Mining Begonia genomic resources to find candidate genes for ecological diversification

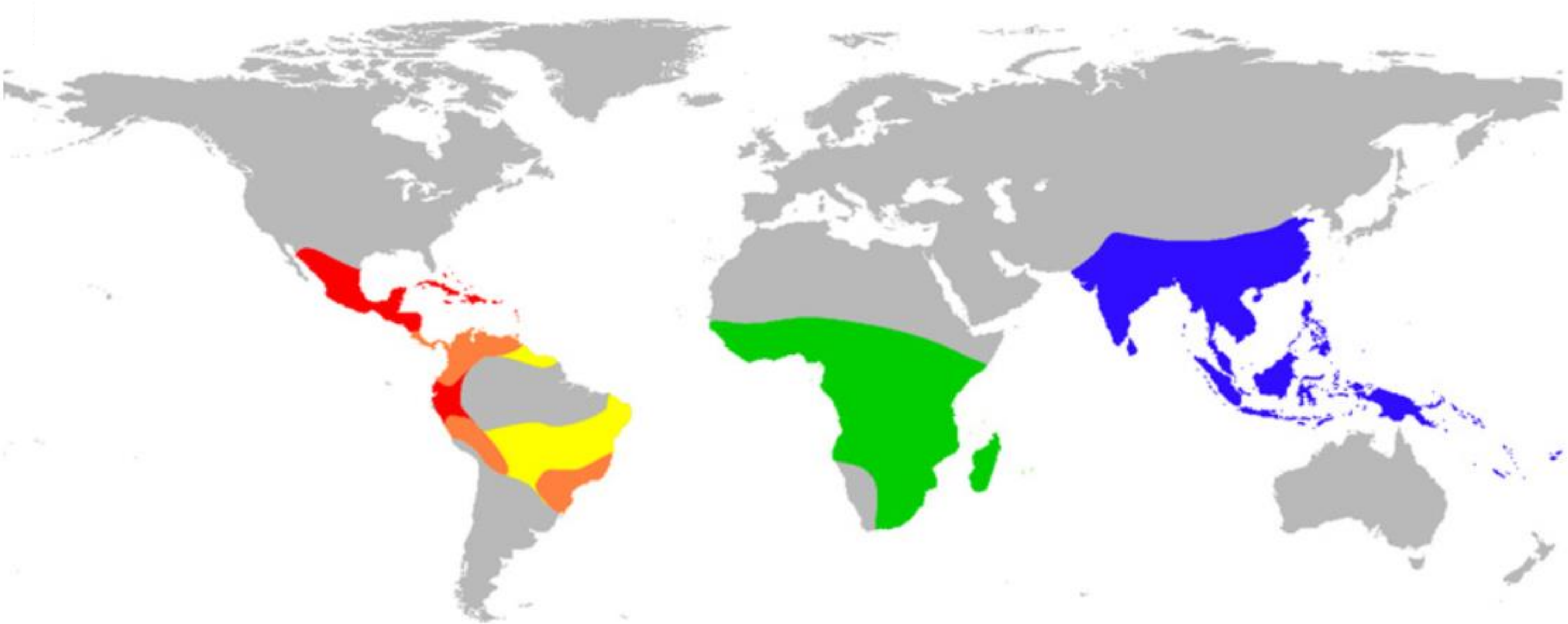
Katie Emelianova

ROYAL
BOTANIC
GARDEN
EDINBURGH

BBSRC
bioscience for the future

THE UNIVERSITY OF EDINBURGH

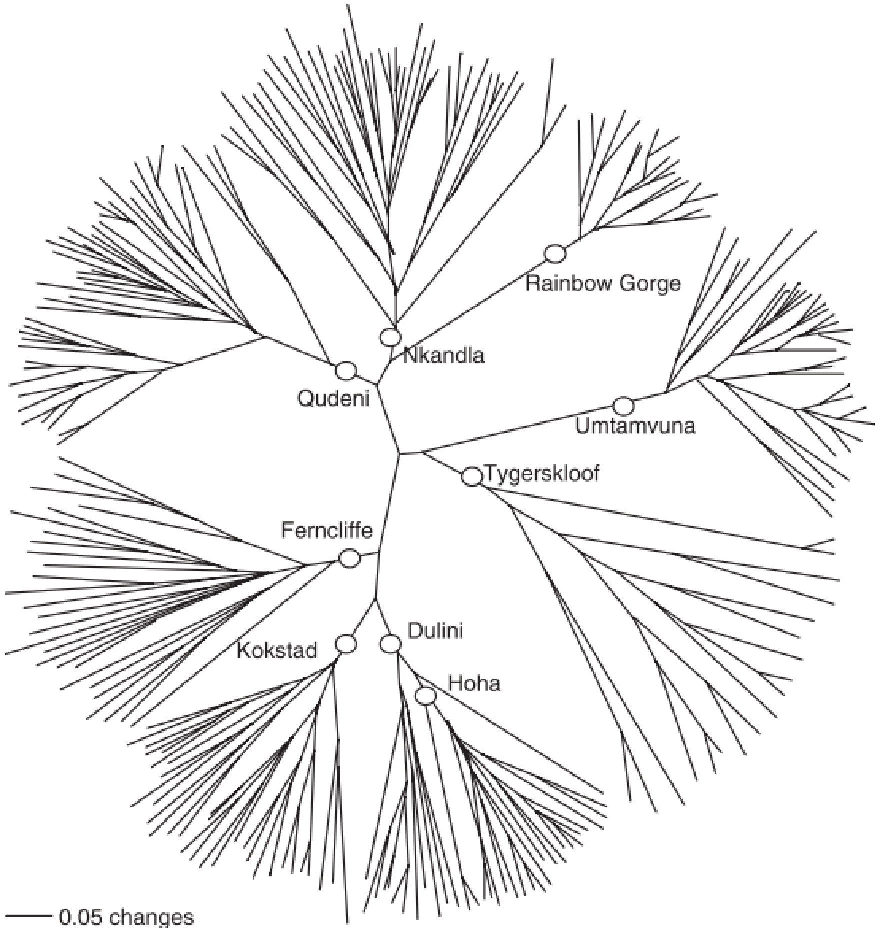
Begonia: the 6th most speciose genus of Angiosperms



Begonia phenotypic diversity

Begonia ecological diversity

Divergence within Begonia populations



Hughes and Hollingsworth, 2008

What has driven phenotypic and ecological diversification in Begonia?

Q: Why is Begonia so diverse?

What ecological pressures have shaped Begonia?

What factors have driven Begonia diversification?

Project Summary:

- A large scale study using reduced representation sequencing data
- Aims to uncover potential candidate genes important in Begonia evolution

Study Species

B. conchifolia

Inhabits wet rainforests
from Costa Rica to
Panama

B. plebeja

Inhabits dry forests
in Mexico and Central
America

Genomic resources

Draft *B. conchifolia* genome: sequenced using Illumina and PacBio platforms

Tissue transcriptomes from *B. conchifolia* and *B. plebeja* (6 tissues, 3 reps)

	***B. conchifolia* Reference Transcriptome**	***B. plebeja* Reference Transcriptome**	***B. conchifolia* Genome**
Number of Sequences	20562	19798	21856
Smallest Sequence	301	301	180
Largest Sequence	15561	15517	14567
Number of Bases	35114785	30477453	32565882
Mean Length	1707.75	1539.42	1489
Number < 200 nt	0	0	98
Number > 1k nt	14052	12101	13415
Number with ORF	17704	16434	18418
N50	2159	2014	1941
N90	910	785	810

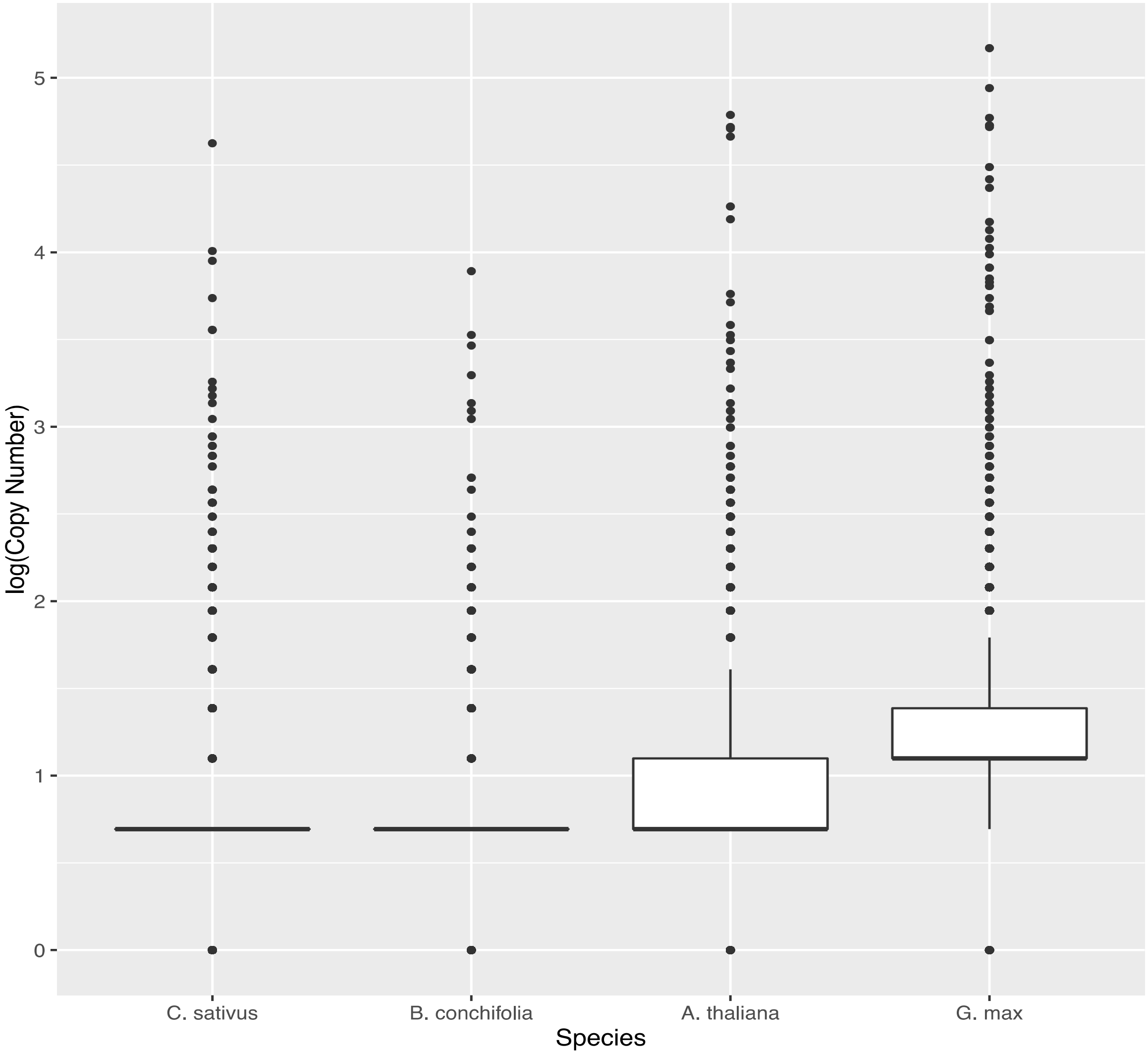
Strategy for mining genomic data

- Cluster *B. conchifolia* genome predicted transcripts into **gene families**

- Align and estimate **pairwise sequence divergence** between gene family members

- Use RNA-seq reads from *B. conchifolia* and *B. plebeja* tissues to compare **divergence of expression profiles** between gene family members

Results: comparing gene family sizes

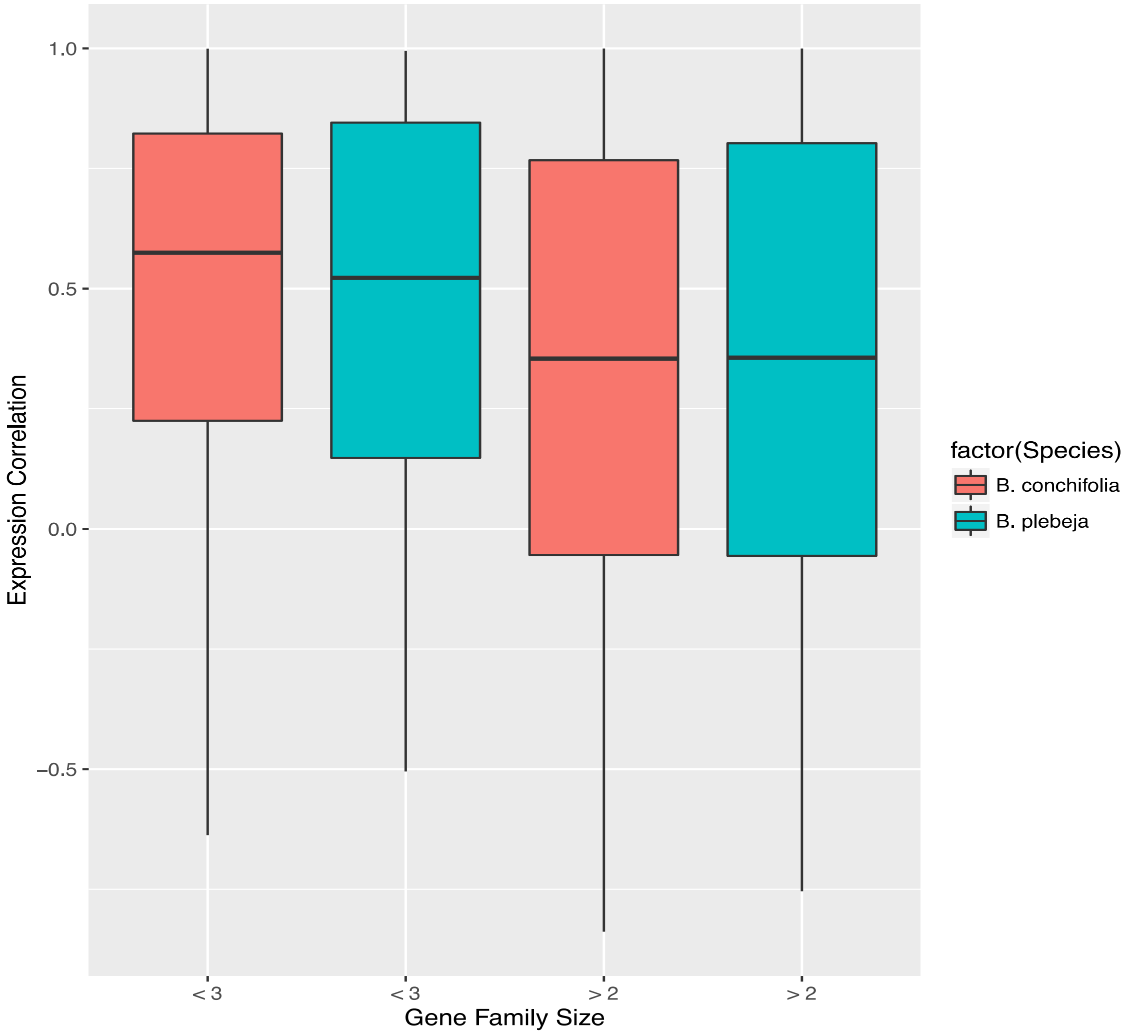


- *Begonia* has fewer expanded gene families than *A. thaliana* and *G. max*
- *C. sativus* similar to *Begonia*, unexpanded

Results: Do more duplicated gene families have greater divergence in expression?

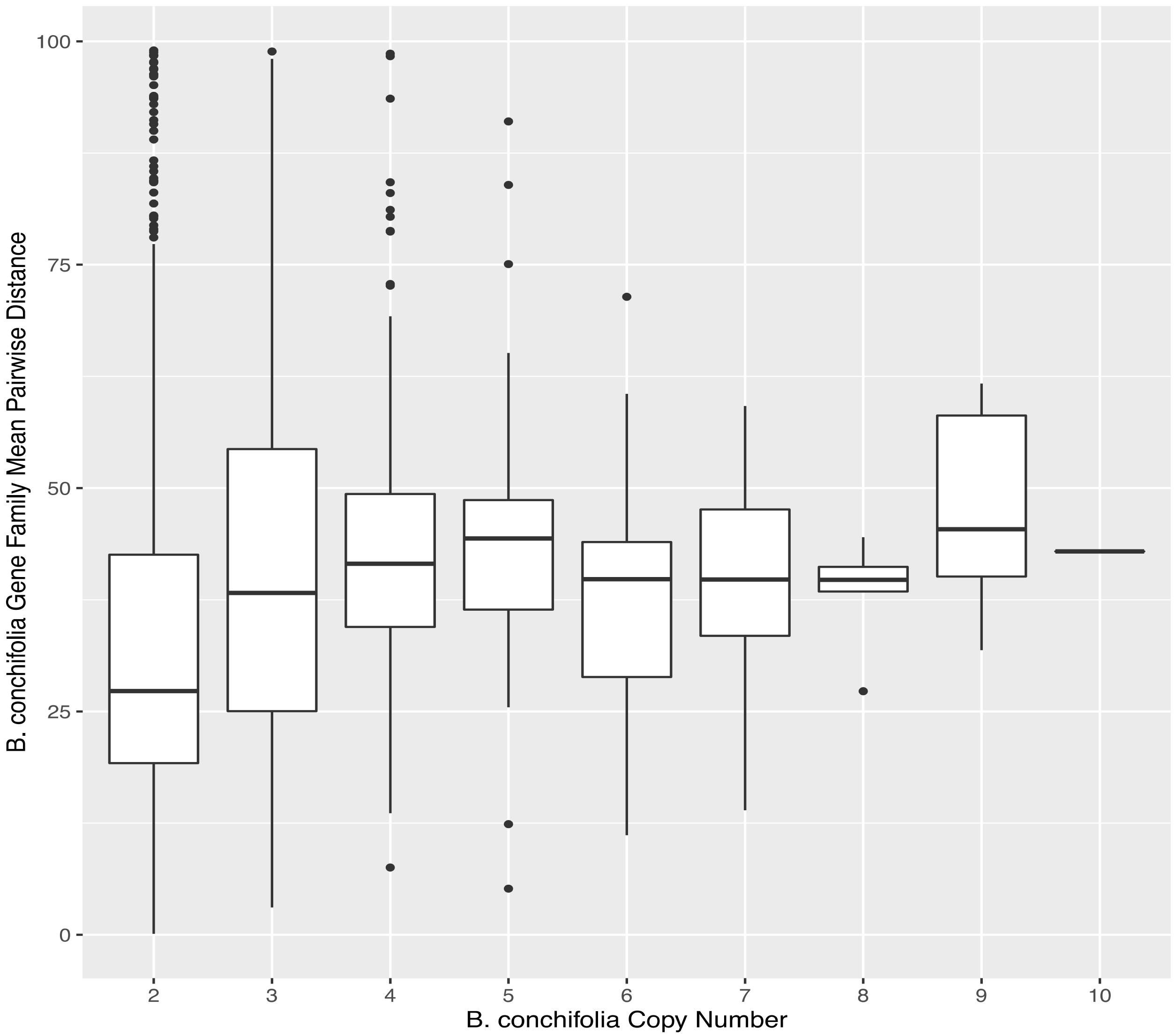
- Greater expression correlation (more similar) in less expanded gene families

- Trend replicated in both species



Results: Do more duplicated gene families have greater sequence divergence?

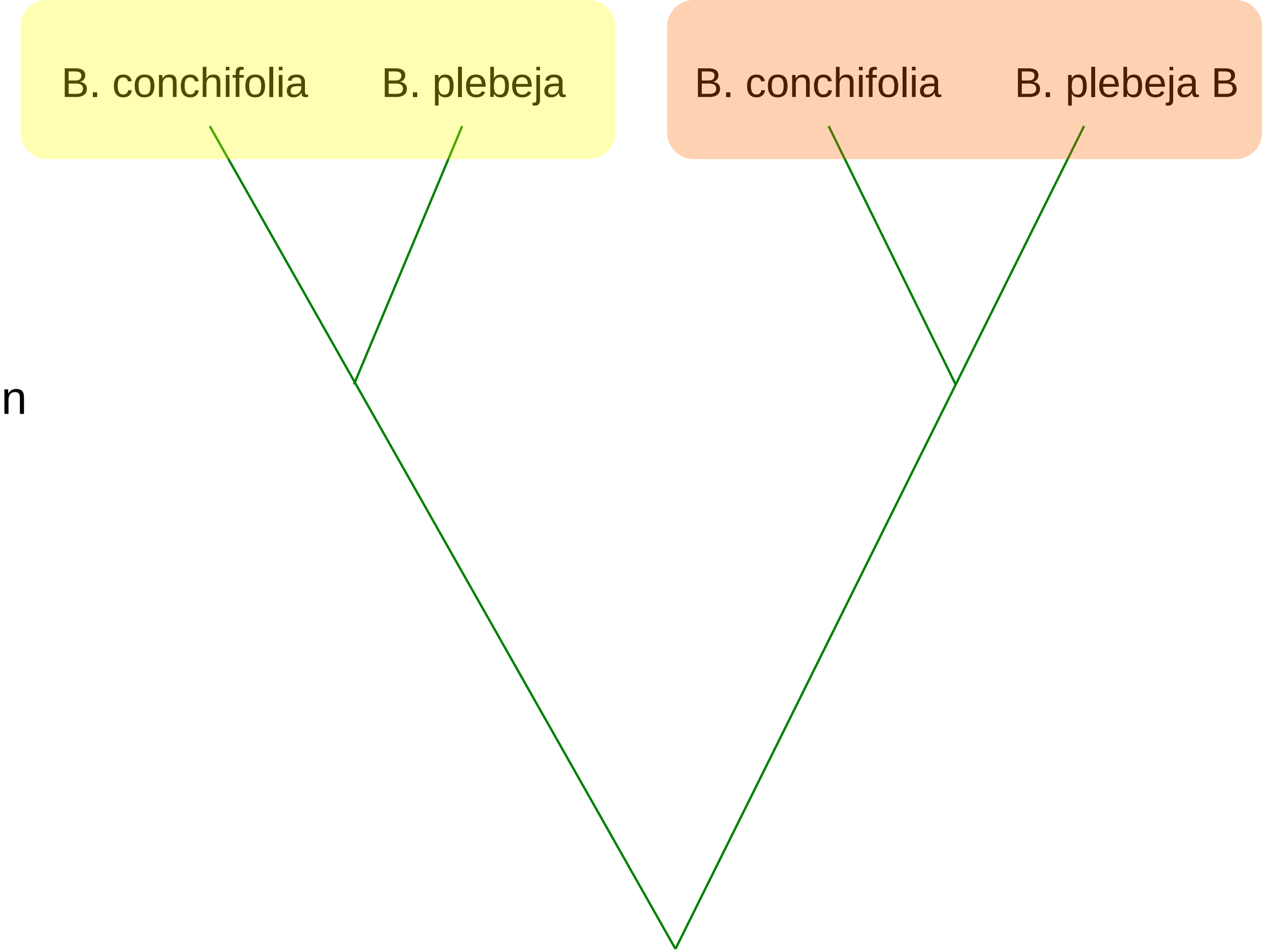
- Gene families which are larger do tend to be more divergent
- Evident in smaller multi-copy gene families
- Does not increase linearly across larger multi-gene families



Comparison of expression correlation

- Within each gene family, get pairwise comparisons of expression across tissues using Pearson correlation coefficient

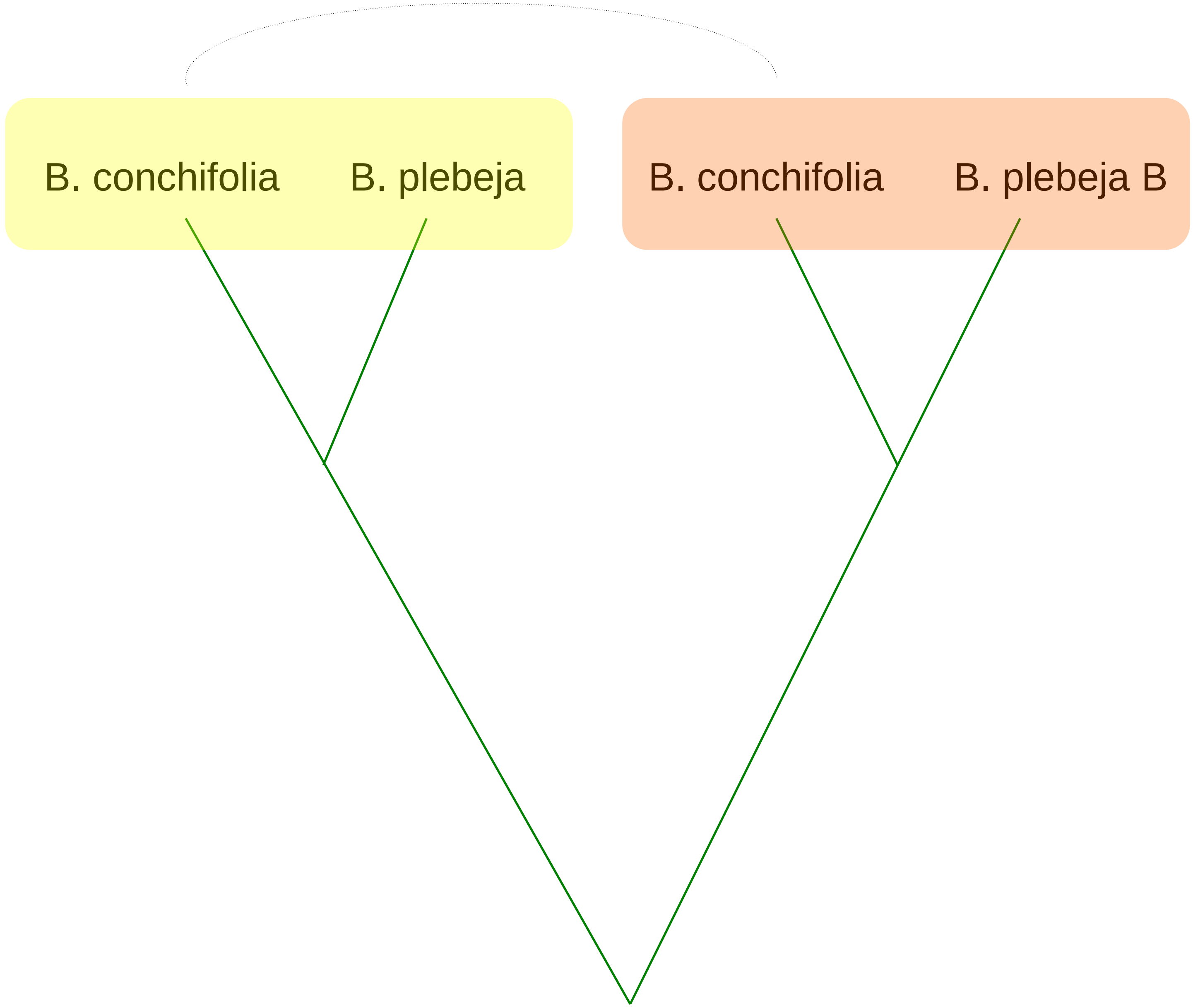
- Compared **within** species and **between** species



B. conchifolia B. plebeja

B. conchifolia B. plebeja B

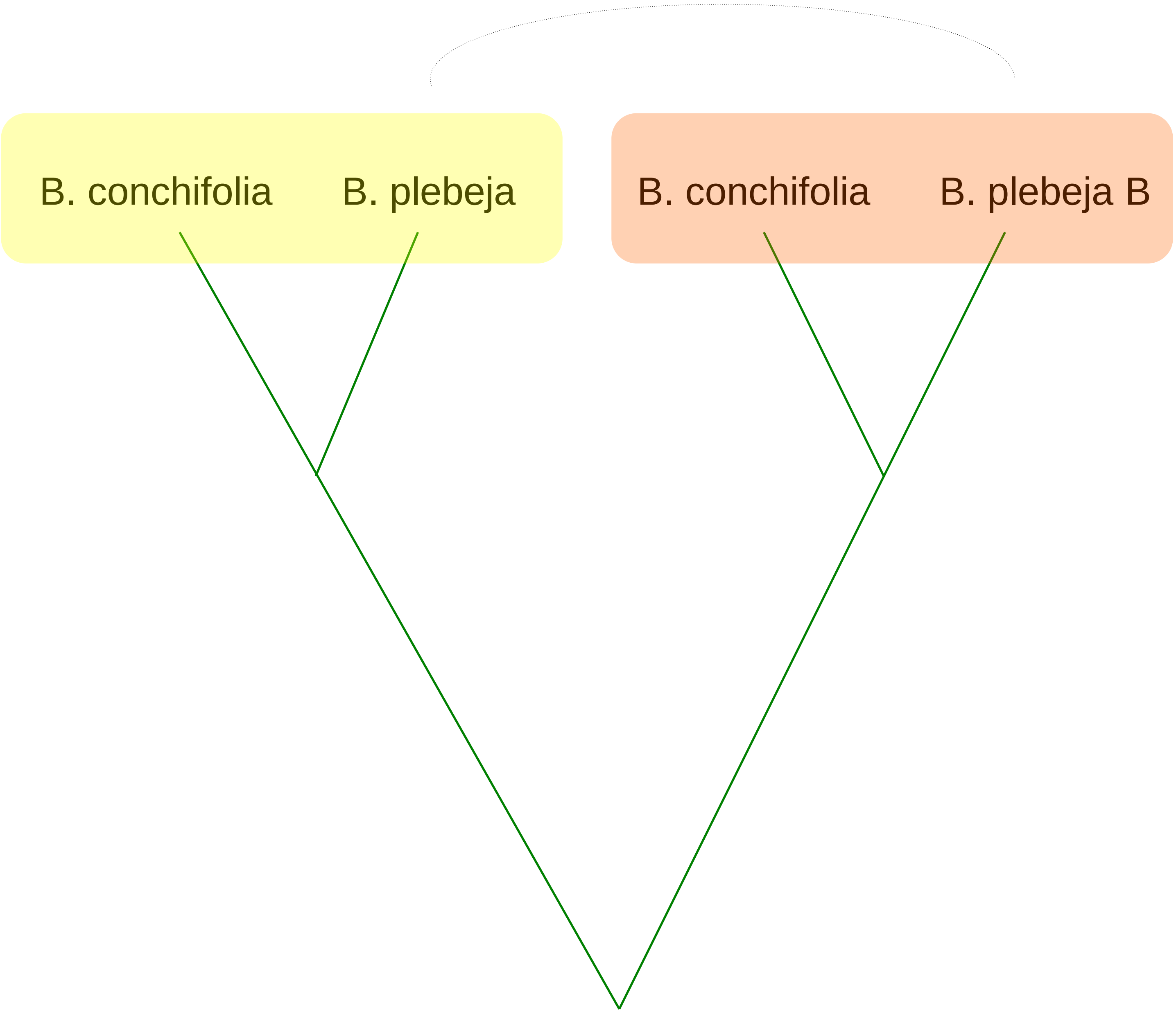
Between species pairs



B. conchifolia B. plebeja

B. conchifolia B. plebeja B

Between species pairs



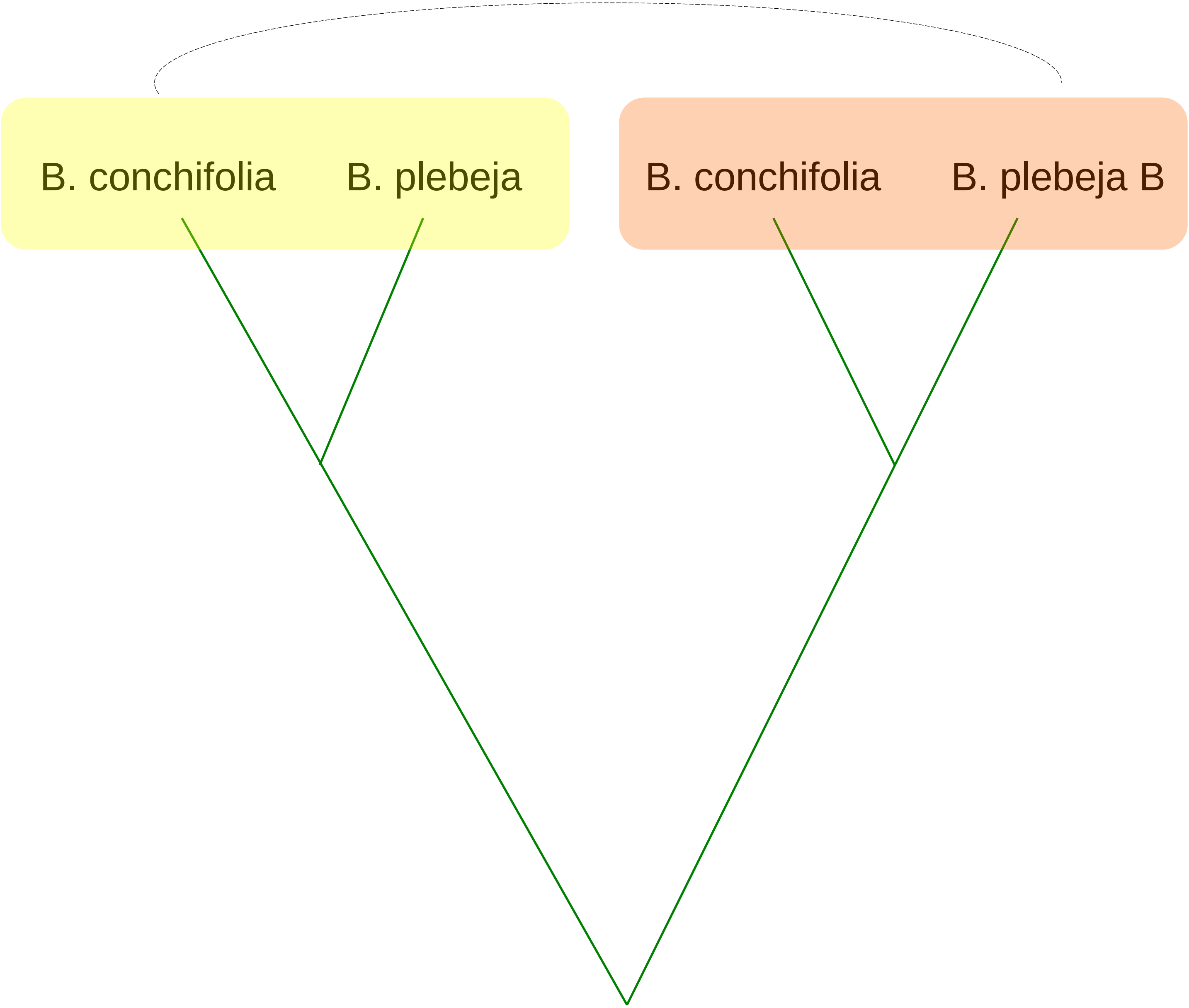
B. conchifolia

B. plebeja

B. conchifolia

B. plebeja B

Between species



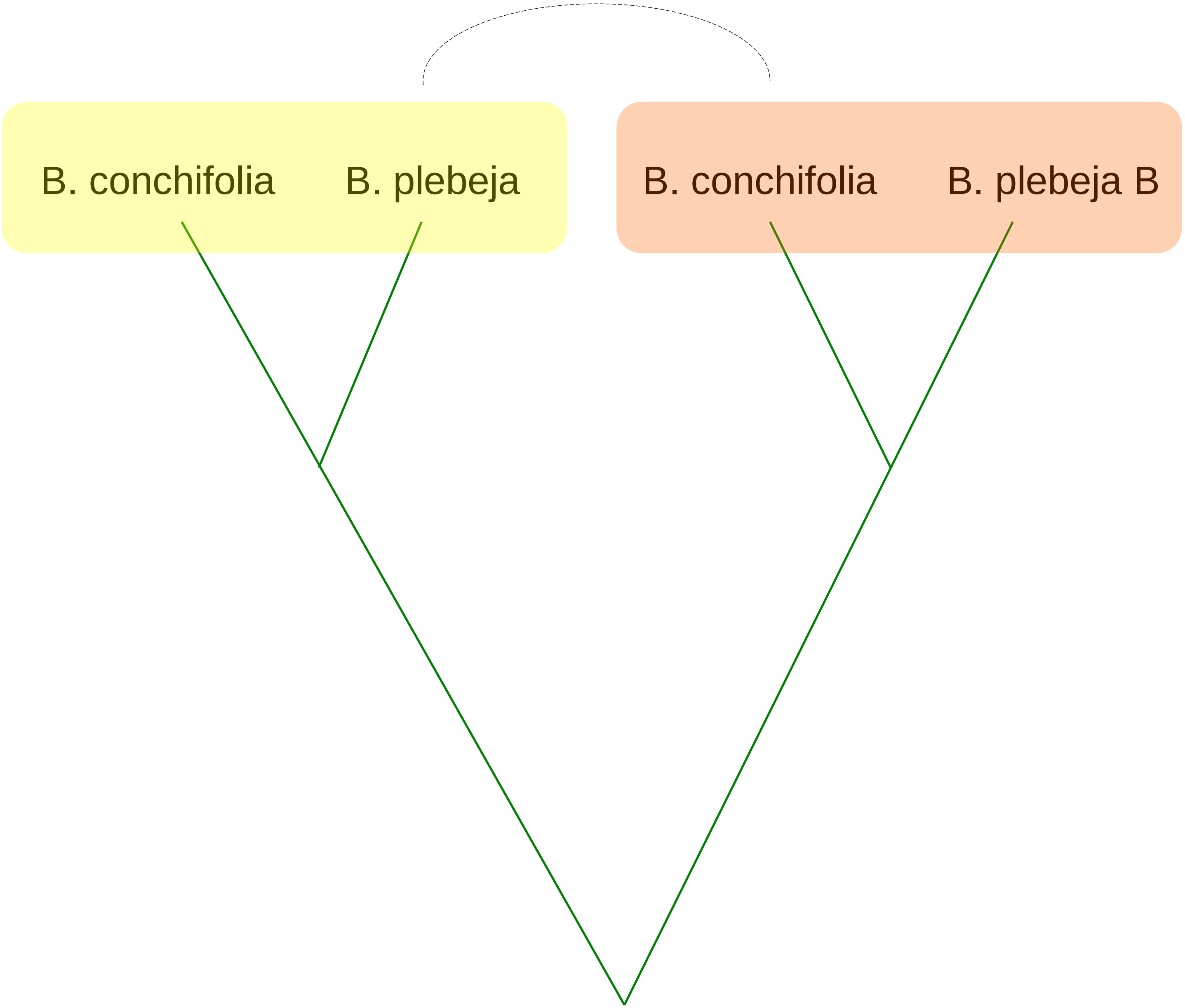
B. conchifolia

B. plebeja

B. conchifolia

B. plebeja B

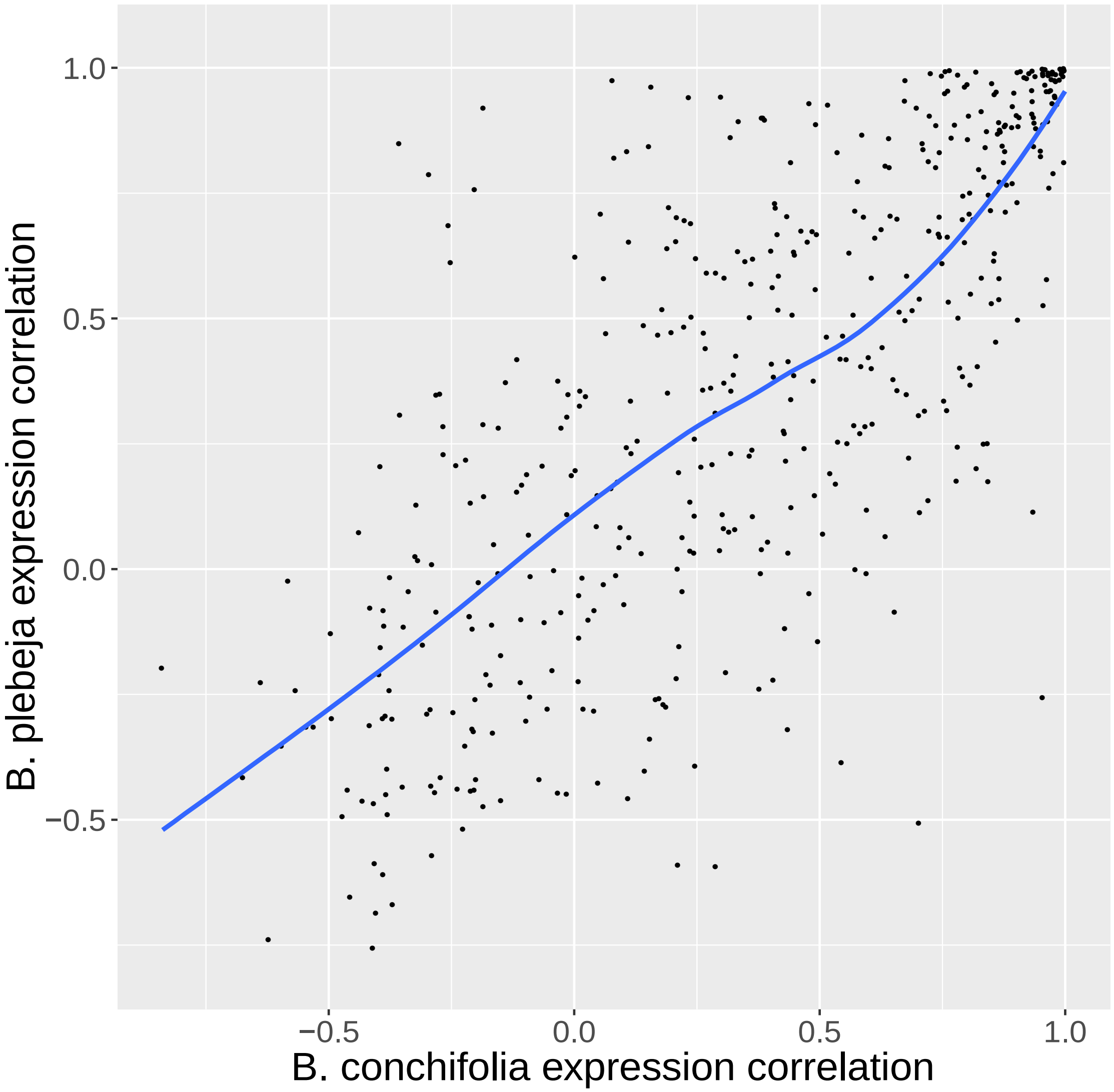
Between species



B. conchifolia B. plebeja

B. conchifolia B. plebeja B

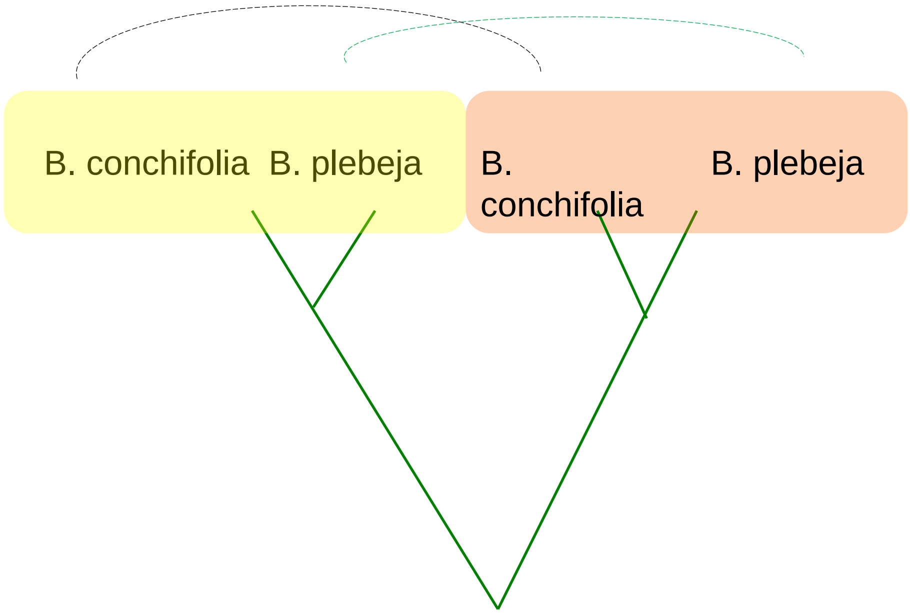
Results - Correlation of expression patterns between study species



- Expression correlation between B. conchifolia gene pairs and B. plebeja gene pairs has linear trend

- Outliers present with bias towards one species

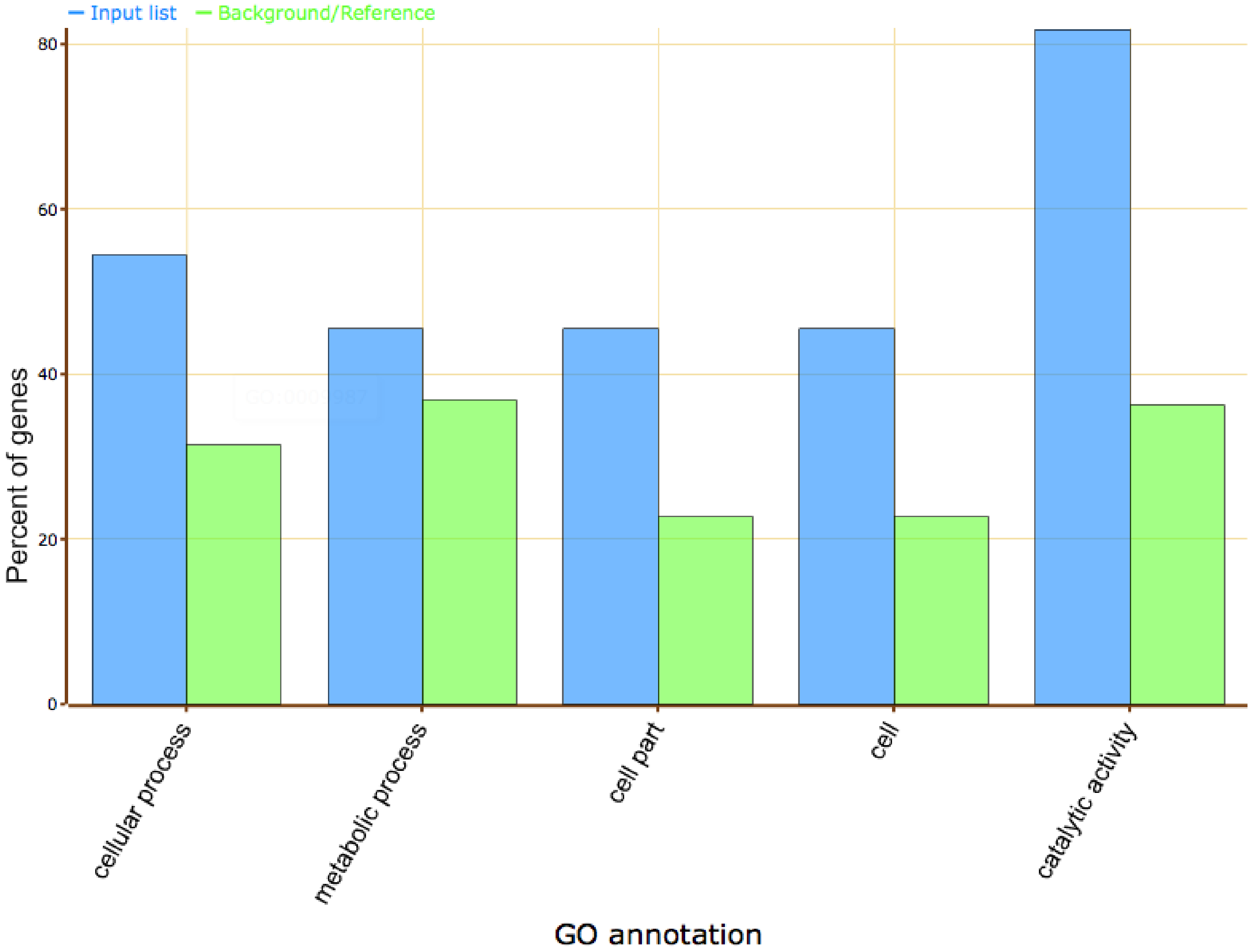
Results – GO terms enriched in between species pairs comparison



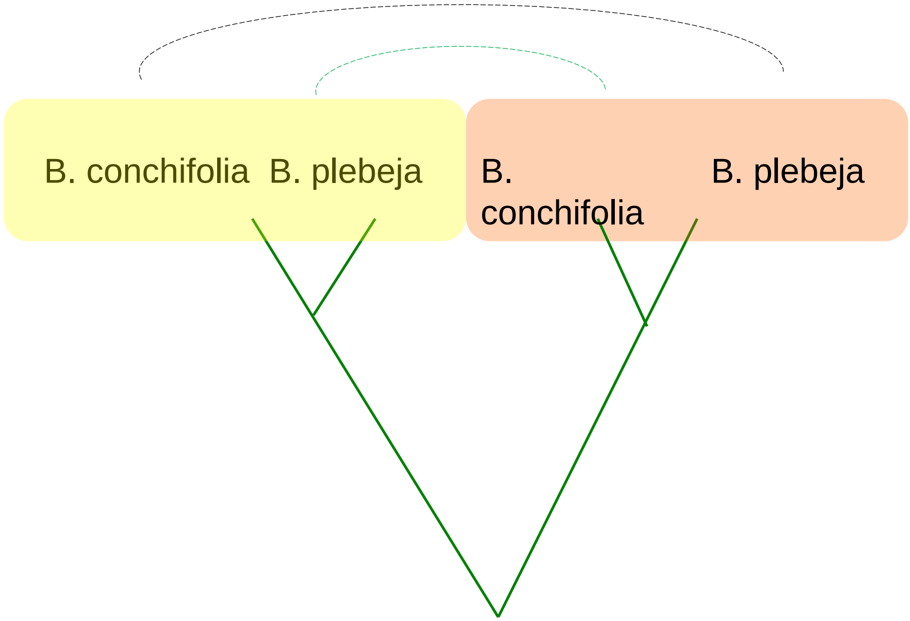
B. conchifolia B. plebeja

B. conchifolia

B. plebeja

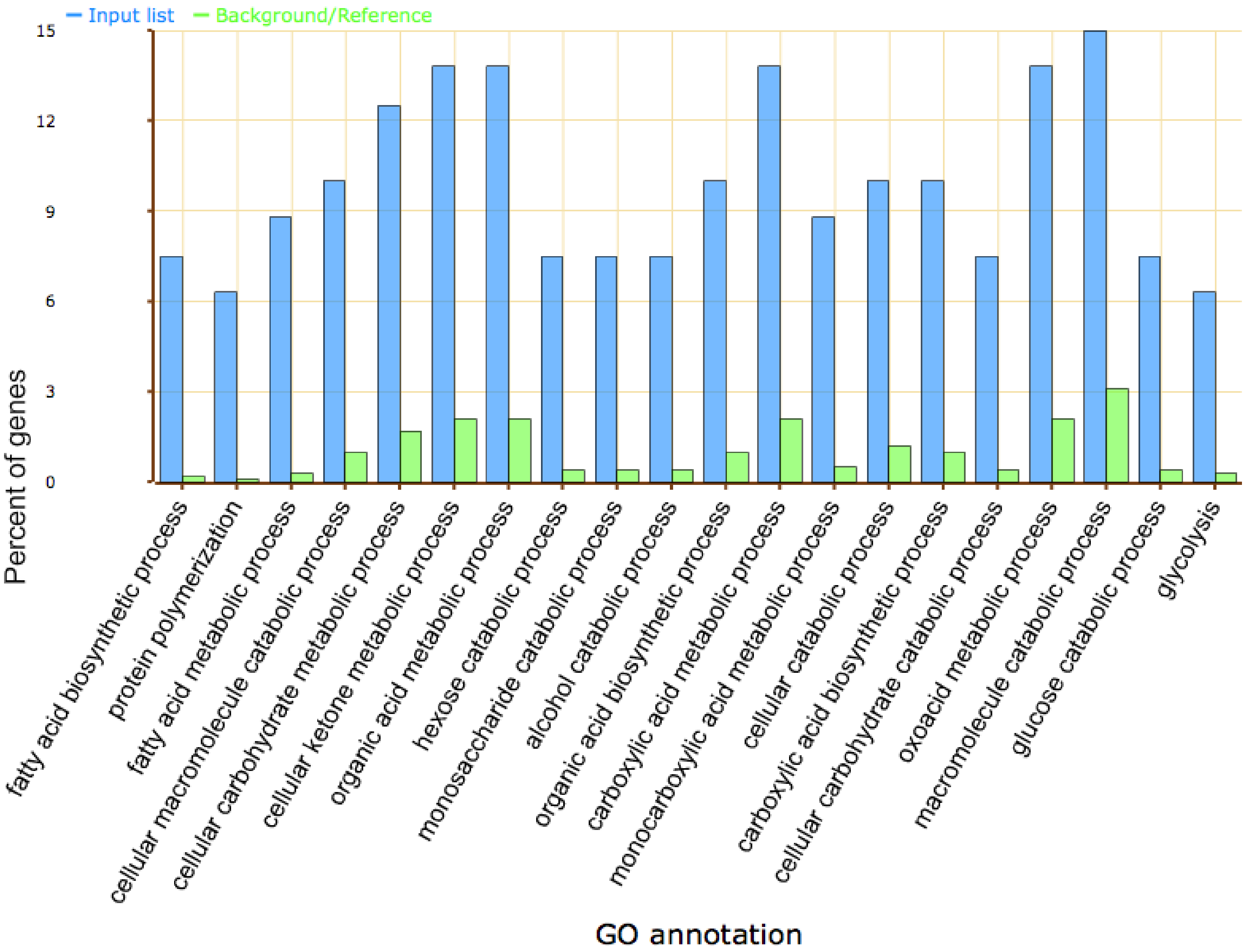


Results – GO terms enriched in between species comparison



B. conchifolia B. plebeja

B. conchifolia B. plebeja



Results – Genes corresponding to enriched GO terms

3-Ketoacyl-CoA Synthase 12-Related – Wax synthesis pathway

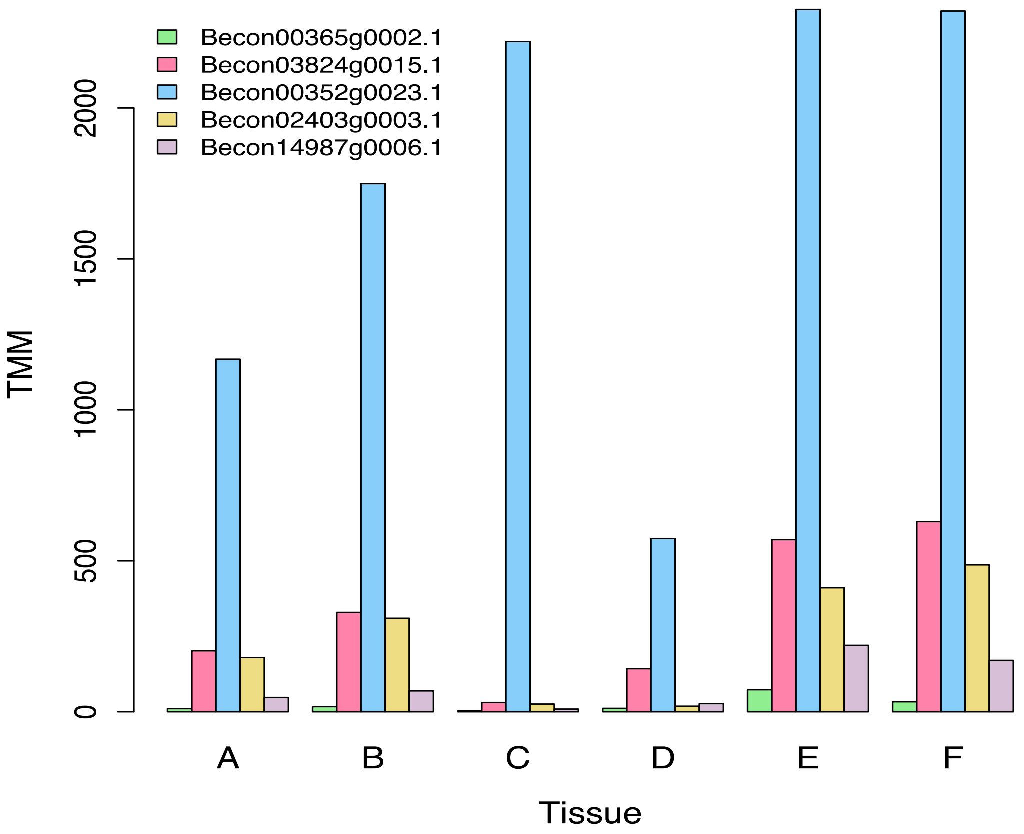
Naringenin-Chalcone Synthase – Anthocyanin biosynthetic pathway

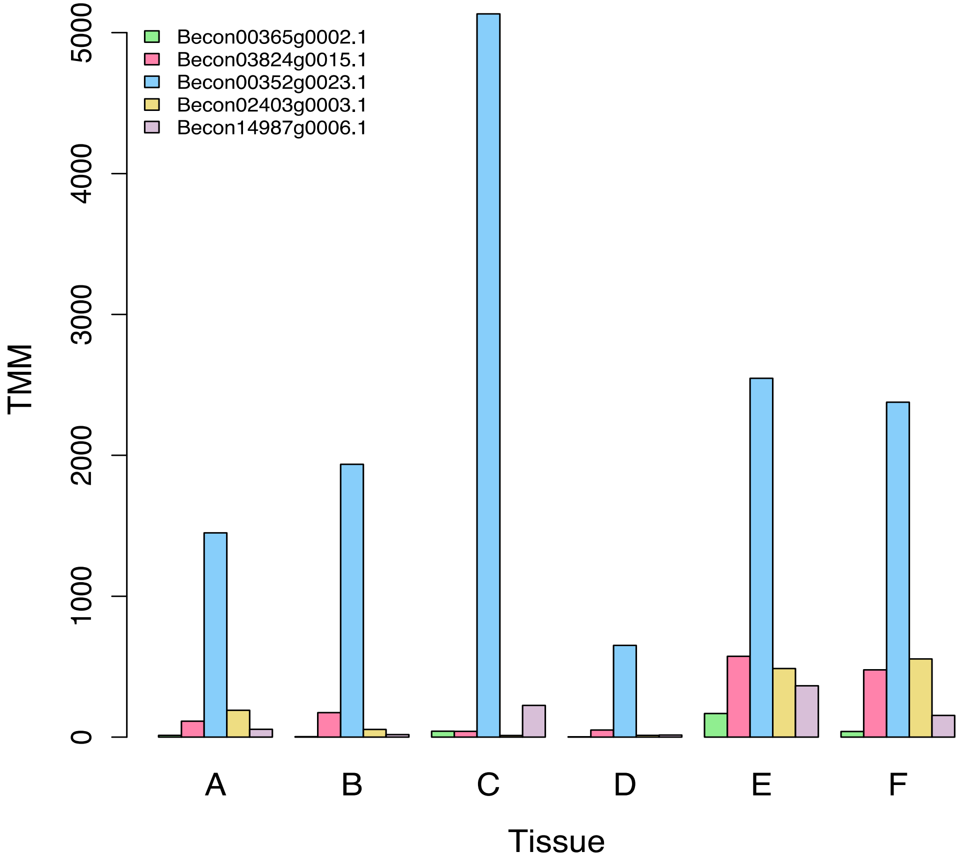
Tubulin – Cell structure

Subtilisin – Protein hydrolysis

Results - CHS expression across tissues

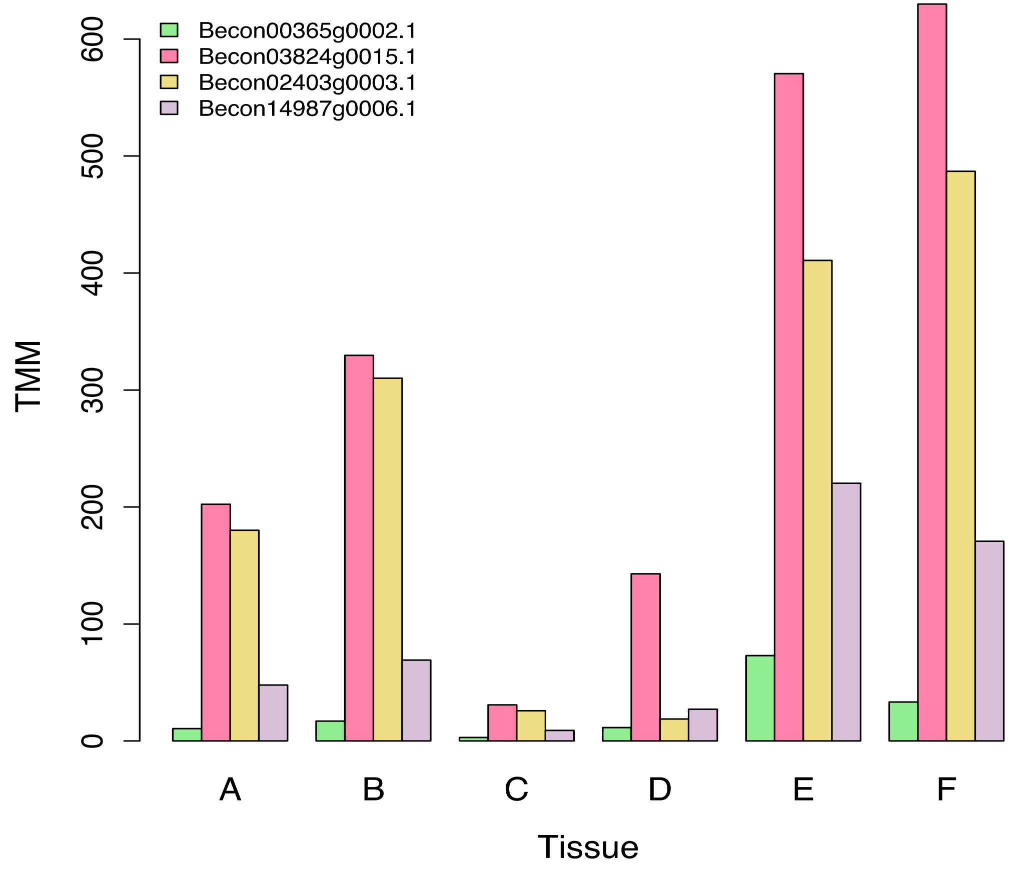
- **A** - Female flower
- **B** - Leaf
- **C** - Male flower
- **D** - Petiole
- **E** - Root
- **F** - Vegetative Bud

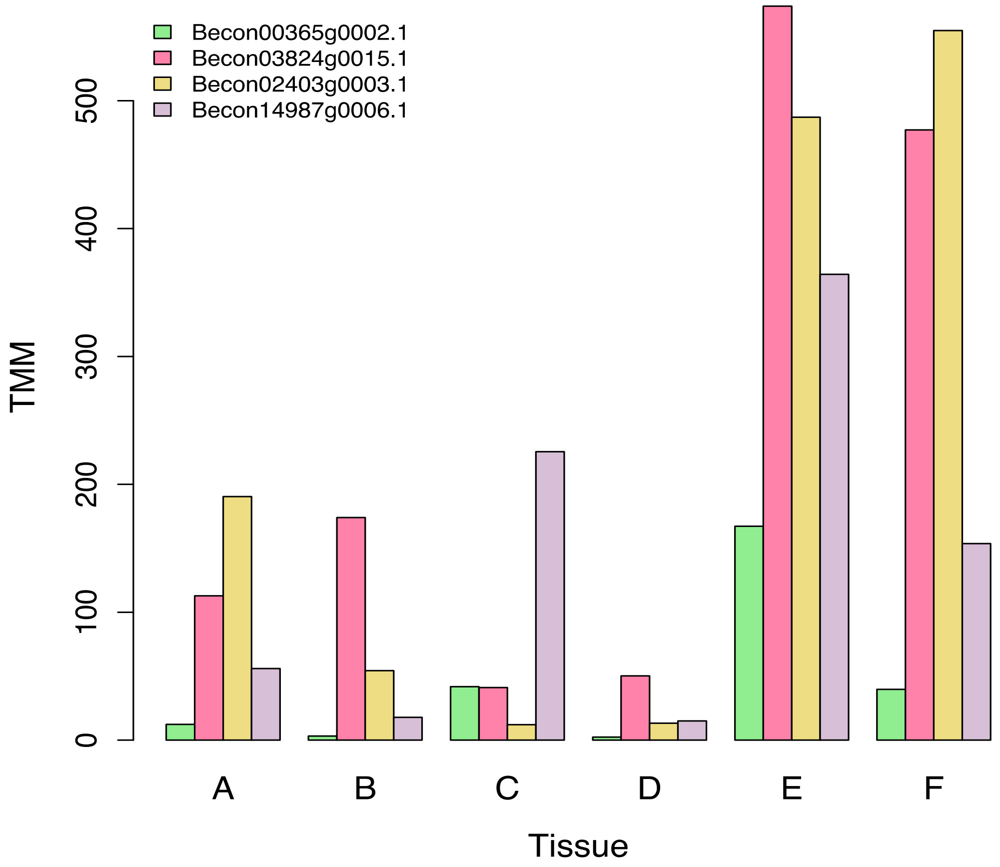




Results - CHS expression across tissues

A - Female flower
B - Leaf
C - Male flower
D - Petiole
E - Root
F - Vegetative Bud





Conclusions

Begonia gene families do not show extensive expansion

Greater expression and sequence diversity is seen with increasing gene family copy number

While linear trend is seen in within species comparison of expression correlation, many genes have expression conservation biased towards one species

Wax cuticle and anthocyanin biosynthesis genes are highly divergent in expression patterns

Many thanks to:

Catherine Kidner

Ian Simpson

Mark Hughes

Michelle Hollingsworth

Laura Forrest

Ruth Hollands

ROYAL
BOTANIC
GARDEN
EDINBURGH

BBSRC
bioscience for the future

THE UNIVERSITY OF EDINBURGH