

A Molecular Phylogeny for *Cercocarpus* H.B.K. (Rosaceae) Using the External Transcribed Spacer of the Nuclear Ribosomal Repeat

Brian D. Vanden Heuvel
C. Randal Linder

Abstract—*Cercocarpus* H.B.K. (Rosaceae) taxa are important members of the plant communities of the western states and Mexico, yet the systematics of this genus are unknown primarily from lack of clear morphological delimitations between taxa. In recent years, molecular data have proven useful for resolving relationships among species and the diversity within species that have otherwise remained elusive. We will report here preliminary data on the phylogenetic utility of two noncoding regions within the nuclear ribosomal DNA (nrDNA) for *Cercocarpus*; the internal transcribed spacer (ITS) region and the external transcribed spacer (ETS) region. The ITS and ETS were amplified and sequenced in five individuals representing four *Cercocarpus* taxa. Maximum parsimony and maximum likelihood phylogenies were created from the two aligned data sets and compared. We found that the ETS has considerably more phylogenetically informative sequence variation than the ITS, which had almost no signal. Further sampling of ETS sequences in *Cercocarpus* taxa produced four main ETS types. The four ETS types were alignable only in the extreme 5' and 3' regions of the ETS sequence. Large regions of the ETS sequences were unalignable between different ETS types. Phylogenetic analysis of a reduced data set including just the regions in common between ETS types suggest that the different ETS types in *Cercocarpus* diverged prior to the origin of the genus, making them paralogous. This deep coalescence requires that a single ETS type be used for reconstruction.

Introduction

Cercocarpus H.B.K. (Rosaceae) is a New World genus composed of montane xerophytic shrubs and trees, found in deserts, chaparral, and mountainous regions of western North America with a center of diversity in northern Mexico. Botanists have paid attention to the ecology and management of *Cercocarpus* in the United States because the evergreen leaves found on most members of the genus have high levels of protein and are an important winter forage for wildlife and livestock in western states (Blauer and others 1975; Davis and Brotherson 1991).

The systematics of *Cercocarpus* are largely unknown; the genus lacks a consensus classification or phylogeny. Since it was first described, *Cercocarpus* has undergone three revisions (Schneider 1905; Rydberg 1913; Martin 1950), each varying widely in the ranks assigned to taxa and the number of species. Much of the current taxonomy of *Cercocarpus* is based on quantitative characteristics involving leaf morphology (Schneider 1905; Rydberg 1913; Martin 1950). Confusion about the number and circumscription of species within *Cercocarpus* arises primarily from the lack of clear morphological delimitations between taxa. Variations in leaf morphology show a continuum within and between taxa, making it difficult to define boundaries among species (Brayton and Mooney 1966; Mortenson 1973). A phylogeny based on leaf characters is poorly supported due to severe homoplasy.

In recent years, molecular data have proved useful for resolving relationships among morphologically similar species and for understanding the diversity within species (Soltis and Soltis 1998 and references therein). A phylogeny and a clear definition of taxa using molecular markers within *Cercocarpus* is critical to an understanding of the evolution and systematics of *Cercocarpus*. Knowledge of the systematics of *Cercocarpus* may enhance its breeding and help prevent undesirable hybridization of species and ecotypes planted in reclaimed areas using bulk seed.

We report here preliminary results on the phylogenetic utility of two noncoding regions within the 18S-26S nuclear ribosomal DNA (nrDNA) repeat for *Cercocarpus*: the internal transcribed spacer (ITS) region and the external transcribed spacer (ETS) region (fig. 1). Both the ITS region and

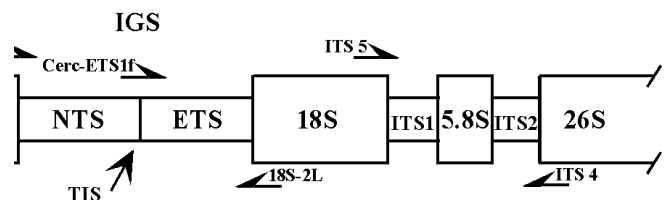


Figure 1—Representation of the 18S-26S ribosomal repeat in the nuclear genome. The genes (26S, 5.8S, and 18S) are shown by the large boxes. Transcription begins at the TIS. Both the ETS and both ITS regions are removed after transcription. The general location and direction of the primers used in this study are shown by arrows and italicized text.

In: McArthur, E. Durant; Fairbanks, Daniel J., comps. 2001. Shrubland ecosystem genetics and biodiversity: proceedings; 2000 June 13–15; Provo, UT. Proc. RMRS-P-21. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Brian D. Vanden Heuvel is a Graduate Student at the University of Texas at Austin, Section of Integrative Biology, Austin, TX 78712. C. Randal Linder is an Assistant Professor, Section of Integrative Biology, Austin, TX 78712.

the ETS region have been used with success in phylogenetic reconstruction at the generic and specific level in other plant groups (for example, Baldwin 1995; Bena 1998; Baldwin and Markos 1998; Linder and others 2000). Recently, the ETS has been shown to contain informative variation in groups where ITS does not (Bena 1998; Baldwin and Markos 1998; Linder and others 2000).

Materials and Methods

Taxon Sampling

For this report, we follow the taxonomy of Martin (1950) with the exception of the Mexican taxa, for which we turned to an unpublished treatment by J. Henrickson (personal communication). Twenty-six taxa were sampled (table 1). In six cases more than a single individual represented a taxon. For each sample, DNA was extracted from the collections using a modified CTAB procedure (Doyle and Doyle 1987). Genomic DNA extractions were cleaned using the Elu-quik DNA purification kit (Schleicher and Schuell) prior to PCR amplification.

Amplification, Sequencing, and Analysis of the ITS Region

To test the phylogenetic utility of the ITS region in *Cercocarpus*, we selected five individuals representing four exemplar taxa of *Cercocarpus* (table 1). This sampling strategy was designed to compare the amount of variation observed between very closely related individuals (two individuals of *C. montanus* var. *montanus*) and more distantly related taxa, as determined by recent work on the leaf architecture of *Cercocarpus* by Lis (1992). The ITS region was amplified using the primers *ITS 5* and *ITS 4* (Innis and others 1990). A negative control was also used that lacked template DNA. Fifty microliter PCRs were performed using 30mM Tricine, pH 8.4, 2mM MgCl₂, 50 mM KCl, 5 percent acetamide, 100 uM of each dNTP, 1.0 unit of Taq polymerase, and 30 nM of each amplification primer. Twenty ng of template DNA was added and the reactions were overlaid with mineral oil and run on an MJ Research Thermal Cycler programmed for a hot start (95 °C, 5 min; 74 °C, 7 min) and 30 cycles of 94 °C for 30 sec, 50 °C for 45 sec, and 72 °C for 1 min. A final extension of 72 °C for 7 min was also included.

Table 1—Collection and ETS type data for the 26 populations of *Cercocarpus* and *Purshia* analyzed. Collectors: JH-James Henrickson, BVH-Brian Vanden Heuvel, BC-Bonnie Crozier. Vouchers are housed at the University of Texas (TEX).

Taxon	Population no.	Collection location and number	ETS type(s)
<i>C. montanus</i> var. <i>montanus</i>	5 ^a	Douglas Co., CO BVH 5	1
	8	Platte Co., WY BVH 8	1
	10 ^a	Custer Co., SD BVH 10	1
<i>C. montanus</i> var. <i>glaber</i>	12	Brewster Co., TX JH 20663	1
	14	Brewster Co., TX JH 20665	1
	15	Brewster Co., TX JH 20666	1
	16	Brewster Co., TX JH 20667	1
	17 ^a	Coahuila, MX JH 20439b	1
<i>C. montanus</i> var. <i>paucidentatus</i>	1	Tamaulipas, MX JH 19218	1
	27C	Eddy Co., NM BVH 27	1
<i>C. montanus</i> var. <i>blancheae</i>	19	Los Angeles Co., CA BVH 123a	2
	20	Los Angeles Co., CA BVH 123b	2
	21	Los Angeles Co., CA BVH 123c	2
	62D	Tulare Co., CA BVH 127a	1
	65C	Tulare Co., CA BVH 127b	2
	65D	Tulare Co., CA BVH 127c	1
<i>C. montanus</i> var. <i>traskiae</i>	52B	Los Angeles Co., CA JH 22312	2
<i>C. montanus</i> var. <i>minutiflorus</i>	57	San Diego Co., CA JH 22311	1, 2
<i>C. fothersgilloides</i> var. <i>mojadensis</i>	18 ^a	Coahuila, MX JH 13156	2
	66	Coahuila, MX BVH and JH 11824	3
<i>C. fothersgilloides</i> var. <i>medranoanus</i>	82	Tamaulipas, MX BVH and JH 22444	1, 3
<i>C. macrophyllus</i>	26	Coahuila, MX BC 1034	1, 2
<i>C. rzedowskii</i>	78	Tamaulipas, MX BVH and JH 22446	1
	81	Tamaulipas, MX BVH and JH 22460	4
<i>C. ledifolius</i> var. <i>intermontanus</i>	45	Box Elder Co., UT BVH 45	1
<i>C. ledifolius</i> var. <i>ledifolius</i>	41	Emery Co., UT BVH 41	1
<i>C. ledifolius</i> var. <i>intricatus</i>	25 ^a	Garfield Co., UT BVH 45	1
<i>Purshia tridentata</i>	34	La Plata Co., CO BVH 34	

^aIndividuals for which the ITS and ETS were compared.

Three μL of the reaction mixture was run on a 1.5 percent agarose gel in a 0.5X TBE buffer. PCR products were visualized with ethidium bromide under UV light. PCR products were purified (QIAquick PCR purification kit, QIAGEN) and sequenced on an ABI Prism 377 automated sequencer using standard reagents at the Core Facility of the Molecular Biology Institute at the University of Texas. Sequences were proofread and then aligned using Clustal W in MegAlign (DNASTar, 1998), followed by adjustment by eye. Aligned ITS sequence data was used to generate phylogenetic trees in PAUP 4.0.0d64 (kindly provided by David Swofford). Gaps were treated as missing data. Trees were constructed using maximum likelihood and maximum parsimony algorithms with the exhaustive search option. The specific parameters for the ML analysis were unequal transition-transversion rates and empirically determined nucleotide frequencies. Support for the monophyly of groups was evaluated using bootstrapping (Felsenstein 1985).

Developing the ETS in *Cercocarpus*

To develop the ETS region for phylogenetic reconstruction in *Cercocarpus*, we amplified the entire intragenic spacer (IGS) of *C. montanus* (collection #5) and *C. intricatus* (collection #25) (table 1) using the primers *1M* and *18S-2L* (Linder and others 2000) present in conserved regions of the 26S and 18S genes, respectively (fig. 1), and following the same PCR parameters listed above except an extension time of 2 min. per cycle. We sequenced from the 3' end of the amplified product in both taxa following the same procedures as for the ITS regions. Because the ETS region was too long to sequence with a single primer, we developed internal primers to sequence through the transcription initiation site (TIS). We identified the TIS based on similarity with the TISs in other plant groups. As a result of its role in transcription of the rDNA repeat, the TIS is under strong selection to stay conserved, making it a good choice for priming throughout *Cercocarpus*. We designed *Cerc-ETS1f* (5'-tataaagggaggcctcatt-3') to include the TIS and have high sequence similarity to both taxa (fig. 1). Using this primer in combination with *18S-2L*, we amplified the ETS reliably for all *Cercocarpus* taxa from which DNA was extracted from fresh material.

The ETS sequence for *Purshia tridentata*, used as the outgroup for this study, was obtained by amplifying the whole IGS as described above for *Cercocarpus*. We then sequenced from the 3' end of the amplified product until we identified the TIS.

For each *Cercocarpus* taxon, one of two strategies was conducted for sequencing the ETS. If the ETS amplification produced a single PCR product, the product was sequenced directly as outlined in the above section. If the ETS amplification produced multiple-length PCR products, we cloned the PCR products using the TOPO-TA cloning kit (Invitrogen). At least 10 colonies were screened for the ETS insert by direct PCR amplification using the same PCR parameters except a 10 min. hot start at 95 °C to lyse the cells. At least two of these amplifications were sequenced for each individual for each ETS length variant.

Comparison of the Phylogenetic Utility of the ITS and the ETS in *Cercocarpus*

The ETS sequences from the same five *Cercocarpus* individuals sampled to test the phylogenetic utility of the ITS were proofread and aligned using Clustal W in MegAlign (DNASTar 1998), followed by adjustment by eye. The five ETS sequences consisted of four ETS type 1 sequences and one ETS type 2 sequence (table 1, fig. 2). Only the 5' and 3' common regions of the ETS sequences (see Results) were included in the analyses (fig. 2). Maximum parsimony and maximum likelihood analyses were carried out in PAUP 4.0.0d64 as described for ITS.

Alignment and Phylogenetic Analysis of the ETS Types

All ETS sequences, regardless of type, were aligned using Clustal W followed by adjustment by eye. Only regions of the ETS sequences that were alignable between different ETS types were included in the analyses (fig. 2). For ETS type 3 sequences, only the 5' common region was included in the aligned sequence matrix. Only the 3' common region of the ETS sequence of *P. tridentata* was included to serve as the outgroup (fig. 2). Therefore, ETS type 3 sequences placement in the phylogeny is based only on its relationship to the

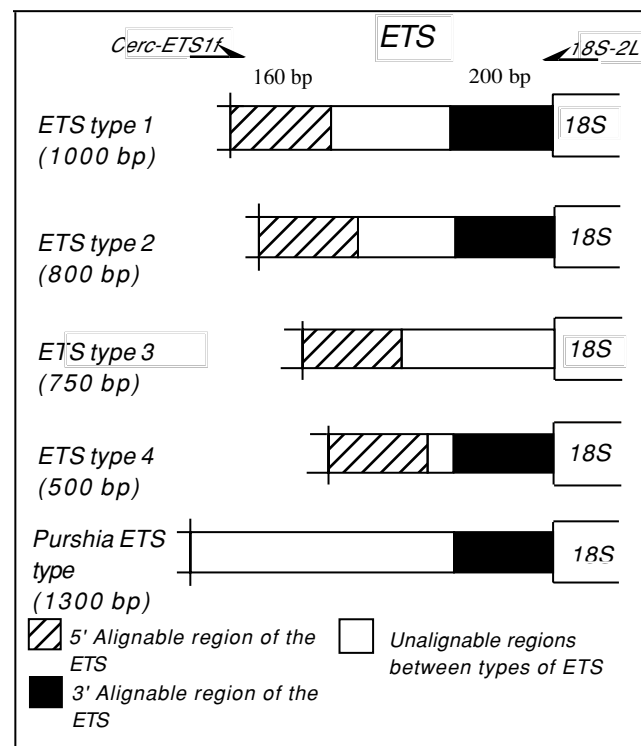


Figure 2—Representation of all four ETS types found in *Cercocarpus* to date and the ETS found in *Purshia tridentata*. Representations begin at the transcription initiation site (TIS) and go to the 18S gene. The hatched boxes indicate the 5' region that is alignable among ETS types. The black boxes show the location of the 3' region that is alignable among ETS types.

other ETS types and not on a direct comparison to the outgroup. The resulting sequence matrix totaled 300 aligned base pairs. Maximum parsimony and maximum likelihood analyses were carried out in PAUP 4.0.0d64 as described for ITS.

Results

Identification of Four ETS Types in *Cercocarpus*

Amplification and sequencing of the ETS region using the primers *Cerc-ETS1f* and *18S-2L* resulted in the identification of four distinct ETSs (table 1, fig. 2). The different ETS types ranged from 1,000 bp to 500 bp in *Cercocarpus* and was 1,300 bp in the outgroup *Purshia*. Three *Cercocarpus* individuals were found to contain multiple ETS types, 16 individuals had ETS type 1 only, four individuals had ETS type 2 only, one individual had ETS type 3 only, and one individual had ETS type 4 only (table 1).

Comparison of the ITS and the ETS Regions

The ITS tree was based on an alignment of 700 bp, which included the ITS1, ITS2, and 5.8S gene (fig. 1). ML and MP analyses produced the same tree topology. Only the MP results are presented here (fig. 3). The MP ITS tree had a length of 2 and very weak bootstrap support for internal branches due to lack of phylogenetically informative variation in the ITS region for the five individuals sampled.

The ETS tree for the same five individuals was based on 360 bp of aligned sequence data in the 5' and 3' common regions (fig. 3). ML and MP resulted in the same tree topology so only the MP results are presented (fig. 3). The ETS tree had a length of 20 steps and considerably higher bootstrap support for internal branches.

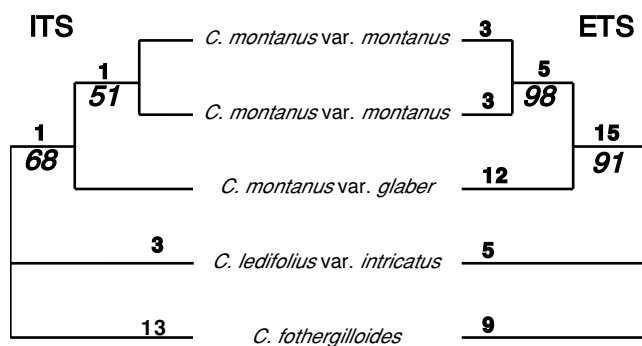


Figure 3—A comparison of MP phylogenies produced from 700 bp of aligned ITS sequences and 360 bp of aligned ETS sequence data. Numbers above branches are the numbers of steps, numbers below are bootstrap values.

Phylogenetic Analysis Using the ETS

The MP analysis of 5' and 3' common regions (360 bp) of 26 individuals of *Cercocarpus* and one individual of *Purshia tridentata* resulted in over 10,000 equally parsimonious trees of length 176. Figure 4 is a 50 percent majority rule consensus tree based upon those trees. Each ETS type formed its own clade, indicating the evolution of the types prior to the origin of the genus. Further, each of the multiple ETS types in *C. montanus* var. *minutiflorus* (pop. 57), *C. fothergilloides* var. *medranoanus* (pop. 82), and *C. macrophyllus* (pop. 26) belonged to their respective "type" clade (fig. 4). ETS type 1 sequences form a well supported clade, which is

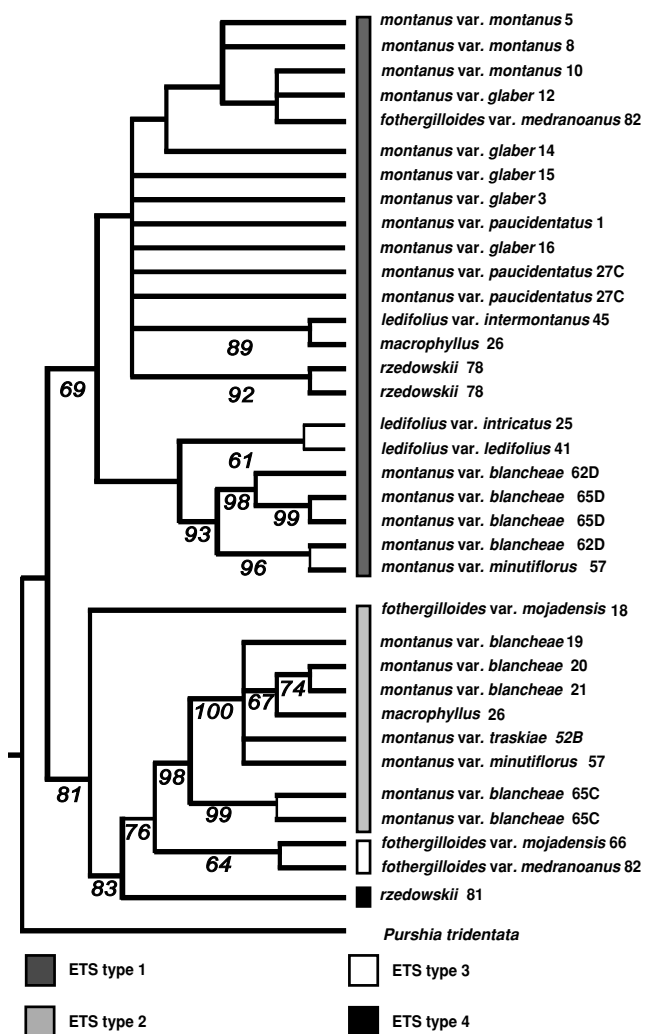


Figure 4—Fifty percent majority rule maximum parsimony consensus tree produced from the 360 bp alignment of the common regions between ETS types. The dark grey box indicates the monophyly of the ETS type 1 sequences, the light grey box indicates ETS sequences of type 2, the white box indicates the monophyly of ETS type 3 sequences, and the black box indicates the placement of the ETS type 4 sequence. Numbers below the branches indicate bootstrap support.

sister to a clade containing ETS types 2, 3, and 4. Because ETS type 3 sequences lack a 3' common region and the outgroup (*P. tridentata*) sequence lacks 5' common region, placement of ETS type 3 sequences in the phylogeny is based on comparison to the other *Cercocarpus* ETS sequences and not to direct comparison to the outgroup (fig. 2). ETS type 1 sequences form a well supported clade, which is sister to a clade containing ETS type 2, 3, and 4 sequences.

Discussion

Comparison of the ITS and the ETS Regions in *Cercocarpus*

The ITS region has been very attractive for phylogenetic reconstruction in plants because it evolves rapidly, it is subject to concerted evolution, and it is easily amplified using universal primers in the flanking genes (Baldwin and others 1995). However, in some studies of recently evolved plant lineages, the ITS region lacks enough phylogenetically informative variation to resolve relationships (Bain and Jansen 1995; Baldwin and others 1995; Francisco-Ortega and others 1997; Schilling and others 1998; Soltis and Kuzoff 1993). This lack of variation in the ITS in some plant groups (Fabaceae and Asteraceae) has led some researchers to look at the ETS for more variation (Bena 1998; Baldwin and Markos 1998; Linder and others 2000). Our results indicate that the ETS is also more phylogenetically informative than the ITS in *Cercocarpus* (Rosaceae). The length of the MP tree increased tenfold from 2 steps for the ITS tree to 20 steps for the ETS tree, even though the aligned ETS sequence data had only half the number of aligned base pairs as the ITS (fig. 3). Also, the bootstrap support for internal nodes increased dramatically in the ETS tree. Our results add to those of others (Bena and others 1998; Baldwin and Markos 1998; Linder and others 2000) that indicate that the ETS can be more phylogenetically useful for resolving the relationships of recently evolved taxa.

So little informative variation in the ITS for *Cercocarpus* adds weight to claims that *Cercocarpus* is a recently evolved group (Martin 1950), because there hasn't been enough time for variation to accumulate in the ITS. Recent origin of *Cercocarpus* would explain why there has been such disagreement in the taxonomy and circumscription of taxa (Schneider 1905; Rydberg 1913; Martin 1950).

Multiple ETS Types and Phylogenetic Analysis Using the ETS in *Cercocarpus*

Multiple ETS types have been found in other plant groups, including three genera in the Asteraceae: *Asteriscus*, *Calycadenia*, and *Helianthus* (Baldwin and Markos 1998; Linder and others 2000). However, the multiple ETS types found in these genera only differed in numbers of sub-repeats and did not show such large scale structural differences. The large structural differences found among the ETS types in *Cercocarpus* may be the result of past DNA mispairing events during replication due to secondary structure. We are currently cloning the ETS from genomic DNA obtained from different *Cercocarpus* individuals with known ETS types to further explore the molecular evolution of the ETS in *Cercocarpus*.

In an earlier study where multiple ETS types were found in individuals (Linder and others 2000), the types were all found to have originated within a species. There was no evidence for deep coalescence of the types. In *Cercocarpus*, we have found evidence for ETS types that have coalescence times earlier than the origination of the genus. ETS types 1, 3, and 4 are monophyletic groups, and ETS type 2 is a very closely related paraphyletic group. This is surprising because concerted evolution of the 18S-26S nrDNA repeat has commonly been invoked to reduce or eliminate the effects of deep coalescence on phylogenetic reconstruction (Moritz and Hillis 1996; Zimmer and others 1980). Within *Cercocarpus*, lack of concerted evolution or lineage sorting followed by hybridization has produced patterns of relationship that do not fit the usual pattern. We do not have enough data to tell which of these scenarios is more likely. In either case, the data suggest that the ETS types are paralogous with an ancient coalescence time. Different *Cercocarpus* ETS types, therefore, should not be used together for phylogenetic analysis. No problem should arise if a single ETS type is used for phylogenetic reconstruction. We believe that, with more sampling, a phylogeny using ETS type 1 will produce a usable nuclear phylogeny for *Cercocarpus*. ETS type 1 was found in all *Cercocarpus* taxa except *C. montanus* var. *traskiae* and *C. fothergilloides* var. *mojadensis* (table 1). Continued sampling of these two taxa may discover usable ETS type 1 sequences to include in the phylogenetic reconstruction.

References

- Bain, J. F.; Jansen, R. K. 1995. A phylogenetic analysis of the aureoid *Senecio* (Asteraceae) complex based on ITS sequence data. *Plant Systematics and Evolution* 195: 209–219.
- Baldwin, B. G.; Sanderson, M. J.; Porter, J. M.; Wojciechowski, M. F.; Campbell, C. S.; Donoghue, J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- Baldwin, B. G.; Markos, S. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution* 10: 449–463.
- Bena, G.; Jubier, M. F.; Olivieri, I.; Lejeune, B. 1998. Ribosomal external and internal transcribed spacers: combined use in the phylogenetic analysis of *Medicago* (Leguminosae). *Journal of Molecular Evolution* 46(3): 299–306.
- Blauer, A. C.; Plummer, A. P.; McArthur, E. D.; Stevens, R.; Giunta, B. C. 1975. Characteristics and hybridization of important Intermountain shrubs. I. Rose family; U.S. Department of Agriculture, Forest Service Research Paper INT-169. 36 p.
- Davis, J. N.; Brotherson, J. D. 1991. Ecological relationships of curlleaf mountain mahogany (*Cercocarpus ledifolius* Nutt.) communities in Utah and implications for management. *Great Basin Naturalist* 51(2): 153–166.
- Brayton, R.; Mooney, H. A. 1966. Population variability of *Cercocarpus* in the White Mountains of California as related to habitat. *Evolution* 20: 383–391.
- Doyle, J. J.; Doyle, J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.
- Francisco-Ortega, J.; Santos-Guerra, A.; Hines, A.; Jansen, R. K. 1997. Molecular evidence for a Mediterranean origin of the Macronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany* 84: 1595–1613.
- Linder, C. R.; Goertzen, L. R.; Vanden Heuvel, B.; Francisco-Ortega, J.; Jansen, R. K. 2000. The complete external transcribed spacer of the 18S-26S rDNA: amplification and phylogenetic utility at

- low taxonomic levels in Asteraceae and closely allied families. *Molecular Phylogenetics and Evolution* 14(2): 285–303.
- Lis, R. 1992. Leaf architecture survey of *Cercocarpus* (Rosaceae) and its systematic significance. *International Journal of Plant Science* 153(2): 258–272.
- Martin, F. L. 1950. A Revision of *Cercocarpus*. *Brittonia* 7(2): 91–111.
- Mortenson, T. H. 1973. Ecological variation in the leaf anatomy of selected species of *Cercocarpus*. *Aliso* 8: 19–48.
- Rydberg, P. A. 1908. Notes on Rosaceae. New York, NY: New York Botanical Gardens.
- Schilling, E. E.; Linder, C. R.; Noyes, R.; Rieseberg, L. H. 1998. Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. *Systematic Botany* 23: 177–187.
- Schneider, C. K. 1905. *Cercocarpus*. *Handbuch der Laubholzkunde*. 529–532.
- Soltis, D. E.; Soltis, P. S. 1998. In: *Molecular systematics of plants II: DNA sequencing*. Soltis, P. S.; Soltis, D. E.; Doyle, J. J. eds. Boston: Kluwer Academic publishers.
- Soltis, P. S.; Kuzoff, R. K. 1993. ITS sequence variation within and among populations of *Lomatium grayi* and *L. laevigatum* (Umbelliferae). *Molecular Phylogenetics and Evolution* 2: 166–170.