Möller et al. 1999

Clokie, Culbas & Cronk

Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study, pp. 375-402.

In: P.M. Hollingsworth, R.M. Bateman, & R.J. Gornall (eds.), Molecular Systematics and Plant Evolution.

REFNO: 3001

KEYWORDS:

Cladistics, Conandron, Genetics, Haberlea, Jancaea, Molecular Systematics, Morphology, Primulina, Ramonda, Saintpaulia, Streptocarpus

Chapter 17

Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study

M. Möller, M. Clokie, P. Cubas and Q. C. B. Cronk

ABSTRACT

We have analysed and characterised the phylogenetic potential of a nuclear developmental gene, *cycloidea* (originally isolated from *Antirrhinum*), involved in the development of floral zygomorphy. We have compared the evolution of part of this putative single copy gene in Old World Gesneriaceae with two contrasting DNA sequence regions, using two sets of data (a 'genus' data set and a 'species' data set); the chloroplast *trnL*(UAA) intron and the spacer between the *trnL* (UAA) 3′ exon and *trnF* (GAA) were relatively conserved and suitable for phylogenetic reconstruction at genus level. The multicopy internal transcribed spacers (ITS1 and ITS2) of nuclear ribosomal DNA in contrast appear to be evolving about five times faster and are suitable for resolution at the species level. The putative homologue of *cycloidea* (*Gcyc*) has an intermediate substitution rate about three times faster than the chloroplast intron/spacer region. However, the level of pairwise sequence divergence of *Gcyc* is higher than that of ITS at very low levels of divergence. This difference in apparent rate of molecular evolution between ITS and *Gcyc* at different levels of the taxonomic hierarchy we attribute to the process of molecular drive in the multicopy ITS. At lower levels of divergence (e.g. between closely related species) fixation of genetic changes in the multicopy ribosomal DNA acts as a restraint on evolutionary rate, whereas third codon position changes in coding single copy nuclear (scnDNA) genes are unconstrained. However, at high levels of divergence (e.g. between genera), scnDNA evolution is more functionally constrained than that of ITS and *Gcyc* therefore varies less. The small restraining effect of concerted evolution is not noticeable at these levels of sequence divergence. All three regions appear to evolve in a clock-like manner and are found to be suitable for phylogenetic reconstruction by parsimony, resulting in the same or similar topologies. We have examined the *Gcyc* sequences of three species that have reverted to actinomorphy from a zygomorphic condition. The gene appears to be intact and therefore, by implication, functional in these species. Furthermore, in one of these clades there has been a reversion back to zygomorphy which also implies that the gene is intact. We therefore suggest that in naturally occurring actinomorphic Gesneriaceae *Gcyc* continues to have a functional role, but zygomorphy is reduced by modifying genes. There is no convincing evidence that *Gcyc* evolves faster in actinomorphic lineages.

In *Molecular systematics and plant evolution* (1999) (eds P. M. Hollingsworth, R. M. Bateman and R. J. Gornall), Taylor & Francis, London, pp. 375–402.

17.1 Introduction

Greater access to single copy nuclear genes is a continuing goal for molecular systematists (Doyle and Doyle, 1999 – this volume). Their advantage over commonly used multicopy genes lies in their apparent higher rate of evolution and their potential to resolve relationships at lower taxonomic levels than the ribosomal internal transcribed spacer (ITS), the fastest evolving sequence commonly used today.

Since the invention of PCR and automated sequencing techniques there has been an almost exponential upsurge in the number of publications in molecular systematics. The great majority of investigations have been carried out using chloroplast or multicopy ribosomal DNA sequence data (Hershkovitz *et al.*, 1999, this volume). Depending on the level of taxonomic distance, different types of DNA sequences are used. Chloroplast genes are generally more conserved than nuclear genes, due to their mode of inheritance (usually uniparental but with exceptions) and lower mutation rate (Hagemann and Schröder, 1989; Reboud and Zeyl, 1994; Tilney-Bassett, 1994; Mogensen, 1996; Ennos *et al.*, 1999 – this volume). Coding genes, due to their functional constraints, are more conserved than intron or spacer sequences. Multicopy genes appear more conserved than single copy genes due to gene conversion events and the forces of concerted evolution (Dover, 1986; Hillis *et al.*, 1991; Elder and Turner, 1995).

Studies of broad scale relationships amongst seed plants are often based on the coding chloroplast gene *rbcL* (Chase *et al.*, 1993). This gene allows good resolution at the family level. At generic level, other chloroplast sequences are often chosen, such as the intron and spacer between transfer RNA the *trnL* (UAA) 5′ exon, *trnL* (UAA) 3′ exon and *trn*F (GAA) (Taberlet *et al.*, 1991). For analyses at species level, multicopy nuclear ITS sequences are frequently used (Baldwin, 1992; Möller and Cronk, 1997a; Hershkovitz *et al.*, 1999 – this volume). However, these sequences have their limitations in resolving relationships amongst very closely related taxa, as is shown in the case of *Saintpaulia* species, where seven species had identical ITS sequences. Thus, the ITS sequence analysis failed to resolve the relationships amongst members of the *Saintpaulia ionantha*-complex (Möller and Cronk, 1997b).

Access to single copy nuclear genes may provide further resolution here, as concerted evolution is not required for establishing mutations. This may result in higher sequence divergence compared to multicopy genes. Recently the gene involved in the expression of zygomorphy of *Antirrhinum* flowers, *cycloidea* (*cyc*), has been isolated (Luo *et al.*, 1996). Its homologue in Gesneriaceae appears to be single copy. Sequence data for the *Antirrhinum cyc* gene and sequence data of a *Saintpaulia* clone allowed the design of Gesneriaceae-specific primer pairs amplifying a single product comprising ~70% of the open reading frame (Fig 17.1). However, it should be noted that as there are two related genes in *Antirrhinum*: *cyc* and *dichotoma* (*dich*) (Luo *et al.*, 1996) and several *cyc*-like ESTs in *Arabidopsis* (Luo *et al.*, 1996; Doebley *et al.*, 1997), as well as a similar gene in maize, *teosinte branched 1* (*tb1*) (Doebley *et al.*, 1997), it is likely that *cyc* is part of a family of genes.

We chose *cyc* partly because in Gesneriaceae flower zygomorphy (and number of fertile stamens) is variable. Flower zygomorphy has evolved and been lost several times independently in angiosperms, e.g. in the Solanaceae and Boraginaceae (Coen and Nugent, 1994). In Gesneriaceae zygomorphy is ancestral, and has been lost

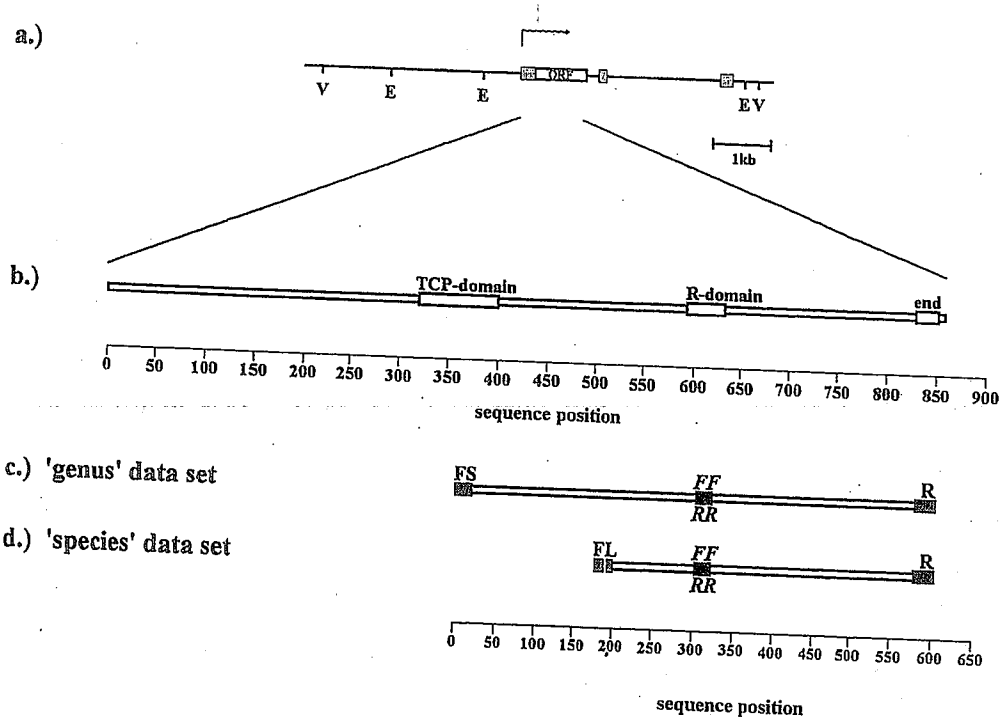


Figure 17.1 a) Structure and map of the *cyc* locus (after Luo *et al.*, 1996). Exons and predicted open reading frame (ORF) are indicated in rectangles; the arrow indicates the direction of transcription; restriction enzyme sites: E – *Eco*RI, V – *Eco*RV. b) The *cyc* ORF and the location of conserved regions (open boxes). Fragment of *cyc* amplified, and the PCR primer positions (closed boxes) for c) the 'genus' and d) for the 'species' data set.

several times independently, for instance in the Asiatic *Conandron*, and in the European *Ramonda* (Möller and Cronk, unpublished). This provides an ideal system for investigating differential gene evolution, and the relationship between sequence divergence of *cyc* homologues in Gesneriaceae and changes in flower morphology.

In this paper we compare sequence data from three DNA sources: from chloroplast DNA (the intron/spacer *trnL+F*), from nuclear DNA (the multicopy nuclear rDNA ITS) and also from the Gesneriaceae homologues of the single copy developmental gene *cyc*. It is essential that there is strict gene homology of the DNA fragments to be compared. We have ensured this by using the following checks: (1) the use of a specific primer pair; (2) sequence similarity including that of shared sequence motifs (the *TCP* domain, *R*-domain and *end*-box [Cubas *et al.*, 1999]) within the open reading frame of *Gcyc*; and (3) checking for congruence between phylogenies derived from independent (nuclear and chloroplast) genes with that derived from *Gcyc*. Any major departure from the topology confirmed by other nuclear and chloroplast genes may indicate homology problems. Two data sets, at the genus level and at the species level, were chosen, firstly to illustrate differences in sequence characteristics; secondly to test the suitability of *Gcyc* sequences for phylogenetic reconstructions; and lastly to investigate the possibility that gene sequence evolution is linked to floral morphology.

17.2 Materials and methods

17.2.1 Plant material

Plant material was from living plants cultivated at the Royal Botanic Garden Edinburgh (RBGE). Identifications were kindly confirmed by B. L. Burtt. For all taxa analysed, voucher herbarium specimens were prepared, flowers were preserved in Kopenhagen mixture in a spirit collection (both deposited in herb. E), and photographs of flowering specimens were taken, and deposited in the RBGE library. For this study a single individual was used to represent each species, which makes the 'species' and 'genus' data sets more directly comparable. Further studies on gene evolution at the species level would ideally choose to use more than one individual per species, as intra-specific variation in DNA sequences is now well known. However, this does not affect the main purpose of the present paper, which focuses on gene evolution rather than species evolution.

17.2.1.1 The 'genus' data set

The ingroup taxa were all from the Didymocarpeae, the largest tribe within Gesneriaceae subfamily Cyrtandroideae (Table 17.1). Representative taxa for zygomorphic flower morphology were the European *Haberlea rhodopensis* and *Jankaea heldreichii*, the Asiatic *Primulina tabacum*, and the African *Saintpaulia velutina*. Additionally, species representing the major growth forms of *Streptocarpus* were selected: the caulescent *Streptocarpus holstii*, the rosulate *Streptocarpus primulifolius* and the unifoliate *Streptocarpus dunnii*. Taxa representing actinomorphic flower types were the Asiatic *Conandron ramondioides*, and two European taxa, *Ramonda myconi* and *Ramonda nathaliae*. Recent classifications of the Gesneriaceae recognise two major subfamilies, the neotropical Gesnerioideae and the chiefly Old

Table 17.1 Accessions of eleven Gesneriaceae taxa examined for sequence variation: 'genus' data set.

No.	Taxon	Origin: distribution	RBGE Accession no.[a]
1	*Sinningia schiffneri* Fritsch	South America	1978 1514
2	*Ramonda myconi* (L.) Rchb	Spain: Pyrenees	1971 1477
3	*Ramonda nathaliae* Pancic & Petrovic, white form	S. Jugoslavia, N. Albania, N.C. Greece	1978 4020
4	*Jankaea heldreichii* Boiss.	Greece: Mt. Olympus	1979 1939
5	*Haberlea rhodopensis* Friv. Symond's form	N.E. Greece, C. and S.C. Bulgaria	1975 4106
6	*Primulina tabacum* Hance	China: Gnangdong, Lian River	1995 1540
7	*Conandron ramondioides* Siebold & Zucc.	Japan:	1969 1267
8	*Saintpaulia velutina* B. L. Burtt	Tanzania: W Usambara Mts., Balangai	1987 2179
9	*Streptocarpus holstii* Engl.	Tanzania: E. Usambara Mts.	1959 2272
10	*Streptocarpus dunnii* Hook. f.	South Africa: Swaziland; N. Mbabane	1994 1745
11	*Streptocarpus primulifolius* Gand.	South Africa: E. Cape, Igoda River	1991 2192

[a] These numbers were also used as voucher numbers.

World Cyrtandroideae (Burtt and Wiehler, 1996). The phylogenetic relationships within and between the subfamilies, based on molecular and morphological cladistic analyses are complicated (Smith, 1996; Smith *et al.*, 1997). Therefore the neotropical *Sinningia schiffneri* was chosen as an outgroup for the genus comparison.

17.2.1.2 The 'species' data set

Nineteen species of *Saintpaulia*, representing all areas of geographical distribution, were chosen for this study (Table 17.2). The outgroup taxon for the species comparison was *Streptocarpus holstii*. This was based on recent results from molecular data where the closest known relatives of *Saintpaulia* are in fact caulescent African species of *Streptocarpus* (Möller and Cronk, 1997a; Smith *et al.*, 1997).

Table 17.2 Accessions of *Streptocarpus* and *Saintpaulia* examined for sequence variation: 'species' data set.

No.	Taxon	Origin: distribution	RBGE Accession no.[a]
1	*Saintpaulia brevipilosa* B. L. Burtt	Tanzania: Nguru Mts., Lulaga, Mt. Kanga	1970 0909
2	*Saintpaulia difficilis* B. L. Burtt	Tanzania: E Usambara Mts., Sigi River, Monga	1987 2176
3	*Saintpaulia diplotricha* B. L. Burtt	Tanzania: NE Usambara Mts., Maweni, Tanga	1987 2172B
4	*Saintpaulia grandifolia* B. L. Burtt	Tanzania: W Usambara Mts., Lutindi	1985 0678
5	*Saintpaulia grotei* Engl.	Tanzania: E Usambara Mts., Amani, Mt Mlinga	1987 2171
6	*Saintpaulia goetzeana* Engl.	Tanzania: Uluguru Mts., Lukwangule Plateau	19971201
7	*Saintpaulia intermedia* B. L. Burtt	Tanzania: E Usambara Mts., Kigongoi	1997 0101
8	*Saintpaulia* cf. *ionantha* H. Wendl.	Tanzania: Tanga, Sigi Caves	1971 0860
9	*Saintpaulia magungensis* E. Roberts	Tanzania: E Usambara Mts., Magunga, Mt. Mlinga	1992 3187
10	*Saintpaulia magungensis* var. *minima* B. L. Burtt	Tanzania: E Usambara Mts., Mavoera estate, Amani	1959 4352
11	*Saintpaulia magungensis* var. *occidentalis* B. L. Burtt	Tanzania: W Usambara Mts.	1985 0680
12	*Saintpaulia nitida* B. L. Burtt	Tanzania: Nguru Mts., Mkobwe, Turiani,	1992 3186
13	*Saintpaulia orbicularis* var. *purpurea* B. L. Burtt	Tanzania: W Usambara Mts., Ambangulu	1958 3586
14	*Saintpaulia pendula* var. *kizarae* B. L. Burtt	Tanzania: NE Usambara Mts., Mt. Mtai, Kizara	1997 0103
15	*Saintpaulia rupicola* B. L. Burtt	Kenya: Kaloleni	1997 0094
16	*Saintpaulia shumensis* B. L. Burtt	Tanzania: W Usambara Mts., Shume	1996 2088
17	*Saintpaulia* Sigi Falls	Tanzania: Tanga, Sigi River	1992 3183
18	*Saintpaulia teitensis* B. L. Burtt	Kenya: Teita Hills, Mbololo Hill	C 3771
19	*Saintpaulia tongwensis* B. L. Burtt	Tanzania: E Usambara Mts., Tongwe Mts.	1985 0668
20	*Saintpaulia velutina* B. L. Burtt	Tanzania: W Usambara Mts., Balangai	1987 2179
21	*Streptocarpus holstii* Engl.	Tanzania: E Usambara Mts.	1959 2272

[a] These numbers were also used as voucher numbers.

17.2.2 DNA extraction

Fresh leaf material was used for total DNA extraction using a modified CTAB procedure of Doyle and Doyle (1987), with no further purification.

17.2.3 PCR amplification and conditions

Chloroplast gene intron and spacer: chloroplast DNA (cpDNA) was PCR amplified using primers c and f of Taberlet *et al.* (1991), amplifying the *trnL*(UAA) intron and the intergenic spacer between the *trnL* (UAA) 3′ exon and *trnF* (GAA). Ribosomal DNA (rDNA) spacer: the complete ITS region, including the 5.8S rDNA gene and both flanking internal transcribed spacers, was PCR amplified, using modified primers (Möller and Cronk, 1997a), based on published data by White *et al.* (1990). Single copy developmental gene (scnDNA): part of the open reading frame (ORF) of *Gcyc* was amplified using forward primer *Gcyc*FS (ATG CTA GGT TTC GAC AAG CC) and the reverse primer *Gcyc*R (ATG AAT TTG TGC TGA TCC AAA ATG) (Fig. 17.1), designed and modified after conserved motifs in *Antirrhinum cyc* sequences (Luo *et al.*, 1996) and a cloned *Saintpaulia* sequence. In *Saintpaulia* and *Streptocarpus* taxa a gene closely related to *Gcyc* was found. A selective forward primer *Gcyc*FL (CAC CCG GAT TCG AGA AAA TC) was designed that in combination with the reverse primer *Gcyc*R exclusively amplified *Gcyc* (Fig 17.1). The PCR reaction mixture and PCR cycle parameters, amplicon quantification and purification are described elsewhere (Möller and Cronk, 1997a).

17.2.4 Sequencing protocol

Cycle sequencing and analysis protocols followed Möller and Cronk (1997a). For each taxon, forward and reverse sequencing reactions were performed for sequence confirmation. Sequencing primers were identical to those used for PCR. Additionally, two shorter reactions were set up using internal primers anchored in highly conserved regions, sequencing the cpDNA intron and intergenic spacer with primers d and e (Taberlet *et al.*, 1991), or ITS 1 and ITS 2 with primers 2G and 3P (Möller and Cronk, 1997a), or *Gcyc* with primers *Gcyc*RR (CTT GAT GCA CAT TTT CTC CTT) and *Gcyc*FF (AAG GAG AAA ATG TGC ATC AAG) from within the PCR amplified products.

17.2.5 Sequence analysis

Sequence boundaries of both rDNA internal transcribed spacers of all taxa were determined as described previously (Möller and Cronk, 1997a). The full length of the PCR amplified fragments of the cpDNA, including the transfer RNA gene, or the whole *Gcyc* gene fragment amplified (~70% of the ORF), including the *TCP* and *R*-domains and the *end*-box, were included in the respective matrices. All matrices were aligned using the CLUSTAL option in the multiple alignment program Sequence Navigator™, version 1.0.1 software package (Perkin Elmer, Applied Biosystems Division, Foster City, CA, USA), followed by manual optimisation. The G + C content was determined by inspection, and transition/transversion ratios calculated using MacClade, version 3.01 (Maddison and Maddison, 1992).

Synonymous and non-synonymous substitutions were determined using the program Molecular Evolutionary Genetics Analysis, version 1.01 (MEGA; Kumar *et al.*, 1993). Sequence divergence among taxa was calculated using the DISTANCE MATRIX option in PAUP, version 3.1.1 (Swofford, 1993), based on unambiguously alignable regions, with adjustment for gaps in pairwise comparisons. Graphics and statistical analyses were produced using the program Statgraphic, version 1.03. Sequences used in this study are available from the authors on request.

17.2.6 Phylogenetic analysis

Phylogenetic trees were generated from unordered character states using PAUP, version 3.1.1 (Swofford, 1993). The genus comparison was analysed using the exhaustive search mode. In view of the large number of taxa included in the species comparison, the following heuristic search strategy was employed to find the most parsimonious trees: 500 replicates of RANDOM addition sequence with no swapping. This was followed by TBR swapping on the resulting trees (Möller and Cronk, 1997a). The options MULPARS, STEEPEST DESCENT, COLLAPSE, and ACCTRAN optimisation were selected.

Bootstrap analyses (Felsenstein, 1985) were performed using PAUP, set to HEURISTIC search option and SIMPLE addition sequence. Bootstrap values were calculated using 200 to 1000 replicates with MAXTREE set to 1000. Decay indices (DI) (Bremer, 1988; Donoghue *et al.*, 1992) for individual clades were obtained by comparing the strict consensus of all equal-length trees up to a maximum of 42 steps longer than the shortest tree, using SIMPLE addition sequence and TBR in PAUP. Descriptive statistics reflecting the amount of phylogenetic signal in the parsimony analyses were given by the consistency index (CI) (Kluge and Farris, 1969), retention index (RI) (Farris, 1989), and the resulting rescaled consistency index (RC) (Swofford, 1993).

All three data sets were used for the 'genus'-level comparison (cpDNA was omitted from the 'species'-level study as the cpDNA showed hardly any variation at this level).

For cpDNA and rDNA sequences, only combined spacer/intron sequence data were subjected to phylogenetic analyses. For one analysis of genus comparison cpDNA, rDNA and scnDNA sequence data matrices were combined. For simplicity, differential weighting schemes were not carried out, and in all analyses character state changes were weighted equally, and gaps were treated as missing data (Soltis and Kuzoff, 1995; Susanna *et al.*, 1995; Downie and Katz-Downie, 1996). Ambiguous regions that allowed alternative alignment interpretations were excluded from phylogenetic analyses (Wojciechowski *et al.*, 1993; Downie and Katz-Downie, 1996).

17.3 Results

17.3.1 The 'genus' data set

17.3.1.1 Sequence comparison between trnL+F, ITS and cyc

The distribution of character changes, base substitutions and indel events, for the cpDNA was higher in the spacer region than in the intron (Fig. 17.2a). The ITS

data matrix showed a similar pattern, although with some more conserved regions within both spacers (Fig. 17.2b), where conservation of secondary structure is observed amongst angiosperms (Liu and Schardl, 1994). These are presumed to be recognition sites important during post-transcriptional processes. In the part of the *Gcyc* gene amplified, the distribution of variable sites was relatively evenly spread, except for approximately 20 to 40bp at the beginning and end of the fragment, representing the loop of the conserved *TCP*-domain found in related genes (Cubas *et al.*, 1999), and a conserved *end*-box, respectively. An extended *R*-domain (Cubas *et al.*, 1999) can be found in the Gesneriaceae, stretching from position 289 to 369, including considerably fewer changes (Fig 17.2c, Fig 17.3). Seemingly less variable regions beyond the *R*-domain up to position 453 and from 613 to 642 were the result of larger insertions in few taxa or the outgroup, respectively (Fig 17.3).

The alignment matrix of the *trnL+F* intron/spacer sequences required the insertion of 25 gaps of 1–65 bp. A 65 bp deletion in the cpDNA spacer was found only in *Streptocarpus primulifolius*. Previous analyses investigating the effects of large deletions on tree topologies found little effect of complete removal of the deletion sites from the matrix (Möller and Cronk, 1997a). As inclusion or exclusion did not affect tree topology, the cpDNA positions were left in the matrix to prevent loss of potential information amongst the other taxa at those positions. The matrix contained 88.2% constant sites, 7.6% autapomorphic sites, but only 4.2% informative sites (Table 17.3).

Due to the taxonomic distance of the taxa included in the 'genus' data set, alignment of the combined ITS 1 and ITS 2 matrix was difficult. Due to alternative alignment interpretation 115 sites had to be excluded. Optimised alignment required the insertion of 51 gaps of 1–10 bp length. Of the remaining 423 sites, 47.0% were constant, 27.9% were unique to individual taxa and 25.1% were informative phylogenetically (Table 17.3).

The amplified *Gcyc* fragment of the diverse genera varied from 530 to 656 bp, and the aligned matrix was 740 bp long, with 39 indels of 3–66 bp length, of which most were informative (34). Of all sites, 64.5% were constant, 18.2% autapomorphies, and 17.3% informative, which was intermediate between the cpDNA and rDNA sequences. Translation to amino acids resulted in a 246 codon matrix, with an increased percentage of variable sites, of which 23.6% were uninformative and 23.2% informative (Table 17.3).

17.3.1.2 Phylogenetic reconstruction using three genes

Parsimony analysis of aligned cpDNA sequences resulted in one most parsimonious tree (Fig. 17.4a). When all sites were included the tree had a length of 133 steps, with a high CI of 0.947, an RI of 0.908 and an RC of 0.860. The average number of nucleotide substitutions per site was low with 0.134, with only two out of 990 sites changing three times, indicating a very low saturation of base mutations. It is therefore unlikely that phylogenetic signal has become obscured by multiple substitutions. Forty-five base substitutions separated the outgroup taxon *Sinningia schiffneri* from the ingroup (Fig. 17.4a). The ingroup taxa form four clades, one consisting of the European taxa *Ramonda nathaliae* and *Jankaea heldreichii* as sister

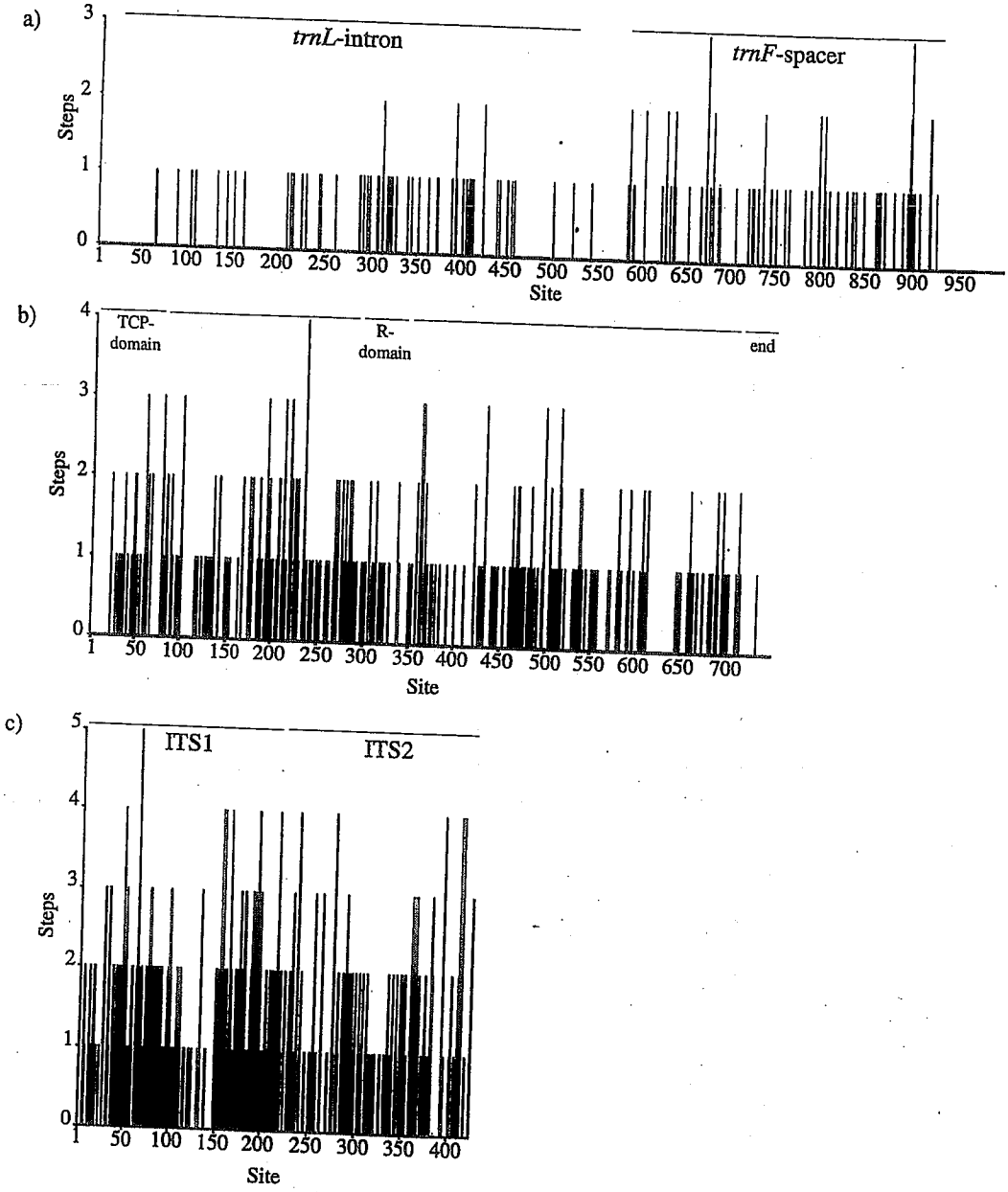


Figure 17.2 The number of steps per base position for the 'genus' data set of a) the *trnL+F*, b) the *Gcyc*, and c) the ITS sequence matrix, illustrating the variation across the gene sequences.

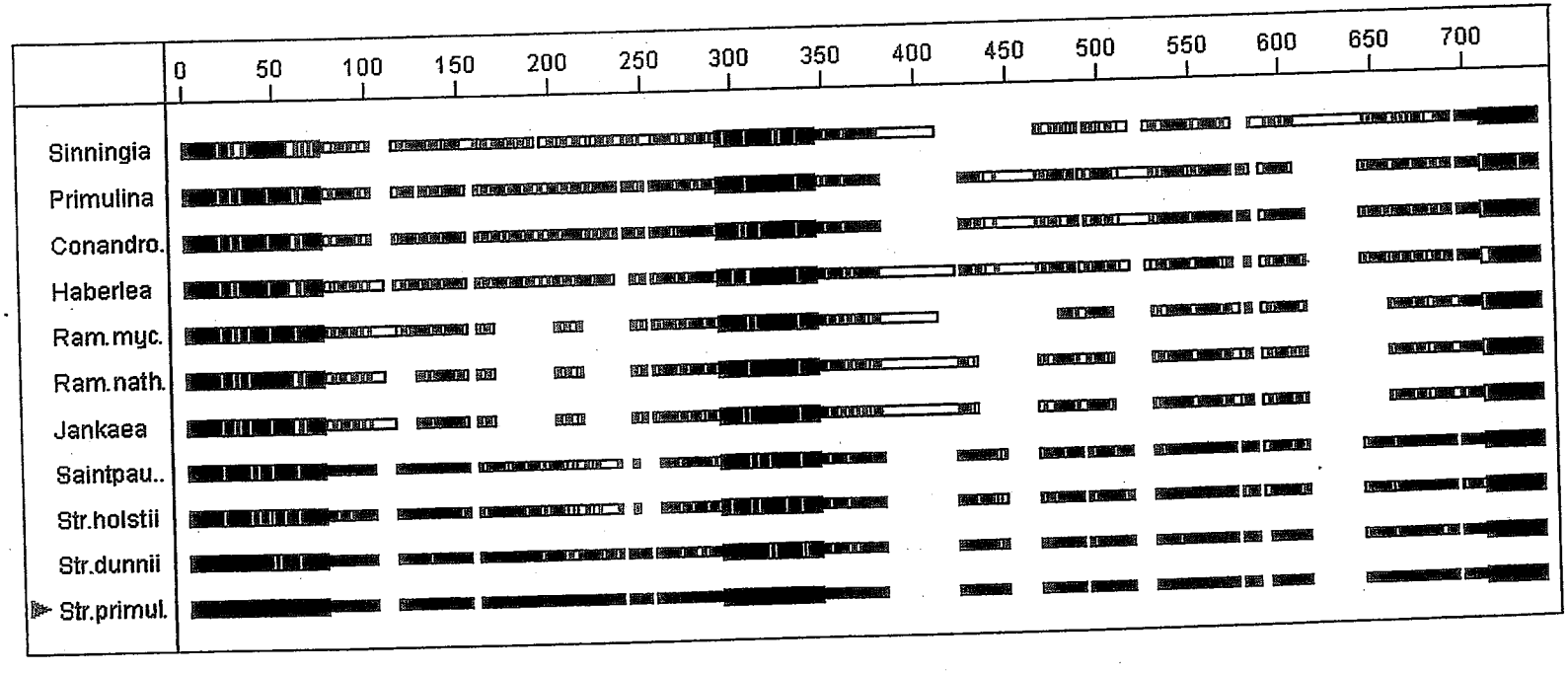


Figure 17.3 Conservation of *Gcyc* DNA sequences between *Streptocarpus primulifolius* and other taxa included in the 'genus' data set. Black = conserved, white = not conserved. Gaps: deletion events. Drawn with MACAW, vers. 2.0.5 (Schuler, 1995).

Table 17.3 Sequence characteristic of *trnL+F*, ITS 1 and ITS 2, and *Gcyc* regions of eleven taxa of Gesneriaceae: 'genus' data set.

Parameter	*trnL+F*	*ITS 1 and ITS 2*	*Gcyc nucleotides*	*Gcyc amino acids*
Length range, bp	870–970	380–390	530–656	176–218
Length mean, bp	946.0	384.4	597.6	198.5
Aligned length, bp	990	538	740	246
Number of sites excluded, bp	–	115	–	–
G+C content mean, %	36.5	56.99	41.0	–
Sequence divergence (in–outgroup)	45–68	109–126	93–136	43–52
Sequence divergence, (in–outgroup) %	4.8–7.0	30.1–34.5	18.4–24.0	23.2–31.0
Sequence divergence (ingroup)	2–45	16–87	27–98	10–46
Sequence divergence, (ingroup) %	0.2–4.8	4.2–23.6	4.9–17.0	5.8–24.2
Number of indels, (informative)	25 (14)	51 (17)	39 (34)	39 (34)
Size of indels, bp	1–65	1–10	3–66	1–22
Number of constant sites (%)	873 (88.2)	199 (47.0)	477 (64.5)	131 (53.2)
Number of variable sites (%)	117 (11.8)	224 (53.0)	263 (35.5)	115 (46.8)
Number of autapomorphic sites (%)	75 (7.6)	118 (27.9)	135 (18.2)	58 (23.6)
Number of informative sites (%)	42 (4.2)	106 (25.1)	128 (17.3)	57 (23.2)
Transitions (min-max)	67–70	218–237	200–213	–
Transversions (min-max)	63–66	143–161	174–187	–
Transitions/transversions	1.07	1.56	1.17	–
Average number of steps per character	0.134	0.898	0.523	0.801

taxa (bootstrap value [BS] = 62%, DI = +1) and *Ramonda myconi* (BS = 90%, DI = +2). *Haberlea rhodopensis* is situated between this clade (BS = 90%, DI = +2) and the rest of the taxa. The Asiatic taxa *Primulina tabacum* and *Conandron ramondioides* form a distinct clade (BS = 100%, DI = +9) and are sister to the African taxa. Within the African group (BS = 100%, DI = +7), the unifoliate *Streptocarpus dunnii* and the rosulate *Streptocarpus primulifolius* (BS = 90%, DI = +2) form a sister clade to the caulescent *Streptocarpus holstii* and *Saintpaulia velutina* (BS = 100%, DI = +7).

One most parsimonious tree was retained after parsimony analysis of unambiguously aligned sequences of both ITS spacers (Fig. 17.4b). The tree length was 380 steps and the CI was 0.779, lower than for cpDNA sequences, indicating more homoplastic changes. The higher homoplasy was also reflected in the higher number of steps per site; 11 sites have changed four times, and one site five times over the tree, with an average of 0.898, indicating a potential saturation and reversals across the matrix. The tree topology, however, was identical to the cpDNA tree (Fig. 17.4a), with similar BS and DI values, except for branches separating *Haberlea rhodopensis* from the other European taxa and the rest of the taxa, and the branch grouping the Asiatic taxa, which collapsed after bootstrap analysis.

Parsimony analysis of the *Gcyc* data matrix resulted in one most parsimonious tree of 387 steps. Estimates of homoplasy were intermediate between cpDNA and rDNA matrices (Fig. 17.4c). Five sites changed four times, and the average number of steps per site was 0.518. The topology of the *Gcyc* tree was similar to the previous trees, except of the position of *Haberlea rhodopensis* which was sister to the other European taxa (BS = 88%, DI = +4) (Fig. 17.4c). Branch support, BS and DI, for residual clades was similar to cpDNA data.

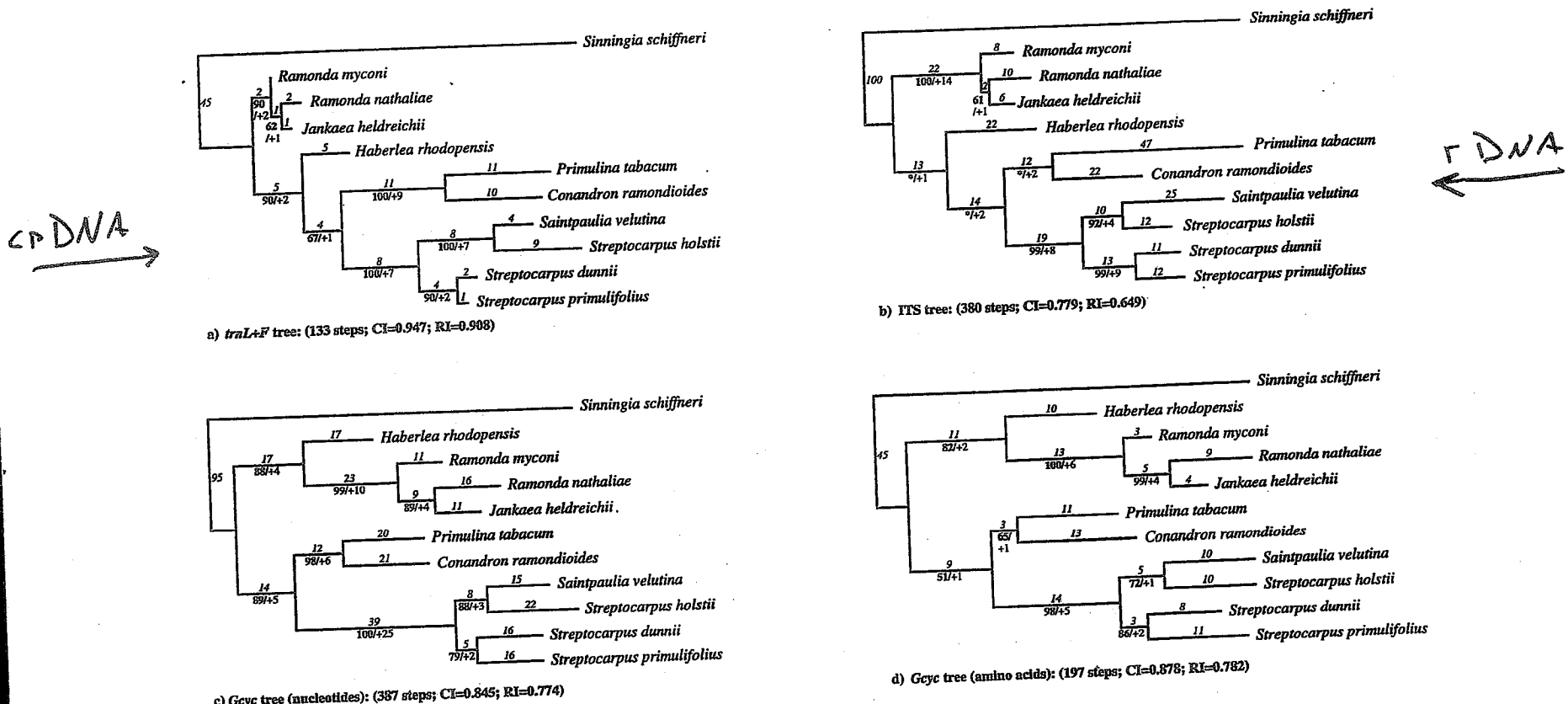


Figure 17.4 Most parsimonious trees for the 'genus' data set based on a) *trnL+F*, b) ITS, c) *Gcyc* nucleotides, and d) *Gcyc* amino acid sequences. Numbers above branches (italic) indicate branchlengths. Numbers below branches (bold) are bootstrap values of 1000 replicates, and numbers in normal print are decay indices.

Phylogenetic analysis of the translated *Gcyc* amino acids by parsimony gave one tree of 197 steps, with similar descriptive values compared to the nucleotide tree (Fig. 17.4d), except for slightly higher consistency values and a higher average number of steps per site, similar to ITS data, indicating a higher saturation to be present compared to the nucleotide data matrix. However, the topology was identical to the nucleotide tree topology. Branch support (as BS) was similar, except for the branches grouping the Asiatic taxa and the Asiatic and African taxa from the rest.

Combining all three data sets gave a matrix of 2275 sites. Parsimony analysis resulted in one most parsimonious tree of 904 steps (Fig. 17.5). The CI was 0.829 and the RC 0.607 (RI = 0.733). The topology was identical to the cpDNA and rDNA trees. Branch support was high (BS = 97–100%, DI = +7–42), except for the branch between the *Ramonda/Jankaea* clade and *Haberlea rhodopensis* and the rest of the taxa (BS = 54%, DI = +1).

17.3.1.3 Comparative pairwise sequence divergence

A translation of *Gcyc* nucleotides into *Gcyc* amino acids increased both the number of informative sites and the divergence in pairwise comparisons, ranging from 5.8% (*Ramonda myconi – Jankaea heldreichii*) to 31.0% (*Sinningia schiffneri – Jankaea heldreichii*) (Table 17.3). This was due to the relation between nucleotide and amino acid divergence in pairwise comparisons ($r = 0.946$; $P > 0.001$) that indicated disproportionally high values for amino acid changes (Fig. 17.6). This is unexpected, as it is believed that most of the variation would be silent, with synonymous sites not altering the amino acid sequence. However, a closer look at the relationship between

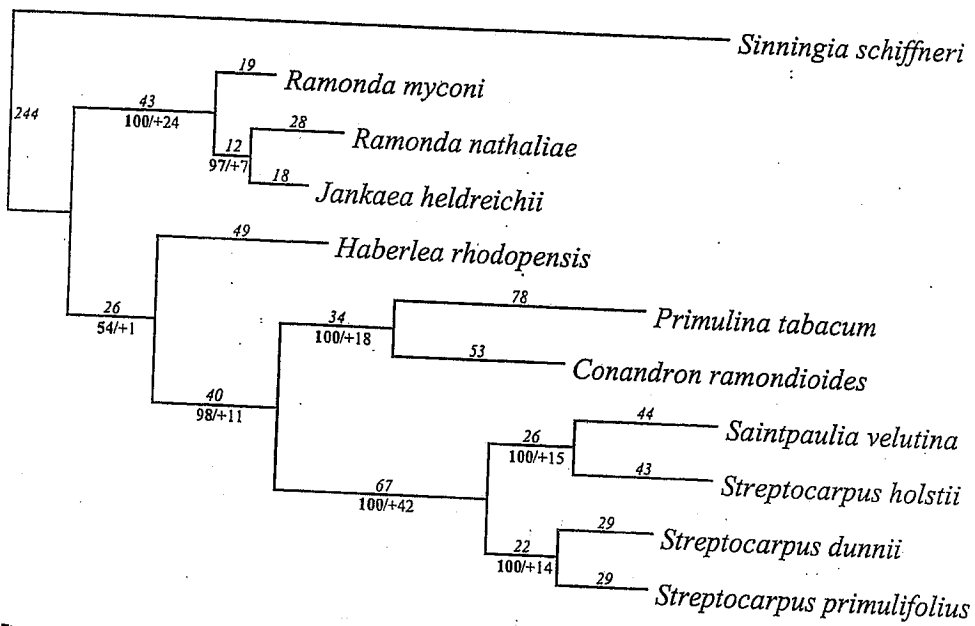


Figure 17.5 Phylogram of the most parsimonious tree for the 'genus' data set of 904 steps length, based on parsimony analysis of the combined *trnL+F*, ITS, and *Gcyc* nucleotides data sets (CI = 0.829; RI = 0.733; RC = 0.607 inclusive uninformative sites). Numbers along branches are as in Figure 17.4.

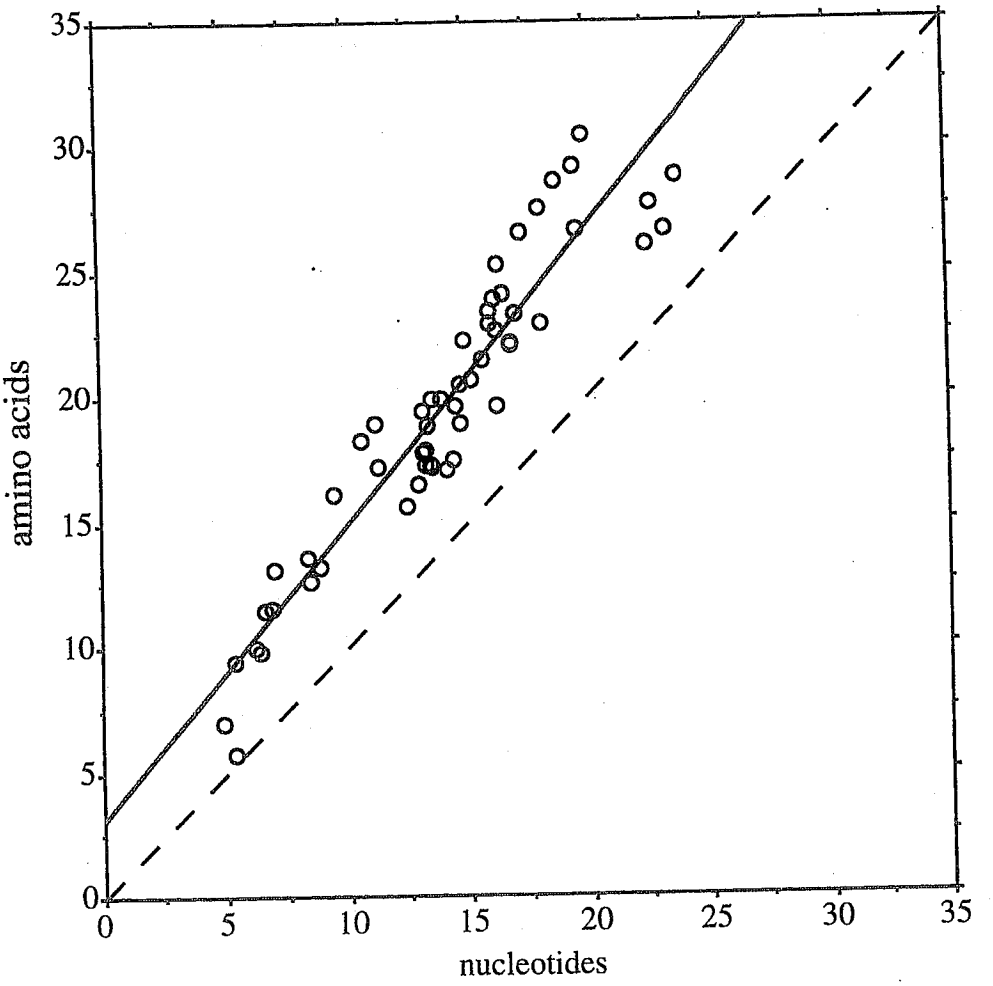


Figure 17.6 Relation between *Gcyc* nucleotides and *Gcyc* amino acid divergence (%) in pairwise sequence comparisons of the 'genus' data set (gaps excluded in pairwise comparisons) (r = 0.946).

synonymous and non-synonymous substitutions revealed a significant correlation (r = 0.860; $P > 0.001$) (Fig. 17.7); the intercept of the regression curve at 4.8% accords with the fact that at shorter taxonomic distances more synonymous substitutions occur than non-synonymous changes. However, with increasing divergence between the taxa, proportionally more non-synonymous substitutions have taken place, indicated by a slope of less than 45° in the regression curve, resulting in proportionally higher values of amino acid changes, thus divergence, compared to nucleotide changes.

Plotted against sequence divergence figures, ITS values are consistently higher (r = 0.876; $P > 0.001$), and *trnL+F* consistently lower (r = 0.86; $P > 0.001$) than *Gcyc* figures (Fig. 17.8). In pairwise sequence comparisons, divergence ranged from 0.2% (*Ramonda myconi – Jankaea heldreichii*) to 7.0% (*Sinningia schiffneri – Primulina*

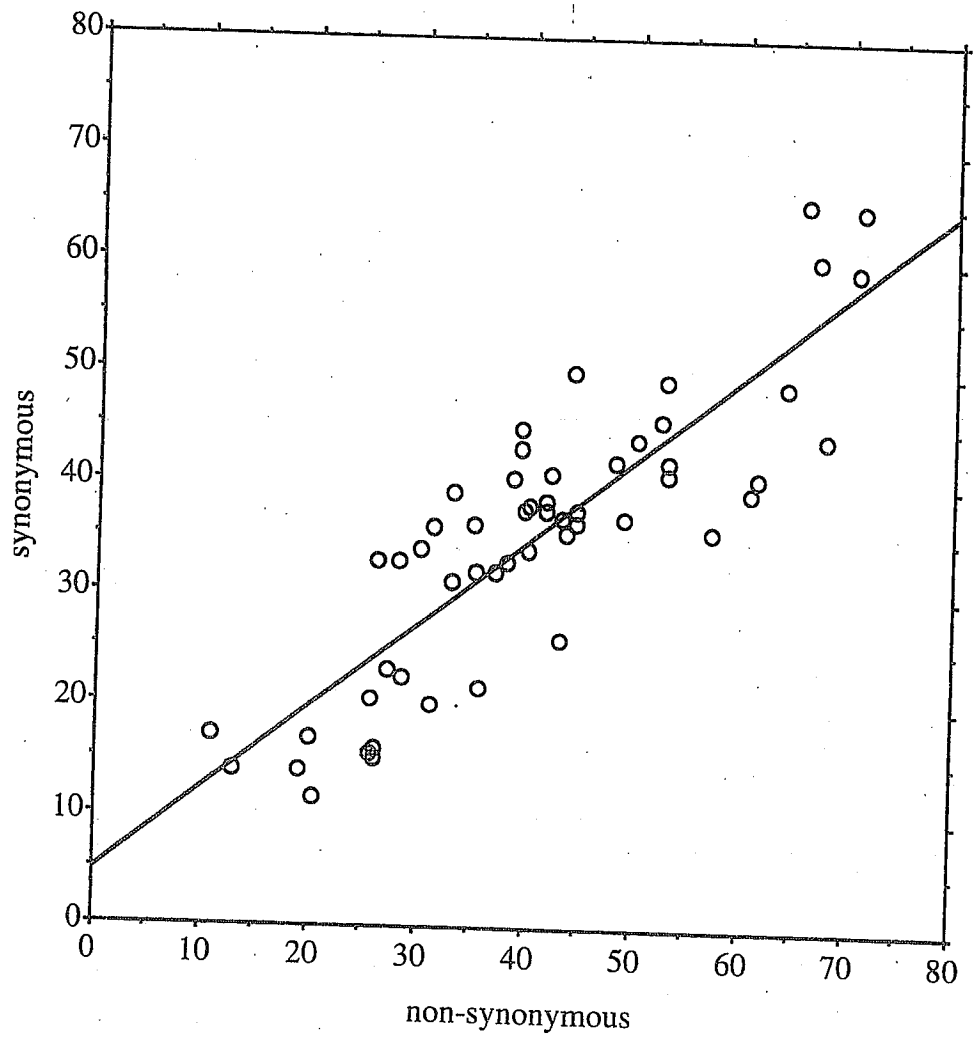


Figure 17.7 Relation between synonymous and non-synonymous changes across the *Gcyc* nucleotide data matrix of the 'genus' data set in pairwise comparisons (gaps excluded in pairwise comparisons) (r = 0.86).

tabacum) in cpDNA, 2% (*Jankaea heldreichii – Ramonda myconi/Ramonda nathaliae*) to 34.5% (*Sinningia schiffneri – Streptocarpus dunnii*) in ITS, and 4.9% (*Ramonda nathaliae – Jankaea heldreichii*) to 23.7% (*Sinningia schiffneri – Streptocarpus dunnii*) in *Gcyc* (Table 17.3).

17.3.2 The 'species' data set

17.3.2.1 Sequence comparison between ITS and Gcyc

Unlike the 'genus' data set the alignment of ITS sequences of the 'species' analysed required the insertion of 17 gaps only, of 1–4 bp length. Not unexpectedly, a high

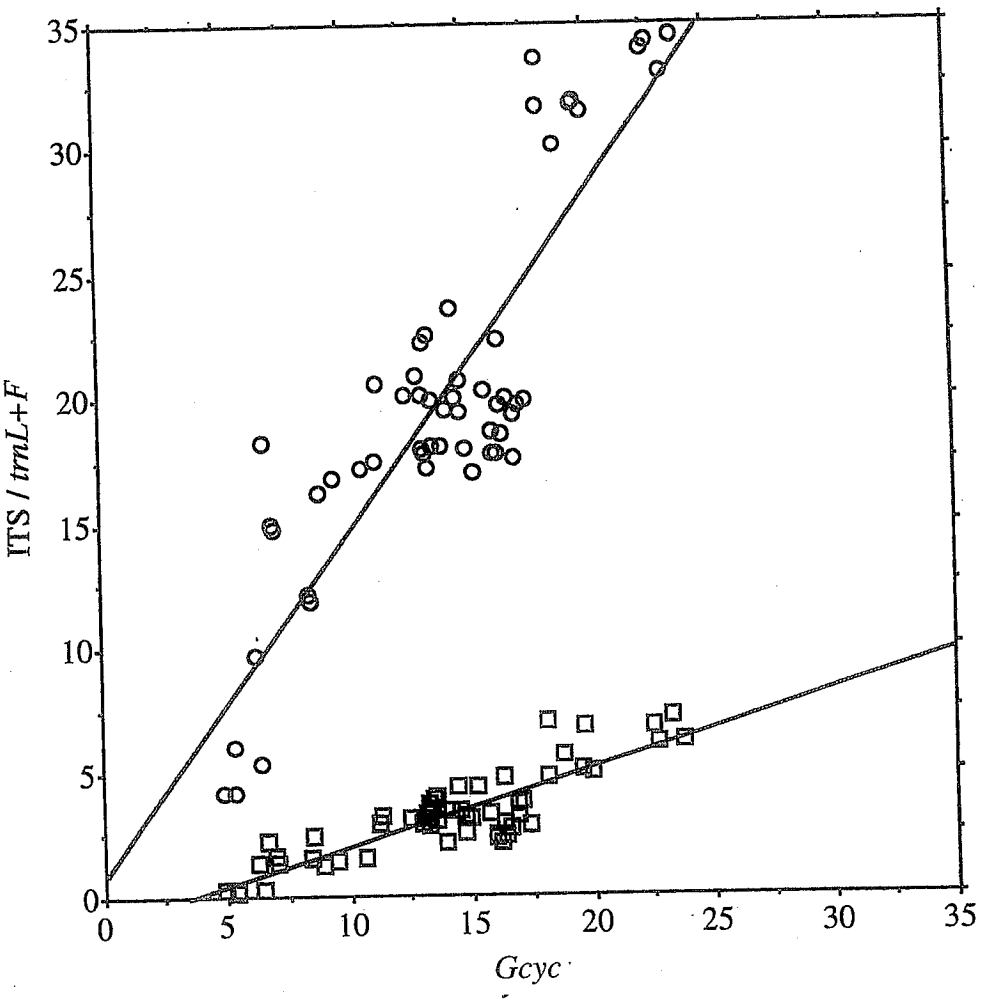


Figure 17.8 Relation between the *Gcyc* nucleotide and *trnL*+F (□, r = 0.86) and ITS (○, r = 0.876) sequence divergences (%) in pairwise comparisons of the 'genus' data set (gaps excluded in pairwise comparisons).

proportion of the unambiguously aligned sites (77.8%) were constant, only 8.7% were potentially informative phylogenetically, and 13.5% were autapomorphies (Table 17.4).

The *Gcyc* gene fragment used for the species comparison was 372 bp long, and required no alignment gaps. Similar to ITS sequences, 82.5% of the sites were constant, but only half as many, 4.9%, were potentially informative (Table 17.4). Translation of the nucleotide matrix to amino acids resulted in 124 amino acids. Compared to the nucleotides, the percentage of constant sites decreased to 70.2%, while the percentages of informative sites and of autapomorphies increased to 8.9% and 21.0%, respectively. *Saintpaulia difficilis*, *S. magungensis*, *S. teitensis*, and *S. grotei* and *S. magungensis* var. *occidentalis* had identical amino acid compositions (Table 17.4).

Table 17.4 Sequence characteristic of ITS 1 and ITS 2, and *Gcyc* regions of *Streptocarpus holstii* and 20 *Saintpaulia* taxa: 'species' data set.

Parameter	*ITS 1 and ITS 2*	*Gcyc nucleotides*	*Gcyc amino acids*
Length range, bp	462–471	–	–
Length mean, bp	466.7	372	124
Aligned length, bp	483	372	124
G+C content mean, %	53.1	42.0	–
Sequence divergence (in–outgroup)	46–58	26–33	14–19
Sequence divergence, (in–outgroup) %	10.0–12.5	7.0–8.9	11.3–15.3
Sequence divergence (ingroup)	0–73	0–19	0–10
Sequence divergence, (ingroup) %	0–15.8	0–5.1	0–8.1
Number of indels, (informative)	17 (7)	0	0
Size of indels, bp	1–4	–	–
Number of constant sites (%)	376 (77.8)	307 (82.5)	87 (70.2)
Number of variable sites (%)	107 (22.2)	65 (17.5)	37 (29.8)
Number of autapomorphic sites (%)	65 (13.5)	47 (12.6)	26 (21.0)
Number of informative sites (%)	42 (8.7)	18 (4.9)	11 (8.9)
Transitions (min)	70	37	–
Transversions (min)	55	30	–
Transitions/transversions	1.27	1.23	–
Average number of steps per character	0.254	0.185	0.331

17.3.2.2 Phylogenetic reconstruction: ITS vs Gcyc

Parsimony analysis of both sets of 'species' data resulted generally in numerous most parsimonious trees. Analysis of aligned ITS1 and ITS2 sequences resulted in four most parsimonious trees, of 123 steps length (CI = 0.919; RI = 0.851; RC = 0.782). The strict consensus of those trees resolved biogeographic relations amongst *Saintpaulia* taxa (Fig. 17.9). *Saintpaulia goetzeana* (BS = 98%, DI = +12) and *Saintpaulia teitensis* (BS = 100%, DI = +10) are separated from the rest, which form a polytomy of three species (*S. intermedia*, *S. pendula*, *S. rupicola*) and two groups, one consisting of *S. magungensis* var. *occidentalis*, *S. brevipilosa* and *S. nitida*, and the second group consisting of the *S. ionantha*-complex (Möller and Cronk, 1997b) and including the rest of the taxa (BS = 64%, DI = +1).

For the *Gcyc* nucleotide matrix 2210 most parsimonious trees of 69 steps length (CI = 0.829; RI = 0.733; RC = 0.608) were obtained. The strict consensus tree was less resolved than the rDNA tree and the clades only weakly supported (Fig 17.10). *Saintpaulia goetzeana* was separated from the rest of the taxa (BS = 95%, DI = +5), but *S. teitensis* nested within the main '*ionantha*'-complex. The rDNA based '*ionantha*'-complex was not completely reflected in the *Gcyc* topology (compare Figs. 17.9 and 17.10): *S.* cf. *ionantha*, *S.* Sigi Falls and *S. diplotricha* formed a group outside the complex (DI = +1); *S. magungensis* was part of a polytomy involving the latter group and the Nguru Mt taxa (*S. brevipilosa*, *S. nitida*) and *S. rupicola* (DI = +1); and *S. pendula* and *S. intermedia* were included in the '*ionantha*'-complex.

Parsimony analysis on the *Gcyc* amino acid data set gave 3890 most parsimonious trees of 41 steps length (CI = 0.829; RI = 0.708; RC = 0.587). As expected, the resolution of the strict consensus tree was very low (data not shown) and only *S. goetzeana* separated from the rest of the taxa which were arranged in a polytomy.

ITS (rDNA) Data

Gcyc Data

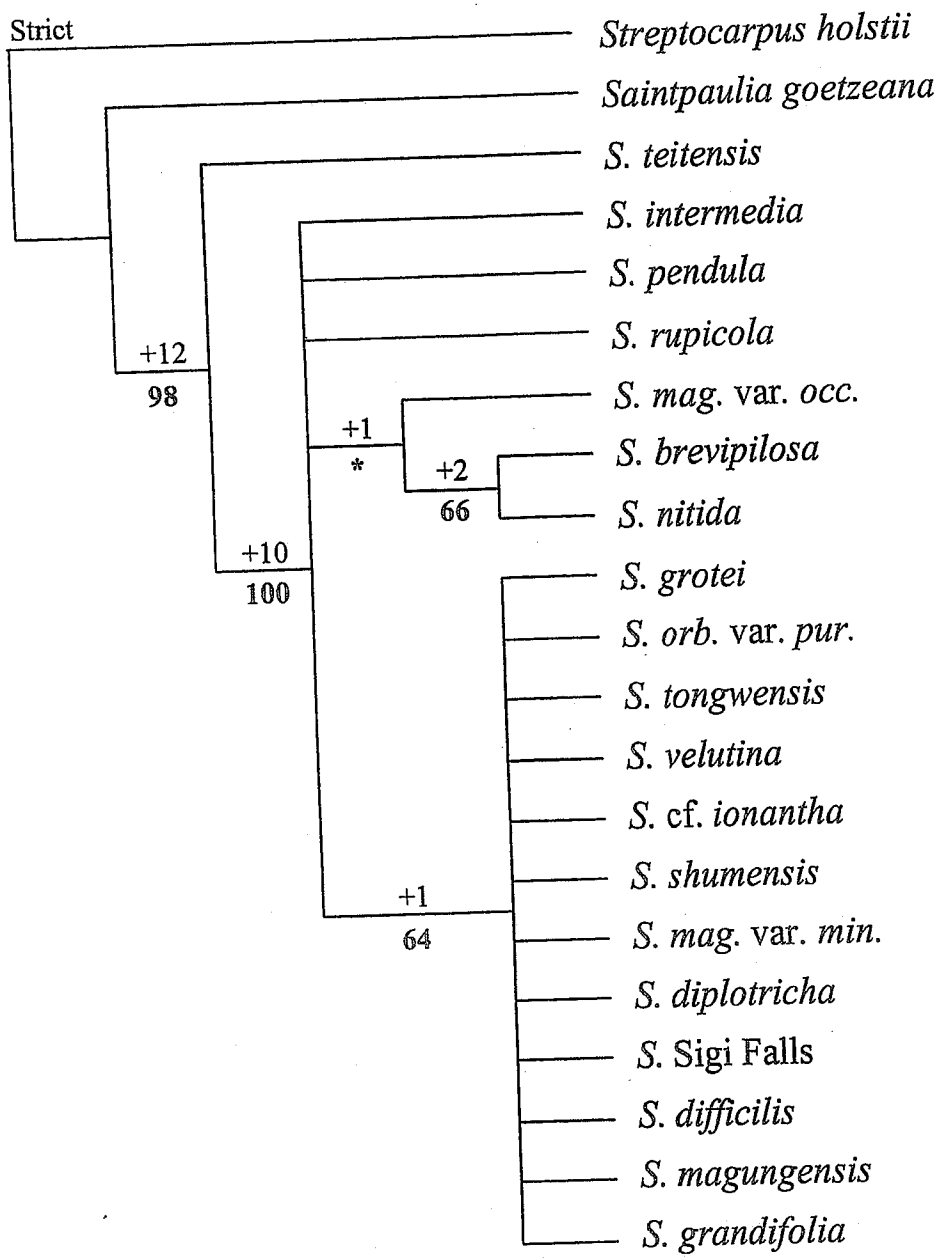


Figure 17.9 Strict consensus tree of 4 most parsimonious trees for 20 *Saintpaulia* taxa of 123 steps length based on ITS 1 and ITS 2 sequence data sets (CI = 0.919; RI = 0.851; RC = 0.782; inclusive uninformative sites). Numbers above branches indicate decay indices. Numbers below branches (**bold** print) indicate bootstrap values of 1000 replicates. * indicates branch that collapses in bootstrap analysis.

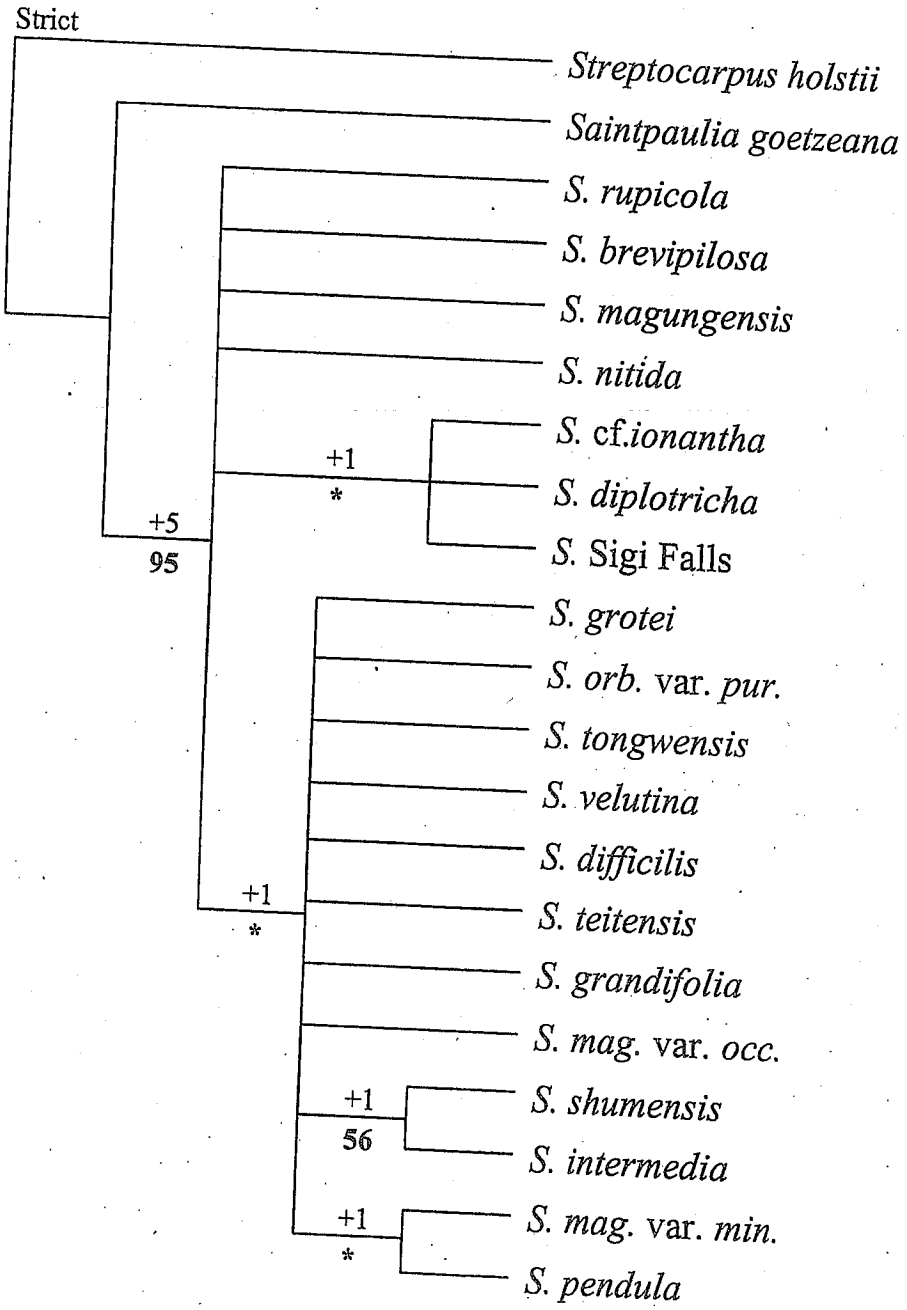


Figure 17.10 Strict consensus tree of 2210 most parsimonious trees for 20 *Saintpaulia* taxa of 69 steps length based on *Gcyc* nucleotide data (CI = 0.829; RI = 0.733; RC = 0.608; inclusive uninformative sites). Numbers above branches indicate decay indices. Numbers below branches (**bold** print) indicate bootstrap values of 200 replicates. * indicates branches that collapse in bootstrap analysis.

17.3.2.3 Comparative pairwise sequence divergence

Within the *Saintpaulia* ingroup accessions, sequence divergence of ITS ranged from 0–15.8% (*S. goetzeana* – *S. nitida*), and from 10.0–12.5% between ingroups and the outgroup. No sequence variation was observed (excluding gap positions) between nine *Saintpaulia* accessions: *S. difficilis*, *S. diplotricha*, *S. grandifolia*, *S. grotei*, *S. magungensis*, *S. magungensis* var. *minima*, S. Sigi Falls, *S. tongwensis*, and *S. velutina*.

Sequence divergence of the *Gcyc* nucleotide matrix within the ingroup was 0–5.1% (*S. goetzeana* – *S. diplotricha*), similar to ITS. However, sequence divergence between in- and outgroup was 7–8.9%, lower than the ITS matrix. *S. grotei* and *S. magungensis* var. *occidentalis* had identical sequences, and those and *S. difficilis*, *S. grotei*, *S. magungensis* var. *minima*, *S. orbicularis* var. *purpurea*, *S. magungensis* var. *occidentalis* and *S. tongwensis*, differed by two sites at most in pairwise sequence divergence comparisons. Maximum sequence divergence was observed between *S. goetzeana* and *S. diplotricha* (Table 17.4).

As in the 'genus' data set, translation of *Gcyc* nucleotides into amino acids resulted in higher divergence percentage values (Table 17.4). Amongst the ingroup taxa sequence divergence was between 0 (*S. grotei* – *S. magungensis* var. *occidentalis*) and 8.1% (*S. goetzeana* – *S. pendula*), and 11.3–15.3% between in- and outgroup.

A graph of the sequence divergence values of pairwise comparisons for rDNA and *Gcyc* indicates a bipartite pattern for the 'species' data set ($r = 0.627$; $P > 0.001$) (Fig. 17.11). At low divergence *Gcyc* was more variable than ITS, up to values of around 2% sequence divergence. Above this divergence ITS showed a greater sequence divergence.

17.4 DISCUSSION

17.4.1 Comparative rates of gene evolution

In common with other chloroplast regions, the *trnL+F* intron/spacer is relatively slowly evolving (although the spacer appears to be noticeably more variable than the intron, probably due to functional constraints on the latter) (Fig. 17.2). The two nuclear genes (*Gcyc* and ITS) both evolve more rapidly than the cpDNA, even though *Gcyc* is a coding region: this is concordant with the general phenomenon that chloroplast genome evolution is conservative. What is more interesting, however, is the rate comparison between *Gcyc* and ITS. At high levels of sequence divergence (inter-genus comparisons) ITS appears to be more divergent (Fig. 17.11). However, at low level of sequence divergence (inter-species comparisons) *Gcyc* appears to be evolving more quickly. How can this be? Our explanation is that this is an artefact of the direct sequencing approach of PCR products employed. Because PCR-based sequencing generated a consensus sequence in a repeat family like ITS (Hershkovitz *et al.*, 1999 – this volume), mutations will not be seen unless they are near complete fixation. At lower levels they will either be scored as polymorphic or not recognised above background noise. The intercept of the line in Fig. 17.11 is therefore a result of the time taken for fixation, both in the genome and in the population. *Gcyc* is a single copy gene and mutations can be fixed more rapidly, so the intercept is at or near zero. At low levels of divergence ITS

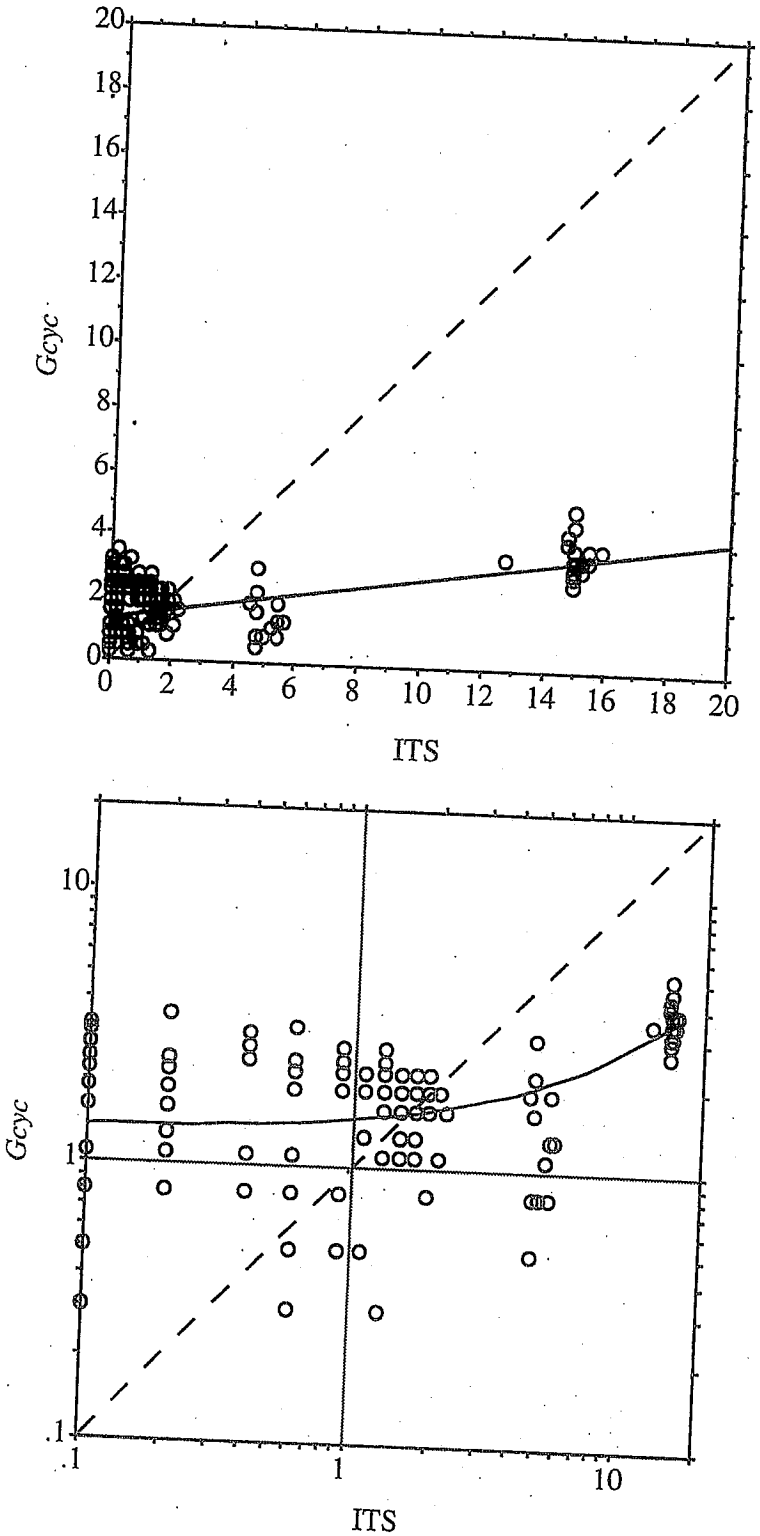


Figure 17.11 Relation between *Gcyc* nucleotide and ITS sequence divergences (%) in pairwise comparisons of the 'species' data set (gaps excluded in pairwise comparisons) (r = 627). Top: linear scale; bottom: log scale.

mutation will not be seen, whereas *Gcyc* mutations will be detected. At higher divergence levels however, ITS can continue to accumulate changes rapidly due to the low functional constraint on ITS sequences (as reflected by the slope of the graph). Although post-transcriptional processes require conservation in ITS at short putative recognition sites (Liu and Schardl, 1994), there is otherwise little functional constraint as preservation of secondary structure does not necessarily require sequence conservation (Mai and Coleman, 1997). On the other hand, *Gcyc* is more functionally constrained. This would be expected to place considerable limits on divergence rates.

17.4.1.1 Rate constancy and molecular clocks

DNA molecules with no functional constraints are believed to evolve linearly over time; they accumulate mutational changes (substitutions, insertion/deletion events) independently from selective pressures, making them useful as 'molecular clocks' (Clegg *et al.*, 1994; Sang *et al.*, 1995). Chloroplast spacers are thought to evolve in a clock-like way (Clegg *et al.*, 1994) as they are not subject to molecular drive and are not considered to be under any strong functional constraint. However, when comparing divergence rates between cpDNA and *Gcyc* it appears that there exists a strong linear relationship, and the divergence rate seems to be constant. The same is true for ITS and *Gcyc* divergence (Fig. 17.8). This indicates that all three sequences evolve in a clock-like manner, and could be used for estimating divergence times if the date of the origin of Gesneriaceae were known in order to provide a calibration point. Unfortunately the relationship between the Gesneriaceae and its closest allied families, e.g. Scrophulariaceae, Bignoniaceae, Lentibulariaceae, or Acanthaceae is not very well established (Chase *et al.*, 1993).

17.4.2 Utility of cyc and ITS for phylogenetic analysis

One consequence of molecular drive is that at very low levels of divergence ITS is not suitable for resolving patterns of hierarchical relationships (Möller and Cronk, 1997b). At this level the rate of synonymous substitutions allows single-copy nuclear genes, such as *Gcyc*, to be phylogenetically more useful (Doyle and Doyle, 1999 – this volume). At higher levels, however, it is a useful conservative source of phylogenetic information, in the same way as *trnL+F*. Its variation pattern brackets that of ITS, being both slower and faster. The similarity of phylogenetic trees derived from *Gcyc*, ITS and *trnL+F* (Fig. 17.4) indicates that all these gene regions have phylogenetic utility. However, any allelic variation in scn genes may cause problems for phylogenetic reconstruction.

17.4.3 Gcyc and the evolution of floral symmetry patterns

Cycloidea is a functional gene involved in establishing floral zygomorphy (Luo *et al.*, 1996), and it is interesting to consider how its evolution varies over morphological transitions to actinomorphy in the Gesneriaceae. If actinomorphy is caused by a loss-of-function mutation at the *Gcyc* locus, we would expect that, relieved from selection, the *Gcyc* gene would evolve rapidly and quickly become frame-

shifted or would contain stop codons. We checked the sequences of the actinomorphic taxa, *Conandron ramondioides*, *Ramonda myconi* and *Ramonda nathaliae*, for frame shifts or stop codons and found none. Furthermore, indels in actinomorphic species are all in multiples of three nucleotides, thus preserving the reading frame. This is a strong indication that the gene in those taxa is still intact and potentially functional. Additionally, there is a reversal to zygomorphy in *Jankaea heldreichii* from an actinomorphic *Ramonda*, further indicating a conservation of functional *Gcyc*.

To investigate the possibility of a differential substitution rate of *Gcyc* for zygomorphic and actinomorphic taxa, two approaches were investigated: (1) the branch lengths of the *Gcyc* tree (Fig. 17.4c) were plotted, for actinomorphic and zygomorphic taxa separately, against the branch lengths of the cpDNA (Fig. 17.4a) and ITS (Fig. 17.4b) trees, respectively (Fig. 17.12). For this analysis it was necessary to alter the tree topology of the *Gcyc* tree to match the tree based on combined data of all three genes (Fig. 17.5); and (2) for a *Gcyc* internal analysis of different rates of evolution between actinomorphic and zygomorphic taxa, separate trees were produced using either the 1st and 2nd or the 3rd codon position only. Trees that matched the topology of the combined data tree were chosen, and the branch lengths of trees derived from 1st and 2nd codons were plotted against those from the 3rd codon position.

In the first case, *Gcyc* vs. ITS or *Gcyc* vs. cpDNA, the branch lengths for actinomorphic taxa were not significantly different from zygomorphic taxa, indicating no differences in the rate of evolution (Fig. 17.12). In the second case, plotting branch lengths for the different codon positions did not indicate a different rate of evolution for actinomorphic taxa either (Fig. 17.13). If the gene is not functional, those taxa would have similar rates at all codon positions, and thus would follow a 45° diagonal line, whereas taxa with coding *Gcyc* genes would have significantly higher rates at the 3rd codon position, as indicated in Fig. 17.13. There is no notable shift in substitution rates over changes to actinomorphy, and no obvious difference in the substitution rate between codon positions (Fig. 17.13) – strongly implying further that the gene is functional and that it is still under selectional constraint.

Naturally occurring reversion to actinomorphy is therefore very different from actinomorphy caused by loss-of-function 'peloric' mutants such as those in *Antirrhinum majus* (Luo *et al.*, 1996). We suggest therefore that *Gcyc* is still expressed, probably at an early stage of development. Early expression of *Gcyc* may be involved in controlling the number of primordia and establishing them in register with the axis of symmetry of the flower (Luo *et al.*, 1996). We suggest that expression has been reduced in the later stages of floral development, which is the stage at which differential growth of the primordia establishes asymmetry (zygomorphy). Thus evolution of actinomorphy from zygomorphy may have come about gradually by selection on modifying genes which progressively alter the expression of *Gcyc*. This idea could be tested by examining the expression pattern of *Gcyc* in secondarily actinomorphic flowers using RNA *in situ* hybridisation. It could also be tested by genetic analysis of the putative modifiers by means of crosses. In the zygomorphic–actinomorphic hybrid *Jankaea heldreichii* × *Ramonda myconi* (× *Jankaemonda vandedemii*) the flowers are always actinomorphic, implying that this trait is dominant. However, the segregation in the F2 has never

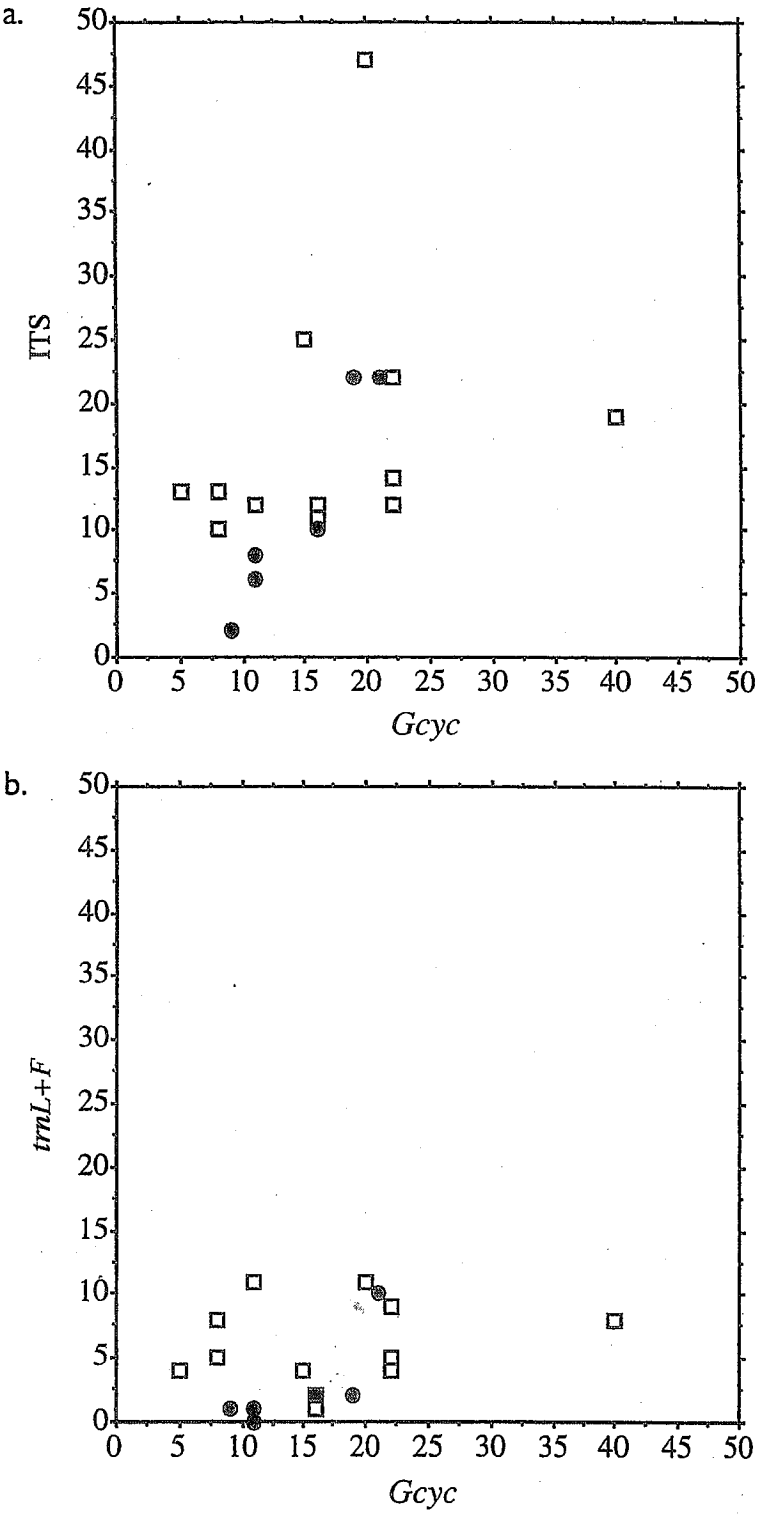


Figure 17.12 Relation between branch lengths for actinomorphic (●) and zygomorphic (□) taxa of the 'genus' data set for trees of *Gcyc* nucleotides and a) ITS or b) *trnL+F*.

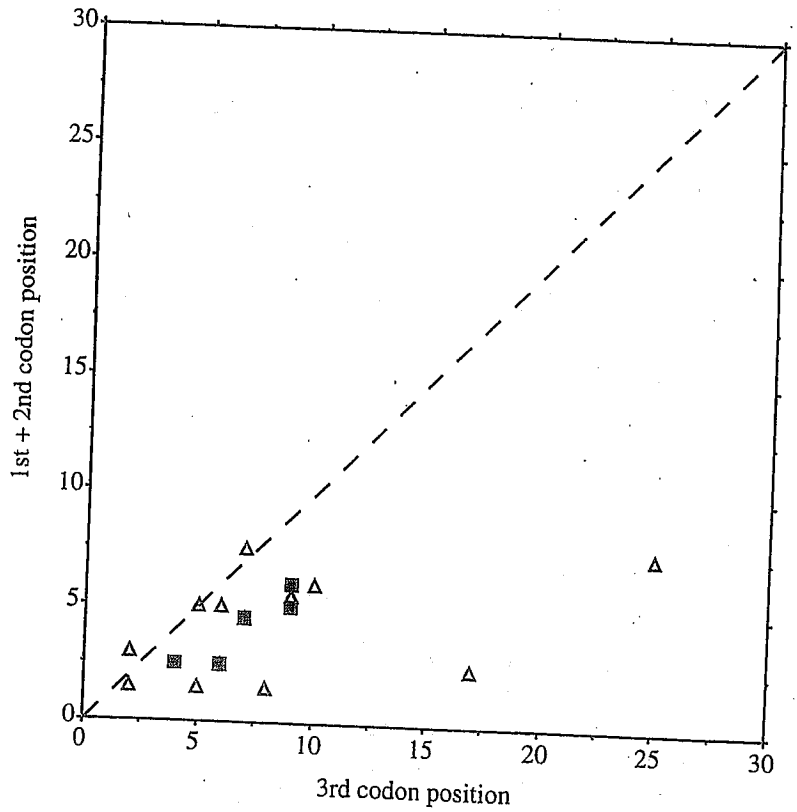


Figure 17.13 Relation between branch lengths for actinomorphic (■) and zygomorphic (△) taxa of trees based on the 1st and 2nd or the 3rd codon position of the *Gcyc* nucleotide 'genus' data set.

been studied because the plants are difficult to grow and the F1 shows a high degree of sterility. In induced peloric mutants of *Antirrhinum* actinomorphy is recessive. Alternative explanations of the intact ORF in the actinomorphic species are: (1) *Gcyc* expression has been reduced in later stages of flower development by mutations in its own regulatory region that still preserve the early expression; (2) the *Gcyc* gene studied is not the orthologue of the *Antirrhinum cyc* gene and the real orthologue is mutated in actinomorphic species; and (3) in Gesneriaceae, the floral asymmetry is generated independently of the *cyc*-like genes. We believe these are less likely hypotheses than the modifier hypothesis presented above. Nevertheless they could be tested by performing *in situ* RNA hybridisations using *Ramonda*. This study reveals the caution which has to be exercised in extrapolating from loss-of-function mutations induced in developmental genes to real evolutionary processes.

ACKNOWLEDGEMENTS

The authors thank Prof. E. Coen, John Innes Centre, Norwich, for kindly supplying materials relating to the *cyc* gene and for much help and advice. We also thank

Prof. D. Schweizer and Dr. A. Bachmeir, Dept. for Cytology and Genetics, Institute of Botany and Botanical Garden, University of Vienna, Austria, for suggestions on data analysis; H.-E. Wichert, Herrenhäuser Gärten, Hannover, Germany, and M.-J. Evans, American Gloxinian and Gesneriad Society, Randolph, NJ, USA, for kind donation of plant material of *Saintpaulia*, and Prof. D. Vokou for sending flower material of *Jankaea*; H. Sluiman, J. Preston and C. Guihal for their technical support; U. Gregory, D. Mitchell, and J. Main for the maintenance and expansion of the Gesneriaceae collection; the Regius Keeper and staff at the Royal Botanic Garden Edinburgh for the use of their excellent research facilities. We thank the Institute of Cell and Molecular Biology, University of Edinburgh, for access to sequencing facilities and N. Preston for assistance. The receipt of a Leverhulme Trust Award, No. F/771/B, is gratefully acknowledged.

REFERENCES

Bakker, F. T., Culham, A. and Gibby, M. (1999) Phylogenetics and diversification in *Perlargonium*, in *Molecular systematics and plant evolution*, (eds P. M. Hollingsworth, R. M. Bateman and R. J. Gornall), Taylor & Francis, London, pp. 353–374.

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.

Bremer, K. (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution*, 42, 795–803.

Burtt, B. L. and Wiehler, H. (1996) Classification of the family Gesneriaceae. *Gesneriana*, 1(1), 1–4.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D. *et al.* (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden*, 80, 528–580.

Clegg, M. T., Gaut, B. S., Learn, G. H. and Morton, B. R. (1994) Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the USA*, 91, 6795–6801.

Coen, E. S. and Nugent, J. M. (1994) Evolution of flowers and inflorescences. *Development* (Suppl.), 107–116.

Cubas, P., Lauter, N., Doebley, J. and Coen, E. (1999) The TCP domain: a motif found in proteins regulating plant growth and development. *The Plant Journal*, 18, 215–222.

Doebley, J., Stec, A. and Hubbard, L. (1997) The evolution of apical dominance in maize. *Nature*, 386, 485–488.

Donoghue, M. J., Olmstead, R. G., Smith, J. F. and Palmer, J. D. (1992) Phylogenetic relationships of Dipsacales based on *rbc*L sequences. *Annals of the Missouri Botanic Garden*, 79, 333–345.

Dover, G. A. (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics*, 2, 159–165.

Downie, S. R. and Katz-Downie, D. S. (1996) A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal spacer sequences. *American Journal of Botany*, 83, 234–251.

Doyle, J. J. and Doyle, J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, 19, 11–15.

Doyle, J. J. and Doyle, J. L. (1999) Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae, in *Molecular systematics and plant evolution*, (eds P. M. Hollingsworth, R. M. Bateman and R. J. Gornall), Taylor & Francis, London, pp. 229–254.

Elder, J. F. and Turner, B. J. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology*, 70, 297–320.

Ennos, R. A., Sinclair, W. T., Hu, X-S. and Langdon. A. (1999) Using organelle markers to elucidate the history, ecology and evolution of plant populations, in *Molecular systematics and plant evolution*, (eds P. M. Hollingsworth, R. M. Bateman and R. J. Gornall), Taylor & Francis, London, pp. 1–19.

Farris, J. S. (1989) The retention index and homoplasy excess. *Systematic Zoology*, 38, 406–407.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39, 783–791.

Hagemann, R. and Schröder, M.-B. (1989) The cytological basis of chloroplast inheritance in angiosperms. *Protoplasma*, 152, 57–64.

Hershkovitz, M. A., Zimmer, E. A. and Hahn, W. J. (1999) Ribosomal DNA sequences and angiosperm systematics, in *Molecular systematics and plant evolution*, (eds P. M. Hollingsworth, R. M. Bateman and R. J. Gornall), Taylor & Francis, London, pp. 268–326.

Hillis, D. M., Moritz, C., Porter, C. A. and Baker, R. J. (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science*, 251, 308–310.

Kluge, A. G. and Farris, J. S. (1969) Quantitative phyletics and the evolution of anurans. *Systematic Zoology*, 18, 1–32.

Kumar, S., Tamura, K. and Nei, M. (1993) *MEGA: molecular evolutionary genetics analysis, version 1.01*. The Pennsylvania State University, University Park, PA.

Liu, J.-S. and C. L. Schardl (1994) A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. *Plant Molecular Biology*, 26, 775–778.

Luo, D., Carpenter, R., Vincent, C., Copsey, L. and Coen, E. (1996) Origin of floral asymmetry in *Antirrhinum*. *Nature*, 383, 794–799.

Maddison, W. P. and Maddison, D. R. (1992) *MacClade, version 3.01*. Sinauer, Sunderland, MA.

Mai, J. C. and Coleman, W. (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution*, 44, 258–271.

Mogensen, H. L. (1996) The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*, 83, 383–404.

Möller, M. and Cronk, Q. C. B. (1997a) Origin and relationships of *Saintpaulia* H.Wendl. (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany*, 84, 956–965.

Möller, M. and Cronk, Q. C. B. (1997b) Phylogeny and disjunct distribution: evolution of *Saintpaulia* (Gesneriaceae). *Proceedings of the Royal Society of London, Series B*. 264, 1827–1836.

Reboud, X. and Zeyl, C. (1994) Organelle inheritance in plants. *Heredity*, 72, 132–140.

Sang, T., Crawford, D. J., Stuessy, T. F. and O, M. S. (1995) ITS sequences and the phylogeny of the genus *Robinsonia* (Asteraceae). *Systematic Botany*, 20, 55–64.

Schuler, G. (1995) *MACAW: Multiple alignment construction & analysis workbench*. National Center for Biotechnology Information (NCBI), Bethesda, Maryland.

Smith, J. F. (1996) Tribal relationships within Gesneriaceae: a cladistic analysis of morphological data. *Systematic Botany*, 21, 497–513.

Smith, J. F., Wolfram, J. C., Brown, K. D., Carroll, C. L. and Denton, S. (1997) Tribal relationships within Gesneriaceae: evidence from DNA sequences of the chloroplast gene *ndh*F. *Annals of the Missouri Botanical Garden*, **80**, 50–66.

Soltis, D. E. and Kuzoff, R. K. (1995) Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution*, **49**, 727–742.

Susanna, A., Jacas, N. G., Soltis, D. E. and Soltis, P. S. (1995) Phylogenetic relationships in tribe Cardueae (Asteraceae) based on ITS sequences. *American Journal of Botany*, **82**, 1056–1068.

Swofford, D. L. (1993) *PAUP: phylogenetic analysis using parsimony, version 3.1.1.* Illinois Natural History Survey, Champaign, IL.

Taberlet, P., Gielly, L., Pautou, G. and Bouvet, J. (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1103–1109.

Tilney-Bassett, R. A. E. (1994) Nuclear controls of chloroplast inheritance in higher plants. *Journal of Heredity*, **85**, 347–354.

White, T. J., Bruns, T., Lee, S. and Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, in *PCR protocols*, (eds M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White), Academic Press, London, pp. 315–322.

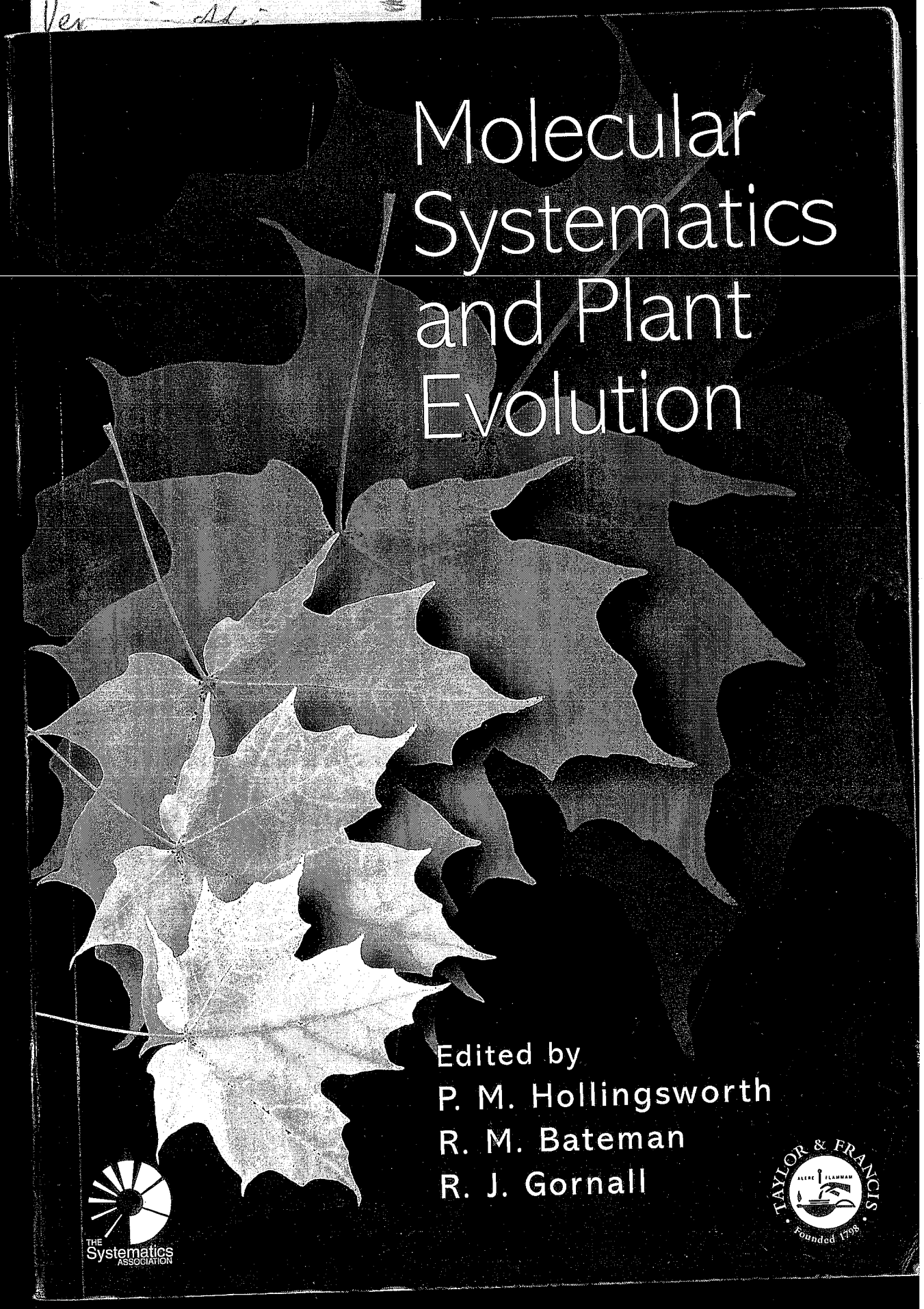
Wojciechowski, M. F., Sanderson, M. J., Baldwin, B. G. and Donoghue, M. J. (1993) Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany*, **80**, 711–722.

The Systematics Association Special Volume Series 57

Molecular Systematics and Plant Evolution

Edited by Peter M. Hollingsworth, Richard M. Bateman and Richard J. Gornall

London and New York


Molecular Systematics and Plant Evolution
Edited by
P. M. Hollingsworth
R. M. Bateman
R. J. Gornall
THE Systematics ASSOCIATION
TAYLOR & FRANCIS
Founded 1798

First published 1999
by Taylor & Francis Limited
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Taylor & Francis Limited is an imprint of the Taylor & Francis Group

Typeset in Sabon by
The Florence Group Stoodleigh, Devon
Printed and bound in Great Britain by
T. J. International, Padstow, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data are available

ISBN 0–7484–0908–4 (pbk)
0–7484–0907–6 (hbk)