6 Secondary Structure of the *trn*L_{UAA} (Group I) Intron and the Molecular Delimitation of Ehretiaceae (Boraginales) [5]

Abstract.

To clarify which taxa belong to Ehretiaceae 29 *trn*L_{UAA} (group I) intron sequences were obtained from *Bourreria*, *Coldenia*, *Ehretia*, *Halgania*, *Lepidocordia*, *Pteleocarpa*, *Rochefortia*, and *Tiquilia*. The molecular analysis includes seven additional sequences from the remaining Boraginales (i.e., Boraginaceae *s.str.*, Cordiaceae, Heliotropiaceae, Hydrophyllaceae *s.str.*, and Lennoaceae). Ehretiaceae in its traditional circumscription (i.e., including *Coldenia* and *Pteleocarpa*) turns out to be polyphyletic: *Coldenia* is assigned to Cordiaceae, and *Pteleocarpa* appears to be a representative of the Gentianales. Lennoaceae (represented by *Pholisma arenarium*) are nested within Ehretiaceae, which constitute a monophyletic assemblage of Boraginales. *Cortesia* is the only taxon of Ehretiaceae which was not available for molecular analysis, however, based on fruit anatomy *Cortesia* is likely to be nested in *Ehretia*.

6.1 Introduction

Ehretiaceae are pantropical in distribution with centers of diversity in Central America, Africa, and East Asia and comprise about 150 species. The plants are subshrubs, shrubs, or trees, with the only exception of herbaceous *Coldenia procumbens*, and otherwise have the typical asterid characters such as tetracyclic, pentamerous flowers with five antesepalous stamens and bicarpellate gynoecia. The fruit is mostly a drupe containing one, two, or four endocarpids with four, two, or one seed(s), respectively (GOTTSCHLING & HILGER 2001).

Traditionally, Ehretiaceae are included in "Boraginaceae" *sensu* GÜRKE (1893), who provided the last comprehensive work on this taxon. "Boraginaceae" *sensu* GÜRKE (1893) are polyphyletic when Hydrophyllaceae and parasitic Lennoaceae are excluded based both on molecular (e.g., BÖHLE & HILGER 1997, FERGUSON 1999, GOTTSCHLING *et al.* 2001) and on morphological data (DIANE *et al.* 2002b). Ehretiaceae together with Cordiaceae, Heliotropiaceae, and Lennoaceae constitute the Primarily Woody Boraginales.

[5] Manuscript in preparation as **GOTTSCHLING M** & HH HILGER (in prep.): Secondary structure of the *trn*L_{UAA} group I intron and the molecular delimitation of Ehretiaceae (Boraginales).

Own contributions: Collecting plants (partly), calculating secondary structures, molecular calculations, writing manuscript, drawings.

The systematic delimitation of Ehretiaceae as well as the phylogenetic relationships within this group has been controversial. Ehretiaceae appear as a morphologically rather heterogeneous group, while the other taxa of the Primarily Woody Boraginales (Cordiaceae, Heliotropiaceae, and Lennoaceae) are monophyletic based on clear apomorphies (e.g., YATSKIEVYCH *et al.* 1986, DIANE *et al.* 2002a, GOTTSCHLING *et al.* in prep. a). Although exclusive characters are not recognised for Ehretiaceae, the naturalness of their core representatives has been generally accepted. The only common character of Ehretiaceae is the more or less bifid style that is a plesiomorphy in Boraginales. MILLER (1989) listed 11 taxa of Ehretiaceae: *Bourreria* (Figs. **1-11** and **1-12**), *Carmona*, *Coldenia*, *Cortesia*, *Ehretia*, *Halgania* (Fig. **1-10**), *Lepidocordia*, *Pteleocarpa*, *Rochefortia*, *Rotula*, and *Tiquilia* (Figs. **1-7** to **1-9**). Two additional taxa assigned to Ehretiaceae (GÜRKE 1893) are now placed in Cordiaceae (*Saccellium*) respectively Globulariaceae (*Poskea*).

MILLER (1989) recognized four probably monophyletic entities in Ehretiaceae. Old World *Coldenia* and New World *Tiquilia* are therefore clearly distinct from the remainder of Ehretiaceae since they developed numerous specialised characters in accord with their habitats. Central American and Northern South American *Lepidocordia* represents another isolated lineage and is said to be intermediate between Heliotropiaceae and Ehretiaceae (JOHNSTON 1950: 175). The Argentinean endemic *Cortesia* and the Australian endemic *Halgania* comprise a questionable third group of shrubby plants of xeric regions. The remaining taxa of Ehretiaceae, *Bourreria* (America and Africa), *Ehretia* (pantropical), and *Rochefortia* (Central America), seem to constitute a rather uniform major group. *Carmona* and *Rotula* have been included in *Ehretia* based both on molecular and on morphological data (GOTTSCHLING & HILGER 2001).

The major aim of this study is the molecular delimitation of Ehretiaceae. We therefore obtained *trn*L_{UAA} (group I) intron sequence data from all taxa of Ehretiaceae with the only exception of monotypic *Cortesia* due to insufficient leaf material. Evaluating the secondary structure of sequences has turned out to be useful both for understanding molecular evolution and for optimising alignments (e.g., AN *et al.* 1999, GOTTSCHLING *et al.* 2001, WOLF *et al.* 2002). The secondary structure of the *trn*L_{UAA} (group I) intron (Fig. **6-1**), with a stem-loop and nine pairing regions, has been found as conserved in cyanobacteria and chloroplasts of green algae and embryophytes (e.g., CECH 1988, KUHSEL *et al.* 1990, CECH *et al.* 1994). While in most non-embryophytes hairpin 8 is short (no more than 10 bp) in embryophytes it is long (e.g., 292 nt in *Nicotiana tabacum*, KUHSEL *et al.* 1990). The secondary structure of the *trn*L_{UAA} (group I) intron is investigated in this study to use additional characters in evaluating the molecular data.

6.2 Materials and Methods

29 species assigned to Ehretiaceae have been investigated (Table **1**, appendix). Furthermore, 6 sequences from Cordiaceae, Heliotropiaceae, Hydrophyllaceae, and Lennoaceae have been included in the alignment. *Borago officinalis* of Boraginaceae *s.str.* was used in the user specified outgroup comparison.

DNA was obtained from silica dried leaves from own field trips and from herbarium specimens of the herbaria B, BSB, F, MO, and NY. DNA extraction, PCR, purification, and sequencing followed standard protocols, which are described in detail in GOTTSCHLING & HILGER (2001). Primers used for amplification and sequencing of the *trn*L_{UAA} intron were those of TABERLET *et al.* (1991). The sequences were manually aligned using Se-Al v2.0a72 (RAMBAUT 2001). The complete data matrix is available on request in NEXUS format.

Common structural elements were initially recognised with the help of mFOLD (JAEGER *et al.* 1989, 1990) by screening for thermodynamically optimal and suboptimal secondary structures. Pairing regions were identified by mutual comparison of sequences in a manual alignment.

Phylogenetic calculation were run on a IBM compatible computer in PAUP* 4.0b1 (SWOFFORD 1998). Minimum evolution (ME) trees were generated using heuristic searches, TBR swapping, and a starting tree obtained by neighbour-joining. A bootstrap analysis (criterion = distance, BS) was estimated based on 1000 replicates. A PUZZLE analysis (criterion = likelihood, P) was performed with 1000 quartet puzzling steps. Likelihood settings from the best-fit model were determined using the AIC-criterion in Modeltest 3.06 (POSADA & CRANDALL 1998). Estimated parameters were used in BS, ME, and P. Gaps were treated as missing in all analyses.

For the sequence of *Pteleocarpa* a NCBI Blast Search (ALTSCHUL *et al.* 1997) has been performed.

6.3 Results

Secondary structure. The secondary structure of the *trn*L_{UAA} (group I) intron is conserved in all species under investigation constituting of a stem-loop with eight pairing regions and four conserved motifs (Fig. **6-1**). Especially pairing region 8 is long (up to approximately 300 nt) in most species under investigation and shows the highest variation between species in the primary nucleotide sequence. In, e.g., *Cordia sebestena* (Fig. **6-2**) pairing region 8 is a linear molecule, and three conserved pairing subregions can be stated (8a to 8c in Figure **6-2**). While 8a and 8b are present in all species under investigations, 8c is lacking in *Pholisma*. In all species 8c is A/T-rich, but is replaced by an exclusively A-rich, non-painring region in *Pholisma*. The number of AT-base pairs is highly variable even between closely allied species and does not follow any phylogenetic pattern.

Molecular delimitation of Ehretiaceae. Figure **6-3** shows the best likelihood tree (–ln = 1606.8000) calculated with the best fit model (GTR + G model; number of substition types: 6; number of distinct data patterns under this model: 169) with bootstrap and PUZZLE support values. Ehretiaceae in the traditional circumscription are a polyphyletic assemblage. *Bourreria*, *Ehretia*, *Halgania*, *Lepidocordia*, *Rochefortia*, and *Tiquilia* together with parasitic Lennoaceae (*Pholisma*) constitute a monophyletic group based both on the best likelihood tree and on the PUZZLE analysis (73 P). *Coldenia* belongs to Cordiaceae as does *Saccellium*. The sequence of

Pteleocarpa is divergent from those of all other Boraginales. Table **5** (appendix) shows the five most similar sequences deposited in GeneBank indicating a relationship among Gentianales.

The relationships within Ehretiaceae are not fully resolved. *Bourreria* (66 BS, 66 P), *Ehretia* (76 BS, 84 P), *Halgania* (61 BS, 75 P), and *Tiquilia* (96 BS, 87 P) are monophyla. *Bourreria*, *Lepidocorpia*, and *Rochefortia* together constitute a monophyletic group (54 P), to which Lennoaceae are the sistergroup. *Lepidocordia* and *Rochefortia* appear to be closely allied (72 BS, 87 P).

6.4 Discussion

Information from secondary structure. It could be shown that investigation of the secondary structure is a useful tool for recognizing homologous base pairings (e.g., GOTTSCHLING *et al.* 2001, WOLF *et al.* 2002). The general secondary structure of the *trn*L_{UAA} (group I) intron proposed by CECH (1988), KUHSEL *et al.* (1990), and CECH *et al.* (1994) is conserved in all species investigated (Fig. **6-1**). Pairing region 8 in Boraginales (and many other asterids) is long compared to Cyanobacteria and green algae (e.g., KUHSEL *et al.* 1990) and is variable both in its primary nucleotide sequence and in its secondary structure. While mFOLD (JAEGER *et al.* 1989, 1990) finds more or less linear structures for, e.g., *Cordia sebestena* (Fig. **6-2**), *Ehretia tinifolia*, and *Heliotropium europaeum*, it suggests a circular stem-loop with several hairpins in *Pholisma arenarium* (not shown). Homologous base pairings are only found in the regions 8a, 8b, and 8c.

Pairing region 8 is long in most species (up to approximately 300 nt), but shorter in *Pholisma*. The reduction of the absolute numbers of nucleotides can be interpreted as an abridgement of pairing regions as has been demonstrated for Heliotropiaceae in the ITS1 transcript (GOTTSCHLING *et al.* 2001, DIANE *et al.* 2002a). Such deletions occur occasionally in the *trn*L_{UAA} intron convergently in different lineages (e.g., in *Hydrolea*, *Wellstedia*, not shown). Pairing subregions 8a and 8b are completely conserved in all species under investigation (Fig. **6-2**). The AT-rich region found in 8c varies highly in length (6 to 13 bp) even between closely allied species and is therefore a source of homoplasy. The AT-rich region should be used with caution in phylogenetic reconstructions.

Molecular delimitation of Ehretiaceae (Figure **6-3**). The polyphyly of Ehretiaceae in their traditional circumscription (e.g., GÜRKE 1893, MILLER 1989) does not come as a surprise since they are a morphologically heterogeneous group with no known apomorphy. However, a monophyletic taxon Ehretiaceae is here retrieved comprising *Bourreria*, *Ehretia*, *Halgania*, Lennoaceae (represented by *Pholisma*), *Lepidocordia*, *Rochefortia*, and *Tiquilia* (73 P). Each of these subordinate taxa are monophyletic based on clear apomorphies and/or molecular data. The relationships within Ehretiaceae are not fully resolved. The close relationship of *Lepidocordia* and *Rochefortia* (72 BS, 87 P) had not been suggested so far, but the dioecy found exclusively in

the taxa *Lepidocordia* and *Rochefortia* of the Ehretiaceae can now be interpreted as apomorphic. Together with *Bourreria* they possibly constitute the largely Caribbean clade of Ehretiaceae (54 P).

For over a century *Coldenia sensu* GRAY (1862) included both New World and Old World species. RICHARDSON (1977) reinstated New World *Tiquilia* as distinct from Old World *Coldenia* primarily based on data from morphological, chromosomal, and ecological studies. The molecular analysis supports this view: While *Tiquilia* is identified as belonging to Ehretiaceae (96 BS, 87 P), *Coldenia* is assigned to Cordiaceae. Since monotypic *Coldenia* does not share apomorphic characters with Cordiaceae such as an undivided endocarp and four stigmatic lobes (GOTTSCHLING *et al.* in prep. a), a sistergroup relationship with the latter group (comprising nearly 300 species) is likely.

The systematic placement of Lennoaceae (*Lennoa* and *Pholisma*, Fig. **1-6**) has been enigmatical. YATSKIEVYCH *et al.* (1986) listed taxa from the Ericales, Boraginales, and Polemoniales as candidates for their closest relatives. Additionally to the results from this study we now have support from seven molecular markers that Lennoaceae are the sistergroup of, or are nested in Ehretiaceae: rps2 (SMITH & DEPAMPHILIS 1998, SMITH *et al.* 2000) and *rbc*L and *ndh*F (OLMSTEAD & FERGUSON 2001) from chloroplasts, cox1 and *atp*A (SMITH *et al.* 2000) from mitochondria, and nuclear ITS1 (SMITH *et al.* 2000, GOTTSCHLING *et al.* 2001). *Lennoa* and *Pholisma* should be included in Ehretiaceae to avoid paraphyly, although this increases the morphological heterogeneity of Ehretiaceae.

The sequence of *Pteleocarpa* is divergent from those of all other Boraginales. Table **5** (appendix) shows the five most similar sequences deposited in GeneBank found by a NCBI Blast Search indicating relationships among Gentianales. The systematic position of *Pteleocarpa* outside the Ehretiaceae has been suggested by many authors (e.g., VELDKAMP 1988, HEUBL *et al.* 1990) and has been also demonstrated in preliminary molecular studies from *rbc*L and *ndh*F (OLMSTEAD & FERGUSON 2001).

Unfortunately, it was impossible to obtain a sequence of Argentinean *Cortesia* that could clarify its systematic position. However, it has turned out that fruit anatomy yields important characters in Ehretiaceae (e.g., MILLER 1989, GOTTSCHLING & HILGER 2001). A four-parted endocarp (e.g., in *Bourreria*, *Ehretia p.p.*, *Rochefortia*, *Tiquilia*) is interpreted as the ancestral character state in Ehretiaceae, while a two-parted endocarp is the result of fusion of two endocarpids ('syn-mericarpy', derived state). Syn-mericarpous endocarps in Ehretiaceae are exclusively found in the *Ehretia* II subclade (GOTTSCHLING & HILGER 2001) and in *Cortesia*, which indicates that *Cortesia* is probably nested in *Ehretia.*

The molecular delimitation of Ehretiaceae provided by this study can now be used to investigate the morphological phylogenies of *Bourreria* (Figs. **1-11**, **1-12**), *Cortesia*, *Ehretia*, *Halgania* (Fig.

1-10), *Lennoa*, *Lepidocordia*, *Pholisma* (Fig. **1-6**), *Rochefortia*, and *Tiquilia* (Figs. **1-7** to **1-9**) as well as the precise relationships of Ehretiaceae within Boraginales.