Elucidation of Terpenoid Biosynthesis in Non-model Plants Utilizing Transcriptomic Data

Henrik Toft Simonsen[*]

Department of Plant and Environmental Sciences, Copenhagen Plant Science Centre, University of Copenhagen, Denmark

[*]**Corresponding author:** Henrik Toft Simonsen, Department of Plant and Environmental Sciences, Copenhagen Plant Science Centre, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark, Tel: +1-650-268-9744; Fax: +1-650-618-1414; E-mail: hts@plen.ku.dk

Abstract

The use of transcriptome sequencing has become very cheap, and it is increasingly being used for enzyme discovery in non-model plants. Especially in plants where no other genomic data is available (being 99.99% of all plants). This short commentary highlights the use of novel sequencing technologies in terpenoid biosynthesis characterization. It is also shown how tissue specific transcriptomic can be useful in this kind of research. Finally, a short overview of general procedures is given together with the perspectives of transcriptomics in plant biochemistry.

Keywords: Terpenoids; Transcriptomic; Non-model plants; Biochemistry

Introduction to Terpenoids

Terpenoids comprise the largest group of specialized plant metabolites, with tens of thousands known structures [1]. Terpenoids have key functions in cellular life (e.g. membrane fluidity, hormones and signalling compounds) in all kingdoms, but the vast majority of the characterized terpenoids are specialized metabolites in plants that influence the fitness of their biosynthesising plant [2]. The specialized plant terpenoids serve as defence compounds, deterrents, and pollinator attractants [3]. The role of the terpenoids has a huge impact on where in the plant the terpenoids are biosynthesized, and tissue specific biosynthesis is an important feature to understand when utilizing transcriptomic data for biosynthesis discovery [4]. Some are synthesized in specialized structures such as trichomes, oil bodies or resin ducts [5,6], whereas some are biosynthesized in general tissues like fruits and roots [7], and others are produced all over the plant [8]. The largest subgroups of terpenoids are the sesqui- and diterpenoids. These subgroups contain molecules that are highly valuable and are used in a range of industrial and medicinal applications [9-12]. For the majority of the described terpenoids their biosynthesis have not been described and no attempts have been made. Specialized terpenoid biosynthesis in general consist of first a terpene synthase that utilize a diphosphate substrate, this is often followed by a cytochrome p450 that decorate the backbone with different oxidations (alcohols, aldehydes, ketones and carboxylic acids) [13,14]. Following the cytochromes P450, there can be from one to 10's of these, enzyme groups such as alcohol dehydrogenases, reductases, and acyl transferases further decorate the molecule to form the final bioactive terpenoid [15]. The diversity of terpenoids is mainly due to the huge diversity within terpene synthases [16] and cytochromes P450 [17] that are involved in the biosynthesis. Recently the branching point in the artemisinin biosynthesis (the reductase DBR2) have been shown to be the key regulator enzym that determine the final product, thus the specific terpenoid profile of a given *Artemisia annua* variety, therefore it is important for the terpenoid diversity within *Artemisia* [14].

Recent observations within sesquiterpenoid biosynthesis studies found that less than ten cytochromes P450 had been described as part of a sesquiterpenoid biosynthesis and so far only within the *CYP71* clan [13]. This number is larger for diterpenoids, but excluding enzymes part of general metabolism, such as gibberellins, lowers this number significantly and limits the described cytochromes to enzymes in the *CYP71* and *CYP85* clans [18,19]. Among the specialized terpenoids the biosynthesis of the sesquiterpene lactone artemisinin is possibly the most well studied [14,20,21] closely followed by the sesquiterpene lactone costunolide [22-26], with several diterpenoids from gymnosperms picking up with an increasing number of publications, especially from the group of Jörg Bohlmann [5,15,27-32]. The use of transcriptome data are greatly facilitating this, and the number of papers utilizing this technology will dramatically increase in the coming years.

Examples of the Use of Transcriptome Data for Terpenoid Biosynthesis Elucidation

The enzyme family, terpene synthases has been known for decades, also before next generation sequencing took off. In 1995 Joe Chappel [33] in one of the first reviews of terpenoid biosynthesis have a chapter on "How little we know about terpenoid biosynthesis". The same year two papers was published on terpene synthases from plants describing the biochemistry and thereby providing an enzyme and transcriptomic sequence that could be used in subsequent BLAST searches [34,35]. These and many successive studies provides the first baits for the initial searches into a newly sequenced plant transcriptome, and today several papers have described the biosynthesis of specific terpenoids along with several on the general parts of terpenoid metabolism [17,36-38].

Within sesquiterpenoids, the artemisinin biosynthesis is the best described, and this was mostly elucidated without transcriptomics available. Subsequently, it has been shown that all five genes involved can be found in the transcriptome of the leaf trichomes [6]. This provided evidence that the trichomes are indeed the cellular compartment that perform the biosynthesis of artemisinin, which has also been confirmed using Gus staining and promoter analysis of the genes involved [14]. This has opened for the use of transcriptomics to

study expression levels of known and unknown genes involved in the biosynthesis in specific tissues. Other sesquiterpenoids biosynthesis has been fully or partly elucidated using transcriptomics. In 2009 and 2010, the transcriptome of the root and fruit of *Thapsia garganica* and the root of *Thapsia laciniata* (syn. *T. villosa*) was sequenced. This showed the presences of several terpene synthases and *CYP71* genes that could be involved on the biosynthesis of terpenoids, especially sesquiterpene lactones [39]. Using both 4-5-4 and Illumina for the sequencing, followed by thorough annotations of the sequenced genes lead to the discovery and characterization of the first step of the biosynthesis of thapsigargin from *Thapsia garganica* [40]. Another example within sesquiterpenoids is the characterization of the biosynthesis of the main component in the sandalwood oil, santalol. Following the transcriptomic analysis of the plant, the sesquiterpene synthase and the cytochrome p450 involved in was discovered in 2008 and 2013 [41,42]. The characterization of the terpene synthases in 2008 did involve cDNA isolation and genome walking, since the transcriptomic data at that time did not provide the same depth and coverage as the later analysis did. This is a general trend seen in non-model plants and not only for terpenoid biosynthesis [43].

Chemical analysis of pine trees have revealed a stunning chemical diversity of the resin. The transcriptomic analysis revealed that this is biosynthesized by only a few genes. The first genes was discovered using traditional degenerate primer design followed by genome walking, but later transcriptomic analysis was adopted. This has led to description of several genes from numerous gymnosperm including *Sikta spruce*, *Abies balsamea* and other pine trees [5,27-32,44-47]. The studies has described the biochemically characterization of a range of terpene synthases mainly found through blast searches that initially started with the general enzymes copalyl synthase and/or ent-kaurene synthase [46]. Subsequently the BLAST searches included cytochromes p450 and this lead to the description of e.g. the multifunctional cytochrome P450 CYP720B4 that is involved in the biosynthesis of a range of conifer defence compounds [5].

Other examples of the use of transcriptomics for the discovery of biosynthesis of terpenoids include that of ginsenoside and saponins [48-50]. These examples are just some of those described in the literature and they include a variety of genes that have been discovered and biochemically characterized. As part of the development, the use of transcriptomic analysis will also lead to phylogenetic studies. This has been seen both within gene families and across plants species. In gene families, the phylogeny of terpene synthases have been studied in several papers. This has led to a useful classification annotated TPSa-h that include plants from bryophytes to angiosperms [16,39]. Within cytochromes P450 the use of phylogeny was established long before next generation sequencing and are continuously used in the annotation of these enzymes [51]. Transcriptomic data also contain information that can be used in phylogenetic studies across species [52].

Lately, tissue and developmental specific transcriptomic analysis has proven to be a very effective tool in the discovery on new biosynthetic genes [53]. In grapes, a developmental study revealed the presence of several new genes that subsequently could be characterized and shown to be involved in the biosynthesis of the peppery aroma of Shiraz wine. This study also revealed that the transcription of specific terpene synthases was depend on the maturation state of the grape, which could be correlated to terpene content during the grape maturation [54,55]. In another example in *Coleus* plants, the use of laser dissection and imaging revealed that special cells in the cork of the root contained the interesting diterpenoids. The transcriptomic analysis of these special cells revealed parts of the biosynthetic machinery of forskolin biosynthesis [4,9]. These studies show that by combining transcriptomic analysis with modern dissection techniques the likelihood of discovering enzymes part of a biosynthetic route is significantly increased. Likewise, when the study is focused on specific tissues in specific developmental stages.

A Useful Approach

Discovery of any new biosynthesis of small molecules should include the following steps in order to be truly successful and utilize the full potential of the transcriptomic analysis. It is important first to establish profound chemical knowledge of the plant, especially where and when the chemical constituents are found. The use of advanced metabolomics tools such as GC-MS and LC-MS is crucial to establish at what developmental stage the constituents are produced. Utilizing advanced microscopic techniques to establish the cellular location of the constituents will show what cells to target. This will establish when and on what tissue to perform transcriptomic analysis in order to get the best coverage of biosynthetic genes. The first transcriptomic analysis should be based on tissues that clearly produce the compounds of interest. Subsequently, this can then be followed by studies that include stress inducement of producing tissues in order to enhance the biological understanding.

It is important early on to obtain very deep coverage of the RNA, since this eventually will prove useful for sorting out chimeric and other miss assemblies. Here the use of the latest developments within transcriptomics should be utilized (these techniques constantly improve and please consult technical reviews to obtain the latest knowledge on this). A first deep transcriptomic dataset open an avenue for later transcriptomic analysis of large sample sets including time-course, developmental and stress studies. Collectively all these studies will provide a solid database that include knowledge on sequence data, expression level, and time of expression.

The obtained database can then be used for BLAST searches of gene families. Here the utilization of minimal datasets that are family specific can be very useful [56]. Discovery of the genes of interest, have to be followed by the design of primers that can confirm that the gene sequence is also found in the living plants by PCR on the original plant material. Only then can one order a synthetic clone of the gene of interest for biochemical characterization. The PCR will confirm the *in-planta* sequence and even with the current technologies and assembly algorithms miss assemblies do occur.

Perspectives

With continuously falling prices on transcriptomic sequencing, increasing depth and coverage in the obtained datasets followed by continuously increasing amount of online available genomic data the perspectives in this field are daunting. The lack of transcriptomic sequence data will never again be the bottleneck in biology and biochemistry. However, the lack of fast biochemical and physiological screening methods will be the bottleneck for the decades to come. From just one transcriptomic analysis at € 2000, one will obtain enough RNA sequence information to keep tens of biochemistry post docs and PhD students busy for many years. Utilization of tissue and developmental specific transcriptomics will significantly shorten the discovery time as shown with both the grape and forskolin studies. Thus, chemical profiling and physiological studies of the plant prior to

transcriptomic analysis is highly recommended to lower the amount of biochemical characterization needed in the later studies.

It is foreseen that numerous studies will utilize transcriptomics throughout plant biology and that the field will explode in publications within the next couple of years.

References

1. http://dnp.chemnetbase.com/.

2. Ashour M, Wink M, Gershenzon (2010) Biochemistry of terpenoids: monoterpenes, sesquiterpenes and diterpenes. Annu Plant Rev 40: 258-303.

3. Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3: 408-414.

4. Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, et al. (2014) Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolin, Is Synthesized in Specialized Root Cork Cells in *Coleus forskohlii*. Plant Physiol 164: 1222-1236.

5. Hamberger B, Ohnishi T, Hamberger B, Séguin A, Bohlmann J (2011) Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. Plant Physiol 157: 1677-1695.

6. Wang W, Wang Y, Zhang Q, Qi Y, Guo D, et al. (2009) Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. BMC Genomics 10: 465.

7. Drew DP, Krichau N, Reichwald K, Simonsen HT (2009) Guaianolides in Apiaceae: perspectives on pharmacology and biosynthesis. Phytochem Rev 8: 581-599.

8. Bohlmann J, Keeling CI (2008) Terpenoid biomaterials. Plant J 54: 656-669.

9. Nielsen MT, Ranberg JA, Christensen U, Christensen HB, Harrison SJ, et al. (2014) Microbial synthesis of the forskolin precursor manoyl oxide in enantiomerically pure form. Appl Environ Microbiol 80: 7258-7265.

10. Simonsen HT, Weitzel C, Christensen SB (2013) Guaianolide sesquiterpenoids - Their pharmacology and biosynthesis. In: Ramawat KG, Merillon JM (eds) Handbook of Natural Products - Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes. Handbook of Natural Products. Springer-Verlag, Berlin, pp. 3069-3098.

11. Caniard A, Zerbe P, Legrand S, Cohade A, Valot N, et al. (2012) Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in *Salvia sclarea* (L.) and their relevance for perfume manufacture. BMC Plant Biol 12: 119.

12. Nguyen TD, MacNevin G, Ro DK (2012) De novo synthesis of high-value plant sesquiterpenoids in yeast. Methods Enzymol 517: 261-278.

13. Weitzel C, Simonsen HT (2015) Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. Phytochem Rev 14: 7-24.

14. Yang K, Monafared R, Wang H, Lundgren A, Brodelius P, et al. (2015) The activity of the artemisinic aldehyde Δ11(13) reductase promoter is important for artemisinin yield in different chemotypes of *Artemisia annua* L. Plant Mol Biol 24: 1-16.

15. Foster AJ, Hall DE, Mortimer L, Abercromby S, Gries R, et al. (2013) Identification of genes in *Thuja plicata* foliar terpenoid defenses. Plant Physiol 161: 1993-2004.

16. 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. Plant J 66: 212-229.

17. Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, et al. (2015) Investigation of terpene diversification across multiple sequenced plant genomes. PNAS 112: E81-E88.

18. Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos T R Soc B 368: 1471-2970.

19. Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, et al. (2011) Cytochromes P450. The Arabidopsis Book: e0144.

20. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496: 528-532.

21. Xiang L, Zeng LX, Yuan Y, Chen M, Wang F, et al. (2012) Enhancement of artemisinin biosynthesis by overexpressing dxr, cyp71av1 and cpr in the plants of Artemisia annua L. Plant Omics 5: 503-507.

22. Ikezawa N, Göpfert JC, Nguyen DT, Kim S-U, O'Maille PE, et al. (2011) Lettuce Costunolide Synthase (CYP71BL2) and Its Homolog (CYP71BL1) from Sunflower Catalyze Distinct Regio- and Stereoselective Hydroxylations in Sesquiterpene Lactone Metabolism. J Biol Chem 286: 21601-21611.

23. Liu Q, Majdi M, Cankar K, Goedbloed M, Charnikhova T, et al. (2011) Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. PloS One 6: e23255.

24. Azarken R, Guerra FM, Moreno-Dorado FJ, Jorge ZD, Massanet GM (2008) Substituent effects in the transannular cyclizations of germacranes. Synthesis of 6-epi-costunolide and five natural steiractinolides. Tetrahedron 64: 10896-10905.

25. Choi JH, Ha J, Park JH, Lee JY, Lee YS, et al. (2002) Costunolide triggers apoptosis in human leukemia U937 cells by depleting intracellular thiols. Japan J Cancer Res 93: 1327-1333.

26. de Kraker JW, Franssen MCR, Joerink M, de Groot A, Bouwmeester HJ (2002) Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory. Plant Physiol 129: 257-268.

27. Hall DE, Zerbe P, Jancsik S, Quesada AL, Dullat H, et al. (2013) Evolution of conifer diterpene synthases: diterpene resin acid biosynthesis in lodgepole pine and jack pine involves monofunctional and bifunctional diterpene synthases. Plant physiol 161: 600-616.

28. Zerbe P, Chiang A, Bohlmann J (2012) Mutational analysis of white spruce (*Picea glauca*) ent-kaurene synthase (PgKS) reveals common and distinct mechanisms of conifer diterpene synthases of general and specialized metabolism. Phytochemistry 74: 30-39.

29. Zerbe P, Chiang A, Yuen M, Hamberger B, Hamberger B, et al. (2012) Bifunctional cis-abienol synthase from *Abies balsamea* discovered by transcriptome sequencing and its implications for diterpenoid fragrance production. J Biol Chem 287: 12121-12131.

30. Hall DE, Robert JA, Keeling CI, Domanski D, Quesada AL, et al. (2011) An integrated genomic, proteomic and biochemical analysis of (+)-3-carene biosynthesis in Sitka spruce (*Picea sitchensis*) genotypes that are resistant or susceptible to white pine weevil. Plant J 65: 936-948.

31. Keeling CI, Madilao LL, Zerbe P, Dullat HK, Bohlmann J (2011) The primary diterpene synthase products of *Picea abies* levopimaradiene/ abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. J Biol Chem 286: 21145-21153.

32. Keeling CI, Weisshaar S, Ralph SG, Jancsik S, Hamberger B, et al. (2011) Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). Bmc Plant Biol 11: 43.

33. Chappell J (1995) Biochemistry and Molecular-Biology of the Isoprenoid Biosynthetic-Pathway in Plants. Annu Rev Plant Phys 46: 521-547.

34. Chen XY, Chen Y, Heinstein P, Davisson VJ (1995) Cloning, expression, and characterization of (+)-delta-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. Arch Biochem Biophys 324: 255-266.

35. Chappell J, Wolf F, Proulx J, Cuellar R, Saunders C (1995) Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme-a Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants. Plant Physiol 109: 1337-1343.

36. Kitaoka N, Lu X, Yang B, Peters Reuben J (2015) The Application of Synthetic Biology to Elucidation of Plant Mono-, Sesqui-, and Diterpenoid Metabolism. 8: 6-16.

37. Kitaoka N, Lu X, Yang B, Peters RJ (2014) The application of synthetic biology to elucidation of plant mono-, sesqui-and diterpenoid metabolism.ssu104.

38. Zi J, Mafu S, Peters RJ (2014) To Gibberellins and Beyond! Surveying the Evolution of (Di)terpenoid Metabolism. Annu Rev Plant Biol 65: 259-286.

39. Drew DP, Dueholm B, Weitzel C, Zhang Y, Sensen C, et al. (2013) Transcriptome Analysis of *Thapsia laciniata* Rouy Provides Insights into Terpenoid Biosynthesis and Diversity in Apiaceae. Int J Mol Sci 14: 9080-9098.

40. Pickel B, Drew DP, Manczak T, Weitzel C, Simonsen HT, et al. (2012) Identification and characterization of a kunzeaol synthase from *Thapsia garganica*: implications for the biosynthesis of the pharmaceutical thapsigargin. Biochem J 448: 261-271.

41. Diaz-Chavez ML, Moniodis J, Madilao LL, Jancsik S, Keeling CI, et al. (2013) Biosynthesis of Sandalwood Oil: *Santalum album* CYP76F cytochromes P450 produce santalols and bergamotol. PloS One 8: e75053.

42. Jones CG, Keeling CI, Ghisalberti EL, Barbour EL, Plummer JA, et al. (2008) Isolation of cDNAs and functional characterisation of two multi-product terpene synthase enzymes from sandalwood, *Santalum album* L. Arch Biochem Biophys 477: 121-130.

43. Vera J, Wheat C, Fescemyer H, Frilander M, Crawford D, et al. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17: 1636 - 1647.

44. Zulak KG, Bohlmann J (2010) Terpenoid Biosynthesis and Specialized Vascular Cells of Conifer Defense. J Integrat Plant Biol 52: 86-97.

45. Robert JA, Madilao LL, White R, Yanchuk A, King J, et al. (2010) Terpenoid metabolite profiling in Sitka spruce identifies association of dehydroabietic acid, (+)-3-carene, and terpinolene with resistance against white pine weevil. Botany 88: 810-820.

46. Keeling CI, Dullat HK, Yuen M, Ralph SG, Jancsik S, et al. (2010) Identification and functional characterization of monofunctional ent-copalyl diphosphate and ent-kaurene synthases in white spruce reveal different patterns for diterpene synthase evolution for primary and secondary metabolism in gymnosperms. Plant physiol 152: 1197-1208.

47. Ro DK, Bohlmann J (2006) Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*): Functional characterization of abietadiene/ levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1). Phytochemistry 67: 1572-1578.

48. Sui C, Zhang J, Wei J, Chen S, Li Y, et al. (2011) Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins. BMC genomics 12: 539.

49. Sun C, Li Y, Wu Q, Luo H, Sun Y, et al. (2010) De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. BMC Genomics 11: 262.

50. Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, et al. (2012) UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant physiology 160: 1881-1895.

51. Nelson DR (2013) A world of cytochrome P450s. Phil Trans R Soc B 368 (1612).

52. Weitzel C, Rønsted N, Simonsen HT (2014) Resurrecting deadly carrots. Towards a revision of *Thapsia* L. (Apiaceae) based on phylogenetic analysis of nrITS sequences and chemical profiles. Bot J Linn Soc 174: 620-636.

53. Lange BM (2014) The Evolution of Plant Secretory Structures and the Emergence of Terpenoid Chemical Diversity. Annu Rev Plant Biol.

54. Drew DP, Andersen TB, Sweetman C, Møller BL, Simonsen HT, et al. (2015) Cloning of cDNAs and in planta characterization of five sesquiterpene synthases from Vitis vinifera (cv. Shiraz) berries. Plant Molecular Biology submitted.

55. Sweetman C, Wong DC, Ford CM, Drew DP (2012) Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genom 13: 691.

56. Hamberger B (2015) P450 family minimal dataset. Resived a copy of Assoc. Prof. Björn Hambergers personal P450 dataset. University of Copenhagen, Personal Communication.