Mol Genet Genomics (2002) 267: 730–745
DOI 10.1007/s00438-002-0709-y

ORIGINAL PAPER

S. Aubourg · A. Lecharny · J. Bohlmann

Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*

Received: 1 February 2002 / Accepted: 27 May 2002 / Published online: 29 June 2002

Abstract A family of 40 terpenoid synthase genes (*AtTPS*) was discovered by genome sequence analysis in *Arabidopsis thaliana*. This is the largest and most diverse group of *TPS* genes currently known for any species. *AtTPS* genes cluster into five phylogenetic subfamilies of the plant *TPS* superfamily. Surprisingly, thirty *AtTPS* closely resemble, in all aspects of gene architecture, sequence relatedness and phylogenetic placement, the genes for plant monoterpene synthases, sesquiterpene synthases or diterpene synthases of secondary metabolism. Rapid evolution of these *AtTPS* resulted from repeated gene duplication and sequence divergence with minor changes in gene architecture. In contrast, only two *AtTPS* genes have known functions in basic (primary) metabolism, namely gibberellin biosynthesis. This striking difference in rates of gene diversification in primary and secondary metabolism is relevant for an understanding of the evolution of terpenoid natural product diversity. Eight *AtTPS* genes are interrupted and are likely to be inactive pseudogenes. The localization of *AtTPS* genes on all five chromosomes reflects the dynamics of the *Arabidopsis* genome; however, several *AtTPS* genes are clustered and organized in tandem repeats. Furthermore, some *AtTPS* genes are localized with prenyltransferase genes (*AtGGPPS*, geranylgeranyl diphosphate synthase) in contiguous genomic clusters encoding consecutive steps in terpenoid biosynthesis. The clustered organization may have implications for *TPS* gene evolution and the evolution of pathway segments for the synthesis of terpenoid natural products. Phylogenetic analyses highlight events in the divergence of the *TPS* paralogs and suggest orthologous genes and a model for the evolution of the *TPS* gene family.

Keywords Gene evolution · Secondary metabolism · Isoprenoid · Terpene cyclase · Prenyl transferase.

Communicated by G. Haughn

S. Aubourg
Unité de Recherche en Génomique Végétale,
Institut National de la Recherche Agronomique (INRA),
FRE-CNRS, 2 Rue Gaston Crémieux,
CP 5708, F-91057 Evry Cedex, France

A. Lecharny
Institut de Biotechnologie des Plantes,
CNRS UMR 8618, Université de Paris-Sud,
Bâtiment 630, F-91405 Orsay Cedex, France

J. Bohlmann (✉)
Biotechnology Laboratory,
Dept. of Botany and Dept. of Forest Sciences,
University of British Columbia,
6174 University Boulevard,
Vancouver V6T 1Z3, B.C., Canada
E-mail: bohlmann@interchange.ubc.ca
Fax: +1-604-8226097

Introduction

Terpenoids (isoprenoids) are a very large and structurally diverse group of natural products (Buckimgham 1998). In plants, gibberellin and brassinosteroid hormones, carotenoid pigments in photosynthesis, and the phytol side-chain of chlorophylls are some well characterized terpenoids with indispensable functions for basic metabolism, growth and development. The majority of terpenoids, however, have functions in a plethora of ecological interactions of plants with other organisms, and have traditionally been referred to as secondary metabolites or as natural products. Terpenoid secondary metabolites are abundant in many essential oils (Lawrence 1992), resins (Rowe 1989) and floral scents (Knudsen et al. 1993; Dudareva and Pichersky 2000). Some terpenoids have roles as phytoalexins in plant-pathogen relationships (Hammerschmidt 1999), as allelopathic inhibitors in plant-plant interactions (Harborne 1991), or as airborne molecules of plant-herbivore multitrophic signalling (Pare and Tumlinson 1999) and plant-plant signalling (Arimura et al. 2000).

Elucidation of the biochemistry and molecular genetics of terpenoid biosynthesis has made rapid progress

in recent years (Cane 1999a) (Fig. 1). The five-carbon biosynthetic building blocks of terpenoids, isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), are formed in two separate and independent pathways. The mevalonate pathway is localized in the cytosol/endoplasmic reticulum (Newman and Chappell 1999); the 2-C-methylerythritol-4-phosphate (MEP) pathway, which proceeds via 1-deoxyxylulose-5-phosphate, occurs in plastids (Eisenreich et al. 1998; Lichtenthaler 1999; Rohmer 1999). Condensation of IPP and DMAPP is catalyzed by prenyltransferases (PTs) and yields three central intermediates of the isoprenoid pathway, geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) (Koyama and Ogura 1999). The hundreds of basic parent skeletons typical of plant terpenoids are then formed from DMAPP, GPP, FPP or GGPP by terpene synthases (TPS) (Bohlmann et al. 1998a; Davis and Croteau 2000). TPS form the hemiterpene isoprene (five carbon atoms) from DMAPP (Miller et al. 2001), monoterpenes (10 carbon atoms) from GPP (Wise and Croteau 1999), sesquiterpenes (15 carbon atoms) from FPP (Cane 1999b), and diterpenes (20 carbon atoms) from GGPP (MacMillan and Beale 1999). These

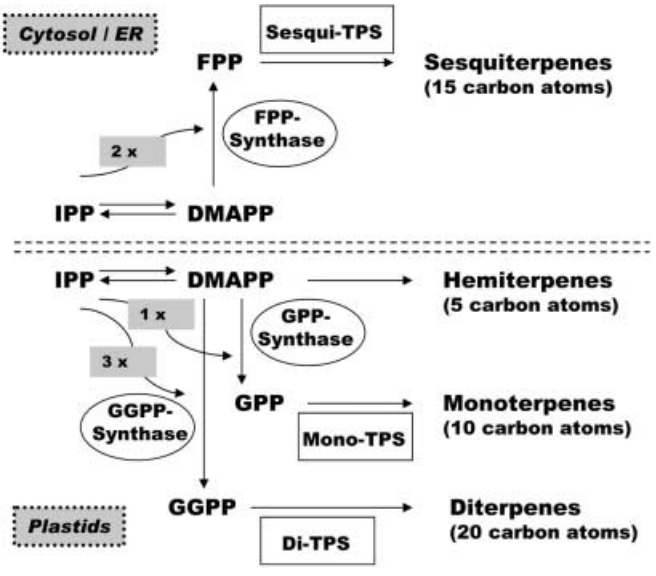

Fig. 1. Scheme of the pathways of terpenoid biosynthesis. The five-carbon precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), are formed via two pathways, the mevalonate pathway in the cytosol/endoplasmic reticulum and the 2-C-methylerythritol-4-phosphate (MEP) pathway, which proceeds via 1-deoxyxylulose-5-phosphate, in plastids. Prenyltransferases, catalyze (1′-4) head-to-tail condensations of DMAPP with one, two or three molecules of IPP, to form geranyl diphosphate (GPP; GPP synthase), farnesyl diphophate (FPP; FPP synthase) and geranylgeranyl diphosphate (GGPP; GGPP synthase), respectively. The terpene synthases (TPS) convert the three prenyl diphosphate intermediates into cyclic and acyclic terpenoid skeletons, yielding 10-carbon monoterpenes (monoterpene synthases), 15-carbon sesquiterpenes (sesquiterpene synthases), and 20-carbon diterpenes (diterpene synthases). Isoprene synthase uses DMAPP as a substrate to form a five-carbon hemiterpene

synthases, with the exception of copalyl diphosphate (CPP) synthase, function through the generation of carbocation intermediates from the respective prenyl diphosphate substrates by a divalent metal ion-dependent reaction (Davis and Croteau 2000).

The plant TPS that yield the low-molecular-weight terpenoids (isoprene, monoterpenes, sesquiterpenes and diterpenes) have a common phylogenetic origin and are members of a diverse TPS family that is divided into six subfamilies (TPS-a through TPS-f; Bohlmann et al. 1998a). Angiosperm *TPS* genes can be divided into two classes, class-I *TPS* genes and class-III *TPS* genes, based on gene architecture (Trapp and Croteau 2001) (see below). All class-I TPS have as a characteristic feature an unusual sequence motif of approximately 200 amino acids (200-aa motif) that is absent in other TPS. Phylogenetic analysis of *TPS* gene structure suggests that an ancestral class I-type *TPS* gene gave rise to all known angiosperm and gymnosperm TPS of the six subfamilies (Trapp and Croteau 2001). TPS enzymes of the same subfamily are similar in sequence and have, in general, similar functions. For example, different angiosperm monoterpene synthases of secondary metabolism belong to a single well defined subfamily (TPS-b) and can be distinguished based on sequence from sesquiterpene synthases and diterpene synthases (TPS-a). However, specific product profiles of members of the same subfamily can be quite diverse and cannot be predicted based on sequence alone (Bohlmann et al. 1998a).

Targeted cloning of individual *TPS* has shown that several plants have multiple *TPS* genes (Bohlmann et al. 2000a). For instance, *Nicotiana tabacum* was estimated to contain 12–15 genes related to the sesquiterpene synthase *epi*-aristolochene synthase (Facchini and Chappell 1992); several different monoterpene synthases exist in *Salvia officinalis* (Wise et al. 1998), and the conifer *Abies grandis* expresses at least eleven TPS, each with a different biochemical function (Bohlmann et al. 1999). The complete *TPS* gene family, however, is not known for any of these species.

Very little is known about low-molecular-weight terpenoid natural products in *Arabidopsis thaliana*. Recently, several terpenoids were found to be emitted as volatiles from *Arabidopsis* upon herbivory by larvae of the cabbage butterfly, *Pieris rapae* (van Poecke et al. 2001). Characterization of a cDNA for myrcene/ocimene synthase, a monoterpene synthase, demonstrated that at least one functional *TPS* gene for terpenoid natural product formation is actively expressed in *Arabidopsis* (Bohlmann et al. 2000b). In this paper, we present a detailed, genome-wide in silico analysis of the remarkably large *AtTPS* gene family, and present a model for the evolution of the complete *AtTPS* family. Surprisingly, many of the previously predicted *AtTPS* coding sequences currently available in the public databases were found to contain errors (see Table 1). Improved prediction of *AtTPS* coding sequences, based on known *TPS* transcripts and improved bioinformatics tools, and sequence motif analyses and phylogenetic