

A Dynamic Interface for Capsaicinoid Systems Biology^{1[C][W][OA]}

Michael Mazourek*, Anuradha Pujar, Yelena Borovsky, Ilan Paran, Lukas Mueller, and Molly M. Jahn

Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853 (M.M.); SOL Genomics Network, Boyce Thompson Institute for Plant Research, Ithaca, New York 14853 (A.P., L.M.); Agricultural Research Organization, Volcani Center, Bet Dagan 50250, Israel (Y.B., I.P.); and College of Agriculture and Life Sciences, University of Wisconsin, Madison, Wisconsin 53706 (M.M.J.)

Capsaicinoids are the pungent alkaloids that give hot peppers (*Capsicum* spp.) their spiciness. While capsaicinoids are relatively simple molecules, much is unknown about their biosynthesis, which spans diverse metabolisms of essential amino acids, phenylpropanoids, benzenoids, and fatty acids. Pepper is not a model organism, but it has access to the resources developed in model plants through comparative approaches. To aid research in this system, we have implemented a comprehensive model of capsaicinoid biosynthesis and made it publicly available within the SolCyc database at the SOL Genomics Network (<http://www.sgn.cornell.edu>). As a preliminary test of this model, and to build its value as a resource, targeted transcripts were cloned as candidates for nearly all of the structural genes for capsaicinoid biosynthesis. In support of the role of these transcripts in capsaicinoid biosynthesis beyond correct spatial and temporal expression, their predicted subcellular localizations were compared against the biosynthetic model and experimentally determined compartmentalization in *Arabidopsis* (*Arabidopsis thaliana*). To enable their use in a positional candidate gene approach in the Solanaceae, these genes were genetically mapped in pepper. These data were integrated into the SOL Genomics Network, a clade-oriented database that incorporates community annotation of genes, enzymes, phenotypes, mutants, and genomic loci. Here, we describe the creation and integration of these resources as a holistic and dynamic model of the characteristic specialized metabolism of pepper.

The characteristic burning sensation produced by hot peppers (*Capsicum* spp.) is caused by capsaicinoids, alkaloid compounds that are synthesized and accumulate in pepper fruit (Nelson and Dawson, 1923). Beyond use as a vegetable or colorant, use as a spice is the major trait for which peppers are known and has driven the global dispersal of this New World crop (Govindarajan, 1985). Ecologically, a function of capsaicinoids in pepper is to deter frugivory by mammals by binding a thermoreceptor, TRPV1, found on nociceptive nerve fibers, thereby creating a sensation of burning pain (Caterina et al., 1997; Tewksbury and Nabhan, 2001). Additionally, capsaicinoids have been found to exhibit antimicrobial properties (Billing and Sherman, 1998; Tewksbury et al., 2008)

A combinatorial biosynthesis is responsible for capsaicinoid diversity. The structure of capsaicin, the

predominant form of the molecule, was solved in 1923 (Nelson and Dawson, 1923; Fig. 1, structure 1a). In 1968, Bennett and Kirby showed that a mixture of several forms of capsaicin (capsaicinoids) was found in hot peppers. These capsaicinoids were shown to share an aromatic portion (Fig. 1, structure 1) derived from Phe as per the contemporary understanding of the phenylpropanoid pathway (Bennett and Kirby, 1968). Leete and Loudon and other groups demonstrated that variable acyl moieties (Fig. 1, structures a–i) were derived from catabolism of amino acids that were subsequently elongated by the fatty acid synthase (Leete and Loudon, 1968; Kopp and Jurenitsch, 1981; Suzuki and Iwai, 1984; Markai et al., 2002), a biosynthesis similar to that known for the membrane lipids of *Bacillus subtilis*. Congeners are found where vanillyl alcohol (Fig. 1, structures 2a–2c) or coniferyl alcohol (Fig. 1, structures 3a and 3b) comprise the aromatic moiety of capsaicinoids in place of vanillylamine, rendering the molecule nonpungent (Kobata et al., 1998, 1999, 2008). Variation in the acyl moiety affects the degree and quality of the pungent sensation (Todd et al., 1977; Krajewska and Powers, 1988; Walpole et al., 1993). Capsaicin and dihydrocapsaicin (Fig. 1, structures 1a and 1b) differ only by the saturation of the acyl moiety and are equally the most potent of the capsaicinoids; changes in the length or branching of the fatty acid from what is found in these two capsaicinoids decrease the potency of the molecule. Modified capsaicinoids have been reported

¹ This work was supported by the National Science Foundation (grant nos. 0116076 and 0417056) and by a gift from Kalsec, Inc.

* Corresponding author; e-mail mm284@cornell.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Michael Mazourek (mm284@cornell.edu).

^[C] Some figures in this article are displayed in color online but in black and white in the print edition.

^[W] The online version of this article contains Web-only data.

^[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.109.136549

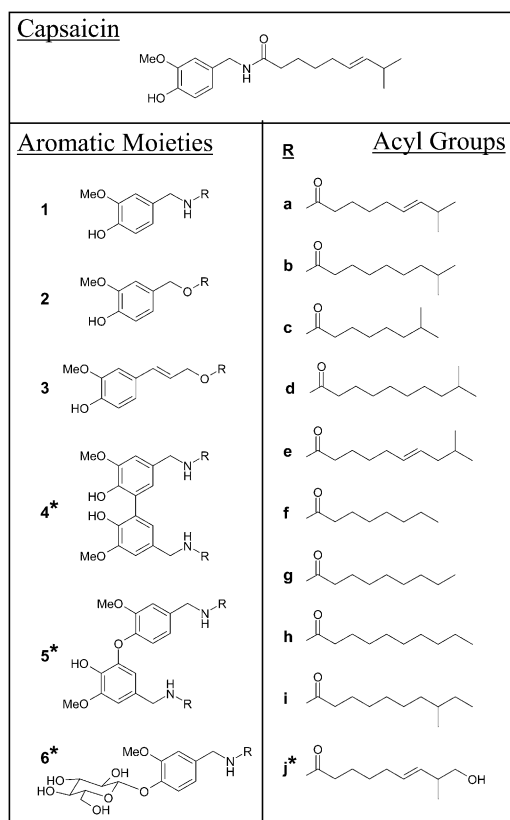


Figure 1. Capsaicinoids and congeners. Capsaicin is most commonly the most abundant capsaicinoid in hot peppers. Other capsaicinoids have been reported with aromatic group 1 paired with various acyl groups (a–j). Ester forms of the molecule have also been found (2a–2c, 3a, and 3b). Other reported capsaicinoids are the result of capsaicin being oxidized (4a and 5a), glycosylated (6a and 6b), or hydroxylated (1j), each of which reduces or eliminates pungency. These compounds are marked with asterisks. The discovery of capsaicinoids 1a to 1i has been reviewed by Suzuki and Iwai (1984). References to the other compounds are found in the text.

where either the aromatic ring is glycosylated (Fig. 1, structures 6a and 6b) or the acyl group is oxidized (Fig. 1, structure 1j), yielding metabolites that are similar to detoxification products reported for capsaicinoids in animal systems (Ochi et al., 2003; Higashiguchi et al., 2006; Reilly and Yost, 2006). The aromatic rings can also be oxidatively coupled between pairs of compounds (Fig. 1, structures 4a and 5a), similar to the bonding found between lignin monomers (Diaz et al., 2004).

The SOL Genomics Network (SGN; <http://www.sgn.cornell.edu>) is a central hub that integrates genomic and biochemical pathway data for the Solanaceae research community. The SGN Web site houses both community-contributed and curated maps and sequence information as well as tools to help researchers link these genomic data to the phenome (Mueller et al., 2005). A component of this resource is SolCyc, a metabolic pathway database for the Solana-

ceae that helps researchers navigate small molecule metabolism, by representing chemical structures, enzymes, cofactors, reactions, and pathways, and linking them with the other SGN resources, such as loci, accessions, and phenotypes (Menda et al., 2008). SGN's clade-oriented approach integrates data from several related organisms, which provides the evolutionary context to study the origins of the biochemical variation in the background of conserved solanaceous genomes.

The creation of a metabolic resource for CapCyc, the *Capsicum*-specific database within SolCyc, is critical to structure biological studies of capsaicinoid biosynthesis around metabolic information that was scattered throughout the literature. While progress has been made in the identification and analysis of genes and proteins found in the historical conceptualization of capsaicinoid biosynthesis (Curry et al., 1999; Kim et al., 2001; Aluru et al., 2003; Lee et al., 2006), the overall capsaicinoid metabolic scheme has not been updated in light of revisions of related pathways in other organisms or considered in terms of its cellular context (Fig. 2). Results from heterologous systems are especially useful in *Capsicum* because it is not a model organism, as judged by the common criteria of generation time, genome size, and molecular resources. The SGN database aids the ability to capitalize on *Capsicum* being a member of the Solanaceae, the most developed comparative genetic system in the dicotyledonous plants. Comparative linkage maps (Livingstone et al., 1999), genetic marker sets (Wu et al., 2006), genes controlling quantitative traits (Zygiel et al., 2005), proteomics data (Rose et al., 2004), microarrays (Alba et al., 2004), EST libraries, and the in-progress genome sequence are all resources being developed in tomato (*Solanum lycopersicum*) that can also be utilized by the *Capsicum* community.

In addition to genomic resources, there is also shared biochemistry among genera of the Solanaceae. Medium-length, branched-chain fatty acids are an unusual metabolite in plants but common in the Solanaceae. Most prominently, they are found as sugar esters in exudates from glandular trichomes that cover the aerial epidermis of plants of the *Nicotiana*, *Datura*, *Petunia*, and *Solanum* genera (Severson et al., 1985; King et al., 1987, 1990; King and Calhoun, 1988; Chortyk et al., 1997). *Capsicum* is not known to have glandular trichomes, but it has glandular regions in the fruit that accumulate branched fatty acids as capsaicinoids (Suzuki and Iwai, 1984; Zamski et al., 1987; Stewart et al., 2007; Fig. 2). Across the Solanaceae, branched-chain fatty acids are derived from branched-chain amino acids via two different elongation mechanisms. *Nicotiana* and *Petunia* extend these primers with acetyl-CoA in one-carbon increments (α -ketoacid elongation) by reiterative cycles of the Leu biosynthetic pathway (Kandra et al., 1990; Kroumova et al., 1994; Kroumova and Wagner, 2003; Slocombe et al., 2008). In contrast, *Solanum*, *Datura*, and *Capsicum* have been shown to elongate the same substrates but

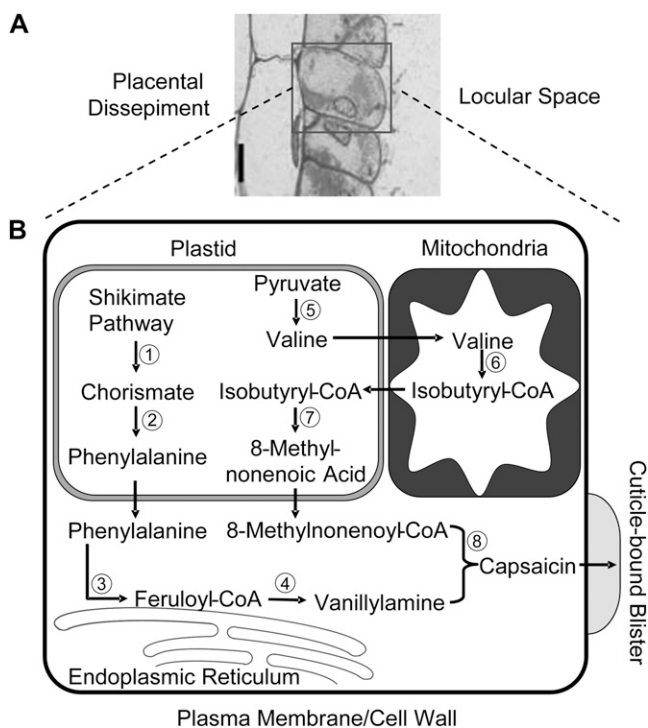


Figure 2. Overview of capsaicinoid biosynthesis in a cellular context. A, Capsaicinoid-producing cells from a hot pepper stained with crystal violet. Bar = 50 μm . This image is reproduced with permission from the Journal of Experimental Botany (Stewart et al., 2007). B, Capsaicinoid biosynthetic pathway modeled in this cell. Chorismate is produced from the shikimate pathway (step 1), which is used to synthesize Phe in the plastid (step 2). In the cytoplasm, associated with the endoplasmic reticulum, Phe is converted to feruloyl-CoA by phenylpropanoid metabolism (step 3) and is likely converted to vanillylamine by an unknown enzyme in an unknown compartment (step 4). Pyruvate is the precursor to Val (step 5), which is exported to the mitochondria and catabolized to isobutyryl-CoA (step 6). Isobutyryl-CoA returns to the plastid, where it is elongated to 8-methylnonenoic acid by the fatty acid synthase (step 7). Export from the plastid is concomitant with formation of the CoA thioester. The location and mechanism of the final condensation and export of capsaicin out of the cell (step 8) are debated. The biosynthesis of capsaicin is shown, other capsaicinoids are formed through variation in steps 5 to 7, and ester forms of the molecule are likely formed by variation in steps 3 and 4. [See online article for color version of this figure.]

in two-carbon increments derived from acetyl-CoA via the fatty acid synthase (Walters and Steffens, 1990; van der Hoeven and Steffens, 2000; Markai et al., 2002; Aluru et al., 2003; Slocombe et al., 2008).

Here, we describe the creation of a comprehensive SolCyc pathway for capsaicinoid biosynthesis. This pathway integrates all information to date on the subbranches of capsaicinoid biosynthesis that are elucidated in pepper and other systems as a testable model for the production of capsaicinoids. As a preliminary test of this hypothesis, we cloned the predicted transcripts from pepper tissue that was actively synthesizing capsaicinoids. These sequences were integrated and annotated in the SGN database and made

available as a comparative, freely accessible, and dynamic resource that will evolve in pace with progress in this area of research and be useful in comparative systems.

RESULTS

CapCyc Model

A testable model for capsaicinoid biosynthesis was developed based on the literature and metabolic databases that incorporates work in *Capsicum*, related genera, and model organisms (Fig. 2). The reactions and their subcellular localization form a metabolic hypothesis for capsaicinoid biosynthesis. This model combines phenylpropanoid and benzenoid metabolisms (Fig. 3) and medium-length, branched-chain fatty acid biosynthesis (Supplemental Fig. S1), which were classically considered as part of capsaicinoid biosynthesis, and further includes Phe (Supplemental Fig. S2) and branched-chain amino acid biosynthesis (Fig. 4; Suzuki and Iwai, 1984). Previously, amino acid synthesis was not included in the capsaicinoid biosynthetic pathway, because amino acids are considered "primary" metabolites and labeled amino acids fed into pepper stems are incorporated into capsaicinoids. However, these amino acids are not "transport" amino acids that are commonly found at high concentration in the phloem (Lea and Ireland, 1999). Furthermore, if we consider the abundance of fruit that can be produced by a single pepper plant and the high level of capsaicinoids that is possible per fruit in some accessions (more than 5% fruit dry weight; Bosland and Baral, 2007), it seems necessary to consider that the biosynthesis of these amino acids would be up-regulated within the capsaicinoid-producing tissue in order to supply a sufficient amount of precursors. In view of the substantial amount of capsaicinoids synthesized during a short developmental period (Iwai et al., 1979), we posit that the model should include not just the substrates but also their production and the regeneration of cofactors. Numerous such considerations and assumptions were part of this process and are detailed in Supplemental Text S1.

Capsaicin and its precursor pathways were added to the SolCyc database (Fig. 5) by adding the pathway to the MetaCyc reference database (Caspi et al., 2006) using the Pathway Tools Pathway Editor functions. Pathway Tools allows visualization of pathways in great detail as well as a display of the integration of individual pathways within the larger metabolic network of an organism in a Pathway Genome Database (PGDB; Caspi et al., 2006; Fig. 5). At present, SolCyc has PGDBs for tomato (LycCoCyc), potato (*Solanum tuberosum*; PotatoCyc), tobacco (*Nicotiana tabacum*; TobaccoCyc), *Petunia* (PetuniaCyc), and pepper (CapCyc), all derived from annotations of unigene sets (Mueller et al., 2005). PGDBs were initially computationally predicted based on sequence annotation and

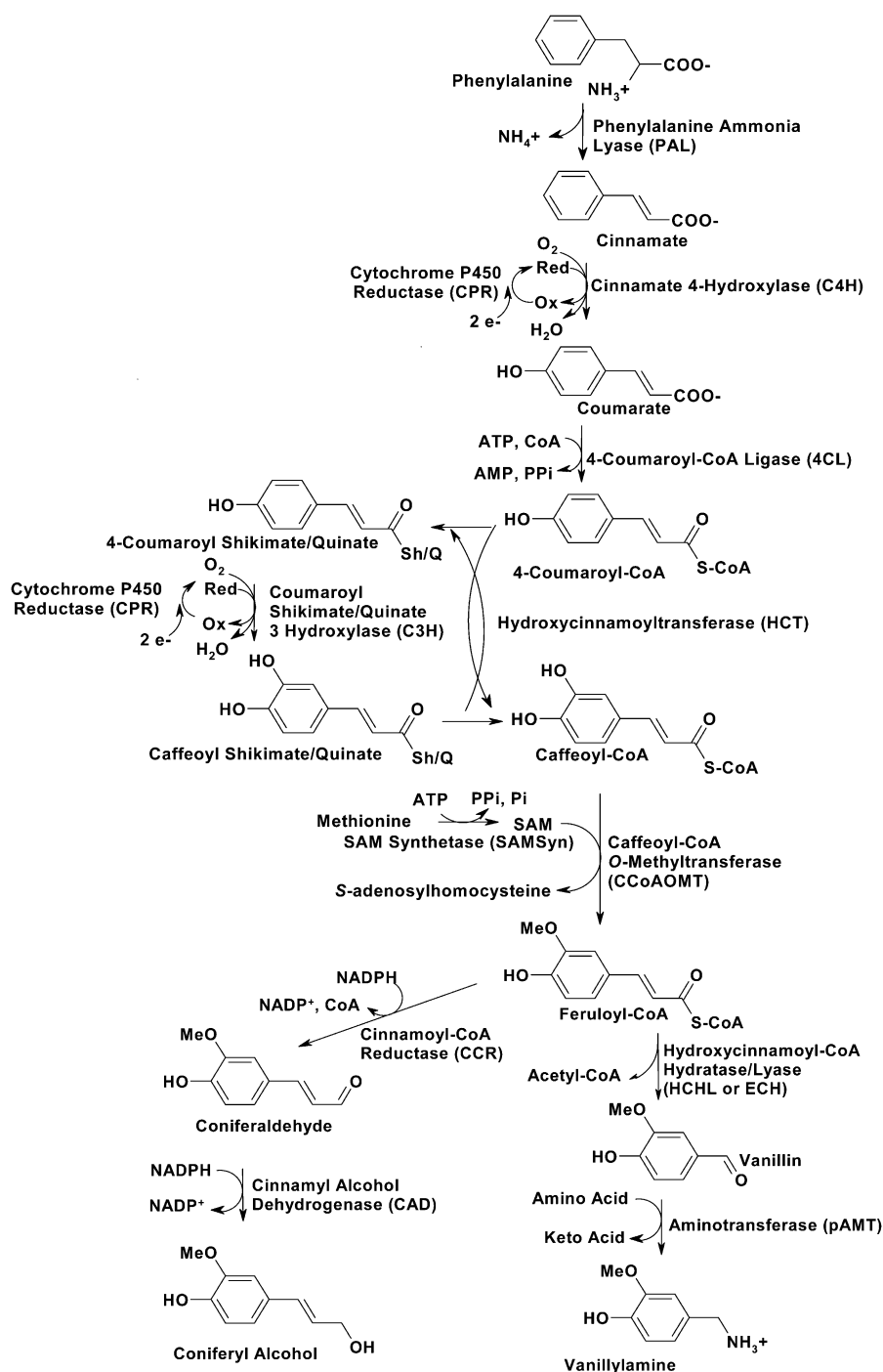


Figure 3. Phenylpropanoid and benzenoid metabolism. Phe is deaminated, hydroxylated, and methylated to produce feruloyl-CoA. The loss of a two-carbon unit and aminotransfer produce a vanillylamine, a direct precursor of capsaicin. Cofactors are to the left of the reaction arrows and enzymes are to the right. Ade, Adenine; e⁻, electron; SAM, S-adenosylmethionine; THF, tetrahydrofolate. Sources are as follows: Bennett and Kirby (1968), Rangoonwala (1969), Kopp et al. (1983), Johnson and Ravishankar (1996), Razal et al. (1996), Gasson et al. (1998), Curry et al. (1999), Lea and Ireland (1999), Achterholt et al. (2000), Wildermuth et al. (2001), Humphreys and Chapple (2002), Podstolski et al. (2002), Hoffmann et al. (2003, 2004), Walton et al. (2003), Winkel (2004), Orlova et al. (2006), and Merali et al. (2007).

the MetaCyc reference database. In CapCyc, the phenylpropanoid pathway was already predicted from this template, but Val degradation, the source of the fatty acid moiety of capsaicinoid molecules, had to be manually annotated based on experimentally verified information from the literature along with other reaction steps that were unique or proposed to be unique to *Capsicum* based on the literature synthesis (Supplemental Text S1). Information gathered for each locus included literature citations, EC numbers, SGN unig-

ene sequences, and Gene Ontology and Plant Ontology annotations.

Candidate Gene Sequences

As an initial test of this model, the genes predicted to participate in capsaicinoid biosynthesis were cloned from cDNA derived from capsaicinoid-producing hot pepper tissue at peak activity. Transcripts for previously described capsaicinoid biosynthetic genes have

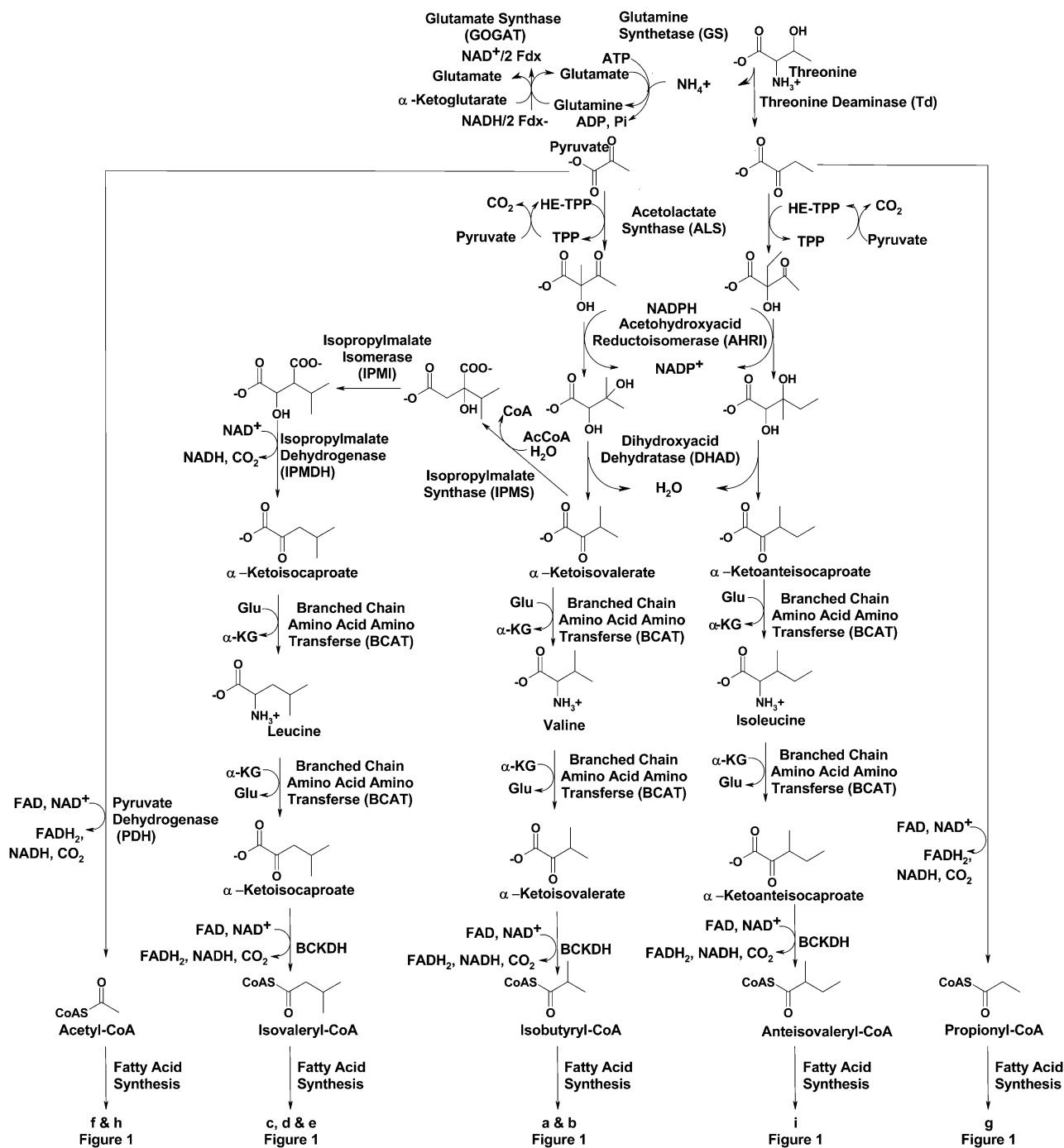


Figure 4. Branched-chain amino acid metabolism leading to capsaicinoids. Amino acids are synthesized in the plastids and then transported to the mitochondria, where they are deaminated and decarboxylated prior to import back into the plastid to serve as primers for fatty acid biosynthesis. Cofactors are to the left of the reaction arrows and products are to the right. α -KG, α -Ketoglutarate; HE-TPP, hydroxyethyl-thiamine pyrophosphate. Sources are as follows: <http://www.kegg.com>, Leete and Loudon (1968), Kopp and Jurenitsch (1981), Walters and Steffens (1990), van der Hoeven and Steffens (2000), Diebold et al. (2002), Li et al. (2003), and Slocombe et al. (2008).

been shown to be highly expressed at 20 DPA and enriched in the placental dissepiment, coinciding with capsaicinoid accumulation (Aluru et al., 2003; Stewart et al., 2005). Recovery of transcripts predicted by our

model from this tissue would be consistent with functions related to capsaicinoid biosynthesis.

Forty-two new transcripts were recovered, corresponding to candidates for 29 previously unaccounted

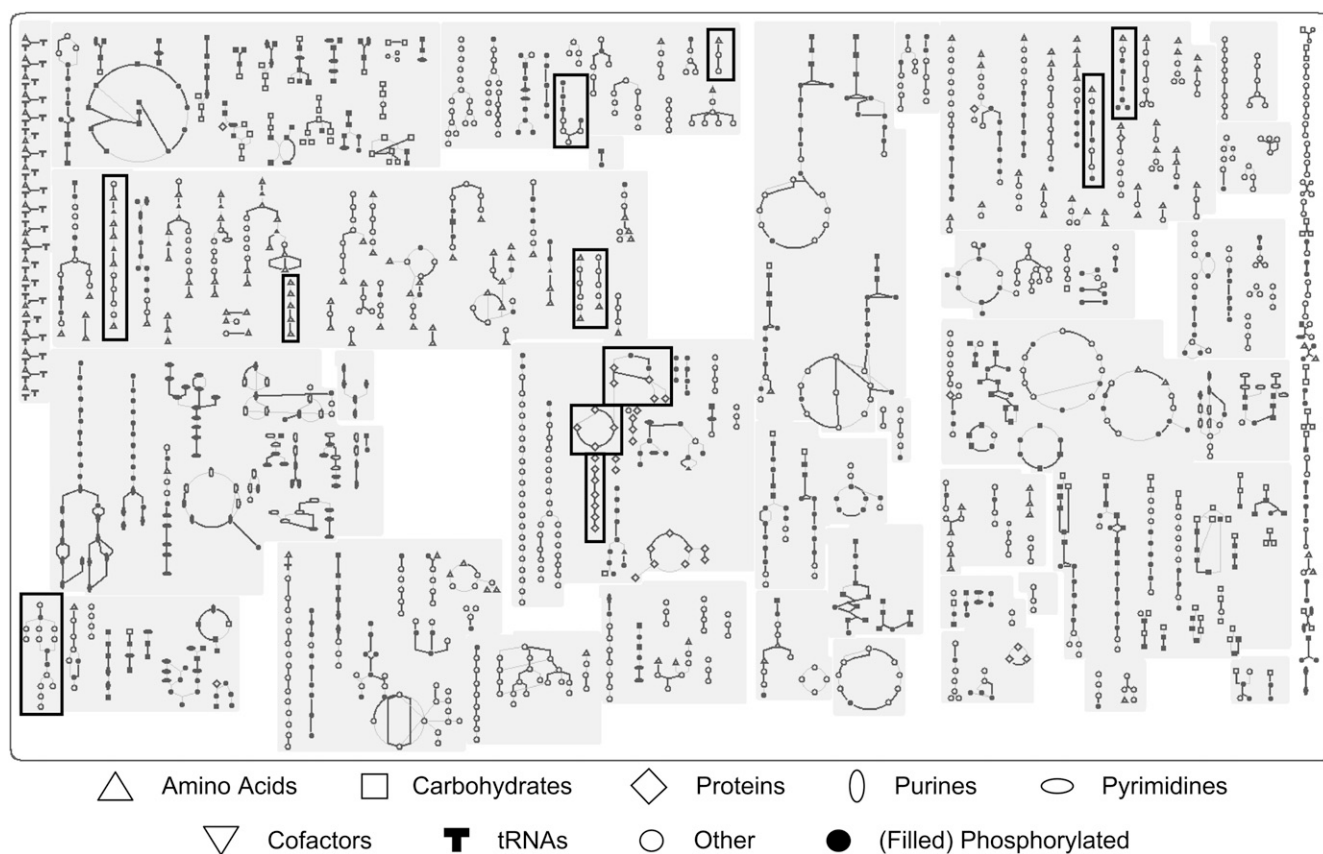


Figure 5. Overview of the *C. annuum* metabolic map (CapCyc). The diverse set of metabolisms that are involved in capsaicinoid biosynthesis are outlined among all of the annotated biosynthetic pathways of pepper. An interactive version is available online (<http://www.sgn.cornell.edu/tools/solcyc/>).

enzymatic steps in this new model of capsaicinoid biosynthesis (Table I). BLASTX searches of the *Arabidopsis* (*Arabidopsis thaliana*) proteome identified homologs for each cloned transcript with greater than 70% amino acid identity, on average, and increased to almost 80% identity after removing the predicted targeting peptides from the alignment. Most cloned transcripts were integrated seamlessly with our model, but there were some exceptions. For two genes, ketoacyl-ACP synthase III and the enoyl-ACP reductase component of the fatty acid synthase (KasIII and ENR; Supplemental Fig. S1), pairs of distinct sequences were recovered. Single nucleotide polymorphisms were also observed in some transcripts, albeit infrequently. The coding sequences of cinnamic acid 4-hydroxylase and Phe ammonia-lyase (C4H and PAL; Fig. 3) were sequenced in their entirety here because previously they were only known as partial sequences from *Capsicum chinense* (Curry et al., 1999). Only the 3' end of NADH-dependent Glu synthetase (NADH-GOGAT; Fig. 5; Supplemental Fig. S2) was obtained, and no transcripts were detected for Thr deaminase or dihydroxy acid dehydratase (TD and DHAD; Fig. 4). The plastid-encoded subunit of the acetyl-CoA carboxylase complex (ACCase β CT; Supplemental Fig.

S1) was cloned from genomic DNA. Two candidates for the acyl-CoA synthetase (Supplemental Fig. S1) that exports fatty acids from the plastid were cloned, both of which were more similar to *Arabidopsis* proteins other than the major plastid acyl-CoA exporter (Schnurr et al., 2002). An array of alternative candidates have been described that will be investigated in further iterations of our model (Shockey et al., 2003). A second homolog was cloned of the putative amino-transferase that is a candidate for catalyzing the formation of vanillylamine from vanillin (Fig. 3). These candidates for vanillylamine synthesis are questionable; pAMT and pAMT2 are the most similar to POP2 in *Arabidopsis* (Fig. 3). POP2 is a γ -aminobutyric acid transaminase that is implicated in pollen tube growth and also in response to stress and nitrogen storage and mobilization (Palanivelu et al., 2003; Miyashita and Good, 2008). One homolog of arogenate dehydratase was cloned from pepper and was most similar to one of the *Arabidopsis* candidates for this activity, ADT6 (Supplemental Fig. S2).

These sequences were deposited in GenBank and annotated in the SGN database as new *Capsicum* unigenes. Loci added to the SGN database were manually curated using the community curation interfaces

Table 1. Capsaicinoid biosynthetic pathway candidate enzymes

n/a, Not applicable (gene not cloned); n.d., not able to be determined.

Name	Abbreviation	GenBank Accession No.	SGN Tomato Unigene	Trichome ESTs/ Total ESTs in Unigene	Most Similar Arabidopsis Locus	Percent Identity	Copy No.	Pepper Chromosome(s)
Phenylpropanoid and benzenoid metabolism								
Chorismate mutase	CM1	EU616556	U331044	0/2	AT3G29200	62	2	2, 1+8
Prephenate aminotransferase	PNT ^a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Arogenate dehydratase	ADT	EU616545	U314126	0/52	AT1G08250	77	2	11
Phe ammonia-lyase	PAL	EU616575	U312621	3/119	AT2G37040	81	3	9 ^b
Gln synthetase	GS2	EU616564	U313256	4/44	AT5G35630	86	3	5
NADH-dependent Glu synthase	NADH-GOGAT ^c	EU616574	U321635	0/8	AT5G53460	n/a	2	3
Ferredoxin-dependent Glu synthase	Fdx-GOGAT	EU616563	No match	n/a	AT5G04140	81	1	n.d.
Cinnamate 4-hydroxylase	C4H	EU620574	U313164	7/95	AT2G30490	86	2	6 ^b
4-Coumaroyl-CoA ligase	4CL	EU616540	U314997	1/39	AT3G21240	70	1	LG A
Hydroxycinnamoyl transferase	HCT	EU616565	U316595	0/25	AT5G48930	79	1	2
<i>p</i> -Coumaroyl shikimate/ quinate 3-hydroxylase	C3H	EU616552	U325121	6/29	AT2G40890	79	1	1+8
Cytochrome P450 reductase	CPR	EU616557	U313813	2/56	AT4G24520	75	1	n.d.
Caffeoyl-CoA 3- <i>O</i> -methyltransferase	CCoAOMT	EU616554	U313985	0/13	AT4G34050	88	2	1+8
S-Adenosylmethionine synthetase	SAMSyn	EU616581	U312579	12/132	AT2G36880	94	3	1+8
Cinnamoyl-CoA reductase	CCR	EU616555	U318660	0/16	AT1G15950	76	2	1+8, LG A
Cinnamyl alcohol dehydrogenase	CAD	EU616553	U324872	0/5	AT4G34230	77	1	1+8
Putative aminotransferase	pAMT ^d	AF085149	U313117	1/103	AT3G22200	74	3	7, 3
Putative aminotransferase	pAMT2	EU616576	U320631	0/10	AT4G39660	80	2	12
Amino acid precursor preparation								
Thr deaminase	TD ^e	n/a	n/a	n/a	AT3G10050 ^f	n/a	n/a	n/a
Acetolactate synthase	ALS	EU616547	U313745	1/48	AT3G48560	72	1	n.d.
Acetohydroxyacid reductoisomerase	AHRI	EU616546	U312795	3/83	AT3G58610	86	2	7
Dihydroxyacid dehydratase	DHAD	n/a	n/a	n/a	AT3G23940 ^f	n/a	n/a	n/a
Isopropylmalate synthase	IPMS	EU616568	U313457	0/8	AT1G74040	72	2	1+8, 6
Isopropylmalate isomerase	IPMI	EU616567	U316862	1/24	AT2G43090	65	3	3, 6, 11
Isopropylmalate dehydrogenase	IPMDH	EU616566	U317294	2/22	AT1G80560	79	1	Unlinked
Branched-chain amino acid aminotransferase	BCAT ^d	AY034379	U319112	0/5	AT1G10070	62	1	4 ^b
Branched-chain α -ketoacid dehydrogenase								
α -Ketoacid decarboxylase E1 α	BCKDH E1 α	EU616548	U320589	0/11	AT1G21400	60	2	6
α -Ketoacid decarboxylase E1 β	BCKDH E1 β	EU616549	U315368	5/41	AT3G13450	82	2	1+8
Dihydrolipoamide transacylase	BCKDH E2	EU616550	U321113	4/10	AT3G06850	57	3	Unlinked
Dihydrolipoamide dehydrogenase	BCKDH E3	EU616551	U314201	0/16	AT3G17240	84	1	1+8
Malonyl-CoA biosynthesis								
Pyruvate dehydrogenase complex								
Pyruvate dehydrogenase E1 α	PDH E1 α	EU616577	U313753	11/59	AT1G01090	80	2	7
Pyruvate dehydrogenase E1 β	PDH E1 β	EU616578	U314162	9/22	AT2G34590	85	3	4
Dihydrolipoamide acetyltransferase	PDH E2	EU616579	U321522	2/9	AT1G34430	73	1	4
Dihydrolipoamide dehydrogenase	PDH E3	EU616580	U319399	2/11	AT4G16155	81	2	Monomorphic

(Table continues on following page.)

Table I. (Continued from previous page.)

Name	Abbreviation	GenBank Accession No.	SGN Tomato Unigene	Trichome ESTs/ Total ESTs in Unigene	Most Similar Arabidopsis Locus	Percent Identity	Copy No.	Pepper Chromosome(s)
Acetyl-CoA carboxylase complex								
α -Carboxyltransferase	α -CT	EU616541	U317741	0/19	AT2G38040	62	2	1+8
β -Carboxyltransferase	β -CT	EU616544	No match	n/a	ATCG00500	60	1	(Not nuclear)
Biotin carboxyl carrier protein	BCCP	EU616543	U317261	1/21	AT5G15530	54	2	n.d.
Biotin carboxylase	BC	EU616542	U327019	0/4	AT5G35360	84	1	5
Fatty acid biosynthesis								
Malonyl-CoA:ACP transacylase	MCAT	EU616573	U316102	3/25	AT2G30200	76	1	n.d.
Ketoacyl-ACP synthase I	KasI ^d	AF085148	U326977	0/4	AT5G46290	70	>8	1+8, etc. ^g
Ketoacyl-ACP synthase III	KasIIIa	EU616569	U316868	10/23	AT1G62640	72	n.d.	n.d.
Ketoacyl-ACP synthase III	KasIIIb	EU616570	U328450	0/3	AT1G62640	67	n.d.	n.d.
Ketoacyl-ACP reductase	KR	EU616561	U315110	0/47	AT1G24360	65	2	6
Hydroxyacyl-ACP dehydratase	DH	EU616560	U319205	1/14	AT5G10160	74	3	1+8, 10
Enoyl-ACP reductase	ENRa	EU616558	U321872	3/8	AT2G05990	73	2	10
Enoyl-ACP reductase	ENRb	EU616559	U321872	3/8	AT2G05990	75	3	1+8
Acyl carrier protein	ACLd	AF127796	U314781	0/49	AT4G25050	55	1	1+8 ^g
Acyl-CoA synthetase 1	ACS1	EU616571	U320040	0/12	AT4G23850	76	1	1+8
Acyl-CoA synthetase 2	ACS2	EU616572	U319774	0/12	AT2G04350	69	1	1+8
Acyl-ACP thioesterase	FatA ^d	AF318288	U319803	0/10	AT4G13050	67	1	6 ^g
Acyl-ACP thioesterase	FatB	EU616562	U318317	2/16	AT1G08510	68	2	9
Acyltransferase 3 (Pun1)	AT3 ^d	AY819029	U331681	1/2	AT3G26040	33	1	2 ^h

^aNo candidate. ^bReported by Blum et al. (2003). ^cLacks 5' sequence. ^dFrom previous studies. ^eNo transcript cloned. ^fPredicted by model but not cloned. ^gReported by Aluru et al. (2003). ^hReported by Stewart et al. (2005).

(Menda et al., 2008). Briefly, each locus was annotated and can be queried with locus symbol, name, description, or enzymatic activity. The subsequent building of CapCyc incorporated links to these new unigenes as an automated function of the database (Menda et al., 2008).

Compartmentalization

The enzymes in our model and/or the reactions they catalyze are known to be restricted to certain organelles (Fig. 2). Previously, only EST sequences for the majority of these genes were available. The cloning of full-length transcripts allowed a secondary test of our model: the consistency of the modeled and predicted subcellular localizations of the enzymes determined using TargetP (Emanuelsson et al., 2007). In support of these predictions, the *in silico* localizations of the pepper proteins were compared with their Arabidopsis homologs (Table II), many of which have experimentally verified locations (Friso et al., 2004; Heazlewood et al., 2007). These predictions also revealed the length of the targeting peptide that is cleaved after organelle import and therefore is informative about the length of the mature, functional protein. The presence of peptides that would target proteins across the thylakoid membrane of the plastid from the stroma was ignored, as none of these en-

zymes would be expected to be found in the thylakoid lumen and the existence of thylakoids on this internal fruit organelle is unknown. There were rare notable discrepancies between modeled and sequence-predicted locations. BCAT, the branched-chain amino acid aminotransferase that is thought to be active in both the synthesis of amino acids in the plastid and their catabolism in the mitochondria, was predicted to be targeted to the mitochondria in pepper, while the closest Arabidopsis homolog is one of several copies in the genome and targeted to the plastid (Diebold et al., 2002). Conflicts between the model, TargetP prediction, and experimental evidence for the Arabidopsis homolog also exist for pAMT and the acyl-CoA synthetases.

Genetic Linkage Mapping in Pepper

Cloned genes were placed on a *Capsicum* linkage map as RFLPs (Fig. 6; Table I). In combination with previous studies (Aluru et al., 2003; Blum et al., 2003; Stewart et al., 2005), these results account for the location of nearly the entire capsaicinoid biosynthetic model in the *Capsicum* genome. The consensus of the markers on this map were amplified fragment length polymorphisms (AFLPs), but some shared markers, such as RFLPs, were included to allow the alignment of this map with other maps (<http://www.sgn.cornell.edu>). The total map length was 1,374 centimorgan and

Table II. Candidate enzyme-targeting predictions and comparison with *Arabidopsis*

C, Chloroplast; ER, endoplasmic reticulum; M, mitochondria; n/a, not applicable; O, other (usually cytoplasm); S, secreted/ER localized; TM, transmembrane domain.

Abbreviation	TargetP Prediction (Score)	Arabidopsis Ortholog Localization	Predicted Targeting Peptide or TM Length (Score)
CM1	C (0.869)	Lumen	58 (0.564)
PNT	n/a	n/a	n/a
ADT	C (0.783)	Stroma	31 (0.555)
PAL	O (0.918)	Cytoplasm	n/a
GS2	C (0.822)	Stroma ^a	44 (0.572)
NADH-GOGAT	n/a	Stroma ^a	n/a
Fdx-GOGAT	C (0.415)	Stroma ^a	52 (0.496)
C4H	S (0.859)	ER ^a	TM 5–24 (0.855)
4CL	O (0.823)	Cytoplasm	n/a
HCT	O (0.699)	Cytoplasm	n/a
C3H	S (0.659)	ER ^a	TM 2–24 (0.932)
CPR	S (0.306)	ER ^b	TM 21–43 (0.853)
CCoAOMT	O (0.734)	Cytoplasm	n/a
SAMSyn	O (0.857)	Cytoplasm	n/a
CCR	O (0.578)	Cytoplasm	n/a
CAD	O (0.835)	Cytoplasm	n/a
pAMT	O (0.918)	Mitochondria ^a	n/a
pAMT2	M (0.907)	Mitochondria ^a	23
TD	n/a	Plastid ^a	n/a
ALS	C (0.978)	Stroma ^a	55 (0.588)
AHRI	C (0.979)	Stroma ^a	46 (0.598)
DHAD	n/a	n/a	n/a
IPMS	C (0.856)	Stroma ^a	41 (0.517)
IPMI	C (0.938)	Stroma ^a	60 (0.571)
IPMDH	C (0.399)	Stroma ^a	34 (0.510)
BCAT	M (0.945)	Chloroplast ^b	19
BCKDH			
BCKDH E1 α	M (0.820)	Mitochondria	46
BCKDH E1 β	M (0.561)	Mitochondria ^a	11
BCKDH E2	M (0.702)	Mitochondria ^a	31
BCKDH E3	M (0.910)	Mitochondria ^a	41
PDH			
PDH E1 α	C (0.895)	Stroma ^a	59 (0.581)
PDH E1 β	C (0.965)	Stroma ^a	32 (0.568)
PDH E2	C (0.761)	Stroma ^a	34 (0.525)
PDH E3	C (0.996)	Plastid ^a	75 (0.583)
ACCase			
α -CT	C (0.951)	Stroma ^a	55 (0.568)
β -CT	N/A	Stroma ^a	N/A
BCCP	C (0.918)	Stroma ^a	12 (0.553)
BC	C (0.853)	Stroma ^a	53 (0.579)
MCAT	C (0.974)	Stroma ^a	64 (0.583)
KasI	C (0.478)	Stroma ^a	60 (0.462)
KasIIIa	C (0.626)	Stroma ^a	73 (0.547)
KasIIIb	C (0.329)	Stroma ^a	39 (0.506)
KR	C (0.919)	Stroma ^a	72 (0.593)
DH	C (0.804)	Stroma ^a	47 (0.573)
ENRb	C (0.728)	Stroma ^a	72 (0.523)
ENRb	C (0.968)	Stroma ^a	73 (0.574)
ACL	C (0.942)	Stroma ^a	48 (0.560)
ACS1	O (0.810)	Cytoplasm	n/a
ACS2	O (0.946)	Cytoplasm	TM 28–50 (0.924)
FatA	C (0.918)	Stroma	55 (0.575)
FatB	C (0.914)	Stroma	53 (0.555)
AT3	C (0.677)	Cytoplasm	47 (0.544)

^aSupported by mass spectrometric identification in *Arabidopsis*.
data.

^bSupported by ectopic expression data.

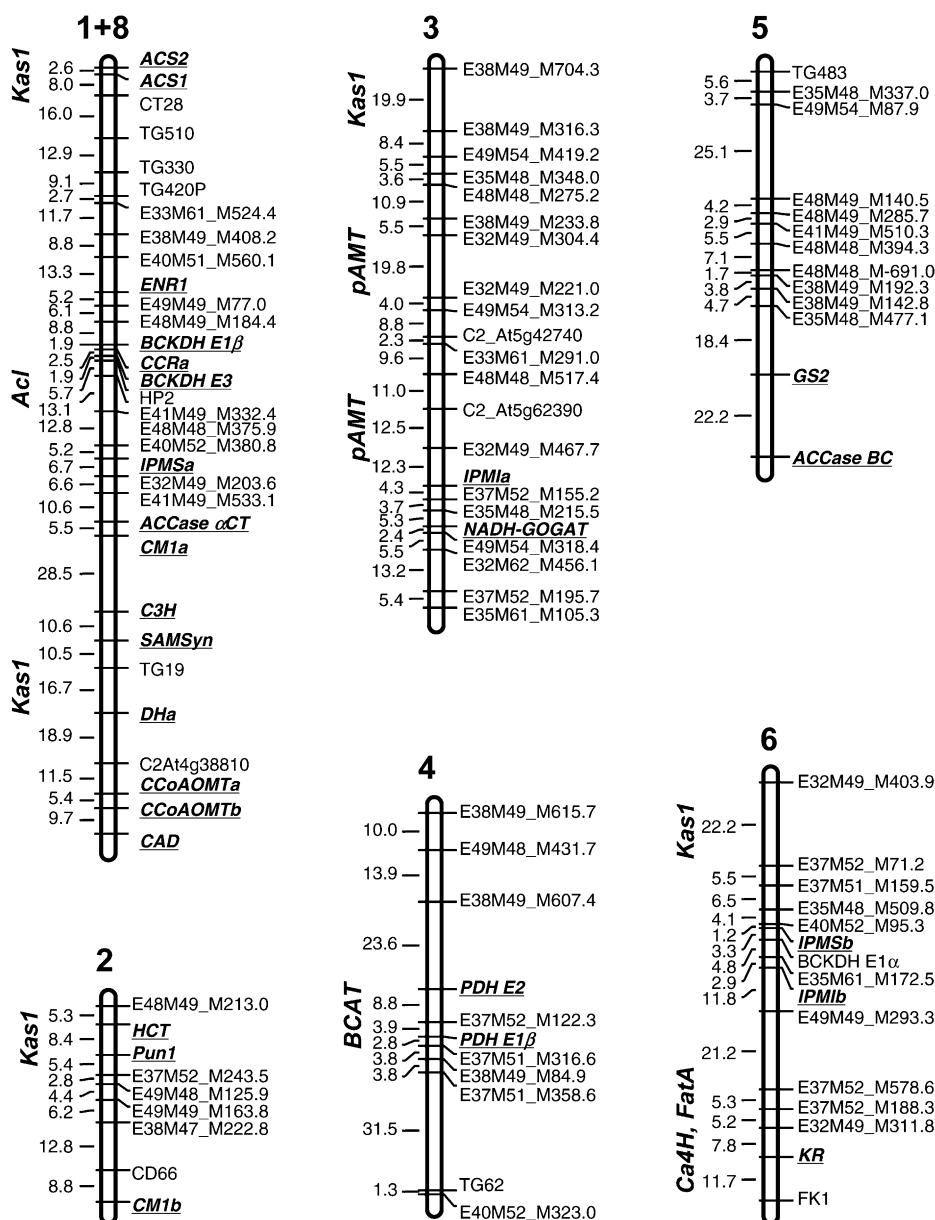


Figure 6. Genetic map of capsaicinoid candidate genes in pepper. This genetic linkage map of pepper represents the 12 chromosomes and one linkage group that could not be assigned to a chromosome. Molecular markers are shown to the right of the chromosome, and candidate genes for capsaicinoid biosynthesis are underlined. Genes mapped in previous studies are shown to the left. Numbers to the left of the chromosomes indicate the centimorgan distances between adjacent markers. Gene names are abbreviated as in Table 1.

consisted of 178 markers separated by an average distance of 7.7 centimorgan. The development of COSII markers in pepper will supersede AFLP and RFLP as an ideal way to link this map with the tomato genome and with other pepper maps (Wu et al., 2006, 2009). The majority of genes were detected as one- or two-copy loci scattered throughout the genome. This is likely an underestimate for several gene families, where paralogs are too divergent for RFLP probe cross-hybridization. For example, there are 272 cytochrome P450 genes in the Arabidopsis genome (Schuler and Werck-Reichhart, 2003), and the two members of this family implicated in capsaicinoid biosynthesis, *C3H* and *C4H*, were detected in pepper as one- and two-copy families, respectively. *CCoAOMT* loci were found as a linked pair of loci on the pepper pseudo-

linkage group that combines chromosomes 1 and 8. Pseudolinkage between these chromosomes is a commonly observed phenomenon in interspecific maps in pepper involving crosses between *C. annuum* and *C. chinense* or *C. frutescens* (Livingstone et al., 1999; Wu et al., 2009). Other homologous loci were found to be linked: two *ACS* loci were found on chromosomes 1 and 8; *HCT* and *AT3* (*Pun1*) are two members of the BAHD superfamily that are involved in capsaicinoid biosynthesis and are linked on chromosome 2. Of the loci that could not be placed on the genetic map, the *CCRb* and *4CL* loci were linked to each other but could not be associated with a pepper chromosome. Several candidates mapped to chromosomes 1 and 8, but one of the arms of chromosome 5 contains none of the candidates mapped to date.

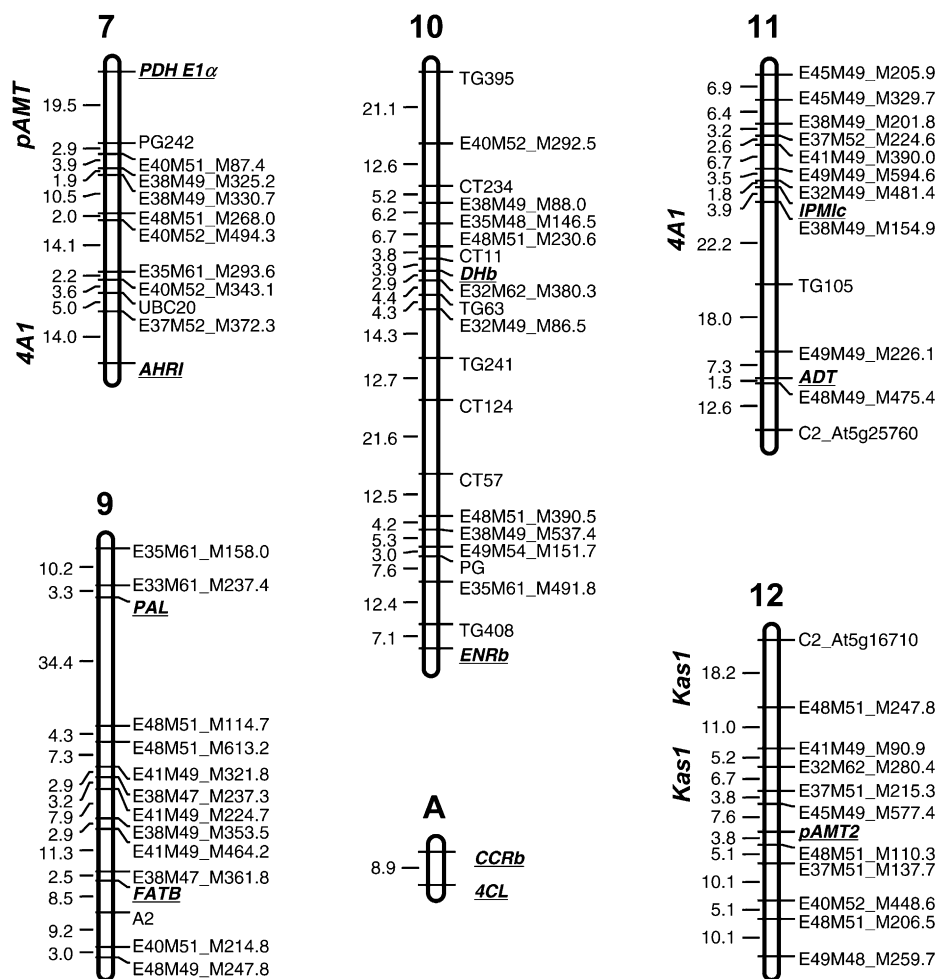


Figure 6. (Continued.)

DISCUSSION

CapCyc Evolution

SGN is a central hub for the Solanaceae research community. The creation of the CapCyc pathway within SGN provides a natural forum for information related to the elucidation of capsaicinoid biosynthesis in *Capsicum*. Many of the reaction steps in the capsaicinoid biosynthetic model have not been demonstrated experimentally in pepper but were instead inferred from Solanaceae or other species. In some metabolisms, such as Phe (Cho et al., 2007) and vanillin biosynthesis (Walton et al., 2003), gaps exist in the model where reaction steps are unknown.

Condensation and Export

Capsaicinoid synthase, the final enzyme of capsaicinoid biosynthesis, is yet to be identified. Assays for capsaicinoid synthase activity have been performed with both CoA-activated fatty acids and free fatty

acids, with ATP, magnesium, and free CoA, ostensibly to supply substrates and cofactors to endogenous acyl-CoA synthetases present in crude extracts. The formation of capsaicinoids from the CoA-activated fatty acids was more than six times greater than with the free fatty acid, leading to an interpretation of a two-step reaction: first, the formation of the acyl-CoA, followed by transfer to vanillylamine (Fujiwaka et al., 1980). However, the study of a number of acyltransferase enzymes of the BAHD superfamily that transfer CoA-activated acyl groups to a variety of metabolites has suggested that capsaicinoids would be formed in this manner (St-Pierre and De Luca, 2000; Ma et al., 2005). The gene that is mutated in bell pepper, *AT3*, which results in a loss of pungency, was recently discovered and found to encode a member of this class of acyltransferase enzymes (Stewart et al., 2005). *AT3* appears to be a hot spot for loss-of-pungency mutations in other species and therefore is believed to be critical in the evolution of pungency (Stewart et al., 2007). While *AT3* is a candidate to encode for capsaicinoid synthase, peppers with loss-of-pungency mu-

tations in *AT3* have been reported to synthesize capsaicinoids when the immediate precursors, vanillylamine and branched-chain fatty acids, are supplied exogenously (Iwai et al., 1977). Another candidate had been reported, *Csy1*, but that report has since been retracted (Prasad et al., 2006, 2007). The final condensation enzyme is still an open question.

The details of the location of the accumulation and the synthesis of capsaicinoids from vanillylamine and fatty acids have also been a matter of much debate. There is a widespread misconception that the site of capsaicinoid accumulation is the pepper seeds (Stewart et al., 2007). Numerous lines of evidence clearly show that the placental dissepiment is the site of this accumulation and biosynthesis by differentiated cells that line the epidermis of this internal fruit structure (Suzuki et al., 1980; Zamski et al., 1987; Stewart et al., 2007; Fig. 2), but there are differing reports of the intracellular site of capsaicinoid synthesis and route of secretion (Suzuki and Iwai, 1984; Zamski et al., 1987). Once out of the cell, capsaicinoids accumulate underneath the cuticle in fluid-filled "blisters" (Stewart et al., 2007), where they are near or in contact with the seeds.

Candidate Genes and Strategies for Nonmodel Genomes

The size of the pepper genome is estimated to be approximately 3,000 Mbp. This is three times the size of the tomato genome and 20 times larger than the *Arabidopsis* genome (Arumuganathan and Earle, 1991). The gene content of these plants is similar, but the amount of intervening repetitive DNA found as heterochromatin is much increased (Van der Hoeven et al., 2002). A common and efficient approach to access genomic information in pepper is through the use of a positional candidate gene approach (Thorup et al., 2000). This employs cloned genes as candidates for phenotypes of interest and tests for association on genetic maps. Candidate gene approaches have the advantage of being very efficient, but they are also dependent on possession of sequences related to the biological process being studied. An advantage of working in an experimental system like the Solanaceae is the ability to use sequences or map positions in a comparative way to associate phenotypes with genes across genera. This candidate gene approach has been very successful in pepper, identifying the gene for loss of pungency in bell pepper and quantitative trait locus candidates for pungency using genes from pepper (Stewart et al., 2005, 2007) and identifying pepper genes that are associated with fruit color loci in tomato (Thorup et al., 2000) and other disease resistance and quality traits (Paran and van der Knaap, 2007). Our collection of pepper sequences related to capsaicinoid biosynthesis will be immediately valuable as candidates for pungency quantitative trait loci in existing pepper populations and will be a resource for discovering the molecular nature of those traits.

A limitation of our approach is that we only identified genes predicted by our model. In addition to not

discovering novel genes, regulatory genes were not included in our study because there was no straightforward way to design primers for their cloning based on homology relationships. Future iterations of this model can include genes that were not predicted to be part of the metabolic pathways by the inclusion of genes that are up-regulated during capsaicinoid biosynthesis. Previously, the spatial and temporal expression pattern of capsaicinoids has been used to identify new candidate genes for this pathway, including regulatory factors (Curry et al., 1999; Kim et al., 2001; Aluru et al., 2003; Blum et al., 2003; Lee et al., 2006). A microarray approach in *Solanum* has already generated many candidates that are up-regulated in trichomes (Slocombe et al., 2008). These resources, and EST libraries in SGN, should be mined for relevance to capsaicinoid biosynthesis, but it should also be kept in mind that differences in mRNA abundance would not necessarily be found for genes involved in other processes or regulated at levels other than transcription.

Comparative Metabolism

The genus *Solanum* includes tomato and its wild relative *S. pennellii*, which, like pepper, produce fleshy fruit as a mode of seed dispersal. Genetic components that suggest the ability to produce capsaicinoids are present and expressed, but capsaicinoids are not. A meta-analysis of genes from the capsaicinoid biosynthetic model was conducted by determining the relative abundance of trichome-derived ESTs in tomato unigenes (Table I). Most of the capsaicinoid transcript candidates were found to have homologs in these libraries. For KasIIIa, FatB, and components of the pyruvate dehydrogenase and branched-chain α -ketoacid dehydrogenase complexes, a substantial portion of the members of the *Solanum* unigene derived from "trichome" libraries. Slocombe et al. (2008) corroborated this expression for a subset of these genes but were limited by the inclusion of less than half of these structural genes on the tomato microarray. The metabolites that are direct upstream precursors to the capsaicinoids, vanillic acid, and fatty acids are present in tomato (and crosses between tomato and *S. pennellii*), vanillin is present in the fruit, and the same medium-length, branched chain fatty acids found in capsaicinoids are abundant (specifically 8-methylnonanoic acid; Fig. 1, structure b; Lewinsohn et al., 2001; Moco et al., 2006). In the future, CapCyc-related sequences will be linked to tomato orthologs, allowing researchers to move between model organisms.

CONCLUSION

We have presented a unified model of the capsaicinoid biosynthetic pathway, the specialized metabolism that is characteristic of hot pepper. In support of

this model, the predicted transcripts were cloned from pepper fruit tissue actively synthesizing capsaicinoids, anchored by homology to Arabidopsis loci (The Arabidopsis Information Resource; www.arabidopsis.org), and referenced to tomato unigenes (SGN; www.sgn.cornell.edu), effectively linking this information in GenBank with two other major plant-specific databases. Because the subcellular compartmentalization of many of these enzymes is known in Arabidopsis, we were able to test whether the enzymes encoded by the pepper transcripts are predicted to be targeted to the proper organelle by homology and the presence of characteristic targeting peptides in the pepper sequences. Map locations were determined as a resource for a positional candidate gene approach. These data are dynamically integrated and accessible at the SGN Web site as a community resource.

MATERIALS AND METHODS

Plant Materials

Seeds for Santa Fe Grande (*Capsicum annuum*), a pungent pepper cultivar that produces capsaicinoids, were generously provided by the Chile Pepper Institute (New Mexico State University, Las Cruces). *C. annuum* is the most cultivated pepper species and is the model pepper species used in SGN. Plants were grown in the Guterman Greenhouse Complex at Cornell University with supplemental lighting and standard greenhouse practices.

Bioinformatics

CapCyc was generated using Pathway Tools software (Karp et al., 2002), developed at SRI International, using MetaCyc (Caspi et al., 2006) as the reference database. Pathway Tools is a comprehensive software package to create, manipulate, edit, visualize, curate, and share pathway information that is written in Common Lisp; it has been applied to several hundred bacterial genomes and about a dozen different plant species. Pathway Tools can be run with a relational back end or based on flatfiles, the latter of which was used in this work. Pathway Tools can be downloaded from the SRI Web site at: <http://biocyc.org/download.shtml>. The PathoLogic module (Karp et al., 2002) of Pathway Tools generated a new PGDB using MetaCyc as a reference database and a *C. annuum* unigene set generated at SGN (Mueller et al., 2005). Annotations for the *C. annuum* unigenes were obtained by transferring annotations of Arabidopsis (*Arabidopsis thaliana*) loci (Berardini et al., 2004) as referenced in the AraCyc database (Mueller et al., 2003). Unigenes were selected that matched Arabidopsis sequences with an E value of less than 1×10^{-20} in BLAST (Altschul et al., 1990). The Pathway Tools editor functionalities were used to validate and manually curate pathways specific to *Capsicum*. Annotations for the loci identified in this work were also added to the SGN locus database using the SGN community curation tools (Menda et al., 2008).

Molecular Biology

Fruit were harvested at 20 DPA to coincide with peak capsaicinoid-related transcript expression, and the placental dissepiment was collected and frozen in liquid nitrogen. Total RNA was extracted using the Qiagen RNeasy kit according to the manufacturer's protocol. 3' and 5' cDNA libraries were generated using the Clontech Smart Race kit and used as the template for PCR amplification of overlapping candidate gene fragments. Primer sequences and RACE product lengths are listed as Supplemental Table S1. PCR products were gel purified (QiaQuick; Qiagen) and cloned into pCR4 TOPO plasmid vectors prior to electroporation into TOP10 *Escherichia coli* (Invitrogen). Plasmid DNA was extracted for sequencing using the QiaPrep kit (Qiagen) and sequenced by the Cornell University Life Sciences Core Laboratories Center.

Candidate Genes

Candidate genes were identified in the literature, the Arabidopsis Lipid Gene Database (<http://lipids.plantbiology.msu.edu/>), and the Kyoto Encyclopedia of Genes and Genomes (<http://www.kegg.com>). Where possible, a corresponding EST was retrieved from the SGN (<http://www.sgn.com>) to design PCR primers. Priority was given to ESTs derived from libraries of capsaicinoid-synthesizing pepper tissue. PCR primers (Integrated DNA Technologies) were arranged to generate overlapping 5' and 3' cDNA fragments. Sequence contigs were assembled from these sequences with Sequencher (Gene Codes). The most probable coding sequence was identified using BLASTX (<http://www.ncbi.nlm.nih.gov>). The resulting sequences (listed in Table I) were deposited in GenBank: chorismate mutase (accession number EU616556), arogenate dehydratase, (EU616545), Phe ammonia-lyase (EU616575), Gln synthetase, (EU616564), NADH-dependent Glu synthase (EU616574), ferredoxin-dependent Glu synthase (EU616563), cinnamate 4-hydroxylase (EU620574), 4-coumaroyl-CoA ligase (EU616540), hydroxycinnamoyl transferase (EU616565), *p*-coumaroyl shikimate/quinic 3-hydroxylase (EU616552), cytochrome P450 reductase (EU616557), caffeoyl-CoA 3-O-methyltransferase (EU616554), S-adenosylmethionine synthetase (EU616581), cinnamoyl-CoA reductase (EU616555), cinnamyl alcohol dehydrogenase (EU616553), putative aminotransferase (EU616576), acetolactate synthase (EU616547), acetoxyacid reductoisomerase (EU616546), isopropylmalate synthase (EU616568), isopropylmalate isomerase (EU616567), isopropylmalate dehydrogenase (EU616566), malonyl-CoA:ACP transacylase (EU616573), ketoacyl-ACP synthase III (EU616569, EU616570), ketoacyl-ACP reductase (EU616561), hydroxyacyl-ACP dehydratase (EU616560), enoyl-ACP reductase (EU616558, EU616559), acyl-CoA synthetase (EU616571, EU616572), acyl-ACP thioesterase (EU616562), the branched-chain α -ketoacid dehydrogenase complex subunits α -ketoacid decarboxylase E1 α (EU616548), α -ketoacid decarboxylase E1 β (EU616549), dihydrolipoamide transacylase (EU616550), and dihydrolipoamide dehydrogenase (EU616551), the pyruvate dehydrogenase complex subunits pyruvate dehydrogenase E1 α (EU616577), pyruvate dehydrogenase E1 β (EU616578), dihydrolipoamide acetyltransferase (EU616579), and dihydrolipoamide dehydrogenase (EU616580), and the acetyl-CoA carboxylase complex subunits ACCase α -carboxyltransferase (EU616541), β -carboxyltransferase (EU616544), biotin carboxyl carrier protein (EU616543), and biotin carboxylase (EU616542).

Prediction of Subcellular Localization

Toward determining the subcellular localization of the candidate enzymes for capsaicinoid biosynthesis, the predicted translations of cloned genes were submitted to TargetP and analyzed as described (Emanuelsson et al., 2007). Those sequences that were predicted to contain a plastid-targeting peptide were further analyzed by BLAST searches (<http://plantbpb.uoregon.edu/>) to identify a putatively orthologous protein whose subcellular localization was experimentally determined as described at <http://ppdb.tc.cornell.edu> (Friso et al., 2004) and <http://www.mitoz.bcs.uwa.edu.au/applications/suba2/index.php> (Heazlewood et al., 2007). Alignment between the predicted pepper proteins and the identified peptides of the mature protein provided a means to further evaluate the accuracy of the predicted targeting peptide cleavage site.

Genetic Mapping

A mapping population was constructed from an interspecific cross of line 1154 (*C. annuum* accession A44750157; Nijmegen Botanical Garden) and PI 152225 (*C. chinense*). An F₂ population of 182 individuals was used to create the map, which consisted of 200 AFLP, RFLP, and COSII markers and candidate genes from the capsaicinoid biosynthetic pathway. The AFLP markers (performed by Nunhems Netherlands) were chosen as a subset of approximately 450 markers distributed throughout the genome. Procedures for RFLP and AFLP analyses and genetic mapping were described previously (Ben Chaim et al., 2001). The genetic map was constructed using MAPMAKER software (Lander et al., 1987). Map distances were computed with the Kosambi mapping function. Chromosome numbers were assigned to linkage groups based on known markers previously mapped to pepper chromosomes.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Fatty acid elongation related to capsaicinoids.

Supplemental Figure S2. Phe biosynthesis.

Supplemental Table S1. RACE primers used for transcript cloning.

Supplemental Text S1. Biosynthetic model assumptions and considerations.

ACKNOWLEDGMENTS

We thank Martha Mutschler, Mary O'Connell, Jocelyn Rose, and Giulia Stellari for providing critical review of the manuscript and helpful discussion, Saj Bashir for the initial cloning of HCT, Charles Stewart Jr. for sharing the micrograph in Figure 2, Nicole Moskal and Elizabeth Cirulli for assistance with candidate sequences, and Aashish Barwale for assistance with references.

Received February 2, 2009; accepted June 13, 2009; published June 24, 2009.

LITERATURE CITED

- Achterholt S, Priefert H, Steinbuechel A (2000) Identification of *Amycolatopsis* sp. strain HR167 genes, involved in the bioconversion of ferulic acid to vanillin. *Appl Microbiol Biotechnol* **54**: 799–807
- Alba R, Fei ZJ, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JKC, et al (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* **39**: 697–714
- Altschul SE, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410
- Aluru MR, Mazourek M, Landry LG, Curry J, Jahn M, O'Connell MA (2003) Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. *J Exp Bot* **54**: 1655–1664
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* **9**: 211–215
- Ben Chaim A, Grube RC, Lapidot M, Jahn M, Paran I (2001) Identification of quantitative trait loci associated with resistance to *Cucumber mosaic virus* in *Capsicum annuum*. *Theor Appl Genet* **102**: 1213–1220
- Bennett J, Kirby GW (1968) Constitution and biosynthesis of capsaicin. *J Chem Soc C* 442–446
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang PF, 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
- Billing J, Sherman PW (1998) Antimicrobial functions of spices: why some like it hot. *Q Rev Biol* **73**: 3–49
- Blum E, Mazourek M, O'Connell M, Curry J, Thorup T, Liu KD, Jahn M, Paran I (2003) Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. *Theor Appl Genet* **108**: 79–86
- Bosland PW, Baral JB (2007) 'Bhut Jolokia': the world's hottest known chile pepper is a putative naturally occurring interspecific hybrid. *HortScience* **42**: 222–224
- Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, et al (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* **34**: D511–D516
- Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD, Julius D (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**: 816–824
- Cho MH, Corea ORA, Yang H, Bedgar DL, Laskar DD, Anterola AM, Moog-Anterola FA, Hood RL, Kohalmi SE, Bernards MA, et al (2007) Phenylalanine biosynthesis in *Arabidopsis thaliana*: identification and characterization of arogenate dehydratases. *J Biol Chem* **282**: 30827–30835
- Chortyk OT, Kays SJ, Teng Q (1997) Characterization of insecticidal sugar esters of *Petunia*. *J Agric Food Chem* **45**: 270–275
- Curry J, Aluru M, Mendoza M, Nevarez J, Melendrez M, O'Connell MA (1999) Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Sci* **148**: 47–57
- Diaz J, Pomar F, Bernal A, Merino F (2004) Peroxidases and the metabolism of capsaicin in *Capsicum annuum* L. *Phytochem Rev* **3**: 141–157
- Diebold R, Schuster J, Daschner K, Binder S (2002) The branched-chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. *Plant Physiol* **129**: 540–550
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protocols* **2**: 953–971
- Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, Wijk KJ (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* **16**: 478–499
- Fujiwake H, Suzuki T, Oka S, Iwai K (1980) Formation and metabolism of pungent principle of *Capsicum* fruits. 7. Enzymatic formation of capsaicinoid from vanillylamine and iso-type fatty-acids by cell-free-extracts of *Capsicum annuum* var *annuum* cv Karayatsubusa. *Agric Biol Chem* **44**: 2907–2912
- Gasson MJ, Kitamura Y, McLauchlan WR, Narband A, Parr AJ, Parsons ELH, Payne J, Rhodes MJC, Walton NJ (1998) Metabolism of ferulic acid to vanillin: a bacterial gene of the enoyl-SCoA hydratase/isomerase superfamily encodes an enzyme for the hydration and cleavage of hydroxycinnamic acid SCoA thioester. *J Biol Chem* **273**: 4163–4170
- Govindarajan VS (1985) *Capsicum* production, technology, chemistry, and quality. 1. History, botany, cultivation, and primary processing. *CRC Crit Rev Food Sci Nutr* **22**: 109–176
- Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: The Arabidopsis subcellular database. *Nucleic Acids Res* **35**: D213–D218
- Higashiguchi F, Nakamura H, Hayashi H, Kometani T (2006) Purification and structure determination of glucosides of capsaicin and dihydrocapsaicin from various *Capsicum* fruits. *J Agric Food Chem* **54**: 5948–5953
- Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinat hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **16**: 1446–1465
- Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J Biol Chem* **278**: 95–103
- Humphreys JM, Chapple C (2002) Rewriting the lignin roadmap. *Curr Opin Plant Biol* **5**: 224–229
- Iwai K, Lee KR, Kobashi M, Suzuki T (1977) Formation of pungent principles in fruits of sweet pepper *Capsicum annuum* var *grossum* during post harvest ripening under continuous light. *Agric Biol Chem* **41**: 1873–1876
- Iwai K, Suzuki T, Fujiwake H (1979) Formation and accumulation of pungent principle of hot pepper fruits: capsaicin and its analogs in *Capsicum annuum* var *annuum* cv Karayatsubusa at different growth stages after flowering. *Agric Biol Chem* **43**: 2493–2498
- Johnson TS, Ravishankar GA (1996) Precursor biotransformation in immobilized placental tissues of *Capsicum frutescens* Mill. I. Influence of feeding intermediate metabolites of the capsaicinoid pathway on capsaicin and dihydrocapsaicin accumulation. *J Plant Physiol* **147**: 481–485
- Kandra L, Severson R, Wagner GJ (1990) Modified branched-chain amino acid pathways give rise to acyl acids of sucrose esters exuded from tobacco leaf trichomes. *Eur J Biochem* **188**: 385–392
- Karp PD, Paley S, Romero P (2002) The Pathway Tools software. *Bioinformatics (Suppl 1)* **18**: S225–S232
- Kim M, Kim S, Kim S, Kim BD (2001) Isolation of cDNA clones differentially accumulated in the placenta of pungent pepper by suppression subtractive hybridization. *Mol Cells* **11**: 213–219
- King RR, Calhoun LA (1988) 2,3-Di-O-acylated and 1,2,3-tri-O-acylated glucose esters from the glandular trichomes of *Datura metel*. *Phytochemistry* **27**: 3761–3763
- King RR, Calhoun LA, Singh RP, Boucher A (1990) Sucrose esters associated with glandular trichomes of wild *Lycopersicon* species. *Phytochemistry* **29**: 2115–2118
- King RR, Singh RP, Boucher A (1987) Variation in sucrose esters from the type-B glandular trichomes of certain wild potato species. *Am Potato J* **64**: 529–534
- Kobata K, Sutoh K, Todo T, Yazawa S, Iwai K, Watanabe T (1999)

- Nordihydrocapsiate, a new capsinoid from the fruits of a nonpungent pepper, *Capsicum annuum*. *J Nat Prod* **62**: 335–336
- Kobata K, Tate H, Iwasaki Y, Tanaka Y, Ohtsu K, Yazawa S, Watanabe T** (2008) Isolation of coniferyl esters from *Capsicum baccatum* L., and their enzymatic preparation and agonist activity for TRPV1. *Phytochemistry* **69**: 1179–1184
- Kobata K, Todo T, Yazawa S, Iwai K, Watanabe T** (1998) Novel capsaicinoid-like substances, capsiate and dihydrocapsiate, from the fruits of a nonpungent cultivar, CH-19 sweet, of pepper (*Capsicum annuum* L.). *J Agric Food Chem* **46**: 1695–1697
- Kopp B, Jurenitsch J** (1981) Biosynthesis of capsaicinoids in *Capsicum annuum* L. var. *annuum*. II. Formation of acid residue of capsaicinoids from L-valine, L-leucine and L-isoleucine. *Planta Med* **43**: 272–279
- Kopp B, Jurenitsch J, Woeginger R, Haunold E** (1983) Biosynthesis of capsaicinoids in *Capsicum annuum* L. var. *annuum*. IV. Incorporation of nitrogen. *Sci Pharm* **51**: 274–277
- Krajewska AM, Powers JJ** (1988) Sensory properties of naturally occurring capsaicinoids. *J Food Sci* **53**: 902–905
- Kroumova AB, Wagner GJ** (2003) Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. *Planta* **216**: 1013–1021
- Kroumova AB, Xie Z, Wagner GJ** (1994) A pathway for the biosynthesis of straight and branched, odd- and even-length, medium-chain fatty acids in plants. *Proc Natl Acad Sci USA* **91**: 11437–11441
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L** (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181
- Lea PJ, Ireland RJ** (1999) Nitrogen metabolism in higher plants. In BK Singh, ed, *Plant Amino Acids: Biochemistry and Biotechnology*. Marcel Dekker, New York, pp 1–48
- Lee JM, Kim S, Lee JY, Yoo JY, Cho MC, Cho MR, Kim B-D, Bahk YY** (2006) A differentially expressed proteomic analysis in placental tissues in relation to pungency during the pepper fruit development. *Proteomics* **6**: 5248–5259
- Leete E, Louden MCL** (1968) Biosynthesis of capsaicin and dihydrocapsaicin in *Capsicum frutescens*. *J Am Chem Soc* **90**: 6837–6841
- Lewinsohn E, Schalechet F, Wilkinson J, Matsui K, Tadmor Y, Nam KH, Amar O, Lastochkin E, Larkov O, Ravid U, et al** (2001) Enhanced levels of the aroma and flavor compound S-linalool by metabolic engineering of the terpenoid pathway in tomato fruits. *Plant Physiol* **127**: 1256–1265
- Li L, Thipyapong P, Breeden DC, Steffens JC** (2003) Overexpression of a bacterial branched-chain alpha-keto acid dehydrogenase complex in Arabidopsis results in accumulation of branched-chain acyl-CoAs and alteration of free amino acid composition in seeds. *Plant Sci* **165**: 1213–1219
- Livingstone KD, Lackney VK, Blauth JR, Van Wijk R, Jahn MK** (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics* **152**: 1183–1202
- Ma XY, Koepke J, Panjekar S, Fritsch G, Stockigt J** (2005) Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J Biol Chem* **280**: 13576–13583
- Markai S, Marchand Patrice A, Mabon F, Baguet E, Billault I, Robins Richard J** (2002) Natural deuterium distribution in branched-chain medium-length fatty acids is nonstatistical: a site-specific study by quantitative 2H NMR spectroscopy of the fatty acids of capsaicinoids. *ChemBioChem* **3**: 212–218
- Menda N, Buels RM, Teclé I, Mueller LA** (2008) A community-based annotation framework for linking Solanaceae genomes with phenomes. *Plant Physiol* **147**: 1788–1799
- Merali Z, Mayer MJ, Parker ML, Michael AJ, Smith AC, Waldron KW** (2007) Metabolic diversion of the phenylpropanoid pathway causes cell wall and morphological changes in transgenic tobacco stems. *Planta* **225**: 1165–1178
- Miyashita Y, Good AG** (2008) Contribution of the GABA shunt to hypoxia-induced alanine accumulation in roots of *Arabidopsis thaliana*. *Plant Cell Physiol* **49**: 92–102
- Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CHR** (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol* **141**: 1205–1218
- Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin CW, Wright MH, Ahrens R, Wang Y, et al** (2005) The SOL Genomics Network: a comparative resource for *Solanaceae* biology and beyond. *Plant Physiol* **138**: 1310–1317
- Mueller LA, Zhang PF, Rhee SY** (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol* **132**: 453–460
- Nelson EK, Dawson LE** (1923) Constitution of capsaicin, the pungent principle of *Capsicum*. III. *J Am Chem Soc* **45**: 2179–2181
- Ochi T, Takaishi Y, Kogure K, Yamauti I** (2003) Antioxidant activity of a new capsaicin derivative from *Capsicum annuum*. *J Nat Prod* **66**: 1094–1096
- Orlova I, Marshall-Colon A, Schnepf J, Wood B, Varbanova M, Fridman E, Blakeslee JJ, Peer WA, Murphy AS, Rhodes D, et al** (2006) Reduction of benzenoid synthesis in petunia flowers reveals multiple pathways to benzoic acid and enhancement in auxin transport. *Plant Cell* **18**: 3458–3475
- Palanivelu R, Brass L, Edlund AF, Preuss D** (2003) Pollen tube growth and guidance is regulated by *POP2*, an Arabidopsis gene that controls GABA levels. *Cell* **114**: 47–59
- Paran I, van der Knaap E** (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* **58**: 3841–3852
- Podstolski A, Havkin-Frenkel D, Malinowski J, Blount JW, Kourteva G, Dixon RA** (2002) Unusual 4-hydroxybenzaldehyde synthase activity from tissue cultures of the vanilla orchid *Vanilla planifolia*. *Phytochemistry* **61**: 611–620
- Prasad BCN, Kumar V, Gururaj HB, Parimalan R, Giridhar P, Ravishankar GA** (2006) Characterization of capsaicin synthase and identification of its gene (*csy1*) for pungency factor capsaicin in pepper (*Capsicum* sp.). *Proc Natl Acad Sci USA* **103**: 13315–13320
- Prasad BCN, Kumar V, Gururaj HB, Parimalan R, Giridhar P, Ravishankar GA** (2007) Retraction. *Proc Natl Acad Sci USA* **105**: 13315
- Rangoonwala R** (1969) On the biosynthesis of capsaicin. *Pharmazie* **24**: 177
- Razal RA, Ellis S, Singh S, Lewis NG, Towers GHN** (1996) Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry* **41**: 31–35
- Reilly CA, Yost GS** (2006) Metabolism of capsaicinoids by P450 enzymes: a review of recent findings on reaction mechanisms, bio-activation, and detoxification processes. *Drug Metab Rev* **38**: 685–706
- Rose JKC, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS** (2004) Tackling the plant proteome: practical approaches, hurdles and experimental tools. *Plant J* **39**: 715–733
- Schnurr JA, Shockey JM, de Boer GJ, Browse JA** (2002) Fatty acid export from the chloroplast: molecular characterization of a major plastidial acyl-coenzyme A synthetase from Arabidopsis. *Plant Physiol* **129**: 1700–1709
- Schuler MA, Werck-Reichhart D** (2003) Functional genomics of P450s. *Annu Rev Plant Biol* **54**: 629–667
- Severson RE, Arrendale RE, Chorlyk OT, Green CR, Thome FA, Stewart JL, Johnson AW** (1985) Isolation and characterization of the sucrose esters of the cuticular waxes of green tobacco leaf. *J Agric Food Chem* **33**: 870–875
- Shockey JM, Fulda MS, Browse J** (2003) Arabidopsis contains a large superfamily of acyl-activating enzymes: phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol* **132**: 1065–1076
- Slocombe SP, Schauvinhold I, McQuinn RP, Besser K, Welsby NA, Harper A, Aziz N, Li Y, Larson TR, Giovannoni J, et al** (2008) Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiol* **148**: 1830–1846
- Stewart C Jr, Kang BC, Liu K, Moore SL, Yoo EY, Kim BD, Paran I, Mazourek M, Jahn MM** (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. *Plant J* **42**: 675–688
- Stewart C Jr, Mazourek M, Stellari GM, O'Connell M, Jahn M** (2007) Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J Exp Bot* **58**: 979–991
- St-Pierre B, De Luca V** (2000) Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Adv Phytochem* **34**: 285–315
- Suzuki T, Fujiwaka H, Iwai K** (1980) Intracellular localization of capsaicin and its analogs capsaicinoid in *Capsicum* fruit. 1. Microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cv. Karayatsubusa. *Plant Cell Physiol* **21**: 839–853
- Suzuki T, Iwai K** (1984) Constituents of red pepper spices: chemistry, biochemistry, pharmacology and food science of the pungent principle

- of *Capsicum* species. In A Brossi, ed, *The Alkaloids: Chemistry and Pharmacology*, Vol 23. Academic Press, Orlando, FL, pp 227–299
- Tewksbury JJ, Nabhan GP** (2001) Seed dispersal: directed deterrence by capsaicin in chilies. *Nature* **412**: 403–404
- Tewksbury JJ, Reagan KM, Machnicki NJ, Carlo TA, Haak DC, Peñaloza AL, Levey DJ** (2008) Evolutionary ecology of pungency in wild chilies. *Proc Natl Acad Sci USA* **105**: 11808–11811
- Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M** (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. *Proc Natl Acad Sci USA* **97**: 11192–11197
- Todd PH, Bensinger MG, Biftu T** (1977) Determination of pungency due to Capsicum by gas-liquid chromatography. *J Food Sci* **42**: 660–667
- Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S** (2002) Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. *Plant Cell* **14**: 1441–1456
- van der Hoeven RS, Steffens JC** (2000) Biosynthesis and elongation of short- and medium-chain-length fatty acids. *Plant Physiol* **122**: 275–282
- Walpole CSJ, Wrigglesworth R, Bevan S, Campbell EA, Dray A, James IF, Masdin KJ, Perkins MN, Winter J** (1993) Analogues of capsaicin with agonist activity as novel analgesic agents: structure-activity studies. 3. The hydrophobic side-chain “C-region”. *J Med Chem* **36**: 2381–2389
- Walters DS, Steffens JC** (1990) Branched chain amino acid metabolism in the biosynthesis of *Lycopersicon pennellii* glucose esters. *Plant Physiol* **93**: 1544–1551
- Walton NJ, Mayer MJ, Narbad A** (2003) Molecules of interest: vanillin. *Phytochemistry* **63**: 505–515
- Wildermuth MC, Dewdney J, Wu G, Ausubel FM** (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**: 562–565
- Winkel BSJ** (2004) Metabolic channeling in plants. *Annu Rev Plant Biol* **55**: 85–107
- Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD** (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theor Appl Genet* **118**: 1279–1293
- Wu FN, Mueller LA, Crouzillat D, Petiard V, Tanksley SD** (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* **174**: 1407–1420
- Zamski E, Shoham O, Palevitch D, Levy A** (1987) Ultrastructure of capsaicinoid-secreting cells in pungent and nonpungent red pepper *Capsicum annuum* L. cultivars. *Bot Gaz* **148**: 1–6
- Zygiar S, Chaim AB, Efrati A, Kaluzky G, Borovsky Y, Paran I** (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. *Theor Appl Genet* **111**: 437–445