Molecular systematics and phytochemistry of

Rehmannia (Scrophulariaceae)

Dirk C. Albach, a Hong-Qing Li, b Nan Zhao, b Søren Rosendal Jensen c

a Institut für Spezielle Botanik, Johannes Gutenberg-Universität Mainz, D-55099 Mainz, Germany.
b School of Life Science, East China Normal University, Shanghai, China.
c Department of Chemistry, The Technical University of Denmark, DK-2800 Lyngby, Denmark.

* Corresponding author. Tel.: +45-45252103; fax: +45-45933968.

E-mail address: srj@kemi.dtu.dk (S.R. Jensen).
Abstract – The relationships between the six known species of *Rehmannia* were investigated. With regard to the content of iridoid glucosides, caffeoyl phenylethanoid glycosides (CPGs) and ionone glucosides, no conclusions could be drawn. Phylogenetic analysis of DNA sequence data (ITS region, *trnL-F* region and *rps16* intron) reveal a well-resolved topology in which *R. glutinosa* and *R. solanifolia* and *R. piasezkii* and *R. elata* are well-supported species pairs. *Rehmannia chingii* is sister to the rest of the genus, which is congruent with its distribution distant to the other species of the genus.

*Keywords: Rehmannia; Scrophulariaceae; Chemotaxonomy; Iridoid glucosides; Caffeoyl phenylethanoid glycosides (CPG); DNA sequence data; Phylogeny*
1. Introduction

*Rehmannia* Libosch. ex Fisch. et Mey. is a small genus of six species (Chin 1998). Five of the six species are endemic to China with *R. glutinosa* (Gaert.) Libosch. ex Fisch. et Mey. (Rg) extending further to Korea and Japan, although it is probably introduced in the latter (Rix 1987). One of the species, *R. glutinosa*, is an important species in traditional Chinese medicine (*Rehmanniae* Radix) used to treat fever and bleeding (Zhu 1998). It is nowadays used in treatments of various hormonal disorders. This importance led to a tremendous amount of chemical work performed on the rhizomes, both fresh and cured. Fresh rhizomes were initially shown to contain sucrose, mannitol and catalpol (Kitagawa et al., 1971). Intensive work was continued, mainly in the 1980's with the work of Oshio and Inouye (1982) as the most outstanding. In this, catalpol was confirmed to be the major iridoid constituent of the rhizomes together with smaller amounts of ajugol, aucubin and five other iridoid di- or triglycosides. More recently, Nishimura et al. (1989) isolated from a large amount of root material a number of the less polar minor iridoid constituents. A large number of other iridoids have also been isolated from cured roots; however, most of these can be considered to artefacts due to degradation of the natural constituents during curing. Iridoids appear to stimulate the production of adrenal cortical hormones with anti-inflammatory effect and sex hormones (Chang and But 1986), thus explaining the effect seen in traditional Chinese medicine. Also a number of caffeoyl phenylethanoid glucosides (CPGs) have been isolated of *R. glutinosa*. Of these verbascoside was by far the major constituent of the roots (Sasaki et al. 1989). In addition to these compounds a number of ionone glucosides have been isolated (Yoshikawa et al., 1986).
Despite the interest in *Rehmannia glutinosa* as a traditional Chinese medicinal plant, little is known about the close relatives. Neither have they been studied phytochemically nor did anyone investigate the relationship of the species with its congeners. Chromosome numbers have only recently become available for all species. *R. glutinosa* (Gao and Zhang, 1984) and *R. solanifolia* (Li, unpubl.) are tetraploid (2n = 4x = 56), all other species are diploid (2n = 2x = 28; Li, unpubl.). There has been some debate about the circumscription of the genus. Two species previously ascribed to the genus, namely *R. rupestris* and *R. oldhamii* were removed to their own genera, *Triaenophora* and *Titanotrichum*, respectively, by Solereder (1909). He noticed the uniloculicidal ovary and type of indumentum of *Titanotrichum* that suggests a relationship with Gesneriaceae, a relationship later supported by DNA-sequence based analyses (Smith et al., 1997; Albach et al., 2001). Regarding *Triaenophora*, Li (1948) refuted the distinctness from *Rehmannia*, although Chin (1998) regarded them as separate following Solereder (1909) based on the lack of secretory cells in leaves and flowers and the bilocular ovary. A further systematic issue concerns the closest relative of *Rehmannia*. Although Solereder (1909) suggested a position in Gesneriaceae based on the loculicidal ovary, *Rehmannia* is usually placed in Scrophulariaceae. The first species was described within *Digitalis* (Gaertner, 1770) and subsequent authors followed in placing *Rehmannia* in Digitalideae (Bentham 1846, von Wettstein 1895, Chin 1998). Phylogenetic analyses of plastid DNA sequences (Oxelman et al. 2005) contradicted such a position and placed *Rehmannia* instead in an uncertain position close to Scrophulariaceae *s. str.* and Orobanchaceae.

In the present study, we were interested in the infrageneric questions regarding phylogenetic relationships among species of *Rehmannia* and phytochemical differences between the species. Phytochemically, we concentrated on sugars and iridoids, which are the main active constituents of *Rehmannia* (Zhu 1998). The molecular phylogenetic part of the study is based
on three DNA regions, the nuclear ribosomal ITS region (3’ end of 18S rDNA, ITS1, 5.8S rDNA, ITS2, 5’ end of 26S rDNA), the trnL-F region (trnL intron, trnL 3’ exon, trnL-F spacer) and the rps16 intron. All regions have been used before in phylogenetic studies of Lamiales and Scrophulariaceae (e.g., Albach et al. 2005; Oxelman et al. 2005).

2. Materials and methods

2.1 Plant material

Material of all species of *Rehmannia* was collected from plants cultivated in the nursery of East China Normal University, Shanghai, China. Voucher information is given in Table 1.

2.2 Phytochemical analysis

$^1$H and $^{13}$C NMR spectra were recorded on a Varian Mercury-300 MHz in D$_2$O or CD$_3$OD using the solvent peak as the internal standard. The isolated compounds were identified by their $^1$H and $^{13}$C NMR spectra by comparison with spectra of known standards (Taskova et al., 2006). Dry leaf material (7 to 11 g) was extracted by blending in EtOH (50 ml). The mixture was brought to the boiling point and then left to stand at room temperature for 3-7 days. After filtering, each extract was evaporated and partitioned in H$_2$O-Et$_2$O (25 ml each); the aqueous layer was concentrated to give the crude extract. The aqueous extracts were processed by preparative chromatography using a Merck Lobar RP-18 column size B. The initial eluent was H$_2$O followed by H$_2$O:MeOH mixtures (15:1 to 4:5) and finally by MeOH. The isolated compounds were identified by means of $^1$H NMR spectroscopy. The content of sugars in the polar fraction was estimated by the intensity of the signals in $^{13}$C NMR spectrum.

2.2.1 *R. solanifolia*
Dry leaves (7.3 g) gave crude extract (1010 mg). Chromatography gave sugars (250 mg – mainly glucose), catalpol (450 mg), impure ajugol (70 mg), caffeoyl phenylethyl glycoside (CPG) mixture (20 mg), verbascoside (140 mg), CPG-mix (30 mg).

2.2.2 R. glutinosa
Dry leaves (9.0 g) gave crude extract (1060 mg). Chromatography gave sugars (250 mg – mainly glucose and ca 30% mannitol), catalpol (70 mg), iridoid mixt. (25 mg), ajugol (40 mg), impure 8-epiloganic acid (20 mg), iridoid mixt. with trace of rehmanioside A (30 mg), verbascoside (130 mg), CPG-mix (70 mg).

2.2.3 R. piasezkii
Dry leaves (7.0 g) gave crude extract (420 mg). Chromatography gave sugars (120 mg – mainly glucose), catalpol (70 mg), impure rehmanionoside A (40 mg), verbascoside (40 mg), CPG-mix (30 mg).

2.2.4 R. elata
Dry leaves 8.4 g) gave crude extract (810 mg). Chromatography gave sugars (160 mg – mainly glucose), catalpol (160 mg), impure ajugol (10 mg), CPG-mix (10 mg), verbascoside (120 mg), CPG-mix (25 mg).

2.2.5 R. chingii
Dry leaves (11.0 g) gave crude extract (980 mg). Chromatography gave sugars (340 mg – mainly glucose), catalpol (12 mg), descaffeoyl verbascoside (15 mg), CPG-mix (30 mg), verbascoside (150 mg), CPG-mix (30 mg).

2.2.6 R. henryi
Dry leaves (7.0 g) gave crude extract (400 mg). Chromatography gave sugars (315 mg – mainly glucose, sucrose), catalpol (10 mg), iridoid mixture (20 mg), CPG-mix (10 mg), verbascoside (5 mg).

2.3 DNA sequence analysis

Total genomic DNA was extracted from dry leaf samples according to the hexadecyltrimethylammonium bromide (CTAB) procedure of Doyle and Doyle (1987) and then washed twice with 70% ethanol. DNA pellets were dried and resuspended in TE buffer. The trnL-F region was amplified with primers c and f of Taberlet et al. (1991). ITS sequences were amplified and sequenced using primers 17SE (Sun et al., 1994) and ITS4 (White et al., 1991). Sequences of the rps16 intron were amplified and sequenced using primers rpsF and rpsR2 (Oxelman et al., 1997). PCR products were run on a 1.0% TBE-agarose gel, cut from the gel, and cleaned using a QIAquick PCR purification and gel extraction kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocols. Sequencing reactions (10 µl) were carried out using 1–2 µl of the taq DyeDeoxy Terminator Cycle Sequencing mix (Applied Biosystems), and unincorporated dye terminators were removed either by cleaning with 4% sodium acetate in ethanol followed by two rounds of cleaning in 70% ethanol. Both strands were sequenced. Sequences were assembled and edited using Sequencher versions 3.0 (Gene codes corporation, Ann Arbor, USA). Assembled sequences were manually aligned prior to analysis.

Matrices were analysed with PAUP* 4.0b10 (Swofford, 1998) using the maximum likelihood criterion. Maximum likelihood analyses were conducted using a model estimated with Modeltest 3.06 (Posada and Crandall, 1998) based on the AIC. Ten runs of random taxon addition (10 replicates) starting from random trees using tree bisection reconnection (TBR) were conducted with MulTrees (keeping multiple shortest trees) in effect and no tree limit.
Outgroups were chosen based on the analysis by Oxelman et al. (2005). Bootstrap percentages were assessed using 1000 replicates and TBR-branch swapping.

Incongruence was tested using the Shimodaira–Hasegawa tests (SH) as implemented in PAUP4.0b (Swofford, 1998) with a model for each of the two data sets estimated using Modeltest 3.06 (Posada and Crandall, 1998; see above) and chosen based on the AIC, full optimisation, and 1000 bootstrap (BS) replicates to compare different trees. Optimal trees (unconstrained) from the maximum likelihood analysis of separate data sets were compared with that of the second analysis as a constraint tree.

3. Results

3.1 Phytochemical analysis
We have investigated leaf material of six species of *Rehmannia*. Glucose was the main carbohydrate present in all species although 30% mannitol was found in *R. glutinosa*. All species contained catalpol as the main iridoid glucoside and also considerable amounts of verbascoside, the latter being the most abundant compound in *R. glutinosa* and *R. chingii*. The iridoid glucoside ajugol was isolated from 3 species, *R. solanifolia*, *R. glutinosa* and *R. elata*, while the ionone glucoside remaionoside A was found in *R. glutinosa* and *R. piasezkii*. Thus, the results of the chemical investigation showed a close relationship among the species of *Rehmannia*, although infrageneric relationships were not resolved.

3.2 DNA sequence analysis
The combined dataset comprises 2506 aligned characters (ITS: 671 characters; *rps*16 intron: 878 characters, *trnL-F* region: 957 characters). Analysis of the combined dataset using the GTR+Γ+I model resulted in a well resolved phylogeny with *Lindenbergia* as sister to *Rehmannia* with high bootstrap support (98%), high support for the monophyly of *Rehmannia*.
(100%). It further shows moderate support for the sister group-relationship of \textit{R. chingii} to the rest of the genus (82% BS) and \textit{R. henryi} as sister to all species except \textit{R. chingii} (83% BS). \textit{Rehmannia solanifolia} and \textit{R. glutinosa} (99% BS) and \textit{R. piasezkii} and \textit{R. elata} (100% BS) are highly supported as sister species. For the analysis of the ITS region alone Modeltest found the GTR+\Gamma model (\Gamma = 0.27) as the optimal model. The optimal tree found three pairs of species in \textit{Rehmannia} with weak to moderate support for almost all nodes. \textit{R. solanifolia}+\textit{R. glutinosa} are sister to \textit{R. piasezkii}+\textit{R. elata} (100% bootstrap support (BS)) and \textit{R. chingii}+\textit{R. henryi}. The result of the two cpDNA regions did not differ (results not shown) and therefore only results from the combined cpDNA analysis are shown. The analysis of the cpDNA regions was conducted using the TVM+I model (I = 0.41). \textit{Rehmannia chingii} is strongly supported (98% BS) as sister to the rest of the genus, whereas \textit{R. henryi} is found in a trichotomy with the two other species pairs found also in the analysis of the ITS region. Differences between the analyses of the ITS region and the cpDNA regions concerned primarily the position of \textit{R. chingii}. Using the Shimodaira-Hasegawa test, the topology found in the analysis of the ITS region was rejected by the cpDNA data set as significantly worse (p = 0.023), whereas the reverse was not the case (p = 0.14). Not surprisingly, the combined analysis (TVM+I model; I = 0.63) also retrieved the same topology as the cpDNA analysis.

4. Discussion

Iridoid glucosides and caffeoyl phenylethanoid glycosides (CPGs) like verbascoside are characteristic for most taxa in Lamiales (Jensen, 1992). However, certain traits are characteristic for each of the families in the order. The chemical profile found for \textit{Rehmannia}, namely: (i) content of mannitol (in at least one species), (ii) catalpol with very small amounts of aucubin present, (iii) ajugol and 6-esters of this combined with (iv) lack of 6-\textit{O-} or 10-\textit{O-}esters of aucubin and catalpol and (v) lack of iridoid aldehydes, can be used for comparison with the chemical traits found in neighbouring families in Lamiales (Table 2). Presence of
mannitol and absence of iridoid aldehydes combined with an apparent lack of aucubin would place *Rehmannia* in an intermediate position among the main families in the order, consistent with the recent results by Oxelman et al. (2005) rather than with the traditional position close to *Digitalis* in Plantaginaceae. Also, the chemical profile of Digitalideae is rather different from that of *Rehmannia* (Taskova et al., 2005).

Phylogenetic analyses of three DNA regions for all six species recognized in *Rehmannia* reveals a moderate to well supported result that is in good agreement with morphological characters. Using four different outgroups chosen among a variety of genera from Scrophulariaceae (Oxelman et al. 2005), monophyly of *Rehmannia* is strongly supported. Future inclusion of *Triaenophora* will be necessary to substantiate this claim.

Within *Rehmannia*, *R. piasezkii* and *R. elata* and *R. glutinosa* and *R. solanifolia* form two highly supported pairs of sister species, which are in good agreement with morphological characters (see below). *Rehmannia chingii* and *R. henryi* either form a clade (ITS; Fig. 2) or are consecutive sister species to the rest of the genus (cpDNA; Fig. 1).

Also, *R. chingii* and *R. henryi* are morphologically a pair of species although obvious differences exist. They live in different mountainous regions, which are separated by the lower Yangtze Valley with the former distributed distant from the distribution (Qinling-Wudangshan-Taihangshan) of the remaining species of the genus, and have different flower colours: purple-red for *R. chingii* (although we have in one instance noticed white flowered plants), and white to yellowish for *R. henryi*. These findings support the idea that they are consecutive sister species, namely, *R. chingii* is a sister to the rest of the species.
We propose a phytochemical progression in the genus, because neither of these two species contained ajugol or rehmannioside, which may only have evolved in the common ancestor of the other four species, in accordance with the monophyly of these four species.

Especially, *R. piasezkii* and *R. elata* are so similar that they are difficult to differentiate morphologically. *R. piasezkii* can reach over 100 cm in height; leaves pinnately partite, or its lobes obtuse with dentate margin, or just several acute lobes with entire margin; Bracteoles present or not; Corolla red to purple-red, the spots located at the throat change largely in density and colour. Sometimes distinct characters appear in the same plant synchronously. *Rehmannia elata* varies in leaf shape, seed morphology, fruit size and pollen wall texture also in some extent but falls into the range of that of *R. piasezkii*. Their seeds are the smallest ones in the genus (about 0.06g/1000 seeds, compared to more than 0.11g/1000 seeds in other species). Germination ratio of crossbred seeds of *R. piasezkii* × *R. elata* is similar to that of the parent plants. So we get the impression that *R. piasezkii* and *R. elata* are conspecific, and should be combined, although they differ somewhat phytochemically.

*Rehmannia solanifolia* is similar to *R. glutinosa* in morphology, though the former bears ovate leaves instead of mainly long-elliptic ones, axillary flowers instead of usually raceme, and 5 to 6-lobed corolla instead of 5-lobed corolla. Its distribution area is very narrow (several spots at the outer part of the generic range) and is overlapped by that of *R. glutinosa*. Both of these species are tetraploid (2n=56) when compared to the other four species. *R. glutinosa* and *R. solanifolia* differ from all other species morphologically. Together with the DNA data, this suggests that *R. solanifolia* is derived from *R. glutinosa*, and the latter may be derived from *R. henryi*; the latter is the only milk white to yellow-flowered species of the genus, and may have contributed this trait to the inner yellow corolla tube in *R. glutinosa*. Our data currently do not allow an inference on whether *R. glutinosa* is an auto- or allopolyploid, although the high
variability of *R. glutinosa* suggests the latter. Detailed analyses (e.g., GISH) is necessary for further understanding of the evolutionary course of the genus. Importantly, the two species seem to be the most iridoid rich with 6% catalpol in dried root material found for *R. solanifolia* and 3-11% reported for *R. glutinosa* (Luo et al. 1994). The present work shows a significantly lower catalpol content in leaves of *R. glutinosa* and suggests that iridoid content may be variable in different varieties and different organs of this species.
References


Li, H L., 1948. A revision of the genus Rehmannia. Taiwania. 1, 71-82, Plates I-V.


Table 1. List of species used in the analyses and their vouchers with locations and their GenBank numbers

<table>
<thead>
<tr>
<th>Species</th>
<th>Voucher</th>
<th>ITS</th>
<th>trnL-F</th>
<th>rps16 intron</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lindenber gia philippensis</em></td>
<td>n.a.</td>
<td>AY911231</td>
<td>AJ608586</td>
<td>AJ609169</td>
</tr>
<tr>
<td><em>Mazus spec.</em></td>
<td>n.a.</td>
<td><em>M. reptans</em> AJ478940</td>
<td><em>M. reptans</em> AJ479004</td>
<td><em>M. stachydifolius</em> AJ609167</td>
</tr>
<tr>
<td><em>Oreosolen spec.</em></td>
<td>n.a.</td>
<td><em>O. wattii</em> AF509817</td>
<td><em>O. wattii</em> AF513357</td>
<td><em>O. unguiculatus</em> AJ609155</td>
</tr>
<tr>
<td><em>Paulownia tomentosa</em></td>
<td>n.a.</td>
<td>AF478941</td>
<td>AJ423122</td>
<td>AJ431051</td>
</tr>
<tr>
<td><em>Rehmannia chingii</em> H. L. Li</td>
<td>Li H.Q. 20040601 (HSNU), Zhejiang, China</td>
<td>DQ069313</td>
<td>DQ856494</td>
<td>DQ856488</td>
</tr>
<tr>
<td><em>Rehmannia elata</em> N. E. Br.</td>
<td>Li H.Q. 20040501 (HSNU), Shanghai, China</td>
<td>DQ069315</td>
<td>DQ856496</td>
<td>DQ856490</td>
</tr>
<tr>
<td><em>Rehmannia glutinosa</em> (Gaert.) Libosch. ex Fisch. et Mey.</td>
<td>Li H.Q. 20040607 (HSNU), Henan, China</td>
<td>DQ069312</td>
<td>DQ856493</td>
<td>DQ856487</td>
</tr>
<tr>
<td><em>Rehmannia henryi</em> N. E. Br.</td>
<td>Li H.Q. 20040605 (HSNU), Hubei, China</td>
<td>DQ272447</td>
<td>DQ856497</td>
<td>DQ856491</td>
</tr>
<tr>
<td><em>Rehmannia piasezkii</em> Maxim.</td>
<td>Li H.Q. 20040603 (HSNU), Hubei, China</td>
<td>DQ069316</td>
<td>DQ856495</td>
<td>DQ856489</td>
</tr>
<tr>
<td><em>Rehmannia solanifolia</em> Tsoong et Chin</td>
<td>Li H.Q. 20040608 (HSNU), Sichuan, China</td>
<td>DQ069314</td>
<td>DQ856492</td>
<td>DQ856486</td>
</tr>
</tbody>
</table>
Table 2. Comparison of chemical characters of *Rehmannia* with those reported from selected families within Lamiales.

<table>
<thead>
<tr>
<th>Compounds Taxon</th>
<th>Reserve carbohydrate&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Aucubin Catalpol</th>
<th>Ajugol</th>
<th>6- or 10-&lt;i&gt;O&lt;/i&gt;-Esters of Auc/Cat</th>
<th>5-OH Iridoids</th>
<th>Iridoid Aldehydes</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Rehmannia</em></td>
<td>Sucr. Glu. (Mannitol)</td>
<td>catalpol</td>
<td>present</td>
<td>not present</td>
<td>present</td>
<td>none</td>
</tr>
<tr>
<td>Scrophulariaceae</td>
<td>Sucr. Glu.</td>
<td>both</td>
<td>common</td>
<td>rare</td>
<td>common</td>
<td>none</td>
</tr>
<tr>
<td>Plantaginaceae</td>
<td>Mannitol, Sorbitol</td>
<td>both</td>
<td>rare</td>
<td>common</td>
<td>rare</td>
<td>rare</td>
</tr>
<tr>
<td>Orobancheae</td>
<td>Mannitol</td>
<td>both</td>
<td>rare</td>
<td>common</td>
<td>common</td>
<td>common</td>
</tr>
<tr>
<td>Lamiaceae&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Sucr. Glu.</td>
<td>catalpol</td>
<td>common</td>
<td>rare</td>
<td>common</td>
<td>none</td>
</tr>
</tbody>
</table>

<sup>a</sup>Sucr. (sucrose); Glu. (glucose). <sup>b</sup>Excluding *Nepeta*, which contains iridoids formed by an alternative pathway.
Fig. 1. Optimal tree from the maximum likelihood analysis of the combined data set, which is identical to the one from the combined cpDNA analysis. Numbers above the branches indicate number of inferred character transformations, those below the branches bootstrap percentages (combined analysis/cpDNA analysis).
Fig. 2. Optimal tree from the maximum likelihood analysis of the ITS data set. Numbers above the branches indicate number of inferred character transformations, those below the branches bootstrap percentages.