Genetics and Molecular Biology, 34, 4, 640-646 (2011)

www.sbg.org.br

Short Communication

Expressed sequence tag analysis of khat (*Catha edulis*) provides a putative molecular biochemical basis for the biosynthesis of phenylpropylamino alkaloids

Jillian M. Hagel[1], Raz Krizevski[2,3], Korey Kilpatrick[4,5], Yaron Sitrit[6], Frédéric Marsolais[4,5], Efraim Lewinsohn[2] and Peter J. Facchini[1]

[1]*Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada.*
[2]*Department of Aromatic, Medicinal and Spice Crops, Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay, Israel.*
[3]*Albert Katz Department of Dryland Agriculture, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer, Israel.*
[4]*Department of Biology, University of Western Ontario, London, Ontario, Canada.*
[5]*Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada, London, Ontario, Canada.*
[6]*The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of The Negev, Beer-Sheva, Israel.*

Abstract

Khat (*Catha edulis* Forsk.) is a flowering perennial shrub cultivated for its neurostimulant properties resulting mainly from the occurrence of (*S*)-cathinone in young leaves. The biosynthesis of (*S*)-cathinone and the related phenylpropylamino alkaloids (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine is not well characterized in plants. We prepared a cDNA library from young khat leaves and sequenced 4,896 random clones, generating an expressed sequence tag (EST) library of 3,293 unigenes. Putative functions were assigned to > 98% of the ESTs, providing a key resource for gene discovery. Candidates potentially involved at various stages of phenylpropylamino alkaloid biosynthesis from L-phenylalanine to (1*S*,2*S*)-cathine were identified.

Key words: khat, (*S*)-cathinone, phenylpropylamino alkaloids biosynthesis, EST library, gene discovery.

Received: June 29, 2011; Accepted: August 17, 2011.

Khat (*Catha edulis*), a perennial flowering shrub native to East Africa and the Arabian Peninsula, has long been cultivated for its neurostimulant properties. Evidence suggests that chewing young khat leaves as a social activity dates back at least a thousand years (Klein *et al.*, 2009) and might even predate the use of coffee (Balint *et al.*, 2009). It was not until 1975 that United Nations laboratories identified the phenylpropylamino alkaloid (*S*)-cathinone as the compound largely responsible for the mild euphoric and anorexic properties of khat (United Nations, 1975). Studies involving the short- and long-term dangers of chewing khat are generally inconclusive, and the physical harm and dependence caused by the plant remain controversial (Mateen and Cascino, 2010). For example, while some studies have linked khat consumption with genotoxic effects in humans (Kassie *et al.* 2001), others have highlighted khat as a potential source of anti-cancer agents (Bredholt *et al.* 2009). Recent evidence has linked khat use with impaired memory and cognitive flexibility (Colzato *et al.* 2011). While khat chewing is a longstanding tradition in parts of East Africa and in the Middle East, possession of khat is illegal in Canada, the United States and parts of the European Union. Fresh khat is a scheduled drug under the controlled substances legislation in Canada and the United States, yet may be imported with proper licensing in Australia and use of the plant is unregulated in the United Kingdom, the Netherlands (Klein *et al.*, 2009) and in Israel (Krizevski *et al.* 2008). A sizable portion of the seven metric tons of licit khat, which is classified as a vegetable in the United Kingdom and therefore not subject to tax, is estimated to travel through Heathrow Airport each week destined for black-market distribution in North America (Klein *et al.*, 2009). Due to controversial and inconsistent domestic policies, and fast-growing communities of East African immigrants

Send correspondence to Peter Facchini. Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N 1N4 Canada. E-mail: pfacchin@ucalgary.ca.

(Gebissa, 2010), khat has become a subject of international concern.

In addition to (*S*)-cathinone, khat accumulates the monoamine alkaloids (1*S*,2*S*)-norpseudoephedrine (cathine) and its diastereomer (1*R*,2*S*)-norephedrine (Krizevski *et al.*, 2007, 2008). *N*-Methylated versions of these compounds such as the sinus decongestants (1*S*,2*S*)-pseudoephedrine and (1*R*,2*S*)-ephedrine are restricted to *Ephedra* spp. and are not found in khat. Together, these phenylalkylamino alkaloids comprise a unique class of compounds derived from L-phenylalanine (Phe) (Figure 1). Early pulse-labeling studies using *Catha edulis* (Leete, 1958) and *Ephedra distachya* (Yamasaki *et al.*, 1969, 1973) established that only the C_6-C_1 component of ephedrine alkaloids is derived from Phe, whereas later studies suggested that the C_2-C_3 unit derives from pyruvic acid (Grue-Sørensen and Spenser, 1988, 1989). A related study suggested that Phe-derived benzoic acid is an intermediate in the formation of phenylpropylamino alkaloids (Grue-Sørensen and Spenser, 1994) although the involvement of benzoyl-CoA (Krizevski *et al.*, 2010) or benzaldehyde cannot be ruled out. The first step in the pathway is catalyzed by L-phenylalanine ammonia lyase (PAL), a well-characterized enzyme in many plants including *Ephedra sinica* (Okada *et al.*, 2008). Although benzoic acid biosynthesis in plants remains unresolved, the propyl side chain of Phe is known to undergo shortening via either β-oxidative or non-β-oxidative routes (Boatright *et al.*, 2004) (Figure 1). Several enzymes involved in benzoic acid biosynthesis including 4-coumaroyl-CoA ligase (4CL; Schilmiller *et al.*, 2009), benzoyl-CoA ligase (BZO1; Kliebenstein *et al.*, 2007), a 3-ketoacyl-CoA thiolase (KAT1; Van Moerkercke *et al.*, 2009) and two distinct dehydrogenases from *Antirrhinum majus* (snapdragon) (BALDH; Long *et al.*, 2009) and *Arabidopsis thaliana* (AAO4; Ibdah *et al.*, 2009) have been isolated in plants that produce floral volatiles or glucosinolates. Alternatively, benzaldehyde might be formed via phenylpyruvate, which is a transamination product of Phe. This route occurs in lactic acid bacteria (Nierop-Groot and de Bont, 1999), but has not been confirmed in plants. Nonetheless, the transamination of Phe to phenylpyruvate has recently been demonstrated, which might lead to the formation of benzaldehyde in melon fruit (Gonda *et al.*, 2010).

An enzyme catalyzing the condensation of benzoic acid, benzoyl-CoA or benzaldehyde with pyruvate has not been detected in khat or other phenylpropylamino alkaloid-producing plants, although the involvement of a ThDP-dependent pyruvate decarboxylase (PDC; EC 4.1.1.1) or an acetohydroxyacid synthase (AHAS; EC 2.2.1.6) have been suggested (Grue-Sørensen and Spenser, 1989). These two distantly related enzymes (Green, 1989) share the common step of pyruvate decarboxylation. However, whereas PDCs generally decompose pyruvate to acetaldehyde and CO_2 (Meyer *et al.*, 2011), AHASs catalyze carboligation reactions forming either acetolactate or acetohydroxybutyrate products as part of branched-chain amino acid biosynthesis (Jaña *et al.*, 2010). Members of both enzyme classes have shown carboligase activity toward benzaldehyde, yielding (*R*)-phenylacetylcarbinol (Engel *et al.*, 2003; Meyer *et al.*, 2011) (Figure 2). In fact, this side reaction of microbial PDCs has gained recent attention as a means of ephedrine production, as (*R*)-phenylacetylcarbinol can be chemically

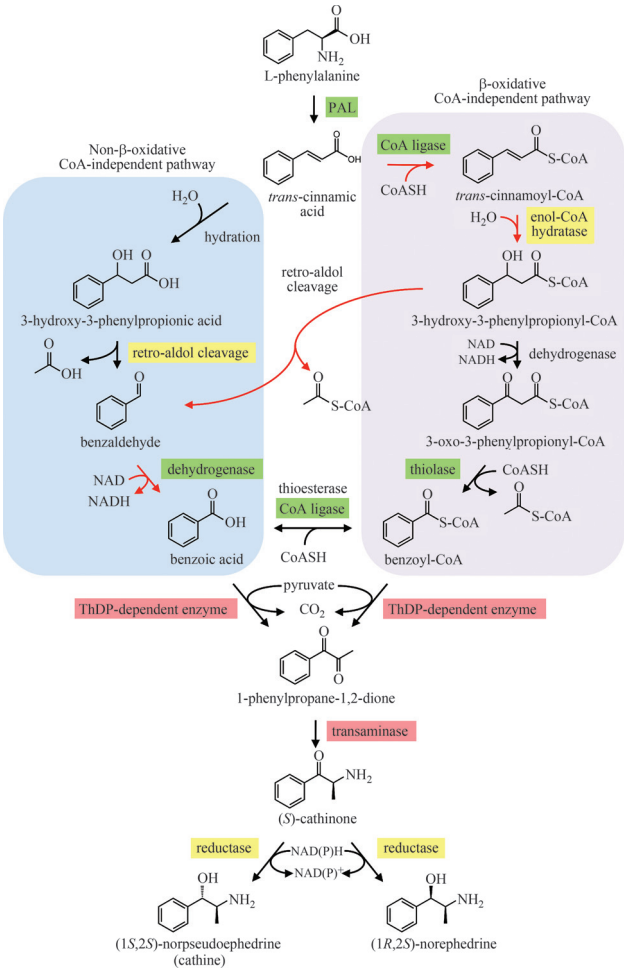


Figure 1 - Proposed biosynthetic routes leading from L-phenylalanine to phenylpropylamino alkaloids in khat. A CoA-independent, non-β-oxidative pathway of L-phenylalanine side chain-shortening is shown in blue, whereas a CoA-dependent, β-oxidative route is shown in purple. Red arrows indicate an alternative CoA-dependent, non-β-oxidative route suggested to occur in some plants (Abd El-Mawla and Beerhues 2002; Boatright *et al.* 2004). Either benzoic acid or benzoyl-CoA undergoes condensation with pyruvate, a reaction putatively catalyzed by a ThDP-dependent enzyme. 1-Phenylpropane-1,2-dione undergoes transamination to yield (*S*)-cathinone, which is reduced to (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine. Activity has been detected for enzymes highlighted in yellow, and corresponding genes are available for enzymes highlighted in green. Enzymes highlighted in red have not been isolated, although EST analysis revealed numerous potential candidates (Table 1). *Catha edulis* ESTs putatively involved in this pathway have been identified for many steps, and candidates are listed in Table 1. Abbreviations: PAL, phenylalanine ammonia lyase; CoA, Coenzyme A; ThDP, thiamine diphosphate; NAD(H), nicotinamide adenine dinucleotide; NADP(H), nicotinamide adenine dinucleotide phosphate.

converted to (1*R*,2*S*)-ephedrine (Meyer *et al.*, 2011). Although most microbial AHASs and PDCs catalyze the stereoselective production of (*R*)-phenylacetylcarbinol from benzaldehyde, engineered PDCs yield both the *R* and *S* enantiomers in addition to other products (Figure 2) (Pohl *et al.*, 1998). It is possible that phenylacetylcarbinol or a related compound is an intermediate in the formation of 1-phenylpropane-1,2-dione in plants. Subsequently, 1-phenylpropane-1,2-dione undergoes transamination to yield (*S*)-cathinone, which is converted to (1*S*,2*S*)-cathine or (1*R*,2*S*)-norephedrine by NADPH-dependent reduction (Figure 1). Although the transamination step has not yet been characterized, (*S*)-cathinone reductase activity was reported recently in *Ephedra sinica* stems (Krizevski *et al.*, 2010) and khat leaves (Krizevski *et al.*, 2007).

Although this pathway has been partially characterized at the biochemical level, no biosynthetic genes involved in the conversion of *trans*-cinnamic acid to phenylpropylamino alkaloids have been isolated. To establish a functional genomics platform aimed at gene discovery, we took the approach of building an EST library from biosynthetically active khat leaves. It was recently shown that the pathway intermediates 1-phenylpropane-1,2-dione and (*S*)-cathinone, and the end products (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine accumulate mainly in young leaves and flowers of khat with lesser quantities in young stems (Krizevski *et al.*, 2007, 2008). In contrast, mature leaves lack (*S*)-cathinone and accumulate only (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine suggesting that phenylpropylamino alkaloid biosynthetic gene expression is highest in young tissues. For this reason young khat leaves were selected for EST analysis. Khat shrubs (*Catha edulis*, Forsk.) were grown in open field conditions using commercial cultivation practices, including drip irrigation and fertilization, at the Newe Ya'ar Research Center in Northern Israel. Young khat leaves approximately 1-3 cm in length were harvested from five-year-old plants during daylight hours in November 2006. Total RNA was isolated using an RNeasy Midi kit (Qiagen) and poly(A)+ RNA was selected using a Dynal Dynabeads kit (Invitrogen). The poly(A)$^+$ RNA was converted to cDNA using the ZAP cDNA synthesis kit (Stratagene) and the resulting clones were unidirectionally inserted into *Eco*RI and *Xho*I sites within the phage vector λUni-ZAPII XR, and packaged by Gigapack III Gold packaging extract (Stratagene). Primary libraries were converted into plasmids by *in vivo* excision, and *Escherichia coli* colonies were randomly transferred to 96-well microtiter plates for automated plasmid preparation using Templiphi Template Amplification kit (GE Healthcare Life Sciences). Twenty randomly chosen plasmid clones were digested using *Eco*R1 and *Xho*1 restriction enzymes for agarose gel electrophoresis analysis to check the insertion rate and average insert length. Sequencing of cDNA inserts was performed using an ABI Prism Big Dye terminator sequencing kit (Applied Biosystems) and an AB 3730 genetic analyzer (Applied Biosystems).

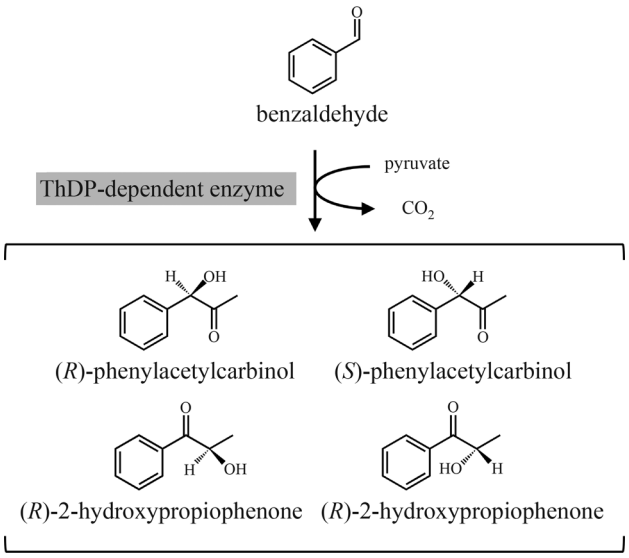


Figure 2 - Carboligation products of benzaldehyde and pyruvate formed by ThDP-dependent AHAS and PDC enzymes in microbes. (*R*)-Phenylacetylcarbinol is formed by AHAS in *Escherichia coli* and PDCs in certain yeast and bacteria. Mutation at a single amino acid in *Zymomonas mobilis* PDC enhanced production of (*S*)-phenylacetylcarbinol and resulted in the formation of both (*R*)- and (*S*)-2-hydroxypropiophenone. Although no evidence is presently available, one or more of these reaction products could be an intermediate in the formation of 1-phenylpropane-1,2-dione in khat.

A total of 4,896 clones were randomly selected from the *C. edulis* library and submitted for unidirectional sequencing from the 5' end using M13 primer. DNA sequencer traces were interpreted and vector and low-quality sequences were eliminated using PHRED (Ewing *et al.*, 1998) and LUCY (Chou and Holmes, 2001), resulting in 4,723 high-quality expressed sequence tag (EST) sequences (96.5%). The ESTs were submitted to GenBank and assigned accession numbers JG723448 through JG728170. Cluster analysis and contig assembly were performed using STACKPACK (Miller *et al.*, 1999), resulting in 3,293 unigenes (Supplemental Table S1). Sequence comparisons were done using the BLAST algorithm (Altschul *et al.*, 1990) with the public sequence databases TAIR Proteins v.8 and UniProt Plants v.14.5. BLAST analysis yielded matches for the majority of unigenes, with only 21 and 56 ESTs finding "no hit" when compared to TAIR Proteins (< 1%) and UniProt (1.7%) databases, respectively (Table S1). These results compare favorably with similar EST-based gene expression studies. For example, an EST-based study of gene expression in flax (*Linum usitatissimum*) seed that used similar homology-search parameter cutoffs (*e.g.* E-value of e-6) revealed a match between only 76.4% of flax unigenes with *Arabidopsis* proteins (Venglat *et al.*, 2011). In another example, analysis of 5,023 unigenes derived from Madagascar periwinkle (*Catharanthus roseus*) yielded a "no hit" rate of 14.2% against GenBank entries, although different annotation parameters were used (Murata *et al.*, 2006).

Khat unigenes that showed significant homology (E-value < e-10) to known proteins of UniProt Plants were selected for Gene Ontology (GO) annotation and mapping to the TAIR database, which is updated on a regular basis (Berardini *et al.*, 2004). GO Annotation analysis assigned a functional category to 2,839 (88%) of the unigenes possessing hits against public databases (Supplemental Table S2). However, to better reflect khat transcripts putatively involved in specialized metabolism, including phenylpropylamino alkaloid biosynthesis, which is a category not included in GO annotations, the *Arabidopsis*-based ontology results were manually verified and reclassified into 8 categories (Figure 3). Although a large proportion of the khat library (33%) appears dedicated to primary metabolism, nearly 5% of the ESTs encoded proteins putatively involved in specialized metabolism. This category includes candidates for enzymes shown in Figure 1, and those putatively involved in flavonoid and terpenoid biosynthesis.

The recent discoveries of biosynthetic genes involved in benzoic acid metabolism facilitated a tBLASTn-based search of the khat EST library for homologues, all of which were detected except for *Arabidopsis thaliana* aldehyde oxidase-4 (AtAAO4) catalyzing the dehydrogenation of benzaldehyde to benzoic acid (Ibdah *et al.*, 2009) (Figure 1, Table 1). However, an EST with extensive similarity (E-value = e-129) to *Antirrhinum majus* benzaldehyde dehydrogenase (AmBALDH) (Long *et al.*, 2009) was identified, suggesting that benzoic acid biosynthesis in khat is more similar to the pathway in snapdragon petals than in Arabidopsis seed since AmBALDH and AtAAO4 likely catalyze the same reaction in non-β-oxidative benzoic acid metabolism. Highly homologous ESTs were also identified using At4CL1 and PhKAT1 as queries suggesting that a

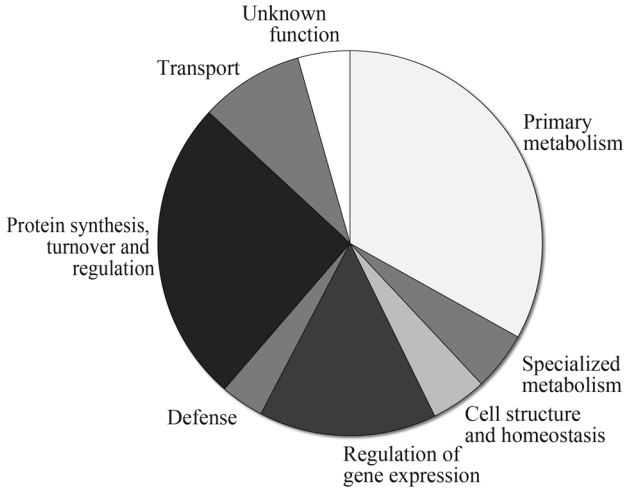


Figure 3 - Functional categorization of expressed sequence tags (ESTs) from *Catha edulis* leaf-derived cDNA library. Assignments were made based on homology to proteins of known function, as evidenced by tBLASTn search results using TAIR Proteins and UniProt Plants databases. ESTs with homology to uncharacterized, putative, or hypothetical proteins (*i.e.* unknown function) comprised 4.4% of the total population.

β-oxidative, CoA-dependent pathway leading to benzoyl-CoA production might also occur in khat leaves (Table 1). An alternative pathway operative in lactic acid bacteria circumvents the PAL-catalyzed production of *trans*-cinnamic acid. In this case, phenylpyruvate, a transamination product of Phe, serves as a precursor to benzaldehyde (Nierop-Groot and de Bont, 1999). An *Arabidopsis* transaminase producing phenylpyruvate from Phe was recently characterized (Prabhu and Hudson, 2010). However, no close homologues were found among khat ESTs.

Beyond benzoic acid biosynthesis, the reactions leading from the formation of 1-phenylpropane-1,2-dione to (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine are not well understood. The recruitment of a ThDP-dependent enzyme for the carboligation of pyruvate with a benzoyl derivative has been proposed (Grue-Sørensen and Spenser, 1989), although the involvement of such an enzyme in ephedrine alkaloid biosynthesis has not been demonstrated. Two ThPD-dependent enzymes isolated from microbes, acetohydroxyacid synthase (AHAS) and pyruvate decarboxylase (PDC), catalyze the conversion of benzaldehyde to (*R*)-phenylacetylcarbinol (Figure 2), an intermediate in the semi-synthetic production of ephedrine alkaloids (Engel *et al.*, 2003; Meyer *et al.*, 2011). In addition to (*R*)-phenylacetylcarbinol, mutant *Zymomonas mobilis* PDCs catalyze the formation of (*S*)-phenylacetylcarbinol, along with *R* and *S* forms of 2-hydroxypropiophenone (Pohl *et al.*, 1998) (Figure 2). The possibility that khat possesses a PDC enzyme with similar catalytic flexibility must also be considered. Enzymatic and molecular characterization of this carboligation step will be necessary to unequivocally establish the biosynthetic precursors of (*S*)-cathinone.

The potential involvement of benzoyl-CoA as a precursor to (*S*)-cathinone has been suggested (Grue-Sørensen and Spenser, 1988, 1989). A ThPD-dependent enzyme could catalyze a carboligation reaction between the benzoyl moiety of benzoyl-CoA and the C_2-C_3 component of pyruvate (Supplemental Figure S1). Similar to reaction schemes proposed for ThPD-dependent enzymes such as AHAS (Jaña *et al.*, 2010, Engel *et al.*, 2003) and PDC, the decarboxation of pyruvate yields a hydroxyethyl-thiamin diphosphate anion/enamine intermediate that would attack the carbonyl carbon of benzoyl-CoA to initiate condensation, release of a CoASH leaving group and the formation of 1-phenylpropane-1,2-dione. A similar, but not identical reaction mechanism involving benzoic acid in lieu of benzoyl-CoA is also possible whereby acid-catalyzed protonation at the carbonyl oxygen would permit nucleophilic attack by the anion/enamine intermediate.

Searching the khat EST library revealed three candidate sequences with homology to ThPD-dependent enzymes putatively involved in the formation of 1-phenylpropane-1,2-dione (Table 1). Unigenes 017_C06-044 and 034_C01-011 annotated as AHAS, reflecting their close homology with characterized plant AHAS enzymes (76%

Table 1 - *Catha edulis* ESTs (CeUniGenes) representing enzymes putatively involved in phenylpropylamino alkaloid biosynthesis. Abbreviations: AAO4, aldehyde oxidase-4; AAT, alanine aminotransferase; ALS, acetolactate synthase; AONS, 8-amino-7-oxononanoate synthase; BALDH, benzaldehyde dehydrogenase; BCAT, branched-chain amino acid aminotransferase; BZO1, benzoate-CoA ligase; 4CL, 4-coumaroyl-CoA ligase; CR, carbonyl reductase; GABA-T, 4-aminobutyrate transaminase; GSA-AT, glutamate-1-semialdehyde aminotransferase; HPA1, histidinol phosphate aminotransferase; KAT1, 3-ketoacyl-CoA thiolase-1; KAR, 3-ketoacyl-CoA reductase; PAL, phenylalanine ammonia lyase; PDC, pyruvate decarboxylase; PORA, protochlorophyllide reductase A; PSAT, phosphoserine aminotransferase; SDR, short-chain dehydrogenase/reductase; SGAT, serine-glyoxylate aminotransferase; TR, tropinone reductase.

CeUniGenes exhibiting sequence similarity to enzymes implicated in plant benzoic acid metabolism				
Enzyme	GenBank accession number and species	Putative activity	CeUniGene ID	E value
PAL1	P35510 *Arabidopsis thaliana*	PAL	036_E05-039	1.00E-142
4CL1	Q42524 *Arabidopsis thaliana*	CoA ligase	CL455Contig1	1.00E-102
			028_E04-024	2.00E-98
KAT1	ACV70032 *Petunia hybrida*	Thiolase	011_D07-057	1.00E-116
			018_F11-085	3.00E-98
BZO1	NP_176763 *Arabidopsis thaliana*	CoA ligase	047_C08-060	8.00E-71
AAO4	NP_563711 *Arabidopsis thaliana*	Dehydrogenase	No hit	N/A
BALDH	ACM89738 *Antirrhinum majus*	Dehydrogenase	044_C05-043	1.00E-129
			034_A10-080	1.00E-86
CeUniGenes annotating as enzymes putatively catalyzing key reactions in alkaloid metabolism				
ALS	Q42768 *Gossypium hirsutum*	carboligation	017_C06-044	1.00E-121
ALS	Q5VB44 *Helianthus annuus*	carboligation	034_C01-011	2.00E-64
PDC	Q9FVF0 *Fragaria ananassa*	carboligation	049_G05-035	1.00E-114
AAT	AT1G70580 *Arabidopsis thaliana*	transamination	CL166Contig1	0
AONS	AT5G04620 *Arabidopsis thaliana*	transamination	CL12Contig1	1.00E-112
GSA-AT	Q84TK5 *Brassica napus*	transamination	021_A12-096	9.00E-94
SGAT	O49124 *Fritillaria agrestis*	transamination	004_F01-005	1.00E-60
PSAT	AT4G35630 *Arabidopsis thaliana*	transamination	045_G09-067	8.00E-73
HPA1	AT5G10330 *Arabidopsis thaliana*	transamination	003_C07-059	9.00E-53
BCAT	Q9SNY8 *Solanum tuberosum*	transamination	029_B03-029	1.00E-81
GABA-T	Q6ZCF0 *Oryza sativa*	transamination	032_C06-044	1.00E-101
PORA	Q41249 *Cucumis sativus*	reduction	CL62Contig1	0
TR-like	AT5G06060 *Arabidopsis thaliana*	reduction	CL440Contig1	1.00E-111
TR-like	AT5G06060 *Arabidopsis thaliana*	reduction	012_H02-002	2.00E-47
SDR	AT3G50560 *Arabidopsis thaliana*	reduction	041_B01-013	3.00E-99
TR-like	ABG22472 *Oryza sativa*	reduction	030_B02-014	7.00E-96
KAR	Q0VH86 *Gossypium hirsutum*	reduction	015_A03-031	6.00E-74
SDR	AT3G26770 *Arabidopsis thaliana*	reduction	046_G06-036	2.00E-53
SDR	Q6DLW2 *Solanum tuberosum*	reduction	014_B06-046	7.00E-83
SDR	AT3G06060 *Arabidopsis thaliana*	reduction	017_D07-057	1.00E-100
CR-like	Q9FI45 *Arabidopsis thaliana*	reduction	039_C07-059	5.00E-25
SDR	Q6DLW2 *Solanum tuberosum*	reduction	020_F10-070	2.00E-54
SDR	AT1G10310 *Arabidopsis thaliana*	reduction	038_H11-081	1.00E-104

and 48%, respectively, compared with the catalytic subunit of *Arabidopsis thaliana* AHAS). The remaining candidate was most similar to PDC. Following carboligation, 1-phenylpropane-1,2-dione undergoes transamination to form the neurostimulant, (*S*)-cathinone. EST analysis yielded a number of putative transaminase/amino transferase candidates (Table 1). Two unigenes (CL166Contig1 and CL12Contig1) were comprised of more than one EST, suggesting a higher expression level than other transaminases. Finally, candidate ESTs possibly involved in the stereospecific reduction of (*S*)-cathinone to either (1*S*,2*S*)-cathine or (1*R*,2*S*)-norephedrine are listed in Table 1. The stereospecific reduction of keto groups to alcohols has been documented in the specialized metabolism of other plants,

such as peppermint (*Mentha* x *piperita*) (Davis *et al.*, 2005), opium poppy (*Papaver somniferum*) and black henbane (*Hyoscyamus niger*) (Ziegler and Facchini, 2008). In each case, reduction to a specific stereoisomer occurs via an enzyme belonging to the short chain dehydrogenase/reductase (SDR) protein family. Interestingly, a bacterial SDR protein was found to reduce *N*-methylated (*S*)-cathinone to (1*S*,2*S*)-pseudoephedrine, but not to (1*R*,2*S*)-ephedrine (Kataoka *et al.*, 2006, 2008), which supports the hypothesis that two distinct SDR enzymes are involved in the formation of (1*S*,2*S*)-cathine and (1*R*,2*S*)-norephedrine, respectively (Krizevski *et al.*, 2010).

Despite its long history, khat has recently become a controversial plant and is regulated, along with its phenylpropylamino alkaloid constituents, as a controlled substance in many Western countries. In contrast, several phenylpropylamino alkaloids are widely available and have a variety of health applications. The biosynthesis of phenylpropylamino alkaloids in khat begins with L-phenylalanine and requires 8-10 steps to yield (1*S*,2*S*)-cathine and its diastereomer (1*R*,2*S*)-norephedrine. Although some steps in benzoic acid metabolism have been recently resolved in Arabidopsis and other plants, the biochemistry of phenylpropylamino alkaloid metabolism in khat remains largely uncharacterized. The annotated EST library provides a snapshot of the khat young leaf transcriptome and establishes a valuable resource for phenylpropylamino alkaloids biosynthetic gene discovery. Candidate cDNAs encoding enzymes that putatively catalyze each step of the pathway were identified, which provides a genomics platform essential for their future characterization.

Acknowledgments

We thank Dustin Cram and Jacek Nowak at the National Research Council – Plant Biotechnology Institute (Saskatoon, Canada) for the bioinformatics analysis. This work was supported through grants from the Binational Agricultural Research and Development Fund (CA-9117-09), and by the Israel Science Foundation (814/06). DNA sequencing and bioinformatics were performed through the Natural Products Genomics Resource (NAPGEN). K. Kilpatrick is co-supervised by Norman P.A. Hüner (Dept. Biology, University of Western Ontario). P.J.F. holds the Canada Research Chair in Plant Metabolic Processes Biotechnology.

References

Abd El-Mawla AMA and Beerhues L (2002) Benzoic acid biosynthesis in cell cultures of *Hypericum androsaemum*. Planta 214:727-733.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Balint EE, Falkay G and Balint GA (2009) Khat – A controversial plant. Wien Klin Wochenschr 121:604-614.

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, *et al.* (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol 135:745-755.

Boatright J, Negre F, Chen X, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D and Dudareva N (2004) Understanding *in vivo* benzenoid metabolism in petunia petal tissue. Plant Physiol 135:1993-2011.

Bredholt T, Dimba EAO, Hagland HR, Wergeland L, Skavland J, Fossan KO, Tronstad KJ, Johannessen AC, Vintermyr OK and Gjertsen BT (2009) Camptothecin and khat (*Catha edulis* Forsk.) induced distinct cell death phenotypes involving modulation of c-FLIP$_L$, McI-I, procaspase-8 and mitochondrial function in acute myeloid leukemia cell lines. Mol Cancer 8:e101.

Chou H-H and Holmes MH (2001) DNA sequence quality trimming and vector removal. Bioinformatics 17:1093-1104.

Colzato LS, Ruiz MJ, van den Wildenberg WPM and Hommel B (2011) Khat use is associated with impaired working memory and cognitive flexibility. PLoS One 6:e20602.

Costelloe SJ, Ward JM and Dalby PA (2008) Evolutionary analysis of the TPP-dependent family. J Mol Evol 66:36-49.

Davis EM, Ringer KL, McConkey ME and Croteau R (2005) Monoterpene metabolism. Cloning, expression, and characterization of menthone reductases from peppermint. Plant Physiol 137:873-881.

Engel S, Vyazmensky M, Geresh S, Barak Z and Chipman DM (2003) Acetohydroxyacid synthase: A new enzyme for chiral synthesis of *R*-phenylacetylcarbinol. Biotechnol Bioengineer 83:833-840.

Ewing B, Hillier L, Wendl MC and Green P (1998) Base-calling of automated sequencer traces using phred I. Accuracy assessment. Genome Res 8:175-185.

Gebissa E (2010) Khat in the Horn of Africa: Historical perspectives and current trends. J Ethnopharmacol 132:607-614.

Gonda I, Bar E, Portnoy V, Lev S, Burger J, Schaffer AA, Tadmor Y, Gepstein S, Giovannoni JJ, Katzir N, *et al.* (2010) Branched-chain and aromatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. fruit. J Exp Bot 61:1111-1123.

Green JBA (1989) Pyruvate decarboxylase is like acetolactate synthase (ILV2) and not like the pyruvate dehydrogenase E1 subunit. FEBS Lett 246:1-5.

Grue-Sørensen G and Spenser ID (1988) Biosynthesis of ephedrine. J Am Chem Soc 110:3714-3715.

Grue-Sørensen G and Spenser ID (1989) The biosynthesis of ephedrine. Can J Chem 67:998-1009.

Grue-Sørensen G and Spenser ID (1994) Biosynthetic route to the *Ephedra* alkaloids. J Am Chem Soc 116:6195-6200.

Ibdah M, Chen Y-T, Wilkerson CG and Pichersky E (2009) An aldehyde oxidase in developing seeds of *Arabidopsis* converts benzaldehyde to benzoic acid. Plant Physiol 150:416-423.

Jãna G, Jiménez V, Villà-Freixa J, Prat-Resina X, Delgado E and Alderete J (2010) Computational study on the carboligation reaction of acetohydroxyacid synthase: New approach on the role of the HEthDP$^-$ intermediate. Proteins 78:1774-1788.

Kassie F, Darroudi F, Kundi M, Schulte-Hermann R and Knasmüller S (2001) Khat (*Catha edulis*) consumption causes genotoxic effects in humans. Int J Cancer 92:329-332.

Kataoka M, Nakamura Y, Urano N, Ishige T, Shi G, Kita S, Sakamoto K and Shimizu S (2006) A novel NADP+ depend-

ent L-1-amino-2-propanol dehydrogenase from *Rhodococcus erythropolis* MAK154: A promising enzyme for the production of double chiral amino alcohols. Lett Appl Microbiol 43:430-435.

Kataoka M, Ishige T, Urano N, Nakamura Y, Sakuradani E, Fukui S, Kita S, Sakamoto K and Shimizu S (2008) Cloning and expression of the L-1-amino-2- propanol dehydrogenase gene from *Rhodococcus erythropolis*, and its application to double chiral compound production. Appl Microbiol Biotechnol 80:597-604.

Klein A, Beckerleg S and Hailu D (2009) Regulating khat – dilemmas and opportunities for the international drug control system. Intl J Drug Policy 20:509-513.

Kliebenstein DJ, D'Auria JC, Behere AS, Kim JH, Gunderson KL, Breen JN, Lee G, Gershenzon J, Last RL and Jander G (2007) Characterization of seed-specific benzoyloxyglucosinolate mutations in *Arabidopsis thaliana*. Plant J 51:1062-1076.

Krizevski R, Dudai N, Bar E and Lewinsohn E (2007) Developmental patterns of phenylpropylamino alkaloids accumulation in khat (*Catha edulis*, Forsk.) J Ethnopharmacol 114:432-438.

Krizevski R, Dudai N, Bar E, Dessow I, Ravid U and Lewinsohn E (2008) Quantitative stereoisomeric determination of phenylpropylamino alkaloids in khat (*Catha edulis* Forsk.) using chiral GC-MS. Israel J Plant Sci 56:207-213.

Krizevski R, Bar E, Shalit O, Sitrit Y, Ben-Shabat S and Lewinsohn E (2010) Composition and stereochemistry of ephedrine alkaloids accumulation in *Ephedra sinica* Stapt. Phytochemistry 71:895-903.

Leete E (1958) Biogenesis of δ-norpseudoephedrine in *Catha edulis*. Chem Ind 33:1088-1089.

Long MC, Nagegowda DA, Kaminaga Y, Ho KK, Kish CM, Schnepp J, Sherman D, Weiner H, Rhodes D and Dudareva N (2009) Involvement of snapdragon benzaldehyde dehydrogenase in benzoic acid biosynthesis. Plant J 59:256-265.

Mateen FJ and Cascino GD (2010) Khat chewing: A smokeless gun? Mayo Clinic Proc 85:971-973.

Meyer D, Walter L, Kolter G, Pohl M, Müller M and Tittmann K (2011) Conversion of pyruvate decarboxylase into an enantioselective carboligase with biosynthetic potential. J Am Chem Soc 133:3609-3616.

Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsyn AA, Broveak TR and Hide WA (1999) A comprehensive approach to clustering of expressed human gene sequence: The sequence tag alignment and consensus knowledge base. Genome Res 9:1143-1155.

Murata J, Brandle JE, Sensen CW and De Luca V (2006) Expressed sequenced tags from Madagascar periwinkle (*Catharanthus roseus*). FEBS Lett 580:4501-4507.

Nierop-Groot MN and de Bont JAM (1999) Involvement of manganese in conversion of phenylalanine to benzaldehyde by lactic acid bacteria. Appl Environ Microbiol 65:5590-5593.

Okada T, Mikage M and Sekita S (2008) Molecular characterization of the phenylalanine ammonia-lyase from *Ephedra sinica*. Biol Pharm Bull 31:2194-2199.

Prabhu PR and Hudson AO (2010) Identification and partial characterization of an L-tyrosine aminotransferase (TAT) from *Arabidopsis thaliana*. Biochem Res Int 2010:549572.

Pohl M, Siegert P, Mesch K, Bruhn H and Grötzinger J (1998) Active site mutants of pyruvate decarboxylase from *Zymomonas mobilis*. Eur J Biochem 257:538-546.

Schilmiller AL, Stout J, Weng J-K, Humphreys J, Ruegger MO and Chapple C (2009) Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. Plant J 60:771-782.

Thomas BC, Rapaka L, Lyons E, Pedersen B and Freeling M (2007) *Arabidopsis* intragenomic conserved noncoding sequence. Proc Natl Acad Sci USA 104:3348-3353.

United Nations (1975) Étude sûr la composition chimique du khat: Recherche sûr la fraction phenylalkylamine. UN document MNAR/11/1975.

Van Moerkercke A, Schauvinhold I, Pichersky E, Haring MA and Schuurink RC (2009) A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. Plant J 60:292-302.

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, *et al.* (2011) Gene expression analysis of flax seed development. BMC Plant Biol 11:e74.

Yamasaki K, Sankawa U and Shibata S (1969) Biosynthesis of ephedrine in *Ephedra*: Participation of a C_6-C_1 unit. Tetrahedron Lett 47:4099-4102.

Yamasaki K, Tamaki T, Uzawa S, Sankawa U and Shibata S (1973) Participation of C_6-C_1 unit in the biosynthesis of ephedrine in *Ephedra*. Phytochemistry 12:2877 2882.

Ziegler J and Facchini PJ (2008) Alkaloid biosynthesis: Metabolism and trafficking. Annu Rev Plant Biol 59:735-769.

Supplementary Material

The following online material is available for this article:

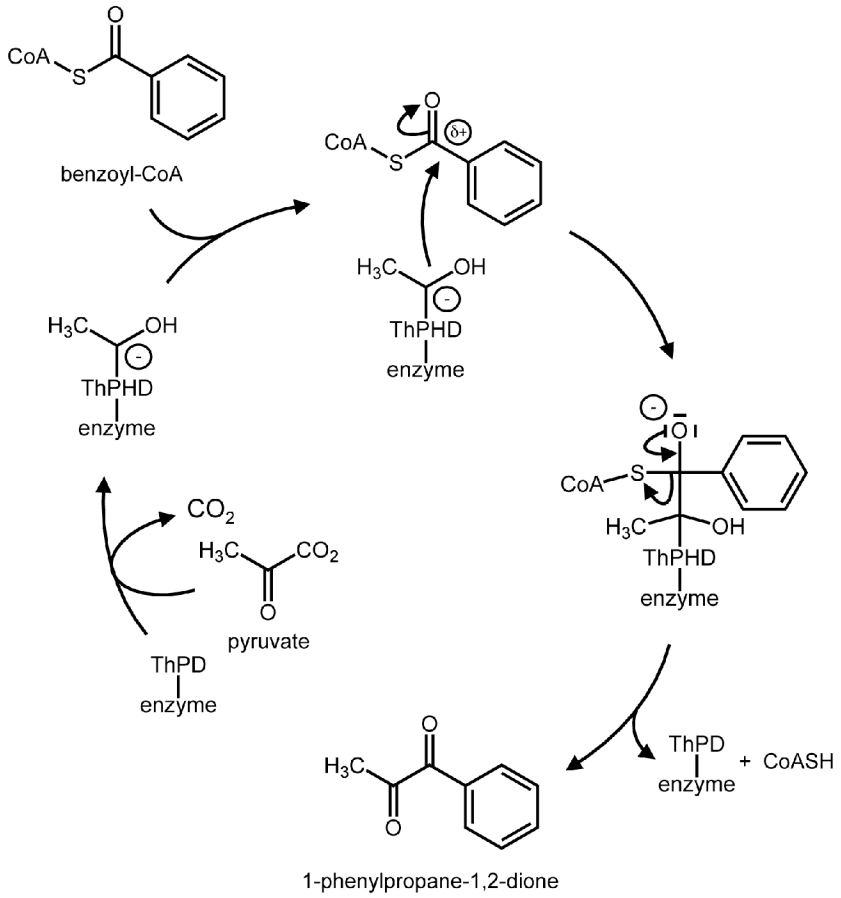
Figure S1 - Proposed reaction mechanism outlining the formation of 1-phenylpropane-1,2-dione from benzoyl-CoA and pyruvate precursors.

Table S1 - Expressed sequence tag analysis summary for the *Catha edulis* leaf library.

Table S2 - Proportion of *Catha edulis* ESTs assigned to various functional categories as defined by the GO Consortium.

This material is part of the online article from http://www.scielo.br/gmb.

Associate Editor: Márcia Pinheiro Margis



Supplemental Figure S1 - Proposed reaction mechanism outlining the formation of 1-phenylpropane-1,2-dione from benzoyl-CoA and pyruvate precursors. The decarboxylation of pyruvate by a ThPD-dependent enzyme yields a hydroxyethyl-thiamin diphosphate anion/enamine intermediate (only anion shown), which initiates nucleophilic attack at the carbonyl carbon of benzoyl-CoA. The resulting tetrahedral adduct undergoes rearrangement, releasing a CoA leaving group and generating 1-phenylpropane-1,2-dione product. Refer to text for details.

Supplemental Table S1 - Expressed sequence tag analysis summary for the *Catha edulis* leaf library.

Standard EST library		Cluster analysis		BLAST results for unigenes	
Total sequences	4723	Total unigenes	3293	Hits against TAIR database	3272
Average length (bp)	710	Contigs	633	Hits against UniProt database	3237
Average GC (%)	44.5	Singletons	2660	Unigenes receiving GO annotation	2839

Supplemental Table S2 - Proportion of *Catha edulis* ESTs assigned to various functional categories as defined by the GO Consortium (http://www.geneontology.org/). The proportion of ESTs in each category is expressed as a % of the total ESTs receiving a GO annotation.

Biological processes GO annotation category	Total (%)	Cellular component GO annotation category	Total (%)	Molecular function GO annotation category	Total (%)
Metabolic process	34.9	Cell and parts	64.9	Catalytic activity	45.6
Cellular process	30.9	Organelle and parts	15.5	Binding	34.4
Response to stimulus	8.5	Macromolecular complex	14.7	Structural molecule activity	7.0
Biological regulation	5.1	Envelope	2.3	Transcription regulator activity	5.8
Localization	4.6	Membrane-enclosed lumen	2.1	Transporter activity	5.6
Establishment of localization	4.6	Extracellular region	0.5	Enzyme regulator activity	1.4
Developmental process	3.3	Extracellular matrix	0.1	Translation regulator activity	1.4
Multicellular organismal process	2.6			Molecule transducer activity	0.7
Reproduction	1.6			Antioxidant activity	0.7
Reproductive process	1.5			Motor activity	0.4
Multi-organism process	1.2			Nutrient reservoir activitly	0.2
Immune system process	0.7			Metallochaperone activity	0.04
Growth	0.2				
Rhythmic process	0.1				
Biological adhesion	0.1				
Pigmentation	0.05				
Total	100	Total	100	Total	100