

Journal: Plant Systematics and Evolution

Article title: The terpene synthase genes of *Melaleuca alternifolia* (tea tree) and comparative gene family analysis among Myrtaceae essential oil crops

Julia Voelker^{1,2}, Ramil Mauleon¹, Mervyn Shepherd¹

¹Faculty of Science and Engineering, Southern Cross University,
Military Road, East Lismore NSW 2480, Australia

² Julia Voelker (corresponding author)

e-mail: Julia.voelker@scu.edu.au

ORCID: <https://orcid.org/0000-0002-7615-0553>

Online Resource 2: Manual TPS annotation evidence

Examples for manual annotations of *TPS* genes in the Apollo genome viewer (see figure below). The first track shows the manual annotation, track two and three show the automatic fgenesh++ and genemarkET+ gene predictions, respectively. Track four (yellow) shows Exonerate evidence based on TPS protein alignments from other species, with black arrows indicating potential splice sites, and where available, track five (grey) shows RNA-seq alignment evidence. a) *TPS* genes classified as type I, complete genes containing both TPS Pfam domains and RNAseq alignment evidence. *MaltTPS025* is shown as example, whereby the last exon was manually adjusted based on Exonerate and RNAseq alignments; b) Complete genes containing both TPS Pfam domains but no RNAseq evidence (type II), with *MaltTPS040* as example. The gene length was adjusted based on Exonerate evidence; c) Genes with both TPS Pfam domains and one premature stop-codon or missing start codon, which might be a point mutation or assembly error (type III). *MaltTPS005* was manually adjusted based on Exonerate alignments to include the RRx8W motif in the first exon, but this resulted in a missing start codon. An alternative annotation is shown on top with a correct start codon, but then the annotation would not agree with spliced TPS protein alignments and the RRx8W motif would be missing; d) Pseudogenes containing only one TPS Pfam domain or multiple stop-codons and frameshifts (type IV). For the shown *Malt-pseudoTPS004*, the Exonerate alignment indicated eight exons with canonical splice sites, and the C-terminal TPS Pfam domain was detected. However, the manual annotation resulted in multiple premature stop codons and frame shifts in this gene model, classifying it as *TPS* pseudogene.

