

TEACHER'S GUIDE

DNA to Darwin Case Study

Darwin's 'abominable mystery'

The origin of flowering plants

Version 1.1

Laura Kelly

Royal Botanic Gardens, Kew

Dean Madden (Ed.)

NCBE, University of Reading



Charles Darwin

Darwin's 'abominable mystery'

Writing to his friend Joseph Hooker in 1879, Charles Darwin famously referred to the origin of flowering plants — the angiosperms — as an 'abominable mystery'. Even today, botanists argue about what the first angiosperms must have looked like. There are two competing hypotheses. One says that the first flowering plants were woody shrubs with large blooms like modern magnolias; the other says that they were almost the opposite — non-woody plants with small flowers, like modern black pepper plants.

DNA evidence from the chloroplast genome is used by students in this *Case Study* to try to solve the century-old mystery.

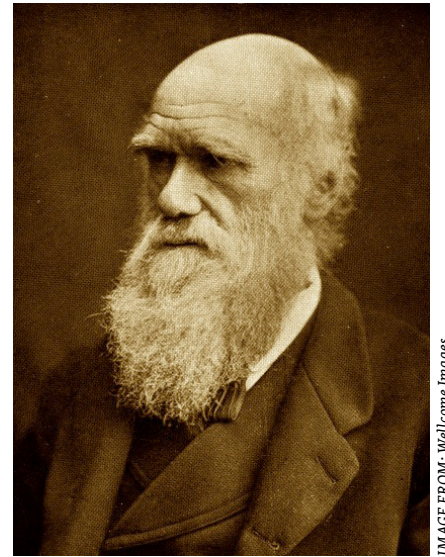


IMAGE FROM: Wellcome Images.

The origin of flowering plants

This exercise is based on the premise that if we can identify the earliest diverging group of extant (living) angiosperms this may provide an indication of what the earliest flowers (or flowering plants) looked like. By understanding the phylogenetic relationships among the groups of angiosperms, scientists hope to explain how this group of plants diversified and rose to its current dominant position amongst the Earth's flora. A better knowledge of phylogeny should also lead to improved understanding of plant development, metabolism, pathology and genomics. Reasons such as these help to explain why the reconstruction of the earliest branches in the angiosperm tree has been a major goal for plant phylogeneticists in recent years.

Outline of the activity

In this *Case Study*, students are given 26 chloroplast DNA (cpDNA) sequences derived from the ribulose 1,5-bisphosphate carboxylase/oxygenase large chain gene. The sequences come from angiosperms and gymnosperms. This gene (codenamed *rbcL*) has commonly been used for phylogeny reconstruction in plants, as cpDNA is easy to obtain and is suitable for examining relationships due to its relatively conserved nature, which facilitates alignment and thus comparison over large taxonomic distances. It is, however, important to bear in mind that this gene only represents a tiny fraction of the genetic information contained within a species. Consequently, we should refer to a tree that is derived from the *rbcL* sequences as a *gene tree*, as it represents the relationships between the sequences themselves (in contrast to a *species tree*, which depicts relationships between species).

We can in practice infer species relationships from the gene tree (which is the aim of this exercise), but much greater confidence could be placed on these inferences if we were to examine multiple datasets and find that they all point to the same arrangement of taxa. In published scientific studies this is precisely the approach that would be adopted: several independent datasets would be constructed from sequences of several different genes and compared. This would identify those relationships that are common to all of the individual gene trees. If the gene trees are in agreement then the datasets are normally combined to allow the production of species tree.

The data provided for use in this *Case Study* is a highly simplified version of the kind used in some of the numerous studies of angiosperm relationships (for a review, see Soltis and Soltis, 2004). Comparison of the results to those obtained from published studies has confirmed, however, that the analysis of this small dataset will reveal realistic relationships between the species included. It should be made clear to students that the published studies involve much more extensive datasets (more species and more genes, both of which are needed to produce reliable results) and more exhaustive methods of analysis.

The exercise simply involves building and interpreting a phylogenetic tree based upon some aligned sequences.

General reading

Reading the story in DNA: A beginner's guide to molecular evolution by Lindell Bromham (2008) Oxford University Press (Paperback) ISBN: 978 0199290918. *An engaging textbook on molecular evolution, which assumes no specialist mathematical knowledge and takes the reader from first principles. Although it's aimed at undergraduates, this superb book contains sufficient detail for PhD students, yet parts will appeal equally to 16–19 year-olds*

Flowering plant families of the world by V. H. Heywood *et al* (2007) Firefly. ISBN: 978 1 55407 206 4. *This is a completely revised edition of the lavishly-illustrated 1978 classic. It is the first popular book to take account of DNA evidence in the classification of angiosperms.*

Scientific publications

Most of these publications require paid access.

- Crepet, W. L. (1998) The Abominable Mystery. *Science*, **282** (5394) 1653–1654. doi: 10.1126/science.282.5394.1653
- Doyle, J.J. and Gaut, B.S. (2000) Evolution of genes and taxa: a primer. *Plant Molecular Biology*, **42** (1) 1–23. doi: 10.1023/A:1006349518932
- Kuzoff, R.K. and Gasser, C.S. (2000) Recent progress in reconstructing angiosperm phylogeny. *Trends in Plant Science*, **5** (8) 330–336. doi:10.1016/S1360-1385(00)01685-X

- Angiosperm Phylogeny Group (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society*, **141**, 399–436.
- Soltis, D.E. and Soltis, P.S. (2004) The origin and diversification of angiosperms. *American Journal of Botany* **91** (10) 1614–1626. doi: 10.3732/ajb.91.10.1614.
- Soltis, P., Soltis, D., and Edwards, C. (2005) Angiosperms. Flowering Plants. Version 3, June 2005. <http://tolweb.org/Angiosperms/20646/2005.06.03>

Websites

Angiosperm Phylogeny Website

www.mobot.org/MOBOT/Research/APweb/welcome.html

A huge site with detailed information about different groups of angiosperms. Includes a page for students.

The Floral Genome Project

www.flmnh.ufl.edu/flowerpower/default.html

An education site that includes some basic information on the evolution of angiosperms. The page for *Amborella* may be of interest, as it gives further illustrations and references.

Tree of Life Web Project

<http://tolweb.org>

This site includes pages on angiosperms and the Magnoliids.

University of California Museum of Paleontology

www.ucmp.berkeley.edu/anthophyta/anthophytafr.html

'The origin of angiosperms is still a mystery'. An introductory article about angiosperm evolution.

University of Illinois

www.life.uiuc.edu/ib/335/Origins/Origins.html

On the origin of the angiosperms. Notes from the university's course on plant systematics, providing brief coverage and some references and additional links.

Rubisco

November 2000 *Molecule of the Month* by David Goodsell

<http://download.rcsb.org/pdb/101/motm.do?momID=11>

doi: 10.2210/rcsb_pdb/mom_2000_11 (PDF Version)

Other useful resources

Botany: A blooming history (2011) BBC4 television series. Three 60-minute programmes. Occasionally available on BBC iPlayer in the UK.

Episode 1, 'A confusion of names', ends with an example of how phylogenetic trees devised from DNA data are being used by Kew scientists to identify possible sources of new anti-malarial drugs.

The private life of plants (1995) BBC Worldwide Ltd.

Region 2, PAL: two DVDs. 2 Entertain Video. ASIN: B0000AISJC.

The complete series presented by David Attenborough. 4 hours 52 minutes.

Episode 3 concerns flowering plants.

Requirements

Software

The software required, *Geneious*, can be downloaded free-of-charge from: www.geneious.com. The software is available for Windows, Macintosh and Linux operating systems. Only the free, 'basic' version of the software is required for this activity.

DNA sequence data

Students will need the *Geneious* document containing 26 aligned DNA sequences from angiosperms and some gymnosperms (which act as an 'outgroup'): **angiosperm.geneious**

Students' worksheets

Students will require copies of Student's Guide, pages 2-10. Each student or working group will benefit from having five different-coloured pencils or pens to mark their tree, but this is not essential.

Presentations

Teachers may find the *QuickTime* animations in the Introduction materials (available separately) and *PowerPoint* or *Keynote* presentations helpful for introducing this exercise. The slide presentation has photographs of all the genera used in the analysis, although it has not always been possible to obtain photographs of the exact species. *QuickTime* may be downloaded free-of-charge from the Apple web site: www.apple.com/quicktime

Educational aims

The activity reinforces students' understanding of DNA structure and how the DNA sequence encodes the amino acids needed to make a polypeptide. It reminds students about plant cell structure and how the chloroplasts have their own DNA.

The principle of alignment and generating an evolutionary tree from molecular data is introduced. The work also presents an ideal opportunity to study plant taxonomy, floral structure, plant reproduction and so on.

Prerequisite knowledge

Students will need to understand the basic structure of DNA. They will need to know that chloroplasts have their own DNA and to be taught the principle of sequence alignment and the generation of phylogenetic trees. The *Introductory Activities*, which are in a separate document, will be useful.

Answers to the questions on the worksheets

Page 8

- a. The DNA sequences are three times longer than the amino acid sequences because three bases are needed in the DNA to specify each amino acid.
- b. There is far less variation in the amino acid sequences than in the DNA ones, because each amino acid can be represented by several DNA codons (group of three bases). There is redundancy in the Genetic Code.
- c. DNA data is better for generating a tree in this case (try it with the amino acid sequence if you wish!).

Page 10

- d. Any of the gymnosperms can be used as an outgroup. They are more closely related to each other than they are to the angiosperm species.
- e. On the far left hand side of the tree, at the root.
- f. The first lineage to diverge contains *Amborella trichopoda*. This group also includes two herbaceous species that are paleoherbs. Some published studies have also resolved a group made up of *Amborella* and Nymphaeaceae as the first lineage to branch off.
- g. *Amborella trichopoda* is in the Magnoliid group; *Nymphaea* and *Cabomba* are in the paleoherb group.
- h. *Amborella* is a small, woody shrub, classified within the woody Magnoliids. Therefore, its placement as the first branch on the angiosperm tree can be seen as supporting the woody Magnoliid hypothesis. However, the inclusion of the two paleoherb species in the grouping does little to help resolve the mystery!
- i. Neither group is particularly well-defined. From the arrangement of taxa within the tree it is evident that the woody Magnoliids and paleoherbs are not monophyletic, i.e. they are not natural groups. For example, *Amborella trichopoda* and *Magnolia grandiflora* (both previously placed in the woody Magnoliids group) are placed far apart from one another. Several species that had been placed within the paleoherbs form a group with the woody Magnoliids. Botanists have now refined the Magnoliids in light of results from molecular phylogenetic analyses. Taxa that were previously contained within the paleoherbs, such as *Piper nigrum*, have now been classified within the Magnoliids. The paleoherbs are no longer distinguished as a group.

Additional comments on the tree

The tree is not very accurate as only a single gene was used to construct it: researchers usually include sequences from multiple genes or DNA regions. Because of the poor level of resolution, we can infer little about the relationships. All four eudicot taxa form a monophyletic group (i.e., they share a single common ancestor). The two monocot taxa are also monophyletic, but we cannot infer anything about their relationship to the eudicots or to the woody Magnoliids. None of the gymnosperms is more closely related to the angiosperms than the other. For this reason the mystery of the origin of flowering plants remains.

The dataset

The data provided has 26 sequences from the *rbcL* gene from gymnosperms and angiosperms. Below is a list of the taxa included in the dataset, followed by the *GenBank* accession number for each sequence (*GenBank* is a database of publicly-available DNA sequences and can be accessed at www.ncbi.nlm.nih.gov). Where the *rbcL* sequence has been derived from the full sequence for the plastid genome of a particular taxa this is noted. All other sequences are partial *rbcL* sequences.

GYMNOSPERMS

<i>Cycas taitungensis</i>	AP009339 (full plastid genomic sequence)
<i>Ginkgo biloba</i>	DQ069500
<i>Pinus koraiensis</i>	AY228468 (full plastid genomic sequence)
<i>Welwitschia mirabilis</i>	AJ235814

ANGIOSPERMS

Paleoherbs

<i>Anemopsis californica</i>	AF197597
<i>Asarum caudigerum</i>	AB205608
<i>Cabomba caroliniana</i>	M77027
<i>Nymphaea alba</i>	AJ627251 (full plastid genomic sequence)
<i>Piper cenocladum</i>	DQ887677 (full plastid genomic sequence)

Eudicots

<i>Buxus microphylla</i>	EF380351 (full plastid genomic sequence)
<i>Dicentra eximia</i>	L37917
<i>Mahonia bealei</i>	L12657
<i>Platanus occidentalis</i>	DQ923116 (full plastid genomic sequence)

Monocots

<i>Acorus calamus</i>	AJ879453 (full plastid genomic sequence)
<i>Asparagus officinalis</i>	AY149374

Magnoliids

<i>Amborella trichopoda</i>	AJ506156 (full plastid genomic sequence)
<i>Annona glabra</i>	AY841596
<i>Austrobaileya scandens</i>	L12632
<i>Canella winterana</i>	AY572265
<i>Chimonanthus praecox</i>	L12639
<i>Drimys granadensis</i>	DQ887676 (full plastid genomic sequence)
<i>Illicium oligandrum</i>	EF380354 (full plastid genomic sequence)
<i>Laurus nobilis</i>	AF197593
<i>Liriodendron tulipifera</i>	DQ899947 (full plastid genomic sequence)
<i>Magnolia grandiflora</i>	AF119180
<i>Trimenia moorei</i>	AY116658

Selection of the data

The 22 angiosperm sequences comprise the ingroup, whilst the four gymnosperm sequences will act as the outgroup. Within the outgroup, one taxon has been included from each of the four major extant lineages of non-flowering seed plants: Cycadophyta; Ginkgophyta; Coniferophyta and Gnetophyta. Within the ingroup are taxa that represent the two groups that are the focus of the hypotheses we are testing (i.e. woody Magnoliids and Paleoherbs). Representatives of the two largest groups of angiosperms, the monocots (which some authors have previously included within the paleoherbs) and eudicots ('true' dicots), are also included.

Results

Depending upon the settings used, diagrams you obtain may look different

