

# Revealing functional and genomic differences between two species of the megadiverse genus *Begonia*

Katie Emelianova<sup>1,2,3</sup>, Mark Hughes<sup>3</sup>, Ian Simpson<sup>2</sup>, Catherine Kidner<sup>2,3</sup>  
<sup>1</sup> Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria  
<sup>2</sup> University of Edinburgh, Kings buildings, Mayfield Road, Edinburgh, EH9 3JZ  
<sup>3</sup> Royal Botanic Gardens Edinburgh, 20a Inverleith Row, EH3 5LR

## *Begonia* is a megadiverse genus

### 1. A widespread genus with a high incidence of endemism

*Begonia*, the 5th largest angiosperm genus, comprises over 2000 species distributed across the tropics except Australia (figure 1). Many species have narrow distributions, and endemism is frequent (figure 2), sometimes to a single hilltop. *Begonia* is also morphologically diverse, particularly in leaf shape and pigmentation patterns, the evolutionary significance of which is not yet fully understood. Small population sizes and poor seed dispersal mechanisms in *Begonia* are thought to increase the role of drift in the evolution of the genus, manifesting in strong intraspecific population structure. Selective forces also shape the evolutionary history of *Begonia*; secondary woodiness, deceit pollination, and tolerance to high altitude provide examples of adaptive traits enabling *Begonia* to survive in new ecological settings. The success of *Begonia* in colonizing much of the tropics and the astounding species and morphological diversity therefore makes it an excellent model to study mechanisms of diversification.



Figure 1. Worldwide distribution of *Begonia*. Colours indicate the major clades of *Begonia*: African (green), Asian (blue), and the two major clades of Neotropical *Begonia* (yellow, NC1; red, NC2). Adapted from Moorlight et al. 2015  
 Figure 2. Histogram of total and endemic *Begonia* species from Southeast Asia, adapted from Hughes, 2008



Figure 3. Photographs of the *Begonia* species used in this study: *B. conchifolia* (A) and *B. plebeja* (B).

### 2. Two closely related but ecologically and morphologically divergent species of *Begonia* provide a good model for identifying signatures of ecological divergence

Our work has focused on two species of South American *Begonia*, *B. conchifolia* (figure 3A) and *B. plebeja* (figure 3B). Despite diverging recently, these two species have striking differences in leaf morphology and habitat; *B. conchifolia* has long-lived fleshy peltate leaves, with a restricted distribution in wet rainforests across southern Mexico and Central America, sometimes establishing itself successfully on open roadbanks. *B. plebeja* has larger, thinner leaves which are deciduous in some populations and often blotched, and is more widespread in seasonally dry, deciduous forests in northern Mexico, tolerating higher levels of insolation than typical for the genus.

## Leveraging functional annotation for a broad overview of species differences

### 1. Comparative methods to identify broad differences

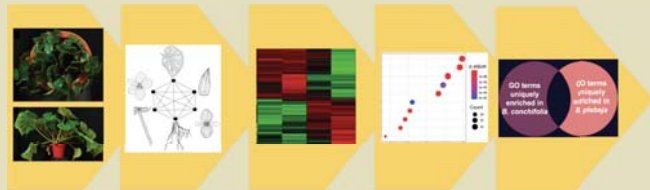


Figure 4. Schematic of the method used to compare transcriptomes of *B. conchifolia* and *B. plebeja*. Libraries of three replicates per tissue per species were mapped back to respective species' reference transcriptomes. Mapped and summarised reads were TMM normalised. Full methods are available in Emelianova and Kidner, 2022 (in Press)

We wanted to identify functional differences between ecologically and morphologically divergent *B. conchifolia* and *B. plebeja* to understand which processes differed the most between the recently diverged species. To this end, we used multi-tissue RNA-seq (figure 4) to find intraspecific differences between tissues, and identified which of these differences were unique to a single species. Species specific differences were categorised at the functional level using GO terms, circumventing problems introduced by similar functions being underpinned by non-homologous genetic components.

### 2. Uniquely enriched GO terms are differentially distributed in *B. conchifolia* and *B. plebeja* tissues

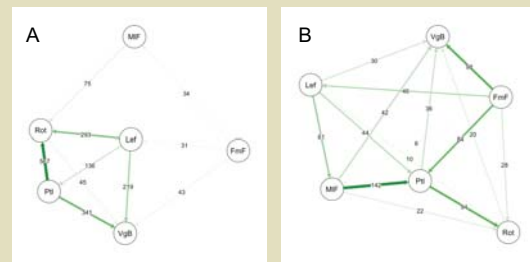


Figure 5. Network graph of the number of genes associated with uniquely enriched GO terms per comparison in *B. conchifolia* (A) and *B. plebeja* (B). Arrows indicate the direction of comparison, and numbers along edges show the number of genes associated with all enriched GO terms in the comparison. Tissue abbreviations are as follows: FrFl = Female flower; Lf = Leaf; MFl = Male flower; Ro = Root; Pt = Petiole; VgB = Vegetative bud.

### 3. Divergent uniquely enriched GO terms between *B. conchifolia* and *B. plebeja*

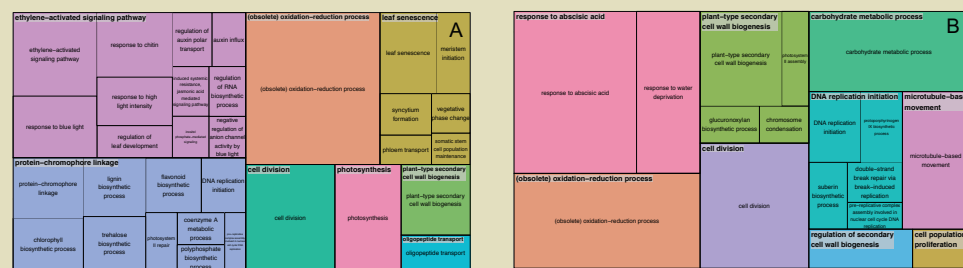


Figure 6. GO terms that are enriched in DE genes for *B. conchifolia* but not for *B. plebeja* (A) and for *B. plebeja* but not *B. conchifolia* (B) across all comparisons. The plots are generated using Revigo (Supek et al., 2011), which visually summarises GO terms, removing redundant terms. Each colour-coded group of rectangles represents a cluster of loosely related terms, the size of each rectangle represents the number of significantly DE genes associated with that term.

We found more *B. conchifolia* loci underlying uniquely enriched GO terms than *B. plebeja* but spread across fewer tissue pairs (figures 5A and 5B). *B. conchifolia* has many uniquely enriched terms in housekeeping functions (figure 6A), with response to abscisic acid and water deprivation among the few environmental response functions. *B. plebeja* has a more active environmental response repertoire; response to light, flavonoid biosynthesis and ethylene, auxin and jasmonic acid mediated pathways featuring among uniquely enriched terms. It is tempting to speculate that the more insolated habitat of *B. plebeja* is reflected in a stronger shade avoidance response than *B. conchifolia*. Our results form the basis for ongoing investigations to understand shade tolerance in *Begonia*, and the role of environmental factors such as light availability in driving diversification in this megadiverse genus.

## Case study shows lineage specific expression and duplication patterns

### 1. Anthocyanins and *Begonia* shade tolerance

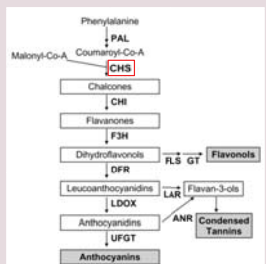


Figure 7. Schematic representing the main enzymes of the anthocyanin biosynthesis pathway. Chalcone synthase is indicated by a red square. Adapted from Takos et al. (2009)

*Begonia* species have diverse growth forms and habitats, one of the most common shared traits is adaptation to low light. Based on our results showing divergent functional processes in light harvesting and photomorphogenesis, we investigated chalcone synthase (CHS), a well-studied member of the anthocyanin biosynthesis pathway (figure 7). Anthocyanins have a key role in protecting plants from photoinhibition after long term and periodic exposure to high UV, thus optimizing photosynthetic capacity. We used our multi-tissue RNA-seq dataset from *B. conchifolia* and *B. plebeja* to identify members of CHS, investigate the evolutionary history of this gene family, and compare gene expression patterns amongst gene copies in both species.

### 2. Lineage specific evolution of CHS in *Begonia*

CHS copies were identified in *B. conchifolia* and *B. plebeja* transcriptomes, and in the *Hillebrandia sanwicensis* genome, the monotypic sister genus of *Begonia*, using the *Arabidopsis thaliana* and *Zea mays* CHS protein sequences. We reconstructed the CHS gene tree (figure 8) finding four orthologous pairs of CHS in *B. conchifolia* and *B. plebeja*, and a putative lineage specific duplication in *B. plebeja*. CHS orthologous groups arising after the divergence of *Begonia* and *Hillebrandia* (figures 8 and 9, groups red, pink and yellow) had evidence of ancestral expression patterns (pink group), putative pseudogenisation in *B. conchifolia* (blue group), and lineage specific duplication in *B. plebeja* (red group). Our findings of dynamic duplication patterns and nascent lineage specific evolution in CHS, a gene family involved in light harvesting and photosynthetic optimization, provides an important basis for further study of CHS and its role in shade tolerance and its importance in *Begonia*'s success in colonising new and challenging environments.

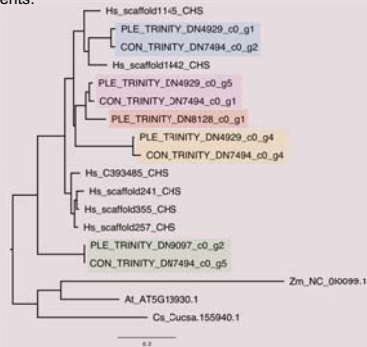


Figure 9. ML tree (RAxML GTR+R substitution model with 1000 bootstrap replicates) of CHS sequences from *B. conchifolia* (CON), *B. plebeja* (PLE), *H. hillebrandia* (H), *C. sativus* (CS), *A. thaliana* (AT) and *Z. mays* (Zm). Pairs of orthologs are highlighted as follows: blue: group 1, pink: group 2, yellow: group 3, green: group 4, red: group 5. For full methods see Emelianova et al. 2021.

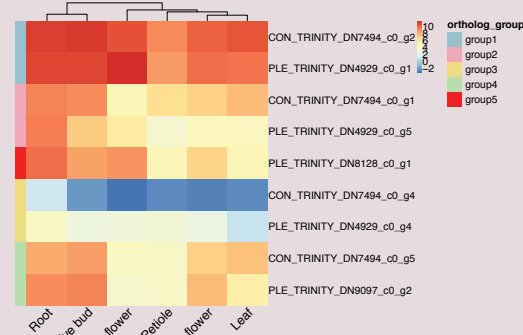


Figure 8. Heatmap of CHS homolog expression in *B. conchifolia* and *B. plebeja*. Library size and composition was accounted for using TMM (trimmed mean of M-values) normalisation, and average FPKM values were calculated for replicates of tissue groups per species. Ortholog groups refer to highlighted homologous CHS pairs in the ML gene tree in figure 9. For full methods see Emelianova et al. 2021.

### 3. Conclusions

- Begonia* is a valuable model for testing evolutionary hypotheses about diversification
- We use modest genomic resources to find divergent signatures of photomorphogenesis and shade tolerance in two species of *Begonia*
- Investigating the photosynthetically relevant gene family chalcone synthase reveals lineage specific evolution since the recent divergence of *B. conchifolia* and *B. plebeja*
- Further work will integrate knowledge about species fitness under different light intensities, and the role of lineage specific gene duplications in determining the range of insolation tolerance in *Begonia*