Family tree of the Rhododendron family (Google translate DK->UK**)**

By Hans Eiberg (Rhododendron I Danmark I 25 år, 1999, 198-199.)

Research into the construction of the family tree of the rhododendron family has now entered a new era, as the methods of DNA technology are introduced. With the help of the new techniques, possibilities are opened for the species of the rhododendron genus to be safely put in place in a system, which is determined based on evolution.

Collections

If we go approx. 80-100 years ago. The plant hunters Veitch (1900-04) collected many new species in the Himalayas. Forrest (from 1910 to 1931), Kingdon-Ward (from 1913-1937), Rock (1923-1932) etc. (see the article by Preben Escherich Holkjær). This brought home plants and huge quantities of seeds, which were imported to England and sown. In addition, herbarium materials were pressed for later studies. These expeditions have been incredibly valuable to future researchers, gardeners, plant collectors and garden owners.

Determination keys.

Researchers have since attempted to determine the species of the material by constructing identification keys to group the species into various families, genera, series, and subseries. This work has been extremely difficult, especially because the collected material contained plants that varied only slightly from one another, but with a few characteristic differences.

In addition, there were plants, which later turned out to be hybrids. Another problem has been that some of the plants that were collected back then could not be found in the wild. Newer taxonomists have cleaned up the previously described species and possibly classified them as hybrids whose evidence of their real existence or parentage could be doubted.

Characteristics

To characterize a species, the ideal would be to find a single 100% certain criterion, but due to lack of this, the botanists set up approx. 8-10 less secure criteria. If the plant then has approx. 80% of these characteristics, the provision will probably be in order. However, some "species" are only characterized by a few uncertain characteristics such as "distribution of hairs" or crown and leaf shape. The weighting of these characteristics has been a point of contention among scholars who support/apply the interim Balfourian system of 44 equal series, versus the later subdivided system proposed by Sleumer and further developed by Davidian, Cullen and Chamberlain.

Linné

found that the structure of the flower in particular (basket, pea, mesh, lip, etc.) was particularly useful as a characteristic for a species to be assigned to a family. This is because the structure and appearance of the flower in particular have played a significant role in the plant's ability to survive. There is an interaction between plants and insects and birds regarding pollination and the exchange of nectar/pollen. To be able to make this specialized and complicated flower has it been necessary that many genes for a long time have been able to cooperate

between themselves to create this relationship.

Other complicated properties may mean less when a plant must determine by species. Here I think of the traits that plants from the dawn of time have solved in the same way. As an example above can be mentioned the common features of the heather order (Ericales): undivided leaves, regular flowers, the stamens placement on the base of the flower and many long axillary leaves. These properties are therefore often worthless to use in a determination rule for the plants that are placed lower in the hierarchy.

Good characteristics.

Examples of complicated properties have been hard to find, if they don't investigated using genetic methods (i.e. inheritance studies), and everything else is only guesswork. The following "good" self-creators are used: the placement of the flower on stem, formation of calyx, petals, dust road, capsule etc., but to a lesser extent indument and hair types". Examples of "bad" characteristics
that is controlled by one or few genes can varies: sticky bumps, color nuances in buds, leaves and flowers and genes, which regulates the distribution of hairdressing.

These morphological differences uncertain characteristics have been used to group the plants into species, the species in series are, series in subseries, etc.

The structure of the family tree

To build a family tree, put the families at the bottom (the trunk), above that the sections (main branches), subsections (branches), the species (the outermost branches) and the varieties are the leaves. This Family Tree" is built based on morphological traits, it is hoped; correspond to the evolutionary development (phylogeny) of the rhododendron genus through the ages within the Rhododendron genus. For division into subgenera (subgenus) and sections have just been shown to be the structures that are complicatedly built (governed that many genes), while subsection and species level are separated by traits that are smaller composite. Properties that are managed only of one gene, are often completely useless for other than description of variants within a species.

Examples of complicated properties have been hard to find, if they are not investigated by genetic methods (ie inheritance studies) and everything else is just guesswork. The following "good" characteristics are used: the position of the flower on the stem, the shape of the calyx, petals, stamen, capsule, etc., but to a lesser extent indument and hair types". Examples of "bad" characteristics that are controlled by one or a few genes can be mentioned: sticky buds, color nuances in buds, leaves and flowers, and genes that regulate the distribution of hairs. These morphological structural differences have been used to group the plants into species. The species into series, series into subseries, etc. Structure of the family tree. To build a family tree the families are placed at the bottom (the stem), above which the sections (main branches), subsections (branches), the species (the outermost branches) and the varieties are the leaves. This family tree, which is built up based on morphological features, is hoped to correspond to the

evolutionary development (phylogeny) of the rhododendron family through the ages.

The researchers have not yet been able to create this family tree with certainty, as they have not agreed on how the characteristics should be weighted. They also do not agree on which species should be included in the series, or whether a specific plant belongs to a species, a hybrid or a variety.

New methods: DNA characteristics.

Now, however, there is help to retrieve genes easily with modern genetic technology. Here, one can consider an incredible number of safe distinguishing marks, such as the differences in a DNA sequence base substitutions), in contrast to the few and more or less sliding uncertain distinguishing marks, which are used within morphology.

The structure of the DNA molecule

The DNA molecule is the hereditary material that contains 4 different chemical letters called A, T, G, C. The 4 letters are in different order.

The DNA molecule often consists of perhaps 1000,000,000 letters and is spiraled up into a chromosome. Such a chromosome can also be divided into perhaps 10,000 genes of approx. 500-50,000 letters (5% of the letters are in the genes)

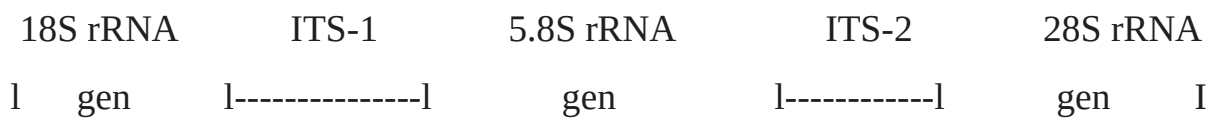
Recent research Molecular geneticists have now agreed to investigate special variable DNA regions on an international level. They will find similarities and differences between plant families and their associated species to create a family tree of the plants evolutionary history. It has been shown that the pieces of a DNA

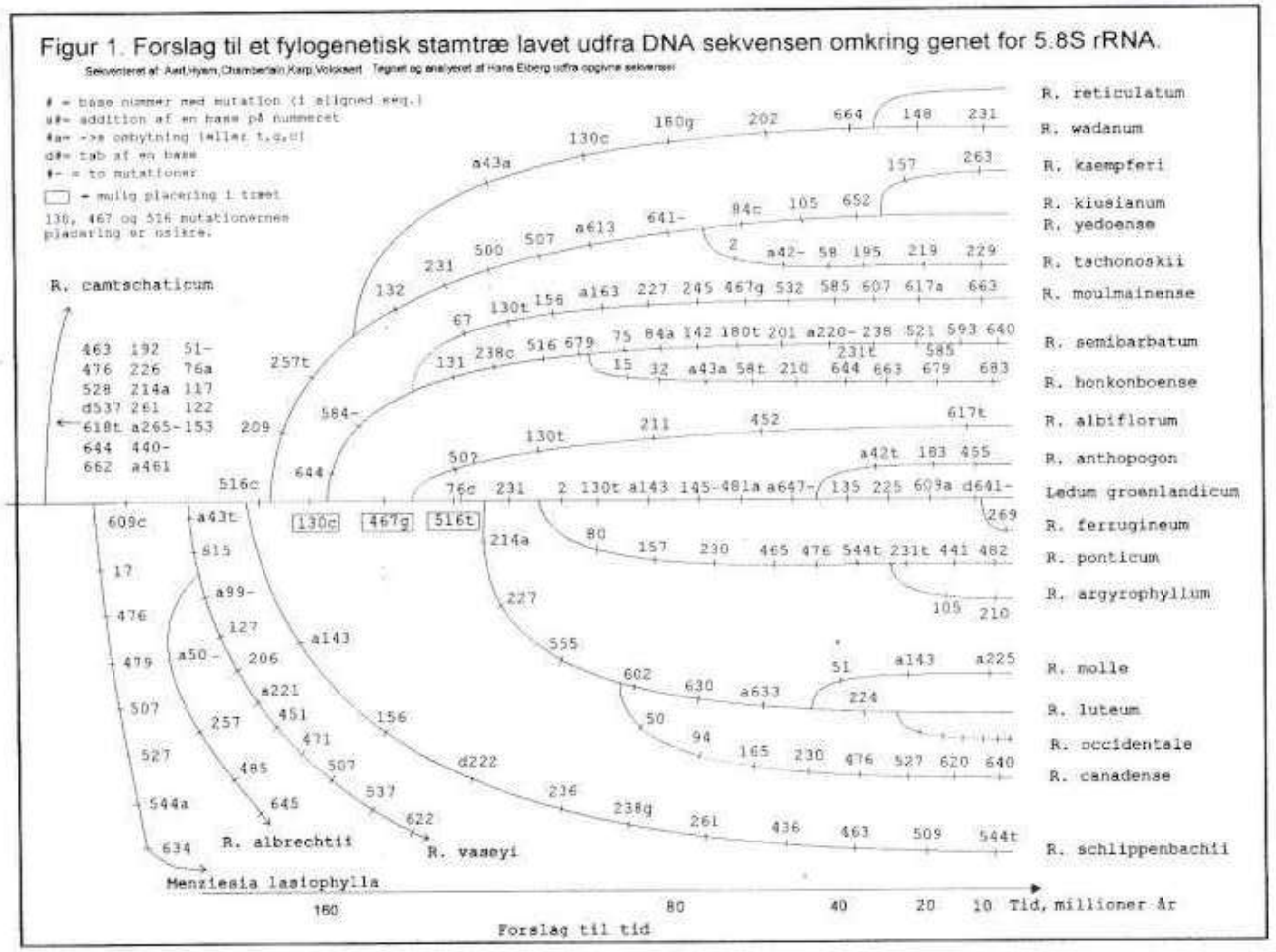
molecule that lie between coded regions (genes/gene parts) vary most from species to species. For every 500 letters in these areas, approx. one letter construction or other change (mutation) occurs every 4 million years. Whereas changes in a gene are observed 10 times as rarely, as the mutation in a gene is most often harmful and will disappear quickly.

DNA materials The first DNA sequences (the order of A,T,G,C) that have been looked at are areas called ITS (ITS-1 and ITS-2) and are marked on the drawing as l------ l. The ITS lies between the genes that code for three components of a ribosome (18S rRNA, 5.8S rRNA and 26S rRNA).
The DNA sequence of ITS can be found on the association's Home page on the internet (under taxonomy). 660 letters have been determined on at least 29 different species and more are gradually added. Also for other plants related to Rhododendron, DNA sequences in this region are shown. There are now several research teams studying DNA variations in ITS, e.g. led by D. Chamberlain (Edinburgh, Figs. 1 & 2), K. Kron (USA) (Fig. 3, published in ARS vol. 52 (2) 1998) and S. M. Scheiber et al. (USA), whose results are included in Fig. 5. Recently, a gene in the chloroplasts (plastids) called matK (ribosomal maturase) has also been sequenced on a large number of rhododendron species by research teams from the USA and Japan (K. Kron 1997 and Y. Kurashige et al.). The family tree (Fig. 4) I have constructed from the matK gene is in good agreement with the family tree constructed from the ITS regions.

Principle of the construction of family trees from DNA sequences

Family trees are constructed according to the principle that for approx. 100-200 million years ago, the ancestors of the rhododendron family had a specific order of letters around the gene for 5.8S rRNA. Mutations/changes in the order may occur, and some of these changes are inherited from generation to generation. If new species arise from a few/single plants by chance isolation, all its descendants can get the mutation in question.





Example of aligned DNA sequences (1-30) from 5 species can be seen below. Alignment means, that the letters are placed one below the other and arranged as well as possible. Then it is necessary to insert spaces (-/fracture) if there is deleted or added some letters in the individual species, for to make it fit as best as possible. The bold letters show the differences from the other species.

If a species is the only one that has this difference, a mutation must have occurred in this one. For *R. ferrugineum* or one of its ancestors has undergone a mutation

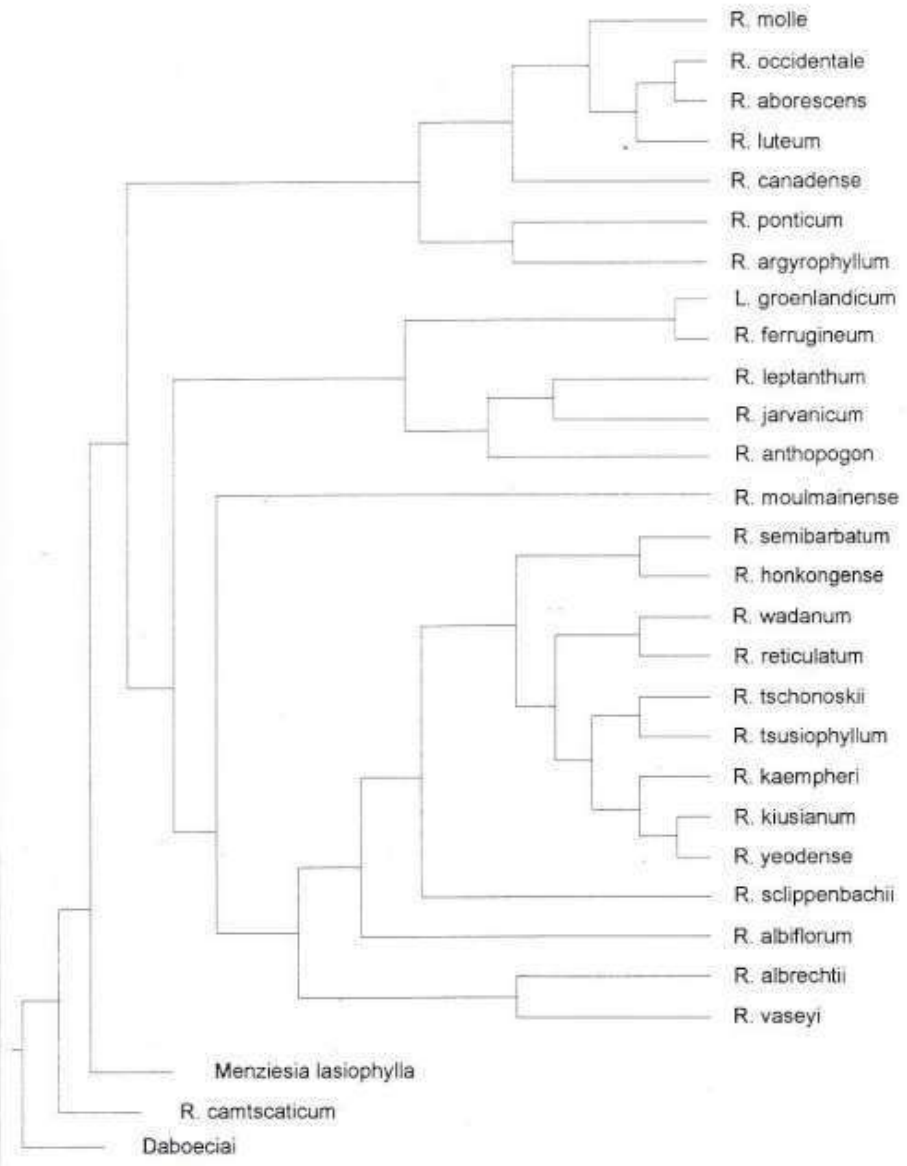
(C -> T) at position 2. *R. ferrugineum* and *R. groenlandicum* are quite similar and therefore closely related.

EDB methods. It is not an easy task to construct one
family tree in hand, and there is also made several EDB software packages (Phylip, Paup etc.) for this purpose. Most of them, however, programs can only analyze a few species and short sequences, if quite ordinary PC are used and inside a reasonable time. I have therefore tried to analyze aligned sequences from 29 species using the computer package Phylip on a "big" Sun-Sparc 20. Results of DNA analyzes at ITS Based on sequences from Chamberlain's group I have found the following results (Rhodo-Ny't l/97). It turns out that *R. ponticum* 's and *R. argyrophyllum* 's DNA is 99% identical in alphabetical order (-5 letters), while between *R. ponticum* and *R. ferrugineum* the similarity is only approx. 96%, but between *R. argyrophyllum* and *R. ferrugineum* only approx. 95%. Hence, can one concludes that *R. argyrophyllum* and *R. ponticum* is most closely related and *R. argyrophyllum* is less closely related to *R. ferrugineum* than *R. ponticum*. You can further conclude that all the species that have exactly the same letter changes, must be very closely related. There are also differences in the length of the gene, as *R. ferrugineum* has lost a letter I compared to the other two elepidotes. Interestingly, *Ledum groenlandicum* is almost identical (-1 letter) to *R. ferrugineum*, which has helped to *Ledum groenlandicum* has been changed to *Rhododendron groenlandicum*. Species like *R. anthopogon, R. molle* and especially *R. camtchaticum* varies most from e.g. *R. ponticum.* I have shown 24 species on figure l. Each number in the figure indicates a place in it aligned sequence, where there is a deviation in relation to the other species. The family board is built

according to the principle that the fewest possible mutations have occurred. It can be seen from figure 1 that there is preserved

R. ponticum TCGA-AACCTGCCAA CAAGCAGAAAACTTG
R. ferrogineum TTGA-AACCTGCCAA CAAGCAGAAAACTAG
L. groenlandicum TTGA-AACCTGCCAA CAAGCAGAAAACTAG
R. camtchaticum TCGA-AACCTGCCAA CAAGCAGAAAAGTTG
Daboecia TCGATAACCTG-----A CGATCAGAAAAGTTG

Figur 2. Forslag til rhododendronfamiliens stamtræ



Beregnet med programmet DNAML (PHYLIP) af H. Eiberg
Udfra sekvenserne på ITS området.

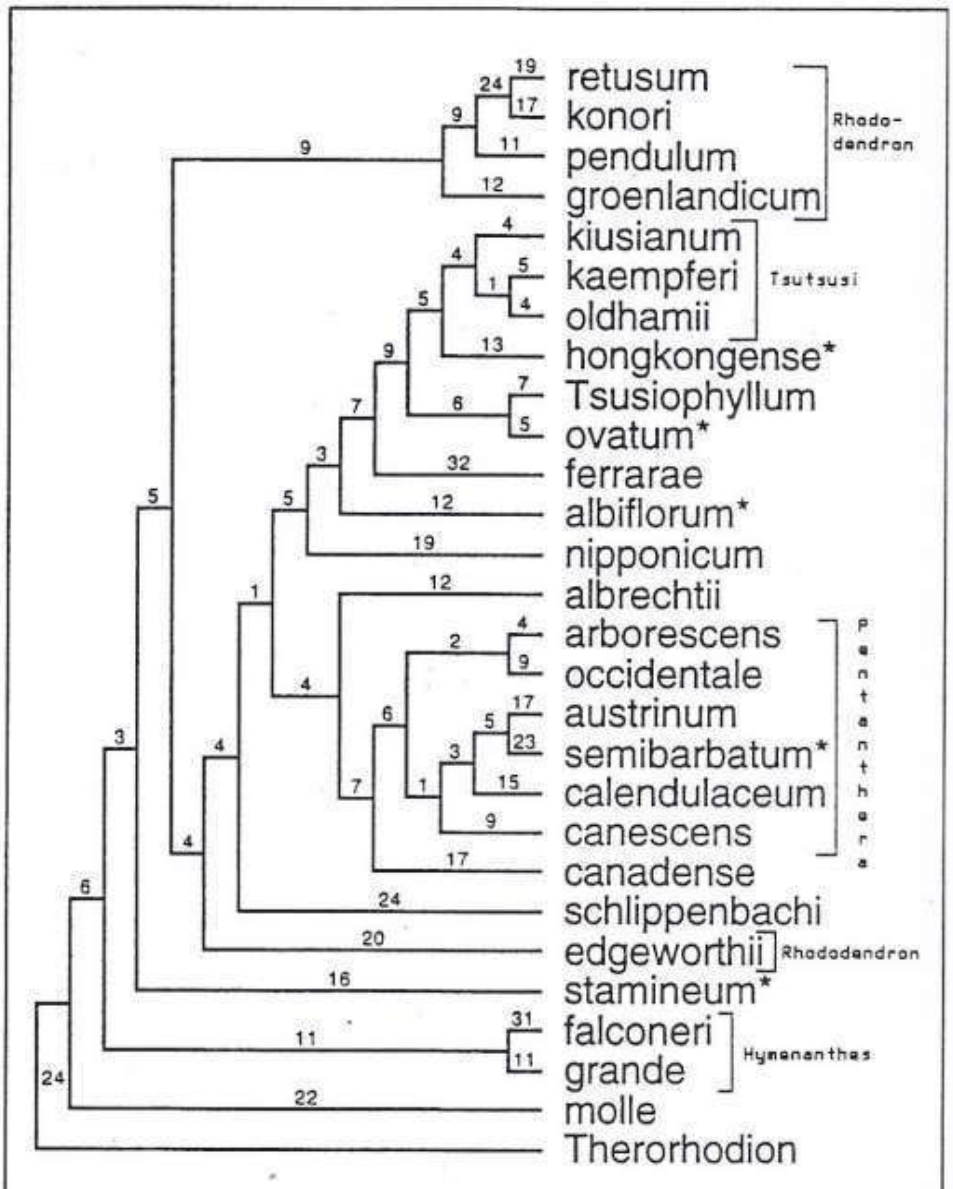


Fig. 3 Et eksempel på et træ beregnet på sekvenser fra adskillige ITS sekvenser fra rhododendron arter. Tallene angiver antallet af forandringer i DNA sekvensen. Stjerner angiver de arter med blomster på sidegrene. Sekventering og analyse af K Kron & SL Johnson (ARS, 52: 70-72. 1998)

Page 193 a photo of R. groenlandicum

survived approx. 7 -20 mutations per species in ITS area since the rhododendron family arose. To be sure of the kinship in the first period requires more analyses, however is carried out. In cases where there are only l-2 mutation(s) that separate two branches from each other there is great uncertainty, and there are several of these in the family tree. *R. vaseyi* which is currently classified in family with *R. canadense* is on the figures together with *R. albrechtii*, which means that *R. vaseyi* is completely wrongly placed in section *Rhodora*. Is *R. camtchaticum* a rhododendron? According to DNA, it is more closely related to Irish heather (*Daboecia*). In addition, *Menzesia lasiophylla* is more similar to a Rhododendron in DNA than *R. camtchaticum*. Based on this can one say that if *R.* camtchaticum is a rhododendron, then *Menzesia* is also one rhododendrons. Since I analyzed 28 species and incorporated *Daboecia*, is far several species put in place (Figure 2). However, *R. moulmainense* and *R. albiflorum* are differently placed in the two figures and are therefore problematic species. In ARS vol. 52(2) 1998, K. Kron has published his results using ITS the area (Fig 3). There are quite a few disagreements, but also many points of similarity with my own previous calculations. Kron has a number of new species included and lacks many of the species which I have used, which makes a together equation more difficult. In particular, Kron points out that the species *R. edgeworthii* is more closely related with azalea than with the other elepidotes and that it may has arisen by hybridisation. That *R. molle* is not related to *R. occidentale*, which I found, shows that there is something wrong. Furthermore, *R. hongkongense, R. semibarbatum* and *R. stramineum* far too far apart in terms of evolution. I believe that it is necessary to analyze both Kron's and Chamberlain's DNA results combined, to find

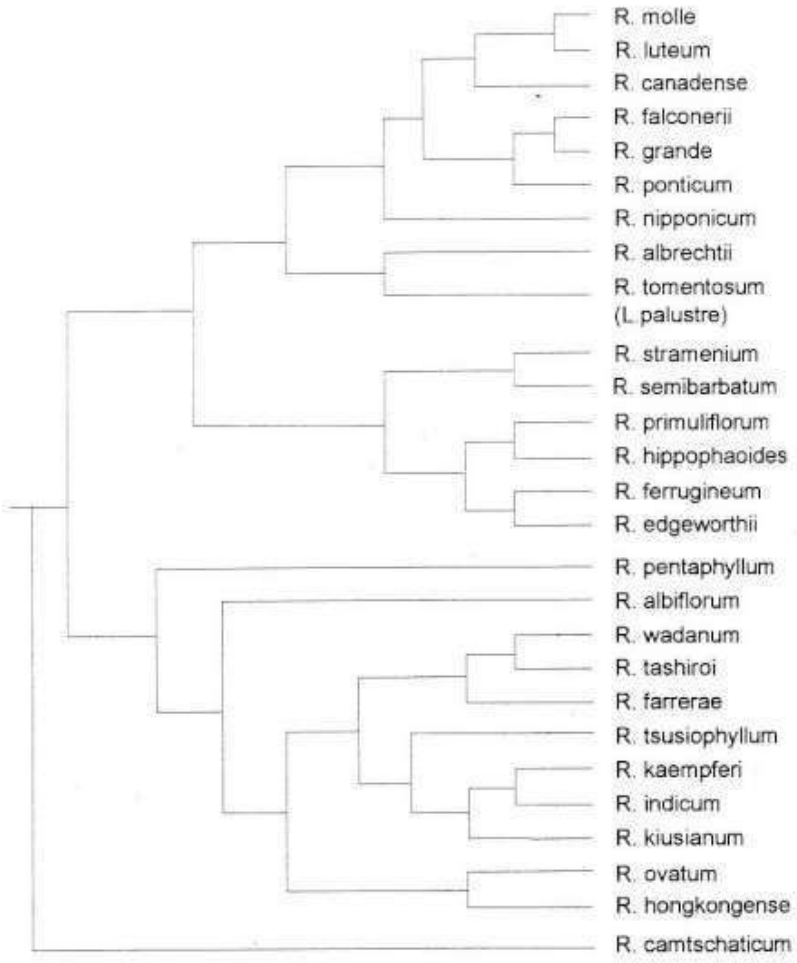
a better explanation. Especially since Kron's results in several areas do not fit with earlier conclusive evidence from mother phylology. Furthermore, one can doubt her results, since that family tree she has constructed, contains far more mutations in relation to the family tree I have constructed (Fig. 1). It suggests that her alignment is uncertain, perhaps too many deletions? Unfortunately, I haven't got hold of Kron's method and basic material, as it has not been sent to Biobase (a public database which, among other things, contains information on DNA sequences).

Most recently, an American research team has (S. M. Scheiber et al. in Press. Horticulture) analyzed 13 different species from *Pentanthera*. Others had previously studied five of these species with one identical result. Pedigree analyzes show that all species from the *Pentanthera* subsection are very closely related (Fig. 5)

Results of DNA analyses on the mat-K gene

DNA is here sequenced on the mat-K gene by Kron (Amer J Bot 84), as well as by a Japanese research team (Y. Kurashige et al. in press J. of Jap. fine). The stem foot (Fig 4), which I have constructed shows that *R. edgeworthii* is closely related to the elepidotes species. The family tree matches them very well results previously found by the morphological studies. Here is the length of it examined gene the same for all the species (1200 bases/letters), as opposed to the region studied on ITS (not a gene)

Figur 4. Forslag til rhododendronfamiliens stamtræ



Beregnet med programmet DNACOMP (PHYLIP) af H. Eiberg udfra sekvenserne af genet matK på plastidkromosomet.

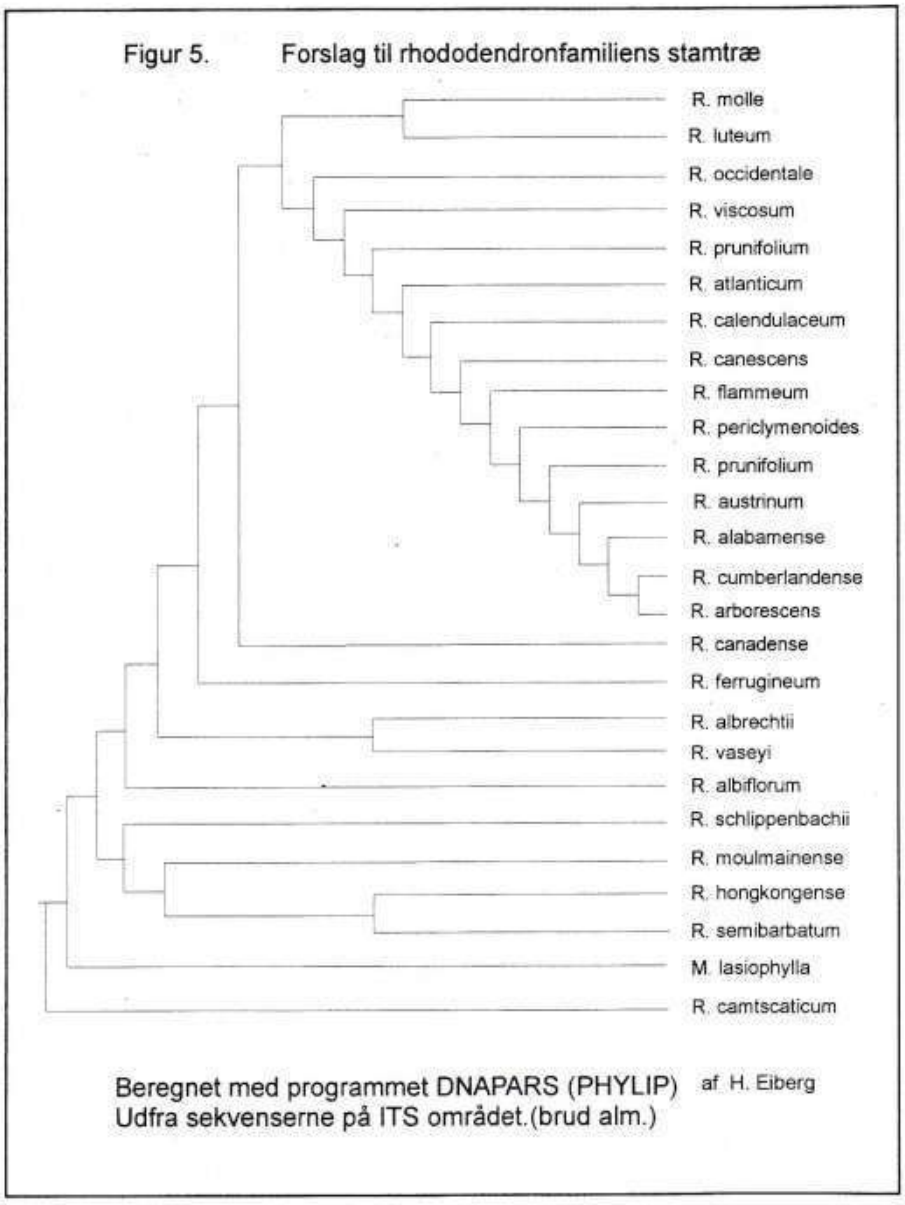


Photo of *R. camtchaticum*

Today it is doubtful whether *R. camtchaticum* is a Rhododendron. Never mind - it is beautiful.

and therefore reduces the possibility of errors by the alignment. Therefore, you can trust better on the oldest branches when this DNA is not analyzed. The mat-K gene is not as effective as ITS when closer related species must be examined, as here are the sequencers on the mat-K gene more the same.

What can we use the family tree for?

Knowledge is always nice to have, and they can learn to explain the evolution. i.e. what plants are there closely related. We can confirm or denied the new classification. The family tree can explain the process of speciation itself. The family tree can provide one rule of thumb regarding which species there crossed. If they are far apart, many of the hybrids (if they can be made) probably be weak. The analysis method can be used for species determination, in that two plants that have significantly different DNA sequences must belong to two different species. DNA sequencing can used for paternity testing, i.e. distinguish between a hybrid and a species. The hybrid will have both the paternal and maternal sequences, which will be different. DNA can be analyzed on a single dried leaf, and from the DNA profile, the plant can be placed in the system. It will be interesting that DNA determines the ancients collected species that formed the basis for the Balfourian system. Are they extinct species or hybrids?

The speciation process of species:

There are different opinions about it the formation of new species. Some believe that new species are formed by hybridization between two different species, and the descendants from this will stabilize into one species by selection. Others incl. undersigned believes that new species are formed from mutation within the same species. Few plants are origin of a new species, which is then propagated hot under constant selection. Speciation is a gradual process, and a species is not constant over time. If the "hybrid theory" is the common method in speciation, it will not be possible to construct a family tree based on the variation in one small gene that corresponds to the systematics created from many morphological characters. From other family tree studies from other plant groups, is inconsistencies were found in a few cases, but the "hybrid theory" is not that common method of speciation.
In other words: if the DNA analysis on the ITS region of a species provides a secure location in the family tree, and this one is completely different from the location which the analysis of the mat-K gene gives, is there signs that the species may have originated by hybridization. If the discrepancy repeats itself when examining two completely new DNA regions, the proof is first at home.

Hybridization

It is no advantage for two species that have specialized for each environment, that form a hybrid. Hybrids and hybrids are frequently found in nature, but they very rarely survive in the long term, also because they set fewer seeds than the species and therefore will be outcompeted. Under I saw the Danish expedition to Sikkim examples of hybrid swarms which were by perishing. Areas with it yellow *R. thomsonii* (*R. thomsonii x R. campylocarpum* F2), *R. decipiens* (here *R. wightii* x *R. hodgsonii*) and *R. sikkimense* (*R. arboreum x R. thomsonii* F2- ?), was protected by

felling competing vegetation. If two species differ from each other by many mutations, so that the relationship between them is distant, then they will not - or will hardly - be able to interbreed as the species have managed to form crossing barriers. There can, however also a single major chromosome mutation (exchange of a large amount of material between two chromosomes), i.e. one rapid speciation, which ensures a hybridization between the very closely related species *Ledum*/ R. *groenlandicum* and *R. ferugineum*.

Uncertainty

The biggest uncertainty in DNA classification is that there may be several different possibilities for aligning the sequences. This can result in an incorrect placement of the species in the tree. There are more programs i.a. ClustalW and Map, which can be used directly on the Internet, where free, can perform this process. Just be aware that the ITS area is not a coding region, and therefore will it be more "allowed" for that to happen additions and deletions of letters more frequent than for the coding regions. By using the default settings will the alignment is done incorrectly, but with correct use makes the analysis far more secure results than those now published (Fig 3). Only approx. 50 different species, and therefore there are many "holes", and here one can only guess to how evolution has taken place.

Conclusion

Results from the mat-K gene have shown that one can reproduce the results from The ITS area, and therefore this form will for analysis revolutionize the taxonomy.

Safe family trees will be constructed when 3-5 gene regions of perhaps 200 species have been analyzed, which represents a so wide selection of the rhododendron species which possible. They are working tenaciously both in England, several places in the USA and in Japan on sequencing different species and different DNA regions. Especially it will be interesting when several of the problematic species and the dried plants from herbariums included in the analyses.
After finishing the analysis, one can hope that the species for the last time are put into system and will preserve their names for posterity. However, I doubt that the researchers can agree to draw the boundaries between the different groups such as families, genera, sections, etc.; here must the morphologic features also come into play. It is difficult to say about our future determination keys will be more difficult to use than they are now. I think that the new research will clarify the current problems.

Referencer:

Aert R., Hyam R,. Chamberlain D, Karp A, Volckaert G. (1996). Upubliceret (fra Biobase, NCBI). Chamberlain DF. (1982). A revision of rhododendron II. Subgenus Hymenanthes. Notes from the Royal Botanic Garden Edinburgh. 39(2): 209-486.

Cullen J. (1980). A revision of rhododendron L Subgenus Rhododendron sections Rhododendron & Pogonanthum. Notes from the Royal Botanic Garden Edinburgh. 39(1): l-207.

Davidian HH. (1982). The rhododendron species I. Lepidotes. Oregon USA. 1-43r.

Eiberg H. (1997). Rhododendronfamiliens stamtræ 2. Rhodo-Nyt 97 1/97: l4-18.

Eiberg H. (1996). Konstruktion af rhododendronfamiliens stamtræ. Rhodo-Nyt 4/96: 19-22.

Kron K. (1997). Phylogenetic relationships of rhododendroideae (Ericaceae). Amer. J. Botany 84(7):973-980.

Kron K. (1998). Using DNA sequences to estimate evolutionary relationships among rhododendron and azaleas. Amer. Rhodo. Soc. 52(2):70-12.

Kurashige Y, Mine M, Kobayashi N, Handa T, Takayanagi T, Yokawa T (1998). J. Japanesee Bot. I trykken (fra Biobase. NCBI).

Scheiber SM, Jarret RL, Robacher CD. (1999). Upubliceret (fra Biobase, NCBD.

Sleumer H. (1949). Ein system der gattung Rhododendron L. Bot. Jahrbücher 74: 511-553.

Synge PM. (1967). The Rhododendron Handbook I. Rhododendron species in general cultivation. The Royal Horticultural Society. London