Plant Cell Physiol. 41(1): 119–123 (2000)

Short Communication

Identification of Multi-Gene Families Encoding Isopentenyl Diphosphate Isomerase in Plants by Heterologous Complementation in *Escherichia coli*

Francis X. Cunningham, Jr. [1, 2] **and Elisabeth Gantt** [1]

[1] *Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, U.S.A.*

Two cDNAs encoding isopentenyl diphosphate isomerase (IPI) in *Adonis aestivalis*, *Arabidopsis thaliana*, and *Lactuca sativa*, and single examples from *Oryza sativa* and *Tagetes erecta* were identified. An analysis of these and other *ipi* leads us to suggest a separate origin for green algal and plant genes and propose that a single gene encodes plastid and cytosolic IPI in plants.

Key words: Carotenoids — Color complementation — Dimethylallyl diphosphate — Isoprenoids — Lycopene — Phylogeny.

The isoprenoids encompass a remarkable variety of essential and secondary products in plants, animals, fungi and bacteria. The five carbon isopentenyl diphosphate (IPP) and its' allylic isomer dimethylallyl diphosphate (DMAPP) are the central metabolites and essential precursors for the biosynthesis of these myriad compounds. The enzyme IPP isomerase (IPI) catalyzes the reversible conversion of IPP to DMAPP (Ramos-Valdivia et al. 1997). A biosynthetic route to IPP from acetyl-CoA via 3-hydroxy-3-methylglutaryl-CoA and mevalonic acid (Fig. 1) is well established (Goldstein and Brown 1990), and has been found in plants, animals, fungi and many bacteria. The more recently recognized deoxyxylulose-5-phosphate (DOXP) pathway (Fig. 1; Lichtenthaler 1999) leading to IPP and/or DMAPP has been found in certain bacteria, in algae, and in the plastids of plants.

Isoprenoids are produced in several different compartments in plant cells (Bach 1995). Geranylgeranyl diphosphate (GGPP) synthase, which utilizes 3 IPP and 1 DMAPP to produce GGPP, has been localized in or shown to be targeted to mitochondria, plastids, the cytosol, and the endoplasmic reticulum in *Arabidopsis thaliana* (Zhu et al. 1997, Okada et al. 1999[3]). Whether IPI is also present in each of these compartments in plants has not been established. A cytosolic location, the location also of the mevalonate pathway leading to IPP in plants, is not in dispute (Bach 1995). Evidence for the occurrence of IPI in the plastids is also ample, and indeed the enzyme has been isolated from the chromoplasts of several plants (see Bach 1995, Ramos-Valdivia et al. 1997). There is no direct evidence, as yet, for the occurrence of IPI in mitochondria, the endoplasmic reticulum, or peroxisomes of plants, but this enzyme has been found to be targeted to the peroxisomes of mammals (Aboushadi et al. 1999).

The GGPP synthases destined for transport to various cell compartments in *A. thaliana* are the products of distinct genes (Zhu et al. 1997, Okada et al. 1999[3]), and at least 7 different GGPP synthase cDNAs or genes (or close homologues) may be discerned in a search of the nucleotide sequence data bases. Do plants also contain a family of genes encoding IPI, with transcripts from distinct genes used to produce polypeptides targeted to specific compartments? To help answer this question we identified, sequenced, and analyzed cDNAs encoding IPI in several plant species.

A complementation strategy earlier employed to isolate green algal and yeast IPI cDNAs (Kajiwara et al. 1997, Sun et al. 1996, 1998) was used to screen for plant cDNAs encoding IPI. Plant cDNA libraries were introduced into a lycopene-accumulating strain of *Escherichia coli*, the cultures were spread on agar plates containing ampicillin to select for transformants, and the occasional colonies that displayed a darker pink color, indicative of an enhanced accumulation of the pink-colored isoprenoid lycopene, were selected. Details of the screening protocol are given elsewhere (Cunningham et al. 1996).

The great majority of cDNAs obtained in screens of the various plant libraries (ca. 90% of more than 200 cDNAs examined by us) predicted polypeptides similar in

Abbreviations: DMAPP, dimethylallyl diphosphate; DOXP, D-deoxyxylulose-5-phosphate; DXR, deoxyxylulose-5-phosphate reductoisomerase; DXS, deoxyxylulose-5-phosphate synthase; GGPP, geranylgeranyl diphosphate; IPP, isopentenyl diphosphate; IPI, isopentenyl diphosphate isomerase.

The nucleotide sequences reported in this paper have been submitted to the GenBank under accession numbers listed in the legend to Figure 2.

[2] Corresponding author: E-mail address fc18@umail.umd.edu; fax number (301) 314-9489.

[3] Okada, K., Saito, T., Nakagawa, T., Kawamukai, M. and Kamiya, Y. (1999) Localization and expression patterns of geranylgeranyl diphosphate synthases in *Arabidopsis thaliana*. Fourth European Symposium on Plant Isoprenoids, April 21–23, Barcelona.

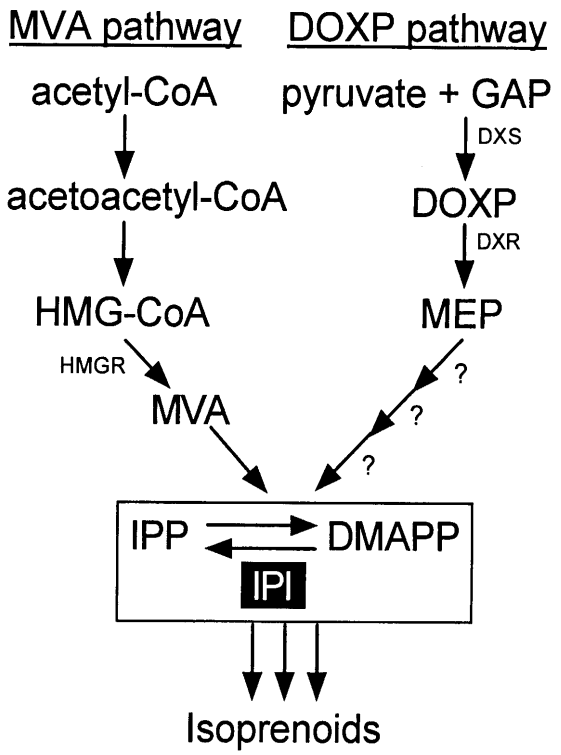


Fig. 1 Isopentenyl diphosphate (IPP) is produced by two different pathways in plants. A route to IPP from acetyl-CoA via mevalonate (MVA) is found in the cytosol. A route from pyruvate and glyceraldehyde-3-phosphate (GAP) via deoxyxylulose-5-phosphate (DOXP) operates in chloroplasts. Later steps in the DOXP pathway remain to be determined and are therefore denoted with question marks. The enzyme IPP isomerase (IPI) interconverts IPP and its' allylic isomer dimethylallyl diphosphate (DMAPP). The DOXP pathway has been shown to lead directly to IPP, but there may be a direct route to DMAPP as well. DXR, DOXP reductoisomerase; DXS, DOXP synthase; GAP; D-glyceraldehyde-3-phosphate; HMG-CoA, hydroxymethylglutaryl coenzyme A; HMGR, HMG-CoA reductase; MEP, 2-C-methyl-D-erythritol-4-phosphate.

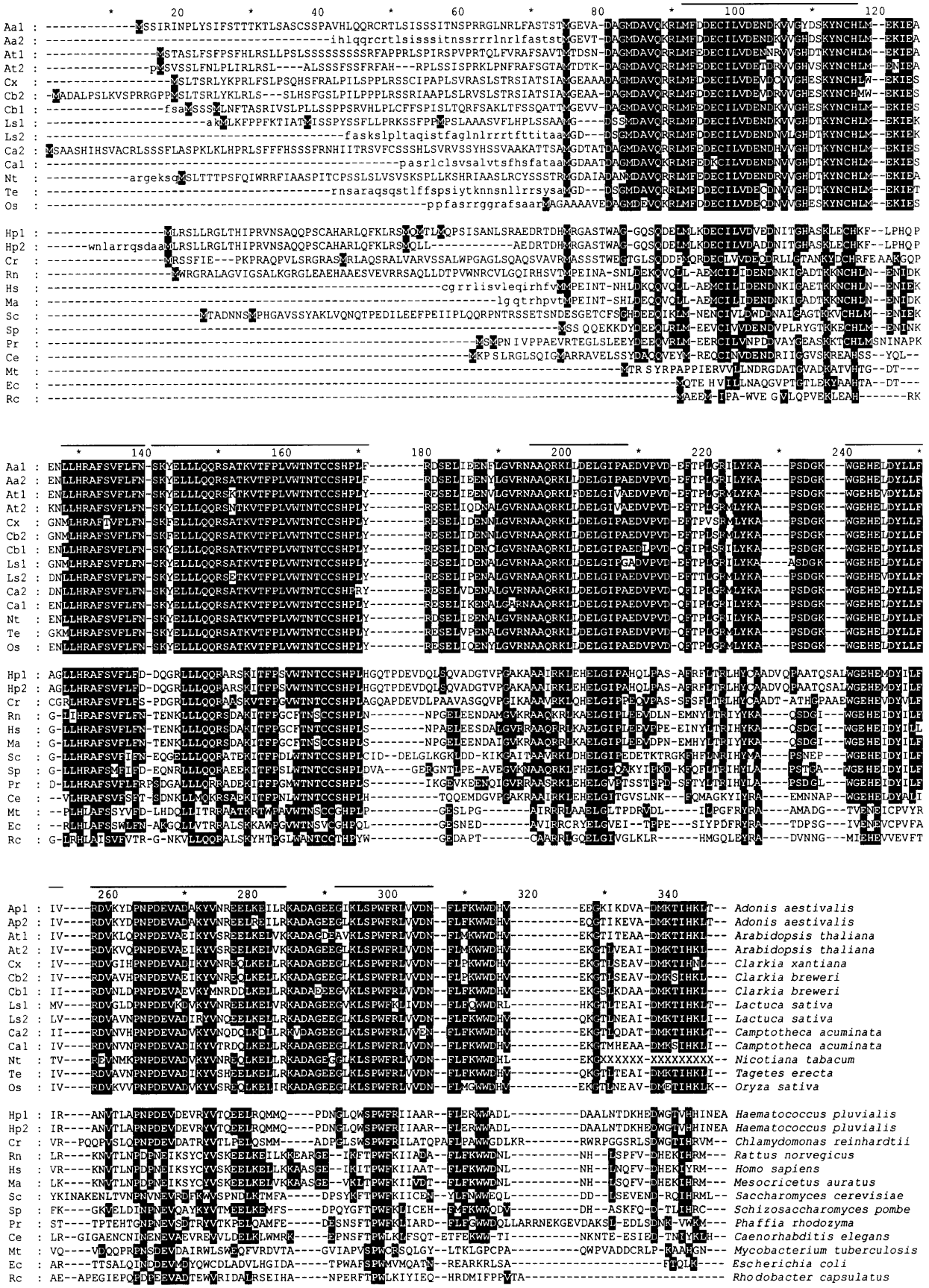
sequence to previously described IPI. Complete sequences for two distinct *Adonis aestivalis* (pheasant's eye) IPI cDNAs, two from *Arabidopsis thaliana* (essentially identical to two described earlier by Campbell et al. 1998), two from *Lactuca sativa* (romaine lettuce), and single examples from *Oryza sativa* (rice) and *Tagetes erecta* (marigold) were ascertained. The predicted amino acid sequences of these IPI are given in Figure 2 along with those for other known and putative IPI.

Guided by the amino acid sequence alignment displayed in Figure 2, the relatedness of DNA sequences encoding IPI was examined. Results of one such analysis are displayed in Figure 3. The 14 plant sequences, including one obtained from a monocot (*Oryza sativa*, rice), grouped together in a tight polytomic cluster well separated from the other IPI sequences. An analysis of the plant sequences alone, using a larger data set (see Fig. 3 legend), resulted in the placement of the *O. sativa ipi* and *C. breweri ipi*1 into a group with the *C. breweri ipi*2 and *C. xantiana ipi* (not shown) and improved bootstrap support for the groupings already shown in Figure 3.

Because at least two compartments in plant cells contain IPP isomerase, we had expected that an analysis of the sequences of the available plant genes and cDNAs would yield at least two separate and distinct groups: e.g. one for the enzyme localized in the cytosol and one for the enzyme targeted to the plastid, and perhaps additional classes for peroxisomal or mitochondrial or endoplasmic reticulum isoforms, if such exist. This was not observed whether analysis was of the nucleotide sequences (using parsimony, maximum likelihood or neighbor-joining; Fig. 3 and data not shown) or of the predicted amino acid sequences (by parsimony or neighbor-joining; data not shown). For those five plant species for which there were two *ipi* cDNAs available, three had isomerases that were most similar in DNA and predicted amino acid sequence to their same species partners (Fig. 3; amino acid sequence identities of 97.4% for the *A. aestivalis*, 91.7% for the *A. thaliana* and 90.0% for the *Camptotheca acuminata* pairs; residues prior to position 81 in the alignment of Fig. 2 were excluded). Sequence identities for the other two pairs, 86.8% for *Clarkia breweri* and 86.5% for *L. sativa*, were also high.

The relatively high DNA and amino acid sequence similarity for all plant IPI (at least 83% amino acid identity

Fig. 2 Alignment of predicted amino acid sequences of isopentenyl diphosphate isomerases. The 14 plant sequences are grouped together above the others. Amino acid residues are in white text on a black background where at least half (14 or more) of the aligned residues in that position are identical. Prospective initiating Met in the first 100 positions are also in white text on a black background. For those cDNAs in which a stop codon, in frame, did not precede the first Met, the predicted amino acids preceding this Met are shown in lowercase letters. The letter "X", used for the last 15 residues of the *N. tabacum* sequence (Nt), denotes missing data. The alignment was obtained by minor adjustment of output produced by ClustalW version 1.7 using a gap opening penalty of 25 and gap extension penalty of 0.05. Horizontal lines above the alignment indicate those regions used for phylogenetic analysis (Fig. 3). GenBank accession numbers: Plants: Aa1, AF188060; Aa2, AF188061; At1, AF188067 (U48961 and U47324 are nearly identical; because these three cDNAs are incomplete at the 5′ end, the first 54 amino acid residues displayed for At1 were obtained by translation of the gene: AB005242: 38477..40287); At2, AF188066 (U49259 is nearly identical) Cb1, X82627; Cb2, U48963 (genomic); Cx, U48962 (genomic); Ls1, AF188062; Ls2, AF188063; Ca2, AF031080; Ca1, AF031079; Nt, Y09634 (Note: we "corrected" the cDNA sequence given under this accession by deleting one of the two G's at bases 491–492 and one of the two A's at bases 521–522); Te, AF188064; Os, AF188065. Green algae: Hp1, AF082325; Hp2, AF082326; Cr, AF082869. Mammals: Rn, AF003835; Hs, X17025; Ma, AF003836. Yeast/Fungi: Sc, J05090; Sp, U21154; Pr, Y15811 (genomic). Nematode: Ce, L15314 (genomic; only the C-terminal 236 amino acids of the 831 amino acid hypothetical protein K06H7.3 are displayed). Eubacteria: Mt, Z95890: 3726..4337 (gene Rv1745c); Ec, U28375: 42879..43427 (orf182). Photosynthetic bacteria: Rc, Z11165: 22575..23105 (orf176).



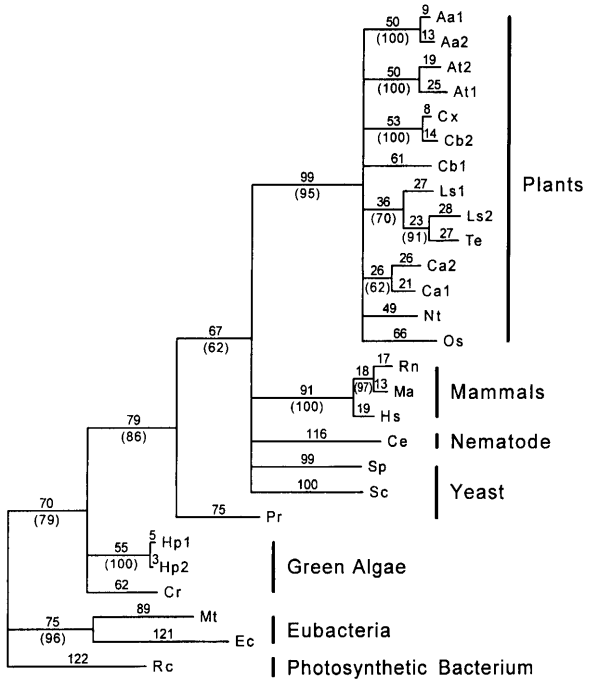


Fig. 3 A phylogeny of isopentenyl diphosphate isomerase. Analysis was by a heuristic search using maximum parsimony (PAUP* version 4.0.0d4; Swofford 1993). The three bacterial sequences were used as the outgroup. DNA sequences were aligned codon by codon in accordance with the alignment of the amino acid sequences displayed in Fig. 2. Regions of uncertainty in the alignment were excluded in the analysis (see the legend of Fig. 2). The data set used for analysis included a total of 411 nucleotide bases (137 codons). For analysis of the plant sequences alone (not shown; see text), a larger data set of 630 nucleotide bases (210 codons; from the conserved D at position 81 to the end of the alignment displayed in Fig. 2) was used. Branch lengths are drawn to scale and the number of nucleotide substitutions is listed above each branch. Bootstrap values greater than 50% (based on 1000 bootstrap replicates) are indicated, in parentheses, below the branches.

with no gaps for a comparison of any two of the plant sequences) and lack of any obvious dichotomy in grouping do not support the concept of distinct ancestral genes or an ancient gene duplication giving rise to contemporary genes encoding plastid, cytosolic, and other IPI isoforms in plants. It can not be discounted that the plant sequences so far identified represent only one of two or more classes of IPI that exist in plant cells. However, we think this is unlikely because these sequences, though only 14 in number, represent an analysis of more than 200 cDNAs obtained after exhaustive searches of a number of plant cDNA libraries.

What then might be the source of plastid, cytosolic and other compartmental isoforms of IPI? An examination of the N terminal region of the predicted amino acid sequences in the alignment of Figure 2 suggests one possibility: translation from two or more of the several well-conserved Met that precede that portion of the polypeptide required for enzymatic activity (beyond residue 95 in the alignment; see Campbell et al. 1998, Sun et al. 1998). The use of alternative Met for initiation of translation is one of several strategies employed by plants to produce polypeptides needed in more than one compartment (reviewed by Small et al. 1998).

Of the 14 predicted plant IPI sequences displayed in Figure 2, nine have extended amino termini with characteristics suggestive of a chloroplast transit peptide. When analyzed with ChloroP (a prediction program entrained on 75 known plastid transit peptides; Emanuelsson, et al. 1999) these nine and two of the three green algal isomerases (Hp2 is the exception) were predicted to be targeted to the chloroplast with cleavage sites forecast between positions 62 and 82 (most after position 72) in the alignment. Three of the remaining five plant sequences (Aa2, Ls2, and Te in Fig. 2) were also predicted to be plastid-targeted (with cleavage sites between 72 and 82) when analysis included the predicted amino acid sequence upstream of the first Met at position 75 in Figure 2. The remaining two plant sequences (Ca1, Os) also do not have a stop codon upstream of the first Met, and may well be of alternative or degraded transcripts that, when full length, encode plastid-targeted polypeptides.

Quite unexpectedly, the three green algal *ipi* cDNA and predicted amino acid sequences were found to be less like the plant sequences than those of genes or cDNAs encoding mammalian IPI (Fig. 3 and data not shown). The predicted *C. reinhardtii* and *H. pluvialis* amino acid sequences, for example, were less than 44% identical (with 7 or 8 gaps in the alignment and residues prior to position 81 in Fig. 2 excluded) to the *L. sativa* IPI2, while a comparison of the human and *L. sativa* sequences yielded an identity of more than 54% (with only 4 gaps). The predicted amino acid sequences of the algal IPI can be seen to have a number of insertions and deletions not found in other IPI (Fig. 2). The relatively low DNA and predicted amino acid sequence identity for plant and green algal IPI argues against a common origin for the plant and green algal genes.

For neither plants nor green algae does it appear likely that those genes so far identified as encoding IPI were acquired from an endosymbiotic plastid progenitor. The genome of the cyanobacterium *Synechocystis* PCC 6803 (Kaneko et al. 1996) does not contain any open reading frames encoding an obvious homologue of the known IPI, nor is there one apparent in the three extrachromosomal elements of this cyanobacterium (Kotani and Tabata 1998[4]). In nine other completed bacterial genomes (*Aqui-*

[4] The nucleotide sequences of the three extrachromosomal elements in *Synechocystis* PCC6803 (pSYSG, pSYSA, and pSYSM) were kindly provided by Dr. Satoshi Tabata of the Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba, 292-0812 Japan.

fex aeolicus, Mycobacterium tuberculosis, Bacillus subtilus, Chlamydia trachomatis, Helicobacter pylori 26695, *Helicobacter pylori* J99, *Haemophilus influenzae, E. coli*, and *Treponema pallidum*) which, like *Synechocystis* PCC 6803, contain open reading frames encoding homologues of the two known DOXP pathway enzymes (DXS and DXR) but not of the mevalonate pathway enzyme hydroxymethylglutaryl-CoA reductase (the genomes were examined using the homology search feature at http://mbgd.genome.ad.jp/), only two (*E. coli* and *M. tuberculosis*) also contain an open reading frame that specifies an IPI homologue. And for these two exceptions the *ipi* may well have been secondarily acquired from a eucaryotic host, as is thought to have occurred for several other *M. tuberculosis* genes (as reported by Winston Hide at the Microbial Genomes III meeting; see *Science* 284: 1305).

The *E. coli ipi* encodes a polypeptide with demonstrated IPI activity, but insertional inactivation of the gene is of little or no apparent consequence to growth of the organism (Hahn et al. 1999). In contrast, inactivation of the single gene encoding IPI in *Saccharomyces cereviseae*, an organism that contains the mevalonate pathway, is lethal (Mayer et al. 1992). The obligate nature of this enzyme in yeast presumably derives from the need to convert IPP, the product of the mevalonate pathway, to the DMAPP that serves as the initial allylic diphosphate primer for long chain isoprenoid biosynthesis. Two recent reports indicate that IPP is also an initial product of the DOXP pathway (the route to IPP does not proceed via DMAPP; Arigoni et al. 1999, McCaskill and Croteau 1999). However, a direct route to DMAPP also, whether by a branch in the pathway or by a catalytically promiscuous terminal step that could yield either DMAPP or IPP, has not been ruled out. Thus, in addition to the obvious possibility of an IPP isomerase with little or no sequence similarity to the known IPI, we must entertain the possibility that isoprenoid production via the DOXP pathway does not require an IPP isomerase.

This work was supported in part by grants from DOE (DEFG0298ER2032) and NSF (MCB 9631257). We thank Matt Cimino of the Department of Cell Biology and Molecular Genetics at the University of Maryland for advice and assistance with the phylogenetic analysis of IPP isomerase. We are grateful to Dr. Satoshi Tabata of the Kazusa DNA Research Institute for providing us with the nucleotide sequences of the three extrachromosomal elements of *Synechocystis* PCC6803 prior to publication.

References

Aboushadi, N., Engfelt, W.H., Paton, V.G. and Krisans, S.K. (1999) Role of peroxisomes in isoprenoid biosynthesis. *J. Histochem. Cytochem.* 47: 1127–1132.

Arigoni, D., Eisenreich, W., Latzel, C., Sagner, S., Radykewicz, T., Zenk, M.H. and Bacher, A. (1999) Dimethylallyl pyrophosphate is not the committed precursor of isopentenyl pyrophosphate during terpenoid biosynthesis from 1-deoxyxylulose in higher plants. *Proc. Natl. Acad. Sci. USA* 96: 1309–1314.

Bach, T.J. (1995) Some new aspects of isoprenoid biosynthesis in plants—a review. *Lipids* 30: 191–202.

Campbell, M., Hahn, F.M., Poulter, C.D. and Leustek, T. (1998) Analysis of the isopentenyl diphosphate isomerase gene family from *Arabidopsis thaliana*. *Plant. Mol. Biol.* 36: 323–328.

Cunningham, F.X. Jr., Pogson, B., Sun, Z., McDonald, K.A., DellaPenna, D. and Gantt, E. (1996) Functional analysis of the β and ε lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. *Plant Cell* 8: 1613–1626.

Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8: 978–984.

Goldstein, J.L. and Brown, M.S. (1990) Regulation of the mevalonate pathway. *Nature* 343: 425–430.

Hahn, F.M., Hurlburt, A.P. and Poulter, C.D. (1999) *Escherichia coli* open reading frame 696 is idi, a nonessential gene encoding isopentenyl diphosphate isomerase. *J. Bacteriol.* 181: 4499–4504.

Kajiwara, S., Fraser, P.D., Kondo, K. and Misawa, N. (1997) Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochem. J.* 324: 421–426.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouch, T., Matsuno, A., Murak, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3: 109–136.

Kotani, H. and Tabata, S. (1998) Lessons from sequencing of the genome of a unicellular cyanobacterium, *Synechocystis* sp. PCC6803. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49: 151–171.

Lichtenthaler, H.K. (1999) The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid synthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 47–65.

Mayer, M.P., Hahn, F.M., Stillman, D.J. and Poulter, C.D. (1992) Disruption and mapping of IDI1, the gene for isopentenyl diphosphate isomerase in *Saccharomyces cerevisiae*. *Yeast* 9: 743–748.

McCaskill, D. and Croteau, R. (1999) Isopentenyl diphosphate is the terminal product of the deoxyxylulose-5-phosphate pathway for terpenoid biosynthesis in plants. *Tetrahedr. Lett.* 40: 653–656.

Ramos-Valdivia, A.C., van der Heijden, R. and Verpoorte, R. (1997) Isopentenyl diphosphate isomerase: a core enzyme in isoprenoid biosynthesis. A review of its biochemistry and function. *Nat. Prod. Rep.* 14: 591–603.

Small, I., Wintz, H., Akashi, K. and Mireau, H. (1998) Two birds with one stone: genes that encode products targeted to two or more compartments. *Plant Mol. Biol.* 38: 265–277.

Sun, Z., Cunningham, F.X., Jr. and Gantt, E. (1998) Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. *Proc. Nat. Acad. Sci. USA* 95: 11482–11488.

Sun, Z., Gantt, E. and Cunningham, F.X., Jr. (1996) Cloning and functional analysis of the β-carotene hydroxylase of *Arabidopsis thaliana*. *J. Biol. Chem.* 271: 24349–24352.

Swofford, D.L. (1993) *PAUP*: Phylogenetic Analysis Using Parsimony.* version 4.0.0d4. Sinauer Assoc., Inc.

Zhu, X.F., Suzuki, K., Saito, T., Okada, K., Tanaka, K., Nakagawa, T., Matsuda, H. and Kawamukai, M. (1997) Geranylgeranyl pyrophosphate synthase encoded by the newly isolated gene GGPS6 from *Arabidopsis thaliana* is localized in mitochondria. *Plant. Mol. Biol.* 35: 331–341.

(Received July 26, 1999; Accepted October 27, 1999)