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 Mauleon[1], Mervyn Shepherd[1]

[1]Faculty 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 5: Screen for organelle-targeting signal peptides

Potential chloroplast transit peptides (cTPs) were predicted using TargetP v2.0 (Almagro Armenteros et al. 2019) and DeepLoc v1.0 (Almagro Armenteros et al. 2017). In plants, sesquiterpenes are produced in the cytosol by TPS-a proteins, while all remaining TPS reactions are located in the chloroplast (Chen et al. 2011). Mitochondrial synthesis of several di- and sesquiterpenoids exists (Vranová et al. 2013) and it has been shown that some plants are capable of producing monoterpenoids in mitochondria, but that the transfer of GDP to mitochondria is limited (Vranová et al. 2013; Dong et al. 2016). Hence, a cTP should be present in most of the predicted TPS sequences, especially the monoTPS. However, cTPs display a high variation in their sequence length and structure (Lee et al. 2008), leading to difficulties in cTP predictions. Using our new TPS annotations for *M. alternifolia*, TargetP v2.0 predicted 13 cTP containing proteins, all of which belong to TPS-b subgroup. One TPS-g and two TPS-b proteins were predicted to contain a potential mitochondrial transit peptide (mTP). The two TPS-b proteins are the TPS-b2 members that are suggested to be one isoprene synthase (MaltTPS043) and one ocimene synthase (MaltTPS042), both of which use GDP as substrate that is produced in the plastidial MEP pathway. TargetP v1.1 has been reported to have difficulties in differentiating cTP and mTP peptides (Almagro Armenteros et al. 2019), and even though TargetP v2.0 uses an improved methodology compared to the previous versions, the mTP prediction in three TPS that are usually located in chloroplasts might be an error. Therefore, a second signal peptide prediction was run with DeepLoc v1 (Almagro Armenteros et al. 2017). The number of predicted cTP containing TPS was increased to 20, including the three sequences that TargetP predicted as mTP, as well as the two TPS-c and TPS-e proteins. Seven TPS-b proteins were still located in the cytoplasm based on the DeepLoc results, indicating that the program either did not recognise the cTP target signal, or that these protein sequences are still incomplete or contain assembly errors. DeepLoc predicted the three TPS-f proteins to be located at the cell membrane, however, the prediction has a low likelihood (40-44%) and is most likely not correct, as no other reports on TPS located in cell membranes could be found.

Calvert et al. (2018) reported only 5 predicted TPS containing a cTP in *M. alternifolia* and 6 in *E. grandis.* In order to compare results based on the same methods, the 37 previously predicted tea tree TPS were also analysed with TargetP v2.0 and DeepLoc v1.0. TargetP detected only two cTP signals and two mTPs, whereas DeepLoc located 9 TPS to the plastids. In conclusion, the advancements in predicting transit peptides, but especially the improved gene models reported in our study, notably increased the number of predicted cTP sequences compared to previous studies.

References

Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H (2019) Detecting sequence signals in targeting peptides using deep learning. Life Science Alliance 2:e201900429. doi:https://doi.org/10.26508/lsa.201900429

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387-3395. doi:https://doi.org/10.1093/bioinformatics/btx431

Calvert J, Baten A, Butler J, Barkla B, Shepherd M (2018) Terpene synthase genes in Melaleuca alternifolia: comparative analysis of lineage-specific subfamily variation within Myrtaceae. Plant Syst Evol 304:111-121. doi:https://doi.org/10.1007/s00606-017-1454-3

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. TPJ 66:212-229. doi:https://doi.org/10.1111/j.1365-313X.2011.04520.x

Dong L, Jongedijk E, Bouwmeester H, Van Der Krol A (2016) Monoterpene biosynthesis potential of plant subcellular compartments. New Phytol 209:679-690. doi:https://doi.org/10.1111/nph.13629

Lee DW, Kim JK, Lee S, Choi S, Kim S, Hwang I (2008) Arabidopsis Nuclear-Encoded Plastid Transit Peptides Contain Multiple Sequence Subgroups with Distinctive Chloroplast-Targeting Sequence Motifs. Plant Cell 20:1603-1622. doi:https://doi.org/10.1105/tpc.108.060541

Vranová E, Coman D, Gruissem W (2013) Network Analysis of the MVA and MEP Pathways for Isoprenoid Synthesis. Annu Rev Plant Biol 64:665-700. doi:https://doi.org/10.1146/annurev-arplant-050312-120116