

## ORIGIN

### Research/technological quality, including any interdisciplinary and multidisciplinary aspects of the proposal

**State of the Art.** Flowering plants (angiosperms) are the dominant life form on land. They drive terrestrial ecosystems by using photosynthesis to produce most of our oxygen and complex nutrients. We depend on them, directly or indirectly, for almost all of our food, in addition to pharmaceuticals, textiles, timber, biofuel and the sheer aesthetic pleasure of flowers and other garden plants (1). Most plant biology research is directed at a few very successful “model plant” systems (especially the “thale cress”, *Arabidopsis thaliana* (2) a model for the dicots; and recently the “false brome” *Brachypodium distachyon* (3) a model for the monocots) that present very small genomes and are simple to cultivate and handle in molecular research, in addition to crop plants with major economical importance in Europe and elsewhere. Substantial recent effort has focused on developing and applying genomic tools for these species [1].

A major difficulty in interpreting vast genomic data sets from single species is that plant genomes are known to be fluid over relatively short evolutionary time spans, due to substantial genomic turnover through gene duplication, diversification and extinction. This is true for most of the major gene families governing organism biology (4). Turnover includes tandem gene duplication events on individual chromosomes and occasional genome duplication events (5-7). Since all plant genes ultimately derive from duplication and subsequent diversification processes (8,9), understanding deep evolutionary history (phylogeny) is important for predicting the function of genes in multigene families (ie most genes). Most plants that are useful to humanity belong to one of two flowering plant angiosperm groups, the monocots and eudicots. We now know that these two major groups diverged from each other close to the origin of the flowering plants (10).

As such, it would be invaluable to develop genomic tools for a model plant lineage that pre-dated this major evolutionary split. This would be broadly useful for plant biology research, as it would allow us to pinpoint which gene families predate vs. postdate the origin of most the world’s major crop plants and to reveal recent vs. ancient genes, permitting inference of gene orthologs, gene copies that are inherited in different plants through speciation events, vs. paralogs, gene copies derived from gene or genome duplication events, which may often gain new functionality (11,12).

**Hydatellaceae: a new window into flowering plant origins.** There is now an excellent candidate plant to develop as a new model plant system for early angiosperms, the genus *Trithuria* (family Hydatellaceae). Dr Graham and his students recently demonstrated that this family originated very close to the origin of flowering-plant phylogeny (13). It is an attractive prospect for development as a new model-plant system [2] because its stem lineage pre-dates the origin of monocots and eudicots (10,13), and the plants are small and easily cultivatable; other early flowering plants are mostly long-lived trees.

The new phylogenetic placement of Hydatellaceae near the origin of all flowering plants (13) re-opens debates on early angiosperm evolution, including the origin of their basic reproductive biology and ecology due to the following characteristics:

- *Trithuria* possesses combined characteristics of flowers and inflorescences in a possibly primitive structure that has been called a “nonflower” (14,15). This structure has an inside-out floral morphology, with pistils surrounding stamens.
- The entire reproductive unit develops centrifugally, in contrast to centripetal or simultaneous development in typical flowers.
- All *Trithuria* species exhibit a pre-fertilization maternal resource allocation to ovules/seeds in contrast to the remaining flowering plants. Post-fertilization allocation has been hypothesized to be a key reproductive innovation associated with the origin and subsequent evolutionary radiation of flowering plants (16).

However, knowledge of species boundaries in *Trithuria* remains unexplored, therefore constraining its use as a new model plant and ultimately the development of genomic tools for detecting orthologs and paralog genes.

The main objective of ORIGIN is to unravel the evolutionary history of this plant, by defining its species boundaries and by facilitating its use as an “early angiosperm” model plant to root studies of multigene family diversification and gene function. To reach this objective, three main goals have been defined.

### Goal 1: Define species boundaries and uncover phylogenetic relationships in *Trithuria* (Hydatellaceae).

## ORIGIN

Under this first goal, ORIGIN will address the following questions:

- What are the genetic limits of *T. submersa* as a species (its taxonomic definition)?
- How is it related to its close relatives (its local phylogenetic relationships)?

The family Hydatellaceae was recently re-classified to comprise only a single genus, *Trithuria*, with 12 distinct species that are found primarily in Australia (17). Several species are currently being investigated for their suitability for cultivation and potential as a model organism, by Dr. Paula Rudall and colleagues at Royal Botanical Gardens, Kew, UK. The most promising of these is *Trithuria submersa*, a southern Australian species that can be grown rapidly from seeds on growth medium. This species is short-lived, very small, tractable to handle in culture, and is primarily self-fertilizing (18). Although its genome size is currently unknown, it is likely to be small given its rapid generation time and extremely small chromosomes (19). These are all attractive features in any model plant system (20).

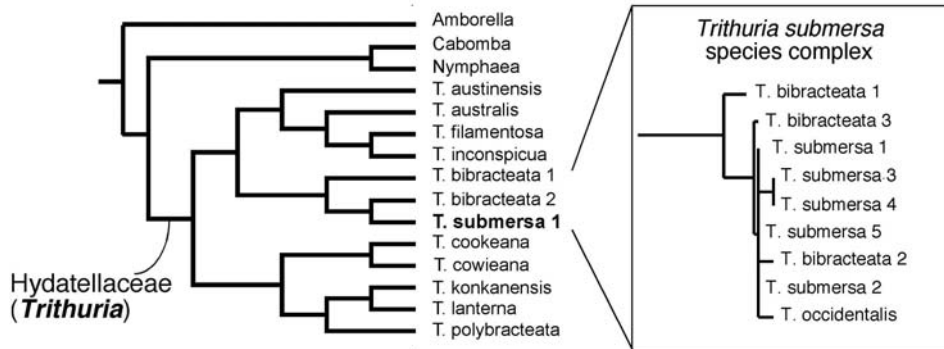


Fig. 1. Preliminary phylogeny of *Hydatellaceae* estimated from chloroplast data. Inset figure shows a sampling based on multiple populations of three species in the *Trithuria submersa* species complex (individuals from distinct populations are indicated with numerals).

*Trithuria submersa* has a large geographic range across much of southern Australia, whereas the two closest relatives, *T. bibracteata* and *T. occidentalis*, occur in localized areas in Western Australia. *Trithuria occidentalis* was previously taxonomically lumped with *T. submersa* based on its overall morphology (21), but a recent morphological study showed that these two species are morphologically distinct from each other (17,22). A preliminary phylogenetic survey based on chloroplast (plastid) genes raises the strong possibility that *T. submersa* is a species that is “nested” phylogenetically inside another species (*T. bibracteata*) and that a third species (*T. occidentalis*) may also be part of this rather poorly-delimited species complex (Fig. 1, inset; 23,24). Thus, it is unclear whether current species delimitation (ie taxonomy) accurately reflects ancestor-descendent relationships (ie, genetic history). Genetic boundaries are arguably at least as important for clarifying species limits as morphology (25). The boundaries of *T. submersa* in the context of its close relatives need to be clarified if it is to be used as a model plant for genetic and evolutionary research. **Thus, ORIGIN, in a first goal, aims to define what are the genetic limits of *T. submersa* as a species (its taxonomic definition), and how it is related to its close relatives (its local phylogenetic relationships).**

### **Goal 2: Resolve a key ambiguity in early plant evolution.**

Under this **second** goal, ORIGIN will address the following question:

- Where does the root of flowering-plant phylogeny belong?

Charles Darwin famously observed that the origin and rapid diversification of flowering plants in recent geological time is an “abominable mystery” (26). One major mystery concerning angiosperm origins was partly resolved with the recognition that the root of flowering-plant phylogeny is at or close to a branch leading to the New Caledonian shrub, *Amborella trichopoda*, and the two water-lily families, Cabombaceae and Nymphaeaceae (27-30). However a persistent and unresolved question concerns the precise placement of the angiosperm root: Is it on the branch leading to *Amborella* alone, or to *Amborella*+water lilies?

The arrowheads in Fig. 2 indicate three statistically equivalent placements for the root of flowering plant phylogeny (31). *Amborella* is a woody shrub, similar to other early lines of angiosperms (the ‘ANITA’ grade) while the water lilies are aquatic herbs, recently recognized to include Hydatellaceae. The new phylogenetic placement of Hydatellaceae near the origin of all flowering plants re-opens debates on the origin of flowering plants (32,33) and also on the possibility that adaptation to aquatic habitats played a role in early angiosperm radiation (34,35). The aquatic reversion has occurred

**ORIGIN**

independently at least 50 times during plant evolution (36), but so far, a model plant system has been lacking to test this hypothesis.

Pinpointing the root of angiosperm phylogeny is crucial for clarifying some of the remaining mysteries concerning flowering-plant origins. It is also critical for understanding the origin and subsequent radiation of gene families that arose close to the origin of flowering plants, and that have since then subsequently diversified spectacularly in major plant groups like the monocots and eudicots. The latter groups include almost all of our major crop plants—eg rice, *Oryza*—and the model plant *Arabidopsis* (see Fig. 2). In summary, through this second defined goal, **ORIGIN** aims to clarify the root of flowering-plant phylogeny by examining large-scale genomic evidence from the nuclear genome.

**Goal 3: Determine the origin of orthologs and paralogous genes on a large evolutionary scale**

Under this third goal, **ORIGIN** will address the following questions:

- What is the origin of the genes that control floral development?

Having solved the genetic boundaries of the early-flowering plant *Trithuria* (Goal 1) and clarified the overall root of the angiosperm tree (Goal 2), **ORIGIN** will facilitate a clearer determination of orthologs vs. paralogs in other plant groups, such as the monocot model plant *Brachypodium*. Ecologically important genes such as the ones that control floral development have been well studied in the current model plant systems (37,38). However, the origin of these genes and therefore the prediction of their function relative to environmental change remain unknown.

Therefore, in this third goal, **ORIGIN** aims to identify the origin of gene families that control flowering in the model plant *Brachypodium* and trace back their evolutionary history relative to *Trithuria* and other angiosperms, identifying the origin of new functionality (i.e., orthologs vs. paralogs) in this model grass and facilitating functional prediction in other cereal crops.

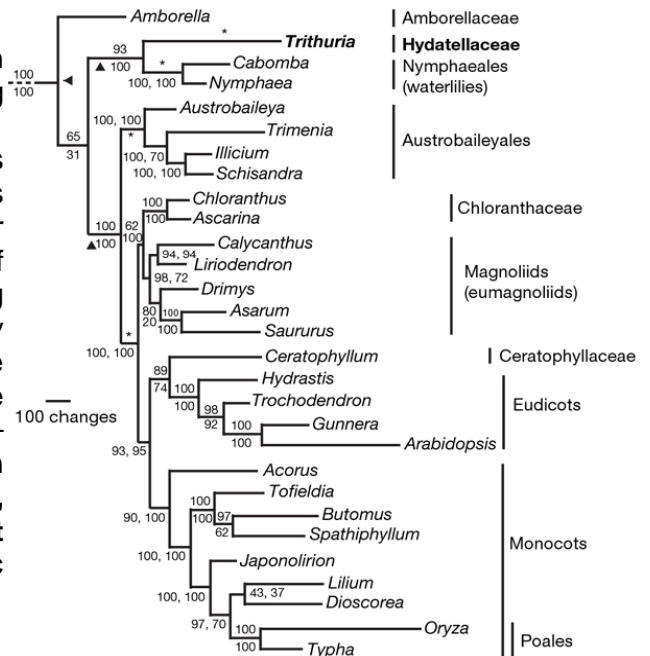


Fig.2. Overview of flowering plant phylogeny based on chloroplast data, showing the early divergence of *Hydatellaceae* and the uncertainty concerning the overall root of the tree (3 equally plausible roots are marked with an arrowhead).