remain (e.g. monophyly of group B).

Within group B, most alleles recovered in the *npGS* analysis could either be assigned to one of four subgroups (B2, B3, B4 or B5) or were clearly recombinants between two alleles of these subgroups, i.e. they had no unique substitutions. However, allele 1 of *S. collinum* was assigned to subgroup B1 and did not show any signs of recombination when examined by eye; the recombination network indicates that it may nevertheless be recombinant. Several other alleles shared unique substitutions with it and have been assumed to be recombinant between this B1 allele and one from another subgroup. Thus, we may not have found the genuine B1 allele (i.e. with no recombination), and possibly the original sequence has been lost due to recombination. In general, the frequent occurrence of recombination may be homogenising the allele pool through the loss of some alleles.

Identifying all the recombinant sequences was not straightforward; GARD analysis and recombination networks were not entirely in agreement about recombination. Schierup and Hein (2000) predicted that phylogenetic trees built on datasets including recombinant sequences would be more star-like with long terminal branches. The exclusion of recombinant sequences from the *ncpGS* dataset did reduce the number of long branches and most terminals fell in one of the major allele groups (groups A, B2, B3, in figure 5). Denser sampling should break up long branches and may help to clarify whether some putative cases of recombination detected are false and if other cases have been undetected.

To date few phylogenetic studies of plants have reported recombination in low-copy nuclear genes, possibly because this issue is not systematically explored, though exceptions include Cinnamoyl CoA reductase in *Eucalyptus* (Myrtaceae, Poke *et al.*, 2006), *PHOT* (phototropin) in *Verbena* (Verbenaceae, Yuan & Olmstead, 2008), and *ncpGS*, *alcohol dehydrogenase* and *LEAFY/FLORICAULA* in *Nicotiana* (Solanaceae, Clarkson *et al.*, submitted; Kelly *et al.* submitted). Recombination will have to be considered as a potential pitfall of using low-copy nuclear genes in future phylogenetic studies.

EVIDENCE FOR THE EXISTENCE OF CRYPTIC SPECIES

Despite recombination, both analyses suggest the existence of two groups of alleles within *Spiraeanthemum*, A and B. Their distinctiveness is supported by a clear phylogenetic distance between them in both the *ncpGS* and *GapC* analysis. Group A is recovered in both, but group B is only well supported in the *GapC* analysis. In the *ncpGS* analysis, the accessions of northern *S. ellipticum* and *S. pubescens* both belong to group A, and those of southern *S. ellipticum* and *S. pubescens* cluster in group B. For *GapC*, where fewer samples were available, the single northern accessions of *S. ellipticum* and *S. pubescens* both fall into clade A, whereas the single accession of southern *S. ellipticum* is found in clade B. Thus,

southern and northern populations of *S. ellipticum* and of *S. pubescens* appear to be genetically distinct. This is particularly clear for *S. ellipticum*, for which distinct northern and southern populations were observed in both gene trees, and two accessions in each geographical areas had a congruent placement in the *ncpGS* tree.

In terms of morphology, *Spiraeanthemum ellipticum* and *S. pubescens* are clearly distinct from one another and from other members of the genus. *Spiraeanthemum ellipticum* has glabrous resinous apical buds and sessile obovate leaves, whereas *S. pubescens* has hairy apical buds and lanceolate leaves with a distinct petiole. Both species can be shrubs or trees in scrub or forest, but *S. ellipticum* is restricted to high elevation (mostly 900-1600m), whereas *S. pubescens* generally occurs at low to medium elevation (mostly 0-700m). No morphological differences that could distinguish the northern and southern populations of either of these species have been found, and the main difference is in their ecology: the southern populations of both are found on ultramafic soils, and the northern populations on non-ultramafic substrates. The lack of morphological differences that correlate with geography, ecology and the patterns observed in molecular markers suggests that these may be examples of cryptic species.

Southern and northern populations of *Spiraeanthemum ellipticum* cannot be considered as sibling species (sensu Bickford *et al.*, 2007) because they are not each other's closest relatives according to the phylogenetic analyses presented here (figures 2 & 3). The large differences between alleles recovered in morphologically identical populations require an explanation. It would be puzzling if the same morphologies had evolved in parallel in groups A and B (figure 6, top diagram), but this is not the only possible scenario nor is it, in our view, the most likely.

An alternative hypothesis is that alleles of both groups A and B were originally present in the pool that included the common ancestors of both *Spiraeanthemum ellipticum* and *S.*

pubescens; parallel lineage sorting in northern and southern populations then fixed alleles of groups A and B as observed here (Fig. 6, middle diagram). This could have occurred through either genetic drift or possibly natural selection driven by differences in the ecological conditions prevailing on non-ultramafic versus ultramafic soils. However, the same pattern of lineage sorting between two a priori unlinked genes seems unlikely, since lineage sorting is stochastic, but it can not be excluded.

A third possibility is that either the southern or the northern population, or both, underwent hybridisation and introgression (Fig. 6, bottom diagram). In this scenario, one or both populations would have acquired alleles from another species growing in the vicinity, and these new alleles would have replaced the ancestral ones of this species. Thus transgression within group A and/or group B would have obscured the histories of both *Spiraeanthemum ellipticum* and *S. pubescens*. If it seems improbable that such a process occurred independently in both regions producing by chance in these patterns, this scenario may be over-simplified because the other southern species with B alleles (*S. meridionale* and *S. pedunculatum*) may have also been involved. There are indeed similarities between some specimens of southern *S. pubescens* and *S. meridionale*, suggesting that species boundaries in these taxa may be porous (Lexer *et al.*, 2009). Whatever the process involved, congruent allelic differences at two presumably unlinked loci between northern and southern populations of *S. ellipticum* lead us to conclude that they are now distinct but cryptic species.

IMPLICATIONS FOR THE EVOLUTION OF THE NEW CALEDONIAN FLORA

Overall, ecology correlated better with our molecular results than does morphology. In the tree based on maximum likelihood (Fig. 4), group A includes only accessions from populations occurring in the north on non-ultramafic soils (*S. ellipticum* p.p., *S. pubescens*

p.p.), and all accessions growing on ultramafic soils have alleles belonging to group B. According to the *ncpGS* phylogenetic tree from which recombinant sequences were excluded, adaptation to ultramafic soils may have evolved only once in *Spiraeanthemum*, in the clade containing groups B2, B3, and B5. This corroborates our findings in *Codia* (Pillon *et al.*, 2009b), which showed that adaptation to ultramafic soils may not be as homoplasious as suggested by de Kok (2002) in other genera.

The complex geology of New Caledonia strongly constrains distributions of plants. Most species are either restricted to ultramafic or non-ultramafic soils, and thus their distributions are generally expected to match one region on the geological map. Species recorded from both soil types represent a small portion of the flora. In Cunoniaceae, ca. 57% of the species are restricted or almost so to ultramafic soils, 28% are restricted to non-ultramafic soils, and only 13% occur on both (Pillon & Hopkins, unpublished). The reported occurrence in taxonomic revisions of species on both ultramafic and non-ultramafic substrates may in some cases be due to misidentification, inadequate label data, or poor species delimitation, although complex local ecological and geological conditions or genuine ecological plasticity is likely to be involved in some instances.

The occurrence of two species of *Spiraeanthemum*, *S. ellipticum* and *S. pubescens*, that are both ecologically and genetically heterogeneous suggests that ecology should be given greater weight when evaluating species boundaries. The New Caledonian flora has largely been monographed using a strictly morphological definition for species, especially in older treatments that preceded most of the ecological studies of this vegetation. The presence of cryptic entities that correlate with soil preferences within a morphologically coherent "species" suggests that fidelity to soils may be higher than thought in the flora of New Caledonia. In *Codia*, species that occur on both soil types all have a hybrid origin (Pillon *et al.*, 2009b), and it is possible that such hybrid species were formed on multiple occasions.

Two of these species (*C. montana* J.R. Forst. & G. Forst. and *C. albicans* Vieill. ex Pamp.) are clearly heterogeneous, but their patterns of morphological variation cannot be readily explained. The same is probably also true for some ecologically variable species in *Geissois*, *Pancheria* and *Weinmannia* (Hopkins & Pillon, unpublished manuscript).

If morphological entities with broad ecological preferences are actually more than one species, then the total number of taxa in the New Caledonian flora may be currently underestimated. It would be valuable to investigate more thoroughly the species with broad ecological preferences because in some cases less obvious morphological variation may correlate with ecology. Such an approach has been applied to *Planchonella* (Sapotaceae) in combination with molecular work (Swenson, Munzinger & Bartish, 2007), and the number of species recognised has significantly increased since the previous revision by Aubréville (1967). The application of genetics to the New Caledonian flora is thus likely to lead to a new form of "taxonomic inflation" (Isaac, Mallet & Mace, 2004). However, it is unclear what proportion of ecologically variable taxa contains multiple species and therefore what is the contribution of cryptic species to the total diversity of the New Caledonian flora.

WHAT IS THE TRUE *SPIRAEANTHEMUM ELLIPTICUM*?

No obvious morphological characters distinguish southern and northern populations in either *Spiraeanthemum ellipticum* or *S. pubescens*. We cannot rule out the possibility that some morphometric or other more detailed study will identify differences in the future, but in general *Spiraeanthemum* offers relatively few morphological characters that differ at species level compared with *Cunonia* or *Pancheria*, for example. The three collections available from northern populations of *S. ellipticum* are inadequate for morphometric analysis as none is in

full flower, and access to the only two localities known in the north (Mont Panié and Mont Colnett) is difficult.

The type of *Spiraeanthemum ellipticum* is *Vieillard 2643* (lectotype: G ex DC, dated 1865; isolectotypes: A, BM, G, GH, K, L, MEL, NY, P, and Z, fide Hoogland 1979). The label gives the collection locality as "Fond de la rivière de Hienguen" (far end of the Hienghène river, presumably meaning the headwaters), which is south of Mont Panié, and no elevation is given. The sheets we have seen at K and P have one to seven fragments comprising leaves, flowers and/or fruits.

However, some doubt exists about the locality and as to whether *Vieillard 2643* represents a single collection. Problems with locality data and mixed collections are common for Vieillard's specimens; his localities were often imprecise, and each of his collection numbers refers to material that he regarded as belonging to the same species and not necessarily to one particular gathering. Nevertheless, Hoogland considered the fragments on different sheets of *Vieillard 2643* were sufficiently similar for him to cite them all as part of the type.

We are doubful about Vieillard's locality because no recent collections of this species have been made from the upper Hienghène river and the maximum elevation of the mountains on either side of the head of the valley (ca. 800 m) is lower than those where the plant has been collected on Monts Panié and Colnett (1000-1200 m, based on *Pillon 347 & 616*, *Schmid 4136*). If the locality is incorrect, it must then be doubtful whether the type came from the northern population of *Spiraeanthemum ellipticum* because we have no firm evidence that Vieillard ever visited Mont Panié or Mont Colnett.

Thus we cannot be sure from the label data that *Vieillard 2643* is in fact from the northern population of *Spiraeanthemum ellipticum*, and we did not attempt to determine its

likely origin by DNA analysis. DNA extraction from this old specimen would be challenging, and successful amplification of a low-copy nuclear gene needed to distinguish between northern and southern populations seems unlikely.

Because no morphological characters currently distinguish southern and northern populations of *Spiraeanthemum ellipticum* and because of problems with the type, we cannot alter its delimitation at present. Nevertheless, for conservation purposes it should be kept in mind that the populations in the two areas represent the results of distinct evolutionary events. This taxon is restricted to high elevation and may thus be considered threatened because of climate change, since the area available for it to occupy will decrease as vegetation zones move up the mountains due to global warming (Pounds, Fogden & Campbell, 1999; Thomas *et al.*, 2004; Thomas, Franco & Hill, 2006; Kelly & Goulden, 2008; Munzinger, McPherson & Lowry, 2008); and the northern population, currently known from only two mountains, is particularly at risk.

Our sampling of *Spiraeanthemum pubescens* was not adequate to draw firm conclusions about whether more than one taxon should be recognised. Because no clear geographical gap exists between northern and southern populations, further sampling is needed, especially in the central part of its distribution.

EVIDENCE FOR HYBRIDISATION

The frequency of recombination in *Spiraeanthemum* suggests that hybridisation has probably been important in the history of this genus. For both genes investigated, we found clearly distinct alleles (excluding recombinants), suggesting that a period of genetic isolation

allowed them to develop and become fixed in some populations. Thus, we think that recombination events occurred during secondary contact between species/populations.

In the *GapC* analysis, the single accession of *Spiraeanthemum collinum* and the accession of *S. pedunculatum* from Bois du Sud each displayed alleles or portion of alleles that belong to both groups A and B, suggesting a hybridisation event between these two population groups. However, this is not reflected in the *ncpGS* analysis where there was neither evidence of recombination events between alleles of group A and B nor any accession exhibiting alleles of both groups. However, both accessions of *S. collinum* and of *S. pedunculatum* displayed a mixture of alleles of subgroup B1, B2 and B3 in the *ncpGS* analysis. The true (ancestral) B1 allele may not have been recovered during this study, but the phylogenetic analysis suggests that it may be closer to group A than to the rest of group B. Thus, these two accessions also appear to have a combination of alleles from group B and something close to group A, which would be a congruent scenario for the two datasets. Furthermore, this would imply that some species might have had a hybrid origin. Similarly, one accession of *S. ellipticum* (Montagne des Sources) and one of *S. pedunculatum* (Pic du Pin) each produced one allele of subgroup B2 and one of subgroup B3, and it is also possible that these two species had hybrid origins. If this were the case, *Spiraeanthemum pedunculatum* might therefore have had two separate origins.

Hybrid speciation has been suspected in several other genera of Cunoniaceae from New Caledonia (Pillon, Hopkins & Bradford, 2008; Pillon *et al.*, 2009b), but considering that recombination has seriously obscured the history of *Spiraeanthemum* and that uncertainty surrounds the allele subgroup B1, hybrid speciation should only be considered as a possibility for this genus. Chromosome counts would also be valuable since they could provide information on the occurrence of polyploids, which are often associated with hybrid

speciation (Mallet, 2007), particularly when parental taxa are distantly related (Paun *et al.*, 2009).

ACKNOWLEDGMENTS

The authors wish to thank the Direction de l'environnement (DENV, Province Sud) and Direction du développement économique et de l'environnement (DDEE, Province Nord) for allowing scientific collections in New Caledonia. We thank staff and students of the Botanical Laboratory at IRD for assistance in the field. Daniel and Irène Létocart, the late Henri Blaffart and the guides of the Dayu-Biik association, Steve Farino and Maurice Wanguene, also provided important assistance in the field and the herbarium. We thank the organisers of the Santo 2006 expedition in Vanuatu (IRD/MNHN/PRO-NATURA) during which material of *Spiraeanthemum macgillivrayi* was collected. Jason Bradford kindly provided some leaf material. We are grateful to Edith Kapinos, Laszlo Csiba, Olivier Maurin, Dion Devey, and Martyn Powell for their help in the Jodrell Laboratory and to Laura Kelly, Nina Rønsted, and Jim Clarkson for discussions about low-copy nuclear genes. Anaïs Cardineau and Cyril Poullain (CNRS) provided access to a Macintosh to carry out data analyses. Two anonymous reviewers made valuable comments on an earlier draft of the manuscript.

REFERENCES

Aubréville, A. 1967. Sapotacées. In: Aubréville, A ed *Flore de la Nouvelle-Calédonie et Dépendances.* Volume 1. Paris: Muséum National d'Histoire Naturelle.

Bickford, D, Lohman, DJ, Sodhi, NS, Ng, PKL, Meier, R, Winkler, K, Ingram, KK, Das, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22:** 148-155.

Bradford, JC. 2002. Molecular phylogenetics and morphological evolution in Cunonieae (Cunoniaceae). *Annals of the Missouri Botanical Garden* **89:** 491-503.

Chase, MW, Hills, HG. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* **40:** 215-220.

Chase, MW, Cowan, RS, Hollingsworth, PM, van den Berg, C, Madriñan, S, Petersen, G, Seberg, O, Jørgensen, T, Cameron, KM, Carine, M, Pedersen, N, Hedderson, TAJ, Conrad, F, Salazar, GA, Richardson, JE, Hollingsworth, ML, Barraclough, TG, Kelly, L, Wilkinson, M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* **56:** 295-299.

Csiba, L, Powell, MP. 2006. DNA extraction protocols. In: Savolainen, V, Powell, MP, Davis, K, Reeves, G, Corthals, A eds. *DNA and tissue banking for biodiversity and conservation: theory, practice and uses.* Pp. 48-51. Richmond: Royal Botanic Gardens, Kew.

de Kok, R. 2002. Are plant adaptations to growing on serpentine soil rare or common? A few case studies from New Caledonia. *Adansonia* **24:** 229-238.

Doyle, JJ, Doyle, JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19:** 11-15.

Emshwiller, E, Doyle, JJ. 1999. Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution* **12:** 310-319.

Guindon, S, Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696-704.

Guindon, S, Lethiec, F, Duroux, P, Gascuel, O. 2005. PHYML online─a web server for fast maximum likelihood-base phylogenetic inference. *Nucleic Acids Research* **33:** W557-W559.

Guruschidze, M, Fritsch, RM, Blattner, FR. 2008. Phylogenetic analysis of *Allium* subg. *Melanocrommyum* infers cryptic species and demands a new sectional classification. *Molecular Phylogenetics and Evolution* **49:** 997-1007.

Hoogland, RD. 1979. Studies in the Cunoniaceae. II. The genera *Caldcluvia*, *Pullea*, *Acsmithia* and *Spiraeanthemum*. *Blumea* **25:** 481-505.

Hoogland, RD. 1987. Studies in the Cunoniaceae. IV. Further notes on New Caledonian *Acsmithia*. *Adansonia* **4:** 393-397.

Hopkins, HCF, Hoogland, RD. 2002. Cunoniaceae. In: Noteboom, HP ed *Flora Malesiana*, Series I, Volume 16: pp 53-165. Leiden: Nationaal Herbarium Nederland.

Huson, DH, Kloepper, TH. 2005. Computing recombination networks from binary sequences. *Bioinformatics* **21:** ii159-ii165.

Huson, DH, Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254-267.

Isaac, NJB, Mallet, J, Mace, GM. 2004. Taxonomic inflation: its influence on macroecology and conservation. *Trends in Ecology and Evolution* **19:** 464-469.

Jaffré, T, Morat, P, Veillon, J-M, MacKee, HS. 1987. Changements dans la végétation de la Nouvelle-Calédonie au cours du Tertiaire : la végétation et la flore des roches ultrabasiques. *Adansonia* **4:** 365-391.

Jaffré, T. 1993. The relationship between ecological diversity and floristic diversity in New Caledonia. *Biodiversity Letters* **1:** 82-87.

Jaffré, T, Morat, P, Veillon, J-M, Rigault, F, Dagostini, G 2001. *Composition and characterisation of the native flora of New Caledonia*. Documents Scientifiques et

Techniques, II 4. Nouméa: Institut de Recherche pour le Développement. www.ird.nc/images/Jaffr%E9%20et%20al,%202001-flore%20nc.pdf

Kelly, AE, Goulden, ML. 2008. Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Science of the USA* **105:** 11823-11826.

Kosakovsky Pond, SL, Posada, D, Gravenor, MB, Woelk, CH, Frost, SDW. 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22:** 3096-3098.

Lexer, C, Joseph, J, Van Loo, M, Prenner, G, Heinze, B, Chase, MW, Kirkup, D. 2009. The use of digital image-based morphometrics to study the phenotypic mosaic in taxa with porous genome. *Taxon* **58:** 349-364.

Mallet, J. 2007. Hybrid speciation. *Nature* **446:** 279-283.

Martínez-Ortega, M, Delgado, L, Albach, DC, Elena-Rosselló, JA, Rico, E. 2004. Species boundaries and phylogeographic patterns in cryptic taxa inferred from AFLP markers: *Veronica* subgen. *Pentasepalae* (Scrophulariaceae) in the western Mediterranean. *Systematic Botany* **29:** 965-986.

Morat, P. 1993. Our knowledge of the flora of New Caledonia: endemism and diversity in relation to vegetation types and substrates. *Biodiversity Letters* **1:** 72-81.

Mort, ME, Crawford, DJ. 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon* **53:** 257-261.

Muellner, AN, Pennington, TD, Chase, MW. 2009. Molecular phylogenetics of Neotropical Cedreleae (mahogany family, Meliaceae) based on nuclear and plastid DNA sequences reveal multiple origins of "*Cedrela odorata*". *Molecular Phylogenetics and Evolution* **52:** 461-469.

Munzinger, J, McPherson, G, Lowry II, PP. 2008. A second species in the endemic New Caledonian genus *Gastrolepis* (Stemonuraceae) and its implications for the

conservation status of high-altitude maquis vegetation: coherent application of the IUCN Red List criteria is urgently needed in New Caledonia. *Botanical Journal of the Linnean Society* **157:** 775-783.

Myers, N, Mittermeier, RA, Mittermeier, CG, da Fonseca, GAB, Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403:** 853-858.

Paun, O, Forest, F, Fay, MF, Chase, MW. 2009. Hybrid speciation in angiosperms: parental divergence drives ploidy. *New Phytologist* **182:** 507-518.

Pelletier, B. 2006. Geology of the New Caledonia region and its implications for the study of the New Caledonian biodiversity. In: Payri, CE, Richer de Forges, B eds. *Compendium of marine species from New Caledonia.* Documents Scientifiques et Techniques, II 7 : 17-30. Nouméa: Centre IRD de Nouméa. www.ird.nc/biodec/downloads/Compendium/Version%20v%E9rrouill%E9e/Pelletier-v.pdf

Pillon, Y, Qamaruz-Zaman, F, Fay, MF, Hendoux, F, Piquot, Y. 2007. Genetic diversity and ecological differentiation in the endangered fen orchid (*Liparis loeselii*). *Conservation Genetics* **8:** 177-184.

Pillon, Y, Hopkins, HCF, Bradford, JC. 2008. Two new species of *Cunonia* (Cunoniaceae) from New Caledonia. *Kew Bulletin* **63:** 419-431.

Pillon, Y, Hopkins, HCF, Munzinger, J, Chase, MW. 2009a. A molecular and morphological survey of generic limits of *Acsmithia* and *Spiraeanthemum* (Cunoniaceae). *Systematic Botany* **34:** 141-148.

Pillon, Y, Munzinger, J, Amir, H, Hopkins, HCF, Chase, MW. 2009b. Reticulate evolution on a mosaic of soils: diversification of the New Caledonian endemic genus *Codia* (Cunoniaceae). *Molecular Ecology* **18:** 2263-2275.

Poke, FS, Martin, DP, Steane, DA, Vaillancourt, RE, Reid, JB. 2006. The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). *Molecular Phylogenetics and Evolution* **39:** 160-170.

Posada, D, Crandall, KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817-818.

Posada, D, Crandall, KA. 2002. The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* **54:** 396-402.

Pounds, JA, Fogden, MPL, Campbell, JH. 1999. Biological response to climate change on a tropical mountain. *Nature* **398:** 611-615.

Proctor, J. 2003. Vegetation and soil and plant chemistry on ultramafic rocks in the tropical Far East. *Perspectives in plant ecology, evolution and systematics* **6:** 105-124.

Rieseberg, LH, Vines, T, Kane, N. 2009. Editorial and retrospective 2008. *Molecular Ecology* **18:** 1-20.

Schierup, MH, Hein, J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156:** 879-891.

Small, RL, Cronn, RC, Wendel, JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17:** 145-170.

Sotuyo, S, Delgado-Salinas, A, Chase, MW, Lewis, GP, Oyama, K. 2007. Cryptic speciation in the *Caesalpinia hintonii* complex (Leguminosae: Caesalpinioideae) in a seasonally dry Mexican forest. *Annals of Botany* **100:** 1307-1314.

Strand, AE, Leebens-Mack, J, Milligan, BG. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* **6:** 113-118.

Swenson, U, Munzinger, J, Bartish, IV. 2007. Molecular phylogeny of *Planchonella* (Sapotaceae) and eight new species from New Caledonia. *Taxon* **56:** 329-354.

Swofford, DL. 2002. PAUP*: *Phylogenetic analysis using parsimony (and other methods)* 4.0 Beta. Florida State University.

Thomas, CD, Cameron, A, Green, RE, Bakkenes, M, Beaumont, LJ, Collingham, YC, Erasmus, BFN, Ferreira de Siqueira, M, Grainger, A, Hannah, L, Hughes, L, Huntley, B, van Jaarsveld, AS, Midgley, GF, Miles, L, Ortega-Huerta, MA, Peterson, AT, Phillips, OL, Williams, SE. 2004. Extinction risk from climate change. *Nature* **427:** 145-148.

Thomas, CD, Franco, AMA, Hill, JK. 2006. Range retractions and extinction in the face of climate warming. *Trends in Ecology and Evolution* **21:** 415-416.

Tuskan, GA, DiFazio, S, Jansson, S, Bohlmann, J, Grigoriev, I, Hellsten, U, Putnam, N, Ralph, S, Rombauts, S, Salamov, A, Schein, J, Sterck, L, Aerts, A, Bhalerao, RR, Bhalerao, RP, Blaudez, D, Boerjan, W, Brun, A, Brunner, A, Busov, V, Campbell, M, Carlson, J, Chalot, M, Chapman, J, Chen, GL, Cooper, D, Coutinho, PM, Couturier, J, Covert, S, Cronk, Q, Cunningham, R, Davis, J, Degroeve, S, Dejardin, A, dePamphilis, C, Detter, J, Dirks, B, Dubchak, I, Duplessis, S, Ehlting, J, Ellis, B, Gendler, K, Goodstein, D, Gribskov, M, Grimwood, J, Groover, A, Gunter, L, Hamberger, B, Heinze, B, Helariutta, Y, Henrissat, B, Holligan, D, Holt, R, Huang, W, Islam-Faridi, N, Jones, S, Jones-Rhoades, M, Jorgensen, R, Joshi, C, Kangasjarvi, J, Karlsson, J, Kelleher, C, Kirkpatrick, R, Kirst, M, Kohler, A, Kalluri, U, Larimer, F, Leebens-Mack, J, Leple, JC, Locascio, P, Lou, Y, Lucas, S, Martin, F, Montanini, B, Napoli, C, Nelson, DR, Nelson, C, Nieminen, K, Nilsson, O, Pereda, V, Peter, G, Philippe, R, Pilate, G, Poliakov, A, Razumovskaya, J, Richardson, P, Rinaldi, C, Ritland, K, Rouze, P, Ryaboy, D, Schmutz, J, Schrader, J, Segerman, B, Shin, H, Siddiqui, A, Sterky, F, Terry, A, Tsai, CJ, Uberbacher, E, Unneberg, P, Vahala, J, Wall, K,

Wessler, S, Yang, G, Yin, T, Douglas, C, Marra, M, Sandberg, G, Van de Peer, Y, Rokhsar, D. 2006. The genome of black cottonwood, Populus trichocarpa Torr. & Gray. *Science* **313:** 1596-1604.

Whittall, JB, Hellquist, CB, Schneider, EL, Hodges, SA. 2004. Cryptic species in an endangered pondweed community (*Potamogeton*, Potamogetonaceae) revealed by AFLP markers. *American Journal of Botany* **91:** 2022-2029.

Yuan, Y-W, Olmstead, RG. 2008. Evolution and phylogenetic utility of the *PHOT* gene duplicates in the *Verbena* complex (Verbenaceae): dramatic intron size variation and footprint of ancestral recombination. *American Journal of Botany* **95:** 1166-1176.

Figure 1. Distribution of the six species of *Spiraeanthemum* belonging to the *S. brongniartianum* group (Pillon *et al.*, 2009a) in New Caledonia. Names of localities sampled and type localities are indicated. The type locality of *S. pedunculatum* ("région du Sud") is imprecise and could match any of the extant populations of this species, and it is therefore not shown. Areas of ultramafic soils are indicated in grey.

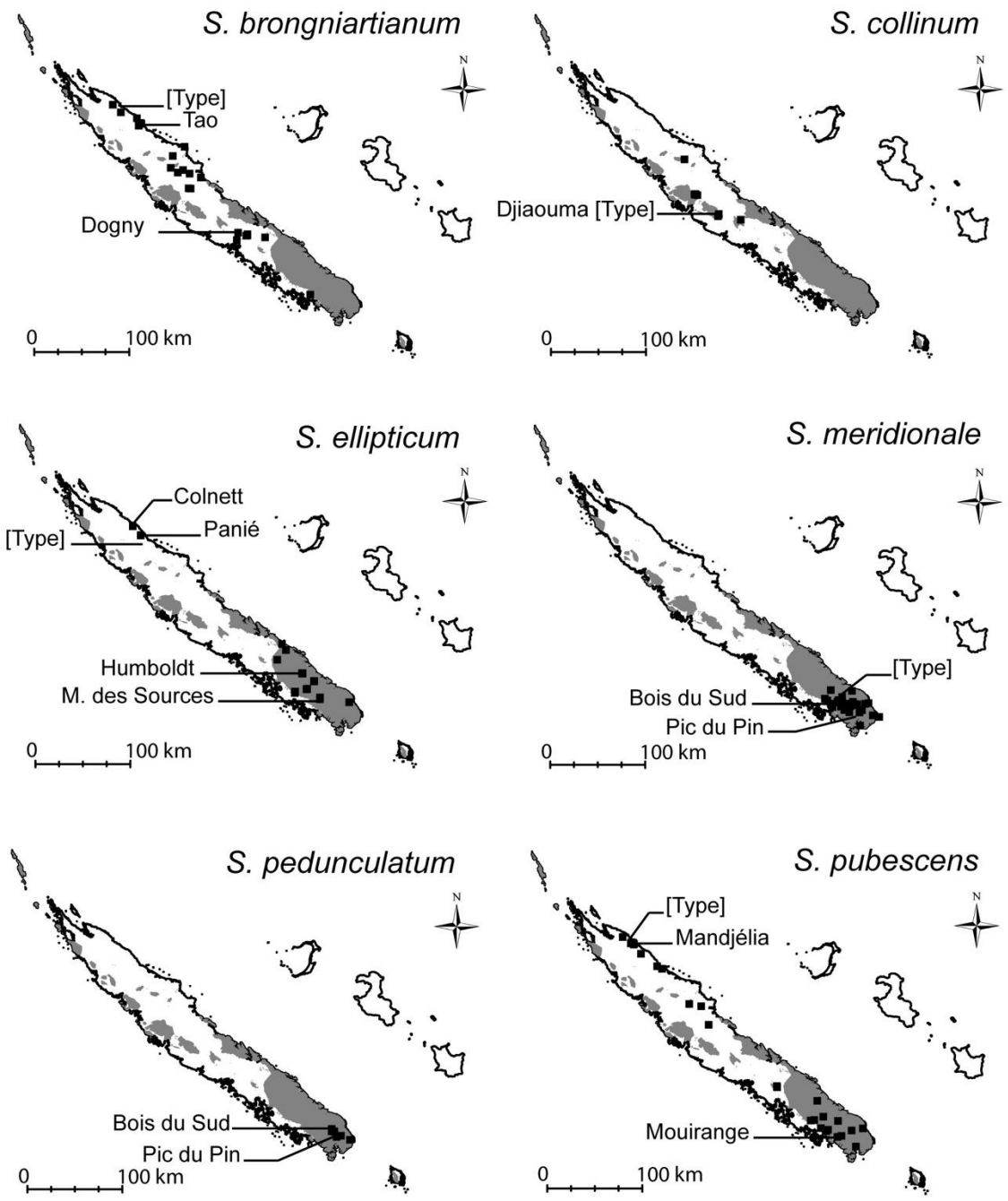


Figure 2. One of the two most parsimonious trees based on the phylogenetic analysis of *GapC* sequences. Numbers above branches are branch lengths, and numbers below branches are bootstrap percentages. Positions for the recombinant allele 1 of *S. peducunlatum* are shown in grey according to which part of the sequence was included in the analysis: part 1 (base 1 to 429), part 2 (base 430 to 906). Numbers in parentheses indicate the number of clones sequenced for each allele in heterozygote accessions. The asterisk indicates a sequence detected as recombinant in the recombination network.

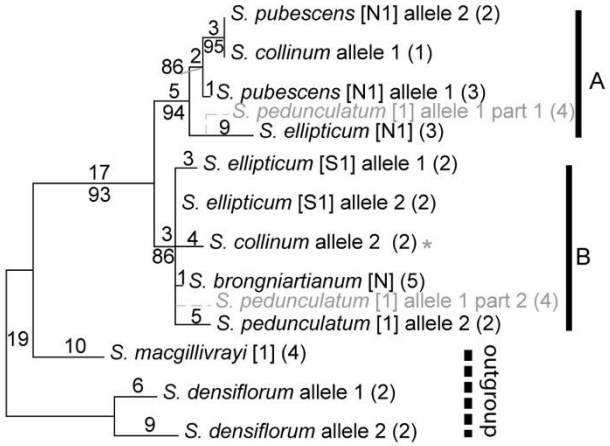


Figure 3. Phylogenetic trees of *Spiraeanthemum* based on *ncpGS* sequence data. Left: one of 195 shortest trees from a maximum parsimony analysis. Numbers above branches are branch lengths, and numbers below branches are bootstrap percentages. Nodes that collapse in the strict consensus are indicated by arrowheads. Right: tree obtained with maximum likelihood analysis. Numbers below branches are bootstrap percentages. Numbers in parentheses indicate the number of clones sequenced for each allele in heterozygote accessions. Grey asterisks indicate alleles that are putatively recombinant: large asterisk = recombination detected by eye; small asterisk = recombination only detected with the recombination network.

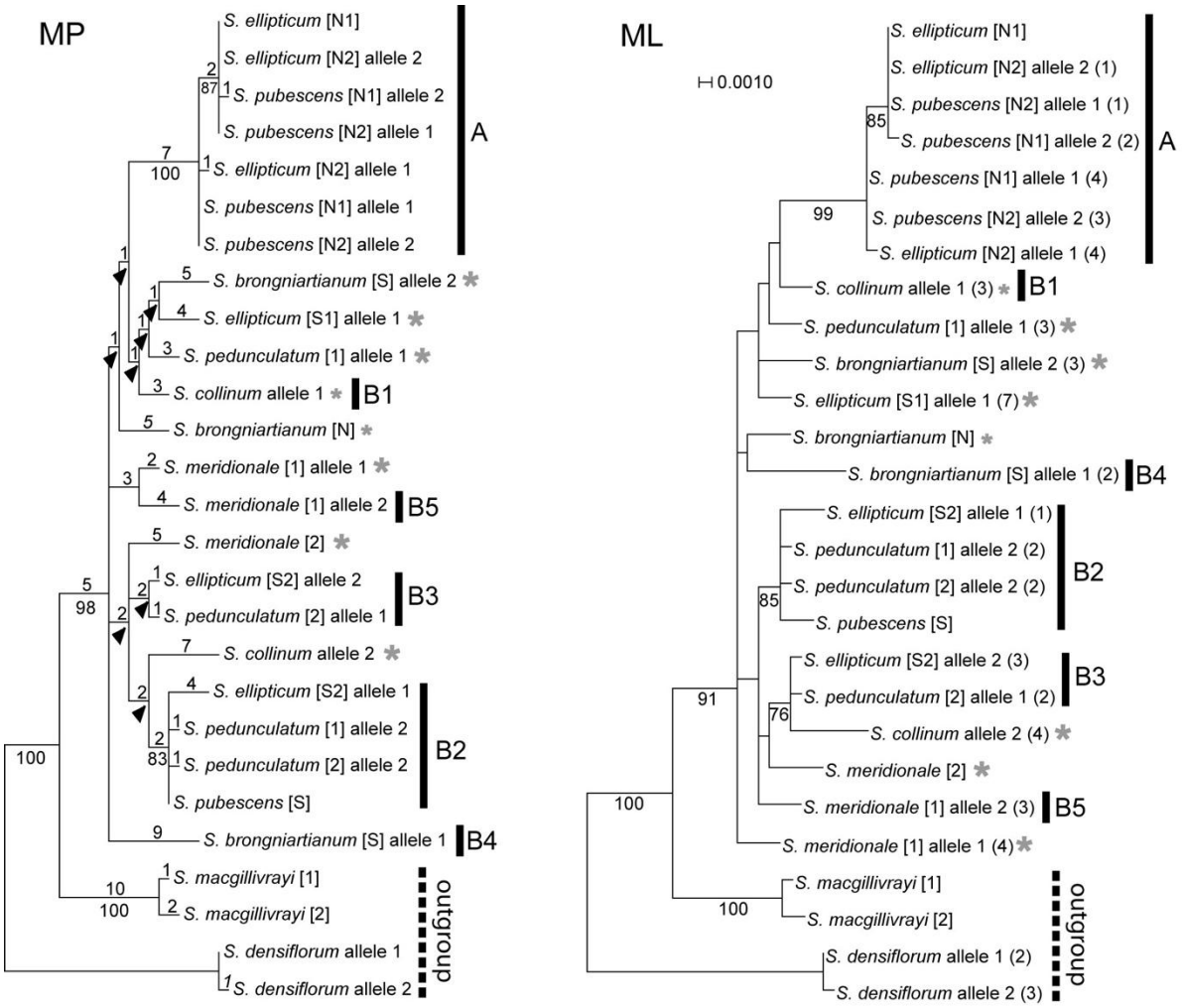


Figure 4. Phylogenetic tree of *Spiraeanthemum* using maximum likelihood based on *ncpGS* sequence data, with recombinant sequences excluded. Numbers below branches are bootstrap percentages.

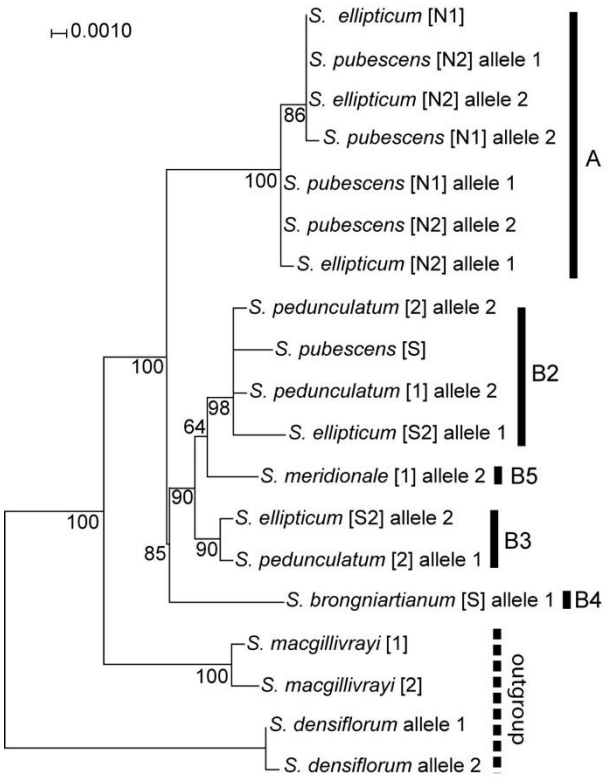


Figure 5. Recombination networks for *Spiraeanthemum* (*S. brongniartianum* group) based on *ncpGS* and *GapC* sequences. Dashed lines represent crossovers. Grey asterisks indicate alleles that are putatively recombinant = large asterisk, recombination detected by eye = small asterisk, recombination only detected with the recombination network.

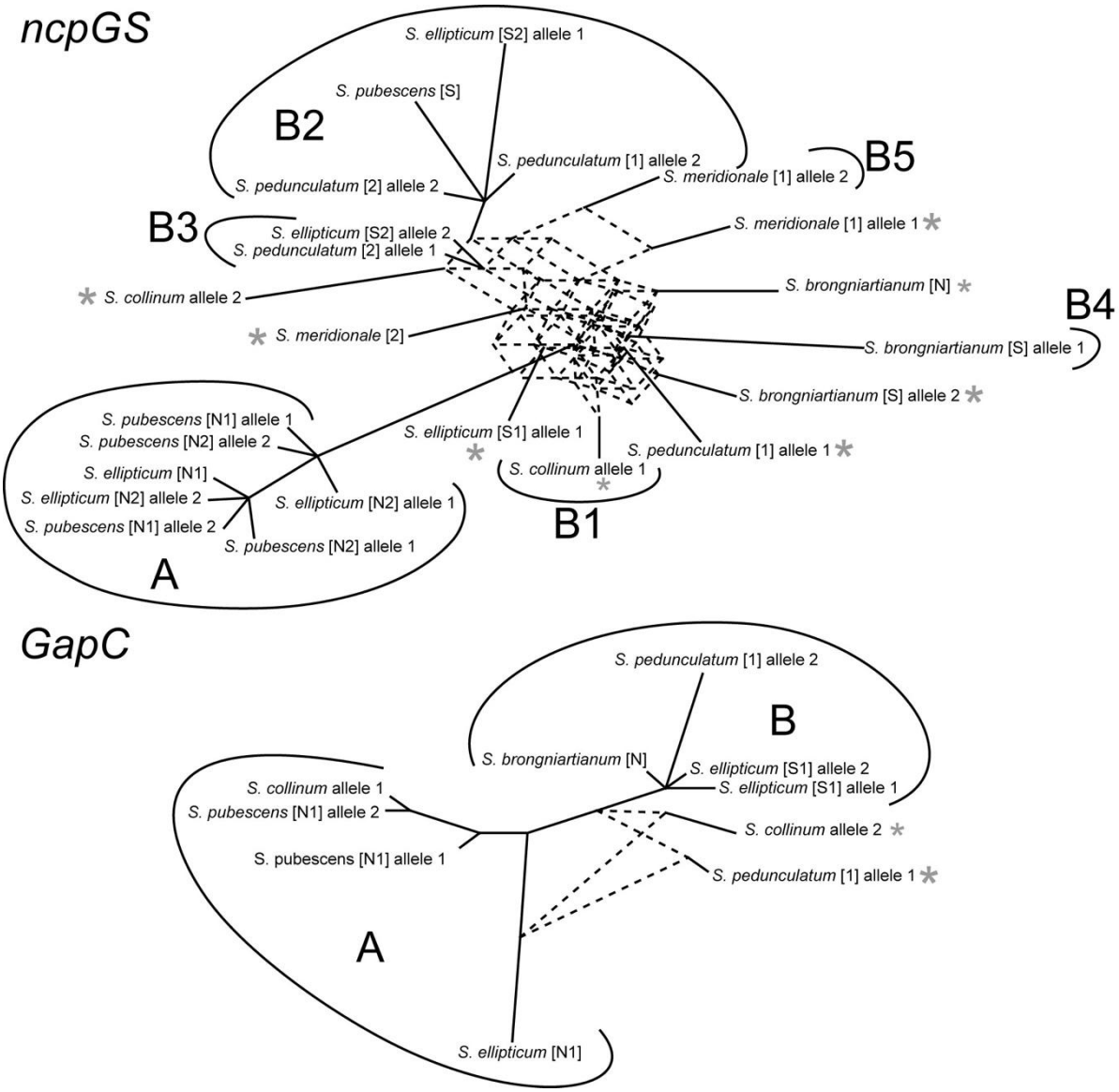


Figure 6. Possible scenarios for the origin of cryptic species in *Spiraeanthemum*. "Other species" refers to *S. brongniartianum*, *S. collinum*, *S. meridionale*, and *S. pedunculatum*. In the third scenario, arrows indicate introgression events, and an alternative hypothesis can be obtained by interchanging *S. ellipticum* and *S. pubescens.*

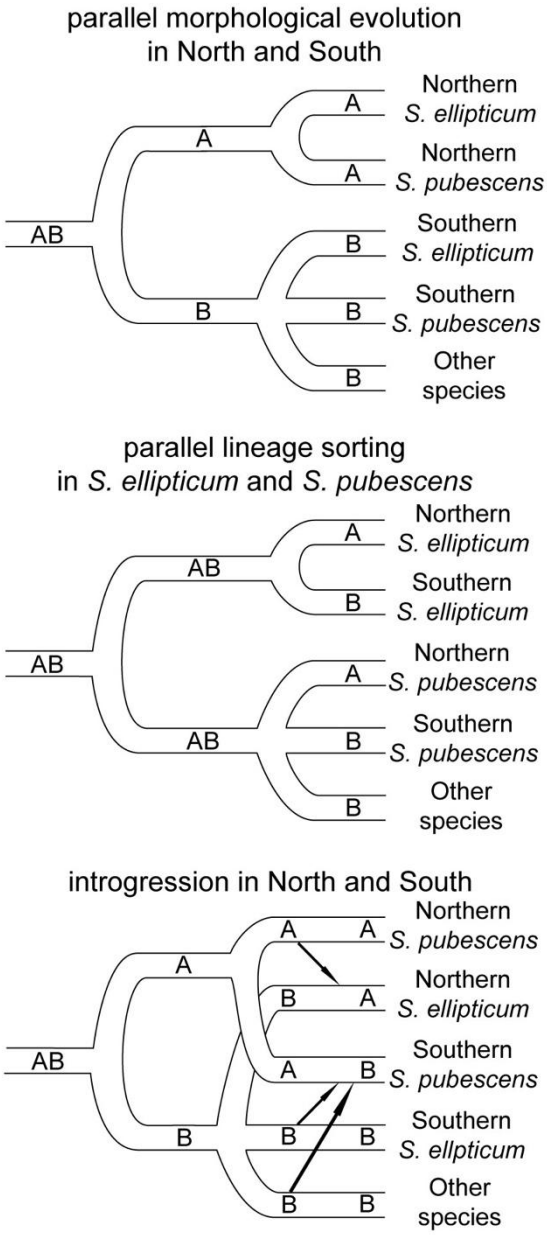


Table 1. Information on the datasets for the two genes (*ncpGS* and *GapC*) included in the study.

	ncpGS	*GapC*
Number of individuals	17	8
Number of haplotypes	24	12
Length of alignment	1209	906
Variable sites	108 (8.9 %)	84 (9.3 %)
Parsimony informative sites	64 (5.3 %)	52 (5.7%)

Table 2. Base composition at the parsimony-informative sites in the *ncpGS* sequences recovered in the species of *Spiraeanthemum* belonging to the *S. brongniartianum* group. The dashes indicate a gap between position 351 and 381. Putative recombination sites are indicated by a vertical line, and the phylogenetic group to which each allele/segment belongs is given in the last column. Asterisks indicate putative recombinant alleles, only detected in the recombination network.

	61	100	111	112	116	165	250	298	307	330	361	400	676	725	761	796	839	858	962	984	1072	1081	1097	1190	group
S. ellipticum [N1]	C	T	A	C	T	G	T	G	G	C	C	T	T	A	C	G	T	T	G	T	A	T	A	C	A
S. ellipticum [N2] allele 2	C	T	A	C	T	G	T	G	G	C	C	T	T	A	C	G	T	T	G	T	A	T	A	C	A
S. pubescens [N1] allele 2	C	T	A	C	T	G	T	G	G	C	C	T	T	A	C	G	T	T	G	T	A	T	A	C	A
S. pubescens [N2] allele 1	C	T	A	C	T	G	T	G	G	C	C	T	T	A	C	G	T	T	G	T	A	T	A	C	A
S. ellipticum [N2] allele 1	T	T	A	C	T	G	T	G	G	C	C	T	T	A	G	G	T	T	G	T	A	T	A	C	A
S. pubescens [N1] allele 1	T	T	A	C	T	G	T	G	G	C	C	T	T	A	G	G	T	T	G	T	A	T	A	C	A
S. pubescens [N2] allele 2	T	T	A	C	T	G	T	G	G	C	C	T	T	A	G	G	T	T	G	T	A	T	A	C	A
S. collinum allele 1	T	T	A	T	T	T	T	G	A	C	C	T	C	A	G	A	A	T	A	C	A	T	A	G	B1*
S. ellipticum [S2] allele 1	T	A	A	C	T	T	T	A	A	T	-	T	C	G	G	G	A	G	A	C	G	T	T	G	B2
S. pedunculatum [1] allele 2	T	A	A	C	T	T	T	A	A	T	-	T	C	G	G	G	A	G	A	C	G	T	T	G	B2
S. pedunculatum [2] allele 2	T	A	A	C	T	T	T	A	A	T	-	T	C	G	G	G	A	G	A	C	G	T	T	G	B2
S. pubescens [S]	T	A	A	C	T	T	T	A	A	T	-	T	C	G	G	G	A	G	A	C	G	T	T	G	B2
S. ellipticum [S2] allele 2	T	T	A	C	T	T	G	G	A	T	T	T	C	G	G	G	A	T	A	C	G	T	T	G	B3
S. pedunculatum [2] allele 1	T	T	A	C	T	T	G	G	A	T	T	T	C	G	G	G	A	T	A	C	G	T	T	G	B3
S. brongniartianum [S] allele 1	T	T	A	C	T	T	T	G	A	T	C	C	C	G	G	G	A	T	A	C	A	C	A	G	B4
S. brongniartianum [N]	T	A	A	C	T	T	T	G	A	C	C	T	C	G	G	G	A	T	A	C	A	C	A	G	B4*
S. meridionale [1] allele 2	T	A	G	C	C	T	T	G	A	T	C	T	C	G	G	G	A	T	A	C	G	T	T	G	B5
S. brongniartianum [S] allele 2	T	A	A	C	T	T	C	G	A	T	C	C	C	A	G	A	A	T	A	C	A	C	A	G	B4+B1+B4
S. collinum allele 2	T	T	A	C	T	T	G	A	A	T	-	T	C	G	G	G	A	T	A	C	G	T	T	G	B3+B2+B3
S. ellipticum [S1] allele 1	T	A	A	C	T	T	T	A	A	T	-	T	C	A	G	A	A	T	A	C	A	T	A	G	B2+B1
S. meridionale [1] allele 1	T	A	G	C	C	T	T	G	A	T	C	T	C	G	G	G	A	T	A	C	A	T	A	G	B5+B1
S. meridionale [2]	T	T	A	T	T	T	T	G	A	C	C	T	C	G	G	G	A	T	A	C	G	T	T	G	B1+B5
S. pedunculatum [1] allele 1	T	A	A	C	T	T	T	G	A	C	C	T	C	A	G	A	A	T	A	C	A	T	A	G	B2+B1

Appendix. List of taxa [accession number], voucher reference (voucher location), locality, and Genbank accession number for *ncpGS*, and *GapC*. Accessions numbers are only given when there are multiple accessions for a given species; southern and northern accessions are indicated with S and N respectively. All localities are in New Caledonia unless indicated (Vanuatu).

Spiraeanthemum brongniartianum Schltr. [S], Bradford *et al.* 1179 (MO,NOU), Plateau de Dogny, GQ332620 & GQ332621, ─; *Spiraeanthemum brongniartianum* Schltr. [N], Pillon *et al.* 87 (K,MO,NOU,P), Mont Panié, Tao, EU867240, GQ332619; *Spiraeanthemum collinum* (Hoogland) Pillon, Pillon *et al.* 464 (K,NOU,P), Mé Maoya, Djiaouma, GQ332622 & GQ332623, GQ332611 & GQ332612; *Spiraeanthemum densiflorum* Brongn. & Gris, Pillon 667 (NOU), Col d'Amieu, EU867243 & EU867244, GQ332615 & GQ332616; *Spiraeanthemum ellipticum* Vieill. ex Pamp. [S1], Pillon *et al.* 20 (K,NOU,P), Mont Humboldt, GQ332625, GQ332607 & GQ332608; *Spiraeanthemum ellipticum* Vieill. ex Pamp. [S2], Pillon & Barrabé 672 (NOU), Montagne des Sources, GQ332626 & GQ332627, ─; *Spiraeanthemum ellipticum* Vieill. ex Pamp. [N1], Pillon *et al.* 347 (K,MO,NOU,NSW,P), Mont Panié, GQ332624, GQ332618; *Spiraeanthemum ellipticum* Vieill. ex Pamp. [N2], Pillon *et al.* 616 (NOU), Mont Colnett, GQ332628 & GQ332629, ─; *Spiraeanthemum macgillivrayi* Seem. subsp. *macgillivrayi* [1], Pillon *et al.* 570 (K,MO,NOU,P,PVNH), VANUATU, Santo, Mont Tabwémasana, EU867245, GQ332617; *Spiraeanthemum macgillivrayi* Seem. subsp. *macgillivrayi* [2], Pillon *et al.* 525 (NOU,P), VANUATU, Santo, Boutmas, EU867246, ─; *Spiraeanthemum meridionale* (Hoogland) Pillon [1], Pillon *et al.* 1 (NOU), Bois du Sud, GQ332631 & GQ332632, ─; *Spiraeanthemum meridionale* (Hoogland) Pillon [2], Pillon & Munzinger 636 (NOU), Pic du Pin, GQ332630, ─; *Spiraeanthemum pedunculatum* Schltr. [1], Pillon *et al.* 437 (NOU), Bois du Sud, GQ332633 & GQ332634,

GQ332613 & GQ332614; *Spiraeanthemum pedunculatum* Schltr. [2], Pillon & Munzinger 637 (K,NOU,P), Pic du Pin, GQ332635 & GQ332636, ─; *Spiraeanthemum pubescens* Pamp. [S], Pillon *et al.* 316 (K,NOU,P), Col de Mouirange, GQ332637, ─; *Spiraeanthemum pubescens* Pamp. [N1], Pillon & Munzinger 82 (K,NOU,P), Mandjélia, EU867241 & EU867242, GQ332609 & GQ332610; *Spiraeanthemum pubescens* Pamp. [N2], Munzinger *et al.* 4014 (NOU,P), Mandjélia, GQ332638 & GQ332639, ─.