

Supplemental Data

The following Supplemental Data files are included in “tpc123620Supplemental.rar”:

[123620Supplemental.pdf \(7.6 MB\)](#)

Supplemental Figure 1. Structural similarity between SFTI-1 and the inhibitory arm of Bowman-Birk Inhibitors.

Supplemental Figure 2. A multiple sequence alignment of PawS1 with preproalbumins from sunflower and compared to other plant species.

Supplemental Figure 3. *In planta* confirmation of PDP-3 in seed peptide extracts of *Tithonia rotundifolia*.

Supplemental Figure 4. Fingerprinted *Helianthus annuus* bacterial artificial chromosome 122C14 using a probe that would bind both *PawS1* and *PawS2*.

Supplemental Figure 5. Alignment of predicted protein sequences for 27 *PawS1* genes aligned against PawS1 from *Helianthus annuus*.

Supplemental Figure 6. Chronogram for 25 genera of Asteraceae subfamily Asteroideae.

Supplemental Figure 7. *In planta* confirmation of PDP-12 within peptide extracts from *Helianthus schweinitzii* seeds.

Supplemental Figure 8. *In planta* confirmation of PDP-4 in seed peptide extracts of *Iostephane heterophylla* seeds.

Supplemental Figure 9. *In planta* confirmation of PDP-5 within peptide extracts made from *Heliopsis helianthoides scabra* seeds.

Supplemental Figure 10. *In planta* confirmation of PDP-14 within peptide extracts made from *Heliopsis helianthoides scabra* seeds.

Supplemental Figure 11. *In planta* confirmation of PDP-10 within peptide extracts made from *Galinsoga quadriradiata* seeds.

Supplemental Figure 12. *In planta* confirmation of PDP-11 within peptide extracts made from *Galinsoga quadriradiata* seeds.

Supplemental Figure 13. *In planta* confirmation of PDP-8 within peptide extracts made from *Philactis zinnioides* seeds.

Supplemental Figure 14. *In planta* confirmation of PDP-8 within peptide extracts made from *Philactis nelsonii* seeds.

Supplemental Figure 15. Structure and activity comparison for SFTI-1 and variant peptides.

Supplemental Figure 16. PDP secondary chemical shifts.

Supplemental Figure 17. Stereo view of the family of 20 structures with highest MOLPROBITY score for each PDP.

Supplemental Figure 18. Molecular surface representation and comparison of PDPs.

Supplemental Figure 19. Trypsin inhibitory assays.

Supplemental Figure 20. Inhibition of insect proteases.

Supplemental Figure 21. *In vivo* processing of PawS1 mutants.

Supplemental Figure 22. The peptide GVLPPMLD encoded by Am-PawL1 is not detectable in peptide extracts of *Arnica montana* seeds.

Supplemental Figure 23. Alignment of predicted protein sequences for PawL1 and selected PawS1 genes.

Supplemental Figure 24. Structural similarity between trypsin inhibitors from a variety of inhibitor families.

Supplemental Table 1. Range of predicted and confirmed peptides from PawS1 genes.

Supplemental Table 2. NMR structure statistics.

Supplemental Table 3. PDP physicochemical properties.

Supplemental Table 4. Summary of findings by Konarev et al. (2002) in-gel trypsin inhibition assays.

Supplemental Table 5. Output statistics of sequencing and assembly quality for the *de novo* transcriptomes of *Helianthus annuus* and *Arnica montana*.

Supplemental Table 6. Models and output statistics of positive selection analysis.

Supplemental Table 7. Newick trees used in PAML analyses of PawS1 regions.

Supplemental Table 8. Exact indel rates and extension probabilities of the different albumin protein regions.

Supplemental Table 9. Primers used in this study.

Supplemental Table 10. ESI-ToF-MS/MS product ions for endo-GluC and trypsin fragments which correspond to Am-PawL1.

Supplemental Table 11. ESI-ToF-MS/MS product ions for the Am-PawL1 small sub-unit (SSU).

Supplemental Methods. Searching Heliantheae GenBank entries for BBIs.

[tpc123620SupplementalDS1.txt \(<0.01 MB\)](#)

Supplemental Dataset 1. BBI sequences used to generate BBI WebLogos.

[tpc123620SupplementalDS2.xls \(<0.01 MB\)](#)

Supplemental Dataset 2: Screening for SFTI-1 and PDPs.

267 species were screened in total (129 species [#8000-8184] in Mylne et al. (2011) denoted by (*) and an additional 138 screened here [#8185-8436], NB. Not all numbers in the sequence are represented). Species screened in this work identified the class of peptides to be more widely dispersed than originally thought, with PDPs being more widely dispersed than SFTI-1. All species are from the Asteraceae except #8016, 8045 and 8047 from the Dipsacaceae. There were hits for SFTI-1 in 33 species, 32 from the genus *Helianthus* and 21 hits for a PDP other than SFTI-1. A plant was deemed positive (+) for SFTI-1 or another PDP represented by a number (by # if there is a predicted PDP from gene screening that has not been confirmed *in planta*) if a product ion of the correct mass and retention time was present. Ast., Asteroideae; Car., Carduoideae; Cic., Cichorioideae; Wun., Wunderlichioideae; Mut., Mutisioideae.; Dip., Dipsacaceae.

[tpc123620SupplementalDS3.docx \(<0.01 MB\)](#)

Supplemental Dataset 3. Voucher information and GenBank accession numbers for sequences used to generate the phylogeny in Fig. 2D

[tpc123620SupplementalDS4 \(<0.01 MB\)](#)

Supplemental Dataset 4. NEXUS format text file of the sequences and alignment used to generate the phylogeny in Fig. 2D

[tpc123620SupplementalDS5.docx \(<0.01 MB\)](#)

Supplemental Dataset 5. Voucher information and GenBank accession numbers for sequences used to generate the phylogeny in Supplemental Fig. 6

[tpc123620SupplementalDS6 \(<0.01 MB\)](#)

Supplemental Dataset 6. NEXUS format text file of the sequences and alignment used to generate the phylogeny in Supplemental Fig. 6

[tpc123620SupplementalDS7.fa \(17.6 MB\)](#)

Supplemental Dataset 7. FASTA format of *de novo* seed transcriptome of *Helianthus annuus* translated into protein sequences from CDS predicted on BLAST results. CL14895 is *PawS1*, Unigene10551 is *PawS2*, Unigene62048 is *PawL1*.

[tpc123620SupplementalDS8.fa \(15.9 MB\)](#)

Supplemental Dataset 8. FASTA format *de novo* seed transcriptome of *Arnica montana* translated into protein sequences from CDS predicted on BLAST results. Unigene22510 is *Am-PawL1*.