Additional File 6

Phylogenetic trees of the fast evolving genes and of their putative orthologues and paralogues .

We used phylogenetic tree analysis to test the possibility that the rates of evolution observed in genes evolving fast along a branch of the phylogeny and in the candidate genes for positive selection were instead the result of a mis-identification of the putative orthologues.

For each candidate gene detected at a false discovery rate threshold of 0.20 (see Table XX), we searched the homologues (and possibly paralogues) in *A. thaliana* and in the two *Cardamine* datasets. BLAST searches identified putative paralogous of six of such genes in at least one among *A. thaliana*, *C. resedifolia* and *C. impatiens*. For these genes we proceeded with the phylogenetic analysis by first searching for the homologues of the *A. thaliana* gene (and, if present, of its paralogues) in other eudicots using BLASTx. Then, sequences of BLAST hits with a reasonable e-value (< 10^{-50}) were aligned using MUSCLE [1], and the subsequent alignment was used to build an unrooted neighbor joining phylogenetic tree using PAUP* [2] (bootstrap values obtained from 1,000 replications). Trees were drawn using Geneious 5.4 [3].

Phylogenetic trees were obtained for the candidate genes AT1G49750 (panel A), AT1G54040 (panel B), AT1G71040 (panel C), AT2G31610 (panel D), AT4G17520 (panel E), and AT5G62680 (panel F).
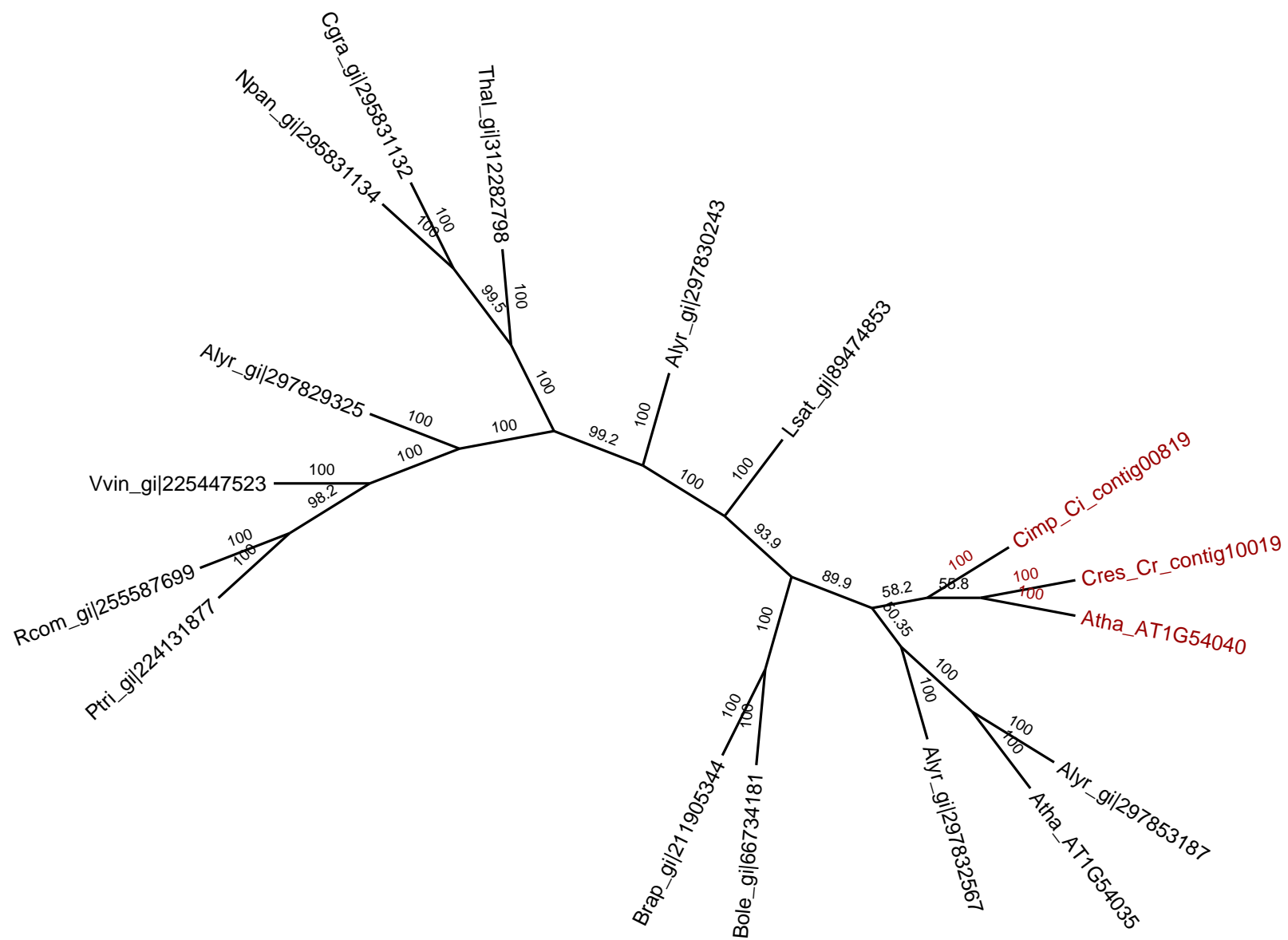
Sequences of the *A. thaliana, C. impatiens* and *C. resedifolia* orthologues used in our analyses are highlighted in red. Sequences were named using a four letter code identifying the species (see legend below), and followed by either the AGI name of the gene (for *A. thaliana*), the name of the contig (for *C. impatiens and C. resedifolia*), or the unique NCBI gene identifier.

Alyr = *Arabidopsis lyrata;* ***Atha*** = *Arabidopsis thaliana;* ***Bole*** = *Brassica oleracea;* ***Brap*** = *Brassica rapa;* ***Cgra*** = *Capsella grandiflora;* ***Cimp*** = *Cardamine impatiens;* ***Cres*** = *Cardamine resedifolia;* ***Ghir*** = *Gossypium hirsutum;* ***Gmax*** = *Glycine max;* ***Lsat*** = *Lepidium sativum;* ***Mtru*** = *Medicago truncatula;* ***Npan*** = *Neslia paniculata;* ***Ptri*** = *Populus trichocarpa;* ***Rcom*** = *Ricinus communis;* ***Stub*** = *Solanum tuberosum;* ***Thal*** = *Thellungiella halophila;* ***Vvin*** = *Vitis vinifera.*

References
1.Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** Nucleic Acids Res 2004, **32**:1792-1797.
2.Swofford DL: **PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.** Sinauer Associates, Sunderland, Massachusetts; 2003.
3.Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. **Geneious v5.4**, Available from http://www.geneious.com

A

B

C

D

Atha_AT5G35530
Alyr_gi297805097
Alyr_gi297816691
Atha_AT3G53870
Cimp_Ci_contig02621
Cres_Cr_contig06562
Alyr_gi297822906
Atha_AT2G31610
Cres_Cr_contig08842
Vvin_gi225444781
Rcom_gi255546002
Gmax_gbBT094691
Gmax_gbBT096639
Mtru_gbBT051429
Mtru_gbBT051240
Stub_gbDQ241842
Rcom_gi255550144
Ptri_gbEF147966
Ptri_gi224088765
Ptri_gi224141684

100
100
100
70.6
92.7
99.9
85.6
100
100
100
100
53
53.1
64.5
65.2
100
67.7
100
99.4
88.9
100
99.4
100
100
99.9
100
91.8
100
100

80.0

E

Cimp_Ci_contig05159

Cres_Cr_contig10404

100

100

100

Alyr_gi297804429

100

100

100

69.2

Atha_AT4G17520

100

Cres_Cr_contig04013

100

100

100

Vvin_gi225468156

Bole_gi89257560

40.0

F

Atha_AT3G47960

Alyr_gi297819405

Thal_gi312283408

Cimp_Ci_contig00876

Cres_Cr_contig03900

Cres_Cr_contig05374

Cimp_Ci_contig00525

Cgra_gi295831274

Atha_AT5G62680

100

100

100

100

100

100

100

99

100

84.6

69.8

100

100

100

100

100

100

Thal_gi312282216

100

100

58.9

100

Gmax_gi11933399

100

100

100

100

100

Gmax_gi11933413

Vvin_gi225456642

Ptri_gi224119139

Rcom_gi255540888

70.0