The Dissertation Committee for Joanna Melinda Padolina Certifies that this is the approved version of the following dissertation:

Phylogenetic reconstruction of *Phalaenopsis* (Orchidaceae) using nuclear and chloroplast DNA sequence data
and
Using *Phalaenopsis* as a natural system for assessing methods to reconstruct hybrid evolution in phylogenetic analyses

Committee:

Beryl B. Simpson, Supervisor

C. Randal Linder, Co-Supervisor

Robert K. Jansen

Donald A. Levin

Ulrich G. Mueller

Phylogenetic reconstruction of *Phalaenopsis* (Orchidaceae) using nuclear and chloroplast DNA sequence data
and
Using *Phalaenopsis* as a natural system for assessing methods to reconstruct hybrid evolution in phylogenetic analyses

by

Joanna Melinda Padolina, B. S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2006

Dedication

This dissertation is dedicated to my family:

to my father for pushing me to achieve a Ph.D.,

to my mother for supporting my choices in life,

to my sisters for their mutual understanding of life, the universe and everything,

to my husband for always being there,

and to my son for putting everything into perspective.

Acknowledgements

Thanks to all of you who have made this journey easier, exciting, and fulfilling.

I would like to acknowledge The University of Texas at Austin, The National Science Foundation, The American Orchid Society, and the PEO International Sisterhood for funding my education and research. I would like to thank Marie Selby Gardens, The New York Botanical Gardens, Bill Tippit, Dwayne Lowder, David Hunt, and Johanna Willink for helping me obtain *Phalaenopsis* species.

I have been blessed with two of the best advisors ever. I want to thank Beryl for all of her practical advice, for returning all of my manuscripts so quickly, and for being such a wonderful role model. I want to thank Randy for his excitement and enthusiasm for whatever the topic is at hand, for his thorough knowledge about everything, and for his understanding and compassion. I would also like to thank the rest of my committee for always being available for questions and advice and for making UT a wonderful place to be a graduate student, Thanks Bob, Ulrich and Dr. Levin!

Greg Plunkett opened his lab to me after I moved to Virginia. Thank you for giving me a lab-away-from-home, for helping me land a job at VCU, and for being such a nice guy.

One of the most important people I need to thank from the bottom of my heart is Tanya Ong. Without her hard work I would not have much data. Thank you Tanya for your late night soirées in the lab. I hope you find the perfect job! I would also like to thank all of the other undergraduate workers and lab techs that have put hours into our projects and into keeping our lab running, especially Mario McDermott and Summer Chartier.

There are many friends I need to thank for everything they have added to my life in Austin. I want to thank Josh for being a great buddy, for getting coffee and Chipotle with me, for road trippin', and for being real smart and telling me what's up. I want to thank Kara for brightening my Austin days and confirming my status as an honorary Texan (no I'm not from Texas, but Texas wants me anyway). I want to thank Linglingay for being a great roommate, a friend, and a sister, and for letting me stay with her whenever I'm back in town. I would like to thank everyone in the Simpson and Linder Lab's these past six years for putting up with me. Especially Josh and Ruth. You two are swell. I would like to thank the X-Gals for helping me through the hard times, for reading manuscripts, and for being the voice of reason.

My mother and my mother-in-law gave me many weeks of childcare so that I could finish writing. I couldn't have done it without their help. Barbara, my sister, is the best person in the world with whom I can commiserate. I thank her for understanding what it is like to be getting a Ph.D. in my family (oi!). I am grateful to have her to kvetch with.

Finally, I would like to thank my husband, Gani, from the bottom of my heart. He is my strength, my conscience, and my light in the darkness.

Phylogenetic reconstruction of *Phalaenopsis* (Orchidaceae) using nuclear and chloroplast DNA sequence data

and

Using *Phalaenopsis* as a natural system for assessing methods to reconstruct hybrid evolution in phylogenetic analyses

Publication No._____

Joanna Melinda Padolina, Ph.D.

The University of Texas at Austin, 2006

Supervisors: Beryl B. Simpson and C. Randal Linder

Two phylogenies of *Phalaenopsis* (Orchidaceae) are presented, one from combined chloroplast DNA data and one from a nuclear actin gene. We used these phylogenies to assess and modify the classification of *Phalaenopsis* and to examine several morphological characters and geographical distribution patterns. Our results support Christenson's (2001) treatment of *Phalaenopsis* as a broadly defined genus that includes the species previously placed in the genera *Doritis* and *Kingidium*. Some of Christenson's subgeneric groups needed to be recircumscribed to reflect a natural classification. We recognized four subgenera and six sections, subgenera *Aphyllae, Parishianae* (with sections *Conspicuum, Delisiosae, Esmeralda,* and *Parishianae*), *Phalaenopsis,* and *Polychilos* (with sections *Fuscatae* and *Polychilos*).

In order to find a set of universally amplifiable, phylogenetically informative, single-copy nuclear regions, we conducted a whole genome comparison of the rice (*Oryza sativa*) and *Arabidopsis thaliana* genomes. We constructed a database of both genomes and searched for pairs of sequences using criteria we felt would ensure primers that would reliably amplify using standard PCR protocols. We tested the most promising 142 primer pairs in the lab on eighteen taxa and found four potentially informative markers in *Phalaenopsis* and one in *Helianthus*. Our results indicated that it will be difficult to find universal nuclear markers, however our database provides an important tool for finding informative nuclear markers within specific groups. The full set of primer combinations is available online at, "The Conserved Primer Pair Project," http://aug.csres.utexas.edu:8080/cpp/index.html.

We used fourteen *Phalaenopsis* species and seven horticultural hybrids to create a real dataset with which to test phylogenetic network reconstruction methods. We tested the performance of Neighbor-Net, implemented in SplitsTree, under four different categories of complexity: one hybrid, two independent hybrids (hybrids with no parents in common), three independent hybrids, and two non-independent hybrids (one parent was shared between hybrids). Neighbor-Net was able to predict accurately the parents of hybrids in only about half of the datasets we tested, and there were so many false positives that it was impossible to distinguish the hybrids from the species. We plan to use this dataset to test methods, such as RIATA and RGNet, when they become available.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Phalaenopsis Bl. is in the family Orchidaceae, subfamily Epidendroideae, tribe Vandeae, subtribe Aeridinae (Cameron et al. 1999), and consists of 63 species that occur primarily as epiphytes throughout South East Asia and the Pacific Islands. Their showy, exotic flowers and ease of cultivation have made *Phalaenopsis* one of the most widely traded groups of horticultural plants throughout the world. *Phalaenopsis* has been the subject of many taxonomic treatments (Reichenbach 1862, 1864; Rolfe 1886; Pfitzer 1889; Rolfe 1917; Sweet 1980; Shim 1982; Christenson 2001), and several of its species groups have been segregated into separate genera by certain authors (Rolfe 1917; Hawkes 1966; Shim 1982). All treatments of *Phalaenopsis* to date have been based solely on morphological characters.

TAXONOMIC HISTORY OF *PHALAENOPSIS*

The first accounting of a *Phalaenopsis* appeared in 1750 and was published by G. E. Rumphius under the name *Angraecum albus majus* (de Wit 1977). In 1752 Osbeck collected a plant similar to Rumphius' *Angraecum* on New Island in Java and sent it to Linnaeus, who named it *Epidendrum amabile* (Linnaeus 1753). *Epidendrum amabile* was transferred to several different genera before the Dutch botanist Blume (1825) erected the genus *Phalaenopsis* after collecting specimens on the island Nusa Kambangan in Indonesia. He established Linnaeus' *E. amabile* as the type specimen renaming it, *Phalaenopsis amabilis*.

Phalaenopsis was a popular group of plants throughout the 19th and 20th centuries and there has been much interest in, and contention about, the proper classification of the

group (TABLE 1.1). Several genera have been named for species that had been (or would be) *Phalaenopsis* but authors who preferred a more broadly defined *Phalaenopsis* have challenged most of these names. Proposed genera included *Polychilos* Breda, *Doritis* Lindl., *Stauroglottis* Schauer, *Kingiella* Rolfe = *Kingidium* Hunt, *Paraphalaenopsis* Hawkes and *Grafia* Hawkes. Of these, *Doritis* and *Kingidium* have been the most controversial.

TABLE 1.1: Classifications associated with *Phalaenopsis*. Genera in bold print, subgeneric classes indicated in parenthases.

Reichenbach 1862	Bentham & Hooker 1882	Rolfe 1886	Pfitzer 1889	Rolfe 1917	Sweet 1968-1969	Shim 1982	Christenson 2001
Phalaenopsis (genus) 1. Lip apex with appendages (unnamed group) 2. Lip apex with no appendages 3. *Polychilos* (Breda) Rchb. f. (added in 1864)	***Phalaenopsis*** (genus) *Euphalaenopsis* Benth. & Hook. (section) *Stauroglottis* (Schauer) Benth. & Hook.	***Phalaenopsis*** (genus) *Euphalaenopsis* (section) *Stauroglottis* *Proboscidioides* Rolfe *Esmeralda* Rchb. f.	***Phalaenopsis*** (genus) *Euphalaenopsis* (section) *Stauroglottis* *Proboscidioides* *Zebrinae* Pfitz. *Antenniferae* Pfitz. (superfluous =*Esmeralda*)	***Phalaenopsis*** (genus) *Euphalaenopsis* (section) *Stauroglottis* *Proboscidioides* Rolfe ***Doritis*** (=section *Esmeralda*) ***Kingiella*** Rolfe	***Phalaenopsis*** (genus) *Phalaenopsis* (section) *Stauroglottis* *Proboscidioides* *Aphyllae* Sweet *Parishianae* Sweet *Polychilos* *Fuscatae* Sweet *Amboinenses* Sweet *Zebrinae* *Zebrinae* (subsection) *Lueddemanniana* Sweet *Hirsutae* Sweet *Glabrae* Sweet ***Kingidium* P.F. Hunt** ***Doritis*** ***Paraphalaenopsis*** Hawkes	***Phalaenopsis*** (genus) *Phalaenopsis* (section) *Stauroglottis* *Paraphalaenopsis* (Hawkes) P. S. Shim ***Polychilos* Breda** *Polychilos* *Kingidium* (P. S. Hunt) Shim ***Doritis***	***Phalaenopsis*** (genus) *Phalaenopsis* (subgenus) *Phalaenopsis* (section) *Stauroglottis* *Esmeralda* *Deliciosae* *Polychilos* *Polychilos* *Amboinenses* *Fuscatae* *Zebrinae* *Aphyllae* *Parishianae* *Proboscidioides* ***Paraphalaenopsis***

Doritis

Lindley (1833) circumscribed the genus *Doritis*, for *Doritis pulcherrima* Lindl. from a dried specimen in poor condition that was collected by Finlayson near Tourane Bay, Cochin China. *Doritis* continues to be a subject of debate to this day. Reichenbach (1879) named several specimens collected from Cochin, China, that are now synonymous with *Doritis pulcherrima*, including *Phalaenopsis esmeralda* Rchb. f., which Rolfe (1886) placed into *Phalaenopsis* section *Esmeralda* Rchb.f. Pfitzer (1889) placed these species into *Phalaenopsis* section *Antenniferae* Pfitz..

The specimens on which Lindley based his description of *Doritis* were generally overlooked due to their condition, until Rolfe (1917) uncovered some additional accessions of Finlayson's original material from the Wallichian Herbarium on loan at Kew. After inspecting these specimens, Rolfe placed *P. esmeralda* into synonymy with *Doritis pulcherrima*, and reduced the other species in section *Esmeralda* (*P. antennifera, P. regnieriana* Rchb.f., and *P. buyssoniana* Rchb.f.) to varieties of *D. pulcherrima*. He also clarified taxonomic issues regarding the characterization of *Doritis*. Several species had been placed in *Doritis*, but once the type species of *Doritis* had been clarified, it became apparent that some of these species, such as *D. bifalcis* and *D. paniculatum*, belonged in other genera. Rolfe felt that five species (*D. taenialis* (Lindl.) J.D. Hook., *D. wightii (*Rchb.f.) Benth. & J.D. Hook., *D. hebe* (Rchb.f.) Schltr., *D. philippinensis* Ames, and *D. steffensii* Schltr.) originally described as *Doritis* were distinct enough in structure from *Doritis pulcherrima* to warrant a new genus, which he named *Kingiella* and characterized by a "spur-like mentum."

J. J. Smith (1933) controversially transferred *Doritis pulcherrima* to *Phalaenopsis* as *Phalaenopsis pulcherrima* (Lindl.) J. J. Sm. and included several taxa in its synonymy.

He noted that many of the differences between *Doritis* and *Phalaenopsis* were adaptations to a terrestrial habitat versus adaptations to an epiphytic habitat. He also pointed out that the 'linear appendages' towards the base of the lip were homologous to the lateral lobes of *Phalaenopsis* flowers and the midlobe of the lip was expanded into three lobules.

Holttum (1965) noted that *Doritis* was closely allied to *Phalaenopsis*, but he felt it distinct enough to be considered a separate genus based on its long column foot with the lateral sepals attached throughout its length, unusual lip morphology, and two pollinia which were so deeply divided they were almost two separate parts (now considered to be four pollinia). Sweet (1980) and Seidenfaden (1988) maintained the genus *Doritis*, on account of its four, globular pollinia which differ from *Phalaenopsis'* two cleft pollinia. Christenson (2001) sunk *Doritis* into synonymy with *Phalaenopsis* in agreement with Smith (1933). He noted that pollinia number had been found to be fairly plastic throughout the Orchidaceae and should not be used as a character for generic circumscription.

Kingidium

Rolfe (1917) described the genus *Kingiella* with five species, which had previously been placed in either *Phalaenopsis* or *Doritis*. Holttum (1966) addressed some of the nomenclatural issues that were causing confusion in *Kingiella*. He modified the description of the genus to describe correctly the anatomical features of the spur, the lip, and the attachment of floral parts. Of the five species recognized by Rolfe, Holttum could only distinguish *Kingiella taenialis* from the other four. He combined the rest into a single variable species, *K. decumbens,* reducing the number of named species in the genus to two. He noted that it would be acceptable to include *Kingiella* as a section of

Phalaenopsis, but kept them as separate genera because Kew Garden and many authors recognized *Kingiella* as distinct. Hunt (1970) changed the name *Kingiella* to *Kingidium* because *Kingiella* was an invalid "orthographic variant" of *Kingella* Tiegh, an earlier established name in the Loranthaceae. Sweet (1980) did not address *Kingidium* species in his treatment of *Phalaenopsis,* except to note that they were not species of *Phalaenopsis*. Shim (1982) treated *Kingidium* as a member of the '*Phalaenopsis* Complex' and placed it in the genus *Polychilos* section *Kingidium*. Shim's classification was never accepted.

Seidenfaden (1988) addressed the taxonomic mess between *Kingidium* and *Phalaenopsis*: regarding *Kingidium*'s status as a separate genus or a section of *Phalaenopsis*, he stated that "it is a subjective judgment what combination of characters may justify generic or sectional circumscription…I have no strong opinion on the matter…but, based on taxonomical considerations, …I prefer the generic solution which also has the advantage not further to add to the confusion." He further transferred two species, *Phalaenopsis braceana* (Hk. f.) Christenson and *P. stobartiana* Rchb. f., to *Kingidium* on the basis of the presence of a spur, which was the decisive character separating *Kingidium* from *Phalaenopsis*. Finally, Seidenfaden described a new species, *Kingidium minus* Seidenf., bringing the total number of species in the genus to five.

Throughout the 1990s Gruss and Rollke published a series of articles moving species from *Phalaenopsis* section *Aphyllae* into *Kingidium,* creating new nomenclatural combinations, and circumscribing addional sections (Gruss and Rollke 1993a, 1993b, 1995a, 1995b, 1996, 1997). Christenson (1995, 2001) brought all species that had been moved into *Kingidium* back into *Phalaenopsis*.

Infrageneric Classification

While the generic circumscription of *Phalaenopsis* has been a subject of heated debate, the classification within the genus has remained relatively stable, changing mostly to accommodate the rapid addition of newly described species. Some of the notable classifications are in TABLE 1.1. Reichenbach (1862) provided the first monograph of *Phalaenopsis*, splitting the genus of eleven species into two unnamed groups, based on the presence or absence of appendages at the base of the lip. He added an additional group, *Polychilos,* to *Phalaenopsis* (Reichenbach 1864), which included the species in Breda's genus *Polychilos*. Bentham (1883) named Reichenbach's groups as *Euphalaenopsis* (with appendages) and *Stauroglottis* (without appendages) after the genus *Stauroglottis* that Schauer erected to describe *Stauroglottis equestris* Schauer (=*Phalaenopsis equestris*). Bentham included *Polychilos* within *Stauroglottis*. Rolfe (1886) published a revision of the genus including 34 species and added two sections to Bentham's classification: *Proboscidioides,* a monotypic section possessing a long, beak-like rostellum, and *Esmeralda*, which was described as being similar to *Stauroglottis,* but differing by the presence of a pair of slender linear appendages on the stalk of the lip bellow the lateral sepals. In 1917, Rolfe revised his classification by moving section *Esmeralda* into synonymy with Lindley's *Doritis,* and he erected the genus *Kingiella*.

Pfitzer (1889) revised *Phalaenopsis*, dividing it into five sections. *Euphalaenopsis, Stauroglottis,* and *Proboscidioides* were similar to Rolfe's classification. Pfitzer named the section *Zebrinae*, which was similar to *Stauroglottis,* differing by having a lip longer than it is wide. He also named section *Antenniferae*, but it was superfluous because it was taxonomically synonymous with section *Esmeralda*.

The next noteworthy revision was eighty years later, when Sweet published a series of articles revising the genus (Sweet 1968, 1969, 1970). These articles were later

collected and published as a bound monograph, *The Genus* Phalaenopsis (Sweet 1980). Therein, he recognized 48 species in nine sections and four sub-sections (TABLE 1.2). Sweet defined the genus *Phalaenopsis* as having two 'more or less' cleft pollinia. He therefore excluded species from section *Esmeralda* and the genus *Kingidium*. Many authors still prefer Sweet's classification.

Sweet (1980) divided the genus as follows. Section *Phalaenopsis* was equivalent to Bentham and Hooker's *Euphalaenopsis*. (Its name was changed to reflect the current International Code of Botanical Nomenclature's rules.) It was characterized by having petals broader than the sepals, a single callus, and appendages at the apex of the lip. In section *Stauroglottis* he placed those species with sepals and petals similar in size, a single callus, and no lip appendages. His concept of *Stauroglottis* possessing only a single callus greatly reduced the size of this section as it had previously been defined. Section *Zebrinae* was as Pfitzer described it, but he further divided the section into four subsections, *Zebrinae, Lueddemannianae*, *Hirsutae*, and *Glabrae*, based on the morphology of the calli and adornments on the midlobe of the lip. Section *Proboscidioides* was the same as Rolfe's monotypic genus and section *Polychilos* contained the group of species that had been described as a distinct genus by Breda. In addition to the five sections already established in *Phalaenopsis*, Sweet described four new sections in his revision: Section *Aphyllae*, section *Parishianae*, section *Fuscatae*, and section *Amboinenses*.

TABLE 1.2. Sweet's (1980) classification of the genus *Phalaenopsis*. Characters marked with † are unique and pertinent only to that group.

Section	Subsection	Species	Description
Amboinenses Sweet		• *P. amboinensis* • *P. javanica* • *P. micholitzii* • *P. gigantea* • *P. robinsonii*	• Sepals and petals similar • Evergreen leaves • †Full, round flowers with broadly elliptic petals • †Fleshy, convex midlobe with a single thin central keel
Aphyllae Sweet		• *P. stobartiana* • *P. wilsonii*	• Sepals and petals similar • Deciduous, bract-like leaves • †Photosynthetic roots
Fuscatae Sweet		• *P. cochlearis* • *P. fuscata* • *P. kunstleri* • *P. viridis*	• Sepals and petals similar • Evergreen leaves • †Recurved margins • †Midlobe of lip concave or flat with one to several fleshy keels or ridges.
Parishianae Sweet		• *P. appendiculata* • *P. gibbosa* • *P. lobbii* • *P. mysorensis* • *P. parishii*	• Sepals and petals similar • Deciduous leaves • †Claw of lip adnate to column-foot at a right angle • †Mobile lip midlobe
Phalaenopsis		• *P. amabilis* • *P. sanderiana* • *P. schilleriana* • *P. stuartiana* • *P.* x*intermedia* • *P.* x*leucorrhoda* • *P.* x*veitchiana*	• Petals much wider than sepals • Evergreen leaves • †Apex of the lip divided into bifid of two cirrhi-like projections
Polychilos		• *P. cornu-cervi* • *P. lamelligera* • *P. mannii* • *P. pantherina*	• Similar sepals and petals • Evergreen leaves • †Apical lobe of lip anchor-shaped
Proboscidioides		• *P. lowii*	• Petals much wider than sepals • Deciduous leaves • †Column with a long proboscis-like extension to the rostellum
Stauroglottis		• *P. celebensis* • *P. equestris* • *P. lindenii*	• Sepals and petals similar • Evergreen leaves • †Peltate callus

TABLE 1.2 continued

Section	Subsection	Species	Description
Zebrinae Sweet	4 sections	14 species	• Sepals and petals similar • Evergreen leaves • †Flowers stellate - petals narrowly obovate or oblanceolate
	Zebrinae Sweet	• *P. corningiana* • *P. speciosa* • *P. sumatrana*	• Calli tuberculate or with denticulate or digitate or superimposed plates. • Column with a cucullate clinandrium (hooded antherbed). • Midlobe of lip pubescent.
	Lueddemannianae Sweet	• *P. fasciata* • *P. fimbriata* • *P. hieroglyphica* • *P. lueddemanniana* • *P. pulchra* • *P. reichenbachiana* • *P. violacea*	• Calli variously tuberculate, never simply bifid. • Column with a non-developed, low, marginate, entire clinandrium. • Midlobe either pubescent or glabrous.
	Hirsutae Sweet	• *P. mariae* • *P. pallens*	• Bifid callus. • Column with a non-developed, low, marginate, entire clinandrium. • Midlobe of lip pubescent.
	Glabrae Sweet	• *P. maculata* • *P. modesta*	• Bifid callus. • Column with a non-developed, low, marginate, entire clinandrium. • Midlobe of lip glabrous.

Shim (1982) published a classification regarding what he termed the '*Phalaenopsis* Complex', in which he included species of *Phalaenopsis, Kingidium, Doritis,* and *Paraphalaenopsis*. In his treatment, he presented a comprehensive review of comparative morphology of the group and used this to hypothesize evolutionary relationships in the complex.

He split the complex into three genera, retaining the genus *Doritis* without alteration as a terrestrial genus with a long rostellum and column foot and four pollinia. Shim considered the genus *Phalaenopsis* to be epiphytes possessing a bilobed callus, ovate anther and two cleft pollinia. He divided the genus into three sections, section *Phalaenopsis* possessing flat leaves and flowers with cirrhi at the apex of the labellum, section *Stauroglottis* with flat leaves and flowers without cirrhi on the apex of the labellum, and section *Paraphalaenopsis* with terete leaves and flowers with or without cirrhi on the apex of the labellum. *Paraphalaenopsis* was previously considered a sister genus to *Phalaenopsis* and was never accepted as a section within the genus (Sweet 1980).

Shim recognized the genus *Polychilos* as possessing a habit similar to *Phalaenopsis*, but possessing flowers with sepals generally wider than petals, toothed side lobes and a callus that may be forked, warty or further divided into plates. He divided the genus into two sections, section *Polychilos* possessed a well-developed secondary callus but no saccate lip base, and section *Kingidium* possessed a saccate lip base but no secondary callus.

In 2001, Christenson published the most recent revision of *Phalaenopsis*, commissioned by the American Orchid Society. Christenson included the species of *Doritis* and *Kingidium* within a more broadly defined *Phalaenopsis*. He recognized 63 species, five subgenera, and eight sections. His classification is discussed below.

CURRENT CLASSIFICATION

This dissertation will use as a starting point the most current classification of *Phalaenopsis*, Christenson's 2001 monograph. Christenson redefined *Phalaenopsis* more broadly than many authors (see Taxonomic History of *Phalaenopsis*) and included *Doritis* and *Kingidium* within it. He deemphasized the taxonomic utility of pollinia number, which historically has been an important taxonomic character throughout the Orchidaceae, but has recently been shown to be fairly plastic.

No single synapomorphy delineates *Phalaenopsis* but rather a suite of vegetative and floral characteristics. *Phalaenopsis* species are monopodial with short stems and succulent, fleshy leaves. The lip of the flower is always three-lobed, and there are one to three calli present at the base of the midlobe between the lateral lobes of the lip.

Christenson divided the genus into five subgenera (TABLE 1.3) based on flower color and pigmentation patterns, lip and callus structure, number of pollinia, and evergreen or deciduous leaves. These subgenera include *Phalaenopsis* (15 species), *Proboscidioides* (1 species), *Polychilos* (35 species) *Aphyllae* (7 species), and *Parishianae* (4 species). Subgenus *Phalaenopsis* was further divided into four sections, *Phalaenopsis* (6 species), *Stauroglottis* (3 species), *Deliciosae* (3 species), and *Esmeralda* (3 species). Subgenus *Polychilos* was also divided into four sections, *Polychilos* (4 species), *Fuscatae* (4 species), *Amboinenses* (22 species), and *Zebrinae* (5 species). This classification, like all classifications of *Phalaenopsis*, was based on morphological characters, as no molecular phylogeny was available to assess evolutionary relationships within the genus.

TABLE 1.3. Christenson's classification of the genus *Phalaenopsis*. Species in bold face are included in this project's analyses. Species marked with an asterisk (*) have not yet been positively identified. Characters marked with † are unique and pertinent only to that group.

Subgenus	Section	Species	Description
Phalaenopsis	4 sections	15 species 1 natural hybrid	• Uniseriate callus. • Essentially unmarked white or pink flowers. • Smooth lateral lobes without a tooth-like ridge (except section *Deliciosae).* • Large or small, evergreen, epiphytic, terrestrial or lithophyic plants. • Sepals and petals similar to subsimilar or petals much broader than sepals • 2 or 4 pollina
	Phalaenopsis	• ***P. amabilis*** • ***P. aphrodite*** • ***P. xintermedia*** • ***P. philippinensis*** • ***P. sanderiana*** • ***P. schilleriana*** • ***P. stuartiana***	• Petals much broader than sepals. • Midlobe of the lip with a pair of appendages at the apex (cirrhi). • Possess a single, prominent, erect, glossy callus. • No teeth on lateral lobes of the lip • 2 cleft pollinia. • Epiphytes.
	Stauroglottis	• ***P. celebensis**** • ***P. equestris*** • ***P. lindenii***	• Similar to section *Phalaenopsis* with smaller flowers • Sepals and petals similar in size • No appendages on the lip apex. • No teeth on lateral lobes of the lip • 2 cleft pollinia. • Epiphytes.
	Deliciosae previously in *Kingidium*	• ***P. chibae*** • ***P. deliciosa*** • *P. mysorensis*	• Sepals and petals similar to subsimilar in size • No appendages on the lip apex • Lateral lobes of the lip with tooth-like flap. • 4 semiglobular pollinia. • Subsaccate lip base. • Epiphytes.
	Esmeralda previously in *Doritis*	• ***P. buyssoniana**** • ***P. pulcherrima*** • *P. regnieriana*	• Sepals and petals similar to subsimilar in size • No appendages on the lip apex • Linear 'appendages' toward the base of the lip are the true lateral lobes of the lip. • Midlobe of the lip expanded into 3 lobules. • 4 globular pollinia. • Terrestrial to lithophytic (with moss) • †Long column foot. • †Long rostellum.

TABLE 1.3 continued

Subgenus	Section	Species	Description
Polychilos	4 sections	35 species	• Biseriate or triseriate callus. • Flowers brightly colored and patterned, fragrant, fleshy and long lasting • Lateral lobes of the lip with a raised tooth. • Large or small, evergreen, epiphytes. • Petals and sepals subequal and subsimilar or petals slightly narrower than petals. • 2 cleft pollinia. • †Persistent, fleshy, chlorophyllous perianth after pollination.
	Polychilos	• ***P. borneënsis**** • ***P. cornu-cervi*** • ***P. mannii*** • ***P. pantherina*** • ***P. thalebanii* (=*P. cornu-cervi*)**	• Petals narrower than sepals. • Triseriate callus. • Flowers yellow with red to brown bars or spots. • †Fleshy flattened rachis (except *P. mannii*). • †Non-fragrant flowers. • †Flowers produced singly in succession over a long period. • †Pair of fleshy knees at the column's base. • †Midlobe of lip transversely anchoriform or lunate.
	Fuscatae	• ***P. cochlearis*** • ***P. fuscata*** • *P. kunstleri* • ***P. viridis****	• Petals and sepals similar to subsimilar and revolute. • Biseriate callus • Pale yellow flowers with brown markings. • †Concave lip with a longitudinal keel.
	Amboinenses	• ***P. amboinensis*** • ***P. bastianii*** • ***P. bellina*** • ***P. doweryënsis*** • ***P. fasciata*** • ***P. fimbriata**** • ***P. floresensis**** • ***P. gigantea**** • ***P. hieroglyphica*** • ***P. javanica*** • ***P. lueddemanniana*** • *P. luteola* • ***P. maculata**** • ***P. mariae**** • ***P. micholitzii*** • ***P. modesta**** • ***P. pallens**** • ***P. pulchra*** • *P. reichenbachiana* • *P. robinsonii* • ***P. venosa*** • ***P. violacea***	• Petals and sepals similar to subsimilar or with petals narrower than sepals. Flowers may be rotuliform or stellate. • Biseriate or Triseriate callus • Flowers variously colored, often with bright pigmentation and bold patterns of stripes, spots, or blotches. • †Anther bed is not hooded
	Zebrinae	• *P. corningiana* • ***P. inscriptiosinensis**** • *P. speciosa* • ***P. sumatrana**** • ***P. tetraspis***	• †Similar to *Amboinenses*, but possessing a cucullate clinandrium (hooded anther bed).

TABLE 2.1 continued

Subgenus	Section	Species	Description
Proboscidioides		• ***P. lowii****	• Biseriate callus • Flowers white to light pink with a dark pink/purple lip • Lateral lobes in the shape of recurved hooks and possessing slightly raised teeth • Small, deciduous, epiphytic plants. • Petals much broader than sepals • 4 semiglobular pollinia. • †Long, beak-like rostellum that is much longer than the stigma, almost as long as the entire column, and almost at a right angle to the column.
Aphyllae previously in *Kingidium*		• ***P. braceana*** • ***P. hainanensis*** • ***P. honghenensis*** • ***P. minus*** • ***P. stobartiana**** • ***P. taenialis*** • ***P. wilsonii***	• Biseriate callus. • Flowers green to pink. • Flap-like tooth on lateral lobes of the lip. • Small, deciduous, epiphytic plants. • Sepals and petals similar to subsimilar in size. • 4 semiglobular pollinia. • †Possesses a prominent to obscure spur. • †Persistent, fleshy, chlorophyllous perianth after pollination.
Parishianae		• *P. appendiculata* • ***P. gibbosa*** • ***P. lobbii*** • ***P. parishii***	• Biseriate callus. • Flowers white and pink or yellow/brown. • Lateral lobes of the lip erect, subparallel, and diverging at the middle to form a U-shaped compound structure. • Small, deciduous, epiphytic plants. • Sepals and petals similar to subsimilar in size. • 4 semiglobular pollinia. • †Prominent swellings at the base of the column (column wings). • †Mobile lip midlobe

MORPHOLOGY OF *PHALAENOPSIS*

Vegetatively, *Phalaenopsis* are all similar. They are monopodial with short or acaulous stems that are completely enclosed by the leaf bases and are often found growing on the vertical axis of their host tree with their leaves hanging pendulously from horizontal stems. Species in the terrestrial section *Esmeralda* have longer, erect stems. Some species of *Phalaenopsis* may form basal plantlets or plantlets arising from a flower spike (a keiki). Their roots are round to oval in cross section, sparsely branched, and possess a velamentous epidermis (Benzing et al. 1983). Often roots will possess elongated root tips that are purple or green in color. The leaves of *Phalaenopsis* are the primary water storage organ, and thus they are thick and succulent. The deciduous species in subgenera *Aphyllae, Parishianae*, and *Proboscidioides* have thinner leaves, though they still appear to be the primary water storage organ. Leaves may be green to grey, purple, or blue-green with green or purple undersides and margins and may produce varying degrees of silver or purple speckles.

Phalaenopsis inflorescences may be small or large, branched panicles or unbranched racemes, arching, erect, or pendant, with many or few flowers that open simultaneously or sequentially. Some of these characters, especially those involved with size, tend to be variable even within an individual depending on the size and vigor of the plant. Other characters, such as simultaneous or sequential bloomers and arching, erect, or pendant inflorescence are important for recognizing species. Many *Phalaenopsis* have long-lived inflorescences, potentially lasting several years after the first round of flowers have senesced, and may produce a keiki or reinitiate growth and form new flower buds.

Flower form, color pigmentation and patterns are important for recognizing species and separating them into their infrageneric groups. Flowers tend to be long

lasting and more or less fleshy. Perianth parts are similar in color and size, however the lateral sepals may be narrower than the petals (subgenus *Polychilos*) or much larger than the petals (section *Phalaenopsis* and subgenus *Proboscidioides*). Subgenera *Aphyllae*, *Parishianae, Phalaenopsis,* and *Proboscidioides* have flowers that are generally white, pink, purple, brown, or green with minimal spotting. Many species in subgenus *Polychilos* possess brightly colored flowers, yellow, orange, pink, purple, and brown, with bold patterns of bars, spots, and splotches.

The lip of *Phalaenopsis* flowers is always three-lobed, and the size and shape of the lobes are important taxonomic characters. The lateral lobes of subgenus *Phalaenopsis* curve to form a cylinder around the column. The apices are rounded and the inner surfaces of the lobes are smooth. Subgenus *Polychilos* has lateral lobes that are subparallel to each other. The lateral lobes of subgenera *Polychilos, Parishianae, Proboscidioides,* and *Aphyllae* all bear an oblique tooth on the inner surface and their apices are variously toothed, as well.

The midlobe of the lip is extremely variable and useful for species identification. Section *Phalaenopsis* has variously shaped tendrils at its apex, *Deliciosae* possesses a notched apex, *Stauroglottis* and many species in sections *Amboinenses* and *Zebrinae* have variously shaped acute apices, while section *Fuscatae* has a rounded apex. *Esmeralda, Parishianae,* and *Polychilos* have obtuse apices resulting from an expanded midlobe. The number and morphology of calli is also taxonomically informative within *Phalaenopsis*. Calli are masses of tissue located at the base of the midlobe between the lateral lobes. There may be one, two, or three calli present. Calli may be bilobed, forked, glandular, fleshy, glabrous, or pubescent.

Like all orchids, the androecium and gymnoecium of *Phalaenopsis* are fused together into a column. The column of *Phalaenopsis* is slightly enlarged at the apex.

The base may also be expanded into a column foot, column knees, or column wings. Section *Zebrinae* is defined by having a hooded clinandrium (anther bed). All other *Phalaenopsis* species have unadorned anther beds. *Phalaenopsis* has a broad stigmatic cavity on the underside of the column that is uniform throughout the genus. The elongate rostellum is held over and parallel to the apex of the stigmatic cavity, with the exception of *P. lowii*, which has a dramatically elongated rostellum. The pollinarium is composed of visciduim, stipe, and pollina. Its morphology is uniform in *Phalaenopsis*, with the exception of *P. lowii* and section *Esmeralda*, which possess elongated stipes.

Phalaenopsis has an inferior ovary that develops into a dehiscent capsule, which opens via six median slits to release a multitude of tiny, dust-like seeds. The fruits of subgenera *Aphyllae* and *Polychilos* are noted for post-pollination chlorophylly of the perianth. In these species, the perianth becomes green and fleshy and remains attached to the ovary throughout the development of the fruit. The perianth of the other species of *Phalaenopsis* withers after pollination, but remains affixed to the ovary.

There may be either two or four pollinia in *Phalaenopsis*. Pollinia number has been a fairly confusing character in *Phalaenopsis*. Historically, *Phalaenopsis* was defined by having two pollinia, more or less cleft. Sweet, in his generic description of *Phalaenopsis* notes that in some cases pollinia are "…cleft, occasionally almost to base." Many species were moved between *Phalaenopsis* and *Kingidium* because some believed there were two deeply cleft pollinia while others felt there were four. Although the number of pollinia in certain species continues to be a subject of contention in *Phalaenopsis*, Christenson deemphasized pollinia number as a generic character, so regardless of whether they have two or four, they are still *Phalaenopsis*. Subgenus *Polychilos* and sections *Phalaenopsis* and *Stauroglottis* obviously have two cleft pollinia. *Esmeralda* bears four, distinct, globular pollinia in two pairs with the larger member of

the pair possessing a slight concavity in which the smaller pollinium fits. Subgenera *Parishianae* and *Aphyllae* and section *Deliciosae* are responsible for the confusion about pollinia number in the genus. These species have four pollinia in two pairs. One member of each pair is smaller than the other, and they are semiglobular with the flat sides pressing against each other. Although this pollinia type is similar in all of the species in *Aphyllae, Parishianae,* and *Deliciosae*, Sweet's classification, which defined *Phalaenopsis* as having two pollinia, placed some of these species in *Phalaenopsis* and some in *Kingidium*.

HYBRIDIZATION AND PHYLOGENETIC RECONSTRUCTION

Hybridization and Speciation

Hybridization is an important process that is responsible for as many as 4% of the speciation events in angiosperms (Otto and Whitton 2000). In order for a new lineage, and eventually a new species, to be formed by the mating of two distinct species, offspring must possess a genome that enables them either to selectively mate with other hybrids rather than parental species (as is the case with most polyploids) or to adapt to a different ecological niche from parental species, and hence become geographically isolated (as seen in diploid hybrid species). Reproductive isolation may be due to either premating or postmating factors. If the hybrid is ecologically adapted to an environment different from its parents, the hybrid is effectively isolated. On the other hand, hybrids may be reproductively isolated from, and sympatric with, their parents if genetic factors (such as ploidy level) prevent them from backcrossing with their parents (Levin 2000).

Hybrid speciation is usually divided into two classes: diploid (homoploid) hybrid speciation and allopolyploid speciation. Diploid hybrids result from normal sexual

reproduction between two different species, where haploid gametes combine to form a diploid offspring. In all known cases of diploid hybridization, both parents have the same base chromosome number, which is to be expected to allow for normal meiosis. For this reason diploid hybrids are difficult to recognize without additional information, such as morphological, cytological, or genetic evidence. It is likely that many diploid hybrid species are as of yet unrecognized because they are often morphologically dissimilar from either parent, and thus appear simply to be non-hybrid diploid species. This phenomenonon was demonstrated by Rieseberg's (1991) desert adapted *Helianthus anomalus*, which he showed to be a hybrid lineage from mesic parents, *H. annuus* and *H. petiolaris*.

Polyploid hybrid speciation, specifically allopolyploid hybrid speciation, is the formation of a hybrid lineage that has a complete diploid set of nuclear chromosomes from both parents. Autopolyploid speciation, the other form of polyploidy, is the formation of a lineage that has two complete diploid sets of chromosomes, but both parents are the same species. While this is a form of polyploid speciation, it is not hybrid speciation because only one parental lineage is involved, resulting in a traditional bifurcating speciation event. Allopolyploid hybrid speciation can occur via at least two different mechanisms. The first mechanism involves unreduced gametes in the parents resulting in two complete sets of chromosomes in the offspring. The second mechanism first involves the formation of a diploid hybrid via normal sexual reproduction and the formation of haploid gametes in parents of two different species. The diploid offspring may not undergo normal meiosis, however, because its chromosomes do not properly pair. If the diploid hybrid produces unreduced gametes forming a second generation, then each chromosome would have a pair and fertility would be restored in the resulting

polyploid offspring. This mechanism may also restore fertility in asexual odd-ploidy hybrids (e.g. triploids giving rise to hexaploid a generation).

Polyploid hybrids are in a sense 'instant species.' They are usually reproductively isolated from their diploid parents since backcrossing to either parent would produce a triploid generation that would most often be sterile. To become a successful species, however, polyploids must be able to reproduce. This may be a problem if they are sympatric with their parents, as many of the initial tetraploids' gametes will be wasted on cross-pollination with diploid parents. To overcome this obstacle, selfing is a convenient method of finding compatible gametes in close proximity. Many polyploids are self-fertile even if their diploid parents are not. Asexual reproduction is also possible for many plants, which can allow survival of the polyploidy until its numbers and density are sufficient to ensure sexual reproduction. Many polyploids also have the advantage of increased growth rate and size, as compared to their diploid parents. These factors have been found to increase the probability of polyploids persisting sympatrically among a larger population of diploids (Baack 2005). Alternatively, polyploids may possess a suite of characters that allow them to exploit a new ecological niche that is removed from their parents, and thus they do not need to compete.

Hybridization and Phylogenetics

Gene trees are contained within the branches of species trees (FIGURE 1.1) (Maddison 1997). A gene tree represents the evolutionary history of a particular gene (or DNA region), including nucleotide substitutions, insertions and deletions, gene losses and duplications, as well as inversions and transpositions. In the absence of recombination, gene trees have a bifurcating pattern of evolution. A species tree can be defined as the pattern of branching and reticulation of lineages due to speciation events. The process of

speciation results in gene copies that are split into different, independent lines of descent. Within each species lineage, contained gene trees continue to branch as they evolve through time. Although gene trees and species trees are not independent from each other, they can differ. This poses a challenge for molecular systematists who often assume that the gene tree is synonymous with the species tree and make taxonomic decisions based on these data. Many instances illustrate that phylogenies based upon different genes can have incongruent topologies for the same set of species (Doyle et al. 2003; Archambault and Bruneau 2004; Hamzeh and Dayanandan 2004; Vanderpoorten et al. 2004; Okuyama et al. 2005).

Incongruence among gene trees can be caused by several factors that can be divided into three general categories: stochastic factors, population level factors, and inter-species factors (Maddison 1997; Nakhleh et al. 2005). Stochastic factors include sampling error, insufficient data, and wrong assumptions or wrong models of evolution. These factors must be dealt with and eliminated early in phylogenetic analyses. Population level factors include lineage sorting, deep coalescence, sampling from orthologous and paralogous gene copies, and meiotic recombination within the region used for reconstruction. These events may lead to incongruent gene and species trees, even if the evolution of the species is following a tree-like branching pattern and the majority of genes should be congruent with the species tree (FIGURE 1.1A). Finally, inter-species factors, such as hybridization and horizontal gene transfer or introgression, are processes that involve two otherwise independent lineages and result in reticulations in the species tree rather than traditional branching (FIGURE 1.1B). In the case of hybridization, no gene tree will recover the true relationship between species, because the genes have evolved in a bifurcating fashion, while the species have evolved in a network-like fashion. However, since the nuclear genome contains DNA from both parental

genomes, it is possible to recover two distinct incongruent topologies (one for each parent) depending on the origin of the DNA regions that were examined.

Incongruence among gene trees is a powerful tool for detecting reticulate evolution. In order to reconstruct phylogenetic histories the processes leading to incongruent gene trees must be elucidated to distinguish population genetic factors from species-level factors. Population level processes would affect only a small part of the genome, so the majority of genes should represent the species tree. Hybridization events should cause the maternal and paternal portions of the genome to have different gene trees, neither of which reflects the true evolutionary history of the hybrid taxon. If we look at a large enough sample of genes, we would expect a more equal distribution of alternate gene trees from hybrid evolution than from population level processes (Maddison 1997; Nakhleh et al. 2005).

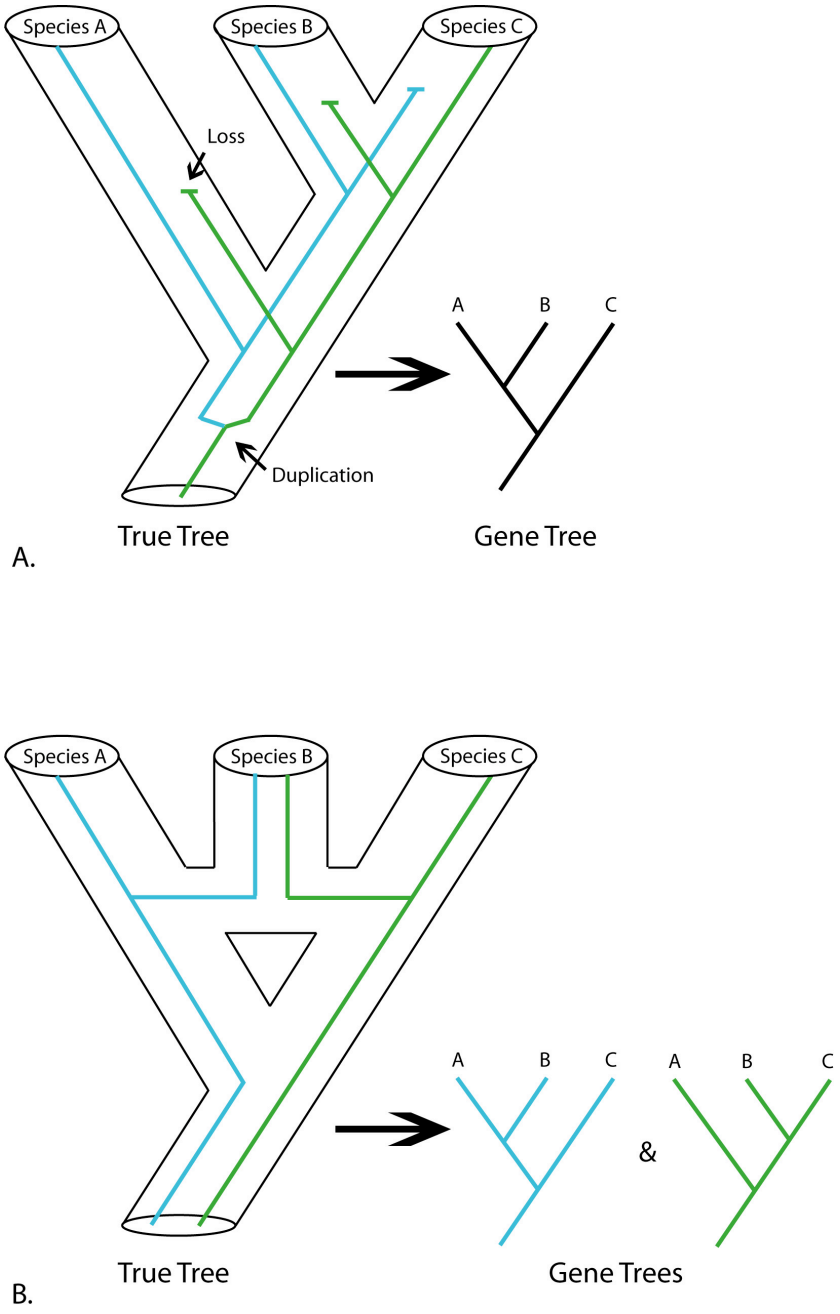


FIGURE 1.1. Species trees containing gene trees (in color) and the phylogenies that would be hypothesized from analyzing the genes. A. The gene tree and species tree are incongruent due to lineage sorting – gene duplication and subsequent losses in certain species. B. Hybridization between species A and C gives rise to species B. Species B possesses copies of both genes. A different topology will be recovered depending on which gene is examined.

OVERVIEW OF THE REMAINING CHAPTERS

The second chapter, "A Molecular Phylogeny of *Phalaenopsis* with implications for the classification of the genus and character evolution," presents the two phylogenies that we reconstructed using combined data from four chloroplast DNA regions and one nuclear marker. These phylogenies were used to assess the current infrageneric classification of *Phalaenopsis,* and addressed the following questions:

1. What constitutes a monophyletic *Phalaenopsis*? Should species previously considered *Doritis* and *Kingidium* be included within a more broadly defined *Phalaenopsis*?

2. Does Christenson's (2001) classification reflect the evolutionary history of the group, and if not, how can we revise *Phalaenopsis* to reflect a natural classification?

3. What evolutionary trends are evident in the following characters: geographic distribution, genome size, pollinia number, deciduous leaves, callus morphology and number, floral coloration, the presence of a raised tooth on the lateral lobes of the lip, a saccate lip base, reflexed tepals, and a hooded anther bed.

The third chapter, "A comprehensive search for phylogenetic markers that are broadly informative in angiosperms," discusses the method we used to find the primers (and the regions they amplify) that were used to reconstruct the nuclear phylogeny in chapter 2 and to assess the methods to detect hybridization in phylogenetic analyses in chapter 4. This was a collaborative project that involved constructing a database of the entire *Arabidopsis* and *Oryza* (rice) genomes and searching for conserved regions suitable as primers for PCR between the two genomes that may be present in *Phalaenopsis* and broadly across the angiosperms. One hundred-and-forty-eight primer

pairs were tested to see if we could amplify single-copy PCR products and obtain sequences with appropriate levels of variation for phylogenetic reconstruction.

The fourth chapter, "Using *Phalaenopsis* species and horticultural hybrids to test methods of reconstructing reticulate evolution in phylogenetic analyses," presents the results of rigorous tests of Neighbor-Net, implemented in SplitsTree, (Bryant and Moulton 2004; Huson and Bryant 2006). The following tests were performed to assess the program's performance under a variety of conditions:

1. No hybrids present.
2. One hybrid with both parents present.
3. Two hybrids, hybridizations independent of each other.
4. Two hybrids, hybridizations not independent of each other.
5. Three hybrids, hybridizations independent of each other.

Five datasets were generated for eight *Phalaenopsis* hybrids and thirteen species of *Phalaenopsis* that were scattered throughout the phylogeny in three of the four major clades. One dataset was from the chloroplast genome and four were from the nuclear genome. Hybrids for this study were selected from previously bred horticultural varieties whose parental species were known and included in the datasets. Chapter 4 represents an ongoing project that will be completed when RIATA and RGNet are ready for evaluation using our data.

LITERATURE CITED

ARCHAMBAULT, A. and A. BRUNEAU. 2004. Phylogenetic utility of the LEAFY/FLORICAULA gene in the Caesalpinioideae (Leguminosae): Gene duplication and a novel insertion. Systematic Botany 29: 609-626.

BAACK, E.J. 2005. To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. Heredity 94: 538-546.

BENTHAM, G. 1883. *Phalaenopsis*. In G. Bentham and J.D. Hooker eds., Genera Plantarum, pp. L. Reeve & Co., London, England.

BENZING, D.H., W.E. FRIEDMAN, G. PETERSON, and A. RENFROW. 1983. Shootlessness, Velamentous Roots, And The Pre-Eminence Of Orchidaceae In The Epiphytic Biotope. American Journal Of Botany 70: 121-133.

BLUME, C.L. 1825. *Phalaenopsis*. Bijdragen Tot de Flora van Nederlandsch Indie 7: 294.

BRYANT, D. and V. MOULTON. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution 21: 255-265.

CAMERON, K.M., M.W. CHASE, W.M. WHITTEN, P.J. KORES, D.C. JARRELL, V.A. ALBERT, T. YUKAWA, H.G. HILLS, and D.H. GOLDMAN. 1999. A phylogenetic analysis of the Orchidaceae: Evidence from rbcL nucleotide sequences. American Journal of Botany 86: 208-224.

CHRISTENSON, E.A. 1995. An overview of the genus *Phalaenopsis*. Orchid Digest Jan-Feb-Mar: 19-22.

_____. 2001. *Phalaenopsis*: a Monograph. Timber Press, Portland, Oregon.

DE WIT, H.C.D. 1977. Orchids in Rumphius' Herbarium Amboinense. In J. Arditti ed., Orchid Biology: Reviews and Perspectives, pp. 47-94. Cornell University Press, Ithaca.

DOYLE, J.J., J.L. DOYLE, and C. HARBISON. 2003. Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: Phylogeny, gene duplication, and ancient polyploidy. Systematic Botany 28: 567-577.

GRUSS, O. and L. ROLLKE. 1993a. Gattung *Kingidium*. Die Orchidee 44: 33-36.

_____. 1993b. Gattung *Kingidium*, Teil 2. Die Orchidee 44: 119-122.

_____. 1995a. *Kingidium wightii* (Rchb.f.) Gruss & Rollke comb. nov., Teil 10. Die Orchidee 46: 23-28.

_____. 1995b. *Kingidium philippinense* (Ames) Gruss & Rollke comb. nov., Teil 11. Die Orchidee 46: 115-118.

_____. 1996. *Kingidium wilsonii* (Rolfe) Gruss & Rollke comb. nov. Die Orchidee 47: 149-155.

_____. 1997. *Kingidium chibae* (T. Yukawa) Gruss & Rollke comb. nov. Die Orchidee 48: 261-263.

HAMZEH, M. and S. DAYANANDAN. 2004. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast tRNT-tRNF region and nuclear rDNA. American Journal of Botany 91: 1398-1408.

HAWKES, A.D. 1966. *Grafia*, a new genus of orchids. Phytologia 13: 305-306.

HOLTTUM, R.E. 1965. Cultivated species of the orchid genus *Doritis* Lindl. Kew Bulletin 19: 207-212.

_____. 1966. The genus *Kingiella*. The Orchid Review 74: 128-129.

HUNT, P.F. 1970. Notes on Asiatic Orchids: V. *Kingidium* P. F. Hunt, nom. nov. Kew Bulletin 24: 97-98.

HUSON, D.H. and D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

HUYNH, T.N.D., J. JANSSON, N.B. NGUYEN, and W.K. SUNG. 2005. Constructing a smallest refining galled phylogenetic network, Research In Computational Molecular Biology, Proceedings, pp. 265-280.

LEVIN, D.A. 2000. The Origin, Expansion and Demise of Plant Species. Oxford University Press, New York.

LINDLEY, J. 1833. *Doritis*, Genera and Species of Orchidaceous Plants, pp. 178. Ridgways, Piccadilly, London.

LINNAEUS, C. 1753. *Epidendrum amabile*. Species Plantarum ed. 1: 953-954.

MADDISON, W.P. 1997. Gene trees in species trees. Systematic Biology 46: 523-536.

NAKHLEH, L., D. RUTHS, and H. INNAN. 2005. Gene trees, species trees, and species networks. Chapman & Hall, Norwell, MA.

OKUYAMA, Y., N. FUJII, M. WAKABAYASHI, A. KAWAKITA, M. ITO, M. WATANABE, N. MURAKAMI, and M. KATO. 2005. Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Molecular Biology and Evolution 22: 285-296.

OTTO, S.P. and J. WHITTON. 2000. Polyploid incidence and evolution. Annual Review Of Genetics 34: 401-437.

PFITZER. 1889. Orchidaceae: *Phalaenopsis* Bl. In A. Engler and K. Prantl eds., Die natürlichen Pflanzenfamilien, Teil II, pp. 211-212. Verlag von Wilhelm Engelmann, Leipzig.

REICHENBACH, H.G. 1862. Uebersicht der Gattungen *Phalaenopsis, Doritis, Stauropsis*. Xenia Orchidacea: 1-8, t. 101; 139-147, t. 151, t. 156.

_____. 1864. Neue Orchideen, *Phalaenopsis pantherina*. Botanische Zeitung 39: 298.

_____. 1879. New species of *Phalaenopsis*. The Gardeners Chronicle n.s., 11: 398.

RIESEBERG, L.H. 1991. Homoploid Reticulate Evolution in *Helianthus* (Asteraceae): Evidence from Ribosomal Genes. American Journal of Botany 78: 1218-1237.

ROLFE, R.A. 1886. A revision of the genus *Phalaenopsis*. The Gardeners Chronicle 26: 168-170; 212-213; 267.

_____. 1917. *Doritis pulcherrima*. The Orchid Review 25: 195-197.

SEIDENFADEN, G. 1988. Orchid Genera in Thailand XIV. Fifty-nine vandoid genera. Opera Botanica 95: 31-34; 182-189; 236-241.

SHIM, P.S. 1982. A new generic classification in the *Phalaenopsis*-complex (Orchidaceae). Malayan Nature Journal 36: 1-28.

SMITH, J.J. 1933. Enumeration of the Orchidaceae of Sumatra and neighboring islands: *Phalaenopsis*. Repertorium specierum novarum regni vegetabilis ("Feddes Repertorium") 32: 364-367.

SWEET, H.R. 1968. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 37: 867-877, 1089-1104.

_____. 1969. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 38: 33-43, 225-239, 321-336, 505-519, 681-694, 888-901.

_____. 1970. *Phalaenopsis* Section *Parishianae*. American Orchid Society Bulletin 39: 1094-1097.

_____. 1980. The Genus *Phalaenopsis*. Orchid Digest, Pomona, CA.

VANDERPOORTEN, A., A.J. SHAW, and C.J. COX. 2004. Evolution of multiple paralogous adenosine kinase genes in the moss genus *Hygroamblystegium*: phylogenetic implications. Molecular Phylogenetics and Evolution 31: 505-516.

Chapter 2: A Molecular Phylogeny of *Phalaenopsis* (Orchidaceae) With Implications for the Classification of the Genus and Character Evolution

INTRODUCTION

Phalaenopsis Bl. is a genus of orchids consisting of 63 species that occur primarily as epiphytes throughout Southeast Asia and the Pacific Islands (Christenson 2001). Their showy, exotic flowers and ease of cultivation have made *Phalaenopsis* one of the most widely traded groups of horticultural plants throughout the world (Griesbach 2002). *Phalaenopsis* has been the subject of many taxonomic treatments and authors have segregated several of its species groups into separate genera (Rolfe 1917; Hawkes 1966; Sweet 1980; Shim 1982; Gruss and Rollke 1993a, 1993b).

Phalaenopsis is in the Orchidaceae, subfamily Epidendroideae, tribe Vandeae, subtribe Aeridinae (Cameron et al. 1999). This dissertation will use the taxonomy of the most current classification of *Phalaenopsis*, Christenson's 2001 monograph. Therein, he redefined *Phalaenopsis* more broadly than many authors (see Taxonomic History of *Phalaenopsis* in chapter 1) and included species previously segregated into the genera *Doritis* Lindl. and *Kingidium* Hunt. He deemphasized the taxonomic utility of pollinia number, which had historically been an important taxonomic character in the Orchidaceae, but has recently been shown to be relatively plastic throughout the family.

The genus *Doritis* has been a controversial and confusing entity for more than a century due to misplaced type specimens, confusing generic definitions, and disagreements about species delimitations. *Doritis* species are terrestrial or lythophytic on moss and morphologically distinct from other *Phalaenopsis* by having an erect growth habit and inflorescence, four globular pollinia, a long column and rostellum, lateral lobes

of the labellum that are greatly reduced into 'antenna' and a labellum midlobe that is expanded into three lobules, two of which mimic the true lateral lobes.

After discovering additional type material of *Doritis* that clarified the definition of the genus, Rolfe erected the genus *Kingiella* Rolfe (now *Kingidium*) and moved several species that had been treated as *Doritis* into this new genus. Species placed in *Kingidium* were small plants with deciduous leaves, and tended to grow in montain forests. They possessed a saccate lip base and four semiglobular pollinia. One of *Phalaenopsis'* defining characters, according to Sweet (1980), was possession of only two cleft pollinia, and so several more *Phalaenopsis* species were moved into the genus *Kingidium* (Seidenfaden 1988; Gruss and Rollke 1993b, 1993a, 1995b, 1995a, 1996, 1997).

A suite of vegetative and floral characteristics rather than a single synapomorphy defines *Phalaenopsis.* *Phalaenopsis* species are monopodial with short stems and succulent, fleshy leaves. The lip of the flower is always three-lobed, and there are one to three calli present at the base of the midlobe between the lateral lobes of the lip.

Christenson divided the genus into five subgenera (TABLE 2.1) based on flower color and pigmentation patterns, lip and callus structure, number of pollinia, and the presence or absence of deciduous leaves. These subgenera include *Polychilos* (35 species), *Phalaenopsis* (15 species, including *Doritis* and some *Kingidium*), *Aphyllae* (7 species, including some *Kingidium*), *Parishianae* (4 species), and *Proboscidioides* (1 species). Subgenus *Polychilos* was further divided into four sections, *Polychilos* (4 species), *Fuscatae* (4 species), *Amboinenses* (22 species), and *Zebrinae* (5 species). Subgenus *Phalaenopsis* was also divided into four sections, *Phalaenopsis* (6 species), *Stauroglottis* (3 species), *Deliciosae* (3 species previously *Kingidium*), and *Esmeralda* (3 species previously *Doritis*). This classification, like all classifications of *Phalaenopsis*, was based on morphological characters.

Fu et al. (1997) conducted the first molecular phylogenetic study of *Phalaenopsis* using RAPD data from 16 species. Their results placed *Doritis* outside of *Phalaenopsis*, and showed *P. equestris* to be closely related to *P. aphrodite*. Cameron et al's (1999) phylogeny of the Orchidaceae included one species of *Phalaenopsis,* which occurred as expected with other species in the subtribe Aeridinae, however this study did not provide any insight into the infrageneric relationships of the group.

Early karyotype analyses found little variation in the number of chromosomes among *Phalaenopsis* species, but wide variation in their size (Woodard 1951; Sagawa 1962; Shindo and Kamemoto 1963). *Phalaenopsis* has a diploid chromosome number of $2n$=38 for all wild species except *Phalaenopsis buyssoniana*, the natural autotetraploid of *P. pulcherrima*, $2n$=76. Shindo and Kamemoto (1963) examined nine species and found they could be split into two groups based on the size of their chromosomes. Seven species, all from the Philippines, had relatively small chromosomes and more or less symmetrical karyotypes. The second group consisted of three species that did not occur in the Philippines. Their chromosomes were found to be two to three times larger and less symmetrical than those of the Philippine species. A more recent study found similar levels of variation in chromosome size (Kao et al. 2001). Flow cytometry studies examining the nuclear DNA content of 19 species of *Phalaenopsis* found a six-fold difference in the amount of DNA from *P. sanderiana* (2.74 pg/diploid nuclear DNA content (C2)) to *P. parishii* (16.61 pg/2C, Lin et al. 2001). It has been hypothesized that small genome size is beneficial to colonizing organisms due to faster growth and development resulting from a shorter mitotic cycle (Evans and Rees 1971; Bennet 1972). A comprehensive study of genome size in Macaronesian angiosperms found that there was significantly smaller genome sizes in Macaronesian endemics than in non-Macaronesian representatives at various taxonomic levels (Suda et al. 2005).

This chapter provides two molecular phylogenies, one constructed from the combined data of four regions of the chloroplast genome and one from an actin gene in the nuclear genome. Chloroplast and nuclear data were not combined because there were multiple sequences of the nuclear region within an individual that were not always in the same clade, whereas the chloroplast data had only one sequence per individual. Rather than eliminating certain nuclear sequences, we chose reconstruct separate nuclear and chloroplast phylogenies. These phylogenies have been used to assess the current classification of *Phalaenopsis*. Specifically, we will address the following questions. Is *Phalaenopsis,* including species previously considered *Doritis* and *Kingidium,* a monophyletic group? Does Christenson's (2001) classification reflect the evolutionary history of the group, and if not, how can we revise *Phalaenopsis* to reflect a natural classification? In addition, we will use the phylogeny obtained from the combined chloroplast DNA data to examine geographic distribution and several characters, including genome size, pollinia number, deciduous leaves, callus morphology and number, and lip and tepal characteristics in an evolutionary context. We will test if smaller genomes are correlated with an island distribution and specifically a Philippine island distribution, as has been hypothesized by Shindo and Kamemoto (1963).

TABLE 2.1. Christenson's classification of the genus *Phalaenopsis*. Species in bold face are included in this project's analyses. Species marked with an asterisk (*) have not yet been positively identified. Characters marked with † are unique and pertinent only to that group.

Subgenus	Section	Species	Description
Polychilos	4 sections	35 species	• Biseriate or triseriate callus. • Flowers brightly colored and patterned, fragrant, fleshy and long lasting • Lateral lobes of the lip with a raised tooth at their apex. • Large or small, evergreen, epiphytes. • Petals and sepals subequal and subsimilar or petals slightly narrower than petals. • 2 cleft pollinia. • †Persistent, fleshy, chlorophyllous perianth after pollination.
	Polychilos	• ***P. borneënsis*** • ***P. cornu-cervi*** • ***P. mannii*** • ***P. pantherina*** • ***P. thalebanii* (=*P. cornu-cervi*)**	• Petals narrower than sepals. • Triseriate callus. • Flowers yellow with brown-red bars or spots. • †Fleshy flattened rachis (except *P. mannii*). • †Non-fragrant flowers. • †Flowers produced singly in succession over a long period. • †Pair of fleshy knees at the column's base. • †Midlobe of lip transversely anchoriform or lunate.
	Fuscatae	• ***P. cochlearis*** • ***P. fuscata*** • *P. kunstleri* • ***P. viridis****	• Petals and sepals similar to subsimilar and revolute. • Biseriate callus • Pale yellow flowers with brown markings. • †Concave lip with a longitudinal keel.
	Amboinenses	• ***P. amboinensis*** • ***P. bastianii*** • ***P. bellina*** • ***P. doweryënsis*** • ***P. fasciata*** • ***P. fimbriata**** • ***P. floresensis**** • ***P. gigantea**** • ***P. hieroglyphica*** • ***P. javanica*** • ***P. lueddemanniana*** • *P. luteola* • ***P. maculata**** • ***P. mariae**** • ***P. micholitzii*** • ***P. modesta**** • ***P. pallens**** • ***P. pulchra*** • *P. reichenbachiana* • *P. robinsonii* • ***P. venosa*** • ***P. violacea***	• Petals and sepals similar to subsimilar or with petals narrower than sepals. Flowers may be rotuliform or stellate. • Biresiate or Triseriate callus • Flowers variously colored, often with bright pigmentation and bold patterns of stripes, spots, or blotches. • †Anther bed is not hooded
	Zebrinae	• *P. corningiana* • ***P. inscriptiosinensis**** • *P. speciosa* • ***P. sumatrana**** • ***P. tetraspis***	• †Similar to *Amboinenses*, but possessing a cucullate clinandrium (hooded anther bed).

TABLE 2.1 continued

Subgenus	Section	Species	Description
Phalaenopsis	4 sections	15 species 1 natural hybrid	• Uniseriate callus. • Essentially unmarked white or pink flowers. • Smooth lateral lobes without a tooth-like ridge (except section *Deliciosae).* • Large or small, evergreen, epiphytic, terrestrial or lithophyic plants. • Sepals and petals similar to subsimilar or petals much broader than sepals • 2 or 4 pollina
	Phalaenopsis	• ***P. amabilis*** • ***P. aphrodite*** • ***P. xintermedia*** • ***P. philippinensis*** • ***P. sanderiana*** • ***P. schilleriana*** • ***P. stuartiana***	• Petals much broader than sepals. • Midlobe of the lip with a pair of appendages at the apex (cirrhi). • Possess a single, prominent, erect, glossy callus. • No teeth on lateral lobes of the lip • 2 cleft pollinia. • Epiphytes.
	Stauroglottis	• ***P. celebensis**** • ***P. equestris*** • ***P. lindenii***	• Similar to section *Phalaenopsis* with smaller flowers • Sepals and petals similar in size • No appendages on the lip apex. • No teeth on lateral lobes of the lip • 2 cleft pollinia. • Epiphytes.
	Deliciosae	• ***P. chibae*** • ***P. deliciosa*** • *P. mysorensis*	• Sepals and petals similar to subsimilar in size • No appendages on the lip apex • Lateral lobes of the lip with tooth-like flap. • 4 semiglobular pollinia. • Subsaccate lip base. • Epiphytes.
	Esmeralda	• ***P. buyssoniana**** • ***P. pulcherrima*** • *P. regnieriana*	• Sepals and petals similar to subsimilar in size • No appendages on the lip apex • Linear 'appendages' toward the base of the lip are the true lateral lobes of the lip. • Midlobe of the lip expanded into 3 lobules. • 4 globular pollinia. • Terrestrial to lithophytic (with moss) • †Long column foot. • †Long rostellum.

TABLE 2.1 continued

Subgenus	Section	Species	Description
Proboscidioides		• ***P. lowii****	• Biseriate callus • Flowers white to light pink with a dark pink/purple lip • Lateral lobes in the shape of recurved hooks and possessing slightly raised teeth • Small, deciduous, epiphytic plants. • Petals much broader than sepals • 4 semiglobular pollinia. • †Long, beak-like rostellum that is much longer than the stigma, almost as long as the entire column, and almost at a right angle to the column.
Aphyllae		• ***P. braceana*** • ***P. hainanensis*** • ***P. honghenensis*** • ***P. minus*** • ***P. stobartiana**** • ***P. taenialis*** • ***P. wilsonii***	• Biseriate callus. • Flowers green to pink. • Flap-like tooth on lateral lobes of the lip. • Small, deciduous, epiphytic plants. • Sepals and petals similar to subsimilar in size. • 4 semiglobular pollinia. • †Possesses a prominent to obscure spur. • †Persistent, fleshy, chlorophyllous perianth after pollination.
Parishianae		• *P. appendiculata* • ***P. gibbosa*** • ***P. lobbii*** • ***P. parishii***	• Biseriate callus. • Flowers white and pink or yellow/brown. • Lateral lobes of the lip erect, subparallel, and diverging at the middle to form a U-shaped compound structure. • Small, deciduous, epiphytic plants. • Sepals and petals similar to subsimilar in size. • 4 semiglobular pollinia. • †Prominent swellings at the base of the column (column wings). • †Mobile lip midlobe

MATERIALS AND METHODS

Sampling

One hundred and one individuals were included in this study (Appendix A, TABLE A.2.1), 53 of the 63 *Phalaenopsis* species (with replicates), including all eight species previously placed in *Kingidium* and two of the three species in *Doritis,* and eight outgroup taxa from seven genera shown to be closely related to *Phalaenopsis* (Cameron et al. 1999; Whitten and Calsward pers com.): two species of *Paraphalaenopsis* and one each from *Sarcoglyphis, Amesiella, Aerides, Renanthera, Ascocentrum,* and *Neofinetia*. The identity of plants obtained through private or commercial growers were confirmed as they bloomed, and vouchers have been placed in The University of Texas Herbarium (TEX).

DNA Extraction, PCR Amplification, and Sequencing

Total DNA was extracted either from fresh or silica dried tissue using the CTAB protocol of Doyle and Doyle (1987) or the Qiagen Plant DNA Extraction Kit. Extracts were cleaned with the QIAGEN QIAEX II Suspension kit if necessary. PCR protocols were modified from Mullis and Faloona (1987). Amplifications were visualized on 1.5% TBE agarose gels and purified with QIAGEN QIAquick PCR purification kits or Sephadex. Cycle sequencing reactions were performed using BigDye Terminator 3.0 and visualized on an MJ BaseStation.

Chloroplast Markers

Four regions from the chloroplast genome were used in order to obtain variation sufficient for phylogenetic reconstruction. Primers for two regions, *trn*D → *trn*E and *mat*K, were found in the literature (Demesure et al. 1995; Whitten et al. 2000). We developed primers for two other regions, *atp*H → *atp*F and *pet*B → *pet*D from the completed maize chloroplast genome available on GenBank (TABLE 2.2) (Maier et al. 1995; Benson et al. 2004). No single region was sufficient to provide complete resolution of *Phalaenopsis* so we combined our data for a total of 3564 aligned base pairs.

***atp*HF**

We developed primers that would amplify the 3' end of *atp*H, most of *atp*F, including an intron at the 5' end, and the intergenic region (FIGURE 2.1.A.). Coding regions specified in analyses included base pairs 1 – 43 and 184 – 394. The ATP synthase proteins, together, form the machinery responsible for H^+ transport across the chloroplast membrane, which generates the proton motive force needed to synthesize ATP. They are encoded by a series of genes, including *atp*H and *atp*F.

***trn*DE**

The *trn*D → *trn*E chloroplast region includes the genes *trn*D, *trn*Y, and *trn*E, and their intergenic regions (FIGURE 2.1.B.). These genes encode tRNA molecules: *trn*D-GUC encodes $tRNA^{Asp}$, *trn*Y-GUA encodes $tRNA^{Tyr}$, and *trn*E-UUC encodes $tRNA^{Glu}$ (Demesure et al. 1995). Coding regions specified in analyses included base pairs 86 – 148 and 352 – 564.

***pet*BD**

We found a region with suitable variation that anchored the forward primer in the 3' end of *pet*B and included most of *pet*D, including its group II intron, and the intergenic

region between *pet*B and *pet*D (FIGURE 2.1.C.). Coding regions were specified in two ways in separate analyses, first excluding the group II intron in the coding sequence (base pairs 195 – 203 and 906 – 1271) and then including the intron in the coding sequence (195 - 1271). Group II introns do not code for amino acids, but they are functionally constrained and do not mutate freely as intergenic regions and group I introns. These genes are part of the small subunit of the cytochrome b_6/f complex. A recent study (Lohne and Borsch 2005) independently found this region to have phylogenetic utility throughout the angiosperms.

***mat*K**

The *mat*K chloroplast gene codes for the protein maturase K, which probably assists in the splicing of chloroplast group II introns (Vogel et al. 1999). It is approximately 1550 base pairs in length and is several times more variable than *rbc*L in most angiosperms (Soltis et al. 1998). All of the *mat*K sequence we used came from the gene; no non-coding regions were included. Four of the primers used for this study were taken from Whitten et al. (2000) and four more were developed by us from sequence generated using the Whitten et al. primers (TABLE 2.2 AND FIGURE 2.1.D.).

Nuclear Marker

Actin gene (6FR)

The nuclear marker used in this study encodes an actin gene that is approximately 825 base pairs in length and contains a noncoding region from position 280 through 799 based on the annotation of matching GenBank sequences (Benson et al. 2004) (TABLE 2.2 AND FIGURE 2.1.E.). Primers for this gene were found by comparing the *Arabidopsis* and *Oryza* genomes and looking for conserved regions (see Chapter 3). The phylogenetic

utility of this gene in other groups is unknown. All PCR products obtained were cloned to ensure single-copy sequencing product. Ten colonies were selected from each individual for a second PCR amplification and only products of the appropriate size were sequenced. Multiple sequence types were obtained for some individuals, and if those sequences clustered together with good support, one was selected to be used in the phylogenetic analysis. Sequences from a single individual that were not monophyletic were retained in analyses.

TABLE 2.2: Primer sequences and sources that were developed for these phylogenies (see FIGURE 2.1).

PRIMER NAME	PRIMER SEQUENCE 5' → 3'
*atp*H	AGC TTT TAT GGA AGC TTT AA
*atp*F	TGA AAT AGA CAA CTC GCA CA
*pet*B F	TCT GCT TAC TGC CGT ATT TAT G
*pet*D 420F	GGC CGG TTC GCT TGA GGA A
*pet*D 463R	CCA CAT CAC TTA TTA TAG GG
*pet*D 980F	CGT TGA AAC TTG AGG AAA TGT C
*pet*D 1160R	CAT CCG GCT CGA GCA GCA AGA ATC
*pet*D R	CGG ATC CGC CGG TTC ACC AAT CA
*mat*K 180F	CTT CGK TGA ATG ATT CTA ACC
*mat*K 339F	TGT CGG ATC TAC TAA TAC CCT
*mat*K 360R	AGG GTA TTA GTA GAT CCG ACA
*mat*K 930R	ATT TCT TAC TAC CAA AGG
Actin gene – 6F	GAT GGA CAG GTG ATC ACC ATT GG
Actin gene – 6R	TAG AAG CAC TTC CTG TGG AC

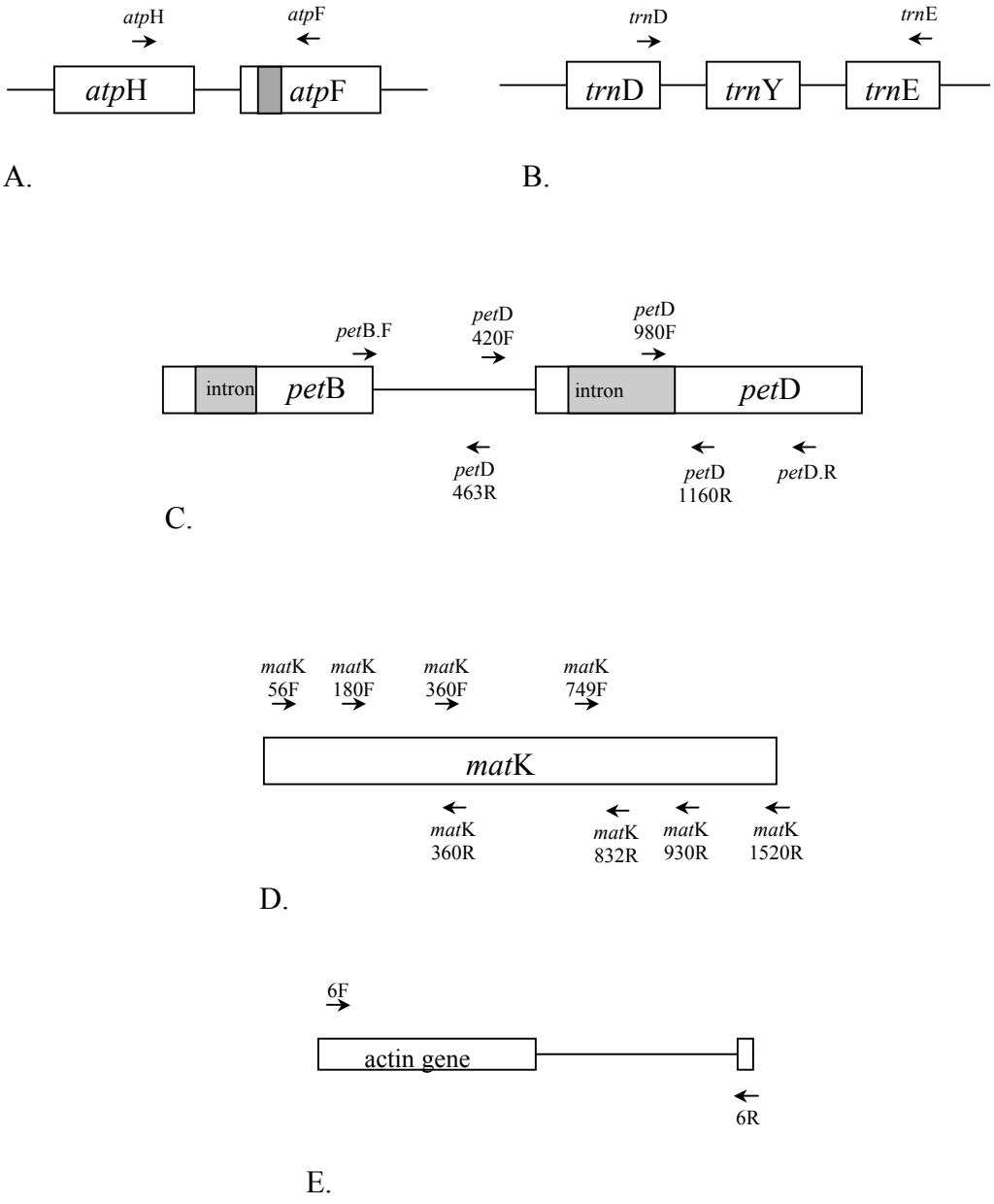


FIGURE 2.1. Map of the chloroplast regions used for phylogenetic reconstruction, and the primers used to amplify them. Novel primer sequences can be found in TABLE 2.2. A. *atp*HF. B. *trn*DE. C. *pet*D. D. *mat*K. E. nuclear actin gene.

Analyses

Sequences were assembled and edited in Sequencher 4.2. Initial alignments were carried out using ClustalX (Thompson et al. 1997), and then modified in MacClade 4.0 (Maddison and Maddison 2000). The four chloroplast markers were analyzed separately, combined, and with either only the coding region or only the noncoding region of the combined data. A partition homogeneity test (Farris et al. 1994) was performed to test for combinability between each chloroplast region using PAUP* 4.0 (Swofford 1996b). The nuclear marker was analyzed separately from the chloroplast data because multiple nuclear sequences for certain individuals made it difficult to combine with chloroplast data. Separate analyses were also conducted using either coding or noncoding regions of the nuclear data. Parsimony analyses were run in PAUP* 4.0 (Swofford 1996b) using a heuristic search with ten random-addition-replicates and TBR branch swapping. A nonparametric bootstrap analysis was performed to determine branch support with 1000 replicates and 1000 trees saved from each replicate.

Bayesian analyses were performed using MrBayes 3.0b4 (Huelsenbeck and Ronquist 2001). Appropriate models of evolution were determined separately for each marker using the likelihood ratio test implemented in ModelTest 3.06 (Posada and Crandall 1998). ModelTest determined that the Felsenstein 81 (F81) + gamma distributed rate variation model was appropriate for *trn*DE and *atp*HF and the general time reversible model (GTR) + invariant sites + gamma distributed rate variation was appropriate for *mat*K, *pet*BD, and the nuclear data set. Data were partitioned in the Bayesian chloroplast analysis with the appropriate model applied to each partition. Four MCMC chains were run for 3,000,000 generations, and one tree was saved every 100 generations. The first 50,000 generations prior to stationarity (as determined by eye)

were discarded as the burn in period for the chloroplast data and the first 3,000 generations were discarded for the nuclear data.

Parametric bootstrap analyses were performed on the combined chloroplast data and on the nuclear data to test whether we could reject the monophylly of subgenus *Phalaenopsis*, including sections *Esmeralda* and *Deliciosae*. A constraint was enforced that included all species in Christenson's (2001) subgenus *Phalaenopsis* in a monophyletic group. Two more tests were performed on the nuclear data to test if Christenson's subgenus *Polychilos* could be rejected as a monophyletic group and if *Polychilos* excluding section *Fuscatae* and *Phalaenopsis gigantea, P. doweryënsis,* and *P. maculata* in section *Amboinenses* could be rejected as a monophyletic group. One thousand replicates of sequences were simulated on the constraint trees with the best likelihood score for each test. A GTR model of evolution with unequal base frequencies and equal rates of among site rate heterogeneity was used to simulate sequences (Rambaut and Grassly 1997; Wilcox 2005). Simulations were analyzed using a heuristic search with ten random addition replicates and TBR branch swapping in PAUP* 4.0 (Swofford 1996b).

Thirteen characters were mapped onto the majority rule consensus tree and the Bayesian tree reconstructed by the chloroplast data using MacClade 4.0. Character states and the data matrix can be found in TABLES 2.3 and 2.4. All characters were treated as unordered character types. Pollinia types are shown in FIGURE 2.2, forked calli are shown in FIGURE 2.3, and raised teeth on the later lobes of the lip are shown in FIGURE 2.4. Character evolution was not analyzed in the context of the nuclear phylogeny because support values for many clades were low and the topologies of the parsimony and Bayesian analyses differed in integral parts of the tree, especially regarding the placement

of section *Esmeralda* and the resolution of subgenus *Polychilos*. In addition, sampling of the nuclear marker was not as thorough as it was for the chloroplast regions.

Genome size was analyzed both as an unordered, binary character and as a continuous character. Values were taken from two flow cytometry studies, (Jones et al. 1998; Lin et al. 2001). Genome size was coded as a binary character with genomes less than 7 pg/ diploid nuclear DNA content (2C) coded as 'small' and genomes greater than 9 pg/2C coded as 'large.' Geographic distributions of taxa were optimized on the chloroplast phylogeny using mainland distribution and island distribution in one test and Philippine distribution and non-Philippine distribution in another test. Genome size was tested as a continuous character in Mesquite 1.06 (Maddison and Maddison 2005) using the pairwise comparisons module (Maddison 2005). The tree was pruned to only those 20 species for which genome size data were available, and pairwise comparisons tests were performed using genome size as the independent variable and geographic distribution as the dependant variable (both Philippine/non Philippine and as island/mainland).

TABLE 2.3. Character states and symbols traced onto the chloroplast phylogeny (for data matrix see TABLE 2.4).

Character (Character letter in TABLE A.2.3)	Character State	Character State Symbol
Genome Size – Binary (A)	<7.0 pg/2C	0
	>9.0 pg/2C	1
	Unknown	?
Geographic Distribution (B)	Mainland	0
	Island	1
Geographic Distribution (C)	Philippine	0
	Non-Philippine	1
Deciduous leaves (D)	Evergreen	0
	Deciduous	1
Pollinia Type (E)	2 cleft pollinia	0
	4 semiglobular pollinia	1
	4 globular pollinia	2
Markings on petals and sepals (F)	Flowers with bright spots, bars, or blotches	0
	Flowers without conspicuous markings	1
Forked callus (G)	None	0
	At least 1	1
Callus number (H)	1	0
	2	1
	3	2
	0	3
Lateral lobes of the lip with a raised tooth (I)	Present	0
	Absent	1
Saccate lip base (J)	Present	0
	Absent	1
Reflexed tepals (K)	Present	0
	Absent	1
Hooded anther bed (L)	Present	0
	Absent	1
Petal/sepal width (M)	Petals and sepals equal or subequal	0
	Petals much wider than sepals	1

TABLE 2.4. Data matrix used to map characters onto combined chloroplast phylogeny. Character names and states defined in TABLE 2.3.

	Characters												
Species	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
Aerides multiflorum	?	0	1	0	0	1	0	3	1	0	1	1	0
Phalaenopsis amabilis	?	0&1	0&1	0	0	1	0	0	1	1	1	1	1
P. amboinensis	1	0	1	0	0	0	1	1	0	1	1	1	0
Amesiella philippinensis	?	1	0	0	0	1	0	3	1	0	1	1	0
P. aphrodite	?	1	0	0	0	1	0	0	1	1	1	1	1
Ascocentrum ampullaceum	?	0	1	0	0	1	0	3	1	0	1	1	0
P. bastianii	?	0&1	0	0	0	0	1	1	0	1	1	1	0
P. bellina	1	1	1	0	0	0	1	1	0	1	1	1	0
P. borneensis	?	1	1	0	0	0	1	2	0	1	1	1	0
P. braceana	?	0	1	1	1	1	1	1	0	0	1	1	0
P. buyssoniana	?	0	1	0	2	1	0	0	1	1	0	1	0
P. celebensis	?	1	1	0	0	1	0	0	1	1	1	1	0
P. chibae	?	0	1	0	1	1	0	0	1	0	0	1	0
P. cochlearis	?	0	1	0	0	1	1	1	0	1	1	1	0
P. cornucervi	0	0&1	1	0	0	0	1	2	0	1	1	1	0
P. deliciosa	?	0&1	0&1	0	1	1	1	1	0	0	1	1	0
P. doweryensis	?	1	1	0	0	0	1	1	0	1	1	1	0
P. equestris var. equestris	0	1	0	0	0	1	0	0	1	1	1	1	0
P. equestris var. rosea	0	1	0	0	0	1	0	0	1	1	1	1	0
P. fasciata3	0	1	0	0	0	0	1	1	0	1	1	1	0
P. fimbriata	?	0&1	1	0	0	1	1	2	0	1	1	1	0
P. floresensis	?	1	1	0	0	1	1	1	0	1	1	1	0
P. fuscata	?	0&1	0&1	0	0	0	1	1	0	1	1	1	0
P. gibbosa	?	0	1	1	1	1	1	1	1	1	0	1	0
P. gigantea	0	1	1	0	0	0	1	1	0	1	1	1	0
P. hainanensis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. hieroglyphica	?	1	0	0	0	0	1	1	0	1	1	1	0
P. honghenensis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. inscriptiosinensis	?	1	1	0	0	0	1	1	0	1	1	0	0
P. intermedia	?	1	0	0	0	1	0	0	1	1	1	1	1
P. javanica	?	1	1	0	0	0	1	1	0	1	1	1	0
P. lindenii	?	1	0	0	0	1	0	0	1	1	1	1	0
P. lobbii	?	0	1	1	1	1	1	1	1	1	0	1	0
P. lowii	?	0	1	1	1	1	1	1	0	1	1	1	1
P. lueddemanniana	0	1	0	0	0	0	1	1	0	1	1	1	0
P. maculata	?	0&1	1	0	0	0	1	1	0	1	1	1	0

P. mannii	1	0	1	0	0	0	1	2	0	1	1	1	0
P. mariae	0	1	0&1	0	0	0	1	1	0	1	1	1	0
P. micholitzii	0	1	0	0	0	1	1	1	0	1	1	1	0
P. minus	?	0	1	1	1	0	1	1	0	0	0	1	0
P. modesta	0	1	1	0	0	1	1	1	0	1	1	1	0
Neofinetia falcata	0	0	1	0	0	1	0	3	1	0	0	1	0
P. pallens	?	1	0	0	0	0	1	1	0	1	1	1	0
P. pantherina	?	0&1	1	0	0	0	1	2	0	1	1	1	0
Paraphalaenopsis laycockii	?	1	1	0	0	1	0	0	1	1	1	1	0
Paraphalaenopsis serpentilingua	?	1	1	0	0	1	0	0	1	1	1	1	0
P. parishii	1	0	1	1	1	1	1	1	1	1	0	1	0
P. philippinensis	?	1	0	0	0	1	0	0	1	1	1	1	1
P. pulcherrima	1	0	1	0	2	1	0	1	1	1	0	1	0
P. pulchra1	0	1	0	0	0	1	1	2	0	1	1	1	0
P. pulchra2	0	1	0	0	0	1	1	2	0	1	1	1	0
Renanthera vietnamica	?	0	1	0	0	0	0	0	1	0	1	1	0
P. sanderiana	0	1	0	0	0	1	0	0	1	1	1	1	1
Sarcoglyphis comberii	?	0	1	0	0	1	0	0	1	1	1	1	0
P. schilleriana	?	1	0	0	0	1	0	0	1	1	1	1	1
P. sp	?	?	?	?	?		?	?	?	?	?	?	?
P. stobartiana	?	0	1	1	1	1	1	1	0	0	1	1	0
P. stuartiana	0	1	0	0	0	1	0	0	1	1	1	1	1
P. sumatrana	0	0&1	1	0	0	0	1	1	0	1	1	0	0
P. taenailis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. tetraspis	?	0	1	0	0	1	1	2	0	1	1	0	0
P. thalebanii	?	0	1	0	0	0	1	2	0	1	1	1	0
P. venosa	1	1	1	0	0	0	1	1	0	1	1	1	0
P. violacea	?	0&1	1	0	0	1	1	1	0	1	1	1	0
P. viridis	?	1	1	0	0	0	1	1	0	1	1	1	0
P. wilsonii	?	0	1	1	1	1	1	1	0	0	1	1	0

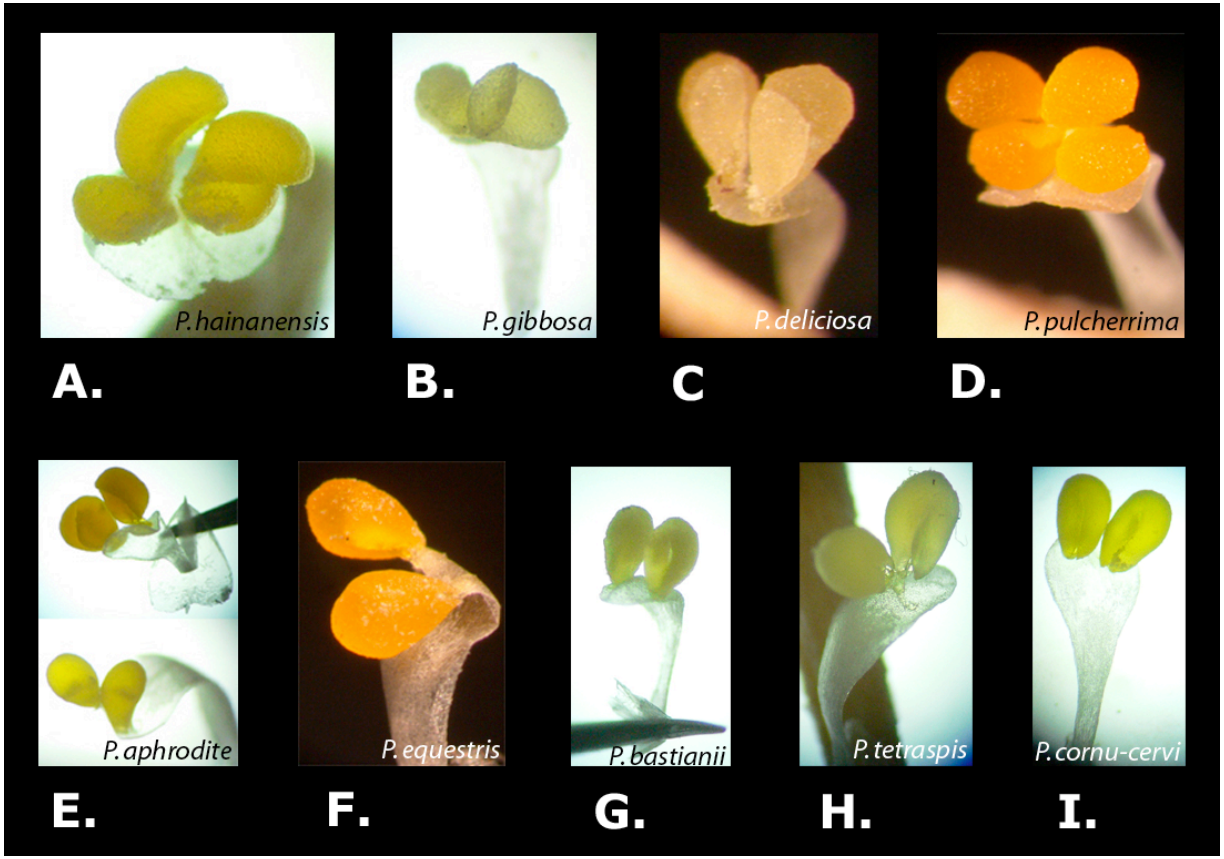


FIGURE 2.2. Examples of pollinia from different *Phalaenopsis* clades. Species in A-D possess four pollinia, and species in E-I possess two pollinia. A. *P. hainanensis* (subgenus *Aphyllae*) B. *P. gibbosa* (subgenus *Parishianae*) C. *P. deliciosa* (subgenus *Phalaenopsis* section *Deliciosae*) D. *P. pulcherrima* (subgenus *Phalaenopsis* section *Esmeralda*) E. *P. aphrodite* (subgenus *Phalaenopsis* section *Phalaenopsis*) F. *P. equestris* (subgenus *Phalaenopsis* section *Stauroglottis*) G. *P. bastianii* (subgenus *Polychilos* section *Amboinenses*) H. *P. tetraspis* (subgenus *Polychilos* section *Zebrinae*) I. *P. cornu-cervi* (subgenus *Polychilos* section *Polychilos*.

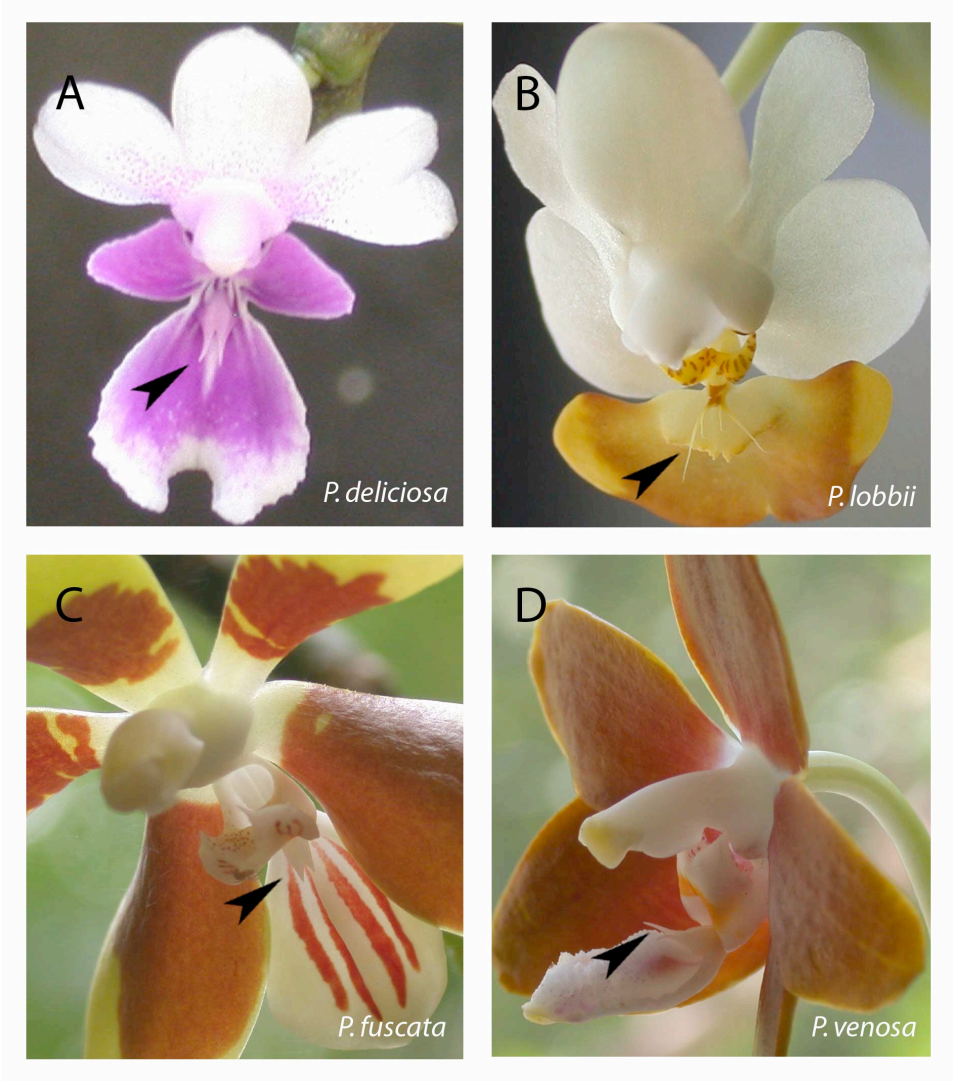


FIGURE 2.3. Examples of forked calli from different *Phalaenopsis* clades. A. *P. deliciosa* (subgenus *Phalaenopsis*, section *Deliciosae*) B. *P. lobbii* (subgenus *Parishianae*) C. *P. fuscata* (subgenus *Polychilos* section *Fuscatae*) D. *P. venosa* (subgenus *Polychilos* section *Amboinenses*)

FIGURE 2.4. Examples of raised teeth on the lateral lobes of the labellum from different *Phalaenopsis* clades. A. *P. deliciosa* (subgenus *Phalaenopsis,* section *Deliciosae*). B. *P. hainanensis* (subgenus *Aphyllae*). A and B possess a flap-like tooth. C. *P. hieroglyphica* (subgenus *Polychilos* section *Amboinenses*) D. *P. cornu-cervi* (subgenus *Polychilos* section *Polychilos*). C and D possess a firm, fleshy tooth.

RESULTS

Chloroplast data

The *atp*HF region used in these analyses consisted of 394 base pairs, the *trn*DE region of 564 base pairs, the *pet*D region of 1271, and the *mat*K region of 1335 base pairs after alignment. Of the 3564 total aligned base pairs, 256 were parsimony informative. After initial analyses, one individual from each species was selected from those species where multiple individuals clustered together with good support, leaving a total of 66 individuals for the chloroplast data set. Parsimony analyses of the combined chloroplast data resulted in 80,004 most-parsimonious trees with a length of 821, consistency index (CI) of 0.721, retention index (RI) of 0.830, and a homoplasy index (HI) of 0.279. Results of the partition homogeneity test (i.e., incongruence length difference test (ILD)) (Farris et al. 1994) indicated that none of the chloroplast markers should be combined with any other. Nevertheless, we have chosen to combine because our markers are of considerably different lengths. Variable length and high rates of heterogeneity have been shown to cause the ILD test to return spurious results (Dowton and Austin 2002). In addition, the chloroplast evolves as a single unit, suggesting that combining chloroplast data should be reasonable. Analyses of individual markers and their statistics can be found in Appendix A (FIGURES A.2.1 – A.2.4 AND TABLE A.2.4). Results of analyzing combined chloroplast data with only coding or only noncoding chloroplast regions can also be found in the appendix (FIGURES A.2.5 – A.2.9).

The combined chloroplast analyses were summarized as a majority rule consensus tree of 80,004 most parsimonious trees (FIGURE 2.5) and as the Bayesian majority-rule tree (FIGURE 2.6). These trees were very similar, differing in that the Bayesian analyses could not resolve the basal branch of the genus *Phalaenopsis* that unite the three clades C2, C3, and C4 in the parsimony analysis, though with no bootstrap support (marked

with an arrow ➢ in FIGURE 2.5). The Bayesian analysis also did not resolve one of the basal branches of clade C1 (subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis*) that was found by the parsimony analysis with low bootstrap support (marked with an arrow ➢ in FIGURE 2.5). Analyses of individual markers and analyses using combined chloroplast data with only coding or noncoding regions were all similar. Where topological differences occurred there was little or no support indicated from either bootstrap or Bayesian posterior probabilities (FIGURES A.2.1-A.2.9).

Analyses recovered a monophyletic genus *Phalaenopsis*, including species previously in *Doritis* and *Kingidium,* consisting of four clades, two highly supported and two with moderate support. The first clade, (C1) was a highly supported clade (bootstrap 100, posterior probability 100%) containing subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis*. The second clade (C2) was moderately supported (bootstrap 83, posterior probability 71%) and represented a monophyletic subgenus *Polychilos*. The third clade (C3) was highly supported (bootstrap 99, posterior probability 100%) and included subgenera *Aphyllae* (excluding *P. minus*) and *Proboscidioides*. The forth clade (C4) had low support (bootstrap 63, posterior probability 71%) and contained subgenus *Parishianae,* subgenus *Phalaenopsis* sections *Esmeralda* and *Deliciosa,* and *P. minus* (subgenus *Aphyllae*). Clades C2-C4 were grouped together in the majority rule consensus tree, however there was no support for this node from either parsimony bootstrap or Bayesian posterior probability.

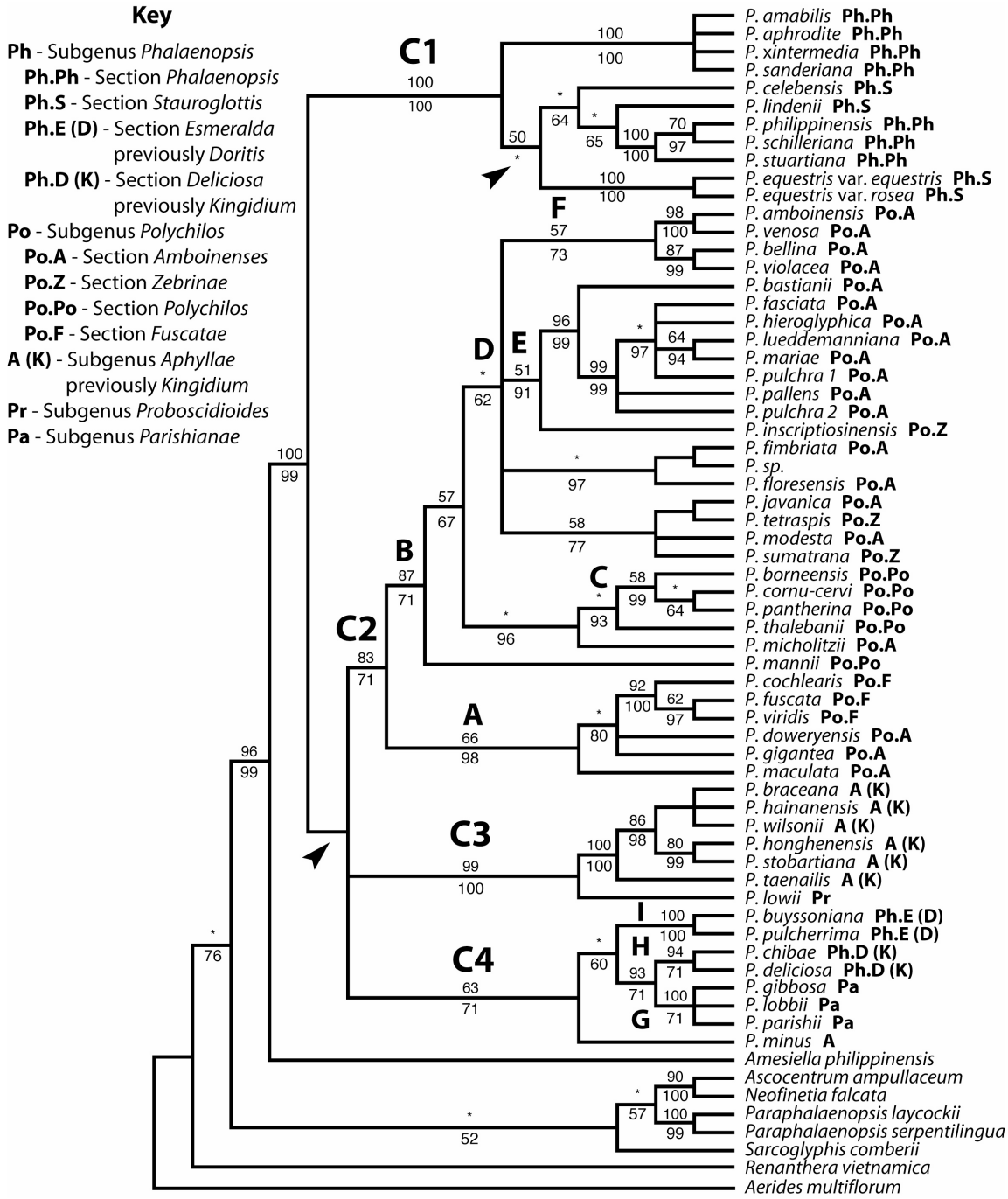


FIGURE 2.5. Majority rule consensus tree of 80,004 most parsimonious trees found for *Phalaenopsis* using combined chloroplast data. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches. Clades C1-C4 represent the major lineages within *Phalaenopsis* recovered by analyses using chloroplast data. The solid arrows ➢ represent branches that are not recovered by the Bayesian analyses.

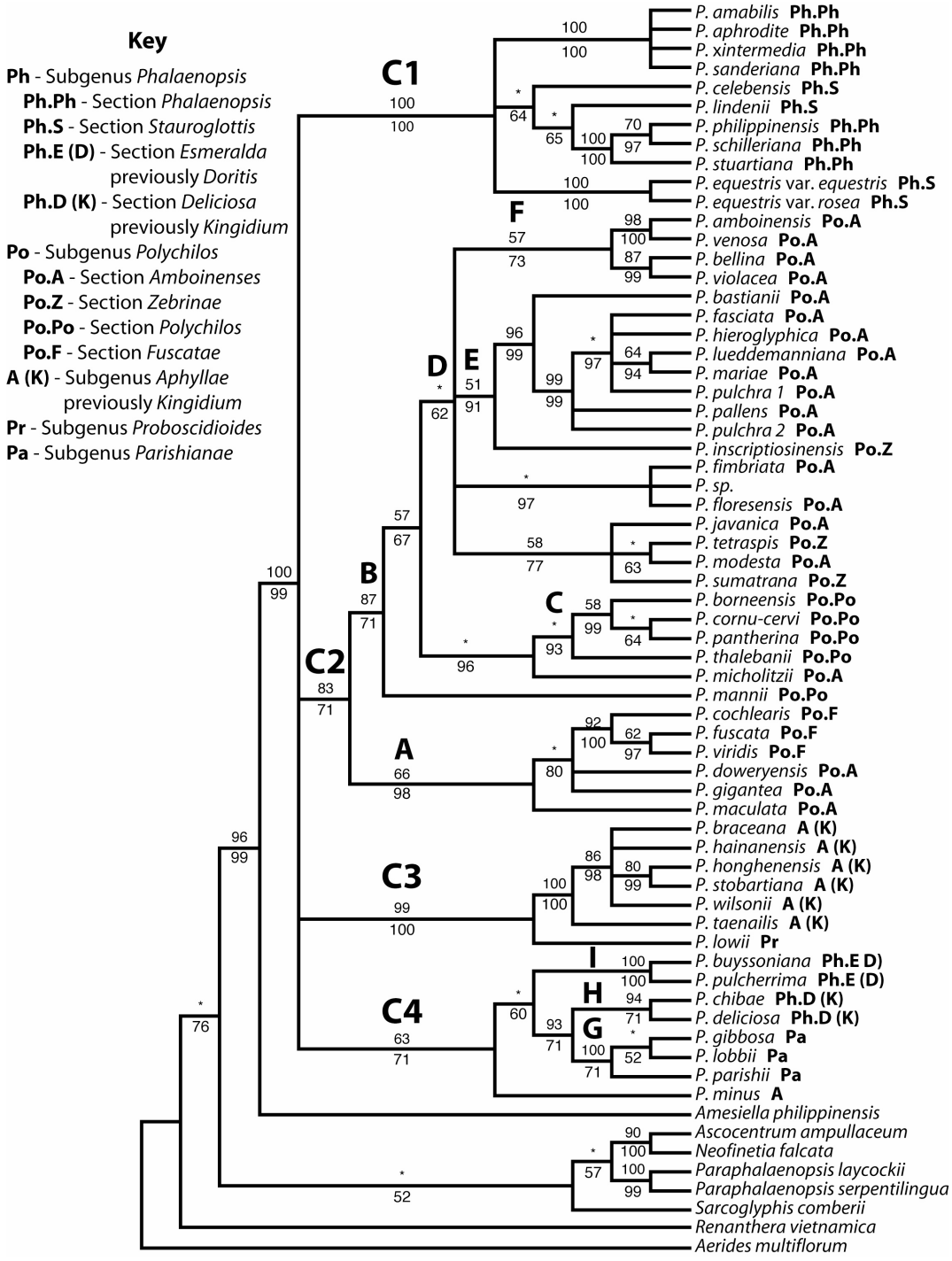


FIGURE 2.6. Majority rule consensus of trees found in Bayesian analysis for *Phalaenopsis* using combined chloroplast data. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches. Clades C1-C4 represent the major lineages within *Phalaenopsis* recovered by analyses using chloroplast data

Clade C1

Subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis* formed a monophyletic group that was recovered by both parsimony and Bayesian analyses. Neither section *Phalaenopsis* nor section *Stauroglottis* was found to be monophyletic. Subgenus *Phalaenopsis* sections *Esmeralda* and *Deliciosa* did not occur in this clade. They were found in clade C4, related to subgenus *Parishianae*. Parametric bootstrap analyses confirmed that subgenus *Phalaenopsis* was not monophyletic if *Esmeralda* and *Deliciosa* were included ($p < 0.001$).

Clade C2

Subgenus *Polychilos* (clade C2) is the largest subgenus of *Phalaenopsis*, and formed a moderately supported monophyletic group with bootstrap support of 83 and a posterior probability of 71%. In the combined analysis section *Fuscatae* formed a highly supported monophyletic group (bootstrap 92, posterior probability 100%) that was nested within a larger clade (C2.A) containing *Phalaenopsis doweryënsis*, *P. gigantea,* and, *P. maculata*, which are currently placed in subgenus *Polychilos* section *Amboinenses*. This clade had low support from the parsimony analysis with a bootstrap value of 66 but high support from the Bayesian analysis, with a posterior probability of 98% and was sister to the rest of the subgenus *Polychilos* (C2.B).

Section *Polychilos*, excluding *P. mannii,* was a monophyletic group (C2.C). It had moderate but insignificant (at $\alpha=0.05$) support from the Bayesian analysis, with a posterior probability of 93%; it was recovered in the majority rule tree from the parsimony analysis but not in the bootstrap consensus. *P. micholitzii* in section *Amboinenses* occurred as sister to clade C2.C. This clade (C2.C + *P. micholitzii*) was highly supported by the Bayesian analysis (posterior probability of 96%). *P. mannii* in section *Polychilos* occurred basal in clade C2.B.

The most derived clade in C2 consisted of a polytomy of four clades with section *Zebrinae* and section *Amboinenses* interdispersed throughout it (C2.D). Neither section formed a monophyletic group. Section *Amboinenses*, the largest section of subgenus *Polychilos*, had species distributed throughout all of the subclades in C2. There does appear to be a core group of *Amboinenses* that formed two monophyletic groups within clade C2.D. One morphologically similar group (C2.E) is sometimes called the '*lueddemanniana* complex,' and has historically been combined into a single species with several varieties. Herein, we will refer to the following species as members of this complex, *P. bastianii, P. fasciata, P. hieroglyphica, P. lueddemanniana, P. mariae, P. pallens,* and *P. pulchra.* These species formed a monophyletic group with high bootstrap support of 96 and a posterior probability of 99%. The second clade (C2.F) contained the species *P. bellina, P. violacea, P. venosa,* and *P. amboinensis*. This group had low support with a bootstrap value of 57 and a posterior probability of 73%.

Clade C3

Subgenus *Aphyllae,* with the exclusion of *Phalaenopsis minus*, formed a clade with bootstrap support of 100 and a posterior probability of 100%. The monotypic subgenus *Proboscidioides* was sister to subgenus *Aphyllae.* Together they formed one of the four major lineages within *Phalaenopsis* (clade C3) with bootstrap support of 99 and a posterior probability of 100%. The majority rule and bootstrap consensus trees from the parsimony analyses and the Bayesian majority rule consensus all placed *P. minus* as the basal taxon in clade C4. However, there were trees in the combined analysis that place *P. minus* basal to clade C3 (*Aphyllae* + *Proboscidioides*) that were of equal length to the trees summarized in the majority rule consensuses (i.e. some of the shortest trees hypothesized *P. minus* as the basal taxon in clade C3, although not the majority).

Clade C4

Clade C4 contained a highly supported, monophyletic subgenus *Parishianae* (C4.G, 100 bootstrap, 71% posterior probability) that was sister to subgenus *Phalaenopsis* section *Deliciosae* (C4.H) (bootstrap 93, posterior probability 71%). Subgenus *Phalaenopsis* section *Esmeralda* (C4.I) was sister to *Parishianae* + *Deliciosae* (C4.G + C4.H), but with weak support. *Phalaenopsis minus*, currently placed in subgenus *Aphyllae*, appeared to be the most basal member of clade C4. This node was only weakly supported, however, and majority rule consensus trees of separate analyses of *atp*HF, *pet*D, and the noncoding chloroplast region (not including the *pet*D group II intron, Appendix A, Figures A.2.1, A.2.4, and A.2.8, respectively) placed *P. minus* in a basal or unresolved position with respect to subgenera *Aphyllae, Polychilos* and *Phalaenopsis.*

Nuclear data

The nuclear data consisted of 827 aligned base pairs, 268 of which were parsimony informative. After initial analyses, one individual from each species was selected from those species where multiple individuals clustered together with good support. If multiple cloned sequences from a single individual clustered together with good support, only one was retained. Final analyses included 55 individuals and 68 sequences. Parsimony analyses resulted in 35,889 most-parsimonious trees with a length of 792, consistency index (CI) of 0.760, retention index (RI) of 0.815, and a homoplasy index (HI) of 0.240. Results of analyses using the complete alignment were summarized as a majority rule consensus tree of 35,889 most parsimonious trees (FIGURE 2.7) and as the majority rule consensus tree of the trees recovered by the Bayesian analysis (FIGURE 2.8). Results from the parsimony and Bayesian analyses were similar overall with two noteworthy differences, the placement of section *Esmeralda* (indicated by an arrow ➢),

and the resolution of clades N1.B, N1.C, N1.D (subgenus *Polychilos* sections *Amboinenses*, *Polychilos,* and *Zebrinae*). Analyses and parsimony statistics of only coding and only non-coding regions can be found in Appendix A (FIGURES A.2.10, A.2.11, AND TABLE A.2.4). Results from the coding region were relatively unresolved, and results from the noncoding region were consistent with those from the complete data set.

The nuclear data, like the chloroplast data, recovered a monophyletic genus *Phalaenopsis* including those species that had been placed in *Doritis* and *Kingidium*. This was supported with a parsimony bootstrap value of 89 and a posterior probability of 100%. Within *Phalaenopsis*, the relationships of the species were not the same as those recovered by the chloroplast data. This was complicated because several individuals, especially within subgenera *Polychilos* and *Aphyllae*, have multiple cloned sequences that fell out in more than one clade. Two major clades were found within *Phalaenopsis*. The first clade (N1) was highly supported with a bootstrap value of 99 and posterior probability of 100%. It contained species from subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis* and most of the species from subgenus *Polychilos*, excluding section *Fuscatae* and two species from section *Amboinenses.* The second clade (N2) contained species from subgenera *Aphyllae, Parishianae, Phalaenopsis* sections *Esmeralda* and *Deliciosae,* and *Polychilos* section *Fuscatae.* It had a bootstrap value of 84 and posterior probability of 100%.

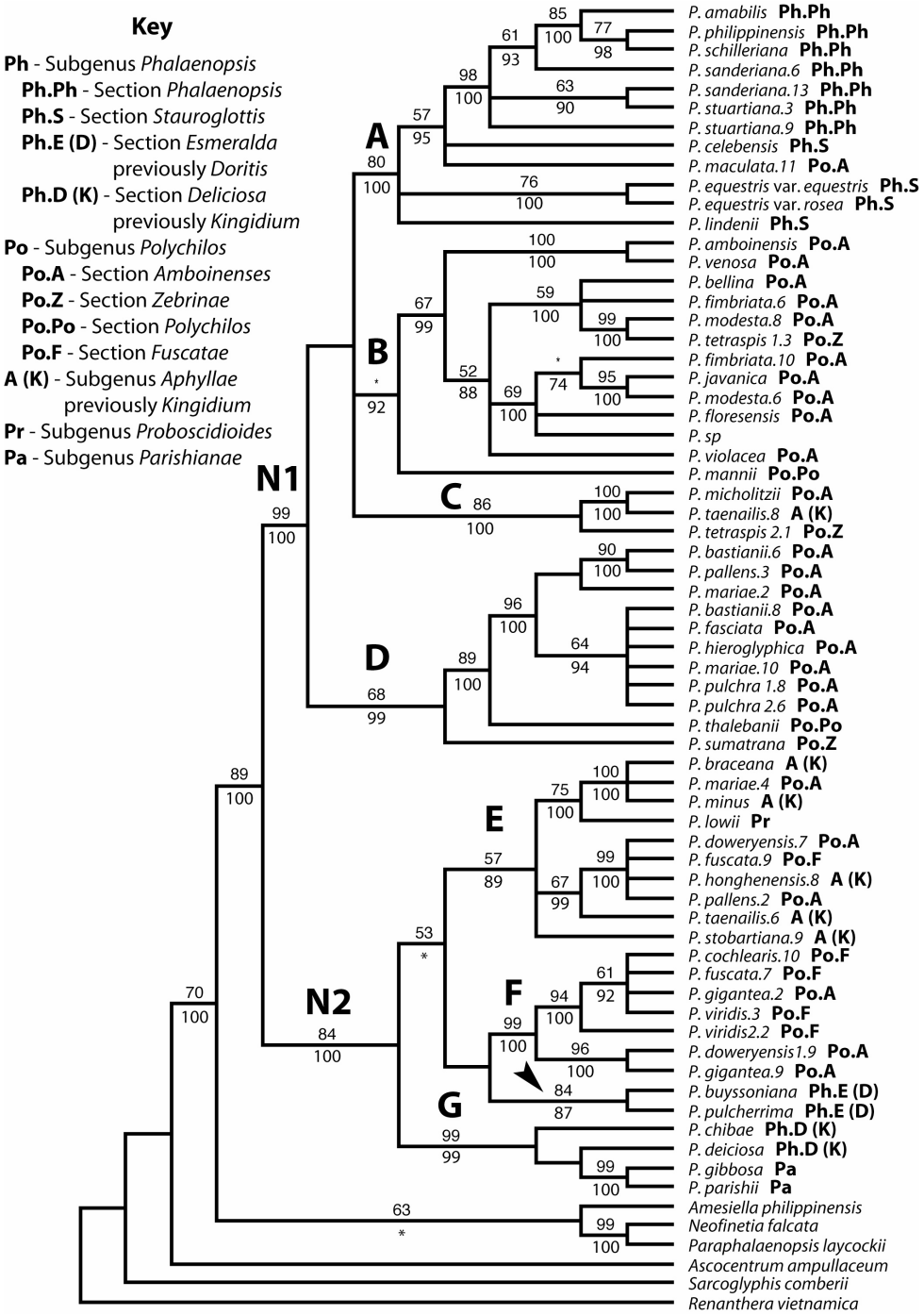


FIGURE 2.7. Majority rule consensus tree of 35,889 most parsimonious trees found for *Phalaenopsis* using nuclear data. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches. A number directly following the species name indicates species that have more than one individual included in the analysis. A number following a '.' indicates individuals with more than one sequence included in analyses.

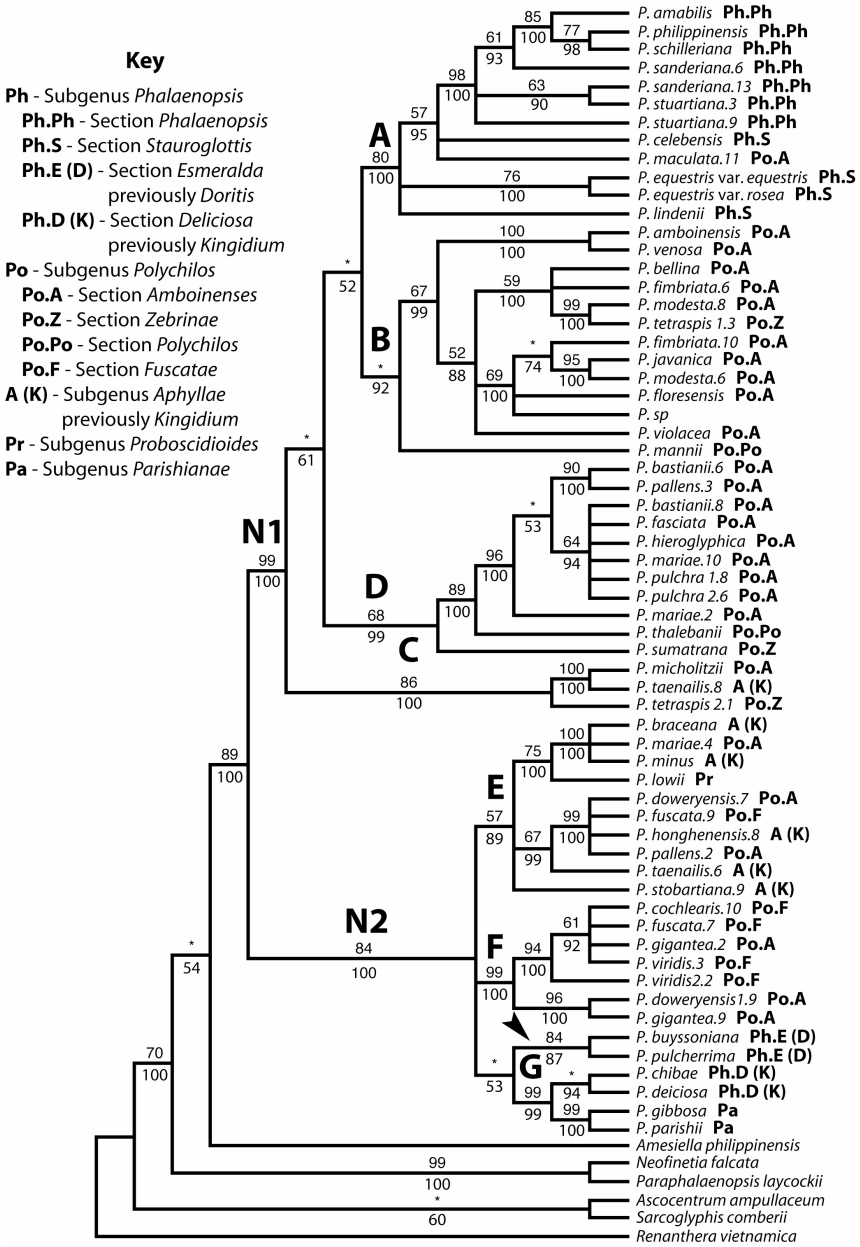


FIGURE 2.8. Majority rule consensus of trees found in Bayesian analysis for *Phalaenopsis* using nuclear data. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches. A number directly following the species name indicates species that have more than one individual included in the analysis. A number following a '.' indicates individuals with more than one sequence included in analyses.

Clade N1

Clade N1 contained four clades (N1.A, N1.B, N1.C, and N1.D) with species from subgenera *Phalaenopsis* and *Polychilos* (FIGURES 2.7 and 2.8). Subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis* formed a highly supported monophyletic group (N1.A) with a bootstrap value of 80 and a posterior probability of 100% and correlated with clade C1 from the chloroplast trees. Section *Phalaenopsis* formed a highly supported monophyletic group (bootstrap 98 and posterior probability 100%), however section *Stauroglottis* occurred as a paraphyletic grade basal to section *Phalaenopsis. Phalaenopsis maculata* was inexplicably included within this clade. Subgenus *Phalaenopsis,* including sections *Esmeralda* and *Deliciosae,* was not monophyletic as was confirmed by parametric bootstrap analyses ($p < 0.001$)

Most of the species in subgenus *Polychilos* were in clades N1 (N1.B, N1.C, and N1.D), including species from sections *Polychilos, Amboinensis,* and *Zebrinae*, while the remaining species in section *Amboinenses* and section *Fuscatae* were in clade N2. The position of the groups N1.B-D with respect to each other differed between the parsimony and Bayesian analyses. These three clades, together, contained the species found in a monophyletic clade C2.B on the chloroplast tree. Parametric bootstrap analyses confirmed that subgenus *Polychilos* did not form a monophyletic group based on the nuclear data as it did in the chloroplast data ($p < 0.001$). If the species in clade N2.F were removed from analyses, the monophyly of the taxa in clades N1.B, N1.C, and N1.D were still significantly non-monophyletic at $\alpha=0.05$ ($p=0.028$), although the best tree under this constraint was only one step longer than the best tree.

Clade N2

Clade N2 contained three clades (N2.E, N2.F, and N2.G) with species from subgenera *Aphyllae, Polychilos, Phalaenopsis,* and *Parishianae* (FIGURES 2.7 and 2.8).

Subgenus *Aphyllae* (clade N2.E) did not occur as a monophyletic group because sequences from species in subgenus *Polychilos* sections *Fuscatae* and *Amboinenses* occurred throughout the clade. When data were reanalyzed omitting the *Polychilos* sequences that fell into the *Aphyllae* clade, the relationships of the species within *Aphyllae* did not concur with the chloroplast data (FIGURE 2.9). *Phalaenopsis minus* was found well-nested within subgenus *Aphyllae* according to the nuclear data, while the chloroplast data tentatively placed it in clade C4 (Figure 2.5 and 2.6). *Phalaenopsis lowii,* in subgenus *Proboscidioides,* occurred well nested within subgenus *Aphyllae* (clade N2.E) according to analyses of the nuclear data, while the chloroplast analyses placed it basal to subgenus *Aphyllae*.

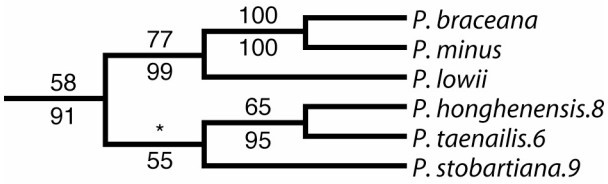


FIGURE 2.9. Majority rule consensus of trees found in Bayesian analysis for *Phalaenopsis* subgenera *Aphyllae* and *Polychilos* section *Amboinensis* using nuclear data and excluding those sequences of Subgenus *Polychilos* section *Fuscatae* that occurred in this clade. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches.

Clade N2.F formed a highly supported monophyletic group with a bootstrap value of 99 and posterior probability of 100%. This clade is represented by subgenus *Polychilos* section *Fuscatae* and several species from section *Amboinenses* and contained the same group of species (except *Phalaenopsis maculata*) as was found in the basal-most group in clade C2 (C2.A) from the chloroplast analyses (FIGURES 2.5 and 2.6). Some of the species in this clade had several sequence types that occurred both within this clade and in clade N2.E.

Clade N2.G included species from subgenus *Phalaenopsis* section *Deliciosae* and subgenus *Parishianae* and was highly supported by bootstrap and posterior probability (99 and 99%, respectively). Species in subgenus *Parishianae* were very difficult to sequence. Sequence length of *Phalaenopsis parishii* and *P. gibbosa* were approximately 1050 base pairs, more than 250 base pairs longer than other *Phalaenopsis* species. Sequences were easily alignable in the coding region but very difficult in the non-coding region. Sequence from *P. chibae* was much shorter than for other *Phalaenopsis*, only 400 base pairs. Most of the noncoding region was missing, with the last sixty base pairs of *P. chibae's* sequence aligning with the end of the other *Phalaenopsis* sequences. *Phalaenopsis deliciosa* had a normal length but was also difficult to align.

Subgenus *Phalaenopsis* section *Esmeralda* (indicated by an arrow ➢) occurred in different clades in the parsimony and Bayesian analyses. The parsimony analysis (FIGURE 2.7) placed *Esmeralda* sister to clade N2.F, which included species in subgenus *Polychilos* sections *Fuscatae* and *Amboinenses*. This placement was not supported by bootstrap or Bayesian posterior probabilities. The Bayesian analysis (FIGURE 2.8) placed section *Esmeralda* sister to the clade N2.G, which contained species in subgenus *Phalaenopsis* section *Deliciosae* and subgenus *Parishianae* with a non-significant posterior probability of 53%. This result, placing *Esmeralda* with *Deliciosae* and *Parishianae*, was corroborated by the chloroplast analyses, and the morphological characters (discussed bellow) seem to ally it with the other species in clade C4.

Character Evolution

Characters were traced onto the chloroplast tree from the Bayesian analyses (FIGURE 2.10) to infer the ancestral state and to discern any phylogenetic trends. The

data matrix and character states can be found in TABLES 2.3 and 2.4. Genome size data were available for nineteen *Phalaenopsis* taxa and one outgroup taxon (Jones et al. 1998; Lin et al. 2001). No statistically significant evolutionary trends in genome size could be discerned. Species in clade C1 (subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis*) have the smallest genomes, between 2 and 4 pg/ diploid nuclear DNA content (2C), which seemed to be consistent throughout this clade. Many species in clade C2 (subgenus *Polychilos*) had intermediate genome sizes between 5 and 7 pg/2c, however there were three species (two independent instances) with genomes larger that 9 pg/2C, two of which had genomes larger than 13 pg/2C. *Phalaenopsis parishii* and *P. pulcherrima*, which both occurred in clade C4, also had genomes greater than 13 pg/2C. A pairwise comparison using continuous genome size data was performed in Mesquite 1.06 (Maddison 2005; Maddison and Maddison 2005) on a phylogeny with only those species for which data were available (FIGURE 2.11). Mesquite found no significant correlation between genome size and either and island or a Philippine distribution.

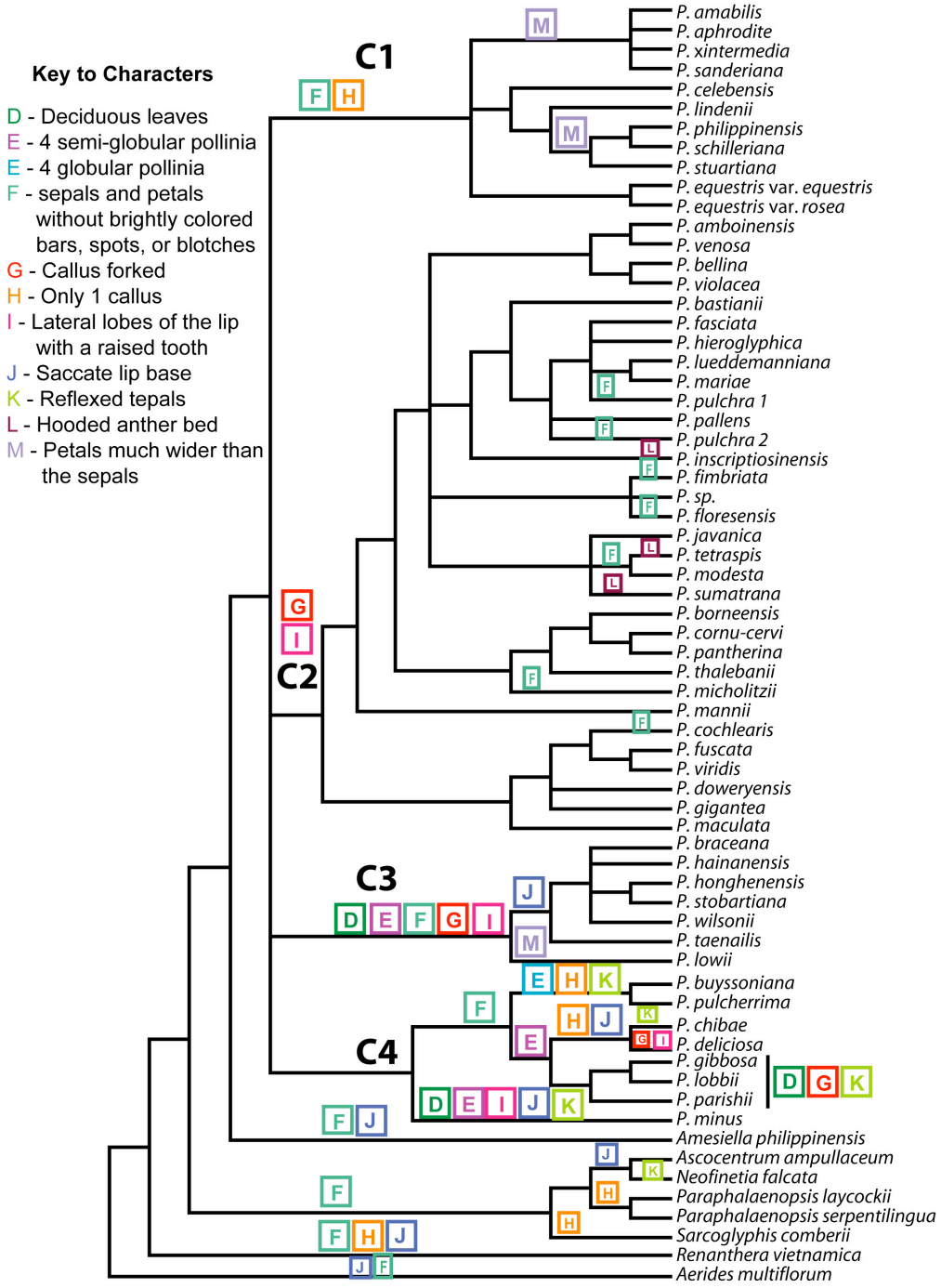


FIGURE 2.10. Majority rule consensus of trees found in Bayesian analysis for *Phalaenopsis* using combined chloroplast data. Certain character states (corresponding with TABLE 2.3 and 2.4) are mapped onto the tree.

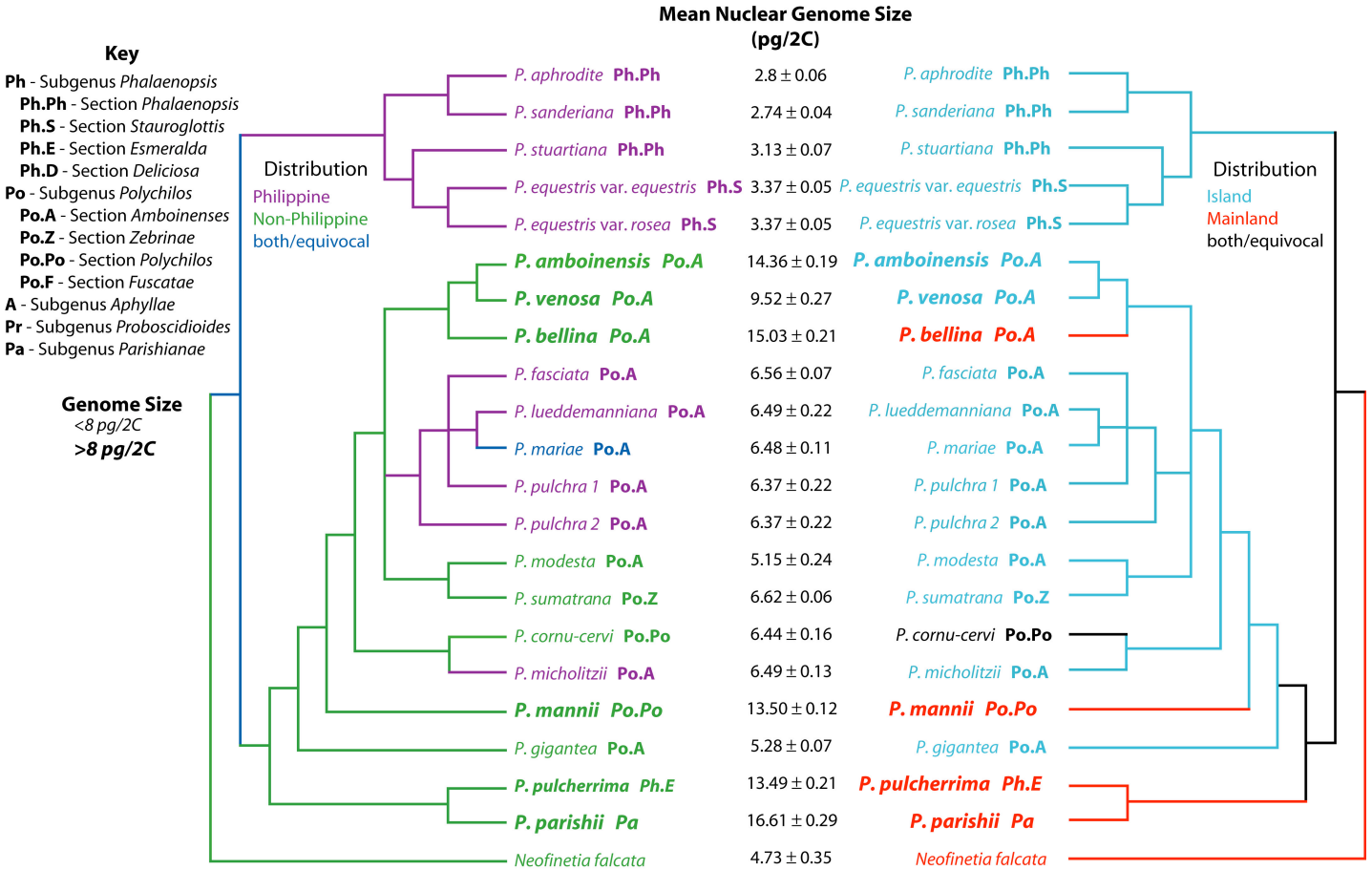


FIGURE 2.11. Majority rule consensus tree of 80,004 most parsimonious trees found for *Phalaenopsis* using combined chloroplast data that was pruned to include only those species for which genome size information is available. Genome size is indicated as a binary character with those taxa smaller than 7 pg/2C in regular font and those species with genomes larger than 9 pg/2C in bold font. In addition, geographic distribution is indicated on the left-hand tree as Philippine taxa in purple and non-Philippine taxa in green and on the right-hand tree as island taxa in blue and mainland taxa in red.

DISCUSSION

Classification

Christenson's (2001) treatment of *Phalaenopsis* as a broadly defined genus that includes the genera *Doritis* and *Kingidium* was supported by the chloroplast and nuclear phylogenies presented here. The morphological differences in these groups that distinguish them from the 'traditional' *Phalaenopsis* are possibly adaptations to their environments. *Doritis* occurs in a terrestrial habit and exhibits an upright, elongated growth habit and erect inflorescences. *Kingidium* species are montane. They are small plants with deciduous leaves that are adaptations to colder, drier climates associated with high altitudes. To treat these as genera separate from *Phalaenopsis* would necessitate breaking *Phalaenopsis* into a minimum of five genera, and would constitute a taxonomic nightmare.

Clades C1 and N1.A

Subgenus *Phalaenopsis* sections *Phalaenopsis* and *Stauroglottis* together form a well-supported monophyletic group (clades C1 and N1.A, FIGURES 2.5-2.8) for all markers and in all reconstructions. This clade should form a subgenus *Phalaenopsis* with no sectional delimitations, since the species in section *Stauroglottis* (*Phalaenopsis equestris, P. lindenii,* and *P. celebensis*) do not form a monophyletic group, according to either the chloroplast or nuclear-based phylogenies. The chloroplast analyses do not indicate that section *Phalaenopsis* is a monophyletic group, although it has high support in the nuclear analyses. The chloroplast analyses put those species with silver mottling on the leaves (*P. philippinensis, P. schilleriana, P. stuartiana, P. celebensis,* and *P. lindenii*) into a monophyletic group. We believe this may be due to lineage sorting of the

chloroplast genome, since morphologically the species in section *Phalaenopsis* are extremely similar. Parametric bootstrap analyses indicate that subgenus *Phalaenopsis* is not a monophyletic group because sections *Esmeralda* (C4.H) and *Deliciosae* (C4.G) do not belong.

This group has many characters that unite it, although the characters do not appear to be synapomorphies. All species for which data are available have a nuclear genome content less than 4.0 pg/2C. No other *Phalaenopsis* has been shown to have a genome this small, however there are many taxa that have not yet been examined. All species in this clade, with the exception of *Phalaenopsis celebensis*, are endemic to the Philippine Islands. Their petals and sepals are generally unmarked, they possess only a single callus that is not forked, and they have no raised teeth on the lateral lobes of the lip. With the exception of chromosome size, all of these diagnostic characters appear to be pleisiomorphic because they are found in other species in the genus as well as in the outgroup taxa.

Clades C2 and N1.B-D, N2.F

Subgenus *Polychilos* formed a monophyletic group according to analyses of the chloroplast data but not the nuclear data. Analyses of the nuclear data excluded section *Fuscatae* and three species in section *Amboinenses* (clade N2.F and C2.A, FIGURES 2.5-2.8) from the other species in the subgenus (clades N1.B, N1.C, and N1.D). Subgenus *Polychilos* is morphologically similar and highly recognizable. It possesses evergreen leaves, two cleft pollinia, and flowers that are usually marked with bright colors and bold patterns. It possesses two or three calli, at least one of which is forked, and it has a raised tooth on the lateral lobes of the lip. We recognized Christenson's (2001) subgenus

Polychilos without changes to the subgeneric circumscription, however we changed the sectional delimitations.

The species in subgenus *Polychilos* have been problematic with regards to delimiting species boundaries and determining subgroups. The sectional relationships within the subgenus were unclear on both the chloroplast- and nuclear-based phylogenies due to lack of resolution and low support for some clades. Neither sections *Polychilos, Amboinenses* nor *Zebrinae* occurred as monophyletic in either analysis. The chloroplast phylogeny delineated two main clades, C2.A and C2.B. The nuclear phylogeny recovered four clades, N1.B, N1.C, N1.D, and N2.F.

Clades C2.A and N2.F were highly supported by both analyses and represented subgenus *Polychilos* section *Fuscatae* and three species from section *Amboinenses*. We propose to recircumscribe section *Fuscatae* and expand it to include *Phalaenopsis doweryënsis, P. gigantea,* and *P. maculata*. *Phalaenopsis maculata* was the basal taxon in clade C2.A and was not included in N2.F. The nuclear phylogeny anomalously placed *P. maculata* nested within clade N1.A. Morphology and the chloroplast phylogeny strongly contradicted this result, however, so we choose to use the results of the chloroplast phylogeny when delimiting this group. We would like to acquire additional taxa of *P. maculata* to add to these analyses to ensure these data are accurate and not a result of an inconspicuous hybrid. It is surprising that *P. maculata* did not occur sister to *P. mariae*, to whom it is morphologically very similar. Species in clades C2.A and N2.F share similar coloration of a creamy-white or yellow background with reddish-brown markings of various sorts. More work will need to be done to determine if there are easily discernable characters that can separate these taxa from the other species in subgenus *Polychilos*.

Clade C2.B did not have a corresponding monophyletic group on the nuclear phylogeny; it was represented by clades N1.B-D. Although parametric bootstrap analyses rejected the monophyly of clades N1.B-D, constraining these clades to be monophyletic resulted in a tree that was only one step longer than our best tree. We choose to expand the definition of section *Polychilos* to recognize all of the species in clades C2.B and N1.B-D in the section. Within section *Polychilos,* one clade represented a highly supported monophyletic group in both phylogenies, the species that we refer to informally as the *lueddemanniana* complex. Historically, species in this group have been treated as a highly variable species, *Phalaenopsis lueddemanniana,* with varieties *hieroglyphica, fasciata, pallens, pulchra,* and others. There has been extensive hybridization within this group by horticulturists, and nomenclatural confusion has obscured the hybrid nature of many plants that are sold as species. We have obtained as many individuals as possible for species in this group to attempt to compensate for this confusion.

Clades C3 and N2.E

Clade C3 (FIGURES 2.5 and 2.6) is a highly supported monophyletic group and represents subgenus *Aphyllae* (excluding *Phalaenopsis minus*) and its sister subgenus *Proboscidioides.* Clade N2.E (FIGURES 2.7 and 2.8) is a weakly supported group and represents a paraphyletic *Aphyllae*. Certain clones from subgenus *Fuscatae* are nested within N2.E, as is *P. lowii* (subgenus *Proboscidioides*). The chloroplast and nuclear analyses place *P. minus* and *P. lowii* in different positions. *Phalaenopsis lowii* has traditionally been placed in its own group within *Phalaenopsis* on account of its long, beak-like rostellum. Other characters, however, ally it with the *Aphyllae*, such as its deciduous habit, its four semi-globular pollinia, and its geographical distribution. The

nuclear analyses nests *P. lowii* within *Aphyllae*, however the low support values and issues of paralogy make this reconstruction suspect. The chloroplast data place *P. lowii* basal to *Aphyllae.* We recognize a more broadly defined subgenus *Aphyllae* that includes *P. lowii*, placing subgenus *Proboscidioides* in its synonymy.

Deciding how to treat *Phalaenopsis minus* was difficult. The combined chloroplast analyses place *P. minus* as the basal taxon in a weakly supported clade C4. Separate analyses of the chloroplast *atp*HF, *mat*K, and the chloroplast noncoding region (including the *pet*D group II intron) (FIGURES A.2.1, A.2.4, and A.2.8, respectively) do not place *P. minus* sister to the other species in C4, and some of the best trees in the combine chloroplast analyses place *P. minus* with subgenus *Aphyllae* (clade C3). The nuclear analyses place *P. minus* nested within *Aphyllae* (N2.G), as is consistent with the current classification. Christenson notes "*Phalaenopsis minus* is an oddball of sorts and does not appear to be closely related to the other species in the subgenus [*Aphyllae*]." We propose that *P. minus* be segregated into a monotypic section, *Conspicuum* Gruss & Rollke stat. nov. that we tentatively place in subgenus *Parishianae*, based on the chloroplast phylogeny and a few morphological characters, such as reflexed sepals and petals. Further studies may show that *P. minus* should be moved back into *Aphyllae*.

All of the species in clades C3 are mainland taxa. They are deciduous with four, semiglobular pollinia. Their flowers are pink to greenish brown without conspicuous markings. They have two calli, one of which is forked, and a flap-like tooth on each of the lateral lobes of the lip. Many of these characters are also shared with species in clade C4, but C3 is distinguishable in that it does not have reflexed petals and sepals.

Clades C4 and N2.G

Subgenus *Parishianae* should be expanded to encompass all the species in clades C4 and N2.G (FIGURES 2.5-2.8), which would include subgenus *Phalaenopsis* section *Deliciosae*, section *Esmeralda,* and according to the chloroplast data *Phalaenopsis minus*. All members of this group possess four pollinia, either semiglobular or globular. Their petals and sepals are generally unmarked (except for *P. minus*, which has purple spots and bars) and strongly reflexed (except *P. deliciosa*). All of these species also possess a lip midlobe that is broadly expanded (with the exception of *P. minus*). Species in *Esmeralda* have a deceptively expanded lip midlobe. What appear to be the lateral lobes are in fact lobules of the midlobe. The true lateral lobes have been reduced to 'antennae.' Reduction of the lateral lobes also occurs in most of the other species in clade C4, with the exception of *P. deliciosa* and *P. minus.*

Sections *Esmeralda* and *Deliciosae* do not belong in subgenus *Phalaenopsis*, as was indicated by the results of the parametric bootstrap analysis. All analyses highly support the placement of section *Deliciosae* sister to subgenus *Parishianae.* We propose that the species in section *Deliciosae* be transferred to subgenus *Parishianae* section *Parishianae*. We also place the species in section *Esmeralda* into subgenus *Parishianae* section *Esmeralda*. The chloroplast-based analyses and the Bayesian analysis of the nuclear data place *Esmeralda* in clade C4, sister to *Deliciosae* + *Parishianae*, but the parsimony analysis of the nuclear data indicate that *Esmeralda* is basal to subgenus *Fuscatae.* There is little to no support for either of these hypotheses, but *Esmeralda* is morphologically more similar to *Deliciosae* and *Parishianae*. They all possess four pollinia, a midlobe of the lip that is expanded, and reflexed sepal and petals. *Phalaenopsis pulcherrima* and *P. parishii* both have large genomes, greater than 13 pg/2C, while *P. gigantea* has a much smaller genome, less than 6 pg/2C. These are the

only species in these clades for which genome size data is available. Finally, the species in clade N2.G, as well as *Esmeralda*, are mainland taxa, while the species in clade N2.F are mostly island taxa. Section *Esmeralda* forms a highly supported monophyletic group in all analyses. It is distinguishable by its many synapomorphies, including four globular pollinia, its unusual lip and column morphology, and its erect inflorescences and vegetative growth habit.

Evolution of Genome Size

Although there appeared to be a trend between genome size and geographic distribution (FIGURE 2.10), there are no Philippine species with large chromosomes, a pairwise comparison found no significant correlation between genome size and geographic distribution, coded as either Philippine/non-Philippine or island/mainland. Perhaps when genome size data are available for more *Phalaenopsis* species a significant trend will be discernible. Genome size does appear to be a useful character for recognizing phylogenetic clades within *Phalaenopsis*, especially subgenus *Phalaenopsis*, with species that possess particularly small genomes and subgenus *Parishianae*, with species that possess particularly large genomes.

Summary

The genus *Phalaenopsis* should include species from *Doritis* and *Kingidium*, as proposed by Christenson's monograph (2001), however some of the subgeneric groups in Christenson's classification needed to be recircumscribed to reflect a natural classification. We have proposed a new classification for *Phalaenopsis* (TABLE 2.5) based on evidence from our molecular phylogenies. We have attempted to retain the

current names as much as possible in the context of a natural classification. Four subgenera and six sections were recognized in our classification. Subgenus *Aphyllae* is similar to Christenson's subgenus, with the exclusion of *Phalaenopsis minus*. Subgenus *Parishianae* has been expanded and includes three sections, *Conspicuum, Esmeralda,* and *Parishianae*. Subgenus *Phalaenopsis* contains no sections and includes the species that were in Christenson's sections *Phalaenopsis* and *Stauroglottis*. Finally, subgenus *Polychilos* contains the same species as were defined by Christenson, but we only recognize two sections, *Fuscatae* and *Polychilos.*

TABLE 2.5. Padolina et al.'s classification of the genus *Phalaenopsis*. Species in bold face are included in this project's analyses. Species marked with an asterisk (*) have not yet been positively identified. Characters marked with † are unique and pertinent only to that group.

Subgenus	Section	Species	Description
Aphyllae 7 species		• ***P. braceana*** • ***P. hainanensis*** • ***P. honghenensis*** • ***P. lowii**** • ***P. stobartiana**** • ***P. taenialis*** • ***P. wilsonii***	• Flowers green to brown to pink. • Biseriate callus. • Flap-like tooth on lateral lobes of the lip. • Small, deciduous, epiphytic plants. • Sepals and petals similar to subsimilar, except *P. lowii*. • 4 semiglobular pollinia. • †Possess a prominent to obscure spur.
Parishianae 11 species	3 sections		
	Conspicuum 1 species (Tentative)	• ***P. minus***	• Flowers white with pink to purple spots and bars. • Biseriate callus. • Lateral lobes of the lip with a falcate triangular tooth. • Small deciduous, epiphytic plants. • Sepals and petals subsimilar and strongly reflexed. • 4 semiglobular pollinia. • †Possesses a spotted column.
	Esmeralda 3 species	• ***P. buyssoniana**** • ***P. pulcherrima*** • *P. regnieriana*	• Essentially unmarked magenta, pink, to white flowers. • Uniseriate callus • Linear 'appendages' toward the base of the lip are the true lateral lobes of the lip. The midlobe is expanded into 3 lobules, the lateral lobules mimic the lateral lobes. No tooth-like appendages on either the lateral lobes or lobules. • Terrestrial to lithophytic with moss. Evergreen or deciduous. • Sepals and petals similar to subsimilar in size • 4 globular pollinia. • †Long column foot. • †Long rostellum.
	Parishianae 4 species	• *P. appendiculata* • ***P. chibae*** • ***P. deliciosa*** • ***P. gibbosa*** • ***P. lobbii*** • *P. mysorensis* • ***P. parishii***	• Flowers white with pink or yellow/brown lip and column (*P. appendiculata* and *P. chibae*, with markings on tepals, as well). • Biseriate or uniseriate callus. • Lateral lobes of lip erect, subparallel, and diverging to form a U-shaped compound structure. No tooth-like appendages. (*P. delisiosa* with rounded laterlobe with a tooth-like flap. • Small, deciduous or evergreen, epiphytic plants. • Sepals and petals similar to subsimilar and strongly reflexed. • 4 semiglobular pollinia. • †Prominent swellings at the base of the column. • †Mobile lip midlobe

TABLE 2.5 continued

Subgenus		Species	Description
Phalaenopsis 9 species 1 natural hybrid		• ***P. amabilis*** • ***P. aphrodite*** • ***P. celebensis**** • ***P. equestris*** • ***P. xintermedia*** • ***P. lindenii*** • ***P. philippinensis*** • ***P. sanderiana*** • ***P. schilleriana*** • ***P. stuartiana***	• Essentially unmarked white or pink flowers. • Uniseriate, prominent, bilobed, glossy callus. • Smooth lateral lobes without a tooth-like ridge. • Large or small, evergreen, epiphytes. • Sepals and petals similar to subsimilar or petals much broader than sepals. • 2 cleft pollinia. • †Some species with a pair of appendages (cirrhi) at the apex of the midlobe of the lip
Polychilos 39 species	2 sections		
	Fuscatae 7 species	• ***P. cochlearis*** • ***P. doweryënsis*** • ***P. fuscata*** • ***P. gigantea**** • *P. kunstleri* • ***P. maculata**** • ***P. viridis****	• Pale yellow flowers with red-brown markings. • Biseriate callus, one of which is bifid. • Lateral lobes of the lip with a raised tooth. • Medium to very large evergreen epiphytes. • Petals and sepals similar to subsimilar and may be revolute. • 2 cleft pollinia • †Concave lip with a longitudinal keel.
	Polychilos 32 species	• ***P. amboinensis*** • ***P. bastianii*** • ***P. bellina*** • ***P. borneënsis*** • *P. corningiana* • ***P. cornu-cervi*** • ***P. fasciata*** • ***P. fimbriata**** • ***P. floresensis**** • ***P. hieroglyphica*** • ***P. inscriptiosinensis**** • ***P. javanica*** • ***P. lueddemanniana*** • *P. luteola* • ***P. mannii*** • ***P. mariae**** • ***P. micholitzii*** • ***P. modesta**** • ***P. pallens**** • ***P. pantherina*** • ***P. pulchra*** • *P. reichenbachiana* • *P. robinsonii* • *P. speciosa* • ***P. sumatrana**** • ***P. tetraspis*** • ***P. thalebanii* (=*P. cornu-cervi*)** • ***P. venosa*** • ***P. violacea***	• Flowers brightly colored and patterned, fragrant, fleshy and long lasting. • Biseriate or triseriate callus, one of which is bifid. • Lateral lobes of the lip with a raised tooth. • Large or small, evergreen, epiphytes. • Petals and sepals equal/subequal and similar/subsimilar or petals slightly narrower than petals. • 2 cleft pollinia.

LITERATURE CITED

BENNET, M.D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London 181: 109-135.

BENSON, D.A., I. KARSCH-MIZRACHI, D.J. LIPMAN, J. OSTELL, and D.L. WHEELER. 2004. GenBank: update. Nucleic Acids Research 32: D23-D26.

CAMERON, K.M., M.W. CHASE, W.M. WHITTEN, P.J. KORES, D.C. JARRELL, V.A. ALBERT, T. YUKAWA, H.G. HILLS, and D.H. GOLDMAN. 1999. A phylogenetic analysis of the Orchidaceae: Evidence from rbcL nucleotide sequences. American Journal of Botany 86: 208-224.

CHRISTENSON, E.A. 2001. *Phalaenopsis*: a Monograph. Timber Press, Portland, Oregon.

DEMESURE, B., N. SODZI, and R.J. PETIT. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4: 129-131.

DOWTON, M. and A.D. AUSTIN. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - The behavior of the incongruence length difference test in mixed-model analyses. Systematic Biology 51: 19-31.

DOYLE, J.J. and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19: 11-15.

EVANS, G. and H. REES. 1971. Mitotic cycles in dicotyledons and monocotyledons. Nature 233.

FARRIS, J.S., M. KALLERSJO, A.G. KLUGE, and C. BULT. 1994. Testing Significance Of Incongruence. Cladistics-The International Journal Of The Willi Hennig Society 10: 315-319.

FU, Y.M., W.H. CHEN, W.T. TSAI, Y.S. LIN, M.S. CHYOU, and Y.H. CHEN. 1997. Phylogenetic studies of taxonomy and evolution among wild species of Phalaenopsis by random amplified polymorphic DNA markers. [Chinese]. Report of the Taiwan Sugar Research Institute 0: 27-42.

GRIESBACH, R.J. 2002. Development of Phalaenopsis Orchids for the Mass Market. ASHS Press, Alexandria, VA.

GRUSS, O. and L. ROLLKE. 1993a. Gattung *Kingidium*. Die Orchidee 44: 33-36.

_____. 1993b. Gattung *Kingidium*, Teil 2. Die Orchidee 44: 119-122.

_____. 1995a. *Kingidium wightii* (Rchb.f.) Gruss & Rollke comb. nov., Teil 10. Die Orchidee 46: 23-28.

_____. 1995b. *Kingidium philippinense* (Ames) Gruss & Rollke comb. nov., Teil 11. Die Orchidee 46: 115-118.

_____. 1996. *Kingidium wilsonii* (Rolfe) Gruss & Rollke comb. nov. Die Orchidee 47: 149-155.

_____. 1997. *Kingidium chibae* (T. Yukawa) Gruss & Rollke comb. nov. Die Orchidee 48: 261-263.

HAWKES, A.D. 1966. *Grafia*, a new genus of orchids. Phytologia 13: 305-306.

HUELSENBECK, J.P. and F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

JONES, W.E., A.R. KUEHNLE, and K. ARUMUGANATHAN. 1998. Nuclear DNA content of 26 orchid (Orchidaceae) genera with emphasis on *Dendrobium*. Annals Of Botany 82: 189-194.

KAO, Y.Y., S.B. CHANG, T.Y. LIN, C.H. HSIEH, Y.H. CHEN, W.H. CHEN, and C.C. CHEN. 2001. Differential accumulation of heterochromatin as a cause for karyotype variation in *Phalaenopsis* orchids. Annals Of Botany 87: 387-395.

LIN, S., H.C. LEE, W.H. CHEN, C.C. CHEN, Y.Y. KAO, Y.M. FU, Y.H. CHEN, and T.Y. LIN. 2001. Nuclear DNA contents of *Phalaenopsis sp.* and *Doritis pulcherrima*. Journal of the American Society for Horticultural Science 126: 195-199.

LOHNE, C. and T. BORSCH. 2005. Molecular evolution and phylogenetic utility of the *pet*D group II intron: A case study in basal angiosperms. Molecular Biology & Evolution 22: 317-332.

MADDISON, W.P. 2005. Pairwise comparisons package for mesquite. http://mesquiteproject.org.

MADDISON, W.P. and D.R. MADDISON. 2000. MacClade 4.0. Sinauer, Sunderland, Massachusetts.

_____. 2005. Mesquite: A modular system for evolutionary analysis. http://mesquiteproject.org.

MAIER, R.M., K. NECKERMANN, G.L. IGLOI, and H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome - gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. Journal of Molecular Biology 251: 614-628.

MULLIS, K.B. and F.A. FALOONA. 1987. Specific synthesis of DNA in vitro via polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335-350.

POSADA, D. and K.A. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

RAMBAUT, A. and N.C. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Computer Applications in the Biosciences 13: 235-238.

ROLFE, R.A. 1917. *Doritis pulcherrima*. The Orchid Review 25: 195-197.

SAGAWA, Y. 1962. Cytological studies of the genus *Phalaenopsis*. American Orchid Society Bulletin 31: 390-398.

SEIDENFADEN, G. 1988. Orchid Genera in Thailand XIV. Fifty-nine vandoid genera. Opera Botanica 95: 31-34; 182-189; 236-241.

SHIM, P.S. 1982. A new generic classification in the *Phalaenopsis*-complex (Orchidaceae). Malayan Nature Journal 36: 1-28.

SHINDO, N. and H. KAMEMOTO. 1963. Karyotype analysis of some species of Phalaenopsis. Cytologia 28: 390-398.

SOLTIS, P.S., D.E. SOLTIS, and J.J. DOYLE. 1998. Molecular Systematics of Plants II: DNA Sequencing. Chapman and Hall, New York.

SUDA, J., T. KYNCL, and V. JAROLIMOVA. 2005. Genome size variation in Macaronesian angiosperms: forty percent of the Canarian endemic flora completed. Plant Systematics And Evolution 252: 215-238.

SWEET, H.R. 1980. The Genus *Phalaenopsis*. Orchid Digest, Pomona, CA.

SWOFFORD, D. 1996. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts.

THOMPSON, J.D., T.J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D.G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.

VOGEL, J., T. BORNER, and W.R. HESS. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. Nucleic Acids Research 27: 3866-3874.

WHITTEN, W.M., N.H. WILLIAMS, and M.W. CHASE. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: Combined molecular evidence. American Journal of Botany 87: 1842-1856.

WILCOX, T.P. 2005. SG Runner: A Mac OS X shell-wrapper for Seq-Gen. http://homepage.mac.com/tpwilcox/SGRUNNER/FileSharing8.html, Long Key, FL.

WOODARD, J.W. 1951. Some chromosome numbers in *Phalaenopsis*. American Orchid Society Bulletin 20: 356-358.

CHAPTER 3: A COMPREHENSIVE SEARCH FOR PHYLOGENETIC MARKERS THAT ARE BROADLY INFORMATIVE IN ANGIOSPERMS

INTRODUCTION

The field of molecular systematics has been growing at a tremendous rate as sequencing costs decrease and high throughput genetic technology increases. In plants, many plastid markers (Shaw et al. 2005) and some nuclear markers, e.g., 18S-26S ribosomal DNA region (Baldwin 1992), have been found to be phylogenetically informative at a variety of taxonomic levels. For many groups, however, it continues to be a challenge to find informative low-copy or single copy nuclear DNA markers. The problem of an insufficient number of nuclear markers is exacerbated in plants and other groups that undergo hybrid speciation because reconstruction of hybrid speciation has been hypothesized to require a large enough number of nuclear markers to distinguish hybrid parentage from other processes such as lineage sorting and meiotic recombination (Linder and Rieseberg 2004). Organellar markers do not typically suffer from these problems since they are usually uniparentally inherited and do not typically recombine (Soltis et al. 1998). It is therefore reasonable to assume in most cases that any gene or genomic region from an organelle shares a common evolutionary history with any other gene or region from the same organelle. Nonetheless, that history may not accurately reflect the phylogeny of the species, and even when it does reflect the phylogenetic history of the species it can only identify one of the parents in hybrid species. In addition, plastid genomes tend to evolve relatively slowly compared to nuclear genomes (Soltis et al. 1998), and in many groups of recently evolved plants it is difficult to find sufficient variation in these markers to reconstruct a well-resolved phylogeny.

Nuclear genes or genomic regions may be composed of several evolutionary histories due to sexual and meiotic recombination, so care must be taken to choose nuclear genes that are orthologous and preferably unrecombined. Low-copy nuclear regions are usually more desirable for studying evolutionary relationships where there has been hybridization, but in many groups of plants, such regions that are phylogenetically informative are difficult to find (Sang 2002). Nuclear ribosomal DNA (nrDNA) has traditionally been used as a marker for phylogenetic reconstruction in plants (Baldwin 1992). It is generally easy to amplify because it occurs in many copies, and usually these copies are identical to each other via concerted evolution. In addition, there are several regions in the nrDNA that are useful at different taxonomic levels – the ribosomal genes, 18S, 5.8S, and 26S, for higher taxonomic levels, and the internal transcribed spacer (ITS) and occasionally the external transcribed spacer (ETS) at lower taxonomic levels. Unfortunately, ITS has been found to be inappropriate for phylogenetic studies in many groups of plants (Bailey et al. 2003). Paralogous copies of ITS have been recovered when PCR products were subcloned and the ITS region has been found to be unalignable in many groups. Deep hybridization events followed by lineage sorting have also been problems when using ITS for phylogenetic reconstruction. In the case of the ETS, for a large number of plant groups it is difficult to amplify or lacks good PCR primers.

To address the lack of nuclear regions available for phylogenetic studies, several approaches have been taken. Expressed sequence tag (EST) libraries have been used to screen for phylogenetically informative nuclear markers (Landais et al. 2003). ESTs are desirable because they are coding regions, and therefore may be under strong selection and contain conserved regions for PCR primers that will amplify across a wide range of taxa. However, at lower taxonomic levels it can be difficult to find sufficient levels of

sequence variation to discern phylogenetic relationships unless noncoding regions can be amplified using the EST coding regions to anchor primers. EST sequences alone do not indicate where introns will be located in an EST, so separate genomic amplifications must be attempted. In the absence of other information about the structure of the genes, such amplifications may be hampered by the presence of long introns that produce fragments too long to amplify easily. The flanking regions of microsatellites have also been used to reconstruct phylogeny (Streelman et al. 1998). However, sequences from the microsatellite-flanking region are usually short and do not include many polymorphisms (Orti et al. 1997). This is due to the fact that microsatellite PCR primers are designed to assess population-level variation and amplify only a negligible portion of the flanking regions. In addition, the primers designed to amplify microsatellites are usually not meant to work over a broad taxonomic range, making their use in groups other than the ones for which they were designed questionable.

Because of these problems with other approaches, we tried a novel approach to search for the full set of single-copy nuclear regions that may amplify broadly across flowering plants and avoid many of the problems with the approaches that have been tried to date. We conducted a whole genome comparison of the rice (*Oryza sativa* L.) and *Arabidopsis thaliana* (L.) Heynh. genomes to search for regions that could be used for priming PCR reactions and then tested the 142 most promising primer combinations in the laboratory. Our choice of these genomes was driven by which genomes are currently available and taxonomic considerations; the two species are relatively distantly related in the Angiosperms. We used the following nested criteria for our search: regions that were highly conserved between the two genomes, the subset of these regions that were found as unique pairs, and the subset of this second class of regions that were spaced appropriately in both species for PCR amplification using *Taq* DNA polymerase.

We present the results of our search here and note that discovery of a large set of universal nuclear regions for phylogenetic reconstruction in plants may not be possible.

MATERIALS AND METHODS

Finding Conserved Primer Pairs

Ample theoretical calculations and empirical experience have demonstrated that in the absence of duplication, pairs of PCR primers approximately 18 to 25 bp in length can be employed reliably to amplify unique regions in the nuclear genome (Hillis et al. 1996). It is also well established that in most cases conservation of identical or nearly identical DNA sequences across long periods of evolutionary time is an indication of stabilizing selective forces due to some essential function coded by the DNA sequence, e.g., a protein coding region or a ribosomal RNA gene (Futuyma 1997). We, therefore, reasoned that comparison of two (or more) distantly related plant nuclear genomes would yield the most complete set of potential unique PCR primer combinations that would be applicable to a broad range of angiosperms. These primer combinations should span a wide range of regions with varying evolutionary rates, and so we also reasoned that this approach would allow us to find primer combinations appropriate for reconstructing phylogenetic relationships of varying evolutionary depths.

To bias our computational search for conserved primer combinations in favor of regions that could be amplified across a broad range of angiosperms and which would amplify orthologous regions, we developed a set of search criteria, which we employed prior to testing primer combinations in the laboratory:

1. Only conserved primer combinations that occurred a single time in both of the genomes were acceptable. This criterion gave us the set of primer combinations most likely to prime single-copy regions.

2. Only regions that were perfect matches between the genomes were accepted as potential primers. This criterion provided the most stringent condition for evolutionary conservation of the primer sequences, making them likely to be conserved in other angiosperm lineages. This criterion also would allow PCRs to be run under the most stringent annealing conditions to lower the incidence of mispriming.

3. Regions of perfect matches between genomes had to be at least 18 bp in length to ensure the priming region would be long enough to design a primer that would favor amplification of a unique region.

4. The distance between potential primer pairs had to be at least 400 bp and no more than 3000 bp in both nuclear genomes (FIGURE 3.1). This criterion helped to ensure the primers would amplify a region long enough to have sufficient variation to be phylogenetically informative, but not so long that it could not be amplified using *Taq* polymerase. Because we were identifying primers using genomic information rather than spliced mRNA sequences (e.g., ESTs) we could have greater confidence that potential primer pairs would be at distances that could amplify.

5. Unique primer combinations where one or both of the primers had low sequence complexity, e.g., primers made of simple sequence repeats, were excluded to prevent the accidental development of ISSR primers and therefore priming of multiple sites in the nuclear genome via mispriming.

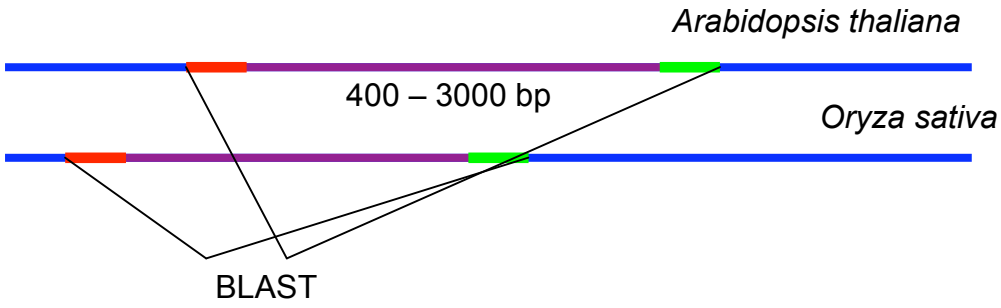


FIGURE 3.1. Specifications of potential primer pairs and the regions that were BLASTed. Matching forward primers are shown in red, matching reverse primers in green, and the intervening gene sequence from each genome is shown in purple.

6. Potential primer pairs were eliminated which were likely to form primer dimers or hairpin loops.

7. Whenever possible, potential primer pairs were verified to amplify regions that were likely to be orthologous by determining if *Arabidopsis* BLAST results of the primers plus their putative amplicon matched orthologous genes in the *Oryza* genome. This criterion could not always be met since some regions spanned uncharacterized ORFs.

At the time of this study, only the *Arabidopsis thaliana* and *Oryza sativa* nuclear genomes had been completely or nearly completely sequenced in plants, so these were the genomes used for comparisons. *Arabidopsis* is a eudicot in the Brassicaceae and rice is a monocot in the Poaceae, making the pair fairly distant in the angiosperm clade and suggesting that paired primers conserved in the two lineages would be conserved in many other angiosperm lineages as well. To identify the full set of potential primers rapidly,

we used MoBIoS, a newly developed biological sequences database system (Miranker et al. 2003), to compare the entire genomes of *A. thaliana* and *O. sativa*.

The following brief explanation of our primer discovery and filtering process is explained in depth in Xu et al. (2004). MoBIoS was used to build a database consisting of the complete set of 18 base pair sequences (18-mers) in the *Arabidopsis* genome (FIGURE 3.2). Next, we queried the database with the rice genome (both the forward sequence and its reverse complement) to identify putative primer pairs using criteria 1-4 above. This initial set of potential primer pairs was then filtered in MoBIoS using criteria 5 and 6. At this point, over 13,000 putative primer pairs were identified. In order to reduce this to a more practical set of primers to work with in the lab, we rank-ordered the primer combinations using the following criteria.

1. Primer pairs needed to be at least 1000 bp apart in both genomes to bias our selection in favor of regions that would consistently be long enough to be phylogenetically informative.

2. Primer combinations separated by at least 1000 bp were sorted according to their complexity as estimated by the compression ratio generated by the LZW compression algorithm (Welch 1984):

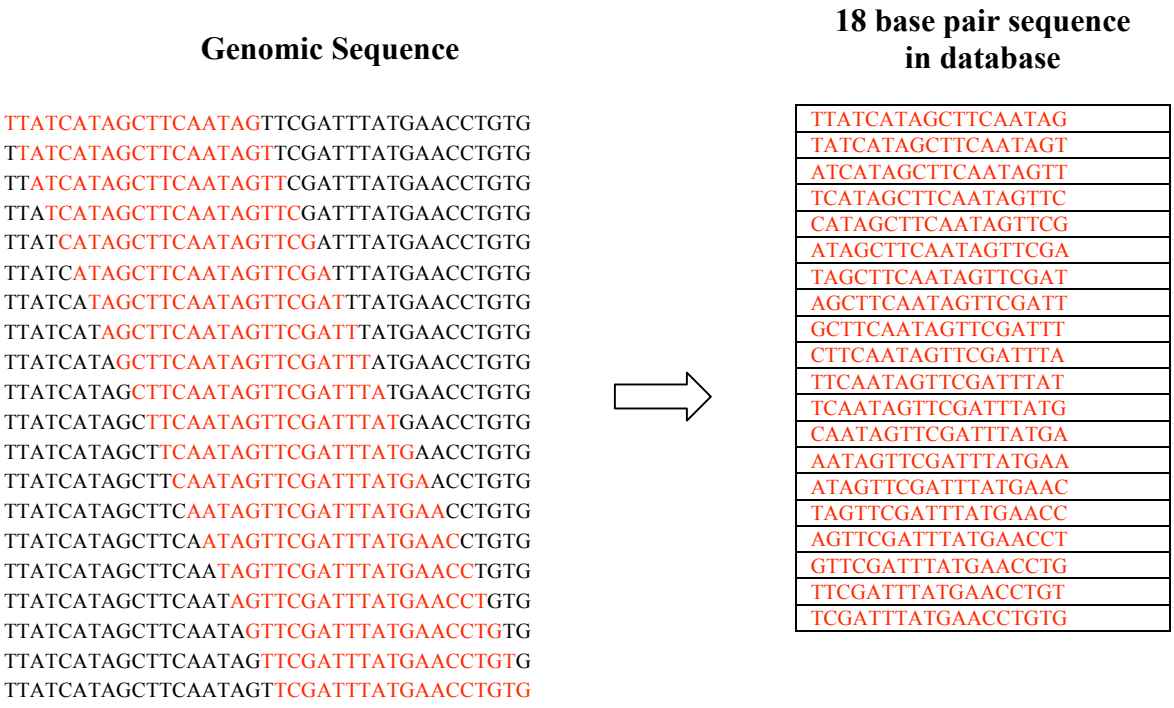


FIGURE 3.2. Decomposing complete genomes into a database consisting of all overlapping18 base pair sequences.

After sorting the resulting set of primer pairs based on complexity, we performed a BLAST search in GenBank (Benson et al. 2004) on each of the top 1000 primer pairs and their intervening sequences (FIGURE 3.1) to assess whether these regions were conserved among other plant genomes. Primer pair candidates were further sorted on (1) the number of taxa found by the BLAST results that matched the primers, (2) the number of taxa whose interprimer sequence had high similarity to the *Arabidopsis* sequence and (3) whether there was any information on the region being BLASTed. Preference was

given to primer pairs that were shown to bracket the same gene or region in one or more GenBank entries. However, since many of the genes in *A. thaliana* and *O. sativa* have not been sequenced in any other plant species, mere absence of a primer pair and/or the intervening sequence was not considered a firm indication that the primer pair would fail to amplify orthologous regions in other angiosperms.

Screening Putative Primers

One hundred forty-two primer pairs were screened in the lab using eighteen taxa for template DNA (TABLE 3.1), including six *Phalaenopsis* (Orchidaceae) species, five *Helianthus* and one *Phoebanthus* (Asteraceae) species, and six species distributed throughout the angiosperms, including *Aquilegia chysantha* (Ranunculaceae), *Geranium maderense* (Geraniaceae), *Peperomia obtusifolii* (Piperaceae), *Reinwardtia indica* (Linaceae), *Tiquilia hispidissima* (Boraginaceae), and *Arabidopsis thaliana* (Brassicaceae) as a positive control. Species chosen from *Helianthus* and *Phalaenopsis* spanned the phylogenetic range of these genera and were diploids considered not to be of hybrid origin. DNA extractions were carried out using the CTAB protocol of Doyle and Doyle (1987) or the Qiagen DNeasy Plant DNA Extraction Kit. Extracts were cleaned with the QIAGEN QIAEX II Suspension kit if necessary. Standard PCR protocols were used (Mullis and Faloona 1987), and 5% dimethyl sulfoxide (DMSO) was added to some reactions. Amplifications were visualized on 1.5% agarose gels with ethidium bromide and purified with QIAGEN QIAquick PCR purification kits or Sephadex columns. Primer pairs were considered promising if they amplified a single band in the majority of the species tested.

Using each of the six species in *Helianthus* and *Phalaenopsis*, we cloned PCR products of selected primers using Invitrogen's TOPO TA cloning kit and picked ten colonies from each individual to be amplified and sequenced. Cycle sequencing reactions were performed using BigDye Terminator 3.0 and were visualized on an MJ BaseStation. Sequences were assembled and edited in Sequencher 4.2. Initial alignments were carried out using ClustalX (Thompson et al. 1997), and then modified by eye in MacClade 4.0 (Maddison and Maddison 2000). Trees for each marker and each genus, including all clones, were constructed using parsimony in PAUP* 4.0 (Swofford 1996b), using a heuristic search with ten random-addition-replicates and TBR branch swapping. Nonparametric bootstrap analyses were performed for each reconstruction with 1000 replicates and 1000 trees saved from each replicate. The resulting trees were examined for potential phylogenetic utility of that marker.

Promising markers had the following characteristics: the clones for each species formed a single clade, only one copy (or type) of gene was recovered, and adequate resolution was present to differentiate the six species in the phylogeny. Multiple sequence 'types' were distinguished when phylogenetic analyses of the sequences of a given primer combination yielded two or more clades of sequences, indicating either multiple alleles at a single copy locus and/or duplicate genes at separate loci. Primer pairs were also considered good in the case where two clearly distinguishable gene types were recovered since a researcher could readily choose to use one or both of the types Primer pairs that met these criteria were considered to amplify orthologous gene regions and were identified as potentially good markers for phylogenetic analysis. After identifying potentially useful markers, we expanded our taxon sampling in *Phalaenopsis* and *Helianthus* to include species from throughout the genera (TABLE 3.2).

TABLE 3.1. Species used to screen putative primer pairs, their taxonomic information (angiosperm clades from Angiosperm Phylogeny Website (Stephens 2005)), and herbarium voucher numbers.

Species	Family	Angiosperm Clade	Voucher information
Phalaenopsis aphrodite	Orchidaceae	Monocots	Padolina A008 (TEX)
Phalaenopsis cornu-cervi	Orchidaceae	Monocots	Padolina A081 (TEX)
Phalaenopsis deliciosa	Orchidaceae	Monocots	Padolina A010 (TEX)
Phalaenopsis equestris	Orchidaceae	Monocots	Padolina A014 (TEX)
Phalaenopsis fasciata	Orchidaceae	Monocots	Padolina A017 (TEX)
Phalaenopsis pulcherrima	Orchidaceae	Monocots	Padolina A006 (TEX)
Helianthus annuus	Asteraceae	Euasterid II	Randal Linder 2AA (TEX)
Helianthus bolanderi	Asteraceae	Euasterid II	Ruth Timme 60 (TEX)
Helianthus cusickii	Asteraceae	Euasterid II	Ruth Timme 61 (TEX)
Helianthus giganteus	Asteraceae	Euasterid II	Ruth Timme 51 (TEX)
Helianthus glaucophyllus	Asteraceae	Euasterid II	Ruth Timme 42 (TEX)
Phoebanthus grandiflorus	Asteraceae	Euasterid II	Orzell & Bridges 14500 (TEX)
Aquilegia chysantha	Ranunculaceae	Basal Eudicots	
Arabidopsis thaliana	Brassicaceae	Rosids: Eurosid II	
Geranium maderense	Geraniaceae	Rosids: Geraniales	
Peperomia obtusifolii	Piperaceae	Magnoliids	
Reinwardtia indica	Linaceae	Rosids: Eurosid I	
Tiquilia hispidissima	Boraginaceae	Euasterid I	

TABLE 3.2. *Phalaenopsis* species and their herbarium voucher information used to construct phylogenetic trees from potentially useful primer pairs and the markers they amplify.

Species	Herbarium Number
Phalaenopsis chibae	Padolina A001 (TEX)
P. pulcherrima 1	Padolina A006 (TEX)
P. amabilis	Padolina A007 (TEX)
P. aphrodite 1	Padolina A008 (TEX)
P. deliciosa	Padolina A010 (TEX)
P. buyssoniana (natural autotetraploid of *P. pulcherrima)*	Padolina A011 (TEX)
P. pulcherrima 2	Padolina A012 (TEX)
P. equestris	Padolina A014 (TEX)
P. fasciata	Padolina A017 (TEX)
P. bastianii	Padolina A028 (TEX)
P. hieroglyphica	Padolina A028 (TEX)
Paraphalaynopsis laycockii (outgroup)	Padolina A047 (TEX)
P. venosa	Padolina A045 (TEX)
P. aphrodite 2	Padolina A059 (TEX)
P. cornu-cervi	Padolina A081 (TEX)
P. amboinensis 1	Padolina A098 (TEX)
P. amboinensis 2	Padolina A099 (TEX)
P. aphrodite 3	Padolina A117 (TEX)
P. amboinensis 3	Padolina A119 (TEX)
P. violacea	Padolina A120 (TEX)

RESULTS

GenBank BLAST Results

The BLAST files of the 'best' 400 primer pairs were examined. Oddly, a large portion (28%) of these pairs bracketed organellar DNA or DNA targeted for the organelles, especially the mitochondria. We eliminated these from our study because it was difficult to determine if they were truly nuclear in origin, a critical condition for their use in reconstructing hybrid evolution. TABLE 3.3 summarizes the information for the 142 markers that we screened in the lab. Ninety-two (65%) of these primer combinations had sequences that matched in species other than *Oryza* or *Arabidopsis* and 87 (61%) had information about the putative genes that the primers would amplify.

TABLE 3.3. Summary of primer pairs screened in the lab, including the ID number (given by MoBIoS), the name of the primer pair (which we assigned to the primers we tested in the lab), the forward and reverse primer sequences (both 5' to 3'), any species that matched the BLASTed region (see FIGURE 3.1), and the putative genes that these regions matched in the BLAST results.

ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
1	1F/1R	GGTTTAGTGAAAATATCAGC	TATGGTTTGAAACAAGCACCT	*Lotus, Glycine, Oryza*	
2	2F/2R	TATCACCCGGGAAAGGCTAATGT	TTTAGTTGCGACACATGGAA	*Oryza, Lotus*	
7	3F/3R	AAACTCAGCTTCGTATTG	TTTGGTTATCATGCTCCAGAGTAA	*Zea, Glycine*	kinase-like protein, serine/threonine
35	4F/4R	AAACTCAACAAGCAGCTGCCTTAC	TGAAAGGCATCACGGACAAGCTTTGC	*Brassica*	Tat binding protein 1
47	5F/5R	ATCCAGAAGGAGTCCACCCTTCA	TCCTTCTGGATGTTGTAGTCGGC	*Drosophila, Antirrhinum, Elaeagnus, Oryza, Saccharum*	polyubiquitin, ubiquitin
48	6F/6R	GATGGACAGGTGATCACCATTGG	TAGAAGCACTTCCTGTGGAC	*Helianthus, Nicotiana, Avena, Mimosa, Solanum*	actin
51	7F/7R	CAGAACAAGAACTCGTCCTACTT	GTATCAGAAACCTTTGGTGA	Pea*, Oat, Lotus, Eleusine, Lupinus, Zinnia, Zea*	beta tubulin
56	8F/8R	CCAAGAGCTTGGATGACATCAGT	TGTCTTACAAGGATATGAG	*Oryza, Xenopus,*	similar to pre-mRNA processing factor 8
57	9F/9R	TCCATGTTGAGCTCCTCGAACCT	CAACACCGTCTTCGGTAA	*Brassica, Spinacia, Lycopersicon, Cucurbita, Pisum…*	heat shock cognate protein (Hsc70) cytosolic heat shock 70 protein
63	10F/10R	CCACGGAAAAGAACTGGGTAGTG	AACAAGGACAACTTGGCTGCTCTTATGGTA	*Lotus, Oryza, Solanum, Flaveria*	glycine dehydrogenase P
69	11F/11R	TTGATCACTTCTGGAGCAACATA	AGAGTGTTGCTATCAAGATGAT	*Oryza*	SNFL3, putative serine/threonine protein kinase
72	12F/12R	CATGCGCAGAAAGTGATA	TAGATGAGTTCTACAGCAAAAT	*Oryza*	
74	13F/13R	CGTCACCGTCTGCGAGATCAAC	TACCTTGAGAAACATCCCA	*Brassica*	
75	14F/14R	TGCCTCTACCGGTATCCACACG	ATCCTTAAGTACTTGAGAAGAAC	*Oryza*	
77	15F/15R	ACGGACAAGAGCAAGCTCGATG	GAGAAAGCCTCGTAGAACTTGTTGTAGTCTTCCTTGTTCTCAGC	*Horedum, Mus, Tritticum, Oryza*	heat shock protein 90
78	16F/16R	AGAGCATCACAGAACCTATCTA	AACAAGGACAACTTGGCTGCTCTTATGGTA	*Lotus, Oryza, Flaveria, Solanum, Pisum, Tritordeum, Avena*	gdcsP gene
80	17F/17R	GCAGAGTCCCTTGCACGTTTTCAAGGAGG	TACCTTTCCGGTGATCAAT	*Oryza, Cicer, Nicotiana, Lotus, Lycopersicon, Medicago, Zea, Beta*	GDP dissociation inhibitor (gdi)

ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
87	18F/18R	TTGTTACCTGTCGAGCTTCAAGGTTTGA	TGGCATCCATTCACTGAGGAAGC	*Oryza, Cucumis, Zea, Solanum, Nicotiana*	ferrochelatase
88	19F/19R	ATACTCCCTCCGTTTCACAAT	AACTTATATTGTGAAACGGAGG	*Oryza*	
91	20F/20R	GATCTGAAGCATCATGTTGAA	TTTATTGATGAAGTGCACACGCT	*Lycopersicon, Brassica, Oryza, Pisum, Triticum, Lotus, Medicago, Zea, Nostoc*	ATP-dependent protease (CD4A) gene ClpC protease (clpC)
93	21F/21R	ATGAAAAAGCTTAAGAAATCTGAGATGCT	TACCTTTCTCCATGGACCTGA	*Mesembryanthemum, Lotus, Oryza, Zea*	protein kinase
97	22F/22R	GCTATGGAGAAGGTTGGCAAAGA	ACCTTTAGCAGCATCATAACC	*Oryza, Zea, Cucrbita, Brassica, Prunus*	mitochondrial chaperonin-60
99	23F/23R	AAGAAATTTGGGCAGTCT	AGAAACACCTCCAATGACCCAAAACTG	*Medicago, Nicotiana, Zea, Oryza, Gossypium*	cellulose synthase
6	24F/24R	GCAACATCTCTTGTGGAT	ATTGTCCACAAGGATAAA		ACC oxidase
8	25F/25R	AACATGTTAGAAATTCCA	ATGGAAACGATTTCTCTT		
10	26F/26R	TTGACCAAGTTTATAGCA	CATGTTTACACGAAAGTA		
11	27F/27R	ATATGAGAAAGCTTTTGC	AGCCACGAGATGGCTCTC		
12	28F/28R	AGGTTGGCAAGATTATCA	TCAGGAAAGTCCCTTGCT		
15	29F/29R	AAGAAATACTGTGAATGCTATCAGGT	AGTATCAGCTTCCGGCCA		DNA binding protein
24	30F/30R	GCTAGCCATGTCCTTGAGGT	TCAGATTTTGGGCTTGCCAAAGATG		
43	31F/31R	CAGATGAGCTGATCAAGACGGCCAAGTACAT	ACCTTTGGGCTGTCGGAGCC		
49	32F/32R	TCTAAGATGTTCTTTGCAGGTAA	AAAGCATATGGTTGTCAC		
50	33F/33R	TCAAACATTGCAATCTTTTGAAC	GCATTTGGGTTTGTCCACAAGAT		
52	34F/34R	TAGGTAAGCAACACCTTGCTAAT	CGTGTGGATACCAGTAGAGGCAC		
53	35F/35R	ATTCACCTGGTACATAAG	TCTGCTTGTGGAGAAGCTAACAA		
54	36F/36R	GCAGGTGTTCCAGCTTCAAATCT	GATGCTTGTCAAAGTGTTCCTCA		
55	37F/37R	AAACCTGAAGTCCAAGAGCCCATTTTC	GGTAAATTATATGCTGGACCTGA		protein kinase
59	38F/38R	GAGAATGTGAAGCAGCGTTGCTT	AACTCTTTGGTCATGGGACACCA		
60	39F/39R	GAAAGGCGCAAACTATTCAAGTA	GCTTGCAGATCAACCTGAAA		
61	40F/40R	AGGATGTGTATTGACTATAGGAA	AATTCTTGGAGCAGCAAAATCCATCT		reverse transcriptase, retroviral element, centromere region
76	41F/41R	TAATGAACAAGCATGATAGGCC	GCTTGCAGATCAACCTGAAA		chromatin remodeling protein SYD (SPLAYED)

ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
81	42F/42R	CTTCATCATCCAAATGTGGTT	AGAAGCTCTGGTGCCATCCATGGTAAGGTTCC		protein kinase family
83	43F/43R	TATCAAGAACAAATATCCTTT	GTGAAGATATACTTCAGACTCTT		retroelement hbc19 integrase gene, centromere region
90	44F/44R	TATGCTGACAATGCTCCGTCAAG	TACCCCTCGAAGAAGGCTGAC		
92	45F/45R	GTATGTGTTTTTACAAAATCAT	TCTAAATATGCATGCTTCAAG		
109	46F/46R	ATGCAGATCTTCGTGAAGAC	TCCTTCTGGATGTTGTAGTCGGC	*Antirrhinum, Phaseolus, Saccharum, Petroselinum, Vicia, Suberites, Nicotiana, Elaeagnus, Avena*	ubiquitin , polyubiquitin, polyubiquitin precursor, tetraubiquitin
110	47F/47R	TCATGGAAAGGTAGATAGCC	GCCGACTACAACATCCAGAAGG	*Oryza, Zea, Sorghum, Pisum*	protein kinase, serine/threonine protein kinase, calcineurin B-like-interacting protein kinase
114	48F/48R	GACTACAACAAGTTCTATGA	GTCTTCTTGCTCGACATGTA	*Hordeum, Triticum, Euphorbia, Zea, Oryza, Mus, Nicotiana, Lycopersicon, Hevea*	cytosolic heat shock protein, heat-shock protein 80
115	49F/49R	ATCCTTCCTTGGATGTCTGA	GAAACAAGCCTTAAGTTTATCATC	*Medicago, Solanum, Oryza*	
117	50F/50R	TCTGAGTTCATCAGCTACCC	GTTGTAGTCTTCCTTGTTCTCAGC	*Hordeum, Triticum, Mus, Oryza, Lycopersicon, Zea, Nicotiana*	cytosolic heat shock protein 90, HSP80-2 protein, heat shock protein 82
120	51F/51R	CCTAGCTTGATGACACCACT	CTTGGACGTATCATGAATGTT	*Nicotiana, Hevea, Pisum, Sorghum*	atp2-1 gene for mitochondrial ATP synthase beta subunit, F1-ATP synthase
122	52F/52R	GAGCCATACAAGGGTATT	AGACCACGGTAGACAATGAT	*Gossypium, Lupinus, Triticum, Oryza, Solanum, Lycopersicon, Maize*	adenine nucleotide translocator 1 (CANT1), ant gene for ADP/ATP translocator, mitochondrial adenine nucleotide translocator (ANT-G1)
126	53F/53R	ATGGTTCTTGATAATGAGGC	CCACGGAACATGGCAGAGGC	*Pisum, Eleusine, Medicago, Citrus, Gossypium, Hordeum, Glycine, Oryza, Triticum*	beta-tubulin 3, beta-tubulin 4 (TUB4)
127	54F/54R	ATGAACATCATGCTTACCAACG	ACGAACCCGCAGATGATCGC	*Brassica, Lotus, Lycopersicon, Oryza*	Brassica napus high-affinity ammonium transporter AMT1;2 mRNA Lotus japonicus amt1.1 gene for ammonium transporter, exons 1-2 L.esculentum mRNA for ammonium transporter Oryza sativa ammonium transporter 1-3 (Osamt1-3) gene
130	55F/55R	ATGGCTTTTACACCATCAGGT	TCCTTTGAAAAGTTTAAGGA	*Oryza, Oryza, Zea*	Oryza australiensis retrotransposon RIRE1 chromosome 3 BAC OSJNBa0037J17 genomic sequence Zea mays chromosome 9S bz genomic region strain McC

ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
131	56F/56R	GTCCTTCGCTTCAGGGCCTG	TTTCATGATAGCAAGGTGCTCCCC	*Brassica, Oenothera, Citrullus, Oryza, Malus, Beta, Triticum, Helianthus, Vicia, Cucumis, Gnetum, Cycas, Ginkgo*	mitochondrial 5S rRNA and ND5 gene for NADH dehydrogenase subunit 5, tRNA-Gly and tRNA-Glx genes; sdh3 pseudogene; and nad9 and atp9 genes
139	57F/57R	AAGGAGGTTTCTCATGAGTGG	GTCTTCTTGCTCGACATGTA	*Hordeum, Triticum, Mus, Zea, Oryza, Lycopersicon, Nicotiana, Nicotiana, Euphorbia, Hevea, Pharbitis*	cytosolic heat shock protein 90 (HSP90), heat shock cognate protein 80 gene, heat shock protein 82, heat shock protein 83 (Hsp83) gene
140	58F/58R	AACATCTCCTGGATTGAGGT	GACAATGAGGCTCTCTACG	*Brassica, Glycine, Gossypium, Oryza, Eleusine, Zea, Triticum, Lotus, Zinnia, Pisum, Avena*	beta-tubulin 5, beta-tubulin 1 (TUB1), beta-7 tubulin (tub7), beta-tubulin 2
143	59F/59R	ATCAACTTGATGAAGTACTCCAT	TACAAGAATGGTCATGTCG	*Zea, Cucumis, Corylus, Glycine, Hordeum, Spinacia, Nicotiana, Oryza, Lycopersicon*	lumenal binding protein cBiPe3, heat shock protein 70 (BiP-related), endoplasmic reticulum HSC70-cognate binding protein precursor (BIP)
144	60F/60R	ACTCTTGAATTTCTATGGCAAGA	AAATTGAACTTCCATCCTGG	*Raphanus, Oryza, Zea*	prolyl-tRNA synthetase
150	61F/61R	ATGGGACAAAAAGATGCTTA	TAGAAGCACTTCCTGTGGAC	*Solanum, Oryza, Nicotiana, Vicium, Gossypium, Avena, Sorghum, Hordeum, Zea, Pyrus, Phalaenopsis, Lactuca, and many others*	actin
151	62F/62R	TATGTTGCACCTGAGGTTCT	CCATCTTTATCAAAGTATG	*Glycine*	calcium dependent protein kinase
152	63F/63R	CTCAACATTGCAGAGAAAGG	ATCTTGAACCACTCGGTGTGG	*Glycine, Medicago, Spinacia, Oryza, Zea*	6-phosphogluconate dehydrogenase
153	64F/64R	AACCGTCCCAACAGCATTGA	GCACCACTGAAGCCTTGGGTG	*Glycine, Oryza, Capsicum*	transitional endoplasmic reticulum ATPase, transitional endoplasmic reticulum ATPase
157	65F/65R	TTTGTCAGTACAGTGGATCC	CTTTGTGGGAATTGGACATA	*Gossypium, Oryza, Populus*	cellulose synthase
161	66F/66R	GCATTTGGAAATGCAAAGACTG	AAGCACCTGTTTGTCTTGAA	*Medicago, Petroselinum, Nicotiana, Oryza, Helianthus*	myosin
163	67F/67R	AAGTTCTTGCGCCGTGTGG	TGCTCAATGACTTCACCCTGAAC	*Brassica, Mesembryanthemum, Nicotiana, Solanum, Gossypium, Lycopersicon, XMokara cv. 'Yellow (orchid), Zea, Glycine, Flaveria, Phaseolus, Oryza*	phosphoenolpyruvate carboxylase
165	68F/68R	GAGTTCATGACATCAATGG	ACCCTGGAGATCGGCATGGG	*Horteum, Citrus, Gossypium, Oryza, Zea, Mesembryanthemum*	vacuolar H+-ATPase B subunit
166	69F/69R	CATTGAGGAGAACAATGCAGG	TGCTCAGGAGCACCTTTGC	*Prunus, Vicia, Juglans, Sesbania, Solanum, Nicotiana, Oryza, Lycopersicon, Zea*	plasma membrane H+-ATPase
167	70F/70R	GCAGAAAAGAGGGGAATC	ACCCTGGAGATCGGCATGGG	*Oryza, Hordeum, Citrus, Mesembryanthemum, Zea*	vacuolar H+-ATPase B subunit
168	71F/71R	TCCTTGAACTTTTCAAAGG	ATTAGCAAGGTGTTGCTTACC	*Oryza, Avena, Triticum*	retrotransposon OARE-1 gag-pol pseudogene

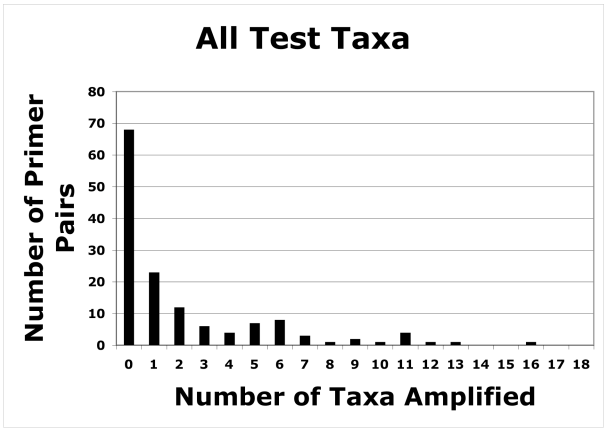
ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
172	72F/72R	AACATCCGTAACATGTCTG	ACAAGCTTGGGAAGGTCAG	*Oryza, Zea, Pisum, Beta, Lycopersicon, Nicotiana, Triticum, Bonnemaisonia (red alga), Filobasidiella (fungus), Euglena,*	elongation factor 2
176	73F/73R	CGGGTTCGAGTCCCGGCAACGG	TGGAACCACTTCATGGCATC	*Oryza, Zea, Nicotiana, Triticum, Cicer, Pyrus*	ribosomal protein L11, ribosomal protein RL5
179	74F/74R	ATGATGCGGAAACCGGAAGG	TCCAAATGCCTGGGATGAAGG	*Brassica, Beta, Oryza, Zea, Triticum, Coix*	homeodomain-leucine zipper protein 8
180	75F/75R	GGATGCATTCCATGCTCAAG	GTTACAAGAGATGTATGCTT	*Gossypium, Nicotiana, Brassica, Citrus, Brassica, Triticum, Euphorbia, Hydrilla, Phaseolus, Zea, Vitis, Glycine, Dendrobium, Saccharum, Clusia, Taxus, Cupressus, Clusia, Mesembryanthemum, Kalanchoe*	phosphoenolpyruvate carboxylase
185	76F/76R	ACAAAAGCACCCTTGGTGTC	GATTTCGAGCCATTCAATGC	*Brassica, Citrus, Gossypium, Carica, Medicago, Oryza, Lotus, Oncidium, Bambusa, Daucus, Triticum*	sucrose synthase
186	77F/77R	GTTTCATTGAGGAGAACAATGC	GCAAAGCCATCAGCTTTCTC	*Nicotiana , Vicia, Prunus, Lycopersicon, Oryza, Zea, Medicago, Kosteletzkya virginica*	plasma membrane H+ ATPase
285	78F/78R	GGTTTTGAGGCTGGTATCTC	TCATCTTAACCATACCAGC	*Arabidopsis, Lycopersicon, Nicotiana, Salsola, Daucus, Malva, Horedeum, Lilium, Saccharum, Zea,*	Elongation factor 1-alpha
282	79F/79R	GACCCGTGAGCACGCTCTTCTTGC	TCATCTTAACCATACCAGC	*Arabidopsis, Lycopersicon, Nicotiana, Salsola, Daucus, Malva, Horedeum, Lilium, Saccharum, Zea,*	Elongation factor 1-alpha
287	80F/80R	ATCATTGACTCCACCACTGG	TCATCTTAACCATACCAGC	*Arabidopsis, Lycopersicon, Nicotiana, Salsola, Daucus, Malva, Horedeum, Lilium, Saccharum, Zea,*	Elongation factor 1-alpha
293	81F/81R	CGAGACCACCAAGTACTACTGC	TCATCTTAACCATACCAGC	*Arabidopsis, Lycopersicon, Nicotiana, Salsola, Daucus, Malva, Horedeum, Lilium, Saccharum, Zea,*	Elongation factor 1-alpha
301	82F/82R	CAAGAGGTCCTTCAAGTACGC	TCATCTTAACCATACCAGC		
368	83F/83R	GCAACTGGGATGACATGGAGAAG	AAGCTTCTCCTTGATGTC	*Brassica, Striga, Gossypium, Solanum, Zea, Stevia, Nicotiana, Musa*	actin
339	84F/84R	AACAAGGAGATCTTCCTCCG	GTAGTCTTCCTTGTTCTCAGC	*Oryza, Lycopersicon, Triticum, Zea*	heat shock cognate protein 80, heat shock protein 82
209	85F/85R	CTGTGACAATGGAACCGGAATGG	TAGAAGCACTTCCTGTGGAC	*Gossypium, Solanum, Sorghum*	actin
379	86F/86R	ATGGTTGAGTACTTTGGTGAGCAG	GCCCAGTCCAAGTAGAAAGC	*Triticum, Solanum, Glycine, Zea*	methionine synthase
222	87F/87R	GTGATCACTACATCAACAATAGC	GTAACAGAACCATAGATCCA	*Populus, Horedeum*	Cellulose synthase
279	88F/88R	GAGCTTCATCCCAAATAAT	AAGTATCTTTTTAAATATG	*Lotus, Arabidopsis*	
348	89F/89R	GGCCCAACCGGGTTCGAACCG	AAGGGTCCATAGCTCAGTG	*Lotus, Trifolium*	

ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
		G			
349	90F/90R	CAGGTTCTCATTTGGGATGT	CCTGAACACAAAGAACAGCAGC	*Medicago*	
262	91F/91R	GTGGTCTTGTCCTCAGCAGA	TGCGTATCATCAACGAGCC	*Oryza*	
203	92F/92R	GGTATTTATGTTACATACAACGG	TCCATTGCAAACCGAACCAT	*Arabidopsis*	
207	93F/93R	CTGAAGAACTATGACCCTC	AGAACCTTCTTCAGCTGGAA	*Zea*	
212	94F/94R	GAGAGCAGCGACACCATCG	CCAGTGAGTGTCTTCACC	*Deschampsia, Zea*	
220	95F/95R	GGATTACCATCCAAGAATGCC	TGGAAGGAACACTCCATGAT	*Oryza*	
255	96F/96R	TGGTGTGATACAAAATCAGT	TCAGGGCTAATATCAGTAT	*Medicago*	
327	97F/97R	GAGGCAATCAGGCAAAGCT	ATTTATGTTTTTAAAACA	*Mus*	
360	98F/98R	GGGATGGATACTAGTAAGTG	GGCTTCTCAAAGTTGTAGTTGC	*Cicer*	GTP-binding protein
371	99F/99R	GTTGCAGTATTCATATGAG	TGAATCTTCAAGCAACTG		
264	100F/100R	TTGATACTCCCTCCGTTTC	ATTGTGAAACGGAGGGAGTA		
265	101F/101R	TTGGGAACTTGAAAGAGCT	AGTCTTGCTAGCCCAAAATC		
266	102F/102R	TTGAGTGTTCTCAAATGGC	TGAACCTGTTGAGAAGCCCA	*Brassica*	arginine methyltransferase
272	103F/103R	TGGTCAAAGTCCTCACCTG	TACAAGAATGGTCATGTCGA	*Hordeum*	
288	104F/104R	TCTGCCAGTACATCTTGCA	GACAAGTCTCACATCTTTGC		
290	105F/105R	TATCACTCCTTTTGGAACCGTAC	GTTTCTGGTCCTTCTTTGA	*Lycopersicon*	gamma-aminobutyrate transaminase subunit precursor isozyme
299	106F/106R	GCTGTTCGACGAAATGCCT	CAAAACCCATCTTTGAAATG	*Sorghum*	
308	107F/107R	TCCTAAAGGCTTGTCAAAG	AAAGGAGTTGTTTGTAAGAA	*Lotus*	
313	108F/108R	TTATTACCTCAGTTCCCCA	GATCAGTTCCCCAAGTGGT	*Populus*	cellulose synthase 6
323	109F/109R	GTCTATGGAGACTTGAGTA	GGTATGTAAATCTGACCATG	*Mesembryanthemum*	potassium transporter
352	110F/110R	ACATTTATCAAAATTGGGCA	CCTTCCCATCATTCCAAAGTCG	*Zea*	
355	111F/111R	CCACTAGTTATGACTATGATGC	CCTGATAAGTCCATTTCTGCC	*Lotus*	
361	112F/112R	CATCAAGAGATAATTCCTC	TAGTTATGGTGAACCTAC	*Actinidia*	sucrose-phosphate synthase
369	113F/113R	GATCTGCACGGTCGGTGG	TGAAATTGATCGTTATGC	*N.plumbaginifolia*	small GTP-binding protein
386	114F/114R	GTGGAGCATTTCAATGCGGA	AACCTGAATCATATGCCA	*Lotus*	
393	115F/115R	CAGCTTATGAGTACATATT	AACCTGAATCATATGCCA		
231	116F/116R	TCATCATTTTGGGGAATGG	AAGCTATTCATCACACCTG		
396	117F/117R	TTTCTTCATGAGAAATTA	TAATTTCTCATGAAGAAA		
331	118F/118R	ATACTCCCTCCGTTTCACA	ATTGTGAAACGGAGGGAGTAT		
303	119F/119R	TGGTTATCTTTGACGGTCA	GTTGAGATCTTCTATACACA		
297	120F/120R	TCCATGTTGAGCTCCTCGAACC	ACCAACACCGTCTTCGGTA	*Oryza*	heat shock protein 70
240	121F/121R	GTGAAGATATACTTCAAACT	TCTTTATTGATGACATTCT		

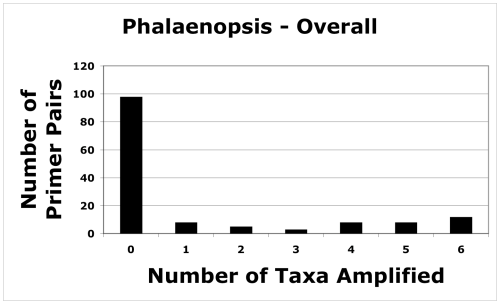
ID #	Primer Name	Primer 1 (forward)	Primer 2 (reverse)	Matching Species	Putative Genes
294	122F/122R	CCTATTGATCTGCAGTCAATTGC	AAGGGTCCATAGCTCAGTG		
239	123F/123R	TCTTTATTGATGACATCCT	AAGATATACTTCAAACTCTT		
121	124F/124R	AACATTCATGATACGTCCAAG	CCTAGCTTGATGACACCACT	*Nicotiana*	gene for mitochondrial ATP synthase beta subunit
122	125F/125R	GAGCCATACAAGGGTATT	GACCACGGTAGACAATGAT		
133	126F/126R	TGCGCTATTTACCAGTGAGTGG	AAAGGGGATGTTTTGTTAAT	*Oryza, Zea*	ribosomal protein S4 type I (rps4)
134	127F/127R	ATTGACTATAGGAAGCTTAA	GGAGCAGCAAAATCCATCT	*Brassica, Pisum, Oenothera, Beta, Triticum, Oryza*	mitochondrial ccb248 gene and partial rps7
142	128F/128R	ATCAACTTGATGAAGTACTCC	ACAAGAATGGTCATGTCGA	*Oryza*	
145	129F/129R	TATGCCAGTGGTCGTACAAC	TTAGAAGCATTTCCTGTG	*Zea*	
154	130F/130R	TTCAAGCATCCAGGTTGCTGGG	GTTGCAAATCCTTCCATCAT	*Medicago, Oryza, Nicotiana*	PDR-like ABC transporter, pleiotropic drug resistance like protein
156	131F/131R	CATGGAAAGTCCACTGTTG	AGATCAATCTTGTTCTGAAG	*Zea, Oryza*	translation initiation factor eIF-2 gamma subunit
159	132F/132R	CAATATTTATGGTACCCATGTCC	GGATTGGTGAAGTTCCATAT	*Oryza, Zea*	NAD-dependent epimerase/dehydratase family
162	133F/133R	ATGAAATCATTCTTGAGGAC	CTACGCCAACAAGGTCGGCCC	*Oryza*	ammonium/proton antiporter
164	134F/134R	CCAACACTCCAGTTCTCGCC	GATGGACTTCCACAACAATG	*Triticum, Oryza, Zea*	glutathione-conjugate transporter AtMRP4, multidrug resistance-associated protein MRP1
169	135F/135R	TGGGGATCCTCAGCTTTTGG	CAGAGTTGCTTCCATTCCT		
170	136F/136R	CACCAACAAACTCTGGAATGC	TTGCTGGCTTCAATATACCT	*Oryza*	valine--tRNA ligase-related protein
173	137F/137R	AGCTCTTCAAGTGTTCC	AAGGAACATTGATCTAGT	*Oryza*	lysine decarboxylase - like protein
174	138F/138R	ACCCGCTGTGAATCAATAAA	CTTAGTGAGGCATGGCGGTG	*Oryza, Homo sapiens, Xenopus, Dictyostelium*	splicing factor
178	139F/139R	TGGAAGACCCATTTGCTT	CAGAGTGTGGGGTTGATTT	*Oryza*	dolichyl-di-phosphooligosaccharide glycotransferase
181	140F/140R	ATCTCACGAATGAGGTGATA	CCTTTGAAAAGTTCAAGGA	*Oryza, Gossypium Raimondii, Amaranthus, Setaria*	retrotransposon reverse transcriptase gene
183	141F/141R	CCACGACTTTCAGGACC	GACCATCTGTAATGATGGA	*Oryza, Botryocladia, Porphyra, Dictyostelium*	RNA polymerase II largest subunit (RPB1)
184	142F/142R	GGTCACTATGAAGAACA	CTGCAACCACATTAGCCTT	*Arabidopsis, Oryza*	retrotransposon gag protein family

Amplifications

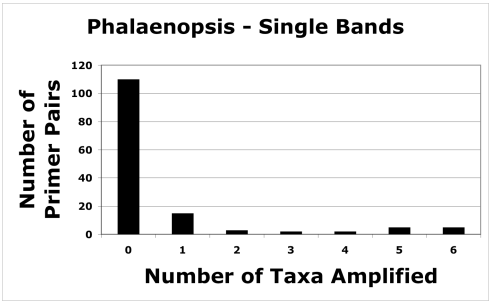
Over 52% of the 142 primer pairs amplified in at least one of the taxa that were screened, however, none of the primer pairs amplified in all eighteen test taxa, and only eight primer pairs amplified over 50% of the test taxa (≥ 9 taxa) (FIGURE 3.3A). For the *Phalaenopsis* species, at least one species was amplified successfully by 31% (44) of the primer pairs and 22% (31) of the primers amplified in at least 3 test taxa (FIGURE 3.3B). However, 17% (24) of the primer pairs yielded multiple bands as visualized by gel electrophoresis in one or more of the *Phalaenopsis* species, reducing the success rate of primers that amplified only a single band in at least one of the test taxa to 23% (32) and primers that would amplify a single band in three or more taxa to 10% (14) (FIGURE 3.3C). For *Helianthus*, only 10% (14) of the primers successfully amplified at least one *Helianthus* species, and 8% (12) of the primers amplified at least three *Helianthus* (FIGURE 3.3D). *Helianthus* did not exhibit the problem of multiple bands that was seen in *Phalaenopsis*, hence rates did not change when only single bands were considered (FIGURE 3.3E). For the remaining species, "the other angiosperms", at least one species could be amplified by 35% (47) of our primer pairs; however, only 5% (7) of the primer pairs were able to amplify more than three taxa (FIGURE 3.3F). Multiple bands were not very common among these species, so success rates dropped only slightly to 32% (45) and 4% (6) for amplification of only single bands in at least one and at least three taxa, respectively (FIGURE 3.3G). Most of the amplification success in the other angiosperms was attributed to *Arabidopsis*. Only 12% (17) of the primers amplified at least one of the non-*Arabidopsis* taxa.



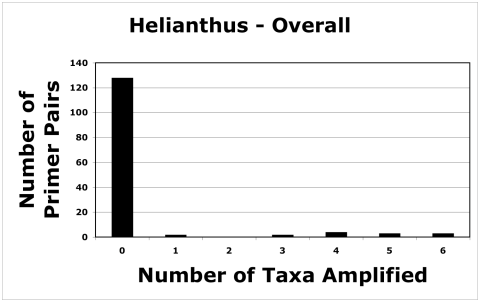
A.



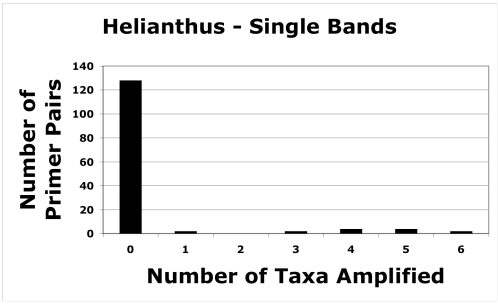
B.



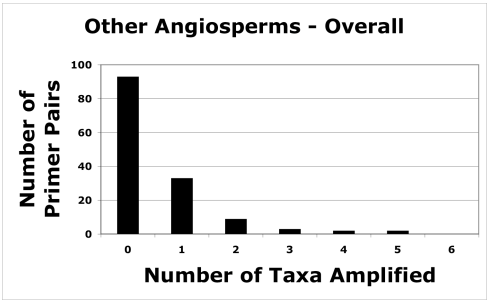
C.



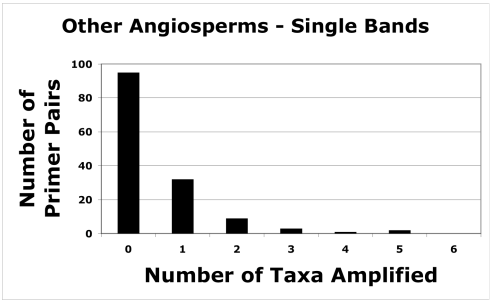
D.



E.



F.



G.

FIGURE 3.3. Histograms of amplification success rates of each primer A. amplification in all 18 test taxa, B. overall amplification in the 6 *Phalaenopsis* species, C. amplification of only single bands for *Phalaenopsis*, D. overall amplification in the 6 *Helianthus* species, E. amplification of only single bands for *Helianthus*, F. overall amplification in the 6 Angiosperms, C. amplification of only single bands for Angiosperms.

Sequencing

Direct sequencing of nearly all of the PCR products proved difficult for both *Phalenopsis* and *Helianthus*, resulting in failed or unreadable reactions. Almost all markers sequenced resulted in multiple types of sequences for each individual. A summary of these results can be seen in TABLES 3.4 (*Phalaenopsis*) and 3.5 (*Helianthus*) and are discussed in detail below.

Phalaenopsis

Initially, the amplicons for nine primer pairs that produced primarily single bands from PCR amplifications and amplified in all of the test cases, were sequenced for *Phalaenopsis* (6F/6R, 49F/49R, 51F/51R, 56F/56R, 57F/57R, 63F/63R, 86F/86R, 91F/91R and 130F/130R). After further sequencing, five of these primer pairs were selected for further development. TABLE 3.4 summarizes the sequence variation found among species and from cloned sequences within an individual for markers 6F/6R, 51F/51R, 57F/57R, 63F/63R, and 86F/86R for up to sixteen *Phalaenopsis* species. After sequencing of several sets of clones, 49F/49R, 91F/91R and 130F/130R were determined to be unsuitable phylogenetic markers because too many different sequence types were present. Even though these markers appeared to produce a single sequence type based on gel electrophoresis, 30-40% of the sequences within an individual for these primers were difficult to align because insertions and deletions were present in different parts of the region in different sequence types. Since the sequence lengths of these markers were fairly short in *Phalaenopsis*, less than 300 base pairs, we decided not to develop more specific primers for one or more of the types. The primer pair 56F/56R was unsuitable because there was insufficient phylogenetically informative variation across species.

TABLE 3.4. Sequence variation within and between individuals of *Phalaenopsis* species, including the number of sequences examined for each species (Seqs/Sp), the number of individuals examined for each species (Inds/Sp), and the number of sequence types found for each species (Types/Sp), as well as pairwise distance statistics between different sequence types within each species and between species.

	Seqs/ Sp.	Ind/ Sp.	Types /Sp.	Pairwise distance statistics: Within species Range	Within species Mean	Within species Median	Among species
6F/6R							
Phalaenopsis amabilis	10	1	2	0 - 0.2695	0.0978	0.0132	
P. amboinensis	24	3	2	0 - 0.3706	0.1464	0.0096	Range
P. aphrodite	13	2	1	0.0013 - 0.0226	0.0122	0.0119	0 – 0.7827
P. bastianii	5	1	1	0.0025 - 0.0151	0.0075	0.0063	
P. buyssoniana	1	1	1				
P. chibae	3	1	2	0.0128 - 0.4373	0.2917	0.4248	
P. cornu-cervi	7	1	1	0 - 0.0213	0.0081	0.0047	Mean
P. deliciosa	7	1	1	0 - 0.0052	0.0026	0.0026	0.2147
P. equestris	9	1	1	0 - 0.0160	0.0065	0.0065	
P. fasciata	14	1	2	0 - 0.3772	0.1055	0.0103	
P. hieroglyphica	7	1	2	0.0025 - 0.5050	0.3048	0.5025	
P. intermedia	12	1	2	0 - 0.0288	0.0139	0.0100	Median
P. pulcherrima	16	2	2	0.0041 - 0.3840	0.1073	0.0291	0.0897
P. venosa	9	1	1	0 - 0.0094	0.0057	0.0054	
P. violacea	8	1	4	0 - 0.6478	0.3825	0.4065	
Paraphalaenopsis	9	1	1	0 - 0.0081	0.0044	0.0041	
51F/51R							
Phalaenopsis amabilis	0	1	0				
P. amboinensis	5	1	1	0 - 0.0221	0.0095	0.0111	Range
P. aphrodite	4	1	1	0 - 0.0577	0.0384	0.0577	0 – 0.7599
P. bastianii	3	1	1	0.0079 - 0.0143	0.0106	0.0095	
P. buyssoniana	3	1	1	0.0032 - 0.0096	0.0064	0.0064	
P. chibae	1	1	1				
P. cornu-cervi	5	1	3	0.0033 - 0.7299	0.5102	0.7222	Mean
P. deliciosa	0	1	0				0.2477
P. equestris	1	1	1				
P. fasciata	4	1	1	0.006 - 0.0165	0.0107	0.0112	
P. hieroglyphica	5	1	1	0.0015 - 0.0079	0.0050	0.0047	
P. intermedia	3	1	2	0.0309 - 0.7536	0.5101	0.7457	Median
P. pulcherrima	2	1	1	0 - 0.0090	0.0090	0.0090	0.0401
P. venosa	0	1	0				
P. violacea	3	1	2	0.0098 - 0.7081	0.4753	0.7079	
Paraphalaenopsis	1	1	1				
57F/57R							
Phalaenopsis amabilis	6	1	1	0 - 0.0076	0.0039	0.0043	
P. amboinensis	9	2	3	0.0022 - 0.2512	0.1728	0.2421	Range
P. aphrodite	9	2	2	0.0021 - 0.2397	0.1461	0.2048	0 – 0.2587
P. bastianii	2	1	1	0.0043 - 0.0043	0.0043	0.0043	
P. buyssoniana	4	1	2	0.0064 - 0.2357	0.1350	0.1426	
P. chibae	4	1	1	0.0021 - 0.0194	0.0108	0.0108	
P. cornu-cervi	6	1	1	0.0011 - 0.0241	0.0108	0.0077	Mean
P. deliciosa	5	1	2	0 - 0.2326	0.0930	0.0032	0.1591

	Seqs/ Sp.	Ind/ Sp.	Types /Sp.	Pairwise distance statistics			
				Within species			Among species
				Range	Mean	Median	
P. equestris	5	1	3	0.0042 - 0.2373	0.1519	0.1877	
P. fasciata	4	1	2	0.0021 - 0.2365	0.1194	0.1198	
P. hieroglyphica	1	1	1				
P. intermedia	9	1	3	0 - 0.2375	0.1219	0.1829	Median
P. pulcherrima	4	1	3	0.0043 - 0.2561	0.1755	0.2519	0.1934
P. venosa	3	1	3	0.1955 - 0.2404	0.2244	0.2372	
P. violacea	6	1	3	0.0075 - 0.2503	0.1382	0.1958	
Paraphalaenopsis	4	1	2	0.0022 - 0.1809	0.0982	0.1008	
63F/63R							
Phalaenopsis amabilis	2	1	2	0.0898 - 0.0898	0.0898	0.0898	
P. amboinensis	3	1	2	0.02167 - 0.032	0.0255	0.0227	Range
P. aphrodite	9	2	3	0.0072 - 0.7368	0.2109	0.0689	0 – 0.7815
P. bastianii	5	1	3	0 - 0.4326	0.2586	0.4226	
P. buyssoniana	2	1	1	0.001 - 0.001	0.0010	0.0010	
P. chibae	4	1	1	0.0021 - 0.0204	0.0113	0.0113	
P. cornu-cervi	5	1	2	0.007 - 0.0755	0.0418	0.0358	Mean
P. deliciosa	6	1	3	0.0082 - 0.7067	0.3964	0.6970	0.1989
P. equestris	4	1	3	0.0078 - 0.0822	0.0566	0.0596	
P. fasciata	3	1	1	0.0089 - 0.0109	0.0098	0.0098	
P. hieroglyphica	1	1	1				
P. intermedia	4	1	3	0.0010 - 0.0719	0.0428	0.0459	Median
P. pulcherrima	8	2	4	0 - 0.0860	0.0496	0.0496	0.0679
P. venosa	1	1	1				
P. violacea	2	1	1	0 - 0	0.0000	0.0000	
Paraphalaenopsis	5	1	2	0.0062 - 0.0560	0.0336	0.0332	
86F/86R							
Phalaenopsis amabilis	2	1	2	0.2763 - 0.2763	0.2763	0.2763	
P. amboinensis	3	1	2	0.0309 - 0.7536	0.5101	0.7457	Range
P. aphrodite	16	2	3	0 - 0.7368	0.2218	0.1465	0 – 0.8233
P. bastianii	5	1	4	0.0059 - 0.4845	0.3136	0.4376	
P. buyssoniana	5	1	3	0.0039 - 0.4444	0.1856	0.0218	
P. chibae	6	1	2	0.0067 - 0.7543	0.2721	0.0505	
P. cornu-cervi	7	1	2	0 - 0.3413	0.0976	0.0023	Mean
P. deliciosa	6	1	1	0 - 0.0094	0.0051	0.0047	0.2159
P. equestris	11	1	2	0 - 0.4819	0.1601	0.0104	
P. fasciata	3	1	2	0.0449 - 0.4622	0.3193	0.4507	
P. hieroglyphica	8	1	2	0 - 0.3246	0.0827	0.0035	
P. intermedia	7	1	2	0 - 0.0523	0.0269	0.0290	Median
P. pulcherrima	14	1	2	0.002 - 0.7655	0.1222	0.0161	0.0987
P. venosa	7	1	2	0.0021 - 0.5178	0.2993	0.4956	
P. violacea	1	1	1				
Paraphalaenopsis	6	1	1	0.0045 - 0.0734	0.0360	0.0419	

The primer pair 6F/6R amplified a region that coded for an actin gene. More than one type of sequence was found for this region in nine of the fifteen *Phalaenopsis* species examined, both of which BLASTed to an actin gene. The common type will be referred to as actin 1 and the uncommon type as actin 2. Actin 1 was approximately 775 bp in length and contained a 451 bp noncoding region, which began at position 317 (GenBank accession number AB180246). Actin 1 was found to be phylogenetically informative, producing a well-resolved tree in *Phalaenopsis* (FIGURE 3.4). Actin 2 sequence products were only alignable with the actin 1 for 280 base pairs. Actin 2 was not pursued for phylogenetic use because it was not found in all of the species that were amplified.

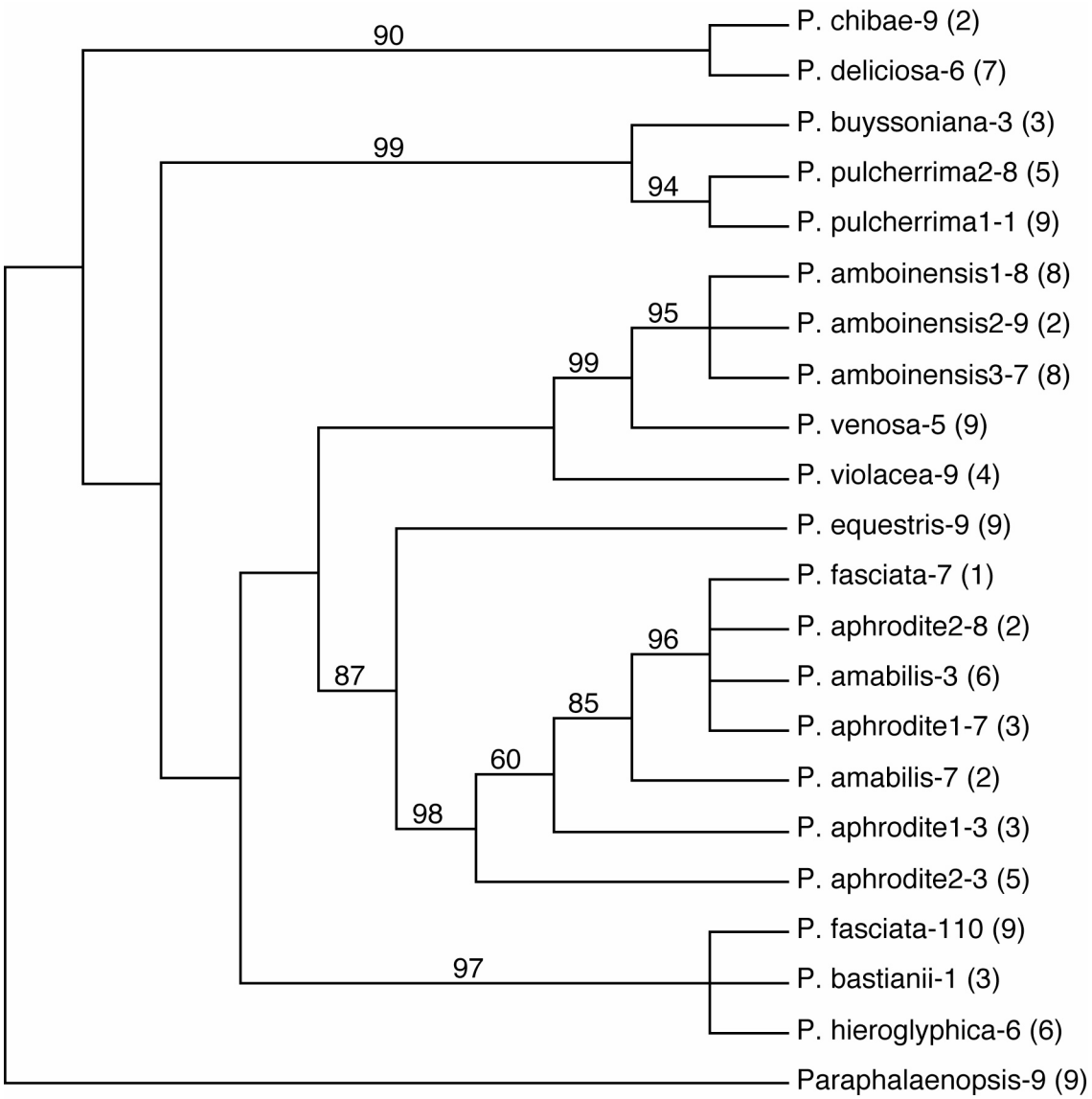


FIGURE 3.4. Maximum parsimony majority rule consensus tree reconstructed from the partial actin 1 gene amplified from primers 6F/6R. Numbers above the branches are nonparametric bootstrap support. The sequence from a single clone from an individual was randomly retained for analysis when cloned sequences from the individual formed a monophyletic group. The number of clones sequenced for each individual is indicated in parentheses.

Primer pair 51F/51R amplified a region that is a portion of an ATP synthase gene (GenBank accession number X02868) and appears to amplify a useful phylogenetic marker for low-level taxonomic groups. The amplicon is about 680 bp in length and contains two introns, one from positions 68 through 153 and one from base pair 227 through 443. This marker was difficult to amplify, and after cloning, many positive colonies yielded no gene product when amplified. Due to these difficulties, there were not as many sequences available for this marker, and *P. amabilis*, *P. deliciosa,* and *P. venosa* produced no sequences. Only three of the thirteen species examined yielded multiple products and we did not find a large number of sequence types to make determination of orthology difficult. The few secondary sequences that were obtained did not match any known genes in GenBank and were completely unalignable with the ATP synthase sequences. However, the small number of sequence types may be an artifact of the lower sample size. The 51F/51R ATP synthase amplicon possessed enough variation to resolve the species-level phylogeny of our subset of *Phalaenopsis* species (FIGURE 3.5). Designing more specific primers for the ATP synthase gene fragment might overcome the difficulties with initial amplification and cloning of 51F/51R.

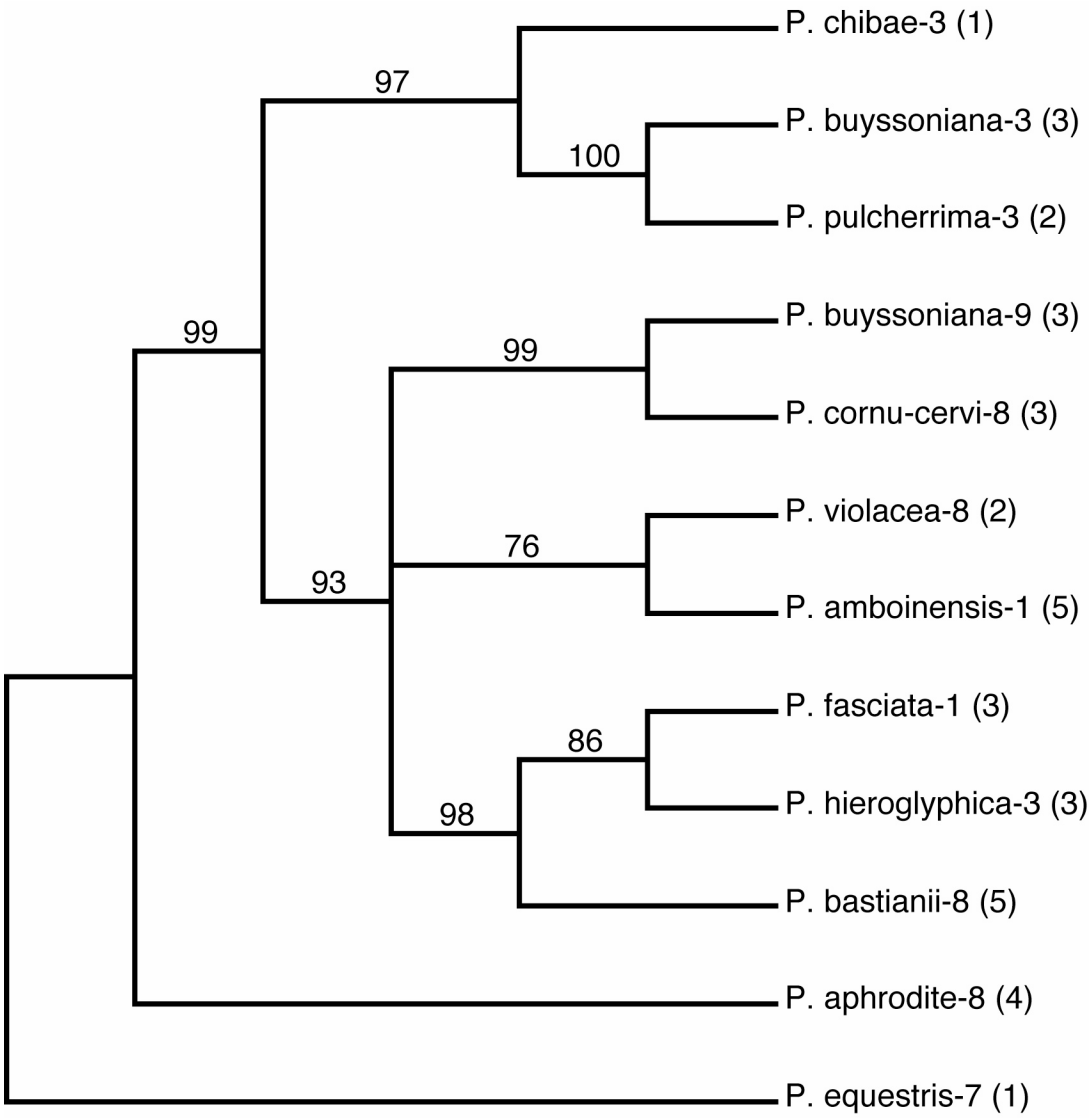


FIGURE 3.5. Maximum parsimony majority rule consensus tree reconstructed from the ATP synthase gene fragment sequenced from primers 51F/51R. Numbers above the branches indicate nonparametric bootstrap support. The sequence from a single clone from an individual was randomly retained for analysis when cloned sequences from the individual formed a monophyletic group. The number of clones sequenced for each individual is indicated in parentheses.

Primer pair 57F/57R amplified heat shock proteins (HSPs) and produced an amplicon of approximately 950 bp. All of the sequences were alignable with each other although it was clear from sequence variation that there were several types. Phylognetic analysis revealed three HSP types (HSP 1, HSP 2, and HSP 3). Type 1 occurred slightly more frequently than types 2 or 3 (FIGURE 3.6). Type 1 one BLASTed to HSP 80, HSP 80-2, HSP 81-2, HSP 82,HSP 90, HSP 90-1, and HSP 90-2 (GenBank accession number M96549, X98582, NM124985, NM124983, Z11920, AY325266, AY368906, and AY368905, respectively), all with e-values less than 10^{-100} and identities greater than 81%. Type 2 BLASTed to HSP 82 (X63195) and HSP 83 (M99431), with e-values less than 10^{-100} and identities greater than 81%, and type 3 to HSP 80, HSP 81-3, HSP 90, HSP 90-1, HSP 90-2, (GenBank accession number X98582, NM124983, AY325266, AY368904, and AY368905, respectively), all with e-values less than 10^{-100} and identities greater than 81%. None of these regions appeared to contain introns based on matching BLAST sequences.

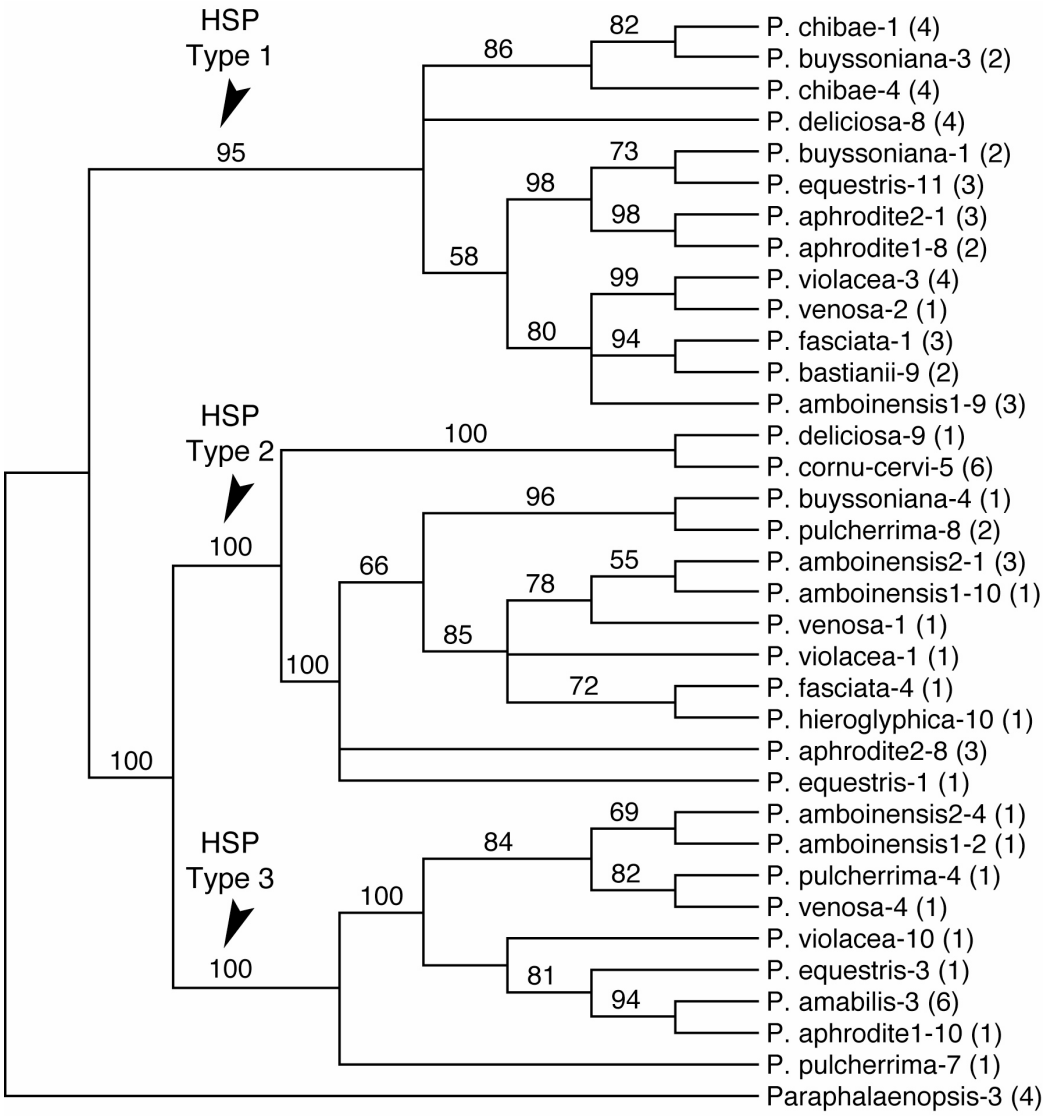


FIGURE 3.6. Maximum parsimony majority rule consensus tree reconstructed from primers 57F/57R, three genes for heat shock proteins. Numbers above the branches indicate nonparametric bootstrap support. Three highly supported clades (indicated by arrows) represent the three sequence types. The sequence from a single clone from an individual was randomly retained for analysis when cloned sequences from the individual formed a monophyletic group. The number of clones sequenced for each individual is indicated in parentheses.

Primer pair 63F/63R amplified a region that is about 1000 base pairs in length and that encoded the metabolic enzyme 6-phosphogluconate dehydrogenase. Although it was apparent from BLAST searches that there were introns present, it was unclear exactly where they occurred because in different taxa they occurred in different locations (GenBank accession numbers U18239, AF061837). Based on the phylogenetic tree (FIGURE 3.7), sequences clustered into two clades, but within a clade sequences from a single individual occurred in several places. It is unclear what evolutionary process is at work, but it is not likely this marker will be useful for phylogenetic investigations.

Primer pair 86F/86R appears to be a useful phylogenetic marker in *Phalaenopsis*. It amplified a region that encoded methionine synthase for all sequenced clones and individuals. The amplicon was only 500 base pairs in length with an intron between positions 227-312 and a stop codon at position 360. Most of the species that were sequence possessed two sequence types, and the variation between the different types occurred primarily within the intron or after the stop codon (GenBank accession number DP000011 type 1, AF439723 type 2) (FIGURE 3.8). Non-coding regions had informative sequence variation, and type 1 and type 2 sequences formed unambiguous monophyletic groups. Type 2 sequences all shared a five base pair deletion at position 230, and insertions at positions 370 (14 or 15 base pairs), 425 (13 base pairs), and 450 (15 base pairs), which make this type easy to identify and eliminate from the alignment.

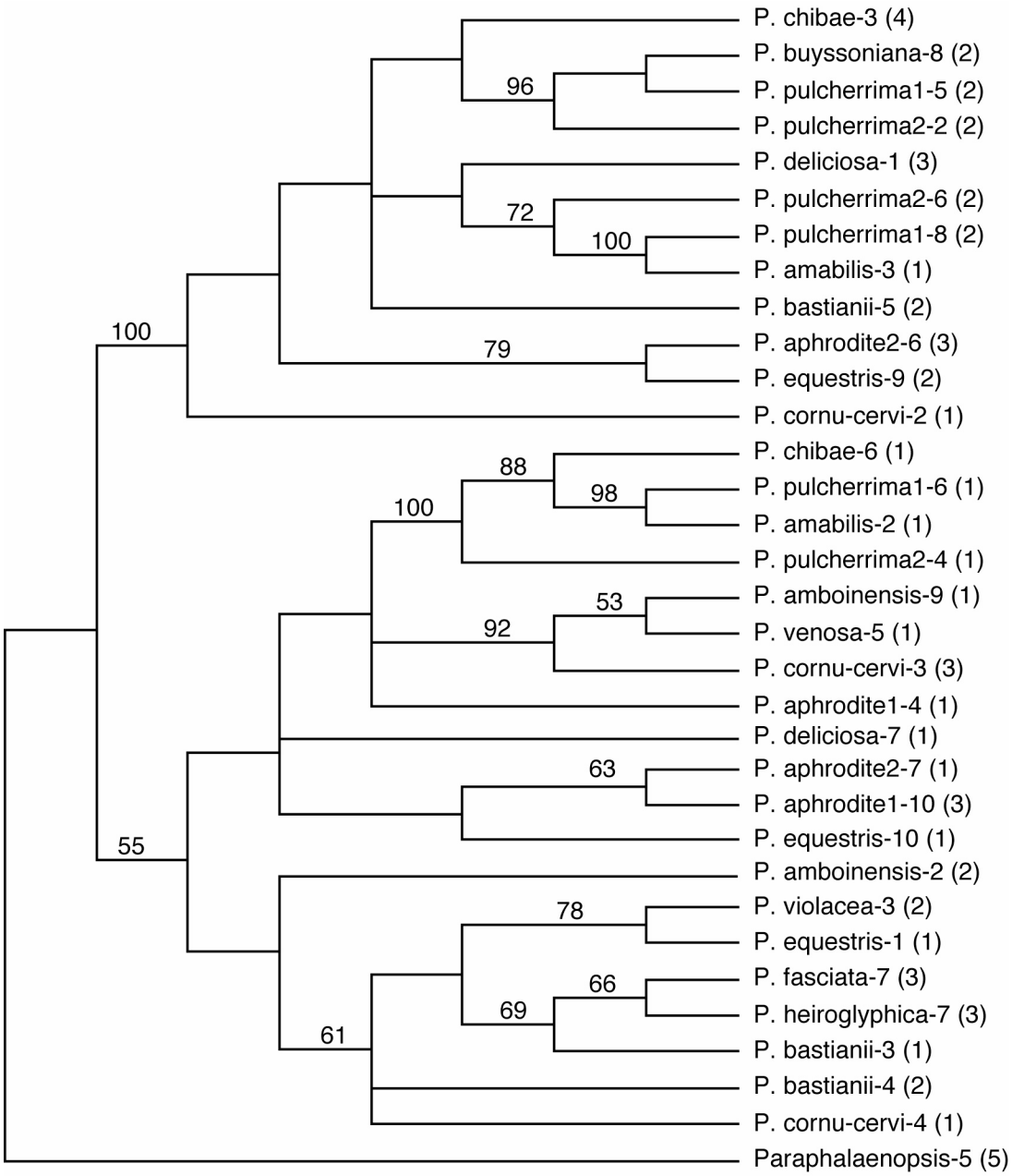


FIGURE 3.7. Maximum parsimony majority rule consensus tree reconstructed from the 6-phosphogluconate dehydrogenase gene sequenced from primers 63F/63R. Numbers above the branches indicate nonparametric bootstrap support. The sequence from a single clone from an individual was randomly retained for analysis when cloned sequences from the individual formed a monophyletic group. The number of clones sequenced for each individual is indicated in parentheses.

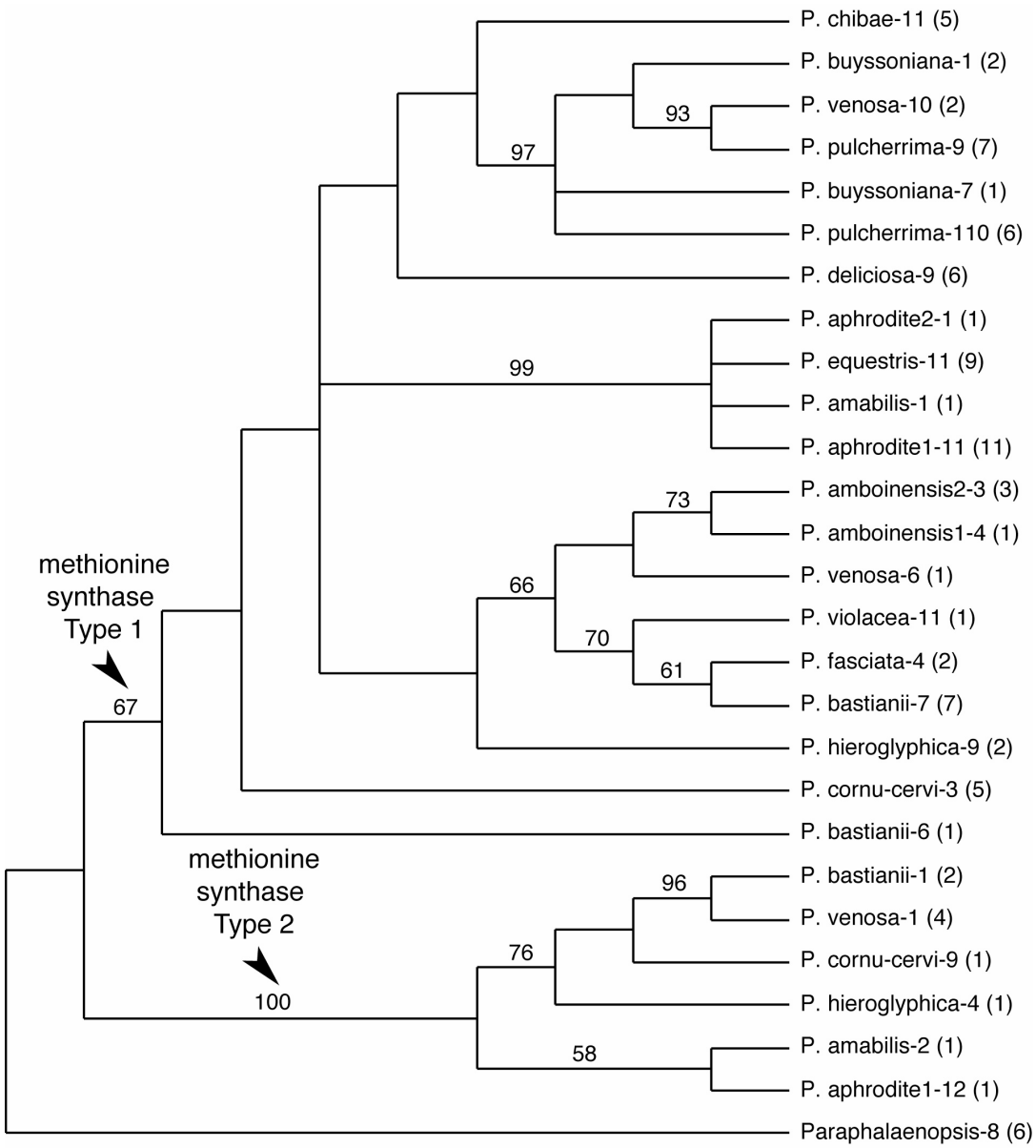


FIGURE 3.8. Maximum parsimony majority rule consensus tree reconstructed from marker 86F/86R, a methionine synthase gene. Numbers above the branches indicate nonparametric bootstrap support. Sequence types 1 and 2 are indicated by the arrows ➤. The sequence from a single clone from an individual was randomly retained for analysis when cloned sequences from the individual formed a monophyletic group. The number of clones sequenced for each individual is indicated in parentheses.

Helianthus

Ten markers were screened for phylogenetic utility in *Helianthus* (TABLE 3.5). Four of these ten also amplified in *Phalaenopsis* (51F/51R, 57F/57R, 86F/86R, 91F91R). However, after further sequencing, only one marker was judged phylogenetically useful in *Helianthus*. Primer pair 85F/85R amplified two different actin types. BLAST results from the TIGR database were used to identify these sequences: the *Helianthus* actin type 1 blasted to NM112046 (actin 11 in *Arabidopsis thaliana*), *Helianthus* actin type 2 blasted to a non-specific actin gene, AF282624. The types were of slightly different lengths (type 1: ~1200 and type 2: ~1300 bp) so after amplification they were distinguishable by gel electrophoresis. Phylogenetic analysis of the actin type 1 indicated that it included either multiple alleles or duplicate genes. It is likely that more than two loci are involved because more than two versions of the actin type 1 were found in some diploid individuals. Actin type 2, in contrast, appears to be a single-copy gene in *Helianthus* (FIGURE 3.9).

36F/36R and 91F/91R did not possess enough variation to resolve specific relationships within *Helianthus*. Although primer pairs 1F/1R, 2F/2R, 30F/30R, 51F/51R, 50F/57R, 61F/61R and 86F/86R each appeared to amplify a single band as seen by gel electrophoresis, sequencing of clones yielded many sequence types, making data collection impractical. In many cases sequences were completely unalignable or phylogenetic trees showed many paralogous types (FIGURE 3.10). TABLE 3.5 illustrates the large number of types found in these markers.

TABLE 3.5. Sequence variation within individuals of *Helianthus,* including the number of sequences examined for each species (Seqs/Sp), and the number of sequence types found for each species (Types/Sp), as well as pairwise distance statistics between different sequence types within each species and between species.

	Seqs/ Sp.	Types/ Sp.	Pairwise distance statistics: Within species Range	Within species Mean	Within species Median	Among species
1F/1R						**Range**
Helianthus annuus	11	4	0.0066 – 0.6012	0.3668	0.3652	0 – 0.6049
H. bolanderi	9	3	0.0022 – 0.5626	0.3423	0.4994	**Mean**
H. cusickii	12	4	0.0011 – 0.6026	0.3641	0.4842	0.3488
H. giganteus	5	3	0.1013 – 0.5845	0.4648	0.5276	**Median**
H. glaucophyllus	9	4	0 – 0.5650	0.2394	0.0777	0.4827
2F/2R						**Range**
H. annuus	6	1	0.0634 – 0.3061	0.1515	0.0836	0.0381 – 0.4447
H. bolanderi	7	4	0.0381 – 0.4272	0.3519	0.4054	**Mean**
H. cusickii	5	3	0.1479 – 0.4414	0.2772	0.1784	0.3451
H. giganteus	4	3	0.1531 – 0.4259	0.3315	0.3527	**Median**
H. glaucophyllus	0	0				0.4054
Phoebanthus	1	1				
30F/30R						**Range**
H. annuus	11	4	0 - 0.7935	0.3149	0.0549	0 – 0.7954
H. bolanderi	17	5	0 - 0.7899	0.2827	0.2907	**Mean**
H. cusickii	7	3	0 - 0.7826	0.2936	0.2668	0.3319
H. giganteus	0	0				**Median**
H. glaucophyllus	0	0				0.2852
36F/36R						**Range**
H. annuus	9	2	0.0012 - 0.0379	0.0189	0.0220	0 – 0.0406
H. bolanderi	9	1	0 - 0.0276	0.0071	0.0025	**Mean**
H. cusickii	7	1	0 - 0.0279	0.0095	0.0038	0.0184
H. giganteus	1	1				**Median**
H. glaucophyllus	6	1	0.0012 - 0.0254	0.0128	0.0091	0.0187
51F/51R						**Range**
H. annuus	4	2	0 – 0.303	0.1924	0.2867	0 – 0.3229
H. bolanderi	11	3	0 – 0.2393	0.0423	0.0035	**Mean**
H. cusickii	12	4	0 – 0.3229	0.1252	0.1125	0.1740
H. giganteus	5	3	0 – 0.3135	0.1912	0.2886	**Median**
H. glaucophyllus	12	4	0 – 0.2554	0.1079	0.0069	0.2242
50F/57R						**Range**
H. annuus	7	4	0.0081 – 0.2412	0.1718	0.2172	0.0009 – 0.3469
H. bolanderi	8	3	0.0009 – 0.2525	0.1580	0.2148	**Mean**
H. cusickii	10	3	0.0009 – 0.2907	0.1239	0.1376	0.1553
H. giganteus	10	4	0.0047 – 0.2482	0.1321	0.1612	**Median**
H. glaucophyllus	8	3	0.0019 – 0.2499	0.1583	0.2236	0.1849

TABLE 3.5. Continued

	Seqs/ Sp.	Types/ Sp.	Pairwise distance statistics			
			Within species			Among species
			Range	Mean	Median	
61F/61R						**Range**
H. annuus	10	3	0 – 0.2740	0.1035	0.0189	0 – 0.6108
H. bolanderi	11	3	0.0009 – 0.0187	0.0112	0.0122	**Mean**
H. cusickii	10	2	0.0009 – 0.5819	0.1216	0.0122	0.0920
H. giganteus	10	2	0 – 0.5934	0.1266	0.0141	**Median**
H. glaucophyllus	11	1	0.0028 – 0.2698	0.0939	0.0235	0.0182
85F/85R						**Range**
H. annuus	2	1	0.00762 - 0.0076	0.0076	0.0076	0.0024 – 0.2518
H. bolanderi	4	2	0.0033 - 0.2406	0.1209	0.1179	**Mean**
H. cusickii	3	1	0.0024 - 0.0040	0.0032	0.0032	0.1414
H. giganteus	1	1				**Median**
H. glaucophyllus	0	0				0.2284
86F/86R						**Range**
H. annuus	6	3	0.0022 – 0.1820	0.0940	0.1184	0 – 0.3376
H. bolanderi	5	4	0.0841 – 0.2632	0.2130	0.2328	**Mean**
H. cusickii	5	2	0.0062 – 0.2224	0.1333	0.2007	0.2353
H. giganteus	2	2		0.2912	0.2912	**Median**
H. glaucophyllus	12	5	0 – 0.3262	0.2340	0.2857	0.2761
Phoebanthus	3	2	0.0512 – 0.6132	0.2044	0.2762	
91F/91R						**Range**
H. annuus	9	3	0.0011 – 0.2141	0.1076	0.1092	0.0011 – 0.2141
H. bolanderi	3	1	0.0048 – 0.0133	0.0097	0.0108	**Mean**
H. cusickii	5	2	0.0055 – 0.1984	0.1181	0.1806	0.1137
H. giganteus	8	2	0.0023 – 0.2003	0.1126	0.1854	**Median**
H. glaucophyllus	3	1	0.0034 – 0.0045	0.0038	0.0034	0.1844

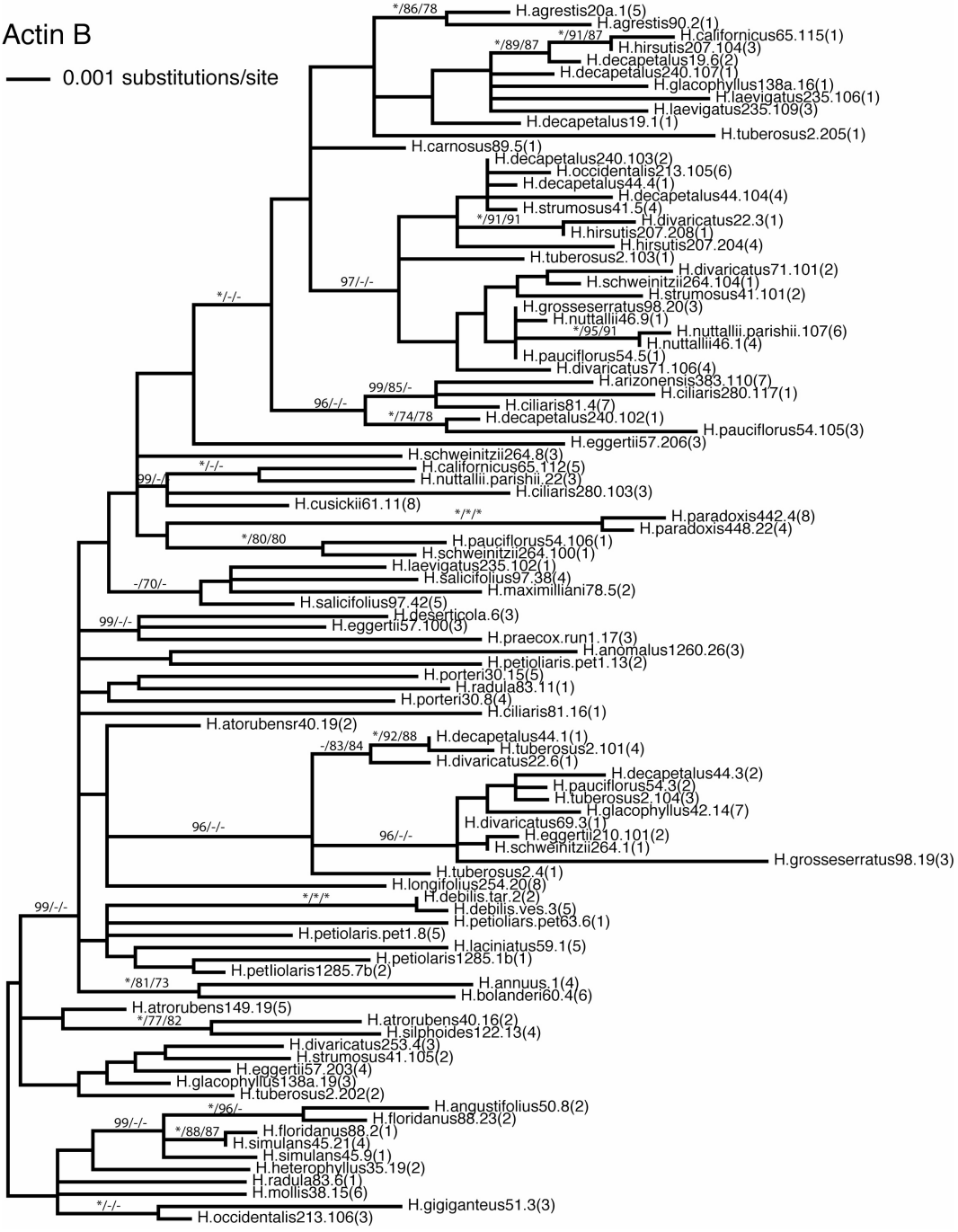


FIGURE 3.9. Maximum parsimony phylogram reconstructed from the *Helianthus* actin 2 gene sequenced using primers 85F/85R. Support values above the branches indicate posterior probabilities from MrBayes/ maximum likelihood scores/ maximum parsimony nonparametric bootstrap values. *Helianthus* species names are directly followed by their voucher number; the number after the period is the clone number, and the number in parentheses is the number of clones that were sequenced. *= 100% support; - = support level below the significance threshold. Significance thresholds are 95% for posterior probabilities and 70% for nonparametric bootstraps.

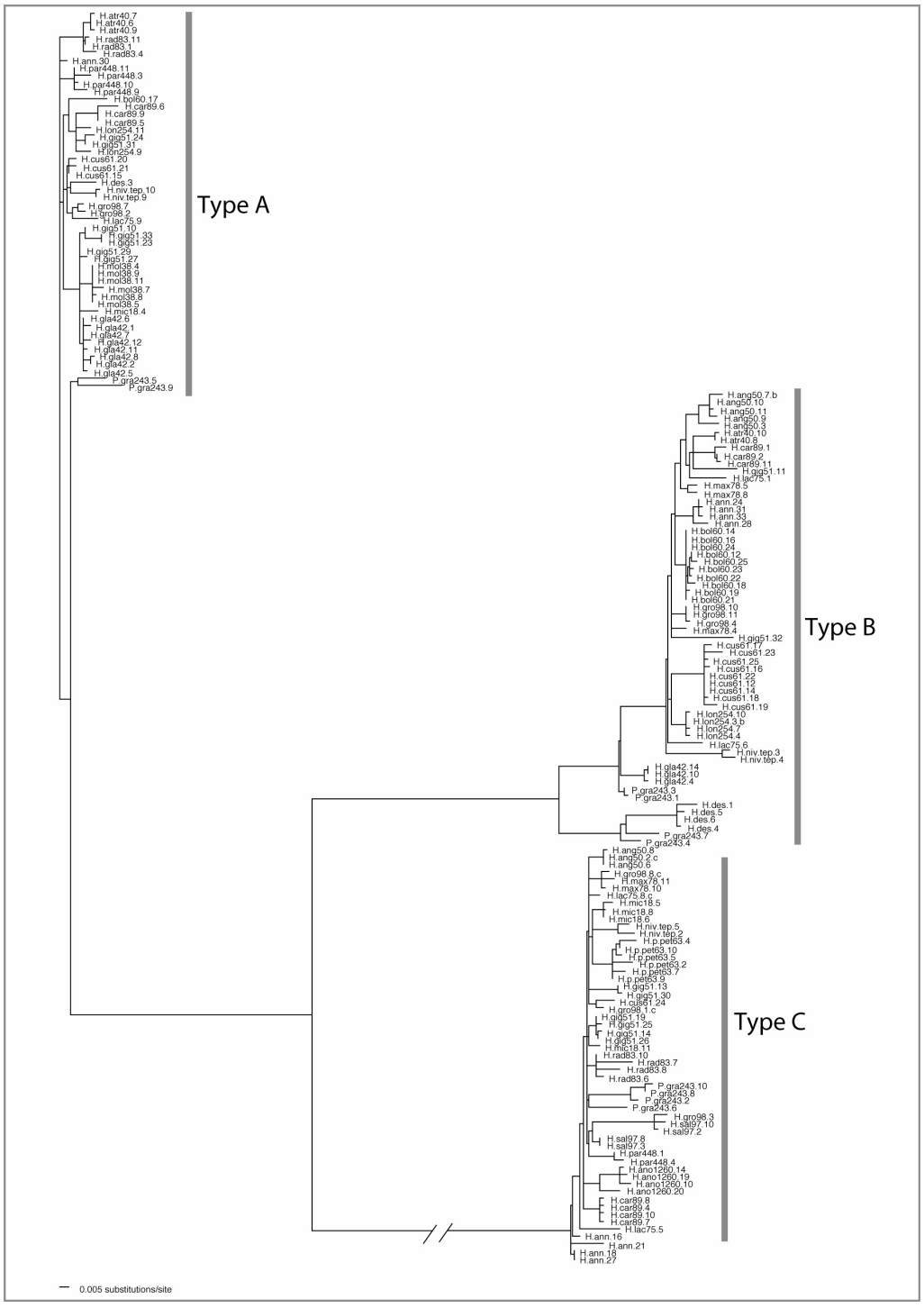


FIGURE 3.10. Maximum likelihood tree of *Helianthus* from the ATP synthase gene sequenced from primers 51F/51R. Three sequence types (A, B, and C) were amplified, and there is evidence for paralogy within each type.

DISCUSSION

The primary aim of this project was, if possible, to identify the largest set of promising single-copy nuclear markers for phylogenetic reconstruction that might be universally amplifiable throughout the angiosperms. We felt that such markers would be generally useful to the systematics community and especially useful for reconstructing relationships in groups that have hybrid species (Linder and Rieseberg 2004; Nakhleh et al. 2004). We reasoned that the most efficient means of finding a comprehensive set of informative nuclear regions would be via whole nuclear genome comparisons of phylogenetically distant flowering plant species. Computationally, we were able to identify a large number of conserved, appropriately spaced, complex primer pairs that occurred only a single time in both *A. thaliana* and *O. sativa* (Xu et al. 2004). However, laboratory tests of 142 of the most promising primer combinations based on computational criteria, revealed limitations in our computational approach due to large differences in the types and rates of evolution in different clades of the flowering plants.

In the absence of extensive optimization trials for PCR, only a relatively small proportion of the putative primer combinations successfully amplified throughout our exemplar angiosperm species. Even more surprising, only 35% of the primer combinations amplified our *A. thaliana* positive control. This low rate of amplification in our positive control could be due to several factors. First, primer design may have been suboptimal in some cases. This is probably not a large issue because primer sequences (1) came directly from the *Arabidopsis* genome, (2) were of a reasonable length, (3) maximized the GC content in the primers, and (4) were screened for primer dimers and hairpin loops. Alternatively, the secondary structure of the genomic DNA may have interfered with amplification. Finally, our PCR reaction conditions may have been suboptimal. Had we tried more PCR amplification conditions we might have been more successful; however, the large number of primer pairs that we tested made optimization

for individual primer pairs impossible. Therefore, some of our "failed" primer combinations might be shown to be usable with additional effort to tune amplification conditions.

Although we did not find primer pairs that would amplify in all of our test taxa, we did find four phylogenetically informative regions for *Phalaenopsis* and one in *Helianthus*. For these two genera, it has been difficult to find DNA regions with suitable variation in commonly used markers of either the plastids (e.g. *mat*K or *rps*16) or the nuclear genome (e.g. ITS). Our discovery of new markers indicates that our approach for finding markers has some merit and could be extended to other groups in which new markers for phylogenetic analysis are needed. Even so, significant effort has to be expended to find a small set of appropriate regions.

Toward this end we have posted the full set of primer combinations in an online database, "The Conserved Primer Pair Project," which is available online at http://aug.csres.utexas.edu:8080/cpp/index.html. "The Conserved Primer Pair Project" (CPP) database incorporates a set of four tools that allow the user to browse the approximately 13,000 primer pairs and input amplification information into the database. These tools include CPPQuery, which is a form-based interface that lets the user search the database for primer pairs that can be used to amplify DNA in the lab. The CPPBrowser is a graphical interface where the user can browse the database along the chromosomes of *Arabidopsis* and rice. CPPInput allows the user to input amplification information into the database for primers that have been screened in the lab. Finally, the CPPWizard program is a direct SQL front-end client to the primer pair database.

This comparative study shows that it is not only difficult to find regions that will amplify broadly throughout the angiosperms, but that, at least in certain groups, we need to be careful to test for orthology of amplified regions. While *Phalaenopsis* had an amplification success rate of 20%, half of these successes included obvious amplification of multiple bands using gel electrophoresis. Furthermore, a portion of those that

appeared to amplify a single band was actually amplifying several sequence products of similar length. These findings suggest that a large portion of the *Phalaenopsis* genome may have undergone duplication events or that many of the loci are single copy but very likely to be heterozygous. Since it is important that only orthologous gene products are used in phylogenetic reconstruction, all new regions will have to be evaluated carefully using at least the phylogenetic methods that we employed.

The 57F/57R primer combination in the case of *Phalaenopsis* was particularly useful since each of the three types was readily distinguishable. A researcher could have as many as three independent markers for reconstruction since there was no evidence in the sequences that concerted evolution was homogenizing them. In the case of *Helianthus*, this DNA region was less informative because there were so many types that a clearly interpretable phylogeny of the types could not be reconstructed.

In short, our study suggests there may be a very limited set of universally amplifiable single-copy orthologous nuclear regions in the angiosperms because the nuclear genome in angiosperms appears to evolve so rapidly in most cases that there are few sequences that will amplify a single type over a broad range of taxa. Nonetheless, because of the large number of putative primer combinations that we identified and our experience with searching for regions in *Helianthus* and *Phalaenopsis*, there are likely combinations of primers in our database that are useful regions for phylogenetic analysis in other groups. There may also be some "universal" primers in our sets that we did not try in the lab. In addition, as more plant whole genome sequences become available our analysis can be repeated using those genomes to trim our putative primer set to a smaller number, which can be assessed in the lab. In its current form, our data can be used to attempt to discover regions that work at whatever taxonomic levels a particular researcher is working.

An important consequence of the lack of a large set of universal regions for phylogenetic reconstruction in angiosperms, is that it makes the development of high

quality supertree methods much more important. Since the angiosperm tree of life will apparently require different genes for different groups of plants, systematists will necessarily need phylogenetic reconstruction methods that can take individual reconstructions from differing DNA regions and assemble them into a larger tree or network.

The tree of life project is an enormous phylogenetic challenge, and will require the use of many different DNA regions. We are hopeful our database will become a tool from which phylogenetically informative genetic markers can be retrieved and into which results from other researchers can be catalogued throughout the breadth of the angiosperms and ultimately all taxonomic groups. It will become ever more powerful as more researchers use it and contribute to it, and it becomes more obvious which primers amplify successfully in different groups of plants.

LITERATURE CITED

BAILEY, C.D., T.G. CARR, S.A. HARRIS, and C.E. HUGHES. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics & Evolution 29: 435-455.

BALDWIN, B.G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Molecular Phylogenetics and Evolution 1: 3-16.

BENSON, D.A., I. KARSCH-MIZRACHI, D.J. LIPMAN, J. OSTELL, and D.L. WHEELER. 2004. GenBank: update. Nucleic Acids Research 32: D23-D26.

DOYLE, J.J. and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19: 11-15.

FUTUYMA, D.J. 1997. Evolutionary Biology. Sinauer Assoc. Inc., Sunderland, MA.

HILLIS, D.M., C. MORITZ, and B.K. MABLE. 1996. Molecular Systematics. Sinauer Associates, Inc., Sunderland, MA.

LANDAIS, I., M. OGLIASTRO, K. MITA, J. NOHATA, M. LOPEZ-FERBER, M. DUONOR-CERUTTI, T. SHIMADA, P. FOUMIER, and G. DEVAUCHELLE. 2003. Annotation pattern of ESTs from *Spodoptera frugiperda* Sf9 cells and analysis of the ribosoma protein genes reveal insect-specific features and unexpectedly low codon usage bias. Bioinformatics 19: 2243-2250.

LINDER, C.R. and L.H. RIESEBERG. 2004. Reconstructing patterns of reticulate evolution in plants. American Journal of Botany 91: 1700-1708.

MADDISON, W.P. and D.R. MADDISON. 2000. MacClade 4.0. Sinauer, Sunderland, Massachusetts.

MIRANKER, D., W. XU, and R. MAO. 2003. MoBIoS: a metric-space DBMS to support biological diversity. Proceedings of the International Conference on Scientific and statistical Database Management Systems.

MULLIS, K.B. and F.A. FALOONA. 1987. Specific synthesis of DNA in vitro via polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335-350.

NAKHLEH, L., T. WARNOW, C.R. LINDER, and K. ST. JOHN. 2004. Reconstructing Reticulate Evolution in Species - Theory and Practive. Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology: 337-346.

ORTI, G., D.E. PEARSE, and J.C. AVISE. 1997. Phylogenetic assessment of length variation at a microsatellte locus. Proceedings of the National Academy of Sciences of the United States of America 94: 10745-10749.

SOLTIS, P.S., D.E. SOLTIS, and J.J. DOYLE. 1998. Molecular Systematics of Plants II: DNA Sequencing. Chapman and Hall, New York.

SANG, T. 2002. Utiliy of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology 37: 121-147.

SHAW, J., E.B. LICKEY, J.T. BECK, S.B. FARMER, W. LIU, J. MILLER, K.C. SIRIPUN, C.T. WINDER, E.E. SCHILLING, and R.L. SMALL. 2005. The Tortoise and the Hare II: Relative Utility of 21 Noncoding Chloroplast DNA Sequences for Phylogenetic Analysis. American Journal of Botany 92: 142-166.

STEPHENS, P.F. 2005. Angiosperm Phylogeny Website. Version 6. http://www.mobot.org/MOBOT/research/APweb/.

STREELMAN, J.T., R. ZARDOYA, A. MEYER, and S.A. KARL. 1998. Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): Evolutionary comparison of microsatellite and single-copy nuclear loci. Molecular Biology And Evolution 15: 798-808.

SWOFFORD, D. 1996. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts.

THOMPSON, J.D., T.J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D.G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.

WELCH, T.A. 1984. A technique for high performance data compression. IEEE Computer 17: 8-19.

XU, W., W.J. BRIGGS, J. PADOLINA, R. TIMME, W. LIU, R.C. LINDER, and D. MIRANKER. 2004. Using MoBIoS' scalable genome join to find conserved primer pair candidates between two genomes. Bioinformatics 20 suppl. 1: i355-i362.

Chapter 4: Using *Phalaenopsis* Species and Horticultural Hybrids to Test Methods of Reconstructing Reticulate Evolution in Phylogenetic Analyses

INTRODUCTION

As so eloquently stated by Theodosius Dobzhansky, "Nothing in biology makes sense except in the light of evolution." A key aspect of evolution is that it is a historical process of ancestor-descendant relationships recorded in the genetics and phenotypes of organisms. Phylogenetic trees are diagrams that represent the evolutionary history of a set of taxa. They are so essential to evolutionary biology that a hypothetical phylogeny is the only illustration in Darwin's "The Origin of Species." Appropriately, phylogenies have become an indispensable tool for researchers in many biological fields, including evolution, ecology, molecular biology, conservation biology, pharmaceutics, and medicine. They are the frameworks on which many hypotheses are based, and accurate conclusions depend on the accuracy of the phylogeny. However, many groups of organisms do not evolve in the tree-like fashion to which current phylogenetic reconstruction methods are limited (Otto and Whitton 2000; Snel et al. 2002).

Hybridization is an important evolutionary mechanism that has led to speciation in many groups of organisms (Rieseberg 1997; Levin 2000). It has been estimated that between 30 and 80% of extant angiosperm species have undergone at least one hybridization event within their evolutionary history (Grant 1981; Funk 1985; Masterson 1994), and that ~2-4% of angiosperm speciation events are associated with polyploidy (Otto and Whitton 2000). Furthermore, reticulate evolution, which also includes horizontal gene transfer and introgression, is widespread in other groups ranging across amphibians, fish, invertebrates, fungi, and prokaryotes (Evans et al. 2001; Loytynoja and Milinkovitch 2001; Tarkhnishvili et al. 2001; Snel et al. 2002). Although most

researchers agree that hybrid speciation is not a rare phenomenon, there are currently limited computational methods with which to infer hybridization in a phylogenetic context (Funk 1985; Rieseberg and Ellstrand 1993; Rieseberg 1997; Posada and Crandall 2001; Linder and Rieseberg 2004; Nakhleh et al. 2004). Current methods for reconstructing phylogenetic relationships, such as maximum parsimony and maximum likelihood, yield bifurcating trees. They have been extensively developed but cannot discern more complex evolutionary events (Hennig 1966; McDade 1990, 1992; Swofford 1996a).

Hybridization and Speciation

We will use the term hybridization to mean cross-fertilization between two distinct lineages (species), which may result in the formation of a new lineage. For our purposes, we do not include mating between unlike types of the same species, another context in which the term hybridization is often used. Often, when two species hybridize, the offspring are not fertile, and no new lineage is produced. Other times, fertile offspring are formed, but they backcross with one or both of their parents. Although a dynamic hybrid zone might be persistent under these circumstances, no new lineage is formed that inhabits a unique environment and is isolated from both parents. These sorts of hybrid zones can be an important means of introducing genetic diversity into populations and may have significant consequences for phylogenetic reconstruction.

Hybridization and Phylogenetic Reconstruction

Several observations have been made regarding the effect of including hybrids in phylogenetic analyses. McDade (1990, 1992) conducted an extensive study to determine

the impact of including hybrids in a traditional phylogenetic tree reconstruction using parsimony. She found that most often a hybrid was placed sister to the clade that included its most derived parent (i.e., the parent that evolved most recently on the phylogeny). Inclusion of the hybrid did not significantly alter the topological structure of the tree, unless its parents were distantly related. Her work illustrated that phylogenetic methods designed to return trees could not distinguish hybrids from normal taxa. Other studies have found that inclusion of hybrids in phylogenetic analyses produces a very large number of different phylogenies each of which appears to be an equally good hypothesis (Hein 1990, 1993). It has also been noticed that hybrid taxa may produce incongruent gene trees when multiple markers, most commonly a plastid and a nuclear marker, are examined (Maddison 1997; Linder and Rieseberg 2004).

Approaches to Reconstructing Reticulate Evolution

Historically, hybridization events have been inferred using a variety of data, including morphological comparisons, ecological and geographical data, secondary chemistry, karyology, allozyme electrophoresis, restriction site data, and breeding and linkage studies involving the creation of artificial hybrids that closely resemble the natural hybrids (Grant 1966; Gallez and Gottlieb 1982; Crawford 1985; Rieseberg 1991, 1997). Whereas many of these data types have contributed strong evidence for the hybrid origin of the species examined, they are often unable to predict hybrid species that have not been previously singled out as a potential hybrid.

More recent methods for detecting hybridization and reconstructing species networks use molecular sequence data. Two general approaches have been used, combined data approaches and incongruence approaches (Linder and Rieseberg 2004).

Combined data approaches input data from multiple, independent genes or loci and look for mixed signals, e.g., non-additivity in the distance matrix. Several methods of this type have been developed that attempt to detect and reconstruct hybrid evolution (Sattath and Tversky 1977; Huson 1998; Bandelt et al. 1999; Xu 2000; Bryant and Moulton 2002), however simulation studies of them suggest they have a high rate of false positives, making them unsatisfactory for practical applications (Linder and Rieseberg 2004; Nakhleh et al. 2004).

Within the combined data approaches, three general methods have been proposed (Linder and Rieseberg 2004; Morrison 2005). First, there are methods that infer a single optimal phylogenetic tree and then add reticulations to the tree in order to optimize certain criteria. Some examples of these methods include reticulograms (Legendre and Makarenkov 2002; Makarenkov and Legendre 2004) and statistical parsimony (Templeton et al. 1992). The second sets of methods infer a set of optimal phylogenetic trees and then attempt to reconcile them using reticulations. Examples of these methods include median networks (Bandelt et al. 1999; Bandelt et al. 2000) and molecular variance parsimony (Excoffier and Smouse 1994). Finally, there are methods that compute splits. These methods identify incompatibilities in the data without explicitly inferring a tree. Instead, they provide a collection of possible resolutions displayed using reticulations and leave the user to determine the causes of the incompatibilities (Bandelt and Dress 1992; Huson 1998; Bryant and Moulton 2002, 2004). All of these methods are analogous to either parsimony- or distance-based tree-building methods and cannot incorporate the complexity of model-based methods, such as maximum likelihood (Morrison 2005). The lack of biological rationale behind these methods is likely to contribute to their shortcomings towards reconstructing hybridization (Linder and Rieseberg 2004).

SplitsTree (Huson 1998; Huson and Bryant 2006) is the most widely used and powerful combined data software package available for network reconstruction. It features several network reconstruction methods based on splits, including Neighbor-Net. As discussed in the previous section, Neighbor-Net computes a set of incompatible splits based on the data in the form of a distance matrix. Huson and Bryant (2006) emphasize that while a phylogenetic tree has a direct evolutionary interpretation (leaves are taxa and branches are speciation events) a splits network does not have such a biological interpretation. Rather, a splits network is an abstract visualization of incompatible splits. The source of incompatibilities may represent reticulation events or other biological processes; alternatively, they may represent homoplasy or simply noisy data. Recent studies (Huson et al. 2005) have shown that there are strong similarities between splits networks and true phylogenetic networks.

Incongruence approaches build what have been referred to as 'true phylogenetic networks' (Morrison 2005), because they attempt to recover the underlying species network by combining the various gene trees contained within it. This is in contrast to the combined data approaches, which build 'character-display networks' that show character conflict, not all of which may be due to reticulate evolution. The objective of the incongruence approach is to find the optimal network with the minimum number of reticulation events necessary to resolve the incongruence between a set of gene trees (Hein 1990; Maddison 1997; Nakhleh et al. 2004). Specifically, a gene tree is inferred for each available dataset; if the gene trees are identical, that is the species tree; if the gene trees are not identical, the minimum network that contains all of the trees is found. Maddison (1997) showed how to do this when the minimum network contained a single reticulation, and Nakhleh et al. (2004) and Huynh et al. (2005) have constructed algorithms that implement this approach.

In their method SpNet, Nakhleh et al. (2004) proposed polynomial time algorithms to infer 'galled networks,' phylogenetic networks where reticulation events are constrained to be evolutionarily independent of each other, which can be used even when topological errors are present in the individual gene trees. This method exploits the fact that for each reticulation event, such as the simple case illustrated in FIGURE 4.1, there are exactly two trees contained within the network. Every gene will evolve down one of the two trees, and the two trees will differ from each other in only one sub-tree (Hein 1990, 1993; Nakhleh et al. 2004). In other words, to reconcile the two trees one branch along with all of its descendent branches is moved to a new location. Biologically, this represents the transfer of genetic material, and mathematically, this represents one rooted-subtree prune-and-regraft (rSPR) operation (Nakhleh et al. 2004). In FIGURE 4.1, the branch of the gene tree containing species B can be pruned and regrafted to recover the alternate gene tree. As the number of independent reticulations increases, the number of trees contained within the network increases exponentially, such that there will be 2^m possible trees contained in a network with m reticulations (Nakhleh et al. 2004). SpNet is not able to reconstruct accurately phylogenetic networks (galled or not) with more than one reticulation. A modified version of SpNet, RIATA, is currently under development. RIATA will be able to reconstruct networks with multiple reticulations (Nakhleh et al. pers com.). Huynh et al (2005) proposed an algorithm that can construct a phylogenetic network with more than one hybrid node in their methods RGNet. They do this by expanding on the algorithms presented by Nakhleh et al. (2004) and allowing more than two input trees to be analyzed. As of April 2006, neither SpNet, RIATA, nor RGNet has been implemented in programs for general use.

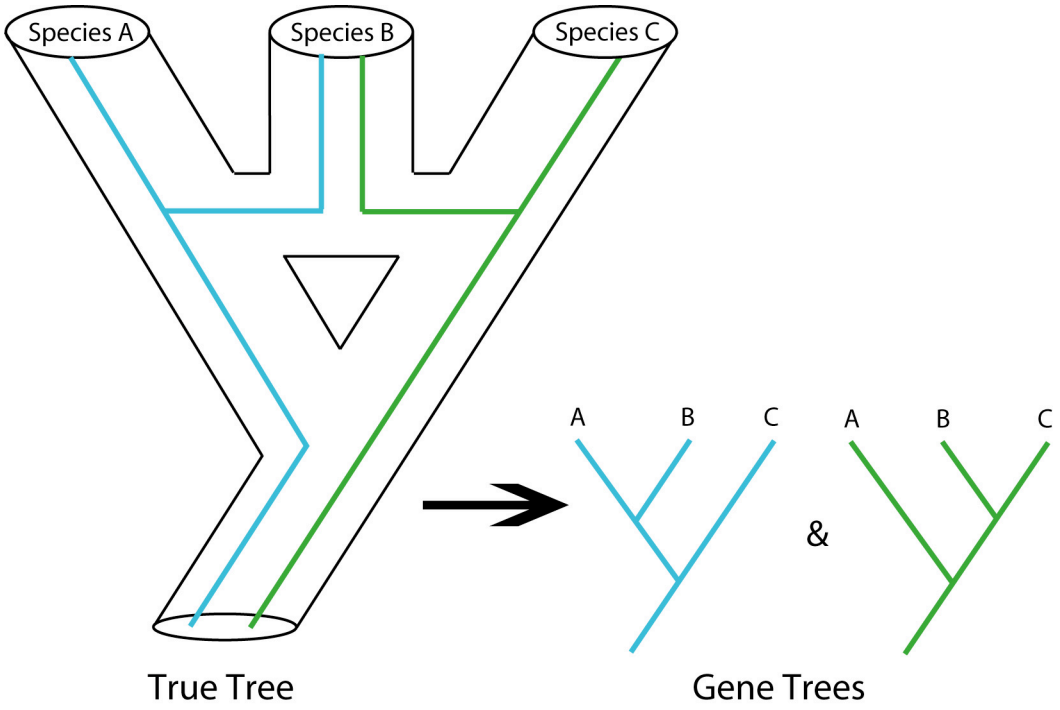


FIGURE 4.1. Species network containing gene trees (in color) and the phylogenies that would be hypothesized from analyzing the genes. Hybridization between species A and C gives rise to species B. Species B possesses copies of both genes. A different topology will be recovered depending on which gene is examined.

Testing Methods for Reconstructing Reticulate Evolution

As more methods for detecting and reconstructing reticulate evolution become available, biological, as well as simulated, datasets are needed to evaluate the strengths and weaknesses of these methods. Nakhleh et al. (2003) have developed tools capable of generating random networks and simulating DNA sequence evolution down networks based on biological criteria. This is important for creating simulated data to test the accuracy of network reconstruction methods. In addition, Linder et al. (2003) have developed an error metric to assess phylogenetic networks based on the Robinson-Foulds

(RF) measure (Robbinson and Foulds 1981) that is commonly used to assess phylogenetic trees. The tripartition metric is useful because it is accurate whether or not hybrid taxa are present, and it conforms to the standard RF measure when there are no hybrid taxa.

Simulation studies are extremely useful, but they cannot always accurately model the complexities of real biological data. In order to evaluate the reliability of these methods further, we have developed a dataset consisting of a combined chloroplast DNA dataset and four independent nuclear DNA datasets from *Phalaenopsis* orchid species and horticultural hybrids. Because the parentage of the hybrids is known, we have constructed several types of phylogenetic networks for which we know the true network. In this chapter, we use these known networks to assess the ability of network reconstruction programs to infer the correct networks using biological data. This chapter examines the results obtained from Neighbor-Net (implemented in SplitsTree) (Huson 1998; Bryant and Moulton 2004; Huson and Bryant 2006). Our data are ready to be implemented by RIATA and RGNet when they become available for use.

MATERIALS AND METHODS

Sampling

Twenty-seven individuals were included in this study (TABLE 4.1), consisting of thirteen *Phalaenopsis* species (with five replicates), one outgroup taxon (*Paraphalaenopsis laycockii*), and eight artificial *Phalaenopsis* hybrids of known parentage. All individuals included were obtained through horticultural sources, either private or commercial growers, or the collections of the Marie Selby Botanical Garden, the New York Botanical Garden, or the University of Florida. The identity of individuals

obtained through private or commercial growers was confirmed as they bloomed, and vouchers have been placed in The University of Texas Herbarium (TEX).

TABLE 4.1. Species and hybrids used to construct phylogenetic trees from combined chloroplast DNA data four nuclear DNA regions and their herbarium voucher information.

Species	**Herbarium Number**
Phalaenopsis amabilis	Padolina A007 (TEX)
P. amboinensis 1	Padolina A098 (TEX)
P. amboinensis 2	Padolina A099 (TEX)
P. amboinensis 3	Padolina A119 (TEX)
P. aphrodite 1	Padolina A008 (TEX)
P. aphrodite 2	Padolina A059 (TEX)
P. aphrodite 3	Padolina A117 (TEX)
P. bastianii	Padolina A028 (TEX)
P. buyssoniana (natural autotetraploid of *P. pulcherrima*)	Padolina A011 (TEX)
P. chibae	Padolina A001 (TEX)
P. deliciosa	Padolina A010 (TEX)
P. equestris	Padolina A014 (TEX)
P. fasciata	Padolina A017 (TEX)
P. hieroglyphica	Padolina A028 (TEX)
P. pulcherrima 1	Padolina A006 (TEX)
P. pulcherrima 2	Padolina A012 (TEX)
P. venosa	Padolina A045 (TEX)
P. violacea	Padolina A120 (TEX)
Paraphalaynopsis laycockii (outgroup)	Padolina A047 (TEX)
Hybrids	
P. amabilis x *P. amboinensis*	Padolina A021 (TEX)
P. buyssoniana x *P. equestris*	Padolina A025 (TEX)
P. equestris x *P. aphrodite*	Padolina A042 (TEX)
P. pulcherrima x *P. equestris*	Padolina A024 (TEX)
P. pulcherrima x *P. fasciata*	Padolina A022 (TEX)
P. pulcherrima x *P. violaceae*	Padolina A051 (TEX)
P. venosa x *P. equestris*	Padolina A023 (TEX)

DNA Extraction, PCR Amplification, and Cycle Sequencing

DNA extractions were carried out using the CTAB protocol of Doyle and Doyle (1987) or the Qiagen DNeasy Plant DNA Extraction Kit. Extracts were cleaned with the

Qiagen QIAEX II Suspension kit if necessary. Standard PCR protocols were used (Mullis and Faloona 1987), and 5% dimethyl sulfoxide (DMSO) was added to reactions. Amplifications were visualized on agarose gels with ethidium bromide and purified with QIAGEN QIAquick PCR purification kits or Sephadex columns.

All nuclear PCR products were cloned using Invitrogen's TOPO TA cloning kit to ensure single-copy sequencing product. Ten colonies were selected from each individual for PCR amplification and only products of the appropriate size were sequenced. Cycle sequencing reactions were performed using BigDye Terminator 3.0 and visualized on an MJ BaseStation.

Chloroplast Markers

Four regions from the chloroplast genome were sequenced for phylogenetic reconstruction. Two chloroplast regions were taken from the literature: *trn*D → *trn*E and *mat*K (Demesure et al. 1995; Whitten et al. 2000). We developed two other regions (see chapter 2 for novel primer sequences), *atp*H → *atp*F and *pet*B → *pet*D, from the completed maize chloroplast genome available on GenBank (accession number NC_001666) (Maier et al. 1995; Benson et al. 2004). No single region was sufficient to provide complete resolution of the *Phalaenopsis* phylogeny so we combined our data for a total of 3564 aligned base pairs. We will refer to this dataset as the 'combined chloroplast regions.'

Nuclear Markers

We developed four nuclear markers for this study by comparing the *Arabidopsis* and *Oryza* genomes and looking for conserved regions (see Chapter 3). The primer pair

6F/6R amplified a region that codes for an actin gene and is approximately 775 base pairs in length. Primer pair 51F/51R amplified a region that is an ATP synthase gene and is about 680 base pairs in length. Due to amplification and sequencing difficulties, several species produced no sequences, including *Phalaenopsis amabilis*, *P. deliciosa, P. venosa,* and *Paraphalaenopsis laycockii*. Primer pair 57F/57R amplified a heat shock protein (HSP) and spanned approximately 950 base pairs. Three sequence types were recovered, each of which formed distinct, monophyletic groups. Only HSP type 2 was used in our analyses because it possessed the most complete taxon sampling. Unfortunately, not all species yielded a type 2 HSP, so we are missing sequences of *P. amabilis, P. hieroglyphica,* and, *P. pulcherrima*. Primer pair 86F/86R amplified a region that codes for methionine synthase and is approximately 500 base pairs in length. A complete data set was obtained for this marker.

Analyses

Sequences were assembled and edited in Sequencher 4.2. Initial alignments were carried out using ClustalX (Thompson et al. 1997), and then edited by hand in MacClade 4.0 (Maddison and Maddison 2000). All four chloroplast markers were combined for all analyses; the nuclear markers were analyzed separately. Only one sequence type was retained for those markers where more than one type was recovered as was determined when a phylogenetic tree of all sequences was reconstructed. When multiple sequences were present for an individual within the retained sequence type, one sequence was randomly selected to be used in analyses using a random number generator (Haahr 2006). Parsimony analyses were run in PAUP* 4.0 (Swofford 1996b) using a heuristic search with ten random-addition-replicates and TBR branch swapping. Bootstrap analyses were

performed to determine branch support each with 1000 replicates and 1000 trees saved from each replicate.

Neighbor-Net (implemented in SplitsTree) was used to generate splits networks using concatenated sequence data as input. Distances were calculated using uncorrected p scores (Hamming distances) and least squares estimation was used to calculate edge lengths. Splits networks were visually analyzed to determine the number of false positives and the number of tests where the parents of hybrids were incorrectly identified for each data set. TABLE 4.2 lists the data sets that were evaluated. Seven tests were conducted adding a single hybrid to the fourteen non-hybrid taxa, ten tests were conducted including two hybrids that were independent of each other (they did not share a parent), four tests were conducted with three hybrids that were all independent of each other, and six tests were conducted including two hybrids that were not independent of each other (each hybrid had one parent in common with the other). Six of the seven hybrids' parents did not occur within the same major clade (i.e. one parent was in one clade and the other parent was in another, FIGURE 4.2) and one hybrid (*Phalaenopsis equestris* X *P. aphrodite*) possessed parents that occurred in the same clade with each other.

Two of the markers we sequenced yielded incomplete taxon sampling due to amplification and sequencing difficulties (ATP synthase gene) and the presence of multiple sequence types, all of which were not recovered in all taxa (HSP gene type 2). We conducted our tests using all five markers, including the two markers with incomplete data sets, and using only the three markers that with complete taxon sampling (the actin gene, the methionine synthase gene, and the combines chloroplast regions).

FIGURE 4.3 briefly illustrates how the splits networks were interpreted. Each band of parallel lines represents a split that groups taxa together. Unlike a phylogeny, each

node does not represent a common ancestor, so the splits network only provides an implicit representation of evolutionary history (Huson and Bryant 2006). FIGURE 4.3.A and 4.3.B represent two sets of splits (marked with arrows and shown in red) that show conflicting interpretations of the placement of the hybrid *Phalaenopsis venosa* X *P. equestris*. The split in FIGURE 4.3.A shows that the hybrid is in a clade with *P. equestris* and FIGURE 4.3.B shows that it is in a clade with *P. venosa* and *P. amboinensis*. Although *P. amboinensis* is included in the split with *P. venosa*, and there is a split that is adjacent to the hybrid, grouping it with only *P. amboinensis* (marked with an arrow), we considered this to be a correct reconstruction of the hybrid taxon by Neighbor-Net because *P. venosa* and *P. amboinensis* were sister taxa with strong support in the *Phalaenopsis* chloroplast phylogeny (FIGURE 4.2). FIGURES 4.3.C and 4.3.D uses the same splits network to illustrate Neighbor-Net's propensity for false positives. These two splits indicate that *P. fasciata* is a hybrid between *P. violacea* and *P. hieroglyphica,* which is incorrect. FIGURES 4.3.E and 4.3.F illustrates a splits network that includes the hybrid *P. amabilis* X *P. amboinensis*. FIGURE 4.3.E shows a split correctly placing the hybrid with its parent, *P. amboinensis*. FIGURE 4.3.F shows a split that places the hybrid with a group containing its other parent, *P. amabilis*, as well as two other species. We scored this as correct because the hybrid's parent was contained in the clade segregated by the split directly adjacent to the hybrid (there were no intervening splits). We chose to interpret the splits networks quite liberally because they do not represent evolutionary histories, but rather unspecified incongruence of data. We are working with our collaborators to develop automated objective methods to evaluate splits networks.

TABLE 4.2. Data sets used to evaluate methods of reconstructing hybrid evolution. Each test included all of the non-hybrid taxa and one, two, or three hybrids. Independent hybrids were hybrids that had no parents in common. Non-independent hybrids shared one parent.

Category	**Test Number**	***Phalaenopsis* Hybrid(s) included**
A. One hybrid	A.1	• *P. amabilis x P. amboinensis*
	A.2	• *P. pulcherrima x P. fasciata*
	A.3	• *P. venosa x P. equestris*
	A.4	• *P. pulcherrima x P. equestris*
	A.5	• *P. buyssoniana x P. equestris*
	A.6	• *P. equestris x P. aphrodite*
	A.7	• *P. pulcherrima x P. violacea*
B. Two independent hybrids	B.1	• *P. amabilis x P. amboinensis* • *P. pulcherrima x P. fasciata*
	B.2	• *P. amabilis x P. amboinensis* • *P. venosa x P. equestris*
	B.3	• *P. amabilis x P. amboinensis* • *P. pulcherrima x P. equestris*
	B.4	• *P. amabilis x P. amboinensis* • *P. buyssoniana x P. equestris*
	B.5	• *P. amabilis x P. amboinensis* • *P. equestris x P. aphrodite*
	B.6	• *P. amabilis x P. amboinensis* • *P. pulcherrima x P. violacea*
	B.7	• *P. pulcherrima x P. fasciata* • *P. venosa x P. equestris*
	B.8	• *P. pulcherrima x P. fasciata* • *P. equestris x P. aphrodite*
	B.9	• *P. venosa x P. equestris* • *P. pulcherrima x P. violacea*
	B.10	• *P. equestris x P. aphrodite* • *P. pulcherrima x P. violacea*
C. Three independent hybrids	C.1	• *P. amabilis x P. amboinensis* • *P. pulcherrima x P. fasciata* • *P. venosa x P. equestris*
	C.2	• *P. amabilis x P. amboinensis* • *P. pulcherrima x P. fasciata* • *P. equestris x P. aphrodite*
	C.3	• *P. amabilis x P. amboinensis* • *P. venosa x P. equestris* • *P. pulcherrima x P. violacea*
	C.4	• *P. amabilis x P. amboinensis* • *P. equestris x P. aphrodite* • *P. pulcherrima x P. violacea*

TABLE 4.2. Continued.

Category	**Test Number**	***Phalaenopsis* Hybrid(s) included**
D. Two non-independent hybrids	D.1	• *P. pulcherrima x P. fasciata* • *P. pulcherrima x P. equestris*
	D.2	• *P. pulcherrima x P. fasciata* • *P. pulcherrima x P. violacea*
	D.3	• *P. venosa x P. equestris* • *P. pulcherrima x P. equestris*
	D.4	• *P. venosa x P. equestris* • *P. equestris x P. aphrodite*
	D.5	• *P. pulcherrima x P. equestris* • *P. equestris x P. aphrodite*
	D.6	• *P. pulcherrima x P. equestris* • *P. pulcherrima x P. violacea*

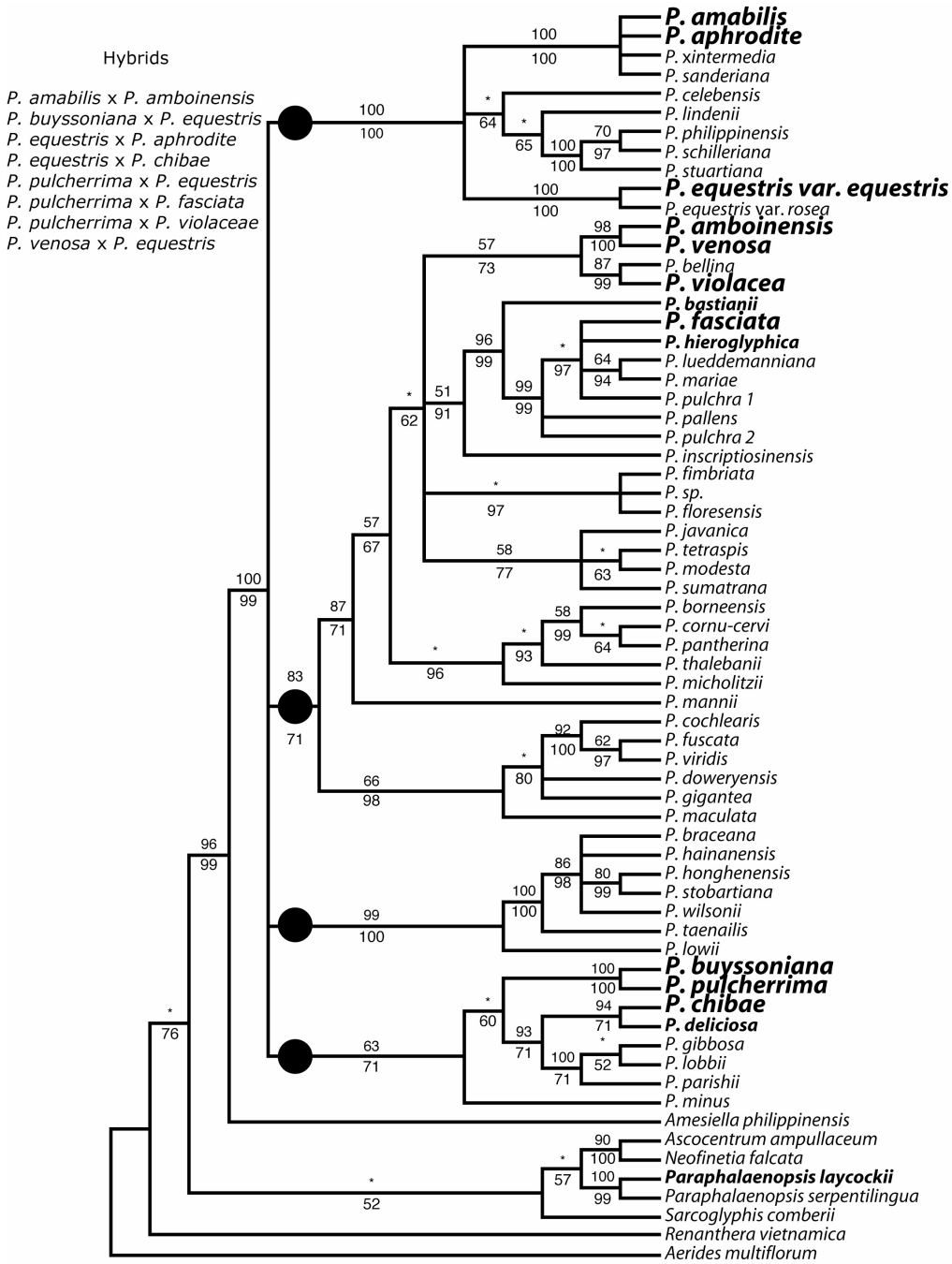


FIGURE 4.2. Phylogeny of *Phalaenopsis* recovered from Bayesian analysis of combined chloroplast data. Species in bold are included in this chapter's analyses. Species in large font are parents of hybrids. The four major clades of *Phalaenopsis* are indicated by a black circle ●. Bootstrap values are indicated above branches and Bayesian posterior probabilities below branches.

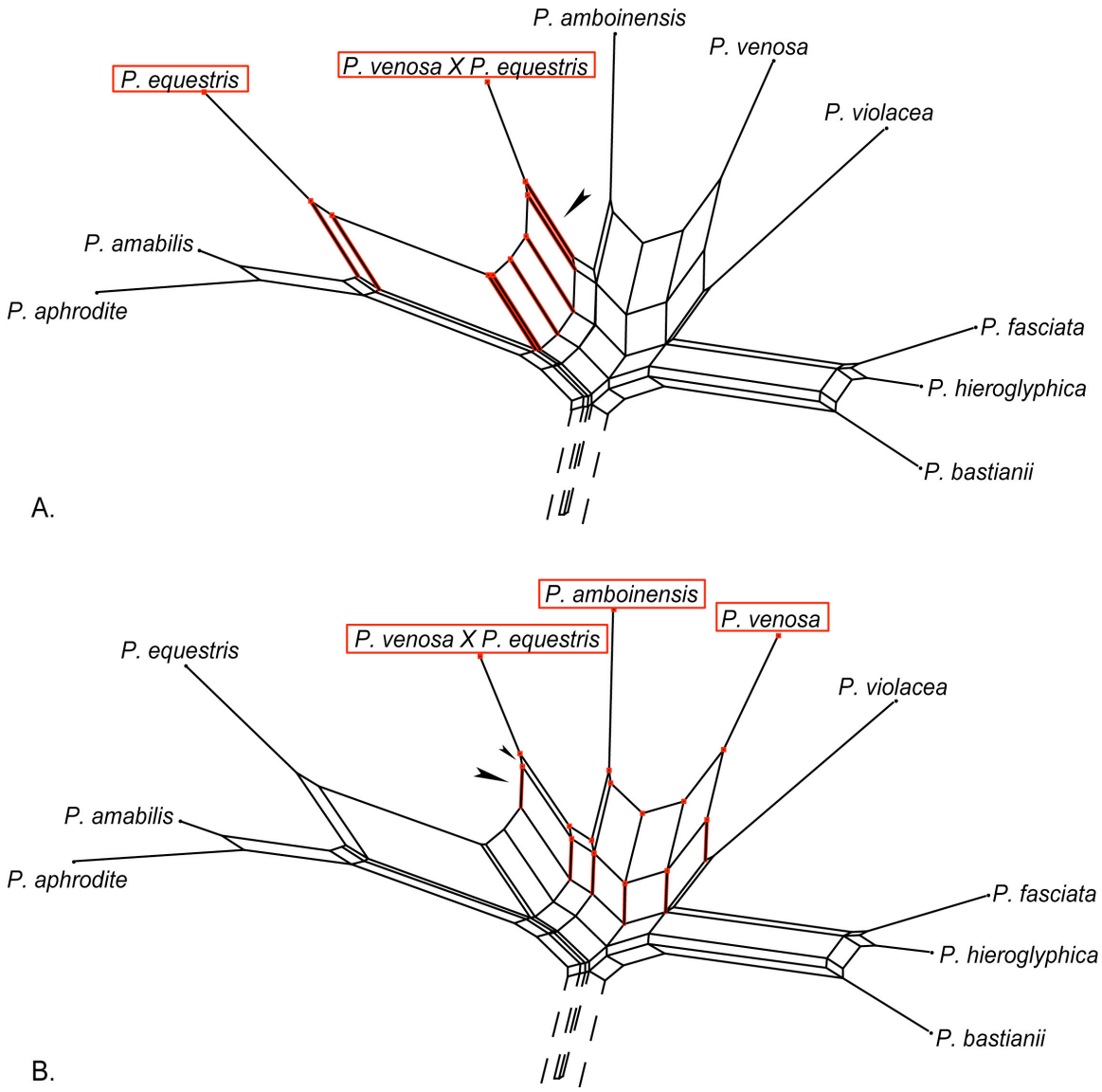
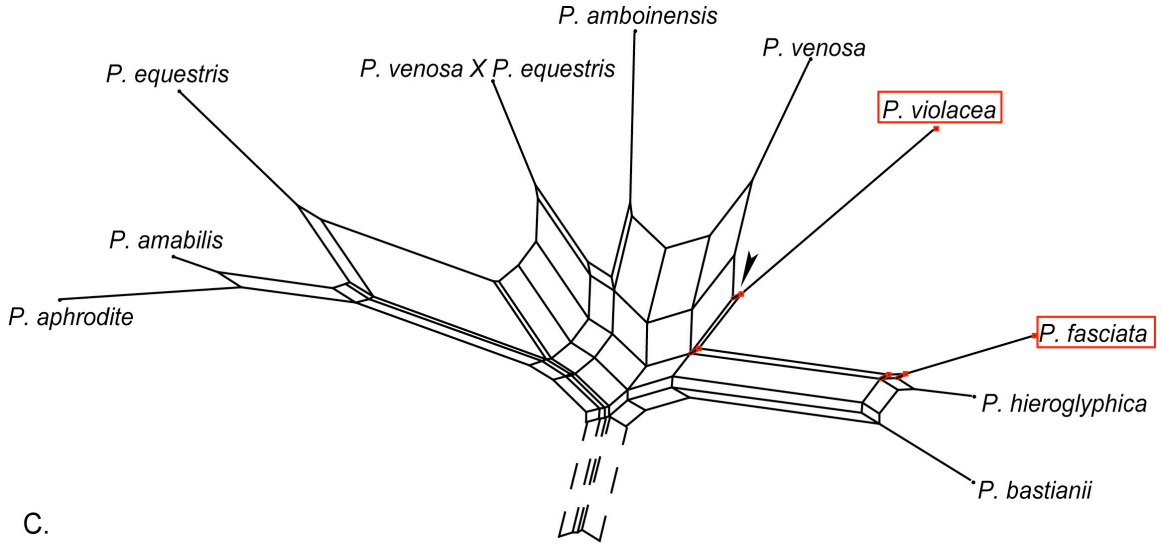


FIGURE 4.3. Part of a splits network recovered from Neighbor-Net using five datasets with some missing data and including one hybrid, *Phalaenopsis venosa* X *P. equestris* (see FIGURE 4.4) in A, B, C, and D, and *P. amabilis* X *P. amboinensis* in E and F. Splits are represented by bands of parallel lines, and arrows ➤ are used to indicate the splits of interest. The splits of interest and all nodes and taxa associated with the split are shown in red.



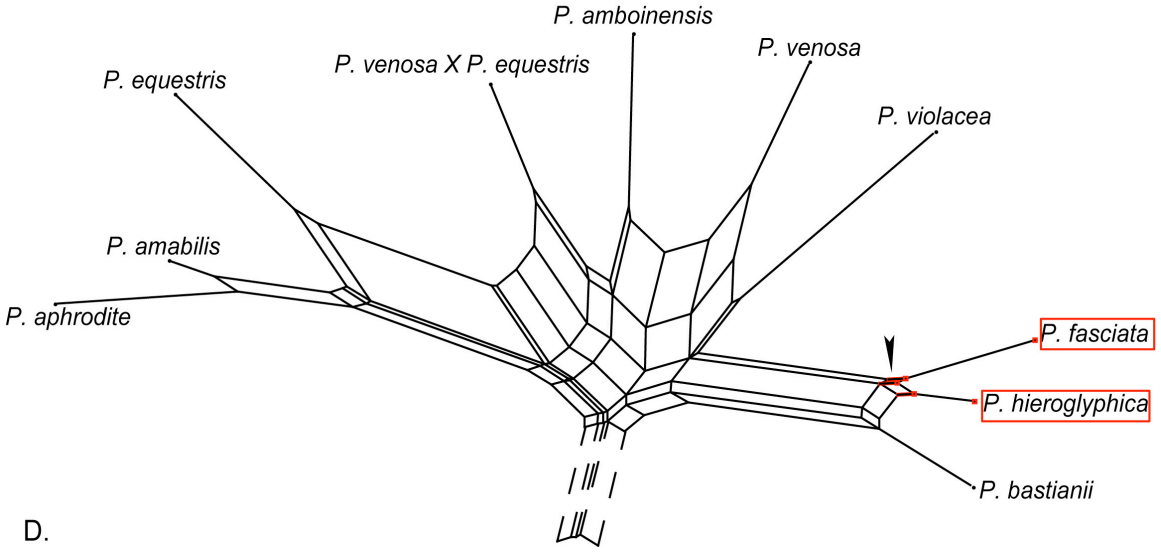
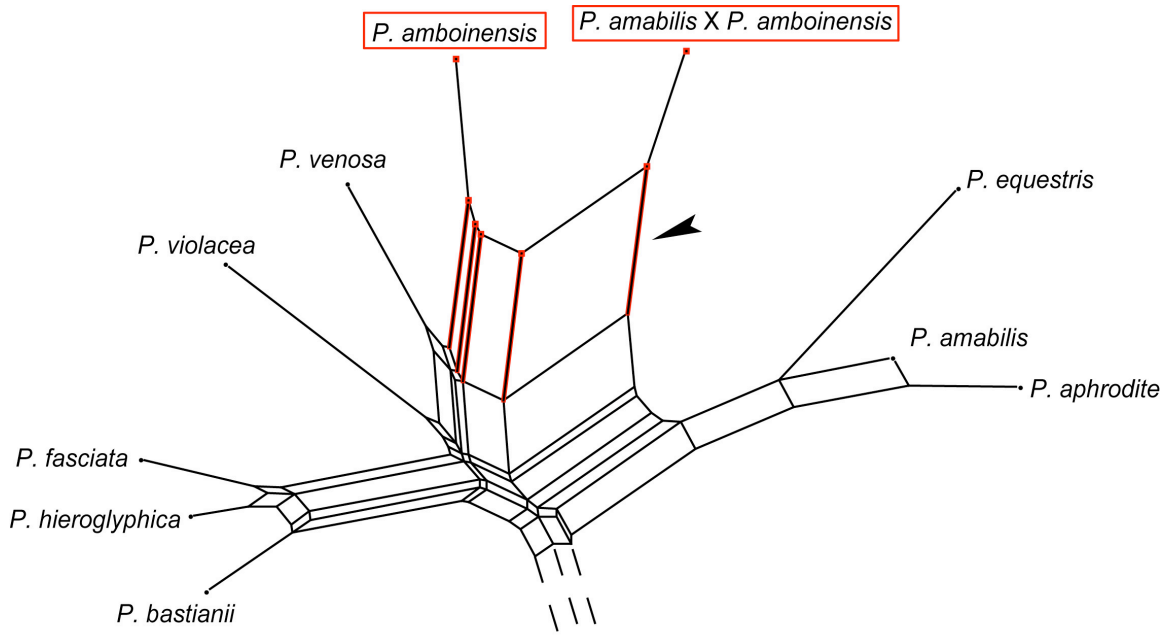
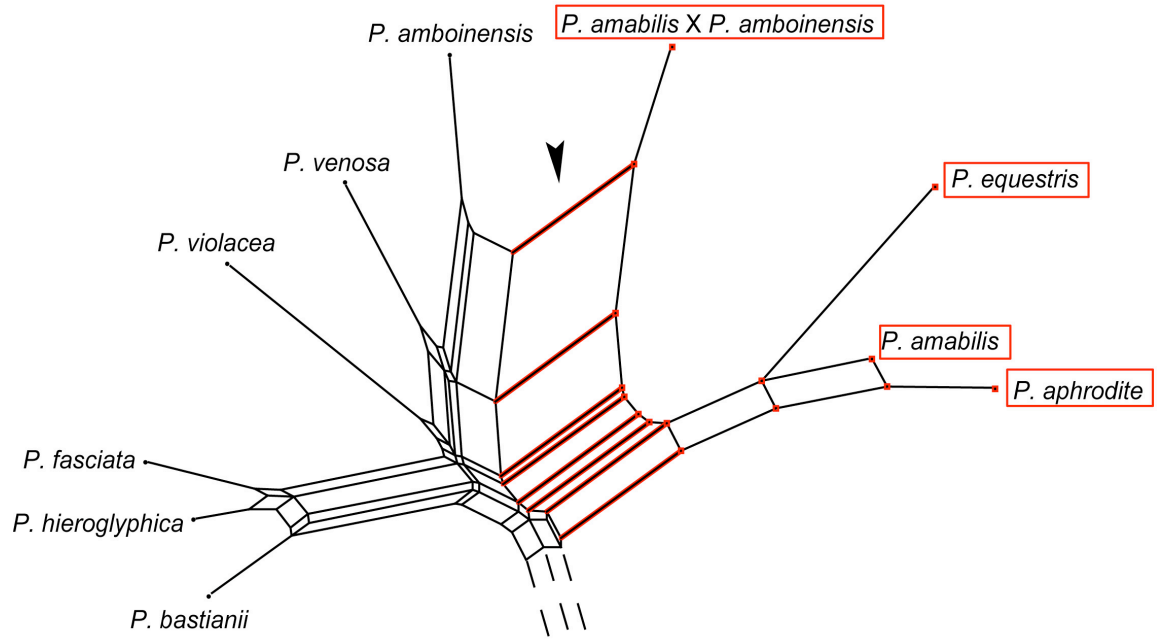


FIGURE 4.3 continued.



E.



F.

FIGURE 4.3 continued.

RESULTS

Neighbor-Net was able to predict accurately the parents of hybrids in only half of the datasets we tested, and there were so many false positives that it was impossible to distinguish the hybrids from the species (FIGURE 4.4). All of the data sets we tested resulted in more than five false positives, assuming that splits represented hybridization and not some other type of conflict in the data. Splits occurred at many different levels, deep within the network as well as at the nodes FIGURES 4.5-4.12. Most of our results indicated that every taxon was a hybrid, and so we decided against using false negatives to report data. Instead, we calculated the error in Neighbor-Net's estimates of the hybrids' parentage. An error would usually consist of one parent of a hybrid not being identified. In all cases, Neighbor-Net correctly predicted at least one of the parents of all of the hybrids. Results were not always consistent when testing with three complete or five incomplete datasets. There were instances in all test categories where analysis with five incomplete datasets would produce an error that was not present when testing with three complete datasets. In addition, in test categories that included more than one hybrid, there were several occasions where analyzing three complete datasets resulted in an error in identifying one hybrid's parents, and analyses of five incomplete datasets would produce an error identifying the parents of a different hybrid.

Overall, 48% (13 tests with errors out of 27 total tests) of the tests using three complete datasets and 59% (16 out of 27) of the tests using five datasets, two of which contained missing data, resulted in at least one error in identifying the parents of the hybrids (FIGURE 4.4). Tests including one hybrid resulted in the least amount of errors; only 29% of the tests (2 out of 7) did not properly identify the hybrid's parents, regardless of whether three complete datasets or five incomplete datasets were used.

Tests with two independent hybrids had error rates of 60% (6 out of 10), regardless of how many datasets were included. Tests including two non-independent hybrids and three independent hybrids had a lower rate of error when three complete datasets were included than when five incomplete datasets were used. Tests with two non-independent hybrids had an error rate of 50% (3 out of 6) and 83% (5 out of 6) and tests with three independent hybrids had an error rate of 50% (2 out of 4) and 75% (3 out of 4), when three complete datasets were included and when five incomplete datasets were included, respectively. Differences between splits networks using five datasets with some missing data and three complete datasets can be seen in FIGURES 4.5 – 4.12 (compare FIGURES 4.5 with 4.6, 4.7 with 4.8, 4.9 with 4.10, and 4.11 with 4.12). Tests with missing data tended to be less accurate and have longer edge lengths than tests without missing data. Overall, our data showed that Neighbor-Net performed poorly, especially when more than one hybrid was present.

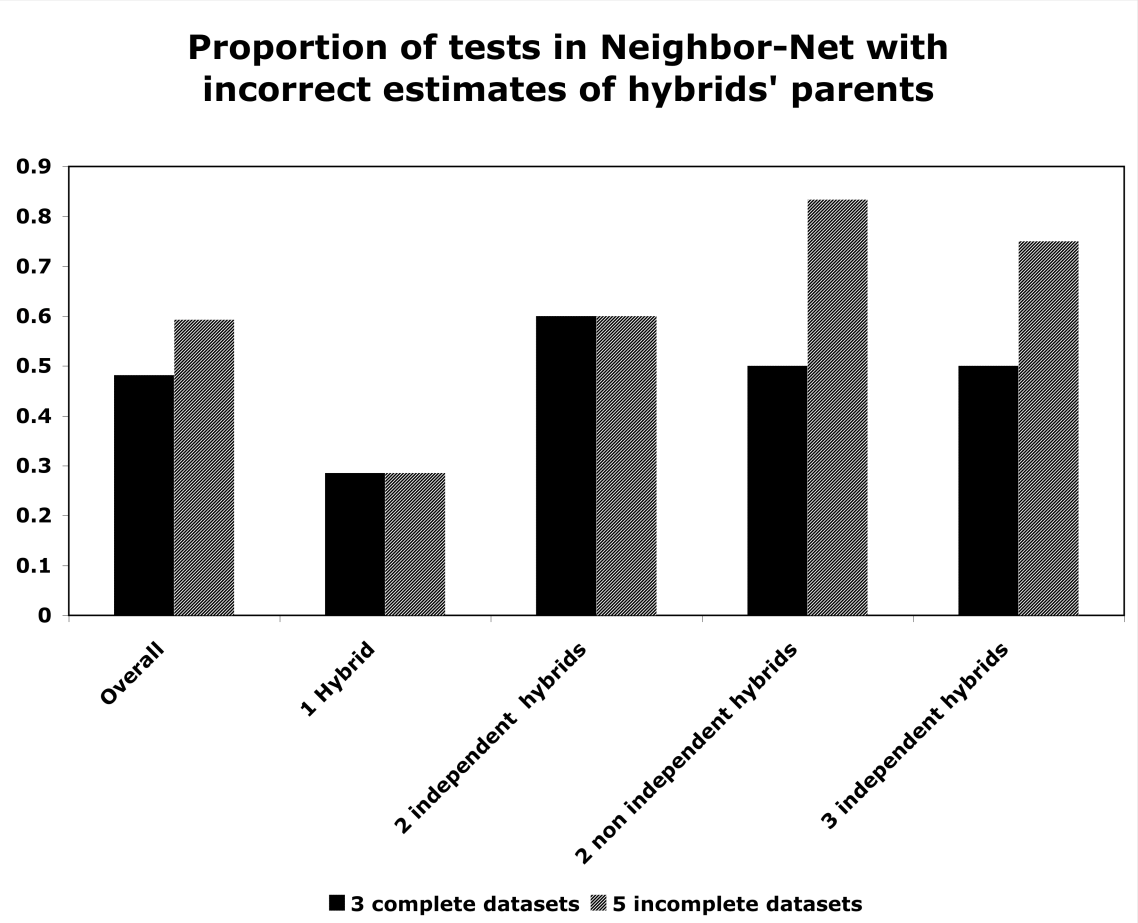


FIGURE 4.4. Proportion of tests conducted in Neighbor-Net with incorrect estimates of the identities of hybrids' parents. Comparison between three complete datasets (in black) and five datasets, two of which are missing data for several taxa (in gray).

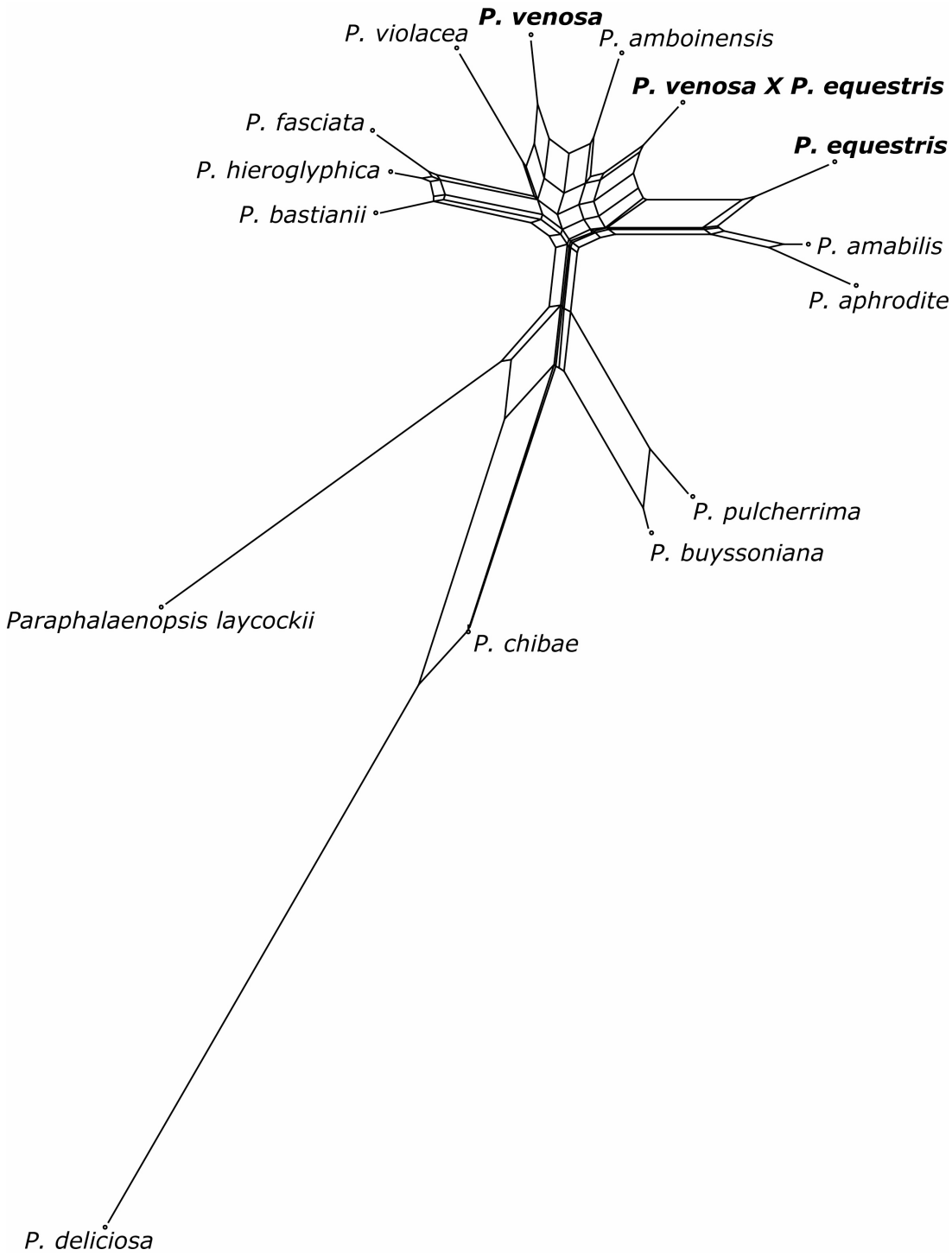


FIGURE 4.5. Splits network recovered from Neighbor-Net using five datasets with some missing data and including one hybrid, *Phalaenopsis venosa* X *P. equestris*. The hybrid taxon and its parent are indicated in bold.

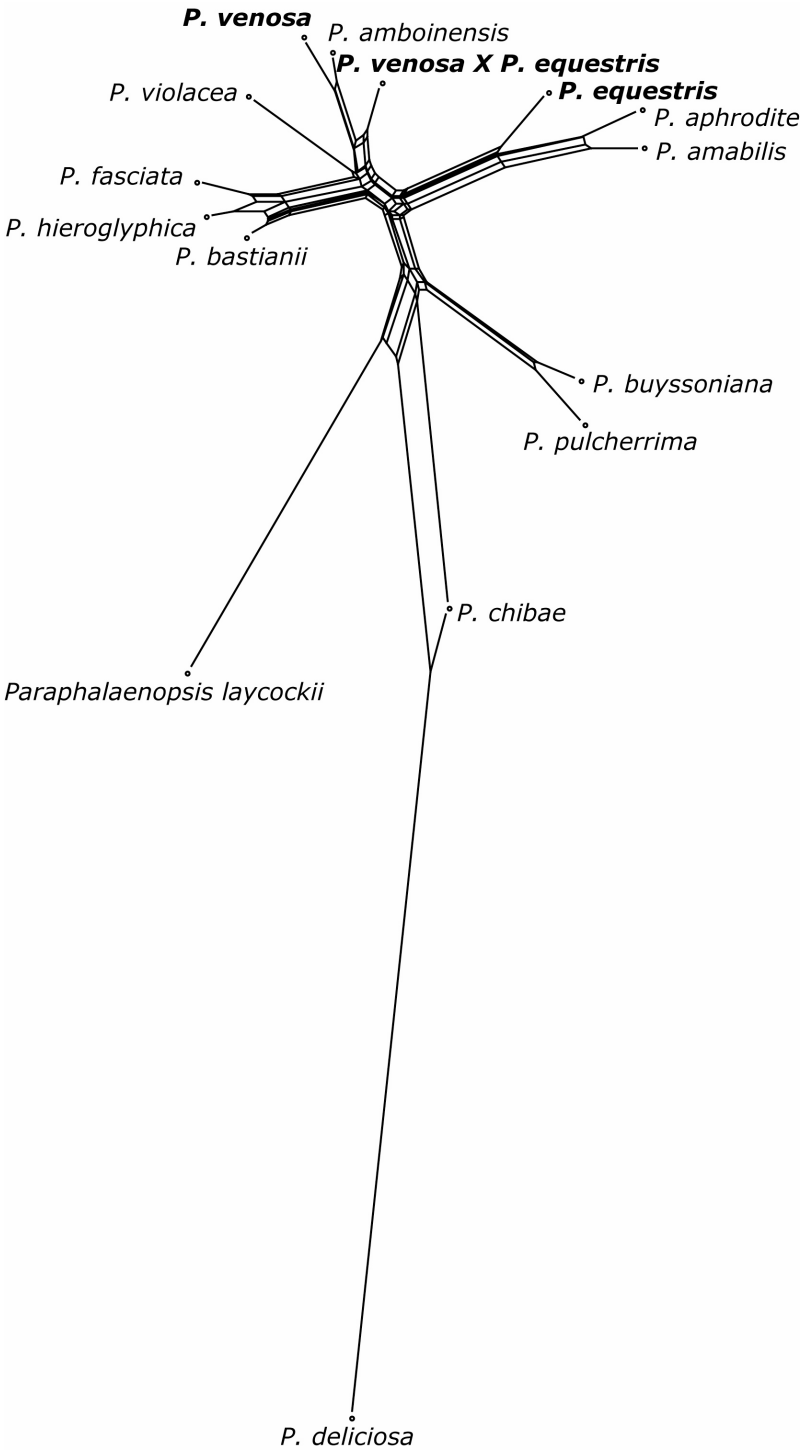


FIGURE 4.6. Splits network recovered from Neighbor-Net using three complete datasets and including one hybrid, *Phalaenopsis venosa* X *P. equestris*. The hybrid taxon and its parent are indicated in bold.

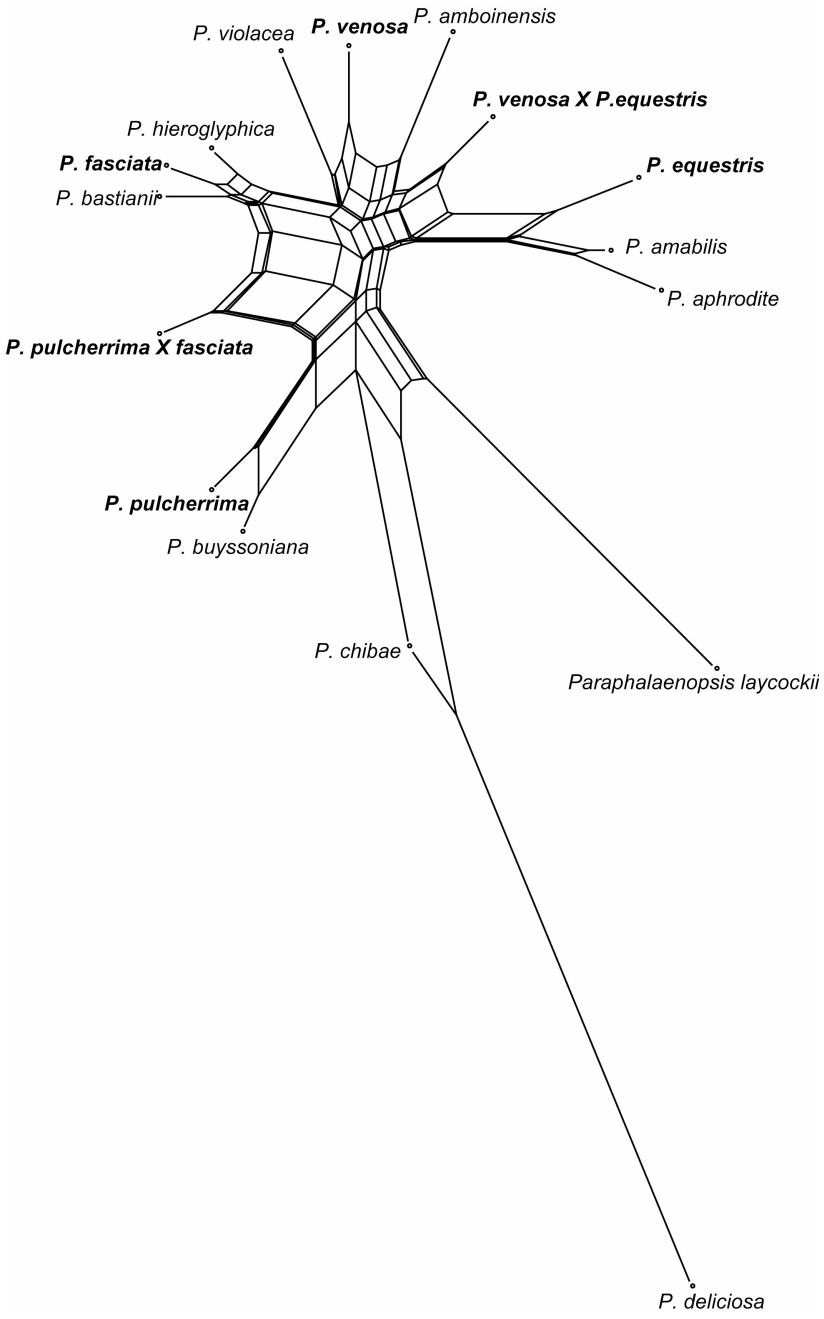


FIGURE 4.7. Splits network recovered from Neighbor-Net using five datasets with some missing data and including two independent hybrids, *Phalaenopsis venosa* X *P. equestris* and *P. pulcherrima* X *P. fasciata*. Hybrid taxa and their parents are indicated in bold.

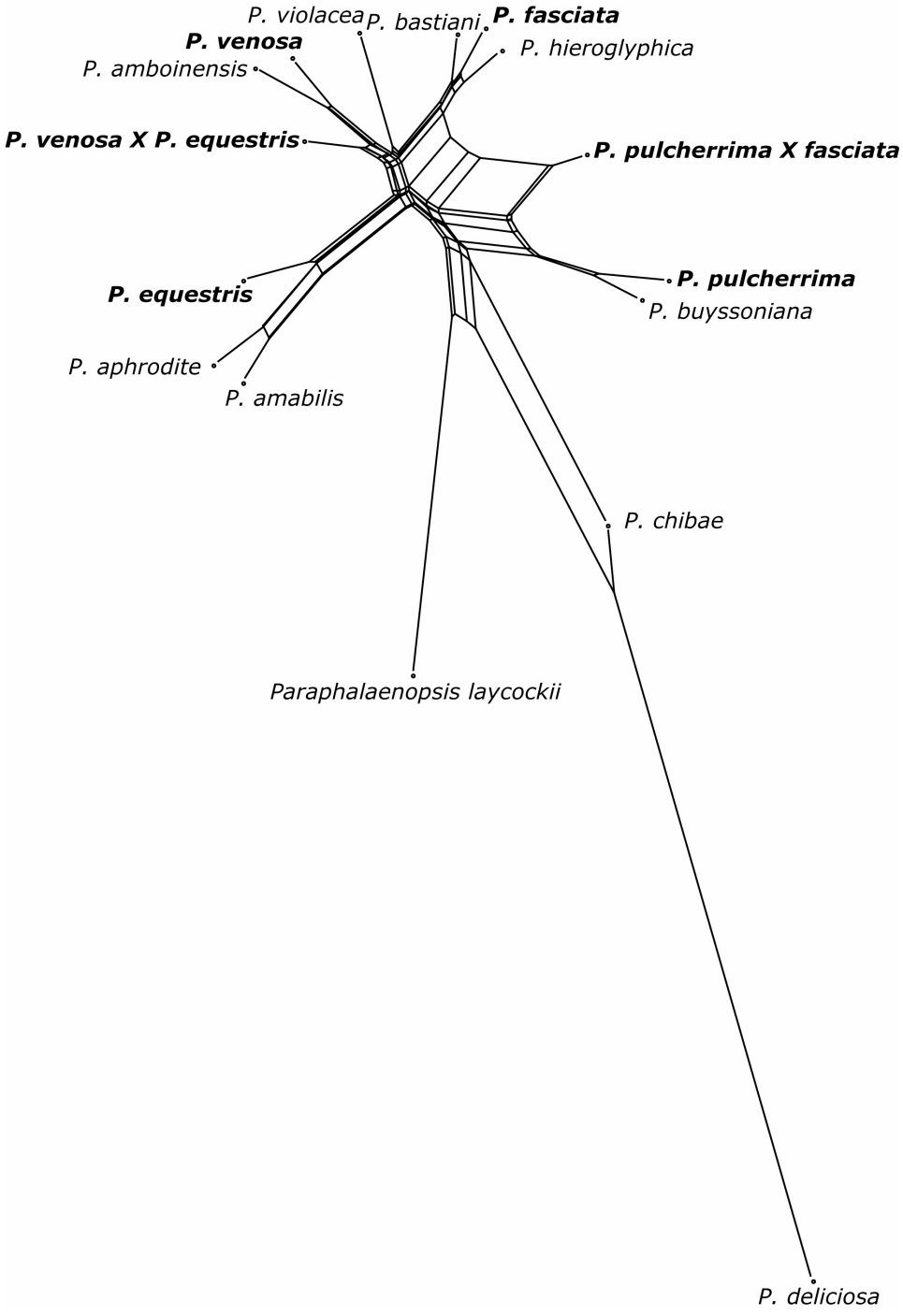


FIGURE 4.8. Splits network recovered from Neighbor-Net using three complete datasets and including two independent hybrids, *Phalaenopsis venosa* X *P. equestris* and *P. pulcherrima* X *P. fasciata*. Hybrid taxa and their parents are indicated in bold.

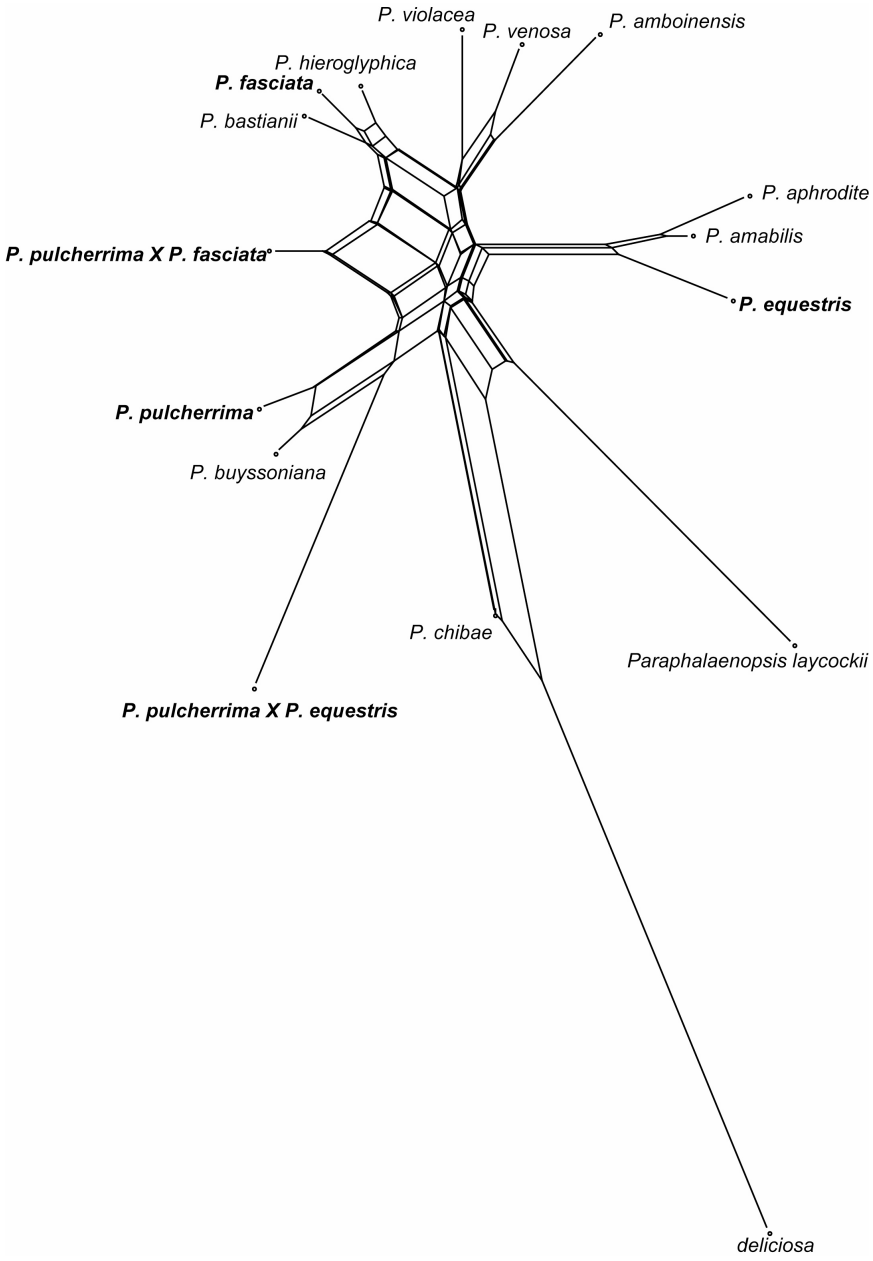


FIGURE 4.9. Splits network recovered from Neighbor-Net using five datasets with some missing data and including two non-independent hybrids, *Phalaenopsis pulcherrima* X *P. equestris* and *P. pulcherrima* X *P. fasciata*. Neighbor-Net did not recover *P. equestris* as a parent of *P. pulcherrima* X *P. equestris*. Hybrid taxa and their parents are indicated in bold.

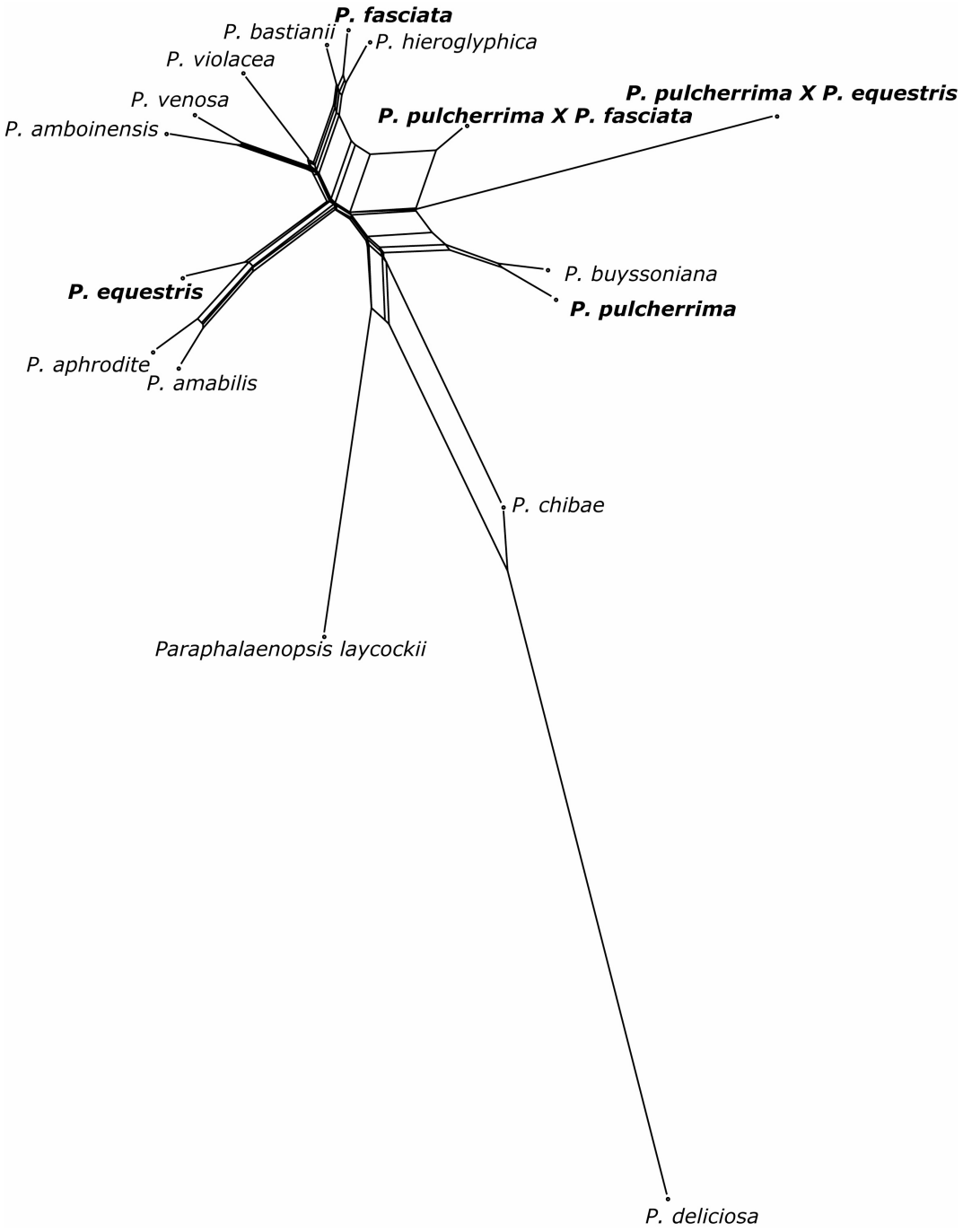


FIGURE 4.10. Splits network recovered from Neighbor-Net using three complete datasets and including two non-independent hybrids, *Phalaenopsis pulcherrima* X *P. equestris* and *P. pulcherrima* X *P. fasciata*. Neighbor-Net did not recover *P. equestris* as a parent of *P. pulcherrima* X *P. equestris*. Hybrid taxa and their parents are indicated in bold.

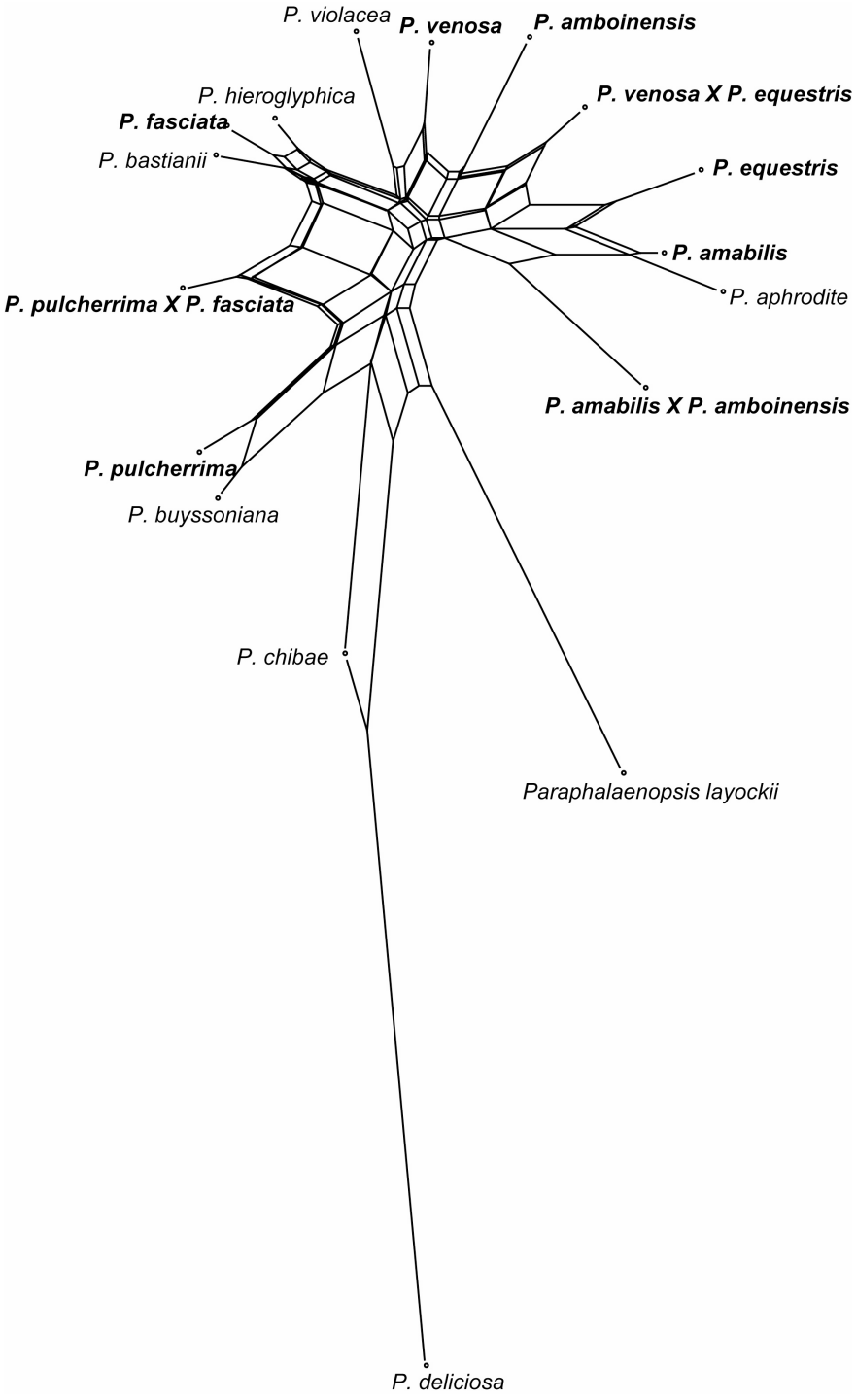


FIGURE 4.11. Splits network recovered from Neighbor-Net using five datasets with some missing data and including three independent hybrids, *Phalaenopsis venosa* X *P. equestris*, *P. pulcherrima* X *P. fasciata*, and *P. amabilis* X *P. amboinensis*. Hybrid taxa and their parents are indicated in bold.

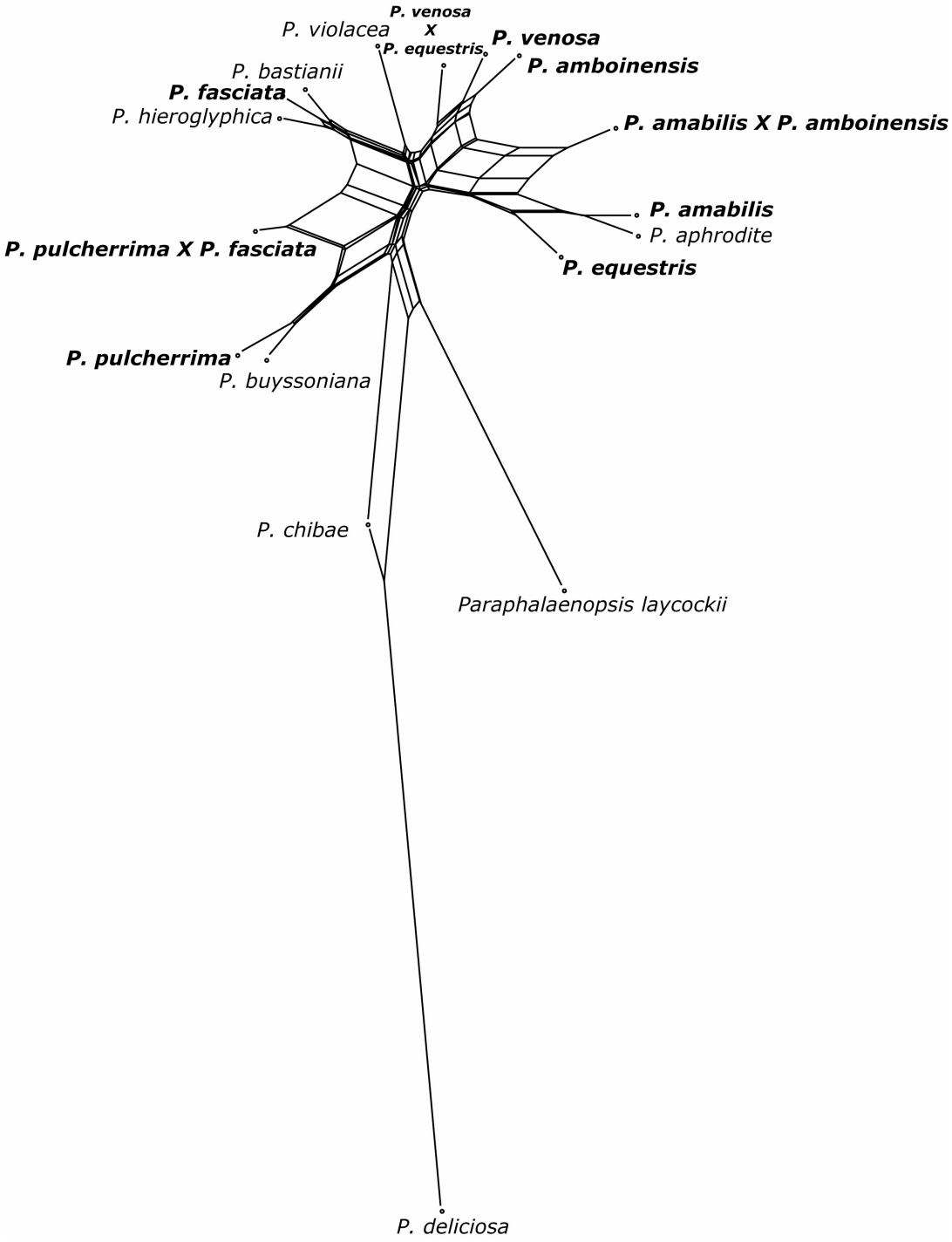


FIGURE 4.12. Splits network recovered from Neighbor-Net using three complete datasets and including three independent hybrids, *Phalaenopsis venosa* X *P. equestris*, *P. pulcherrima* X *P. fasciata*, and *P. amabilis* X *P. amboinensis*. Hybrid taxa and their parents are indicated in bold.

DISCUSSION

Our results lead us to advise against using Neighbor-Net for detecting hybridization. Neighbor-Net returned networks with so many splits (false positives) that interpreting the results was confusing. In addition, it was not able to identify reliably the parents of hybrid taxa. Nakhleh et al. (2004) performed simulation studies to assess the reliability of Neighbor-Net (Bryant and Moulton 2004; Nakhleh et al. 2004) and showed that Neighbor-Net had a high false positive rate on both trees and networks, regardless of sequence length.

Our results indicated that using less data with no missing taxa might provide more accurate results than adding additional, incomplete datasets. Perhaps with cleaner data, Neighbor-Net would perform better, but most sequences from real organisms (as opposed to simulated data) are going to have a certain level of 'noise.' As previously noted, however, Bryant and Moulton (2004) intended Neighbor-Net to provide "a snapshot of the data that can guide a more detailed analysis," and so perhaps it is unfair to analyze these data under the assumption that a split represents a hybridization event. However, it is important to show that using Neighbor-Net to predict hybridization should be used with caution. Useful methods will need to differentiate between biological processes that give similar signals to hybridization, such as lineage sorting, meiotic recombination, and stochastic homology. These difficulties are a continuing challenge for designing programs that detect hybrid taxa and their parents.

Preliminary results using the program RIATA (Nakhleh et al. pers com.) also gave a high number of false positives. It is thought that this might be due to the fact that the majority rule consensus trees that are being used as input are unresolved in different parts of the trees. For example (FIGURE 4.13), the three consensus trees recovered from the actin gene (FIGURE 4.13 A), the methionine synthase gene (FIGURE 4.13 B), and the

combined chloroplast regions (FIGURE 4.13 C) using only *Phalaenopsis amabilis* X *P. amboinensis* as a hybrid give three different, mostly compatible resolutions. The problem is that there are polytomies of different species and clades (marked with an arrow ≻) in the three different resolutions. These differences are causing the algorithms to predict false hybridization events. Simulated sequences did not encounter these problems because no polytomies were present in the model trees. These findings underscore the need for testing methods on real as well as simulated data. Useful methods will have to be able to deal with partially unresolved trees, as this is a common occurrence in phylogeny reconstruction from real organisms.

Once RGNet and RIATA are available we will test their accuracy using the datasets that we have developed. Our findings suggest that it may be difficult to design an algorithm that can accurately discern true hybrid species from non-hybrid species since other biological processes produce a signal similar to hybridization. We would like to create additional datasets from other families of plants on which to test these methods. The development of accurate methods for reconstructing hybrid evolution is critical if we are to accurately reconstruct the tree of life.

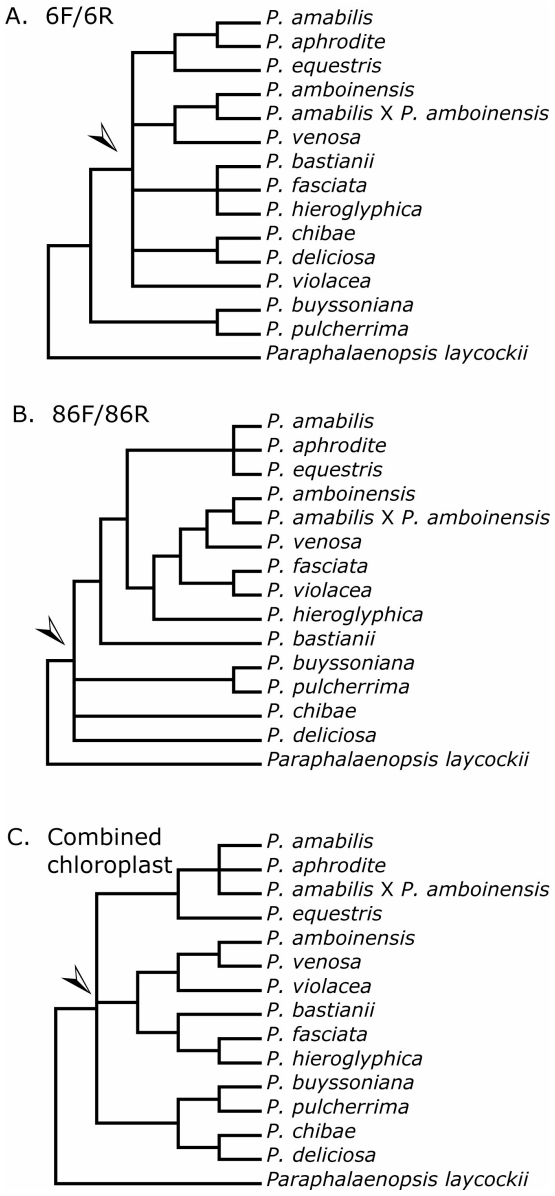


FIGURE 4.13. Three phylogenies recovered from datasets including one hybrid, *Phalaenopsis amabilis* X *P. amboinensis*: A. Actin gene, B. methionine synthase gene, C. combined chloroplast regions. Each phylogeny has a different polytomy, marked with an arrow ➢.

LITERATURE CITED

ARCHAMBAULT, A. and A. BRUNEAU. 2004. Phylogenetic utility of the LEAFY/FLORICAULA gene in the Caesalpinioideae (Leguminosae): Gene duplication and a novel insertion. Systematic Botany 29: 609-626.

BAACK, E.J. 2005. To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. Heredity 94: 538-546.

BAILEY, C.D., T.G. CARR, S.A. HARRIS, and C.E. HUGHES. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics & Evolution 29: 435-455.

BALDWIN, B.G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Molecular Phylogenetics and Evolution 1: 3-16.

BANDELT, H.J. and A.W.M. DRESS. 1992. A canonical decomposition-theory for metrics on a finite-set. Advances in Mathematics 92: 47-105.

BANDELT, H.J., P. FORSTER, and A. ROHL. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.

BANDELT, H.J., V. MACAULAY, and M. RICHARDS. 2000. Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Molecular Phylogenetics and Evolution 16: 8-28.

BENNET, M.D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London 181: 109-135.

BENSON, D.A., I. KARSCH-MIZRACHI, D.J. LIPMAN, J. OSTELL, and D.L. WHEELER. 2004. GenBank: update. Nucleic Acids Research 32: D23-D26.

BENTHAM, G. 1883. *Phalaenopsis*. In G. Bentham and J.D. Hooker eds., Genera Plantarum, pp. L. Reeve & Co., London, England.

BENZING, D.H., W.E. FRIEDMAN, G. PETERSON, and A. RENFROW. 1983. Shootlessness, Velamentous Roots, And The Pre-Eminence Of Orchidaceae In The Epiphytic Biotope. American Journal Of Botany 70: 121-133.

BLUME, C.L. 1825. *Phalaenopsis*. Bijdragen Tot de Flora van Nederlandsch Indie 7: 294.

BRYANT, D. and V. MOULTON. 2002. NeighborNet: An agglomerative method for the construction of planar phylogenetic networks, Algorithms in Bioinformatics, Proceedings, pp. 375-391.

_____. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution 21: 255-265.

CAMERON, K.M., M.W. CHASE, W.M. WHITTEN, P.J. KORES, D.C. JARRELL, V.A. ALBERT, T. YUKAWA, H.G. HILLS, and D.H. GOLDMAN. 1999. A phylogenetic analysis of the Orchidaceae: Evidence from rbcL nucleotide sequences. American Journal of Botany 86: 208-224.

CHRISTENSON, E.A. 1995. An overview of the genus *Phalaenopsis*. Orchid Digest Jan-Feb-Mar: 19-22.

_____. 2001. *Phalaenopsis*: a Monograph. Timber Press, Portland, Oregon.

CRAWFORD, D.J. 1985. Electrophoretic data and plant speciation. Systematic Botany 10: 405-416.

DE WIT, H.C.D. 1977. Orchids in Rumphius' Herbarium Amboinense. In J. Arditti ed., Orchid Biology: Reviews and Perspectives, pp. 47-94. Cornell University Press, Ithaca.

DEMESURE, B., N. SODZI, and R.J. PETIT. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4: 129-131.

DOWTON, M. and A.D. AUSTIN. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - The behavior of the incongruence length difference test in mixed-model analyses. Systematic Biology 51: 19-31.

DOYLE, J.J. and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19: 11-15.

DOYLE, J.J., J.L. DOYLE, and C. HARBISON. 2003. Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: Phylogeny, gene duplication, and ancient polyploidy. Systematic Botany 28: 567-577.

EVANS, B.J., J. SUPRIATNA, and D.J. MELNICK. 2001. Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. Evolution 55: 1686-1702.

EVANS, G. and H. REES. 1971. Mitotic cycles in dicotyledons and monocotyledons. Nature 233.

EXCOFFIER, L. and P.E. SMOUSE. 1994. Using allele frequencies and geographic subdivision to reconstructG gene trees within a species - Molecular variance parsimony. Genetics 136: 343-359.

FARRIS, J.S., M. KALLERSJO, A.G. KLUGE, and C. BULT. 1994. Testing Significance Of Incongruence. Cladistics-The International Journal Of The Willi Hennig Society 10: 315-319.

FU, Y.M., W.H. CHEN, W.T. TSAI, Y.S. LIN, M.S. CHYOU, and Y.H. CHEN. 1997. Phylogenetic studies of taxonomy and evolution among wild species of Phalaenopsis by random amplified polymorphic DNA markers. [Chinese]. Report of the Taiwan Sugar Research Institute 0: 27-42.

FUNK, V.A. 1985. Phylogenetic patterns and hybridization. Annals of the Missouri Botanical Gardens 72: 681-715.

FUTUYMA, D.J. 1997. Evolutionary Biology. Sinauer Assoc. Inc., Sunderland, MA.

GALLEZ and L.D. GOTTLIEB. 1982. Genetic evidence for the hybrid origin of the diploid plant Stephanomeria diegensis. Evolution 36: 1158-1167.

GRANT, V. 1966. Selection for vigor and fertility in the progeny of a hihly sterile species hybrid in *Gillia*. Genetics 53.

_____. 1981. Plant Speciation. Columbia University Press, New York.

GRIESBACH, R.J. 2002. Development of Phalaenopsis Orchids for the Mass Market. ASHS Press, Alexandria, VA.

GRUSS, O. and L. ROLLKE. 1993a. Gattung *Kingidium*. Die Orchidee 44: 33-36.

_____. 1993b. Gattung *Kingidium*, Teil 2. Die Orchidee 44: 119-122.

_____. 1995a. *Kingidium wightii* (Rchb.f.) Gruss & Rollke comb. nov., Teil 10. Die Orchidee 46: 23-28.

_____. 1995b. *Kingidium philippinense* (Ames) Gruss & Rollke comb. nov., Teil 11. Die Orchidee 46: 115-118.

_____. 1996. *Kingidium wilsonii* (Rolfe) Gruss & Rollke comb. nov. Die Orchidee 47: 149-155.

_____. 1997. *Kingidium chibae* (T. Yukawa) Gruss & Rollke comb. nov. Die Orchidee 48: 261-263.

HAAHR, M. 2006. random.org - True random number service. Distributed Systems Group, Department of Computer Science. University of Dublin, Trinity College.

HAMZEH, M. and S. DAYANANDAN. 2004. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast tRNT-tRNF region and nuclear rDNA. American Journal of Botany 91: 1398-1408.

HAWKES, A.D. 1966. *Grafia*, a new genus of orchids. Phytologia 13: 305-306.

HEIN, J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. Mathematical Biosciences 98: 185-200.

_____. 1993. A heuristic method to reconstruct the history of sequences subject to recombination. Journal of Molecular Evolution 36: 396-405.

HENNIG, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana, Illinois.

HILLIS, D.M., C. MORITZ, and B.K. MABLE. 1996. Molecular Systematics. Sinauer Associates, Inc., Sunderland, MA.

HOLTTUM, R.E. 1965. Cultivated species of the orchid genus *Doritis* Lindl. Kew Bulletin 19: 207-212.

_____. 1966. The genus *Kingiella*. The Orchid Review 74: 128-129.

HUELSENBECK, J.P. and F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

HUNT, P.F. 1970. Notes on Asiatic Orchids: V. *Kingidium* P. F. Hunt, nom. nov. Kew Bulletin 24: 97-98.

HUSON, D.H. 1998. SplitsTree: A program for analyzing and visualizing evolutionary data. Bioinformatics 14: 68-73.

HUSON, D.H. and D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

HUSON, D.H., T. KLOPPER, P.J. LOCKHART, and M.A. STEEL. 2005. Reconstruction of reticulate networks from gene trees, Research in Computational Molecular Biology, Proceedings, pp. 233-249.

HUYNH, T.N.D., J. JANSSON, N.B. NGUYEN, and W.K. SUNG. 2005. Constructing a smallest refining galled phylogenetic network, Research In Computational Molecular Biology, Proceedings, pp. 265-280.

JONES, W.E., A.R. KUEHNLE, and K. ARUMUGANATHAN. 1998. Nuclear DNA content of 26 orchid (Orchidaceae) genera with emphasis on *Dendrobium*. Annals Of Botany 82: 189-194.

KAO, Y.Y., S.B. CHANG, T.Y. LIN, C.H. HSIEH, Y.H. CHEN, W.H. CHEN, and C.C. CHEN. 2001. Differential accumulation of heterochromatin as a cause for karyotype variation in *Phalaenopsis* orchids. Annals Of Botany 87: 387-395.

LANDAIS, I., M. OGLIASTRO, K. MITA, J. NOHATA, M. LOPEZ-FERBER, M. DUONOR-CERUTTI, T. SHIMADA, P. FOUMIER, and G. DEVAUCHELLE. 2003. Annotation pattern of ESTs from *Spodoptera frugiperda* Sf9 cells and analysis of the ribosoma protein genes reveal insect-specific features and unexpectedly low codon usage bias. Bioinformatics 19: 2243-2250.

LEGENDRE, P. and V. MAKARENKOV. 2002. Reconstruction of biogeographic and evolutionary networks using reticulograms. Systematic Biology 51: 199-216.

LEVIN, D.A. 2000. The Origin, Expansion and Demise of Plant Species. Oxford University Press, New York.

LIN, S., H.C. LEE, W.H. CHEN, C.C. CHEN, Y.Y. KAO, Y.M. FU, Y.H. CHEN, and T.Y. LIN. 2001. Nuclear DNA contents of *Phalaenopsis sp.* and *Doritis pulcherrima*. Journal of the American Society for Horticultural Science 126: 195-199.

LINDER, C.R. and L.H. RIESEBERG. 2004. Reconstructing patterns of reticulate evolution in plants. American Journal of Botany 91: 1700-1708.

LINDER, C.R., B.M.E. MORET, L. NAKHLEH, J. PADOLINA, J. SUN, A. THOLSE, R. TIMME, and T. WARNOW. 2003. An error metric for phylogenetic networks, Technical Report TR-CS-2003-2026, pp. University of New Mexico, Albuquerque, NM.

LINDLEY, J. 1833. *Doritis*, Genera and Species of Orchidaceous Plants, pp. 178. Ridgways, Piccadilly, London.

LINNAEUS, C. 1753. *Epidendrum amabile*. Species Plantarum ed. 1: 953-954.

LOHNE, C. and T. BORSCH. 2005. Molecular evolution and phylogenetic utility of the *pet*D group II intron: A case study in basal angiosperms. Molecular Biology & Evolution 22: 317-332.

LOYTYNOJA, A. and M.C. MILINKOVITCH. 2001. Molecular phylogenetic analyses of the mitochondrial ADP-ATP carriers: The Plantae/Fungi/Metazoa trichotomy revisited. Proceedings of the National Academy of Sciences of The United States of America 98: 10202-10207.

MADDISON, W.P. 1997. Gene trees in species trees. Systematic Biology 46: 523-536.

_____. 2005. Pairwise comparisons package for mesquite. http://mesquiteproject.org.

MADDISON, W.P. and D.R. MADDISON. 2000. MacClade 4.0. Sinauer, Sunderland, Massachusetts.

_____. 2005. Mesquite: A modular system for evolutionary analysis. http://mesquiteproject.org.

MAIER, R.M., K. NECKERMANN, G.L. IGLOI, and H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome - gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. Journal of Molecular Biology 251: 614-628.

MAKARENKOV, V. and P. LEGENDRE. 2004. From a phylogenetic tree to a reticulated network. Journal of Computational Biology 11: 195-212.

MASTERSON, J. 1994. Stomatal Size In Fossil Plants - Evidence For Polyploidy In Majority Of Angiosperms. Science 264: 421-424.

MCDADE, L. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44: 1685-1700.

_____. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46: 1329-1346.

MIRANKER, D., W. XU, and R. MAO. 2003. MoBIoS: a metric-space DBMS to support biological diversity. Proceedings of the International Conference on Scientific and statistical Database Management Systems.

MORRISON, D.A. 2005. Networks in phylogenetic analysis: new tools for population biology. International Journal for Parasitology 35: 567-582.

MULLIS, K.B. and F.A. FALOONA. 1987. Specific synthesis of DNA in vitro via polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335-350.

NAKHLEH, L., D. RUTHS, and H. INNAN. 2005. Gene trees, species trees, and species networks. Chapman & Hall, Norwell, MA.

NAKHLEH, L., T. WARNOW, C.R. LINDER, and K. ST. JOHN. 2004. Reconstructing Reticulate Evolution in Species - Theory and Practive. Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology: 337-346.

NAKHLEH, L., J. SUN, T. WARNOW, C.R. LINDER, B.M.E. MORET, and A. THOLSE. 2003. Towards the development of computational tools for evaluating phylogenetic network reconstruction methods. Proceedings of the Eighth Pacific Symposium on Biocomputing (PSB03): 315-326.

OKUYAMA, Y., N. FUJII, M. WAKABAYASHI, A. KAWAKITA, M. ITO, M. WATANABE, N. MURAKAMI, and M. KATO. 2005. Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Molecular Biology and Evolution 22: 285-296.

ORTI, G., D.E. PEARSE, and J.C. AVISE. 1997. Phylogenetic assessment of length variation at a microsatellte locus. Proceedings of the National Academy of Sciences of the United States of America 94: 10745-10749.

OTTO, S.P. and J. WHITTON. 2000. Polyploid incidence and evolution. Annual Review Of Genetics 34: 401-437.

PFITZER. 1889. Orchidaceae: *Phalaenopsis* Bl. In A. Engler and K. Prantl eds., Die natürlichen Pflanzenfamilien, Teil II, pp. 211-212. Verlag von Wilhelm Engelmann, Leipzig.

POSADA, D. and K.A. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

_____. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology & Evolution 16: 37-45.

RAMBAUT, A. and N.C. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Computer Applications in the Biosciences 13: 235-238.

REICHENBACH, H.G. 1862. Uebersicht der Gattungen *Phalaenopsis, Doritis, Stauropsis*. Xenia Orchidacea: 1-8, t. 101; 139-147, t. 151, t. 156.

_____. 1864. Neue Orchideen, *Phalaenopsis pantherina*. Botanische Zeitung 39: 298.

_____. 1879. New species of *Phalaenopsis*. The Gardeners Chronicle n.s., 11: 398.

RIESEBERG, L.H. 1991. Homoploid Reticulate Evolution in *Helianthus* (Asteraceae): Evidence from Ribosomal Genes. American Journal of Botany 78: 1218-1237.

_____. 1997. Hybrid origins of plant species. Annual Review of Ecology and Systematics 28: 359-389.

RIESEBERG, L.H. and N.C. ELLSTRAND. 1993. What can molecular and morphological markers tell us about plant hybridization? Critical Reviews in Plant Science 12: 213-241.

ROBBINSON, D.R. and L.R. FOULDS. 1981. Comparison of phylogenetic trees. Mathematical Biosciences 53: 131-147.

ROLFE, R.A. 1886. A revision of the genus *Phalaenopsis*. The Gardeners Chronicle 26: 168-170; 212-213; 267.

_____. 1917. *Doritis pulcherrima*. The Orchid Review 25: 195-197.

SAGAWA, Y. 1962. Cytological studies of the genus *Phalaenopsis*. American Orchid Society Bulletin 31: 390-398.

SANG, T. 2002. Utiliy of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology 37: 121-147.

SATTATH, S. and A. TVERSKY. 1977. Additive Similarity Trees. Psychometrika 42: 319-345.

SEIDENFADEN, G. 1988. Orchid Genera in Thailand XIV. Fifty-nine vandoid genera. Opera Botanica 95: 31-34; 182-189; 236-241.

SHAW, J., E.B. LICKEY, J.T. BECK, S.B. FARMER, W. LIU, J. MILLER, K.C. SIRIPUN, C.T. WINDER, E.E. SCHILLING, and R.L. SMALL. 2005. The Tortoise and the Hare II: Relative Utility of 21 Noncoding Chloroplast DNA Sequences for Phylogenetic Analysis. American Journal of Botany 92: 142-166.

SHIM, P.S. 1982. A new generic classification in the *Phalaenopsis*-complex (Orchidaceae). Malayan Nature Journal 36: 1-28.

SHINDO, N. and H. KAMEMOTO. 1963. Karyotype analysis of some species of Phalaenopsis. Cytologia 28: 390-398.

SMITH, J.J. 1933. Enumeration of the Orchidaceae of Sumatra and neighboring islands: *Phalaenopsis*. Repertorium specierum novarum regni vegetabilis ("Feddes Repertorium") 32: 364-367.

SNEL, B., P. BORK, and M.A. HUYNEN. 2002. Genomes in flux: The evolution of archaeal and proteobacterial gene content. Genome Research 12: 17-25.

SOLTIS, P.S., D.E. SOLTIS, and J.J. DOYLE. 1998. Molecular Systematics of Plants II: DNA Sequencing. Chapman and Hall, New York.

STEPHENS, P.F. 2005. Angiosperm Phylogeny Website. Version 6. http://www.mobot.org/MOBOT/research/APweb/.

STREELMAN, J.T., R. ZARDOYA, A. MEYER, and S.A. KARL. 1998. Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): Evolutionary comparison of microsatellite and single-copy nuclear loci. Molecular Biology And Evolution 15: 798-808.

SUDA, J., T. KYNCL, and V. JAROLIMOVA. 2005. Genome size variation in Macaronesian angiosperms: forty percent of the Canarian endemic flora completed. Plant Systematics And Evolution 252: 215-238.

SWEET, H.R. 1968. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 37: 867-877, 1089-1104.

_____. 1969. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 38: 33-43, 225-239, 321-336, 505-519, 681-694, 888-901.

_____. 1970. *Phalaenopsis* Section *Parishianae*. American Orchid Society Bulletin 39: 1094-1097.

_____. 1980. The Genus *Phalaenopsis*. Orchid Digest, Pomona, CA.

SWOFFORD, D. 1996a. Phylogenetic inference. In D. Hillis ed., Molecular Systematics, pp. 407-514. Sinauer Associates, Sunderland, MA.

_____. 1996b. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts.

TARKHNISHVILI, D., A. HILLE, and W. BOHME. 2001. Humid forest refugia, speciation and secondary introgression between evolutionary lineages: differentiation in a Near Eastern brown frog, *Rana macrocnemis*. Biological Journal of the Linnean Society 74: 141-156.

TEMPLETON, A.R., K.A. CRANDALL, and C.F. SING. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA Sequence Data, 3. Cladogram estimation. Genetics 132: 619-633.

THOMPSON, J.D., T.J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D.G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.

VANDERPOORTEN, A., A.J. SHAW, and C.J. COX. 2004. Evolution of multiple paralogous adenosine kinase genes in the moss genus *Hygroamblystegium*: phylogenetic implications. Molecular Phylogenetics and Evolution 31: 505-516.

VOGEL, J., T. BORNER, and W.R. HESS. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. Nucleic Acids Research 27: 3866-3874.

WELCH, T.A. 1984. A technique for high performance data compression. IEEE Computer 17: 8-19.

WHITTEN, W.M., N.H. WILLIAMS, and M.W. CHASE. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: Combined molecular evidence. American Journal of Botany 87: 1842-1856.

WILCOX, T.P. 2005. SG Runner: A Mac OS X shell-wrapper for Seq-Gen. http://homepage.mac.com/tpwilcox/SGRUNNER/FileSharing8.html, Long Key, FL.

WOODARD, J.W. 1951. Some chromosome numbers in *Phalaenopsis*. American Orchid Society Bulletin 20: 356-358.

XU, S. 2000. Phylogenetic analysis under reticulate evolution. Molecular Biology & Evolution 17: 897-907.

XU, W., W.J. BRIGGS, J. PADOLINA, R. TIMME, W. LIU, R.C. LINDER, and D. MIRANKER. 2004. Using MoBIoS' scalable genome join to find conserved primer pair candidates between two genomes. Bioinformatics 20 suppl. 1: i355-i362.

Appendix A

TABLE A.2.1. List of taxa included in this study, their source, and voucher or accession number. Source abbreviations: Hort = commercial horticultural establishment; UF = University of Florida, Department of Botany; Selby = Marie Selby Botanical Garden, Sarasota, FL; NYBG = The New York Botanical Garden.

Taxa	Source	Voucher/Accession Number
Aerides multiflorum	Hort	Padolina A114 (TEX)
Amesiella philippinensis	UF	C 295
Ascocentrum ampullaceum	Hort	Padolina A141 (TEX)
Neofinetia falcata	Hort	Padolina A113 (TEX)
Paraphalaenopsis laycockii	Hort	Padolina A047 (TEX)
Paraphalaenopsis serpentilingua	Hort	Padolina A046 (TEX)
Phalaenopsis stobartiana	Hort	Padolina A131 (TEX)
Phalaenopsis amabilis	Hort	Padolina A007 (TEX)
Phalaenopsis amboinensis 1	Hort	Padolina A098 (TEX)
Phalaenopsis amboinensis 2	Hort	Padolina A099 (TEX)
Phalaenopsis amboinensis 3	Hort	Padolina A119 (TEX)
Phalaenopsis aphrodite 1	Hort	Padolina A008 (TEX)
Phalaenopsis aphrodite 2	Selby	1994-0178E
Phalaenopsis aphrodite 3	Hort	Padolina A117 (TEX)
Phalaenopsis bastianii	Hort	Padolina A028 (TEX)
Phalaenopsis bellina 1	Hort	Padolina A044 (TEX)
Phalaenopsis bellina 2	Selby	1994-0149B
Phalaenopsis bellina 3	Hort	Padolina A085 (TEX)
Phalaenopsis borneënsis	Hort	Padolina A107 (TEX)
Phalaenopsis braceana	Hort	Padolina A100 (TEX)
Phalaenopsis buyssoniana 1	Hort	Padolina A011 (TEX)
Phalaenopsis buyssoniana 2	Hort	Padolina A112 (TEX)
Phalaenopsis celebensis	Hort	Padolina A080 (TEX)
Phalaenopsis chibae	Hort	Padolina A001 (TEX)
Phalaenopsis cochlearis	Hort	Padolina A133 (TEX)
Phalaenopsis corningiana	Selby	1994-0177C
Phalaenopsis cornu-cervi 1	Hort	Padolina A081 (TEX)
Phalaenopsis cornu-cervi 2	Selby	EYMC 717-1
Phalaenopsis deliciosa	Hort	Padolina A010 (TEX)
Phalaenopsis doweryënsis 1	Hort	Padolina A097 (TEX)

Taxa	Source	Voucher/Accession Number
Phalaenopsis doweryënsis 2	Hort	Padolina A128 (TEX)
Phalaenopsis equestris 1	Hort	Padolina A013 (TEX)
Phalaenopsis equestris 2	Hort	Padolina A136 (TEX)
Phalaenopsis equestris var. rosea	Hort	Padolina A014 (TEX)
Phalaenopsis fasciata 1	Hort	Padolina A017 (TEX)
Phalaenopsis fasciata 2	Selby	1982-0137B
Phalaenopsis fasciata 3	Hort	Padolina A139 (TEX)
Phalaenopsis fimbriata	Hort	Padolina A103 (TEX)
Phalaenopsis floresensis	Hort	Padolina A086 (TEX)
Phalaenopsis fuscata	Hort	Padolina A092 (TEX)
Phalaenopsis gibbosa 2	Hort	Padolina A130 (TEX)
Phalaenopsis gigantea 1	Hort	Padolina A026 (TEX)
Phalaenopsis gigantea 2	NYBG	3044/94
Phalaenopsis hainanensis	Hort	Padolina A132 (TEX)
Phalaenopsis hieroglyphica 1	Hort	Padolina A029 (TEX)
Phalaenopsis hieroglyphica 2	Selby	1982-0140B
Phalaenopsis hieroglyphica 3	NYBG	2789/95
Phalaenopsis honghenensis 1	Hort	Padolina A095 (TEX)
Phalaenopsis honghenensis 2	Hort	Padolina A094 (TEX)
Phalaenopsis inscriptiosinensis	Hort	Padolina A105 (TEX)
Phalaenopsis javanica 1	Hort	Padolina A096 (TEX)
Phalaenopsis javanica 2	Hort	Padolina A122 (TEX)
Phalaenopsis kunstleri	Hort	Padolina A093 (TEX)
Phalaenopsis lueddemanniana 2	Hort	Padolina A104 (TEX)
Phalaenopsis lindenii	Hort	Padolina A088 (TEX)
Phalaenopsis lobbii 3	NYBG	NYBG
Phalaenopsis lobbii 1	Hort	Padolina A018 (TEX)
Phalaenopsis lobbii 2	UF	C 261
Phalaenopsis lowii 1	Hort	Padolina A019 (TEX)
Phalaenopsis lowii 2	Selby	2000-0202A
Phalaenopsis lowii 3	Hort	Padolina A118 (TEX)
Phalaenopsis lueddemanniana 1	Hort	Padolina A003 (TEX)
Phalaenopsis maculata	Hort	Padolina A082 (TEX)
Phalaenopsis mannii 1	Hort	Padolina A002 (TEX)
Phalaenopsis mannii 2	NYBG	NYBG
Phalaenopsis mariae 1	Selby	1992-0327C
Phalaenopsis mariae 2	Hort	Padolina A124 (TEX)
Phalaenopsis micholitzii 1	Hort	Padolina A041 (TEX)
Phalaenopsis micholitzii 2	Selby	1994-0157B
Phalaenopsis minus 1	Hort	Padolina A040 (TEX)

Taxa	Source	Voucher/Accession Number
Phalaenopsis minus 2	Hort	Padolina A030 (TEX)
Phalaenopsis modesta 1	Hort	Padolina A027 (TEX)
Phalaenopsis modesta 2	Selby	1994-0159A
Phalaenopsis modesta 3	Hort	Padolina A135 (TEX)
Phalaenopsis pallens 1	Hort	Padolina A083 (TEX)
Phalaenopsis pallens 2	Selby	1982-0131C
Phalaenopsis pallens 3	Hort	Padolina A123 (TEX)
Phalaenopsis pantherina	Hort	Padolina A089 (TEX)
Phalaenopsis parishii 1	UF	178 FLAS
Phalaenopsis parishii 2	Hort	Padolina A115 (TEX)
Phalaenopsis philippinensis 1	Hort	Padolina A009 (TEX)
Phalaenopsis philippinensis 2	NYBG	NYBG
Phalaenopsis pulcherrima 1	Hort	Padolina A006 (TEX)
Phalaenopsis pulcherrima 2	Hort	Padolina A012 (TEX)
Phalaenopsis pulchra 1	Hort	Padolina A050 (TEX)
Phalaenopsis pulchra 2	Hort	Padolina A079 (TEX)
Phalaenopsis sanderiana	Hort	Padolina A084 (TEX)
Phalaenopsis schilleriana 1	Hort	Padolina A015 (TEX)
Phalaenopsis schilleriana 2	NYBG	NYBG
Phalaenopsis sp.	Hort	Padolina A106 (TEX)
Phalaenopsis speciosa	Selby	1994-0155A
Phalaenopsis stobartiana 1	Hort	Padolina A087 (TEX)
Phalaenopsis stuartiana	Hort	Padolina A016 (TEX)
Phalaenopsis sumatrana	Hort	Padolina A110 (TEX)
Phalaenopsis taenialis	Hort	Padolina A101 (TEX)
Phalaenopsis tetraspis 1	Hort	Padolina A005 (TEX)
Phalaenopsis tetraspis 2	Hort	Padolina A138 (TEX)
Phalaenopsis thalebanii	Hort	Padolina A137 (TEX)
Phalaenopsis venosa	Hort	Padolina A045 (TEX)
Phalaenopsis violacea 1	Hort	Padolina A120 (TEX)
Phalaenopsis violacea 2	Hort	Padolina A129 (TEX)
Phalaenopsis violacea 3	Selby	1993-0221A
Phalaenopsis viridis 1	Hort	Padolina A121 (TEX)
Phalaenopsis viridis 2	Hort	Padolina A134 (TEX)
Phalaenopsis wilsonii 2	UF	331 FLAS
Phalaenopsis X*intermedia*	Hort	Padolina A042 (TEX)
Renanthera vietnamica	Hort	Padolina A140 (TEX)
Sarcoglyphis comberrii	UF	C 296

TABLE A.2.2. Character states and symbols traced onto the chloroplast phylogeny (for data matrix see Table 9).

Character (Character letter in Table A.2.3)	Character State	Character State Symbol
Genome Size – Binary (A)	<7.0 pg/2C	0
	>9.0 pg/2C	1
	Unknown	?
Geographic Distribution (B)	Mainland	0
	Island	1
Geographic Distribution (C)	Philippine	0
	Non-Philippine	1
Deciduous leaves (D)	Evergreen	0
	Deciduous	1
Pollinia Type (E)	2 cleft pollinia	0
	4 semiglobular pollinia	1
	4 globular pollinia	2
Markings on petals and sepals (F)	Flowers with bright spots, bars, or blotches	0
	Flowers without conspicuous markings	1
Forked callus (G)	None	0
	At least 1	1
Callus number (H)	1	0
	2	1
	3	2
	0	3
Lateral lobes of the lip with a raised tooth (I)	Present	0
	Absent	1
Saccate lip base (J)	Present	0
	Absent	1
Reflexed tepals (K)	Present	0
	Absent	1
Hooded anther bed (L)	Present	0
	Absent	1
Petal/sepal width (M)	Petals and sepals equal or subequal	0
	Petals much wider than sepals	1

TABLE A.2.3. Data matrix used to map characters onto combined chloroplast phylogeny. Character names and states defined in Table A.2.2.

	Characters												
Species	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
Aerides multiflorum	?	0	1	0	0	1	0	3	1	0	1	1	0
Phalaenopsis amabilis	?	0&1	0&1	0	0	1	0	0	1	1	1	1	1
P. amboinensis	1	0	1	0	0	0	1	1	0	1	1	1	0
Amesiella philippinensis	?	1	0	0	0	1	0	3	1	0	1	1	0
P. aphrodite	?	1	0	0	0	1	0	0	1	1	1	1	1
Ascocentrum ampullaceum	?	0	1	0	0	1	0	3	1	0	1	1	0
P. bastianii	?	0&1	0	0	0	0	1	1	0	1	1	1	0
P. bellina	1	1	1	0	0	0	1	1	0	1	1	1	0
P. borneensis	?	1	1	0	0	0	1	2	0	1	1	1	0
P. braceana	?	0	1	1	1	1	1	1	0	0	1	1	0
P. buyssoniana	?	0	1	0	2	1	0	0	1	1	0	1	0
P. celebensis	?	1	1	0	0	1	0	0	1	1	1	1	0
P. chibae	?	0	1	0	1	1	0	0	1	0	0	1	0
P. cochlearis	?	0	1	0	0	1	1	1	0	1	1	1	0
P. cornucervi	0	0&1	1	0	0	0	1	2	0	1	1	1	0
P. deliciosa	?	0&1	0&1	0	1	1	1	1	0	0	1	1	0
P. doweryensis	?	1	1	0	0	0	1	1	0	1	1	1	0
P. equestris var. equestris	0	1	0	0	0	1	0	0	1	1	1	1	0
P. equestris var. rosea	0	1	0	0	0	1	0	0	1	1	1	1	0
P. fasciata3	0	1	0	0	0	0	1	1	0	1	1	1	0
P. fimbriata	?	0&1	1	0	0	1	1	2	0	1	1	1	0
P. floresensis	?	1	1	0	0	1	1	1	0	1	1	1	0
P. fuscata	?	0&1	0&1	0	0	0	1	1	0	1	1	1	0
P. gibbosa	?	0	1	1	1	1	1	1	1	1	0	1	0
P. gigantea	0	1	1	0	0	0	1	1	0	1	1	1	0
P. hainanensis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. hieroglyphica	?	1	0	0	0	0	1	1	0	1	1	1	0
P. honghenensis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. inscriptiosinensis	?	1	1	0	0	0	1	1	0	1	1	0	0
P. intermedia	?	1	0	0	0	1	0	0	1	1	1	1	1
P. javanica	?	1	1	0	0	0	1	1	0	1	1	1	0
P. lindenii	?	1	0	0	0	1	0	0	1	1	1	1	0
P. lobbii	?	0	1	1	1	1	1	1	1	1	0	1	0
P. lowii	?	0	1	1	1	1	1	1	0	1	1	1	1
P. lueddemanniana	0	1	0	0	0	0	1	1	0	1	1	1	0
P. maculata	?	0&1	1	0	0	0	1	1	0	1	1	1	0

P. mannii	1	0	1	0	0	0	1	2	0	1	1	1	0
P. mariae	0	1	0&1	0	0	0	1	1	0	1	1	1	0
P. micholitzii	0	1	0	0	0	1	1	1	0	1	1	1	0
P. minus	?	0	1	1	1	0	1	1	0	0	0	1	0
P. modesta	0	1	1	0	0	1	1	1	0	1	1	1	0
Neofinetia falcata	0	0	1	0	0	1	0	3	1	0	0	1	0
P. pallens	?	1	0	0	0	0	1	1	0	1	1	1	0
P. pantherina	?	0&1	1	0	0	0	1	2	0	1	1	1	0
Paraphalaenopsis laycockii	?	1	1	0	0	1	0	0	1	1	1	1	0
Paraphalaenopsis serpentilingua	?	1	1	0	0	1	0	0	1	1	1	1	0
P. parishii	1	0	1	1	1	1	1	1	1	1	0	1	0
P. philippinensis	?	1	0	0	0	1	0	0	1	1	1	1	1
P. pulcherrima	1	0	1	0	2	1	0	1	1	1	0	1	0
P. pulchra1	0	1	0	0	0	1	1	2	0	1	1	1	0
P. pulchra2	0	1	0	0	0	1	1	2	0	1	1	1	0
Renanthera vietnamica	?	0	1	0	0	0	0	0	1	0	1	1	0
P. sanderiana	0	1	0	0	0	1	0	0	1	1	1	1	1
Sarcoglyphis comberii	?	0	1	0	0	1	0	0	1	1	1	1	0
P. schilleriana	?	1	0	0	0	1	0	0	1	1	1	1	1
P. sp	?	?	?	?	?		?	?	?	?	?	?	?
P. stobartiana	?	0	1	1	1	1	1	1	0	0	1	1	0
P. stuartiana	0	1	0	0	0	1	0	0	1	1	1	1	1
P. sumatrana	0	0&1	1	0	0	0	1	1	0	1	1	0	0
P. taenailis	?	0	1	1	1	1	1	1	0	0	1	1	0
P. tetraspis	?	0	1	0	0	1	1	2	0	1	1	0	0
P. thalebanii	?	0	1	0	0	0	1	2	0	1	1	1	0
P. venosa	1	1	1	0	0	0	1	1	0	1	1	1	0
P. violacea	?	0&1	1	0	0	1	1	1	0	1	1	1	0
P. viridis	?	1	1	0	0	0	1	1	0	1	1	1	0
P. wilsonii	?	0	1	1	1	1	1	1	0	0	1	1	0

TABLE A.2.4. Parsimony statistics for analyses of data from the chloroplast and nuclear genomes. Trees for these analyses can be seen in Figures A.2.1 – A.2.11.

Marker	Number of characters	Number of parsimony informative characers	Number of trees saved	Length of trees	CI	HI	RI
Combined chloroplast data	3564	256	80,004	821	0.72	0.28	0.83
*atp*HF	394	31	100,000	99	0.74	0.25	0.81
*trn*DE	564	51	100,000	134	0.72	0.28	0.86
*pet*D	1271	86	70,000	248	0.78	0.22	0.88
*mat*K	1335	97	100,000	324	0.71	0.29	0.81
Chloroplast coding regions (not including the petD group II intron)	2240	151	100,000	501	0.72	0.28	0.80
Chloroplast noncoding regions (including the petD group II intron)	1324	108	288	317	0.74	0.27	0.88
Chloroplast coding regions (including the petD group II intron)	2942	20	23,654	655	0.73	0.27	0.83
Chloroplast noncoding regions (not including the petD group II intron)	622	53	164	165	0.72	0.28	0.85
Combined nuclear data	827	268	35,889	792	0.76	0.24	0.82
Nuclear coding regions	307	48	100,000	195	0.84	0.16	0.67
Nuclear noncoding regions	520	214	100,000	821	0.76	0.24	0.81

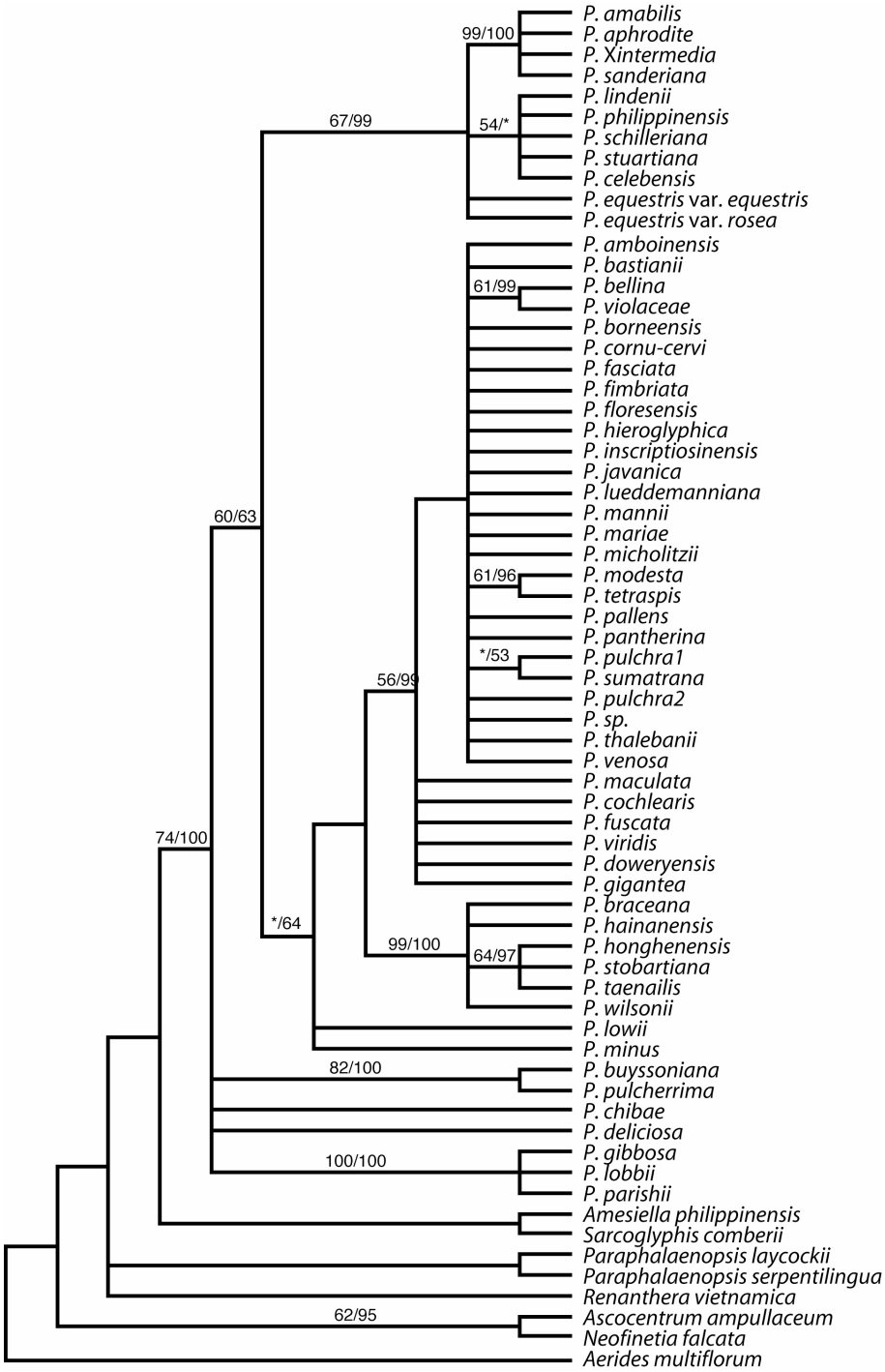


FIGURE A.2.1. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* using the *atp*HF chloroplast data. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

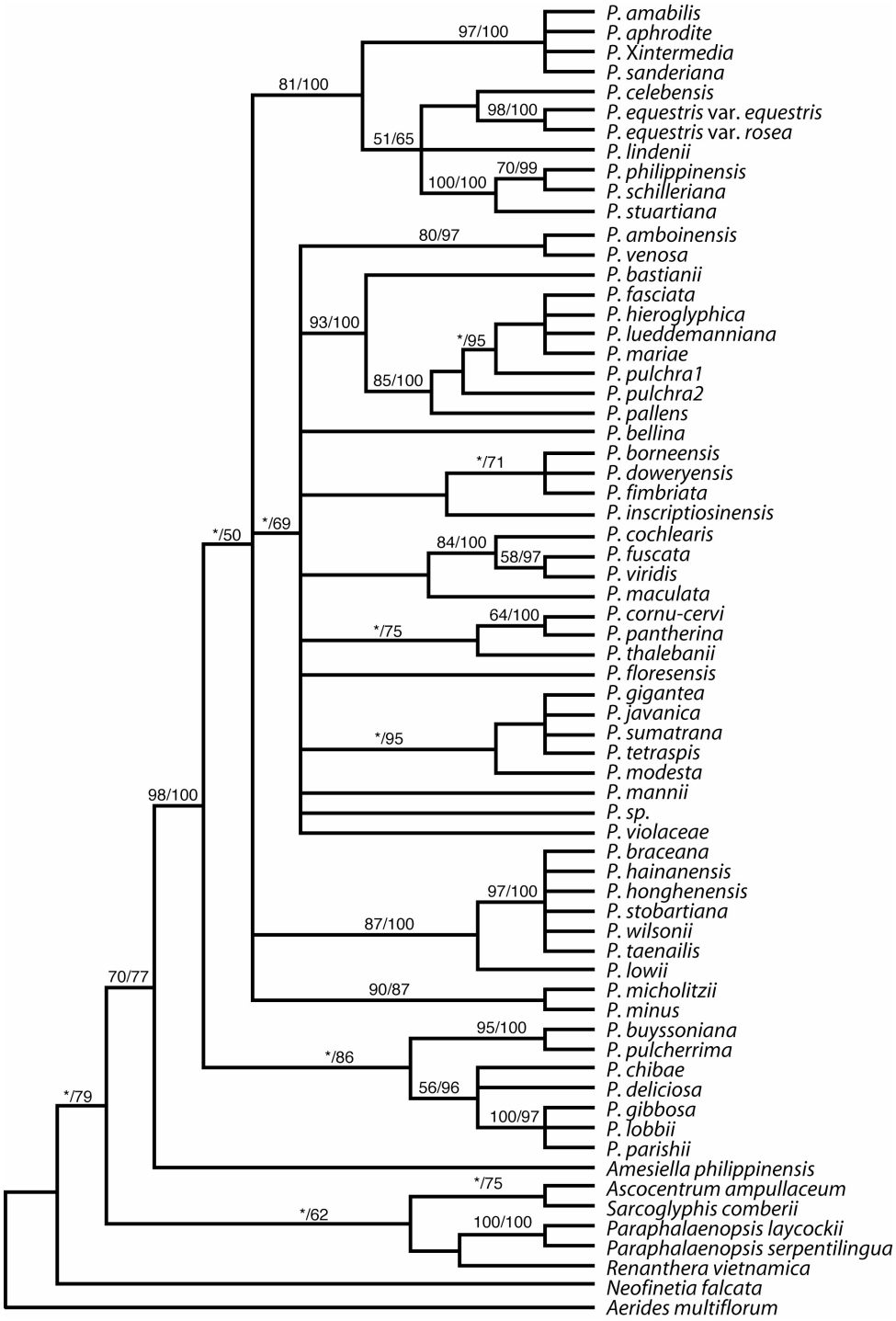


FIGURE A.2.2. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* using the *trn*DE chloroplast data. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

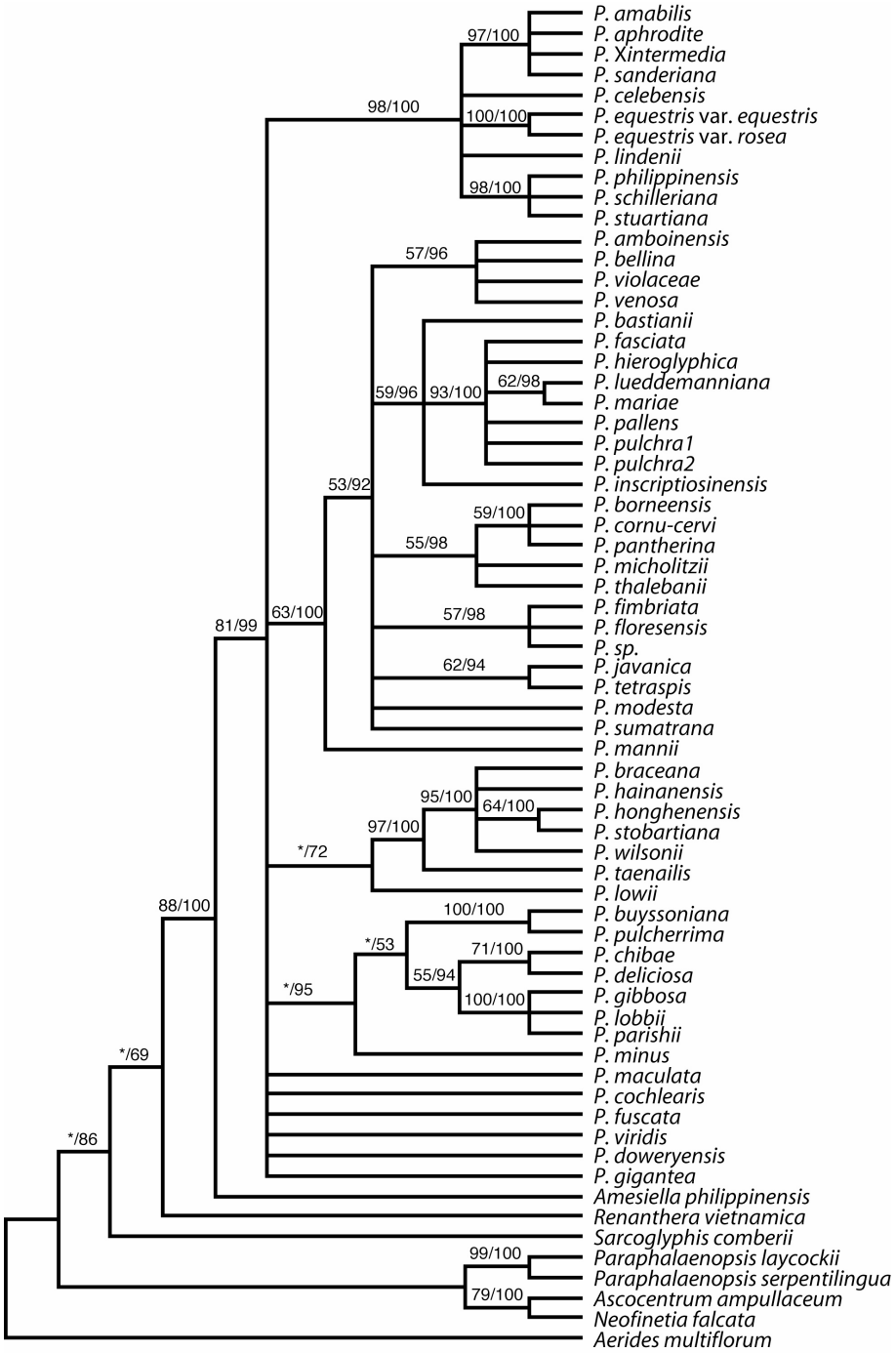


FIGURE A.2.3. Majority rule consensus tree of 70,000 most parsimonious trees found for *Phalaenopsis* using the *pet*D chloroplast data. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

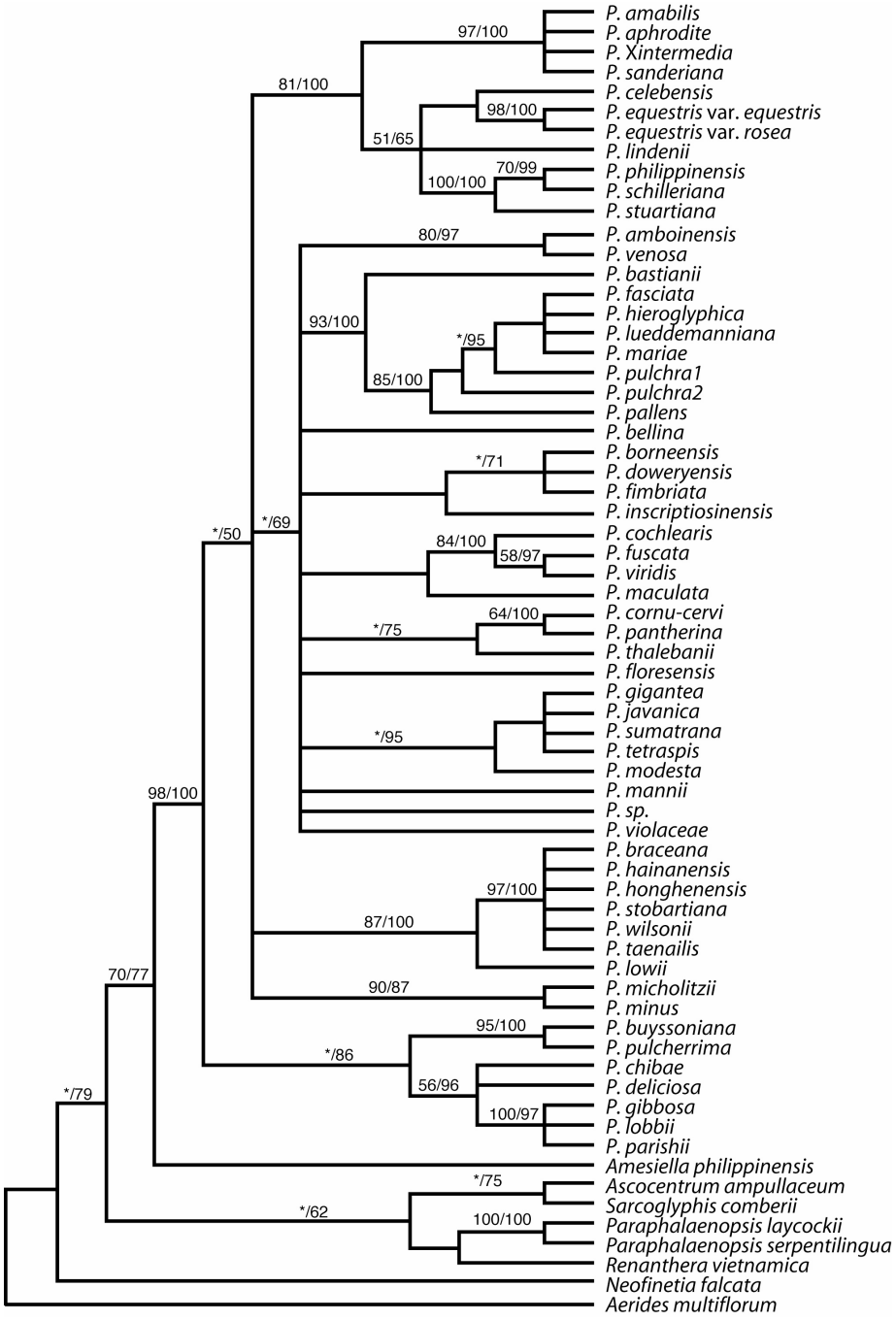


FIGURE A.2.4. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* using the *mat*K chloroplast data. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

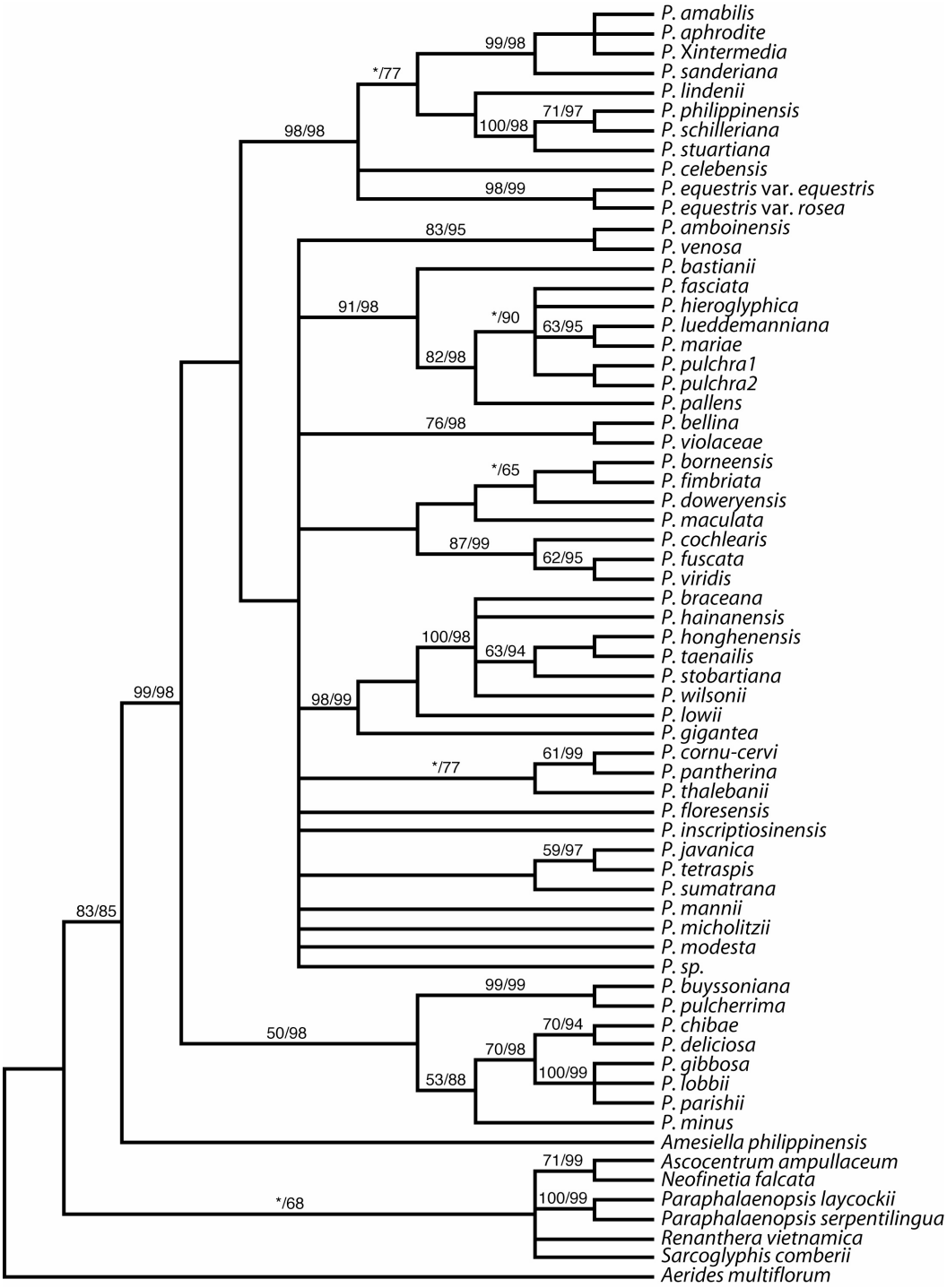


FIGURE A.2.5. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the chloroplast coding regions (I). Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

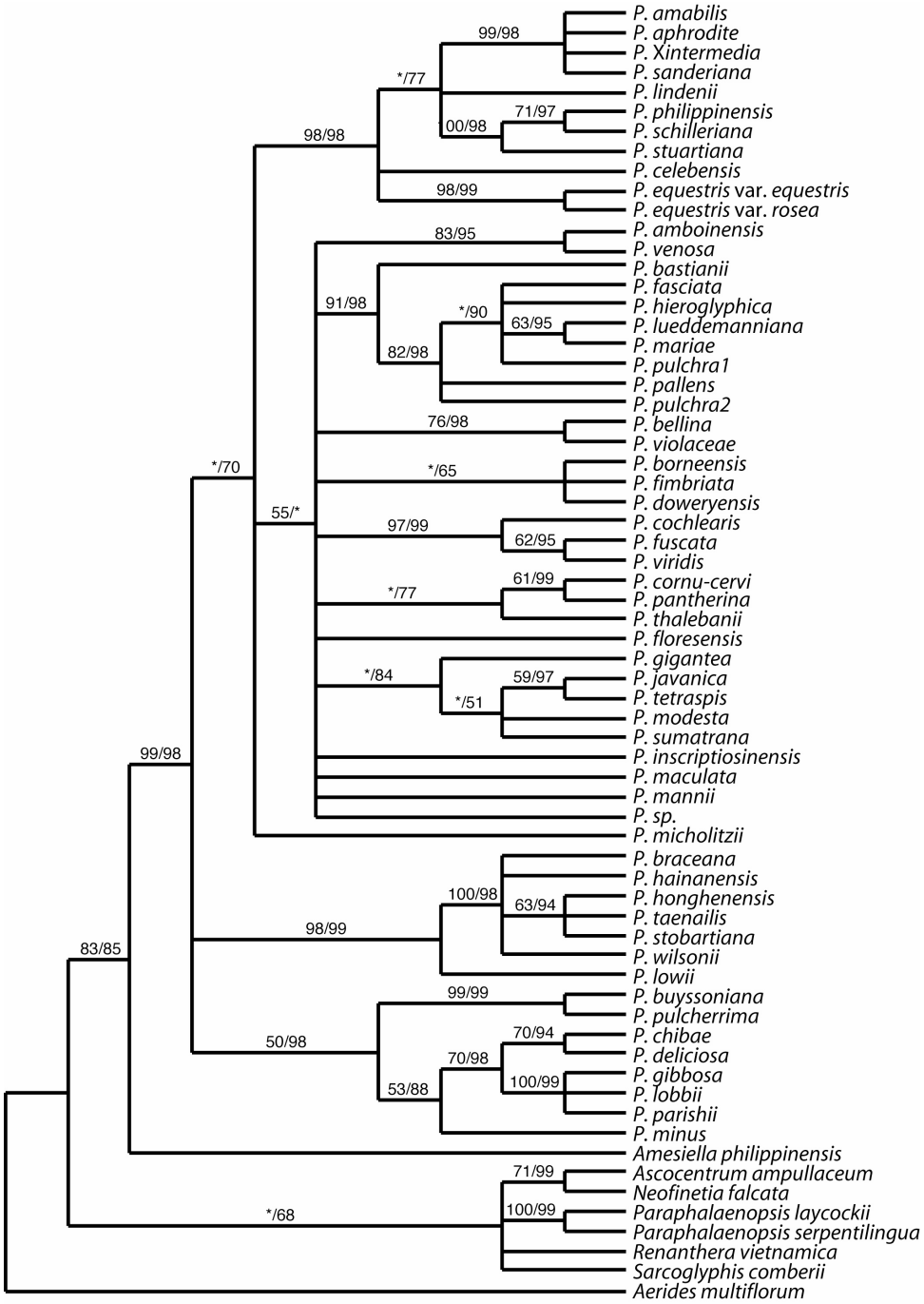


FIGURE A.2.6. Majority rule consensus tree of trees found for *Phalaenopsis* analyzed under Bayesian analyses (first 25,000 generations discarded as burn-in period) using the chloroplast coding regions (I). Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

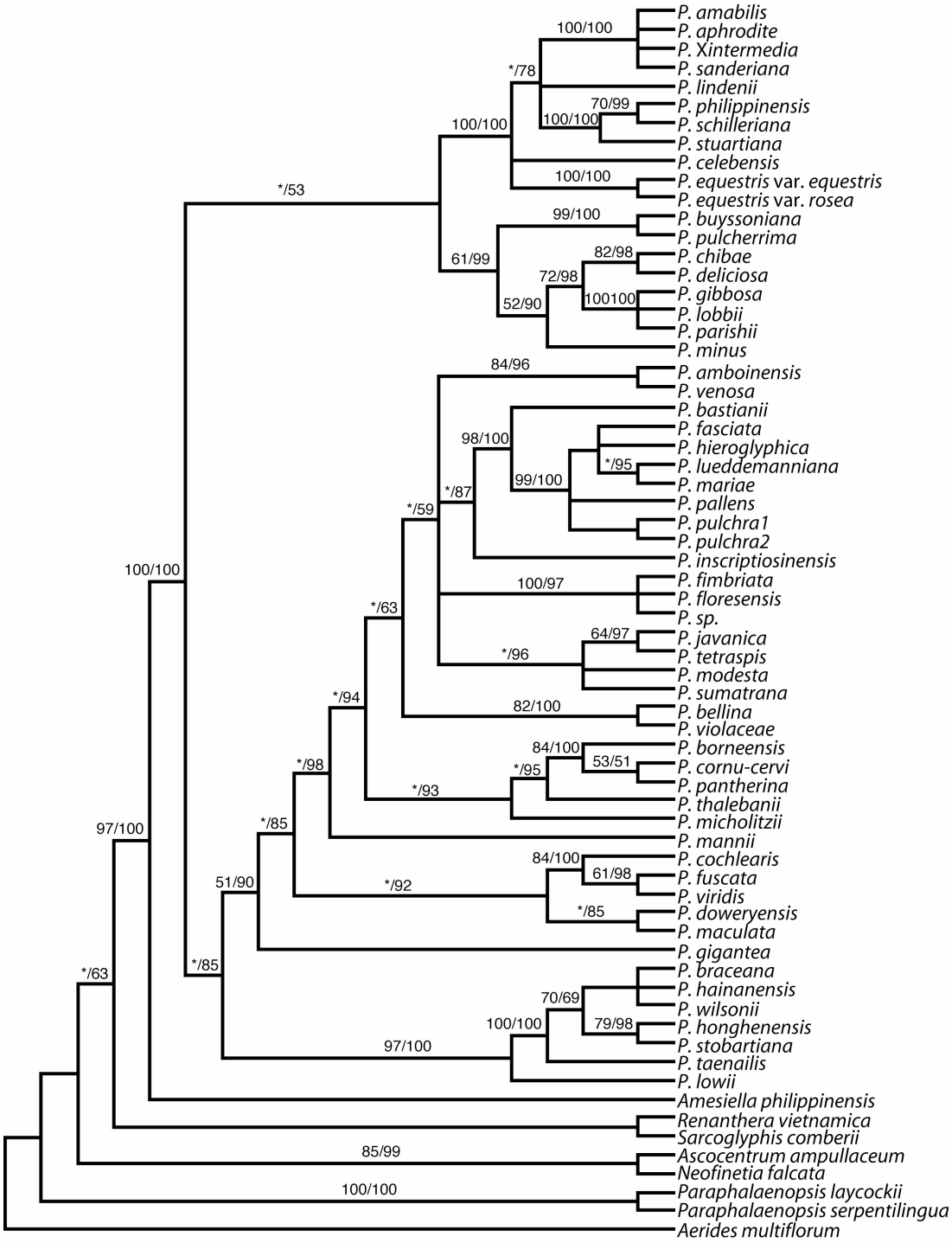


FIGURE A.2.7. Majority rule consensus tree of 23,654 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the chloroplast coding regions (II). Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

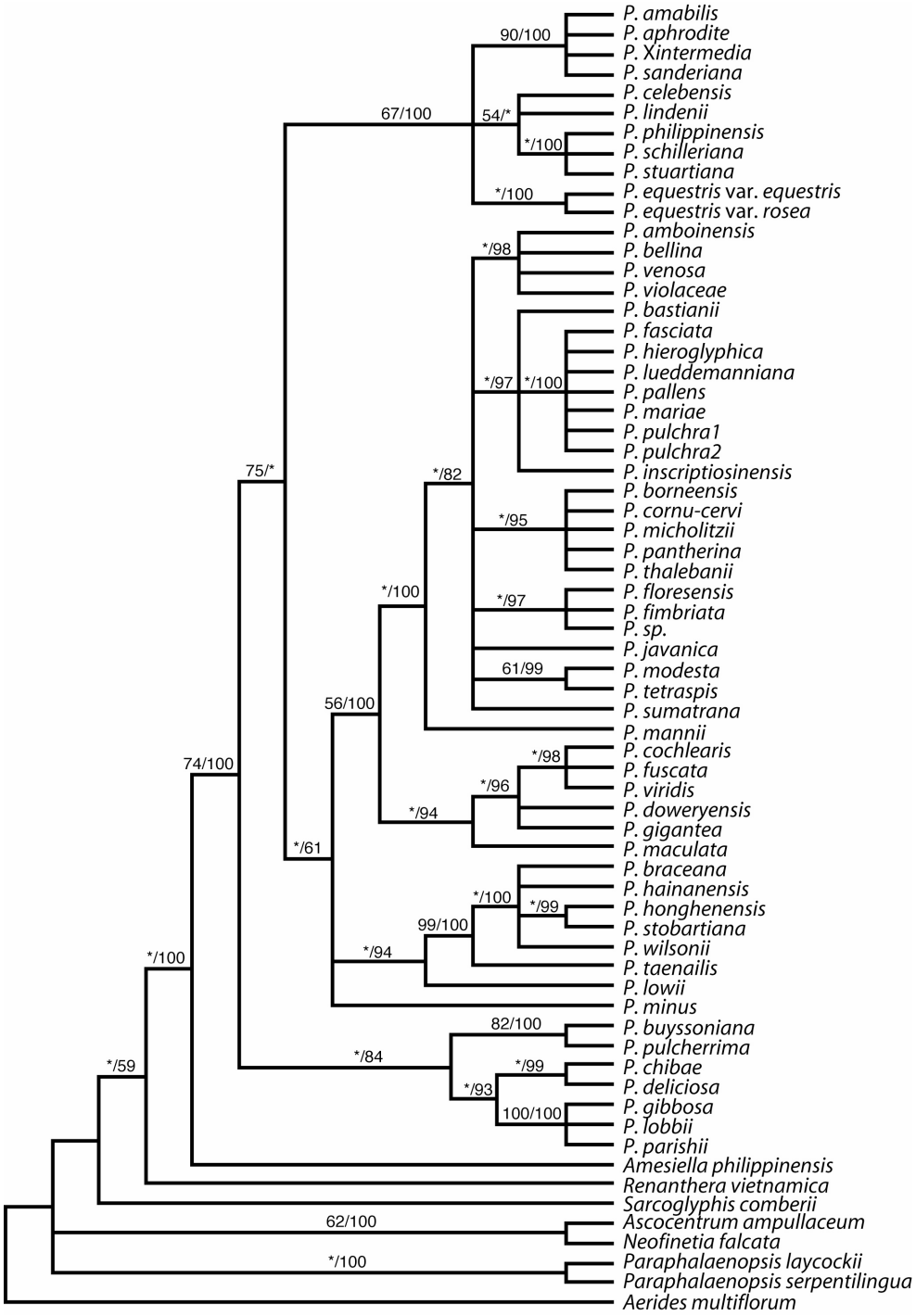


FIGURE A.2.8. Majority rule consensus tree of 288 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the chloroplast noncoding regions (I). Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

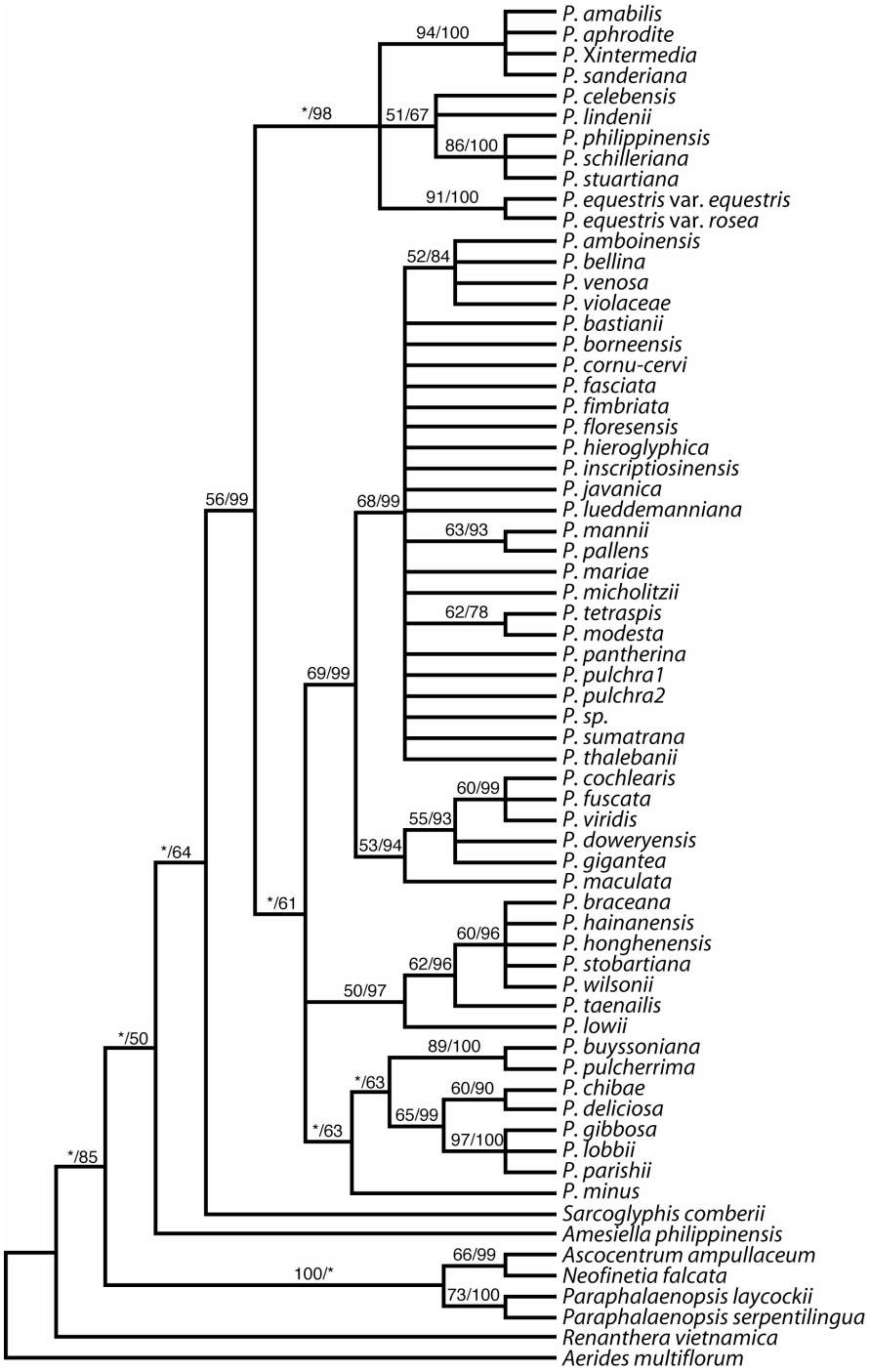


FIGURE A.2.9. Majority rule consensus tree of 164 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the chloroplast noncoding regions (II). Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis.

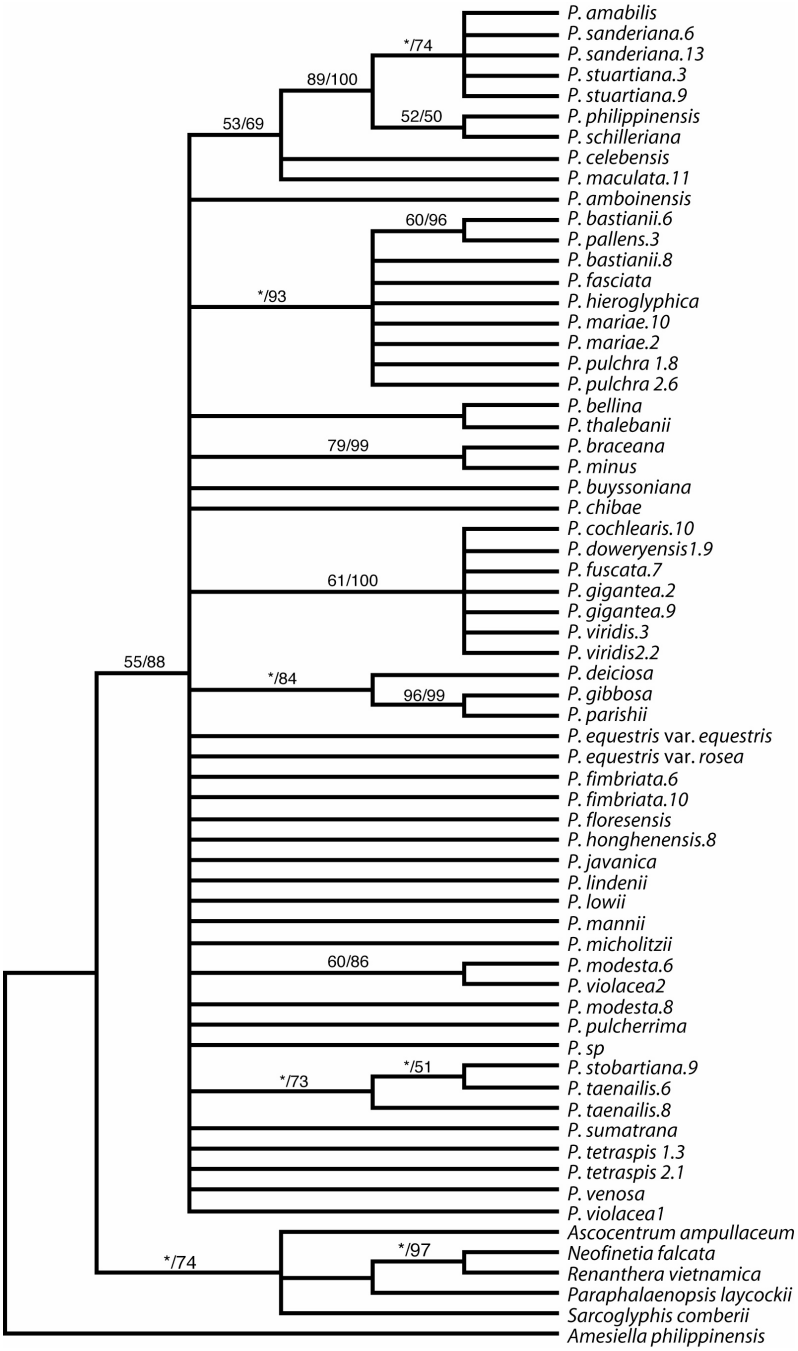


FIGURE A.2.10. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the nuclear actin gene coding regions. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis. A number following a '.' indicates individuals with more than one sequence included in analyses.

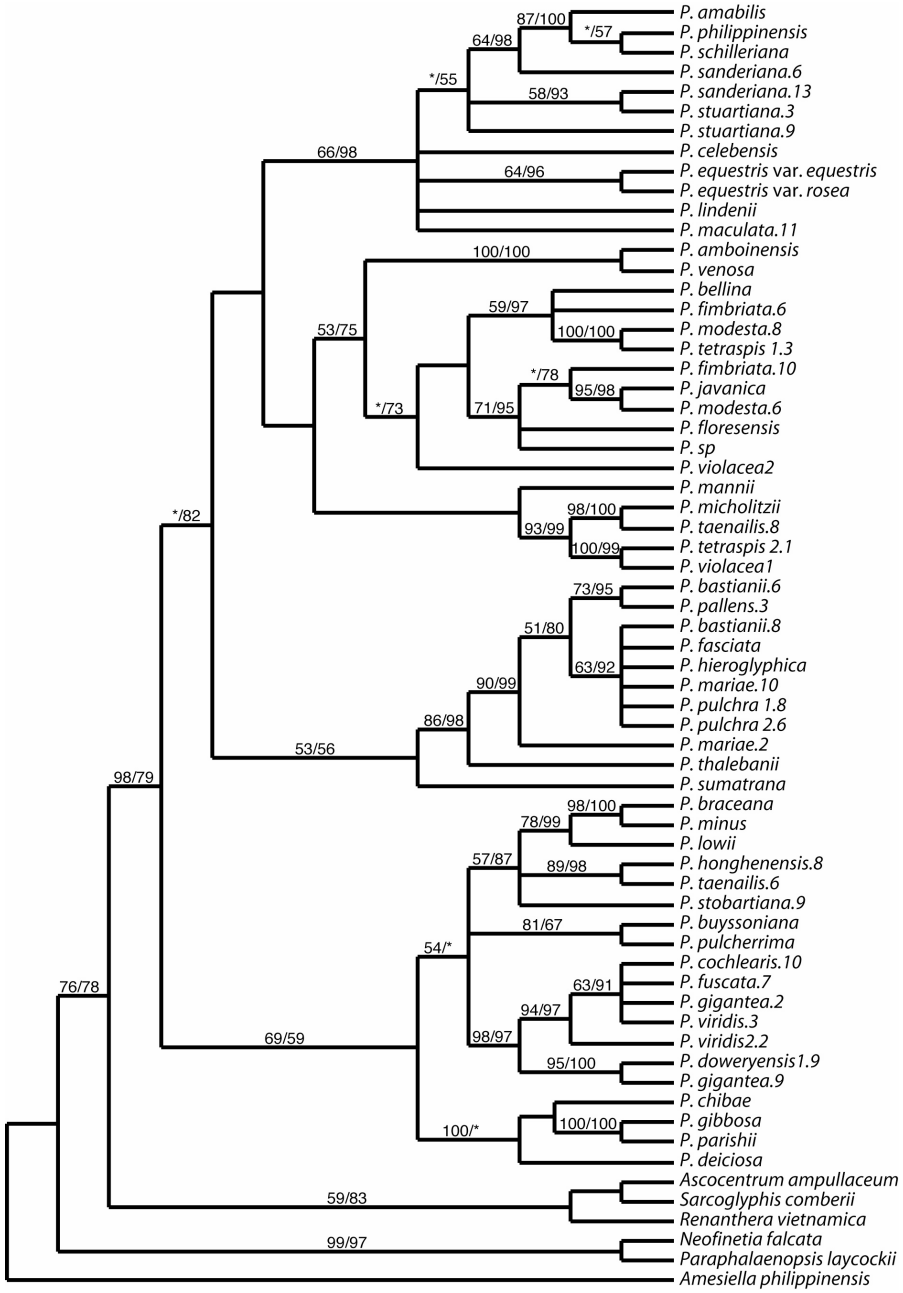


FIGURE A.2.11. Majority rule consensus tree of 100,000 most parsimonious trees found for *Phalaenopsis* analyzed under parsimony using the nuclear actin gene noncoding regions. Bootstrap/Bayesian posterior probabilities are indicated above branches. A number directly following the species name indicates species that have more than one individual included in the analysis. A number following a '.' Indicates individuals with more than one sequence included in analyses.

Appendix B

TABLE B.3.1. Amplification results. Overall success and successes for only single-band amplifications as visualized by gel electrophoresis are tallied for each primer in *Phalaenopsis, Helianthus*, and the six other Angiosperms: *Peperomia, Aquilegia, Geranium, Arabidopsis, Reinwardtia,* and Tiquilia. See Table 3.3 for information on each primer.

ID #	Primer Name	Amplification in Phalaenopsis (out of 6 possible successes)		Amplification in Helianthus (out of 6 possible successes)		Amplification in Others (out of 6 possible successes)	
		Overall Success	Single Bands	Overall Success	Single Bands	Overall Success	Single Bands
1	1F/1R	4	0	6	5		
2	2F/2R			5	5		
7	3F/3R						
35	4F/4R	5	0				
47	5F/5R						
48	6F/6R	6	6				
51	7F/7R						
56	8F/8R						
57	9F/9R						
63	10F/10R						
69	11F/11R					1	1
72	12F/12R					1	1
74	13F/13R	2	0			1	1
75	14F/14R					1	0
77	15F/15R	4	0			2	2
78	16F/16R	1	1				
80	17F/17R						
87	18F/18R	2	1				
88	19F/19R						
91	20F/20R						
93	21F/21R	2	2				
97	22F/22R						
99	23F/23R	4	0				
6	24F/24R						
8	25F/25R						
10	26F/26R						
11	27F/27R						
12	28F/28R					1	1
15	29F/29R			3	3		
24	30F/30R	5	5	4	4	2	2
43	31F/31R						
49	32F/32R					1	1
50	33F/33R						
52	34F/34R	4	3				
53	35F/35R						
54	36F/36R			6	6		

ID #	Primer Name	Amplification in Phalaenopsis (out of 6 possible successes)		Amplification in Helianthus (out of 6 possible successes)		Amplification in Others (out of 6 possible successes)	
		Overall Success	Single Bands	Overall Success	Single Bands	Overall Success	Single Bands
55	37F/37R	3	1			1	1
59	38F/38R	3	3				
60	39F/39R						
61	40F/40R			1	1	1	1
76	41F/41R						
81	42F/42R	1	1			1	1
83	43F/43R						1
90	44F/44R	4	1				
92	45F/45R						
109	46F/46R	5	0			4	0
110	47F/47R						
114	48F/48R						1
115	49F/49R	6	1				
117	50F/50R	6	1	6	6	1	1
120	51F/51R	6	6	5	5	1	1
122	52F/52R	5	0				
126	53F/53R					1	1
127	54F/54R					1	1
130	55F/55R	5	0				
131	56F/56R	6	1	4	4	1	1
139	57F/57R	6	6	4	4	1	1
140	58F/58R	1	1			4	4
143	59F/59R						
144	60F/60R						
150	61F/61R	5	5	3	3	3	3
151	62F/62R	1	1				
152	63F/63R	6	6				
153	64F/64R	6	0				
157	65F/65R					2	2
161	66F/66R					2	2
163	67F/67R			1	1	2	2
165	68F/68R	4	0			1	1
166	69F/69R	4	2			1	1
167	70F/70R	6	5			1	1
168	71F/71R					1	1
172	72F/72R	6	0				
176	73F/73R	2	1			1	1
179	74F/74R						
180	75F/75R						
185	76F/76R						
186	77F/77R	3	0				
285	78F/78R					2	2
282	79F/79R						
287	80F/80R					2	2
293	81F/81R					2	2

ID #	Primer Name	Amplification in Phalaenopsis (out of 6 possible successes)		Amplification in Helianthus (out of 6 possible successes)		Amplification in Others (out of 6 possible successes)	
		Overall Success	Single Bands	Overall Success	Single Bands	Overall Success	Single Bands
301	82F/82R					2	2
368	83F/83R	5	5			1	1
339	84F/84R					1	1
209	85F/85R			4	4	5	5
379	86F/86R	6	6	5	5	5	5
222	87F/87R	4	4			3	3
279	88F/88R						
348	89F/89R						
349	90F/90R	2	2				
262	91F/91R	5	4			3	3
203	92F/92R					1	1
207	93F/93R	1	1			1	1
212	94F/94R	6	5			1	1
220	95F/95R	1	1				
255	96F/96R					1	1
327	97F/97R						
360	98F/98R					1	1
371	99F/99R					1	1
264	100F/100R						
265	101F/101R						
266	102F/102R					1	1
272	103F/103R					1	1
288	104F/104R						
290	105F/105R					1	1
299	106F/106R						
308	107F/107R						
313	108F/108R						
323	109F/109R						
352	110F/110R						
355	111F/111R	1	1				
361	112F/112R	1	1				
369	113F/113R						
386	114F/114R						
393	115F/115R						
231	116F/116R						
396	117F/117R						
331	118F/118R						
303	119F/119R						
297	120F/120R						
240	121F/121R						
294	122F/122R						
239	123F/123R						
121	124F/124R						
122	125F/125R						
133	126F/126R						

ID #	Primer Name	Amplification in Phalaenopsis (out of 6 possible successes)		Amplification in Helianthus (out of 6 possible successes)		Amplification in Others (out of 6 possible successes)	
		Overall Success	Single Bands	Overall Success	Single Bands	Overall Success	Single Bands
134	127F/127R						
142	128F/128R						
145	129F/129R						
154	130F/130R						
156	131F/131R						
159	132F/132R						
162	133F/133R						
164	134F/134R						
169	135F/135R						
170	136F/136R						
173	137F/137R						
174	138F/138R						
178	139F/139R						
181	140F/140R						
183	141F/141R						
184	142F/142R						
	Totals	171	90	57	56	78	73
	Percent	20.1%	10.6%	6.7%	6.6%	9.2%	8.6%
		Combined success of overall amplification		Combined success rate of amplification of single bands			
	Total	306		219			
	Percent	12.0%		8.6%			

References

ARCHAMBAULT, A. and A. BRUNEAU. 2004. Phylogenetic utility of the LEAFY/FLORICAULA gene in the Caesalpinioideae (Leguminosae): Gene duplication and a novel insertion. Systematic Botany 29: 609-626.

BAACK, E.J. 2005. To succeed globally, disperse locally: effects of local pollen and seed dispersal on tetraploid establishment. Heredity 94: 538-546.

BAILEY, C.D., T.G. CARR, S.A. HARRIS, and C.E. HUGHES. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics & Evolution 29: 435-455.

BALDWIN, B.G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Molecular Phylogenetics and Evolution 1: 3-16.

BANDELT, H.J. and A.W.M. DRESS. 1992. A canonical decomposition-theory for metrics on a finite-set. Advances in Mathematics 92: 47-105.

BANDELT, H.J., P. FORSTER, and A. ROHL. 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16: 37-48.

BANDELT, H.J., V. MACAULAY, and M. RICHARDS. 2000. Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Molecular Phylogenetics and Evolution 16: 8-28.

BENNET, M.D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London 181: 109-135.

BENSON, D.A., I. KARSCH-MIZRACHI, D.J. LIPMAN, J. OSTELL, and D.L. WHEELER. 2004. GenBank: update. Nucleic Acids Research 32: D23-D26.

BENTHAM, G. 1883. *Phalaenopsis*. In G. Bentham and J.D. Hooker eds., Genera Plantarum, pp. L. Reeve & Co., London, England.

BENZING, D.H., W.E. FRIEDMAN, G. PETERSON, and A. RENFROW. 1983. Shootlessness, Velamentous Roots, And The Pre-Eminence Of Orchidaceae In The Epiphytic Biotope. American Journal Of Botany 70: 121-133.

BLUME, C.L. 1825. *Phalaenopsis*. Bijdragen Tot de Flora van Nederlandsch Indie 7: 294.

BRYANT, D. and V. MOULTON. 2002. NeighborNet: An agglomerative method for the construction of planar phylogenetic networks, Algorithms in Bioinformatics, Proceedings, pp. 375-391.

_____. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. Molecular Biology and Evolution 21: 255-265.

CAMERON, K.M., M.W. CHASE, W.M. WHITTEN, P.J. KORES, D.C. JARRELL, V.A. ALBERT, T. YUKAWA, H.G. HILLS, and D.H. GOLDMAN. 1999. A phylogenetic analysis of the Orchidaceae: Evidence from rbcL nucleotide sequences. American Journal of Botany 86: 208-224.

CHRISTENSON, E.A. 1995. An overview of the genus *Phalaenopsis*. Orchid Digest Jan-Feb-Mar: 19-22.

_____. 2001. *Phalaenopsis*: a Monograph. Timber Press, Portland, Oregon.

CRAWFORD, D.J. 1985. Electrophoretic data and plant speciation. Systematic Botany 10: 405-416.

DE WIT, H.C.D. 1977. Orchids in Rumphius' Herbarium Amboinense. In J. Arditti ed., Orchid Biology: Reviews and Perspectives, pp. 47-94. Cornell University Press, Ithaca.

DEMESURE, B., N. SODZI, and R.J. PETIT. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4: 129-131.

DOWTON, M. and A.D. AUSTIN. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - The behavior of the incongruence length difference test in mixed-model analyses. Systematic Biology 51: 19-31.

DOYLE, J.J. and J.L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19: 11-15.

DOYLE, J.J., J.L. DOYLE, and C. HARBISON. 2003. Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: Phylogeny, gene duplication, and ancient polyploidy. Systematic Botany 28: 567-577.

EVANS, B.J., J. SUPRIATNA, and D.J. MELNICK. 2001. Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. Evolution 55: 1686-1702.

EVANS, G. and H. REES. 1971. Mitotic cycles in dicotyledons and monocotyledons. Nature 233.

EXCOFFIER, L. and P.E. SMOUSE. 1994. Using allele frequencies and geographic subdivision to reconstructG gene trees within a species - Molecular variance parsimony. Genetics 136: 343-359.

FARRIS, J.S., M. KALLERSJO, A.G. KLUGE, and C. BULT. 1994. Testing Significance Of Incongruence. Cladistics-The International Journal Of The Willi Hennig Society 10: 315-319.

FU, Y.M., W.H. CHEN, W.T. TSAI, Y.S. LIN, M.S. CHYOU, and Y.H. CHEN. 1997. Phylogenetic studies of taxonomy and evolution among wild species of Phalaenopsis by random amplified polymorphic DNA markers. Report of the Taiwan Sugar Research Institute 0: 27-42.

FUNK, V.A. 1985. Phylogenetic patterns and hybridization. Annals of the Missouri Botanical Gardens 72: 681-715.

FUTUYMA, D.J. 1997. Evolutionary Biology. Sinauer Assoc. Inc., Sunderland, MA.

GALLEZ and L.D. GOTTLIEB. 1982. Genetic evidence for the hybrid origin of the diploid plant Stephanomeria diegensis. Evolution 36: 1158-1167.

GRANT, V. 1966. Selection for vigor and fertility in the progeny of a hihly sterile species hybrid in *Gillia*. Genetics 53.

_____. 1981. Plant Speciation. Columbia University Press, New York.

GRIESBACH, R.J. 2002. Development of Phalaenopsis Orchids for the Mass Market. ASHS Press, Alexandria, VA.

GRUSS, O. and L. ROLLKE. 1993a. Gattung *Kingidium*. Die Orchidee 44: 33-36.

_____. 1993b. Gattung *Kingidium*, Teil 2. Die Orchidee 44: 119-122.

_____. 1995a. *Kingidium wightii* (Rchb.f.) Gruss & Rollke comb. nov., Teil 10. Die Orchidee 46: 23-28.

_____. 1995b. *Kingidium philippinense* (Ames) Gruss & Rollke comb. nov., Teil 11. Die Orchidee 46: 115-118.

_____. 1996. *Kingidium wilsonii* (Rolfe) Gruss & Rollke comb. nov. Die Orchidee 47: 149-155.

_____. 1997. *Kingidium chibae* (T. Yukawa) Gruss & Rollke comb. nov. Die Orchidee 48: 261-263.

HAAHR, M. 2006. random.org - True random number service. Distributed Systems Group, Department of Computer Science. University of Dublin, Trinity College.

HAMZEH, M. and S. DAYANANDAN. 2004. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast tRNT-tRNF region and nuclear rDNA. American Journal of Botany 91: 1398-1408.

HAWKES, A.D. 1966. *Grafia*, a new genus of orchids. Phytologia 13: 305-306.

HEIN, J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. Mathematical Biosciences 98: 185-200.

_____. 1993. A heuristic method to reconstruct the history of sequences subject to recombination. Journal of Molecular Evolution 36: 396-405.

HENNIG, W. 1966. Phylogenetic Systematics. University of Illinois Press, Urbana, Illinois.

HILLIS, D.M., C. MORITZ, and B.K. MABLE. 1996. Molecular Systematics. Sinauer Associates, Inc., Sunderland, MA.

HOLTTUM, R.E. 1965. Cultivated species of the orchid genus *Doritis* Lindl. Kew Bulletin 19: 207-212.

_____. 1966. The genus *Kingiella*. The Orchid Review 74: 128-129.

HUELSENBECK, J.P. and F. RONQUIST. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

HUNT, P.F. 1970. Notes on Asiatic Orchids: V. *Kingidium* P. F. Hunt, nom. nov. Kew Bulletin 24: 97-98.

HUSON, D.H. 1998. SplitsTree: A program for analyzing and visualizing evolutionary data. Bioinformatics 14: 68-73.

HUSON, D.H. and D. BRYANT. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

HUSON, D.H., T. KLOPPER, P.J. LOCKHART, and M.A. STEEL. 2005. Reconstruction of reticulate networks from gene trees, Research in Computational Molecular Biology, Proceedings, pp. 233-249.

HUYNH, T.N.D., J. JANSSON, N.B. NGUYEN, and W.K. SUNG. 2005. Constructing a smallest refining galled phylogenetic network, Research In Computational Molecular Biology, Proceedings, pp. 265-280.

JONES, W.E., A.R. KUEHNLE, and K. ARUMUGANATHAN. 1998. Nuclear DNA content of 26 orchid (Orchidaceae) genera with emphasis on *Dendrobium*. Annals Of Botany 82: 189-194.

KAO, Y.Y., S.B. CHANG, T.Y. LIN, C.H. HSIEH, Y.H. CHEN, W.H. CHEN, and C.C. CHEN. 2001. Differential accumulation of heterochromatin as a cause for karyotype variation in *Phalaenopsis* orchids. Annals Of Botany 87: 387-395.

LANDAIS, I., M. OGLIASTRO, K. MITA, J. NOHATA, M. LOPEZ-FERBER, M. DUONOR-CERUTTI, T. SHIMADA, P. FOUMIER, and G. DEVAUCHELLE. 2003. Annotation pattern of ESTs from *Spodoptera frugiperda* Sf9 cells and analysis of the ribosoma protein genes reveal insect-specific features and unexpectedly low codon usage bias. Bioinformatics 19: 2243-2250.

LEGENDRE, P. and V. MAKARENKOV. 2002. Reconstruction of biogeographic and evolutionary networks using reticulograms. Systematic Biology 51: 199-216.

LEVIN, D.A. 2000. The Origin, Expansion and Demise of Plant Species. Oxford University Press, New York.

LIN, S., H.C. LEE, W.H. CHEN, C.C. CHEN, Y.Y. KAO, Y.M. FU, Y.H. CHEN, and T.Y. LIN. 2001. Nuclear DNA contents of *Phalaenopsis sp.* and *Doritis pulcherrima*. Journal of the American Society for Horticultural Science 126: 195-199.

LINDER, C.R. and L.H. RIESEBERG. 2004. Reconstructing patterns of reticulate evolution in plants. American Journal of Botany 91: 1700-1708.

LINDER, C.R., B.M.E. MORET, L. NAKHLEH, J. PADOLINA, J. SUN, A. THOLSE, R. TIMME, and T. WARNOW. 2003. An error metric for phylogenetic networks, Technical Report TR-CS-2003-2026, pp. University of New Mexico, Albuquerque, NM.

LINDLEY, J. 1833. *Doritis*, Genera and Species of Orchidaceous Plants, pp. 178. Ridgways, Piccadilly, London.

LINNAEUS, C. 1753. *Epidendrum amabile*. Species Plantarum ed. 1: 953-954.

LOHNE, C. and T. BORSCH. 2005. Molecular evolution and phylogenetic utility of the *pet*D group II intron: A case study in basal angiosperms. Molecular Biology & Evolution 22: 317-332.

LOYTYNOJA, A. and M.C. MILINKOVITCH. 2001. Molecular phylogenetic analyses of the mitochondrial ADP-ATP carriers: The Plantae/Fungi/Metazoa trichotomy revisited. Proceedings of the National Academy of Sciences of The United States of America 98: 10202-10207.

MADDISON, W.P. 1997. Gene trees in species trees. Systematic Biology 46: 523-536.

_____. 2005. Pairwise comparisons package for mesquite. http://mesquiteproject.org.

MADDISON, W.P. and D.R. MADDISON. 2000. MacClade 4.0. Sinauer, Sunderland, Massachusetts.

_____. 2005. Mesquite: A modular system for evolutionary analysis. http://mesquiteproject.org.

MAIER, R.M., K. NECKERMANN, G.L. IGLOI, and H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome - gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. Journal of Molecular Biology 251: 614-628.

MAIER, R.M., K. NECKERMANN, G.L. IGLOI, and H. KOSSEL. 1995. Complete sequence of the maize chloroplast genome - gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. Journal of Molecular Biology 251: 614-628.

MAKARENKOV, V. and P. LEGENDRE. 2004. From a phylogenetic tree to a reticulated network. Journal of Computational Biology 11: 195-212.

MASTERSON, J. 1994. Stomatal Size In Fossil Plants - Evidence For Polyploidy In Majority Of Angiosperms. Science 264: 421-424.

MCDADE, L. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44: 1685-1700.

_____. 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46: 1329-1346.

MIRANKER, D., W. XU, and R. MAO. 2003. MoBIoS: a metric-space DBMS to support biological diversity. Proceedings of the International Conference on Scientific and statistical Database Management Systems.

MORRISON, D.A. 2005. Networks in phylogenetic analysis: new tools for population biology. International Journal for Parasitology 35: 567-582.

MULLIS, K.B. and F.A. FALOONA. 1987. Specific synthesis of DNA in vitro via polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335-350.

NAKHLEH, L., D. RUTHS, and H. INNAN. 2005. Gene trees, species trees, and species networks. Chapman & Hall, Norwell, MA.

NAKHLEH, L., J. SUN, T. WARNOW, C.R. LINDER, B.M.E. MORET, and A. THOLSE. 2003. Towards the development of computational tools for evaluating phylogenetic network reconstruction methods. Proceedings of the Eighth Pacific Symposium on Biocomputing (PSB03): 315-326.

NAKHLEH, L., T. WARNOW, C.R. LINDER, and K. ST. JOHN. 2004. Reconstructing Reticulate Evolution in Species - Theory and Practive. Proceedings of the Eighth Annual International Conference on Research in Computational Molecular Biology: 337-346.

OKUYAMA, Y., N. FUJII, M. WAKABAYASHI, A. KAWAKITA, M. ITO, M. WATANABE, N. MURAKAMI, and M. KATO. 2005. Nonuniform concerted evolution and chloroplast capture: Heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). Molecular Biology and Evolution 22: 285-296.

ORTI, G., D.E. PEARSE, and J.C. AVISE. 1997. Phylogenetic assessment of length variation at a microsatellte locus. Proceedings of the National Academy of Sciences of the United States of America 94: 10745-10749.

OTTO, S.P. and J. WHITTON. 2000. Polyploid incidence and evolution. Annual Review Of Genetics 34: 401-437.

PFITZER. 1889. Orchidaceae: *Phalaenopsis* Bl. In A. Engler and K. Prantl eds., Die natürlichen Pflanzenfamilien, Teil II, pp. 211-212. Verlag von Wilhelm Engelmann, Leipzig.

POSADA, D. and K.A. CRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

_____. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology & Evolution 16: 37-45.

RAMBAUT, A. and N.C. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Computer Applications in the Biosciences 13: 235-238.

REICHENBACH, H.G. 1862. Uebersicht der Gattungen *Phalaenopsis, Doritis, Stauropsis*. Xenia Orchidacea: 1-8, t. 101; 139-147, t. 151, t. 156.

_____. 1864. Neue Orchideen, *Phalaenopsis pantherina*. Botanische Zeitung 39: 298.

_____. 1879. New species of *Phalaenopsis*. The Gardeners Chronicle n.s., 11: 398.

RIESEBERG, L.H. 1991. Homoploid Reticulate Evolution in *Helianthus* (Asteraceae): Evidence from Ribosomal Genes. American Journal of Botany 78: 1218-1237.

_____. 1997. Hybrid origins of plant species. Annual Review of Ecology and Systematics 28: 359-389.

RIESEBERG, L.H. and N.C. ELLSTRAND. 1993. What can molecular and morphological markers tell us about plant hybridization? Critical Reviews in Plant Science 12: 213-241.

ROBBINSON, D.R. and L.R. FOULDS. 1981. Comparison of phylogenetic trees. Mathematical Biosciences 53: 131-147.

ROLFE, R.A. 1886. A revision of the genus *Phalaenopsis*. The Gardeners Chronicle 26: 168-170; 212-213; 267.

_____. 1917. *Doritis pulcherrima*. The Orchid Review 25: 195-197.

SAGAWA, Y. 1962. Cytological studies of the genus *Phalaenopsis*. American Orchid Society Bulletin 31: 390-398.

SANG, T. 2002. Utiliy of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology 37: 121-147.

SATTATH, S. and A. TVERSKY. 1977. Additive Similarity Trees. Psychometrika 42: 319-345.

SEIDENFADEN, G. 1988. Orchid Genera in Thailand XIV. Fifty-nine vandoid genera. Opera Botanica 95: 31-34; 182-189; 236-241.

SHAW, J., E.B. LICKEY, J.T. BECK, S.B. FARMER, W. LIU, J. MILLER, K.C. SIRIPUN, C.T. WINDER, E.E. SCHILLING, and R.L. SMALL. 2005. The Tortoise and the Hare II: Relative Utility of 21 Noncoding Chloroplast DNA Sequences for Phylogenetic Analysis. American Journal of Botany 92: 142-166.

SHIM, P.S. 1982. A new generic classification in the *Phalaenopsis*-complex (Orchidaceae). Malayan Nature Journal 36: 1-28.

SHINDO, N. and H. KAMEMOTO. 1963. Karyotype analysis of some species of Phalaenopsis. Cytologia 28: 390-398.

SMITH, J.J. 1933. Enumeration of the Orchidaceae of Sumatra and neighboring islands: *Phalaenopsis*. Repertorium specierum novarum regni vegetabilis ("Feddes Repertorium") 32: 364-367.

SNEL, B., P. BORK, and M.A. HUYNEN. 2002. Genomes in flux: The evolution of archaeal and proteobacterial gene content. Genome Research 12: 17-25.

SOLTIS, P.S., D.E. SOLTIS, and J.J. DOYLE. 1998. Molecular Systematics of Plants II: DNA Sequencing. Chapman and Hall, New York.

STEPHENS, P.F. 2005. Angiosperm Phylogeny Website. Version 6. http://www.mobot.org/MOBOT/research/APweb/.

STREELMAN, J.T., R. ZARDOYA, A. MEYER, and S.A. KARL. 1998. Multilocus phylogeny of cichlid fishes (Pisces: Perciformes): Evolutionary comparison of microsatellite and single-copy nuclear loci. Molecular Biology And Evolution 15: 798-808.

SUDA, J., T. KYNCL, and V. JAROLIMOVA. 2005. Genome size variation in Macaronesian angiosperms: forty percent of the Canarian endemic flora completed. Plant Systematics And Evolution 252: 215-238.

SWEET, H.R. 1968. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 37: 867-877, 1089-1104.

_____. 1969. A revision of the genus *Phalaenopsis*. American Orchid Society Bulletin 38: 33-43, 225-239, 321-336, 505-519, 681-694, 888-901.

_____. 1970. *Phalaenopsis* Section *Parishianae*. American Orchid Society Bulletin 39: 1094-1097.

_____. 1980. The Genus *Phalaenopsis*. Orchid Digest, Pomona, CA.

SWOFFORD, D. 1996a. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusetts.

_____. 1996b. Phylogenetic inference. In D. Hillis ed., Molecular Systematics, pp. 407-514. Sinauer Associates, Sunderland, MA.

TARKHNISHVILI, D., A. HILLE, and W. BOHME. 2001. Humid forest refugia, speciation and secondary introgression between evolutionary lineages: differentiation in a Near Eastern brown frog, *Rana macrocnemis*. Biological Journal of the Linnean Society 74: 141-156.

TEMPLETON, A.R., K.A. CRANDALL, and C.F. SING. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA Sequence Data, 3. Cladogram estimation. Genetics 132: 619-633.

THOMPSON, J.D., T.J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D.G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882.

VANDERPOORTEN, A., A.J. SHAW, and C.J. COX. 2004. Evolution of multiple paralogous adenosine kinase genes in the moss genus *Hygroamblystegium*: phylogenetic implications. Molecular Phylogenetics and Evolution 31: 505-516.

VOGEL, J., T. BORNER, and W.R. HESS. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. Nucleic Acids Research 27: 3866-3874.

WELCH, T.A. 1984. A technique for high performance data compression. IEEE Computer 17: 8-19.

WHITTEN, W.M., N.H. WILLIAMS, and M.W. CHASE. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: Combined molecular evidence. American Journal of Botany 87: 1842-1856.

WILCOX, T.P. 2005. SG Runner: A Mac OS X shell-wrapper for Seq-Gen. http://homepage.mac.com/tpwilcox/SGRUNNER/FileSharing8.html, Long Key, FL.

WOODARD, J.W. 1951. Some chromosome numbers in *Phalaenopsis*. American Orchid Society Bulletin 20: 356-358.

XU, S. 2000. Phylogenetic analysis under reticulate evolution. Molecular Biology & Evolution 17: 897-907.

XU, W., W.J. BRIGGS, J. PADOLINA, R. TIMME, W. LIU, R.C. LINDER, and D. MIRANKER. 2004. Using MoBIoS' scalable genome join to find conserved primer pair candidates between two genomes. Bioinformatics 20 suppl. 1: i355-i362.

Vita

Joanna Melinda Padolina, known to her friends and family as Anneke, was born in Gainesville, FL on February 11, 1978 to Dr. Stephen G. Schulman and Mrs. Johanna F. Tjeenk Willink. She was raised in Gainesville with her two sisters, Barbara and Christina Schulman. Anneke attended the honors program at The University of North Florida for one semester before she transferred to The University of Florida. She graduated from The University of Florida *summa cum laude*, earning the degree of Bachelors in Science in Botany. She worked at The Plant Shoppe for three years while she was in college. Immediately after her graduation in 2000, Anneke moved to Austin, TX to begin her Ph.D. program in botany at The University of Texas at Austin. In 2003, she taught Plant Biology at St. Edward's University in Austin, TX. Anneke met her husband, Isagani Padolina, in the botany department at The University of Texas. They were married in Austin on March 8, 2003, and their first son, Liam Isaac, was born on December 16, 2005.

Permanent address: 1501 Pritchard Terrace, Richmond, VA 23235

This dissertation was typed by Joanna M. Padolina.