

Supplementary Materials: Identification and Evolution of Functional Alleles of the Previously Described Pollen Specific Myrosinase Pseudogene *AtTGG6* in *Arabidopsis thaliana*

Lili Fu, Bingying Han, Deguan Tan, Meng Wang, Mei Ding and Jiaming Zhang

```

1 ATGGCAATTCACAAAAGCTCACTACTCTTTAGCCATCATTTGTCGTTCTTTTGGCCGTTTCGAGGAGCCAAAAGTATGCAATCCAGAATGCAAGGCCAAAAG Tsu
1 ATGGCAATTCACAAAAGCTCACTACTCTTTAGCCATCATTTGTCGTTCTTTTGGCCGTTTCGAGGAGCCAAAAGTATGCAATCCAGAATGCAAGGCCAAAAG Col
101 AACCCCTTCAACTGCGACAAAACCTCTGCATTCAACCGAAATGGATTTCCCAACAA--TTTACTTTTGGTGAGCTACCTTGCATACCAGgtttacata Tsu
101 AACCCCTTCAACTGCGACAAAACCTCTGCATTCAACCGAAATGGATTTCCCAACAAATTTTACTTTTGGTGAGCTACCTTGCATACCAGgtttacata Col
199 tatcttcgcatgatagcttttaggggattgatttctagctataaacatbtccatataactttctcttacaatatccaaagctttgttctatcagGTTGA Tsu
201 tatcttcgcatgatagatttaggggattgatttctagctataaacatbtccatataactttctcttacaatatccaaagctttgttctatcagGTTGA Col
299 AGGTGCAGCAGATAGAGCACTTAATGGATGGGACTATTTCACATAGATAATCCAGgtcataaacatcatttaatatattatttataacatgcactt Tsu
301 AGGTGCAGCAGATAGAGCACTTAATGGATGGGACTATTTCACATAGATAATCCAGgtcataaacatcatttaatatattatttatta-catgcactt Col
399 catgtatgaaaaagataaataaagaatataatattcattcattcattcctggattggttatcacagAAGAGTTTCAGATCGCATATAGGAGACCTTGCT Tsu
400 catgtatgtaaaagataaataaagaatataatattcattcattcattcctggattggttatcacagAAGAGTTTCAGATCGCATATAGGAGACCTTGCT Col
499 TGTAAATTCGTATGATCTTTATAAGgtgaaaggctatcaatgacgaatgtgtacataggtgtttgttcaatttttcttagtttaatttaatttaaga Tsu
500 TGTAAATTCGTATGATCTTTATAAGgtgaaaggctatcaatgacgaatgtgtacataggtgtttgttcaatttttcttagtttaatttaatttaaga Col
599 tttttcatataaaccttgaaacagGATGATGTTAAACTACTAAAAGAAATGAACGTTCAAGCATAACCGATTCTCAATAGCATGGTCAAGGGCTTACCAAgT Tsu
600 tttttcatataaaccttgaaacagGATGATGTTAAACTACTAAAAGAAATGAACGTTCAAGCATAACCGATTCTCAATAGCATGGTCAAGGGCTTACCAAgT Col
699 aaatggatcttcatgttttcatcctaataataaataagaattcgtaaaaatttcaactccatcactaaacttatttggatataatgctaaatttagAGG Tsu
700 aaatggatcttcatgttttcatcctaataataaataagaattcgtaaaaatttcaactccatcactaaacttatttggatataatgctaaatttagAGG Col
799 GCAGACTAATGGAGGAGTCGACGAGAAATGGGATCACATACTACAACAACTTCATTAACGAGTTGAAAGCAAATGgtaaattcctaattcaagtgcttaa Tsu
800 GCAGACTAATGGAGGAGTCGACGAGAAATGGGATCACATACTACAACAACTTCATTAACGAGTTGAAAGCAAATGgtaaattcctaattcaagtgcttaa Col
899 ttaatcatattttaaactcgttttagtatgtttgtgatataaaacatttaacttcatgaatacagGCATAGAACCATTGTGACTATAATTCATTGGGA Tsu
900 ttaatcatattttaaactcgttttagtatgtttgtgatataaaacatttaacttcatgaatacagGCATAGAACCATTGTGACTATAATTCATTGGGA Col
999 TGTTCCTCA-GACITTAGAAGACGAATATGGAGGCTTCTTAAGCCACGATAGTgttaagtggttctatattctctcttagcttagatataatgtaac Tsu
1000 TGTTCCTCA-GACTTTAGAAGACGAATATGGAGGCTTCTTAAGCCACGATAGTgttaagtggttctatattctctcttagcttagatataatgtaac Col
1098 atcgcatacagatagtgctgtcctaataatgatgatacataaataagGGAGGACTTCAAAAATTACGCTGAGCTTCTATTCCAAAGATTCCGGAGACAGAGT Tsu
1100 atcgcatacagatagtgctgtcctaataatgatgatacataaataagGGAGGACTTCAAAAATTACGCTGAGCTTCTATTCCAAAGATTCCGGAGACAGAGT Col
1198 TAAATTTTGGATCACAATAAATCAGCCGTACTCTCTCGCTGTCAAAAGTTATGGAGATGGACAGTATCCACCGGGAAGGTGCAGTCACTGTAATTTGGA Tsu
1200 TAAATTTTGGATCACAATAAATCAGCCGTACTCTCTCGCTGTCAAAAGTTATGGAGATGGACAGTATCCACCGGGAAGGTGCAGTCACTGTAATTTGGA Col
1298 GGAGATTCGGAAGTGAACCTTATATAGTTGGACATCAGCAACTTCTGGCTCATATGGAAGCTGTATCTTTATACCGAAAAGATATCAGgtatacatac Tsu
1300 GGAGATTCGGAAGTGAACCTTATATAGTTGGACATCAGCAACTTCTGGCTCATATGGAAGCTGTATCTTTATACCGAAAAGATATCAGgtatacatac Col
1398 cgtaaactcataaatttctcttgcgtcgataactactttgtggtatatttgagctataatcaagaactacttaacagAAATTTCAAGGTGGTAAAGATAG Tsu
1400 cgtaaactcataaatttctcttgcgtcgataactactttgtggtatatttgagctataatcaagaactacttaacagAAATTTCAAGGTGGTAAAGATAG Col
1498 GAACGACATTTGATCGGTAGATGGTTCATCCCACTAAATGAACATAACGATCTCGACAAGGCTGCTGCAAAACGAGAAATTCGATTTTCCGTTGGCTGgtA Tsu
1500 GAACGACATTTGATCGGTAGATGGTTCATCCCACTAAATGAACATAACGATCTCGACAAGGCTGCTGCAAAACGAGAAATTCGATTTTCCGTTGGCTGgtA Col
1598 tatatgttcaactgatggattac-aaaacatt---atgtatataataaaacatgctttgcaatattaatt-atgtatattttcttacaatagtgaagGT Tsu
1600 tatatgttcaactgatggattacaaaacatttaattatgtatataataaa-catgctgtgcaatattaattatgtatattttcttacaatagtgtagGT Col
1692 TCTTGGATCCACTGGTGTACGACAAATATCCAAAGATAATGAGAGACATGCTAGGAGATAGATTGCCAAAATTCACGCCGAGCAATCAGCTTTACTTAA Tsu
1699 TCTTGGATCCACTGGTGTACGACAAATATCCAAAGATAATGAGAGAC----TAGGAGATAGATTGCCAAAATTCACGCCCAAGCAATCAGCTTTACTTAA Col
1792 AGGATCACTTGTATTTTCTAGGGTTGAACATATTACGTTACAAGATATGCAACCTCACAGACCTCCTCCAATGCCGACACAAATAGTGTCTTAAACCGATTCA Tsu
1795 AGGATCACTTGTATTTTCTAGGGTTGAACATATTACGTTACAAGATATGCAACCTCACAGACCTCCTCCAATGCCGACACAAATAGTGTCTTAAACCGATTCA Col
1892 GGAGTTACAATTTGGATgtatgtcaaacatttgtcgtattttagtggatttttccatataaatacaccaaaactaatcttttataattgggtttcagTTGA Tsu
1895 GGAGTTACAATTTGGATgtatgtcaaacatttgtcgtattttagtggatttttccatataaatacaccaaaactaatcttttataattgggtttcagTTGA Col
1992 GCGAAATGGAGTTTCTATTGGTGTAAAGgtcagtagtcaactttgttctcgtatattttttccagatgtgaaagatttcaagacatttagttgattttt Tsu
1995 GCGAAATGGAGTTTCTATTGGTGTAAAGgtcagtagtcaaacatttgttctcgtatattttttccagatgtgaaagatttcaagacatttagttgattttt Col
2092 ttttatttgcgagcttctacagGCTCCAAAGCTTCTCTACTATCCACCAGGATTCGTCAGATTTCAAATCACATCAAAAACAGAGTACAAGRATCCATT Tsu
2095 ttttatttgcgagcttctacagGCTCCAAAGCTTCTCTACTATCCACCAGGATTCGTCAGATTTCAAATCACATCAAAAACAGAGTACAAGRATCCATT Col
2192 AACCTACATCACCAAAACGgtatattgttcatcacaatgtttaaactttaaaccatgaaaaactatttgtgttagttacaatcttatgtcaatttgt Tsu
2178 AACCTACATCACCAAAACGgtatattgttcatcacaatgtttaaactttaaaccatgaaaaactatttgtgttagttacaatcttatgtcaatttgt Col
2292 agGAGTTGTGATGCTGATTTTGGAAAACGTAACGATCGCAATGCTCTTGGCCGATAATGGACGAATTCAAATTTCAATGCAGCCATCTTTCTTGTCTCAAA Tsu
2278 agGAGTTGTGATGCTGAT-----CGCAAATGCTCTTGGCCGATAATGGACGAATTCAAATTTCAATGCAGCCATCTTTCTTGTCTCAAA Col
2392 TGGCBAATTTGagtgagtttaataatataatcgtatattatgtgattactgtctctcggaga-tataatatttaattcaat-ccaaactatatttttaaaa Tsu
2361 TGGCBAATTTGagtgagtttaataatataatcgtatattatgtgattactgtctctcggagaatataaatttatattcatatacaaaactattttcttaaaa Col
2489 tatttaaataggGATGGATGCAACGTAGCAGGATATTTTGGCGTGGTCAATGATGGACAATTAAGAAATTCGAAATGGTTACACTCTCCGGTTTGGTATGAA Tsu
2461 tatttaaataggGATGGATGCAACGTAGCAGGATATTTTGGCGTGGTCAATGATGGACAATTAAGAAATTCGAAATGGTTACACTCTCCGGTTTGGTATGAA Col
2589 TTGGGTCAACTTCACTAATCCCTGCTGATCGGAGAGAAAAGCTTCTGGAAAATGGTTCTCTAGGTTCAATGCAAAAATAA 2667 Tsu
2561 TTGGGTCAACTTCACTAATCCCTGCTGATCGGAGAGAAAAGCTTCTGGAAAATGGTTCTCTAGGTTCAATGCAAAAATAA 2639 Col

```

Figure S1. Sequence comparison between the functional (from Tsu-1) and the disabled allele (from Col-0) of *AtTGG6* gene in *Arabidopsis thaliana*. The InDels leading to frame-shift mutations in Col-0 are highlighted in yellow. The exons are shown in capital letters, and the introns are shown in lower case.

Table S1. *AtTGG6* alleles from different ecotypes and their accession numbers in GenBank.

| Allele | Accession Number |
|---|-------------------------|
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ag-0</i> | KU301827 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Bsch-0</i> | KU301828 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Gel-1</i> | KU301829 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_La-0</i> | KU301830 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ler-1</i> | KU301831 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Mr-0</i> | KU301832 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Oy-0</i> | KU30183a3 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Tsu-1</i> | KU301834 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ty-0</i> | KU301835 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Wil-1</i> | KU301836 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Aa-0</i> | KU301837 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ba-1</i> | KU301838 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Col-0</i> | KU301839 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Col-4</i> | KU301840 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Col-J</i> | KU301841 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Cvi-0</i> | KU301842 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Edi-0</i> | KU301843 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Gre-0</i> | KU301844 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_JM-1</i> | KU301845 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_JM-2</i> | KU301846 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Kas-1</i> | KU301847 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Mir-0</i> | KU301848 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Mz-0</i> | KU301849 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Pr-0</i> | KU301850 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Stw-0</i> | KU301851 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ts-6</i> | KU301852 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Tul-0</i> | KU301853 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Van-0</i> | KU301854 |
| <i>AtTGG6_genomic-fll.sqn AtTGG6_Ws-0</i> | KU301855 |
| <i>AITGG4.sqn AITGG4</i> | KU301856 |
| <i>AITGG4.sqn AITGG5</i> | KU301857 |
| <i>AITGG4.sqn AITGG6</i> | KU301858 |
| <i>AITGG4.sqn AITGG45</i> | KU301859 |