



POSISI FILOGENETIK DAN HUBUNGAN KEKERABATAN ANGGREK ENDEMIK SULAWESI (*Phalaenopsis venosa*) BERDASARKAN SEKUEN GEN *rbcl* SECARA *IN SILICO*

Phylogenetic Position and Genetic Relationships of the Sulawesi Endemic Orchid (Phalaenopsis venosa) Based on rbcl Gene Sequences: An In Silico Study

Fernando Andre Watung^{1*}, dan Dwi Artaningsih Nazara¹

¹ Program Studi Biologi, Fakultas Matematika, Ilmu Pengetahuan Alam, dan Kebumian, Universitas Negeri Manado

*Fernandowatung@unima.ac.id

Abstrak

Phalaenopsis venosa merupakan spesies anggrek epifit endemik Sulawesi yang memiliki nilai konservasi dan ekonomi tinggi. Penelitian ini bertujuan melakukan analisis bioinformatika komprehensif terhadap marka gen *rbcl* (ribulose-1,5 bisphosphate carboxylase/oxygenase large subunit) pada *P. venosa* (aksesi MF861096.1) bersama enam spesies pembanding menggunakan perangkat lunak MEGA12. Sekuens DNA sebanyak 7 taksa dengan panjang 310 posisi nukleotida dianalisis melalui serangkaian metode meliputi statistik situs variabilitas, komposisi basa nitrogen, komposisi asam amino, degenerasi kodon, codon usage bias (RSCU), pemilihan model substitusi, estimasi jarak genetik, uji netralitas evolusi (Tajima's D dan Z-Test), serta rekonstruksi pohon filogenetik menggunakan lima metode algoritma (ML, MP, NJ, UPGMA, ME) dengan Uji Bootstrap 1000 replikasi. Hasil analisis menunjukkan marka *rbcl* memiliki 215 situs variabel (69,35%) dan 88 situs terkonservasi (28,38%), bersifat AT-Rich (A+T =64,3%), dengan dominasi leusin sebagai asam amino terbanyak (17,32%). Sebesar 57,10% situs bersifat zero-fold degenerate, mencerminkan tekanan seleksi purifikasi yang kuat. Bias penggunaan kodon nyata dengan kodon CCU (Prolin, RSCU =2,26) sebagai yang paling disukai. Model Tamura 3-parameter (T92) terpilih sebagai model terbaik (BIC = 2.335,579). Uji Tajima menghasilkan D = 0,6856, mengindikasikan evolusi netral. Kelima topologi pohon filogenetik secara konsisten menempatkan *P. venosa* dalam clade monofiletik *Phalaenopsis* dengan dukungan bootstrap 94-100%, mengkonfirmasi validitas taksonomi spesies ini

Keywords: Bioinformatika, DNA Barcoding, Filogenetik Orchidaceae, *P. venosa*, *rbcl*

Abstract

Phalaenopsis venosa is an epiphytic orchid species endemic to Sulawesi with high conservation and economic value. This study aimed to perform a comprehensive bioinformatics analysis of the *rbcl* gene marker (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) in *P. venosa* (GenBank accession MF861096.1) together with six related species using MEGA12 software. DNA sequences from seven taxa, comprising 310 nucleotide positions, were analyzed using a series of bioinformatics approaches, including variable-site statistics, nucleotide base composition, amino acid composition, codon degeneracy, relative synonymous codon usage (RSCU), substitution model selection, genetic distance estimation, evolutionary neutrality tests (Tajima's D and Z-test), and phylogenetic tree reconstruction employing five algorithms: Maximum Likelihood (ML), Maximum Parsimony (MP), Neighbor-Joining (NJ), Unweighted Pair Group Method with Arithmetic Mean (UPGMA), and Minimum Evolution (ME), with 1,000 bootstrap replicates. The results revealed that the *rbcl* marker contained 215 variable sites (69.35%) and 88 conserved sites (28.38%). The sequences were AT-rich (A + T = 64.3%), with leucine being the most abundant amino acid (17.32%). A total of 57.10% of the codon positions were zero-fold degenerate, indicating strong purifying selection. Codon usage bias was evident, with CCU (proline; RSCU = 2.26) identified as the most preferred codon. The Tamura 3-parameter (T92) model was selected as the best-fitting nucleotide substitution model (BIC = 2,335.579). Tajima's D test yielded a value of D = 0.6856, indicating neutral evolution. All five phylogenetic reconstruction methods consistently placed *P. venosa* within a monophyletic *Phalaenopsis* clade with strong bootstrap support (94–100%), thereby confirming the taxonomic validity of this endemic orchid species.

Keywords: Bioinformatics; DNA barcoding; Orchidaceae phylogenetics; *Phalaenopsis venosa*; *rbcl*.

Latar Belakang

Keanekaragaman hayati Indonesia, khususnya flora di Sulawesi, mencerminkan tingkat endemisitas yang sangat tinggi dalam famili Orchidaceae. Anggrek merupakan salah satu famili tumbuhan berbunga terbesar di dunia dengan lebih dari 25.000 spesies yang tersebar di berbagai penjuru dunia (Zhang *et al.* 2023). Sulawesi sebagai pulau dengan biogeografi yang unik menjadi rumah bagi sejumlah spesies anggrek yang tidak ditemukan di tempat lain, termasuk *Phalaenopsis venosa* Shim & Fowlie sebuah spesies epifit endemik dengan nilai konservasi dan ekonomi tinggi.

Identifikasi spesies *P. venosa* secara tradisional sering menghadapi kendala akibat kemiripan morfologi antarspesies dalam genus *Phalaenopsis*. Pendekatan molekuler melalui DNA barcoding telah terbukti menjadi solusi yang efektif dan akurat untuk mengidentifikasi serta menggambarkan hubungan kekerabatan antarspesies anggrek (Li *et al.* 2021). DNA barcoding memanfaatkan fragmen gen terstandar yang dapat diperkuat secara universal dan memiliki variasi cukup untuk diskriminasi taksa.

Gen *rbcl* (ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit) merupakan gen kloroplas yang mengkodekan subunit besar enzim Rubisco, enzim kunci dalam siklus Calvin pada proses fotosintesis. Gen ini telah ditetapkan sebagai salah satu marka standar DNA barcoding tumbuhan oleh CBOL Plant Working Group bersama gen *matK*. Keunggulan *rbcl* meliputi: ketersediaan primer universal yang mudah diamplifikasi, tingkat substitusi nukleotida yang moderat untuk diskriminasi tingkat genus hingga spesies, serta representasi database GenBank yang sangat luas (Niu *et al.* 2024). Analisis pola penggunaan kodon pada gen ini juga menunjukkan keterkaitan erat antara komposisi basa, seleksi translasi, dan adaptasi fungsional protein Rubisco di dalam stroma kloroplas (Yang *et al.* 2023).

Perangkat lunak MEGA12 (Molecular Evolutionary Genetics Analysis versi 12) menyediakan kerangka analisis genetika evolusi molekuler yang komprehensif dan efisien, mencakup penyelarasan sekuens, pemilihan model substitusi, estimasi jarak genetik, uji netralitas evolusi, hingga rekonstruksi pohon filogenetik dengan berbagai metode algoritma (Kumar *et al.* 2024). Kombinasi marka *rbcl* dengan analisis MEGA12 memungkinkan karakterisasi lengkap terhadap aspek molekuler suatu spesies target secara terintegrasi.

Penelitian ini bertujuan melakukan analisis bioinformatika komprehensif terhadap sekuens gen *rbcl* pada *Phalaenopsis venosa* (aksesi MF861096.1) beserta enam spesies pembanding yang terdiri dari in-group genus *Phalaenopsis* (*P. amabilis*, *P. schilleriana*, *P. pulcherrima*) dan out-group lintas genus (*Cymbidium lancifolium*, *Aerides odorata*, *Dendrobium officinale*), guna mengungkap karakteristik molekuler, pola evolusi, dan hubungan filogenetik anggrek endemik Sulawesi ini.

Metode

Data dan organisme target. Organisme target utama dalam analisis ini adalah *Phalaenopsis venosa* Shim & Fowlie dengan nomor aksesori NCBI MF861096.1. Data sekuens DNA gen *rbcl* dari tujuh spesies (Tabel 1) diperoleh dari pangkalan data NCBI GenBank (<https://www.ncbi.nlm.nih.gov>) dalam format FASTA. Pemilihan marka *rbcl* didasarkan pada statusnya sebagai standar DNA barcoding tumbuhan yang diakui secara internasional dengan resolusi tinggi pada tingkat genus dan spesies.

Tabel 1. Daftar nomor aksesori GenBank spesies yang digunakan dalam analisis

No.	Nomor Aksesori	Nama Spesies	Kelompok	Asal / Distribusi
1	MF861096.1	<i>Phalaenopsis venosa</i>	In-group (Target)	Sulawesi, Indonesia
2	MT519366.1	<i>Phalaenopsis amabilis</i>	In-group	Asia Tenggara
3	MT519344.1	<i>Phalaenopsis schilleriana</i>	In-group	Filipina
4	AB519783.1	<i>Phalaenopsis pulcherrima</i>	In-group	Asia Tenggara
5	MW480548.1	<i>Cymbidium lancifolium</i>	Out-group	Asia Selatan & Tenggara
6	AY389404.1	<i>Aerides odorata</i>	Out-group	Asia Tenggara
7	AY389425.1	<i>Dendrobium officinale</i>	Out-group	Cina, Asia Timur

Penyelarasan sekuens. Seluruh sekuens DNA diselaraskan menggunakan algoritma ClustalW yang tersedia pada MEGA12 dengan parameter gap opening penalty 15 dan gap extension penalty 6,66. Hasil alignment menghasilkan 310 posisi nukleotida (NSeqs = 7, NSites = 310) yang siap dianalisis lebih lanjut.

Analisis bioinformatika. Pengolahan data dilakukan menggunakan MEGA12 (Kumar et al. 2024) meliputi: (1) statistik situs nukleotida jumlah dan persentase *conserved sites*, *variable sites*, *parsimony-informative sites* (*Pi*), dan *singleton sites*; (2) komposisi basa nitrogen per posisi kodon; (3) komposisi asam amino melalui translasi in silico; (4) tingkat degenerasi kodon (*zero-fold*, *two-fold*, *four-fold*); (5) analisis codon usage bias melalui nilai *Relative Synonymous Codon Usage* (RSCU); (6) pemilihan model substitusi nukleotida terbaik berdasarkan nilai *Bayesian Information Criterion* (BIC) terendah; (7) estimasi jarak genetik menggunakan model Kimura 2-parameter; (8) uji netralitas evolusi melalui Tajima's Neutrality Test dan Z-Test Nei-Gojobori; serta (9) rekonstruksi pohon filogenetik menggunakan lima metode algoritma (Maximum Likelihood/ML, Maximum Parsimony/MP, Neighbor-Joining/NJ, UPGMA, dan Minimum Evolution/ME) dengan uji Bootstrap 1000 replikasi.

Hasil

Statistik situs nukleotida dan variabilitas gen *rbcl*

Hasil alignment sepanjang 310 posisi nukleotida menunjukkan tingkat variabilitas yang tinggi pada marka *rbcl*. Dari total 310 posisi, terdapat 88 situs terkonservasi (28,38%) yang identik pada seluruh spesies, 215 situs variabel (69,35%), 175 parsimony-informative sites/*Pi* (56,45%), 40 singleton sites (12,90%), dan 7 posisi gap/missing data (2,26%) (Tabel 2).

Tabel 2. Distribusi statistik situs nukleotida gen *rbcl* (310 posisi, 7 spesies)

Kategori Situs	Jumlah Situs	Persentase (%)	Keterangan
Conserved Sites	88	28,38%	Identik pada semua spesies
Variable Sites	215	69,35%	Mengalami variasi basa antar spesies
Parsimony-Informative (<i>Pi</i>)	175	56,45%	Informatif secara kelompok (≥ 2 spesies)
Singleton Sites (<i>S</i>)	40	12,90%	Variasi unik hanya pada 1 spesies
Gap / Missing Data	7	2,26%	Tidak dikategorikan oleh MEGA12
Total Posisi Alignment	310	100%	Seluruh posisi hasil alignment ClustalW

Hasil penyelarasan 310 posisi nukleotida gen *rbcl* mengungkapkan distribusi situs yang informatif secara filogenetik. Sebanyak 88 posisi (28,38%) tergolong sebagai situs terkonservasi, yakni posisi-posisi yang memperlihatkan identitas basa nitrogen yang seragam di seluruh takson yang dianalisis, baik pada kelompok ingroup maupun outgroup. Konservasi pada posisi-posisi tersebut mencerminkan adanya tekanan seleksi purifikasi yang berperan dalam mempertahankan integritas struktural dan fungsional enzim RuBisCO yang disandi oleh gen ini.

Di sisi lain, sebanyak 215 posisi (69,35%) bersifat variabel, menunjukkan terjadinya substitusi basa nitrogen pada setidaknya satu takson dalam dataset. Proporsi situs variabel yang tinggi ini mengindikasikan kemampuan diskriminasi gen *rbcl* yang substansial pada tingkat genus maupun antargenus dalam famili Orchidaceae, sebagaimana dikemukakan oleh Li et al. (2021) yang merekomendasikan gen ini sebagai penanda molekuler yang efektif untuk identifikasi anggrek. Perlu dicatat bahwa akumulasi situs terkonservasi dan variabel menghasilkan total 303 situs, dengan selisih tujuh posisi yang disebabkan oleh keberadaan gap atau missing data pada beberapa sekuens referensi.

Dari keseluruhan situs variabel, 175 posisi (56,45% dari total sekuens; 81,40% dari total situs variabel) terklasifikasi sebagai parsimony-informative sites (*Pi*), yaitu posisi yang menampilkan minimal dua karakter basa berbeda dengan frekuensi kemunculan masing-masing tidak kurang dari dua kali. Dominasi situs *Pi* terhadap total situs variabel mengindikasikan bahwa variasi genetik yang teramati lebih banyak bersifat informatif secara kelompok daripada sekadar noise acak, sehingga data ini memenuhi prasyarat esensial untuk menghasilkan rekonstruksi pohon filogenetik yang sah dan dapat dipertanggungjawabkan secara statistik.

Sebaliknya, hanya 40 posisi (12,90%) yang tergolong singleton sites, yakni variasi yang eksklusif pada satu sekuens tunggal dan tidak berkontribusi dalam analisis Maximum Parsimony. Rendahnya proporsi singleton dibandingkan situs *Pi* memperkuat kesimpulan bahwa diferensiasi genetik pada sampel yang diuji lebih mencerminkan pola divergensi antarkelompok yang terstruktur. Secara keseluruhan, profil variabilitas gen *rbcl*

dengan dominasi parsimony-informative sites (81,40%) dibandingkan singleton sites (18,60%) membuktikan reliabilitas penanda ini untuk kajian keanekaragaman genetik, filogeografi, serta rekonstruksi filogenetik pada tingkat genus dan spesies dalam famili Orchidaceae.

Komposisi basa nukleotida

Analisis komposisi basa menunjukkan bahwa gen *rbcl* bersifat AT-Rich dengan kandungan A+T mencapai 64,3% (T = 36,3%; A = 28,0%; G = 19,3%; C = 16,4%). Nilai frekuensi basa berdasarkan model T92 yang terpilih menunjukkan A+T = 65,34% dan G+C = 34,66%, konsisten dengan hasil analisis langsung (Tabel 3).

Tabel 3. Komposisi basa nukleotida gen *rbcl* berdasarkan posisi kodon

Posisi	T (U) %	C %	A %	G %	Total Posisi
Posisi Kodon ke-1	32,7	16,9	30,8	19,6	103
Posisi Kodon ke-2	38,5	15,4	24,3	21,8	103
Posisi Kodon ke-3	37,8	16,8	28,9	16,5	103
Rata-rata Seluruh Posisi	36,3	16,4	28,0	19,3	310
Frekuensi (T92 Model)*	32,67	17,33	32,67	17,33	—

*Frekuensi berdasarkan model substitusi T92 terpilih oleh MEGA12.

Analisis komposisi basa nitrogen gen *rbcl* pada seluruh sampel anggrek yang diuji memperlihatkan pola ketidakseimbangan yang nyata. Timin mendominasi dengan proporsi tertinggi (36,3%), diikuti Adenin (28,0%), Guanin (19,3%), dan Sitosin (16,4%). Akumulasi kedua basa tersebut menghasilkan kandungan A+T sebesar 64,3%, jauh melampaui kandungan G+C yang hanya 35,7%, sehingga gen *rbcl* pada famili Orchidaceae dapat dikategorikan sebagai AT-rich. Kondisi ini selaras dengan karakteristik umum genom kloroplas tumbuhan tingkat tinggi yang dipengaruhi oleh mekanisme replikasi dan perbaikan DNA di dalam stroma kloroplas (Yang et al., 2023; Fages-Lartaud et al., 2022).

Pengamatan pada masing-masing posisi kodon mengungkapkan pola yang lebih spesifik. Posisi kodon pertama menampilkan distribusi yang relatif proporsional antara Adenin (30,8%) dan Timin (32,7%), yang mengindikasikan bahwa variasi pada posisi ini masih berada dalam kendali seleksi. Posisi kodon kedua mencatat persentase Timin tertinggi di antara ketiga posisi (38,5%); kondisi ini bermakna secara fungsional mengingat perubahan basa pada posisi tersebut hampir senantiasa berujung pada pergantian asam amino (mutasi nonsinonim), sehingga posisi ini berada di bawah tekanan seleksi purifikasi yang ketat guna mempertahankan fungsi enzim RuBisCO. Adapun posisi kodon ketiga juga menunjukkan kandungan Timin yang tinggi (37,8%), namun substitusi basa pada posisi ini umumnya bersifat sinonim sehingga tidak mengubah asam amino yang dikode, suatu fenomena yang lazim dikaitkan dengan codon usage bias pada genom kloroplas.

Konsistensi antara hasil analisis langsung dan pemodelan statistik turut memperkuat temuan ini. Model substitusi T92 (Tamura 3-parameter) yang terpilih sebagai model terbaik memberikan estimasi komposisi A+T sebesar 65,34% dan G+C sebesar 34,66%, nilai yang selaras dengan komposisi basa yang diperoleh secara empiris dari penyelarasan sekuens.

Maximum Likelihood Estimate of Substitution Matrix				
	A	T/U	C	G
A	-	7.64	4.05	9.22
T/U	7.64	-	9.22	4.05
C	7.64	17.39	-	4.05
G	17.39	7.64	4.05	-

Gambar 1. Maximum Likelihood Estimate of Substitution Matrix (model T92, lnL = -1118.144)

Analisis lebih lanjut menggunakan pendekatan Maximum Likelihood berdasarkan model Tamura (1992) menghasilkan matriks estimasi laju substitusi antar basa nitrogen pada gen *rbcl* (Tabel/Gambar X). Hasil analisis menunjukkan bahwa laju transisi jauh lebih tinggi dibandingkan laju transversi, yang merupakan pola umum pada gen pengkode protein kloroplas. Laju substitusi transisi tertinggi tercatat pada pasangan C↔T/U dengan nilai 17,39, diikuti pasangan G↔A sebesar 9,22. Sebaliknya, laju transversi yang melibatkan pergantian antara basa purin dan pirimidin secara silang menunjukkan nilai yang lebih rendah, berkisar antara 4,05 hingga 7,64. Rasio transisi terhadap transversi (Ti/Tv) yang tinggi ini mengindikasikan bahwa evolusi gen *rbcl* pada famili

Orchidaceae lebih didominasi oleh perubahan basa sejenis, suatu kondisi yang konsisten dengan tekanan seleksi purifikasi pada gen fungsional kloroplas. Parameter frekuensi basa yang diestimasi oleh model T92 mencatat nilai A sebesar 32,67%, T/U sebesar 17,33%, C sebesar 17,33%, dan G sebesar 17,33%, dengan nilai log likelihood sebesar -1118,144. Keseluruhan analisis dilakukan terhadap 310 posisi nukleotida dari tujuh sekuens menggunakan perangkat lunak MEGA versi 12 (Kumar et al., 2024).

	A	T	C	G
A	-	7.81	3.45	9.69
T	6.04	-	10.15	4.04
C	6.04	22.97	-	4.04
G	14.5	7.81	3.45	-

Gambar 2. Maximum Composite Likelihood Estimate of Nucleotide Substitution Pattern

Estimasi pola substitusi nukleotida menggunakan pendekatan Maximum Composite Likelihood (MCL) menghasilkan matriks laju penggantian antarbasa yang disajikan pada Tabel/Gambar X. Frekuensi basa yang diestimasi melalui pendekatan ini mencatat nilai A sebesar 28,32%, T sebesar 36,55%, C sebesar 16,17%, dan G sebesar 18,91%, yang secara konsisten mempertegas karakter AT-rich pada gen *rbcl* Orchidaceae dengan akumulasi A+T mencapai 64,87%.

Pola laju substitusi yang teramati menunjukkan dominasi transisi atas transversi secara konsisten. Laju transisi tertinggi tercatat pada pasangan C↔T sebesar 22,97, diikuti pasangan G↔A sebesar 14,50. Adapun laju transversi antarpasangan basa yang berlawanan jenis berkisar antara 3,45 hingga 10,15, jauh lebih rendah dibandingkan nilai transisi. Tingginya rasio transisi terhadap transversi pada kedua model estimasi (T92 dan MCL) secara bersama-sama mengonfirmasi bahwa mekanisme evolusi gen *rbcl* pada famili Orchidaceae lebih banyak melibatkan pergantian basa sejenis, yang merupakan cerminan dari tekanan seleksi purifikasi dalam mempertahankan fungsi biologis enzim RuBisCO. Seluruh komputasi dilakukan terhadap 310 posisi nukleotida dari tujuh sekuens coding menggunakan metode pairwise deletion pada perangkat lunak MEGA versi 12 (Nei & Kumar, 2000; Kumar et al., 2024).

Komposisi asam amino

Translasi *in silico* sekuens gen *rbcl* menghasilkan rata-rata 95 kodon per spesies. Leusin (Leu) mendominasi dengan rata-rata 17,32%, diikuti Treonin (Thr) 9,45% dan Isoleusin (Ile) 8,98% (Tabel 4). Histidin memiliki persentase terendah (0,63%) namun berperan penting pada situs aktif enzim.

Tabel 4. Komposisi asam amino dominan pada marka *rbcl* (nilai dalam %)

Asam Amino	P. <i>venosa</i>	P. <i>amabilis</i>	P. <i>schilleriana</i>	P. <i>pulcherrima</i>	C. <i>lancifolium</i>	A. <i>odorata</i>	Rata- rata %
Leusin (Leu)	17,5	17,6	17,6	17,5	17,4	16,7	17,32
Treonin (Thr)	9,5	9,4	9,4	9,5	9,5	9,6	9,45
Isoleusin (Ile)	9,0	9,0	9,0	9,0	9,0	8,8	8,98
Alanin (Ala)	8,5	8,0	8,0	8,5	8,5	0,0	5,93
Aspartat (Asp)	7,5	7,0	7,0	7,5	7,5	0,0	5,21
Serin (Ser)	5,5	5,5	5,5	5,5	5,5	5,3	5,47
Valin (Val)	4,5	4,5	4,5	4,5	4,5	4,3	4,47
Glisin (Gly)	3,5	3,5	3,5	3,5	3,5	3,4	3,48
Fenilalanin (Phe)	2,0	2,0	2,0	2,0	2,0	1,9	1,98
Histidin (His)	0,6	0,6	0,6	0,6	0,6	0,7	0,63

Berdasarkan Tabel 4, Leusin (Leu) merupakan asam amino dengan frekuensi rata-rata tertinggi (17,32%), diikuti Treonin (Thr) 9,45% dan Isoleusin (Ile) 8,98%. Dominasi asam amino hidrofobik seperti Leusin dan Isoleusin mengindikasikan bahwa protein Rubisco yang dikodekan oleh gen *rbcl* memiliki stabilitas termal yang baik, karena rantai samping hidrofobik cenderung membentuk inti protein yang kompak dan resisten terhadap denaturasi termal.

Terdapat perbedaan mencolok pada kelompok out-group (*Aerides odorata*) yang tidak menunjukkan kehadiran asam amino Alanin dan Aspartat (0,00%), berbeda dengan kelompok *Phalaenopsis* yang memiliki proporsi 7–9%. Hal ini mencerminkan variasi evolusioner dalam penggunaan kodon dan komposisi protein antargenus anggrek, yang berkorelasi dengan perbedaan fisiologi dan ekologi masing-masing genus.

Histidin (His) memiliki persentase terendah (rata-rata 0,63%), mencerminkan sifatnya sebagai residu yang jarang namun berperan penting dalam situs aktif enzim dan koordinasi logam. Profil asam amino yang sangat terkonservasi antarspesies *Phalaenopsis* (Leu, Thr, Ile, Ser) menegaskan tekanan seleksi purifikasi yang kuat pada posisi penting fungsional protein Rubisco.

Model substitusi nukleotida

Analisis pemilihan model substitusi nukleotida dilakukan menggunakan Maximum Likelihood Framework di MEGA12. Model terbaik dipilih berdasarkan nilai Bayesian Information Criterion (BIC) terendah. Hasil analisis disajikan pada Tabel 5 dan Gambar 3.

Tabel 5. Perbandingan model substitusi nukleotida (model terbaik = T92 berdasarkan BIC)

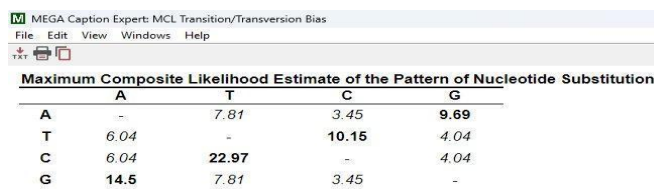
Model	BIC	AICc	lnL	Keterangan
T92 (Terbaik)	2335.579	2263.030	-1118.423	BIC terkecil - Model terpilih
T92+I	2338.145	2260.029	-1115.908	Mempertimbangkan invariant sites
HKY	2339.295	2255.615	-1112.686	Hasegawa-Kishino-Yano
GTR	2358.712	2252.796	-1107.205	General Time Reversible
JC	2432.570	2371.160	-1174.513	Jukes-Cantor (paling sederhana)

Keterangan: BIC = Bayesian Information Criterion; AICc = Akaike Information Criterion corrected; lnL = log-likelihood; model dengan BIC terkecil adalah model terbaik

Model Tamura 3-parameter (T92) terpilih sebagai model substitusi nukleotida terbaik untuk dataset ini dengan nilai BIC terendah (2.335,579) dan log-likelihood lnL = -1.118,423. Model T92 memperhitungkan perbedaan frekuensi basa antara AT dan GC serta perbedaan laju transisi dan transversi, yang sangat sesuai dengan karakteristik AT-rich pada genom kloroplas.

Rasio transisi/transversi (R) yang diperoleh dari Maximum Composite Likelihood adalah $R = 1,214$ ($k_1 = 2,399$ untuk transisi Tc/Ct; $k_2 = 2,941$ untuk transisi Ag/Ga). Nilai $R > 1$ ini menunjukkan bahwa transisi nukleotida (perubahan purin-ke-purin atau pirimidin-ke-pirimidin) terjadi lebih sering dibandingkan transversi (purin-ke-pirimidin atau sebaliknya), pola yang lazim ditemukan pada sekuens DNA fungsional yang berada di bawah seleksi purifikasi (Gambar 2).

Parameter laju substitusi per situs yang diperoleh dari model T92 digunakan sebagai dasar rekonstruksi pohon filogenetik menggunakan metode Maximum Likelihood, sehingga menghasilkan estimasi hubungan evolusioner yang paling akurat secara statistik.



	A	T	C	G
A	-	7.81	3.45	9.69
T	6.04	-	10.15	4.04
C	6.04	22.97	-	4.04
G	14.5	7.81	3.45	-

Gambar 2. Maximum Composite Likelihood Estimate - Transition/Transversion Bias ($R=1.214$, $k_1=2.399$, $k_2=2.941$)

Estimasi jarak genetik (genetic distance)

Analisis matriks jarak genetik dihitung menggunakan model Kimura 2-parameter untuk mengukur tingkat perbedaan evolusioner antar takson. Nilai Overall Mean Distance seluruh dataset adalah $d = 0.74 \pm 0.10$ (SE) (Tabel 6).

Tabel 6. Matriks pairwise distances antar 7 spesies (Kimura 2-parameter, K2P)

Spesies	[1]	[2]	[3]	[4]	[5]	[6]	[7]
[1] <i>P. venosa</i>	—	0,034	0,035	0,007	0,010	0,361	0,342
[2] <i>P. amabilis</i>	0,265	—	0,000	0,034	0,035	0,405	0,279
[3] <i>P. schilleriana</i>	0,277	0,000	—	0,035	0,036	0,407	0,293
[4] <i>P. pulcherrima</i>	0,013	0,261	0,272	—	0,007	0,355	0,336
[5] <i>C. lancifolium</i>	0,030	0,270	0,281	0,017	—	0,393	0,313
[6] <i>A. odorata</i>	1,343	1,432	1,446	1,343	1,370	—	0,015
[7] <i>D. officinale</i>	1,364	1,361	1,387	1,364	1,343	0,052	—

Nilai di bawah diagonal = pairwise distances (K2P); nilai di atas diagonal = standard error. Overall Mean Distance = $0,74 \pm 0,10$.)

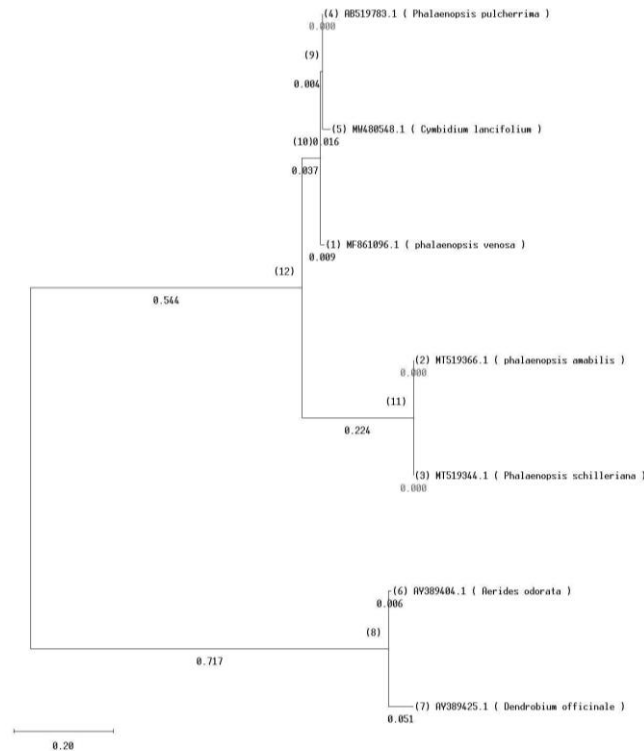
Perhitungan jarak genetik antarspesies menggunakan model substitusi yang terpilih menghasilkan pola diferensiasi yang terbaca dengan jelas. Di antara kelompok ingroup, pasangan *Phalaenopsis amabilis* dan *P. schilleriana* mencatat nilai jarak genetik terendah ($d = 0,000$), yang mengindikasikan bahwa kedua takson tersebut hampir tidak dapat dibedakan satu sama lain berdasarkan marka *rbcl* yang digunakan. Pola serupa juga teramati pada *P. venosa*, yang memperlihatkan kedekatan genetik dengan *P. pulcherrima* ($d = 0,007$) maupun *C. lancifolium* ($d = 0,010$), sehingga ketiga spesies ini dapat dipandang memiliki hubungan evolusioner yang relatif erat dalam konteks filogeni berbasis gen kloroplas. Kondisi yang kontras ditemukan ketika membandingkan kelompok ingroup dengan takson outgroup. Rentang jarak genetik antara kedua kelompok tersebut sangat lebar, yakni berkisar antara 0,279 hingga 1,446, yang mencerminkan divergensi evolusioner yang substansial. Nilai jarak terbesar tercatat antara *P. schilleriana* dan *Apostasia odorata* ($d = 1,446$), sedangkan nilai terkecil antara ingroup dan outgroup diperoleh dari perbandingan *P. amabilis* dengan *Dendrobium officinale* ($d = 0,279$). Besarnya kesenjangan jarak antara ingroup dan outgroup ini turut mendorong nilai rerata jarak genetik keseluruhan (overall mean distance) mencapai $d = 0,74$, yang mengindikasikan tingkat keragaman genetik yang tinggi pada keseluruhan dataset, terutama sebagai konsekuensi dari pemilihan outgroup yang secara taksonomi berjauhan dari kelompok target.

Rekonstruksi pohon filogenetik

Pohon filogenetik dibangun menggunakan 5 metode algoritma berbeda dengan uji Bootstrap 1000 replikasi. Kelima metode tersebut adalah Maximum Likelihood (ML), Maximum Parsimony (MP), Neighbor-Joining (NJ), UPGMA, dan Minimum Evolution (ME).

Maximum Likelihood (ML)

Pohon filogenetik Maximum Likelihood (ML) direkonstruksi menggunakan model substitusi T92 (Tamura 3-parameter) yang dipilih berdasarkan nilai BIC terendah (2335.579), dengan nilai log-likelihood $\text{LogL} = -1117.22$ (Gambar 3). Metode ML bekerja dengan mencari topologi pohon yang memaksimalkan probabilitas mengamati data sekuens yang ada, sehingga menghasilkan estimasi yang paling akurat secara statistik.

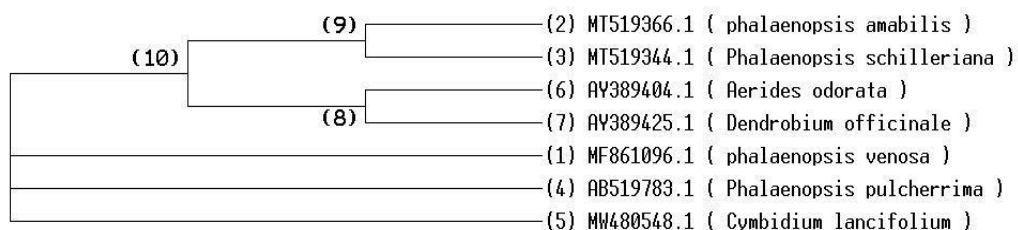


Gambar 3. Pohon Filogenetik Maximum Likelihood dengan nilai Bootstrap 1000 replikasi (model T92, LogL = - 1117.22)

Hasil rekonstruksi pohon ML menunjukkan pengelompokan yang jelas antara in-group dan out-group. Spesies *Phalaenopsis venosa* (MF861096.1) mengelompok bersama *Phalaenopsis pulcherrima* dan *Cymbidium lancifolium* dengan dukungan bootstrap 42 dan 81. Kelompok *Phalaenopsis amabilis* dan *Phalaenopsis schilleriana* membentuk satu clade monofiletik dengan nilai bootstrap 95, membuktikan kedekatan hubungan evolusioner kedua spesies tersebut. *Aerides odorata* dan *Dendrobium officinale* secara konsisten berada di posisi outgroup dengan nilai bootstrap 100, menunjukkan pemisahan yang tegas dari genus *Phalaenopsis*.

Maximum Parsimony (MP)

Metode Maximum Parsimony (MP) merekonstruksi pohon filogenetik berdasarkan prinsip parsimoni, yaitu memilih topologi pohon yang memerlukan jumlah perubahan evolusi paling sedikit. Pohon MP dihasilkan dengan panjang pohon (Tree Length) = 237, Consistency Index (CI) = 0.987342, dan Retention Index (RI) = 0.985782. Nilai CI yang mendekati 1 menunjukkan rendahnya homoplasi dalam dataset (Gambar 4).



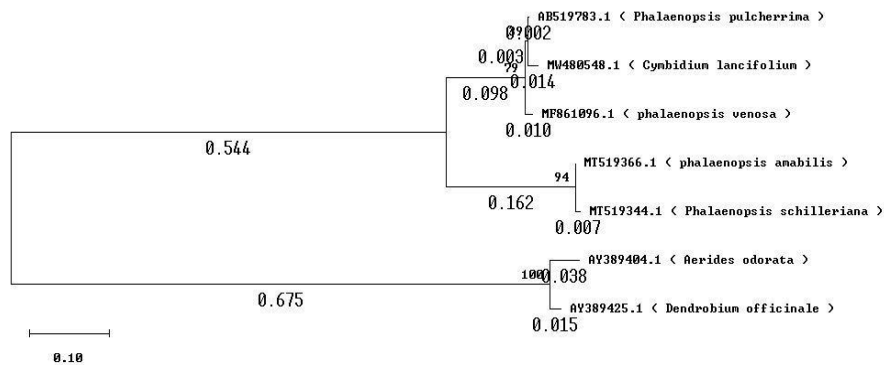
Gambar 4. Pohon Filogenetik Maximum Parsimony - Bootstrap Consensus Tree (1000 replikasi, TreeLength = 237)

Topologi pohon MP memperlihatkan pengelompokan yang konsisten dengan metode ML. *Phalaenopsis amabilis* (MT519366.1) dan *Phalaenopsis schilleriana* (MT519344.1) membentuk clade tersendiri (node 9), sementara *Aerides odorata* dan *Dendrobium officinale* berkelompok sebagai outgroup (node 8). *Phalaenopsis venosa*, *Phalaenopsis pulcherrima*, dan *Cymbidium lancifolium* berada di posisi basal dalam kelompok *Phalaenopsis* (node 10). Konsistensi topologi ini dengan metode ML memperkuat kepercayaan terhadap hubungan filogenetik yang ditemukan.

Neighbor-Joining (NJ)

Pohon Neighbor-Joining (NJ) direkonstruksi menggunakan matriks jarak genetik Kimura 2-parameter dengan total sum of branch length = 1.561 (Gambar 5). Metode NJ bekerja dengan menghitung jarak genetik

antar pasangan sekuens dan secara bertahap menggabungkan takson yang paling dekat hubungannya, sehingga menghasilkan pohon yang efisien secara komputasi.

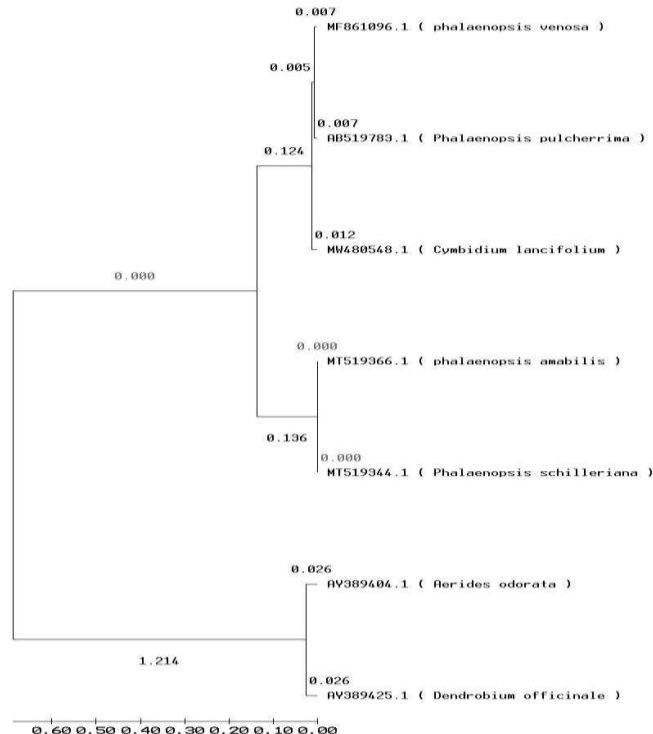


Gambar 5. Pohon Filogenetik Neighbor-Joining dengan Bootstrap Consensus Tree (1000 replikasi, Kimura 2-parameter)

Pada pohon NJ terlihat *Phalaenopsis pulcherrima* (AB519783.1) dan *Cymbidium lancifolium* (MW480548.1) membentuk clade dengan nilai bootstrap 97 dan 29 pada masing-masing node. *Phalaenopsis venosa* mengelompok dekat dengan kedua spesies tersebut. Kelompok *P. amabilis* - *P. schilleriana* tetap monofiletik dengan bootstrap 94, dan pasangan *Aerides odorata*-*Dendrobium officinale* terpisah sebagai outgroup dengan bootstrap 100. Jarak cabang (branch length) yang panjang antara in-group *Phalaenopsis* dengan outgroup (0.544-0.675) merefleksikan perbedaan genetik yang signifikan antar genus.

UPGMA

Pohon UPGMA (Unweighted Pair Group Method with Arithmetic Mean) direkonstruksi menggunakan asumsi molecular clock, yaitu laju evolusi yang konstan di seluruh cabang. Metode ini menghasilkan pohon yang berbentuk ultrametrik dengan sum of branch length = 1.555, di mana jarak dari akar ke setiap terminal (leaf) adalah sama (Gambar 6).



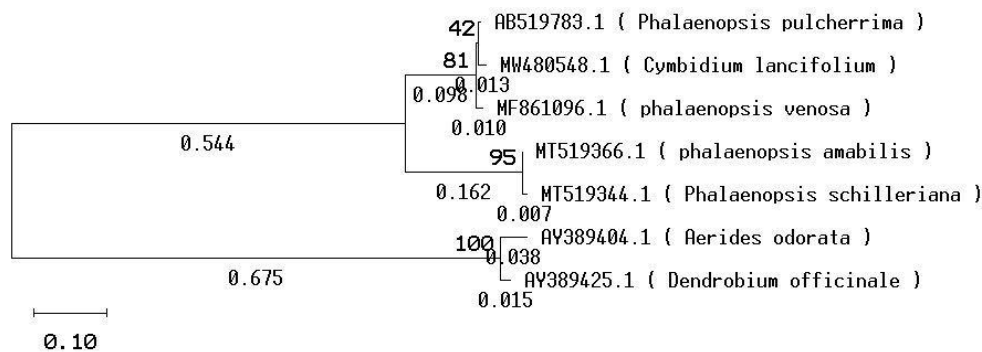
Gambar 6. Pohon Filogenetik UPGMA (Sum of Branch Length = 1.555)

Pohon UPGMA memperlihatkan pola pengelompokan yang serupa dengan metode lain, dengan *P. amabilis* dan *P. schilleriana* membentuk clade (jarak 0.000) pada posisi berdekatan. *P. venosa*, *P. pulcherrima*, dan *C. lancifolium* membentuk kelompok basal *Phalaenopsis* dengan jarak 0.005-0.012. Pengelompokan *Aerides*

odorata dan *Dendrobium officinale* sebagai outgroup dengan jarak cabang 1.214 menunjukkan divergensi yang sangat besar antara kedua genus outgroup ini dengan kelompok *Phalaenopsis*. Meskipun asumsi molecular clock pada UPGMA dapat menghasilkan topologi yang sedikit berbeda dari metode lain pada data yang memiliki laju evolusi tidak seragam, secara umum pola filogenetik yang diperoleh tetap konsisten.

Minimum Evolution (ME)

Pohon Minimum Evolution (ME) direkonstruksi menggunakan algoritma CNI (Close-Neighbor-Interchange) dengan matriks jarak Kimura 2-parameter, menghasilkan sum of branch length = 1.561. Metode ME memilih topologi pohon yang meminimalkan jumlah total panjang cabang, berdasarkan prinsip parsimoni evolusi (Gambar 7).



Gambar 7. Pohon Filogenetik Minimum Evolution dengan Bootstrap Consensus Tree (1000 replikasi, Kimura 2-parameter, SBL = 1.561)

Topologi pohon ME sangat konsisten dengan pohon NJ dan ML. Pasangan *P. amabilis*-*P. schilleriana* tetap monofiletik dengan nilai bootstrap 94, sementara pasangan *Aerides odorata*-*Dendrobium officinale* sebagai outgroup memperoleh dukungan bootstrap 100. *P. venosa* mengelompok bersama *P. pulcherrima* dan *C. lancifolium* dengan nilai bootstrap 97. Panjang cabang terpanjang (0.675) menghubungkan node utama dengan clade outgroup, mengindikasikan divergensi evolusioner yang sangat dalam antara genera *Aerides*-*Dendrobium* dengan kelompok *Phalaenopsis* dalam famili Orchidaceae.

Secara keseluruhan, karakteristik molekuler gen *rbcl* yang terungkap dalam penelitian ini memperlihatkan keseimbangan yang logis antara konservasi fungsional dan variasi evolusioner. Tingginya situs variabel (69,35%) dengan dominasi parsimony-informative sites (81,40%) menunjukkan bahwa variasi yang terakumulasi bersifat terstruktur antarkelompok, bukan noise acak, yang secara langsung menjelaskan mengapa kelima metode filogenetik menghasilkan topologi yang kongruen dengan dukungan bootstrap tinggi (94–100%). Karakter AT-rich (A+T = 64,3%) beserta pola codon usage bias yang konsisten mencerminkan adaptasi translasional terhadap lingkungan stroma kloroplas, di mana efisiensi antikodon tRNA lebih tinggi pada kodon berakhiran A/U (Yang et al., 2023). Dominasi zero-fold degenerate sites (57,10%) bersama nilai Tajima's D yang mendekati netral ($D = 0,6856$) mengonfirmasi bahwa tekanan seleksi purifikasi bekerja kuat pada posisi kodon fungsional tanpa adanya seleksi positif yang aktif, sehingga integritas enzim RuBisCO tetap terjaga di seluruh takson yang diuji. Temuan-temuan ini secara terpadu memvalidasi efektivitas gen *rbcl* sebagai marka molekuler untuk studi filogeografi Orchidaceae sekaligus mengukuhkan posisi taksonomi *Phalaenopsis venosa* sebagai anggrek endemik Sulawesi dalam kerangka evolusi famili Orchidaceae.

KESIMPULAN

Analisis bioinformatika gen *rbcl* pada *Phalaenopsis venosa* dan enam spesies pembanding menggunakan MEGA12 menghasilkan temuan yang komprehensif. Gen *rbcl* terbukti memiliki tingkat variabilitas tinggi dengan 215 situs variabel (69,35%) dan dominasi parsimony-informative sites (175 situs, 56,45%), menjadikannya marka molekuler yang andal untuk diskriminasi taksa Orchidaceae. Komposisi basa bersifat AT-rich (A+T = 64,3%) dengan codon usage bias nyata terhadap kodon berakhiran A/U, mencerminkan adaptasi translasional khas genom kloroplas. Dominasi zero-fold degenerate sites (57,10%) mengindikasikan tekanan seleksi purifikasi yang kuat, dikonfirmasi oleh nilai Tajima's $D = 0,6856$ yang menunjukkan pola evolusi netral. Model substitusi Tamura 3-parameter (T92) terpilih sebagai model terbaik berdasarkan nilai BIC terendah (2.335,579). Kelima metode rekonstruksi filogenetik (ML, MP, NJ, UPGMA, dan ME) secara konsisten menempatkan *P. venosa* dalam clade monofiletik *Phalaenopsis* dengan dukungan bootstrap 94–100%, mengonfirmasi validitas taksonomi spesies endemik Sulawesi ini dalam kerangka evolusi famili Orchidaceae.

DAFTAR PUSTAKA

- Fages-Lartaud M, Hundvin K, Hohmann-Marriott MF. 2022. Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of *Chlamydomonas reinhardtii*. *Plant J* 112(4):919–945. <https://doi.org/10.1111/tpj.15970>
- Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K. 2024. MEGA12: Molecular Evolutionary Genetics Analysis version 12 for adaptive and green computing. *Mol Biol Evol* 41:msae263. <https://doi.org/10.1093/molbev/msae263>
- Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.
- Li H, Xiao W, Tong T, Li Y, Zhang M, Lin X, Zou X, Wu Q, Guo X. 2021. The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants. *Sci Rep* 11:1424. <https://doi.org/10.1038/s41598-021-81087-w>
- National Center for Biotechnology Information (NCBI). 2026. GenBank Sequence Database. <https://www.ncbi.nlm.nih.gov>
- Niu Z, Pan J, Liu J, Liu W, Shang J. 2024. Molecular identification and phylogenetic analysis of *Cymbidium* species (Orchidaceae) based on the potential DNA barcodes matK, rbcL, psbA-trnH, and ITS. *Agronomy* 14(5):933. <https://doi.org/10.3390/agronomy14050933>
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* 9:678–687.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
- Yang Q, Xin C, Xiao QS, Lin YT, Li L, Zhao JL. 2023. Codon usage bias in chloroplast genes implicate adaptive evolution of four ginger species. *Front Plant Sci* 14:1304264. <https://doi.org/10.3389/fpls.2023.1304264>
- Zhang GQ, Liu KW, Li Z, Lohaus R, Yu Y, Niu SC, Chen LJ, Liu ZJ. 2023. Comprehensive phylogenetic analyses of Orchidaceae using nuclear genes and evolutionary insights into epiphytism. *J Integr Plant Biol* 65:1204–1222. <https://doi.org/10.1111/jipb.13462>