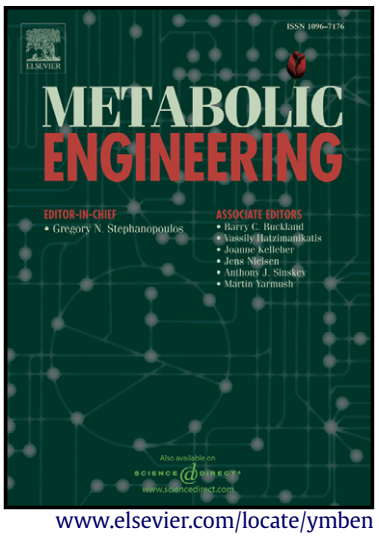

www.elsevier.com/locate/ymben

Identification of novel isoprene synthases through genome mining and expression in *Escherichia coli*

Marja Ilmén, Merja Oja, Anne Huuskonen, Sangmin Lee, Laura Ruohonen, Simon Jung

PII: S1096-7176(15)00098-1
DOI: http://dx.doi.org/10.1016/j.ymben.2015.08.001
Reference: YMBEN1026

To appear in: *Metabolic Engineering*

Received date: 27 January 2015
Revised date: 15 June 2015
Accepted date: 3 August 2015

Cite this article as: Marja Ilmén, Merja Oja, Anne Huuskonen, Sangmin Lee, Laura Ruohonen, Simon Jung, Identification of novel isoprene synthases through genome mining and expression in *Escherichia coli*, *Metabolic Engineering,* http://dx.doi.org/10.1016/j.ymben.2015.08.001

Identification of Novel Isoprene Synthases through Genome Mining and Expression in *Escherichia coli*

Marja Ilmén[1], Merja Oja[1], Anne Huuskonen[1], Sangmin Lee[2], Laura Ruohonen[1] and Simon Jung[2]*

[1]*VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Espoo, Finland*

[2]*Global Technology, SK innovation, Daejeon, Republic of Korea*

**Corresponding author Tel.: +82 42 609 8024; Fax: +82 42 609 8929.*

E-mail address: jungmoongeun@sk.com

Abstract

Isoprene is a naturally produced hydrocarbon emitted into the atmosphere by green plants. It is also a constituent of synthetic rubber and a potential biofuel. Microbial production of isoprene can become a sustainable alternative to the prevailing chemical production of isoprene from petroleum. In this work, sequence homology searches were conducted to find novel isoprene synthases. Candidate sequences were functionally expressed in *Escherichia coli* and the desired enzymes were identified based on an isoprene production assay. The activity of three enzymes was shown for the first time: expression of the candidate genes from *Ipomoea batatas*, *Mangifera indica*, and *Elaeocarpus photiniifolius* resulted in isoprene formation. The *Ipomoea batatas* isoprene synthase produced the highest amounts of isoprene in all experiments, exceeding the isoprene levels obtained by the previously known *Populus alba*

and *Pueraria montana* isoprene synthases that were studied in parallel as controls.

Keywords

isoprenoid, terpene, genome mining, homology-based screening

Abbreviations

IspS, isoprene synthase; DMAPP, dimethylallyl pyrophosphate; MEP, methylerythritol phosphate; MEV, mevalonate; IPTG, isopropyl β-D-1-thiogalactopyranoside; CDW, cell dry weight; TSP, terpene synthase; TSP-b, terpene synthase protein family clade b; MSA, multiple sequence alignment; RNA-seq, RNA sequencing.

1 Introduction

Isoprene is a naturally produced organic five-carbon chemical compound (C5). It has significant commercial potential, as it has been used to manufacture products ranging from rubber to adhesives. Isoprene, i.e. 2-methyl-1,3-butadiene, and other isoprenoids have been used in the pharmaceutical, nutraceutical, flavour, fragrance, and rubber industries (Ajikumar et al., 2010; Davies et al., 2014; Lemuth et al., 2011; Zhan et al., 2014). The industrial supply of isoprene is limited to petrochemical-based sources. Until now, commercially viable quantities of isoprene have been obtained by direct isolation from petroleum C5 cracking fractions or by dehydration of C5 isoalkanes or isoalkenes. However, the increasing global demand for isoprene

calls for novel sources of isoprene (Bentley and Melis, 2012; Davies et al., 2014; Lv et al., 2014; Miller et al., 2001; Xue and Ahring, 2011; Zhao et al., 2011), and the industry has a strong interest in finding a commercially viable and environmentally sustainable production process (Choi et al., 2015; Whited et al., 2010). Therefore, the development of microbial isoprene production is gaining increased interest.

Two alternative biosynthetic pathways, the methylerythritol phosphate (MEP) pathway (Zhao et al., 2013) and the mevalonate (MEV) pathway (Miziorko, 2011), produce isoprenoid precursors in living cells. The MEV pathway exists in Archaea, most eukaryotes and some bacteria, while most bacteria and chloroplasts synthesize isoprenoids via the MEP pathway (Chandran et al., 2011; Zhao et al., 2013). In green plants both pathways co-exist (Lohr et al., 2012). Isoprene synthase (EC 4.2.3.27) catalyses the enzymatic conversion of dimethylallyl pyrophosphate (DMAPP) to isoprene. Genes coding for isoprene synthases (IspS) are presently known only from the plant kingdom, although some bacteria are also known to produce isoprene (Hess et al., 2013).

Isoprene synthases have been characterized from relatively few plant families. Isoprene synthase sequences were first identified from grey poplar *Populus alba* (Miller et al., 2001; Sasaki et al., 2005), and the kudzu vine *Pueraria montana var. lobata* (Sharkey et al., 2005). Recently, Sharkey et al. (2013) identified novel isoprene synthases using a homology based database search (sequences from *Eucalyptus globulus* and *Melaleuca alternifolia*) and a

homology-based cloning approach (sequences similar to *P. alba* from *Populus* and *Salix* genera and sequences similar to *P. montana* from *Robinia pseudoacacia* and *Wisteria sp.).* Beatty et al. (2013) filed a patent application that describes IspS sequences from Fabales (similar to *P. montana*), including IspS from *Arachis hypogaea*. The patent also gives a sequence for the IspS from oak, *Quercus petraea* (Schnitzler et al., 1996). The kudzu and poplar enzmes have been most extensively studied. Whited et al. (2010) suggested that the kinetic properties of these enzymes are not optimal. Identification of new isoprene synthases may provide enzymes with improved kinetics and benefit the construction of microbial cell factories for isoprene production.

Microbes, e.g. *E. coli* and cyanobacteria, have been engineered to produce isoprene from renewable raw materials such as sugars (Whited et al., 2010) , or carbon dioxide (Kallas et al, 2013; Lindberg et al., 2010). In addition to IspS, components of the MEP or MEV pathway have been overexpressed in the recombinant hosts, resulting in enhancement of isoprene production relative to the strain expressing IspS alone (Bentley et al., 2013; Kallas et al, 2013; Yang et al., 2012; Zhao et al., 2011). The aim of the present work was to identify novel isoprene synthases that perform well in a microbial host producing isoprene. Our approach was to identify candidate genes coding for isoprene synthase using sequence based homology searches. The retrieved sequences (9123) were grouped to separate the candidate IspS genes from other terpene synthases, and the candidates were screened for the presence of identified

IspS signatures (Sharkey et al., 2013). The top 9 candidates were chosen for *in vivo* functional testing for isoprene production in *E. coli.*

2 Material and methods

2.1 Sequence homology searches for novel isoprene synthases

To identify novel candidate isoprene synthase enzymes, homology based database searches were conducted. Public databases were queried with sequences reported to have isoprene synthase activity. These included:

- well characterized IspS sequences from *P. alba* (Uniprot: Q50L36; Miller et al. (2001)), *Populus canescens* (Uniprot: Q9AR86, PDB: 3N0G; (Köksal et al., 2010)) and *P. montana var. lobata* (Uniprot: Q6EJ97; (Sharkey et al., 2005)),
- IspS sequences from Patent WO 2013/166320 A1 (Beatty et al., 2013): *A. hypogaea* (SEQ ID 3), *Glycine max* (SEQ ID 5 and SEQ ID 7), *Mucuna pruriens* (SEQ ID 9), *Cajanus cajans* (SEQ ID 11) and *Q. petraea* (SEQ ID 13),
- sequences from Sharkey et al. (2013): isoprene synthases from *Wisteria sp.* (GenBank: AEK70969), *R. pseudoacacia* (GenBank: AEK70968), *M. alternifolia* (GenBank: AAP40638), *E. globulus* (GenBank: BAF02831), *Salix sp.* (GenBank: AEK70970), and
- bifunctional enzymes shown to have some isoprene synthase activity: myrcene synthase from *Humulus lupulus* (GenBank: ACI32638; (Sharkey

et al., 2013) and a 2-methyl-3-buten-2-ol (MBO) synthase from *Pinus sabiniana* (GenBank: AEB53064; (Gray et al., 2011)).

The homology based searches were conducted against Uniprot (SwissProt and TrEMBL) and GenBank protein databases (nr, pat and env_nr) using blastp, and against GenBank nucleotide databases (tsa_nt, env_nt and pat) using tblastn. The GenBank pat databases contain sequences from patents, deposited by the U.S. Patent and Trademark Office (Benson et al., 2013). GenBank tsa_nt and env_nt databases contain DNA sequences assembled based on shotgun sequencing data, such as RNA sequencing or metagenomics sequencing (Benson et al., 2013). The sequences derived from plant transcriptomics projects are a particularly promising source of candidate isoprene synthases. Sequences with E-value smaller or equal to 1e-30 were extracted in each case. Nucleotide sequences were translated to protein sequences with the GeneWise (Birney et al., 2004) program using the corresponding query sequence as a guide in the translation.

To provide reference sequences for annotating the blast search results, members of the "Terpene synthase, metal-binding domain" protein family

(Pfam[1]: PF03936, Interpro[2]: IPR005630) were retrieved from SwissProt (a manually curated subset of Uniprot) based on the InterPro domain annotations.

To remove redundancy among the retrieved sequences the sequences were clustered using the CD-HIT algorithm (Li and Godzik, 2006), using the command "cd-hit -c 0.8". After the clustering, each cluster contains sequences that are more than 80% identical to each other at the amino acid level. One representative sequence was selected from each cluster with preference given to sequences with annotation information. The remaining sequences were kept in reserve. The CD-HIT algorithm was also used with higher identity thresholds ("-c 0.95" or "-c 0.9") for subsets of the full data set.

2.2 Multiple sequence alignment and phylogenetic tree construction

The retrieved sequences were analysed based on a multiple sequence alignment (MSA) and a phylogenetic tree. The MSA was created by aligning the protein sequences to the terpene synthase protein family Pfam motif (PF03936) using the HMMer program (Eddy, 1998). The MSA was used as an input for the phylogenetic tree reconstruction algorithm FastTree. Bioperl scripts were used to convert MSAs between formats (stockholm, phylip or nexus). Geneious

[1] pfam.xfam.org

[2] www.ebi.ac.uk/interpro

software was used for visualization and manipulation of the MSAs and phylogenetic trees (Fig. 1). R programming (www.r-project.org) was used to manipulate lists of candidate sequences extracted from phylogenetic trees in Geneious. The phylogenetic trees were created from the reduced data set (i.e. CD-HIT clustering result), and when a subset of such a tree was taken we used programs written in R to bring back the sequences that had previously been set aside in the CD-HIT clustering step. The HMMer alignment and FastTree phylogenetic tree reconstruction algorithm were subsequently used again for the new smaller data set.

The final phylogenetic tree in Fig. 2 was created using Geneious Tree Builder and the final MSA alignment in Fig. 3 was created in Geneious using ClustalW.

2.3 Strains

E. coli TOP10 (Invitrogen) was used for cloning and plasmid propagation and *E. coli* BL21(DE3)Star (Invitrogen) for isoprene production. LB medium supplemented with 100 µg/ml ampicillin was used in all cultivations.

2.4 Cloning of IspS genes

Known IspS genes from *P. montana* (Uniprot: Q6EJ97, codon optimized for expression in *Synechocystis*; (Lindberg et al., 2010)), *P. alba* (Uniprot: Q50L36), and *A. hypogaea* (SEQ ID 3 in patent WO13166320 A1) were used as positive controls in parallel with the TSP-b protein family members chosen as putative isoprene synthase candidates after identification through homology

based searches, from *Ipomoea batatas, Elaeocarpus photiniifolius, Morus notabilis, Dahlia pinnata, Mangifera indica, Fragaria vesca subsp. vesca, Medicago sativa, Sesamum indicum,* and *Eucalyptus grandis.* The ChloroP1.1 program (http://www.cbs.dtu.dk/services/ChloroP/) was used to predict chloroplast targeting sequences in the candidate IspS genes. IspS genes, without the predicted chloroplast targeting sequence and with starting methionine added, as listed in Table1, were codon optimized to enable expression in either *E. coli* or *Synechocystis* sp. PCC6803 and were synthesized by Genscript (Hong Kong). The *D. pinnata* sequence appeared incomplete and therefore the sequence coding for amino acids MTARRSANYQ, deduced from IspS multiple sequence alignment, was added to the N-terminus. The IspS coding region was cloned between the T7 promoter and terminator in vector pET-Duet-1 (Novagen). A StrepII-tag was added to the N-terminus of the *E. photiniifolius* IspS by incorporating a sequence that is translated to MASWSHPQFEK.

2.5 Functional expression of IspS genes in E. coli

For isoprene synthase expression and functionality assays, expression plasmids were transformed into *E. coli* strain BL21(DE3)Star. Selected transformants and the parental strain were pregrown over night at 30℃ (or 37℃) in liquid LB medium (for transformants, supplemented with 100 µg/ml ampicillin). Cultures were diluted 1:50 into prewarmed medium and cultivated at

30℃ until the OD_{600} reached 0.6-0.7. Enzyme expression was induced with addition of 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside). Cultivations for isoprene analyses were carried out in sealed 22 ml head-space bottles at 30℃ or 37℃ with 1 or 2 ml medium with 230-250 rpm shaking. Biomass was measured as optical density (OD) at 600 nm.

2.6 Determination of isoprene

An automatic solid-phase microextraction (SPME) method was applied for isoprene analysis by using a divinylbenzene/carboxen/PDMS (DVB/CAR/PDMS) fibre (2 cm). The sampler was a Gerstel MPS system, which was connected to an Agilent 7890A gas chromatograph (GC) and a 5975C mass selective detector (MSD). An HP-InnoWax column (60 m, 0.25 mm ID, phase thickness 0.15 µm) was used for the analysis. For the fibre, the pre incubation time was 1 min and the incubation temperature was 40 ℃. The extraction time was 20 min with 250 rpm shaking and the desorption time was 8 min. The GC oven temperature program was from 40 ℃ (4 min) to 70 ℃ (5 ℃/min) and the column was cleaned by raising the temperature to 200 ℃ (30 ℃/min) for two minutes. The total run time was 16.3 min. The analyses were done in splitless mode and the temperature of the injector was 250 ℃. Helium was used as carrier gas at 1.2 mL/min. The MS data was collected at a range of m/z 35 to 300 and isoprene was identified by comparing the mass spectrum to that in the NIST08 library. The base peak is m/z 67, mass peak m/z 68 and other major fragments m/z 53 and 39. The m/z 67 peak was used for

quantification. Calibration curves were determined by spiking pre-cooled isoprene analytical standard (Fluka 59240, CAS 78-79-5) solution in ethanol into 2ml LB medium at various concentrations from 2 to 84.7 ng/2ml in 22 ml head space bottles. The curves proved to be linear with r2 values of 0.998-0.999. The quantification limit was estimated to be 0.5 ng/ml.

Isoprene production is expressed in volumetric units (ng/ml) or normalized to biomass assuming that OD=1 corresponded to 0.33 g/L dry cell weight (Sauer et al., 1999).

2.7 Preparation of E. coli whole cell extracts and detection of isoprene synthase on SDS-PAGE gels

Samples were taken from parallel headspace bottles. An aliquot from each sample was centrifuged, the supernatant discarded and the pellet suspended in 2x Laemmli sample buffer with Orange G dye (instead of bromophenol blue) and β-mercaptoethanol. Samples were heated 5 min at 95 ℃ and stored at -20 ℃. Before analyses, samples were thawed, heated 5 min at 95 ℃, and centrifuged 1 min, at 13200 rpm. Usually 10 µl of supernatant was loaded per lane. Proteins were separated on Criterion TGX 4-20% gel (Bio-Rad, or similar) and stained with GelCode Blue stain (Thermo Scientific).

3 Results and Discussion

3.1 Sequence homology based search for novel isoprene synthases

A homology based screening approach was used to identify novel isoprene synthase genes, using previously characterized isoprene synthases as query sequences. The GenBank tsa_nt database proved to be particularly useful. Several good candidate isoprene synthases were found from the plant RNA-seq derived sequences deposited in tsa_nt.

In total, 9123 sequences were retrieved: 278 sequences from Uniprot/SwissProt, 1989 from Uniprot/TrEMBL, 3953 from GenBank nucleotide databases, and 2905 from GenBank protein databases. These numbers include duplicate sequences, because the same sequence may be present in several databases. To remove redundant sequences, the retrieved sequences were clustered into clusters that contain sequences that are more than *x* % identical to each other at the amino acid level. In the first phase of analysis, *x* was set to 80%, and clustering reduced the number of sequences to 1929. The query sequences, and additional IspS sequences from the *Populus* genus, were added to the set of retrieved sequences before functional annotation of the sequences, bringing the total to 1945 sequences.

The retrieved sequence data contained various terpene synthases, not only isoprene synthases. This is mainly because the e-value threshold used in the BLAST search was intentionally set to be quite non-restrictive, but also because the databases were queried with bi-functional enzymes as well as mono-

functional isoprene synthases. The known isoprene synthases belong to the large terpene synthase (TPS) family. Besides isoprene synthases (C5), this protein family contains monoterpene (C10) synthases, sesquiterpene (C15) synthases and diterpene (C20) synthases (Chen et al., 2011). Overall, enzymes of the TPS family give rise to thousands of different types of terpenes. The family has been split into seven clades on the basis of phylogeny (Chen et al., 2011). Isoprene synthases belong to the TPS-b subfamily, together with monoterpene synthases such as β-ocimene synthase or limonene synthase. Many TSP-b enzymes produce cyclic monoterpenes (Chen et al., 2011; Sharkey et al., 2013).

To understand the diversity of functions among the retrieved terpene synthases a multiple sequence alignment (MSA) and a phylogenetic tree (Fig. 1) were created. The functional annotations of the sequences included in the tree were analysed to identify the terpene synthase sub families. All previously reported IspS sequences fall into the subtree containing sequences with annotations indicating that they are TPS-b subfamily members. The bifunctional sequences from *H. lupulus* and *P. sabiniana*, reported to have isoprene synthase activity, do not group together with other IspS sequences, as illustrated in Fig. 1. The TPS-b branch was analysed in more detail to identify subtrees that may contain isoprene synthases. Based on the annotations of the sequences, only the branch containing the currently known isoprene synthases is likely to contain good isoprene synthase candidates. For the next phase of analysis, sequences within this branch were extracted, also including sequences that had previously

been grouped together with these in the redundancy reduction step. A total of 618 sequences were included in the set of tentative IspS candidates.

The set of tentative IspS candidates was analysed using a similar workflow as for the full set of retrieved terpene synthases: a MSA and phylogenetic tree were created. The sequence features of the candidates were evaluated based on the MSA. The locations of the substrate binding amino acids were annotated based on the structure of PDBID:3N0G from *P. canescens* (Uniprot: Q9AR86). An “isoprene score” as defined by Sharkey et al. (2013) was also used to annotate the sequences. The “isoprene score” is based on conserved amino acids that are specific to isoprene synthases within the TPS-b family. The “isoprene score" amino acids are F338, S445, F485, and N505 (numbers based on *P. alba* sequence). The isoprene score is computed by counting how many of these four key amino acids are present in the sequence of interest. The best isoprene score is thus 4, and minimum 0. These “isoprene score” amino acid positions are indicated in the multiple sequence alignment of the finally selected IspS candidates (Fig. 3).

The MSA of the tentative IspS sequences was analysed, especially focusing particularly on the conserved amino acids and “isoprene score” amino acids, to reduce the list of IspS candidates. Several sequences were removed either because they had too low isoprene score or because they were more than 95% identical to previously characterized sequences. Some sequences for enzymes with confirmed functions other than isoprene synthase were not discarded, but

were included for reference: tricycle synthases from *Medicago truncatula* (Uniprot: Q5UB07) and *Lotus japonicus* (Uniprot: Q672F7), and β-ocimene synthases from *Vitis vinifera* (Uniprot: A5BLS5; (Martin et al., 2010) and *Matricaria recutita* (Uniprot: I6RE61). Eventually, 32 candidate isoprene synthases, 4 sequences with confirmed other functions, and 24 reference isoprene synthase sequences remained.

The 32 candidate isoprene synthase sequences were aligned and analysed in detail by comparing them to the 28 reference sequences. A phylogenetic tree is shown in Fig. 2 and a multiple sequence alignment showing key amino acid positions in some of the selected IspS candidates is shown in Fig. 3. This multiple sequence alignment shows that many of the candidate sequences are missing F338, an amino acid thought be important for isoprene synthase activity based on the structure of *P. canescens* IspS and Sharkey et al. (2013). To test whether F338 is critical, we selected some candidate sequences in which this phenylalanine is replaced with serine, valine or threonine. To compensate for the fact that the size of amino acid in position F338 is likely to be critical we selected candidates that have a relatively large isoleucine at V341, the next position that is on the same side of the α-helix as F338 in the 3D structure (*P. canescens*). Most of the previously characterized IspS have valine at this position. In total, nine isoprene synthases were selected for testing. The best candidates, *I. batatas* and *E. photiniifolius,* have an isoprene score of 4. The *M. indica* sequence has isoprene score of 3, but still has the key amino acid F338, and was thought likely to be a functional isoprene synthase. The remaining

candidates, from *M. sativa, F. vesca subsp. vesca, M. notabilis, D. pinnata, S. indicum* and *E. grandis* are missing amino acid F338, but still have a relatively large amino acid at this position.

3.2 Isoprene synthase functionality

The supply of isoprene precursors in the *E. coli* screening host relies solely on the native MEP pathway. The candidate IspS enzymes were tested for the ability to catalyze the conversion of DMAPP to isoprene *in vivo*. The enzymes were not tagged because tags may interfere with the activity of the IspS, as has been observed previously (Zurbriggen et al., 2012). *E. coli* BL21(DE3)Star cells transformed with the *IspS* expression vectors were cultivated in 2ml LB medium in sealed 22 ml head-space bottles and enzyme expression was induced with 0.5 mM IPTG.

3.2.1 Isoprene production by E. coli expressing known IspS

The *P. montana* and *P. alba* IspS produced isoprene while, somewhat surprisingly, the *A. hypogaea* IspS did not produce a significant amount of isoprene compared to the *E. coli* host as illustrated in Fig. 4A. *A. hypogaea* IspS activity has been reported, but its relatively low affinity for the substrate (Beatty et al., 2013) may explain why isoprene was not detectably produced in the present study where the assay conditions differed from those of Beatty et al. (2013). The ability to produce isoprene *in vivo* requires the isoprene synthase to efficiently compete for isoprene precursors that are also substrates for other cellular reactions.

Isoprene was produced abundantly even without IPTG addition, thus it is obvious that the T7 promoter was also substantially active under non-inducing conditions. Unexpectedly, in some cases more isoprene was detected in the absence than in the presence of IPTG. This may be explained by the fact that the non-induced cultures grew to a significantly higher cell density overnight than the induced cultures. The number of cells is expected to have a strong effect on the accumulation of a product, especially when the half-life of the enzyme is short or the catalytic rate is low.

3.2.2 Isoprene production by E. coli expressing novel IspS candidates

IspS candidates of *I. batatas* and *E. photiniifolius* have the highest isoprene score, 4, and both produced isoprene (Fig. 4A). The other candidates have a lower score. The *I. batatas* enzyme performed remarkably well in isoprene production, relative to the reference enzymes. Figure S1 shows an example of GS-MS analysis of head-space gases from the *E. coli* host and transformant expressing *I. batatas IspS*.

The candidate sequence from *M. indica*, score 3, has the key amino acid F338, but the conserved N505 is replaced by Ser. As illustrated in Fig. 4A and 4B, F338 was present in the isoprene producing candidates of *I. batatas, E. photiniifolius* and *M. indica,* whereas the six other candidates without F338 did not produce isoprene in the *E. coli* test system. It is known that *M. indica* leaves emit isoprene (Harley et al., 2004), thus the genome is also expected to contain a gene that codes for an isoprene synthase.

In comparison, in the candidates that did not produce detectable isoprene (Fig. 4B) F338 is replaced by either Ser (*M. sativa, F. vesca, D. pinnata*), Val (*M. notabilis, S. indicum*) or Thr (*E. grandis*) but they contain N505, with the exception of *S. indicum* that contains Lys at position 505. However, the *M. sativa and E. grandis* IspS polypeptides were not detected by SDS-PAGE, and therefore it is not possible to conclude that these two enzymes cannot produce isoprene.

Because the goal was to identify IspS enzymes that perform well when functionally expressed, we did not study in more detail those enzymes that were not good at producing isoprene in *E. coli* cellular environment. The possibility remains that some of the tested candidates may be able to catalyse isoprene formation under different conditions, as has been reported for the *A. hypogaea* enzyme (Beatty et al., 2013) that did not produce detectable isoprene in the present work although the *A. hypogaea* polypeptide was clearly visible on a SDS-PAGE gel (not shown). However, based on sequence homology comparisons, the proteins that did not produce isoprene are also less likely to be isoprene synthases. Table 2 summarizes the amino acid sequence features and experimental results.

The new isoprene synthases of *I. batatas, E. photiniifolius* and *M. indica* are not very similar to previously characterized IspS, as indicated in Fig. 5. They share the highest similarity to *Q. petraea* IspS with 55%, 60% 65% amino acid identity, for *I. batatas, E. photiniifolius* and *M. indica,* respectively. Furthermore, the new

IspS are less than 64% identical to each other at the amino acid level. Each of the new IspS originates from a different plant order. Previously, IspS have been only identified in rosids: orders Malpighiales (*P. alba*), Fabales (*P. montana*), Myrtales (*E. globulus*), and Fagales (*Q. petraea*). *E. photiniifolius* (order Oxalidales) and *M. indica* (order Sapindales) are also rosids, but the IspS from *I. batatas* is the first IspS from an asterid (order Solanales).

3.2.3 Introduction of F338 into M. notabilis and E. grandis proteins

Since the presence of F338 appeared to correlate with isoprene production, T338F and V338F mutations were generated in the *E. grandis* (species known to produce large quantities of isoprene) and the *M. notabilis* (sequence annotated as isoprene synthase in GenBank) proteins that did not produce detectable isoprene, in order to test whether the single amino acid change would result in isoprene production. *E. coli* expressing the mutated *E. grandis* or *M. notabilis* genes did not produce detectable isoprene (data not shown). The mutated *M. notabilis* protein was abundantly produced based on the SDS-PAGE gel, but the *E. grandis* protein was poorly expressed (data not shown). Thus it was not possible to conclude whether the latter enzyme was active or not.

We conclude that a simple mutation introducing F338 was not enough to introduce IspS functionality to a TSP-b family sequence. The other phenylalanine from the isoprene score amino acids defined by Sharkey et al., F485, was present in all known and candidate IspS. These two phenylalanines,

F338 and F485, are critical in reducing the size of the substrate binding site so that the larger terpene synthase substrates (geranyl diphosphate, farnesyl diphosphate and geranylgeranyl diphosphate) do not fit into the active site. The second Phe (F485) is missing from the *H. lupulus* bi-functional myrcene synthase that had lower isoprene synthase activity than other IspS (Sharkey et al., 2013). In addition to these phenylalanines, W317 and Y565, which are almost fully conserved in the TPS family, are essential in limiting the size of the substrate binding pocket (Köksal et al., 2010). All known and candidate IspS also had Y565, but *Q. petraea* has a phenylalanine instead of a tryptophan at the position corresponding to *P. alba* W317.

Of the other isoprene score amino acids, N505 has been reported to be critical in determining the ion requirements of the TPSs (Sharkey et al., 2013). Terpene synthases that do not require ions have a positively charged lysine in this position, others have serine or asparagine. In our study we found that N505 is not required for IspS activity, because the *M. indica* IspS has a serine at this position.

Sharkey et al. (2013) include S445 in the isoprene score and state that other TPS-b proteins most commonly have Val or Ile in the middle of a triple serine motif. However, based on our results S445 is the first position of the triple serine motif and is almost fully conserved in the protein family. The middle position of the triple serine motif, S446, does not seem to be critical for isoprene production, because the best IspS, from *I. batatas*, has valine at this position. Out of the

four isoprene score amino acids, F338, F485, S445 (or S446), and N505, only F338 and F485 are present in all functional IspS known to date.

3.2.4 Tagging the E. photiniifolius protein

The *E. photiniifolius IspS* protein was expressed at a low level based on SDS-PAGE analyses, but activity was clearly observed. This may suggest that the specific activity of the protein is relatively high, but that the protein is not very stable. In order to see whether tagging the *E. photiniifolius* IspS would be beneficial, a StrepII-tag was added onto either the N- or C- terminus of the protein. The N-terminally tagged protein could readily be seen on SDS-PAGE gel of *E. coli* whole cell extracts. As Fig. 4C and 4D illustrate, the introduction of the N-terminal tag increased the amount of *E. photiniifolius* IspS protein in the cell extracts, resulting in increased isoprene production relative to the non-tagged protein.

The calculated molecular weights of the IspS polypeptides range from 62 to 65 kD and they are visualized on SDS-PAGE gel between the 50-75 kD MW markers. A single band was seen for the *P. alba* (64 kD) and *I. batatas* (62 kD) IspS, whereas for the *M. indica* protein two bands were seen, one corresponding to the size of a full length polypeptide and the other to a lower molecular weight polypeptide, as shown in Fig. 4C. This suggested that the protein was somewhat unstable.

3.2.5 Isoprene production rate

The novel enzymes and the *P. alba* IspS were further assessed for isoprene production by taking samples after shorter incubation times, at 2 and 5 h (Fig. 6A), in addition to 24 h. Because the amount of isoprene produced from some 2 ml cultures was too high to be accurately measured after only 5 h, isoprene production was also measured from 1 ml cultures. The strains expressing *P. alba* and *I. batatas* IspS produced more isoprene at 37 ℃ than at 30 ℃ (Fig. S2). Isoprene production rate of the *P. alba* and *I. batatas* IspS enzymes at 37 ℃ was determined more accurately and samples taken at 1 h intervals. As shown in Fig. 6B, the rate of isoprene production by *I. batatas* IspS was 40 µg/L•h during the first 3 h, which was two times more than that of *P. alba* IspS. *E. coli* expressing the *P. montana IspS* or, alternatively, the P*. alba IspS* have been reported to produce 0.4 mg/L isoprene in 18 h (Zurbriggen et al., 2012; 22 µg/L•h) or 0.80 mg/L in 24 h (Yang et al., 2012; 33 µg/L•h), respectively.

Until now, the different isoprene synthases have mainly been studied individually and few studies have reported comparative data. The present work compared the performance of the well-known *P. montana* and *P. alba* enzymes directly with novel isoprene synthases in the same genetic background. It can be expected that enhancement of isoprenoid biosynthetic pathway in strains expressing the *I. batatas IspS* will lead to increased isoprene production, as has been shown for strains expressing the *P. alba* or *P. montana* IspS (Whited et al., 2010; Yang et al., 2012; Zurbriggen et al., 2012).

4 Conclusions

Sequence homology searches enabled identification of genes potentially encoding isoprene synthases. The corresponding amino acid sequences were examined for the presence of key amino acids in the active site of known IspS to identify the most probable candidates. Selected genes were expressed in *E. coli* and introduction of genes encoding the three highest ranking candidates resulted in isoprene production by the cells, indicating that our approach was effective in predicting isoprene synthases among uncharacterized terpene synthases. The main strength of the bioinformatics approach used in this work was the broadness of the search. Several databases that are not yet part of the main GenBank databases were searched to find novel isoprene synthase candidates. All novel IspS were found from GenBank tsa_nt, the database containing RNA-seq derived sequences. These sequences are not included in the main nucleotide (nt) and protein (nr) databases at GenBank. Additionally, the bioinformatics approach used in this work benefited from a custom-made data management workflow that allowed easy handling of the thousands of retrieved BLAST hits and included automated creation of alignments and phylogenetic trees. The same genome mining workflow can be used for any enzyme family of interest.

The three enzymes from *I. batatas, M. indica* and *E. photiniifolius* shown to have isoprene synthase activity are the first IspS representatives from the taxonomic groups Solanales, Sapindales and Oxalidales, respectively, and their overall similarity to each other and to previously known IspS sequences is

relatively low, 65% or less. The performance of *I. batatas* IspS compares favourably with the established isoprene synthases and warrants further study.

Acknowledgements

We thank Sunghyup Lee, Myunghan Noh, Anna-Liisa Ruskeepää and Tuulikki Seppänen-Laakso for isoprene analytics, and Dr. Marilyn Wiebe for critical reading of the manuscript.

References

Ajikumar, P.K., Xiao, W.-H., Tyo, K.E.J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science 330, 70–4. doi:10.1126/science.1191652

Beatty, M.K., Hayes, K., Hou, Z., Meyer, D.J., Nannapaneni, K., Rife, C.L., Wells, D.H., Zastrow-Hayes, G.M., 2013. The Legume Isoprene Synthase For Production of Isoprene. WO13166320.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. Nucleic Acids Res. 41, D36–42. doi:10.1093/nar/gks1195

Bentley, F.K., García-Cerdán, J.G., Chen, H.C., Melis, A., 2013. Paradigm of Monoterpene (β-phellandrene) Hydrocarbons Production via

Photosynthesis in Cyanobacteria. Bioenergy Res. 6, 917–929. doi:10.1007/s12155-013-9325-4

Bentley, F.K., Melis, A., 2012. Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms. Biotechnol. Bioeng. 109, 100–9. doi:10.1002/bit.23298

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res. 14, 988–95. doi:10.1101/gr.1865504

Chandran, S.S., Kealey, J.T., Reeves, C.D., 2011. Microbial production of isoprenoids. Process Biochem. 46, 1703–1710. doi:10.1016/j.procbio.2011.05.012

Chen, F., Tholl, D., Bohlmann, J., Pichersky, E., 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 66, 212–29. doi:10.1111/j.1365-313X.2011.04520.x

Choi, S., Song, C.W., Shin, J.H., Lee, S.Y., 2015. Biorefineries for the production of top building block chemicals and their derivatives. Metab. Eng. 28, 223–239. doi:10.1016/j.ymben.2014.12.007

Davies, F.K., Jinkerson, R.E., Posewitz, M.C., 2014. Toward a photosynthetic microbial platform for terpenoid engineering. Photosynth. Res. 265–284. doi:10.1007/s11120-014-9979-6

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–63.

Gray, D.W., Breneman, S.R., Topper, L. a, Sharkey, T.D., 2011. Biochemical characterization and homology modeling of methylbutenol synthase and implications for understanding hemiterpene synthase evolution in plants. J. Biol. Chem. 286, 20582–90. doi:10.1074/jbc.M111.237438

Harley, P., Vasconcellos, P., Vierling, L., Pinheiro, C.C.D.S., Greenberg, J., Guenther, A., Klinger, L., Almeida, S.S. De, Neill, D., Baker, T., Phillips, O., Malhi, Y., 2004. Variation in potential for isoprene emissions among Neotropical forest sites. Glob. Chang. Biol. 10, 630–650. doi:10.1111/j.1529-8817.2003.00760.x

Hess, B.M., Xue, J., Markillie, L.M., Taylor, R.C., Wiley, H.S., Ahring, B.K., Linggi, B., 2013. Coregulation of Terpenoid Pathway Genes and Prediction of Isoprene Production in *Bacillus subtilis* Using Transcriptomics. PLoS One 8. doi:10.1371/journal.pone.0066104

Kallas, T; Nelson, M, Singsaas, E., 2013. Methods for Isoprene and Pinene Production in Cyanobacteria. U.S. Patent Application 13/952,071.

Köksal, M., Zimmer, I., Schnitzler, J.-P., Christianson, D.W., 2010. Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. J. Mol. Biol. 402, 363–73. doi:10.1016/j.jmb.2010.07.009

Lemuth, K., Steuer, K., Albermann, C., 2011. Engineering of a plasmid-free *Escherichia coli* strain for improved in vivo biosynthesis of astaxanthin. Microb. Cell Fact. 10, 29. doi:10.1186/1475-2859-10-29

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–9. doi:10.1093/bioinformatics/btl158

Lindberg, P., Park, S., Melis, A., 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. Metab. Eng. 12, 70–9. doi:10.1016/j.ymben.2009.10.001

Lohr, M., Schwender, J., Polle, J.E.W., 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. Plant Sci. 185-186, 9–22. doi:10.1016/j.plantsci.2011.07.018

Lv, X., Xie, W., Lu, W., Guo, F., Gu, J., Yu, H., Ye, L., 2014. Enhanced isoprene biosynthesis in *Saccharomyces cerevisiae* by engineering of the native

acetyl-CoA and mevalonic acid pathways with a push-pull-restrain strategy. J. Biotechnol. 186, 128–36. doi:10.1016/j.jbiotec.2014.06.024

Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T., Bohlmann, J., 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol. 10, 226. doi:10.1186/1471-2229-10-226

Miller, B., Oschinski, C., Zimmer, W., 2001. First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. Planta 213, 483–487. doi:10.1007/s004250100557

Miziorko, H.M., 2011. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. Arch. Biochem. Biophys. 505, 131–143. doi:10.1016/j.abb.2010.09.028

Sasaki, K., Ohara, K., Yazaki, K., 2005. Gene expression and characterization of isoprene synthase from *Populus alba*. FEBS Lett. 579, 2514–8. doi:10.1016/j.febslet.2005.03.066

Sauer, U., Lasko, D.R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wüthrich, K., Bailey, J.E., 1999. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. J. Bacteriol. 181, 6679–88.

Schnitzler, J.P., Arenz, R., Steinbrecher, R., Lehning, A., 1996. Characterization of an isoprene synthase from leaves of *Quercus petraea* (Mattuschka) Liebl. Bot. Acta 109, 216–221.

Sharkey, T.D., Gray, D.W., Pell, H.K., Breneman, S.R., Topper, L., 2013. Isoprene synthase genes form a monophyletic clade of acyclic terpene synthases in the TPS-B terpene synthase family. Evolution 67, 1026–40. doi:10.1111/evo.12013

Sharkey, T.D., Yeh, S., Wiberley, A.E., Falbel, T.G., Gong, D., Fernandez, D.E., 2005. Evolution of the isoprene biosynthetic pathway in kudzu. Plant Physiol. 137, 700–12. doi:10.1104/pp.104.054445

Whited, G.M., Feher, F.J., Benko, D.A., Marguerite, A., Cervin, Gopal K. Chotani, 1 Joseph C. McAuliffe, 1 Richard J. LaDuca, 1, Eliahu A. Ben-Shoshan, 1 and Karl J. Sanford1, 2010. Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering Abstract. Ind. Biotechnol. 6, 152–163.

Xue, J., Ahring, B.K., 2011. Enhancing isoprene production by genetic modification of the 1-deoxy-d-xylulose-5-phosphate pathway in *Bacillus subtilis*. Appl. Environ. Microbiol. 77, 2399–405. doi:10.1128/AEM.02341-10

Yang, J., Zhao, G., Sun, Y., Zheng, Y., Jiang, X., Liu, W., Xian, M., 2012. Bio-isoprene production using exogenous MVA pathway and isoprene synthase in *Escherichia coli*. Bioresour. Technol. 104, 642–7. doi:10.1016/j.biortech.2011.10.042

Zhan, X., Zhang, Y.-H., Chen, D.-F., Simonsen, H.T., 2014. Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and α/β-santalene. Front. Plant Sci. 5, 636. doi:10.3389/fpls.2014.00636

Zhao, L., Chang, W., Xiao, Y., Liu, H., Liu, P., 2013. Methylerythritol phosphate pathway of isoprenoid biosynthesis. Annu. Rev. Biochem. 82, 497–530. doi:10.1146/annurev-biochem-052010-100934

Zhao, Y., Yang, J., Qin, B., Li, Y., Sun, Y., Su, S., Xian, M., 2011. Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. Appl. Microbiol. Biotechnol. 90, 1915–22. doi:10.1007/s00253-011-3199-1

Zurbriggen, A., Kirst, H., Melis, A., 2012. Isoprene Production Via the Mevalonic Acid Pathway in *Escherichia coli* (Bacteria). BioEnergy Res. 5, 814–828. doi:10.1007/s12155-012-9192-4

Tables

Table 1. List of genes expressed in *E. coli.* For sequences identified in this work, the GenBank or RefSeq accession number and NCBI gi are given.

Source organism (abbreviation)	**Original sequence (Uniprot, from a patent, or NCBI gi (GenBank))**	**Sequence present in the optimized expression construct**
P. montana (P.mon)	Q6EJ97	SkIspS (Lindberg et al., 2010) aa 1-569
P. alba (P.alb)	Q50L36	aa 38 - 595
A. hypogaea (A.hyp)	S3_Ahypogaea WO2013/166320	aa 1-546
I. batatas (I.bat)	345720078 (JP105673.1)	aa 49-588
E. photiniifolius (E.pho)	388282537 (FX134022.1)	aa 41-579
M. notabilis (M.not)	587936327 (EXC23171.1)	aa 47-598
D. pinnata (D.pin)	629169945 (GBDN01008783.1)	aa 1-534 preceded by MTARRSANYQ
M. indica (M.ind)	617864104 (GBCV01019670.1)	aa 43-596
F. vesca subsp. vesca (F.ves)	470140611 (XP_004306033.1)	aa 1-563
M. sativa (M.sat)	585498671 (GAFF01118088.1)	aa 49-586
S. indicum (S.ind)	357338539 (JP645796.1)	aa 1-556
E. grandis (E.gra)	629080625 (KCW47070.1)	aa 27-589

Table 2. Qualitative summary of amino acid sequence features and experimental results from expression of different IspS candidates in *E. coli*.

Source organism	IspS score	Presence of F338	Isoprene detected	Protein detected
Populus alba	4	+	+	+
Pueraria montana	4	+	+	+
Arachis hypogaea	4	+	-	+
Ipomoea batatas	4	+	+	+
Elaeocarpus photiniifolius	4	+	+	+
Mangifera indica	3	+	+	+
Fragaria vesca	3	-	-	+
Morus notabilis	3	-	-	+
Dahlia pinnata	3	-	-	+
Medicago sativa	3	-	-	-
Eucalyptus grandis	3	-	-	-
Sesamum indicum	2	-	-	+

Figure Captions

Fig. 1. Phylogenetic tree of the retrieved terpene synthases and known IspS sequences (total 1945 sequences). The TPS-b subtree (indicated in black) contains known isoprene synthases (indicated in red or dark pink) and some sequences unreliably annotated as IspS (in orange). The two known bifunctional enzymes are indicated in blue. Sequences from the red oval were analysed in more detail.

Fig. 2. Phylogenetic tree of potential candidate sequences together with the reference enzymes. Reference IspS are coloured red. Candidate IspS sequences with and isoprene score 4 (*I. batatas* and *E. photiniifolius*) and that from *M. indica,* which had a score of 3 and contained F338, are coloured green. Previously patented candidates are in orange, and previously published candidates in light blue. Sequences with confirmed alternative functions are shown in purple and the bi-functional *H. lupulus* sequence in pink. Grey sequences do not have the required conserved amino acids and are unlikely to be isoprene synthases. Sequences in black are missing some of the key amino acids, and have an isoprene score of 3, but may still have isoprene synthase activity.

Fig. 3. Multiple sequence alignment of candidate IspS. The known isoprene synthase sequences and 4 sequences that are not isoprene synthases (indicated as tricyclene and β-ocimene synthases) are included for reference.

The sequence order follows the order in Fig. 2, and the sequence names are coloured the same way as in Fig. 2. The alignment shows only the sequence around the substrate binding amino acids. The blue boxes below the alignment indicate the amino acids belonging to the “isoprene score” amino acids. Red blocks (below the *P. canescens* sequence) indicate gaps, i.e. places where sequence has been removed. The pink arrows (below the *P. canescens* sequence list the amino acid sequence positions (*P. alba* sequence numbering) of each segment. The single amino acid segment is F420, and the three amino acid segment is S445-S447. The amino acid sequence has been coloured using RasMol colouring scheme that shows similar amino acids with similar colours.

Fig. 4. Isoprene production by *E. coli* expressing different isoprene synthases. Previously known isoprene synthases of *P. montana*, *P. alba* and *A. hypogaea* are indicated in red, the novel candidates in blue, and the negative control strains w/o IspS in grey. Isoprene was measured after 24h incubation at 30 ℃ in the presence (dark coloured bars) or absence (light coloured bars) of IPTG. **A**. Isoprene (ng/ml). **B**. Isoprene (ng/mg CDW). The absence of F338 in the IspS sequence is indicated by the yellow colour. **C**. Coomassie-stained SDS-PAGE analysis of proteins from *E. coli* whole cell extracts. Total protein was isolated from non-induced (w/o IPTG) and IPTG induced cells after 2 h incubation. The bands appearing in the induced samples corresponding to IspS are indicated by red diamonds. **D**. Isoprene (ng/mg CDW). The average results obtained from three cultivations are shown. Error bars indicate standard error of the mean (n=3-6).

Fig. 5. Amino acid identity percentages of IspS sequences and reference sequences. The database identifiers are from Uniprot, GenBank or RefSeq. For the candidate IspS sequences (from this study), the isoprene score value is given at the end of the sequence name.

Fig. 6. Isoprene production at 37 °C by *E. coli* expressing different IspS. **A.** Isoprene (ng/ml) measured 2 h (dark coloured bars) and 5 h (light coloured bars) after IPTG induction in IPTG induced samples (+) and in non-induced samples (-). **B**. Isoprene production at 37 °C by *E. coli* expressing *I. batatas* (blue solid diamonds) and *P. alba* (red solid rectangles) IspS. Parallel cultivations were set up and sampled hourly between 1 and 6 hours and then at 23 h after IPTG induction. The average of two samples is shown and the error bars indicate standard error of the mean. OD_{600} of the *E. coli* cultures expressing the *I. batatas* (open diamonds) and *P. alba* (open rectangles) *IspS* at 2, 5 or 23 h is indicated.

Supplementary Figure Captions

Fig S1. Isoprene measurements. Examples of GC-MS chromatograms from headspace of *E. coli* BL21(DE3) host (left panel), *E. coli* expressing *I. batatas* IspS (middle panel), and isoprene standard (right panel).

Fig S2. Isoprene production at 37 ℃ and 30 ℃. *E. coli* expressing *I. batatas* (blue) and *P. alba* (red) IspS were analysed for isoprene production after 2 h (dark blue and dark red bars) and 5 h (light blue and light red bars) incubation at 37 ℃ and 30 ℃. Error bars indicate standard error of the mean (n=2).

Tables

Table 1. List of genes expressed in *E. coli.* For sequences identified in this work, the GenBank or RefSeq accession number and NCBI gi are given.

Source organism (abbreviation)	**Original sequence (Uniprot, from a patent, or NCBI gi (GenBank))**	**Sequence present in the optimized expression construct**
P. montana (P.mon)	Q6EJ97	SkIspS (Lindberg et al., 2010) aa 1-569
P. alba (P.alb)	Q50L36	aa 38 - 595
A. hypogaea (A.hyp)	S3_Ahypogaea WO2013/166320	aa 1-546
I. batatas (I.bat)	345720078 (JP105673.1)	aa 49-588
E. photiniifolius (E.pho)	388282537 (FX134022.1)	aa 41-579
M. notabilis (M.not)	587936327 (EXC23171.1)	aa 47-598
D. pinnata (D.pin)	629169945 (GBDN01008783.1)	aa 1-534 preceded by MTARRSANYQ
M. indica (M.ind)	617864104 (GBCV01019670.1)	aa 43-596
F. vesca subsp. vesca (F.ves)	470140611 (XP_004306033.1)	aa 1-563
M. sativa (M.sat)	585498671 (GAFF01118088.1)	aa 49-586
S. indicum (S.ind)	357338539 (JP645796.1)	aa 1-556
E. grandis (E.gra)	629080625 (KCW47070.1)	aa 27-589

Table 2. Qualitative summary of amino acid sequence features and experimental results from expression of different IspS candidates in *E. coli*.

Source organism	IspS score	Presence of F338	Isoprene detected	Protein detected
Populus alba	4	+	+	+
Pueraria montana	4	+	+	+
Arachis hypogaea	4	+	-	+
Ipomoea batatas	4	+	+	+
Elaeocarpus photiniifolius	4	+	+	+
Mangifera indica	3	+	+	+
Fragaria vesca	3	-	-	+
Morus notabilis	3	-	-	+
Dahlia pinnata	3	-	-	+
Medicago sativa	3	-	-	-
Eucalyptus grandis	3	-	-	-
Sesamum indicum	2	-	-	+

Highlights

- Three uncharacterized terpene synthases were shown to code for active isoprene synthases.
- Production of isoprene by the novel and previously known isoprene synthases was assessed in *E. coli*.
- The newly identified enzyme from *Ipomoea batatas* produced isoprene effectively.
- *I. batatas* IspS is an attractive option for development of microbial isoprene factories.

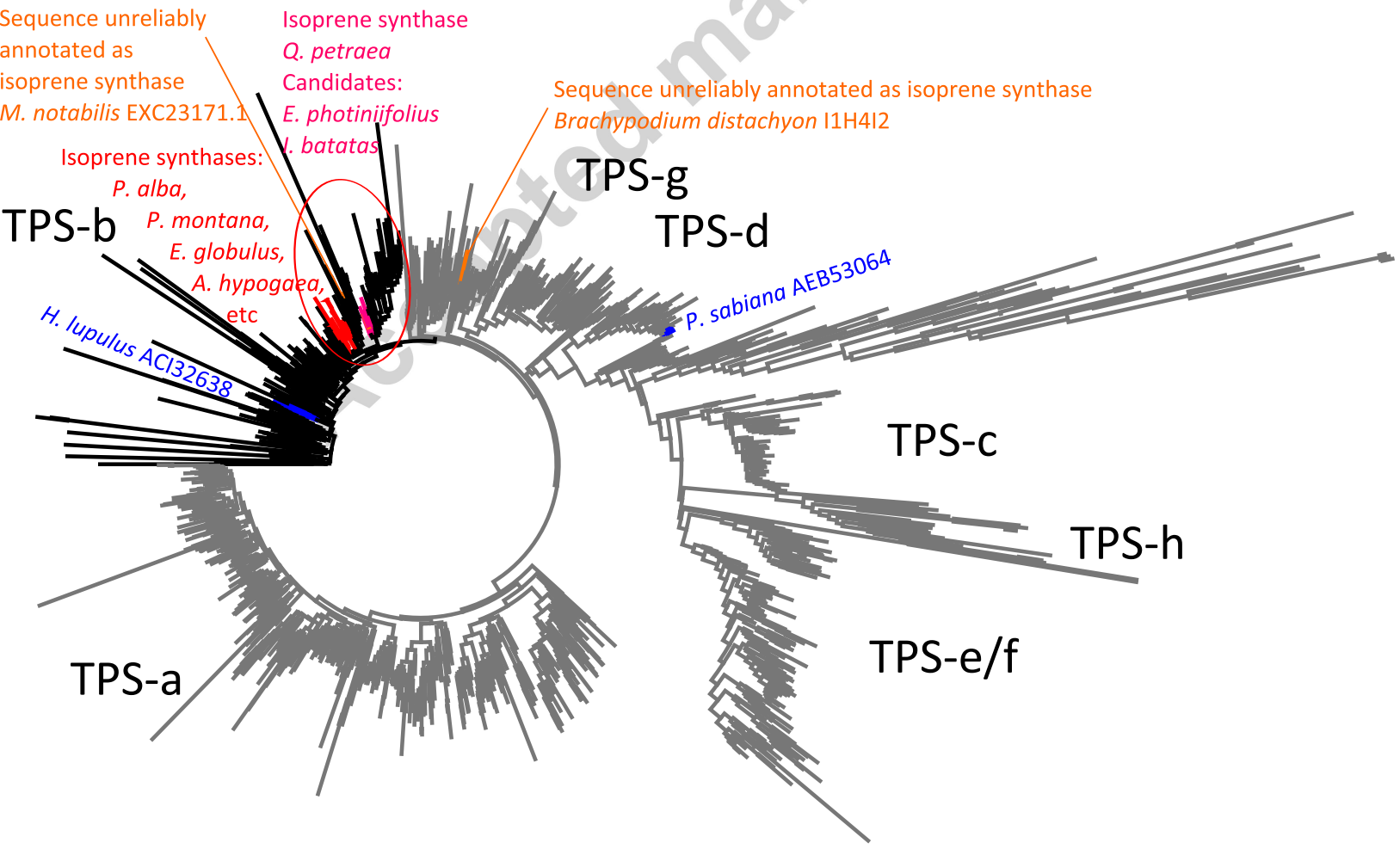

Figure 2

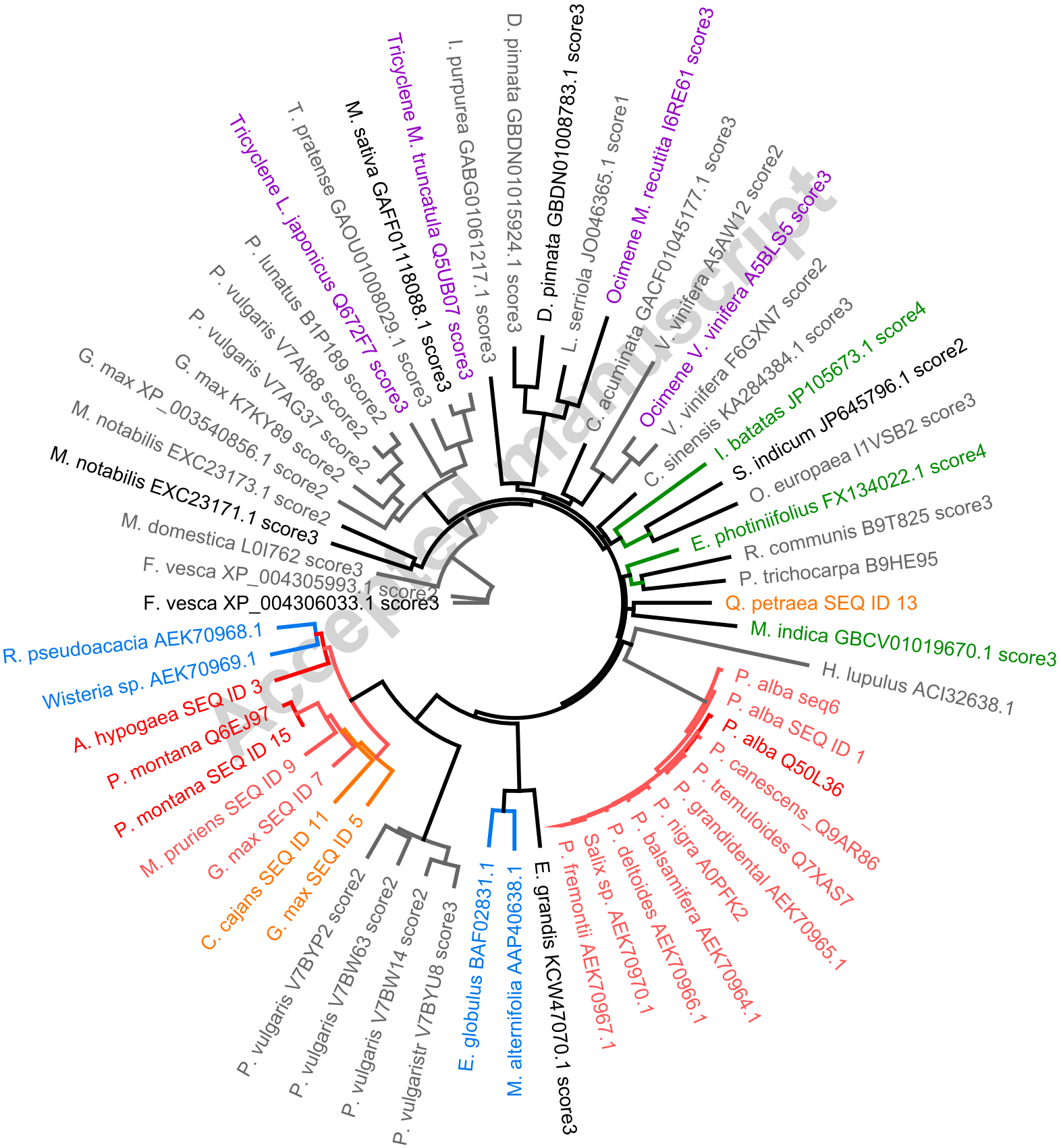

Figure 3

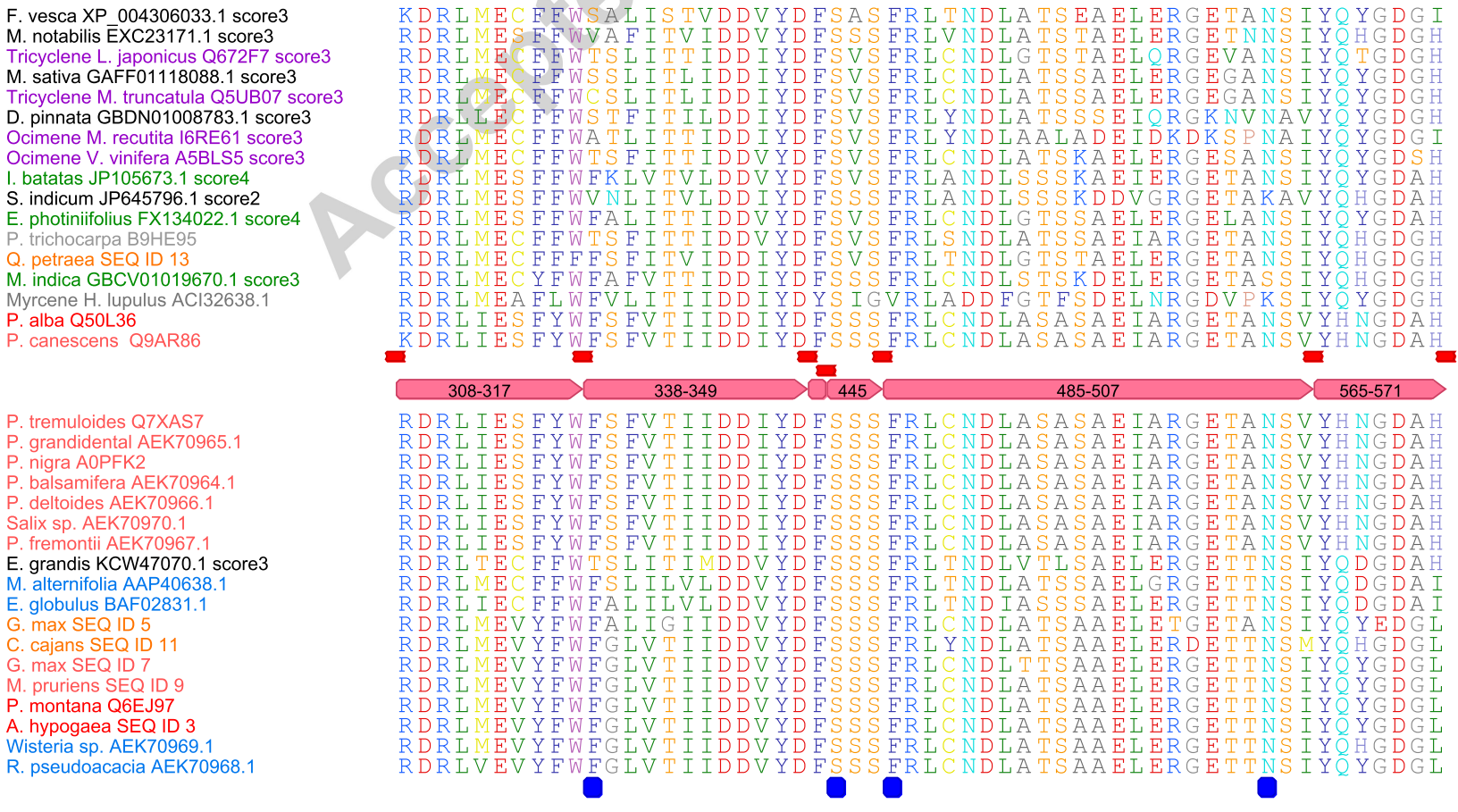

Figure 4

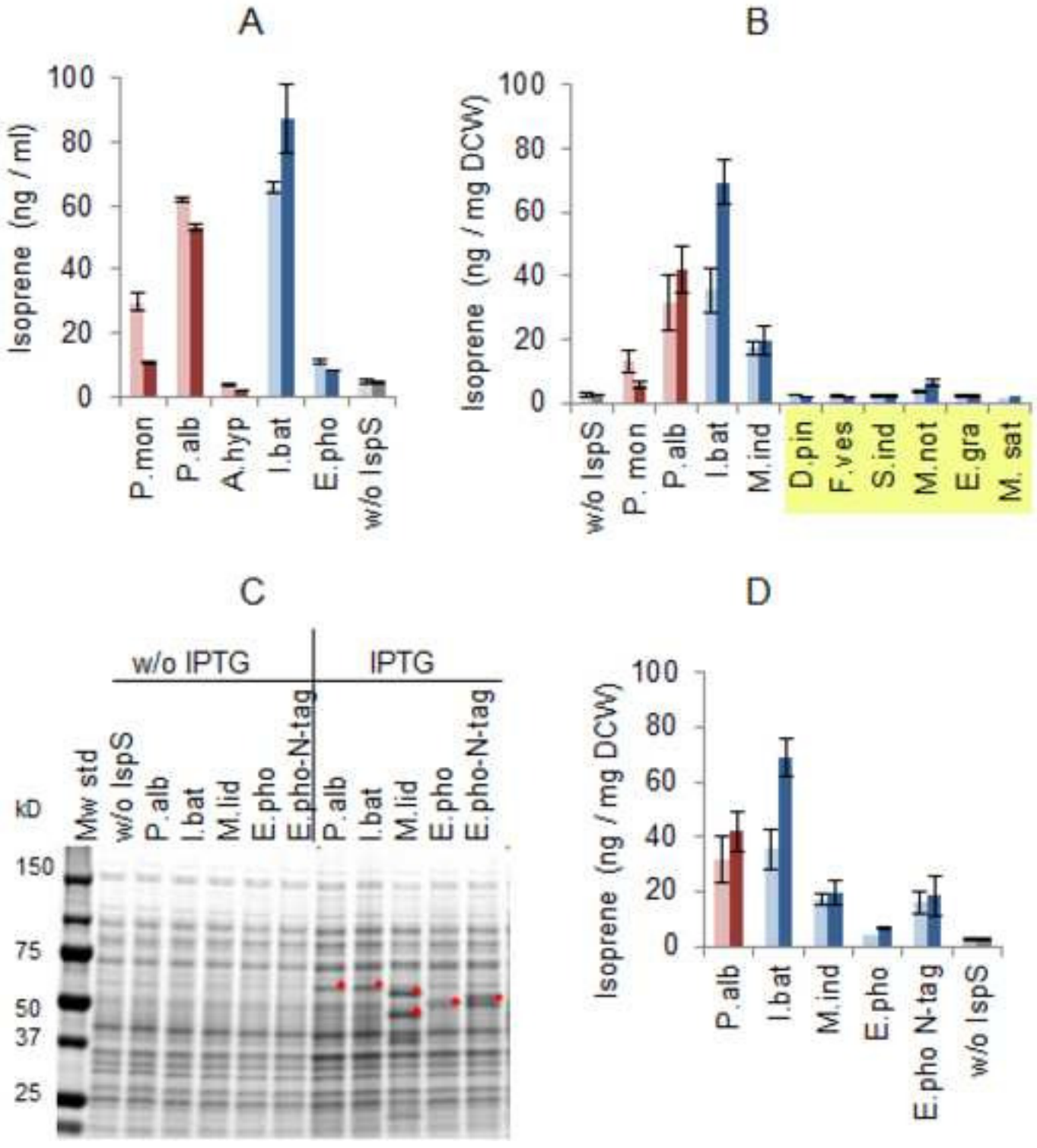

Figure 5

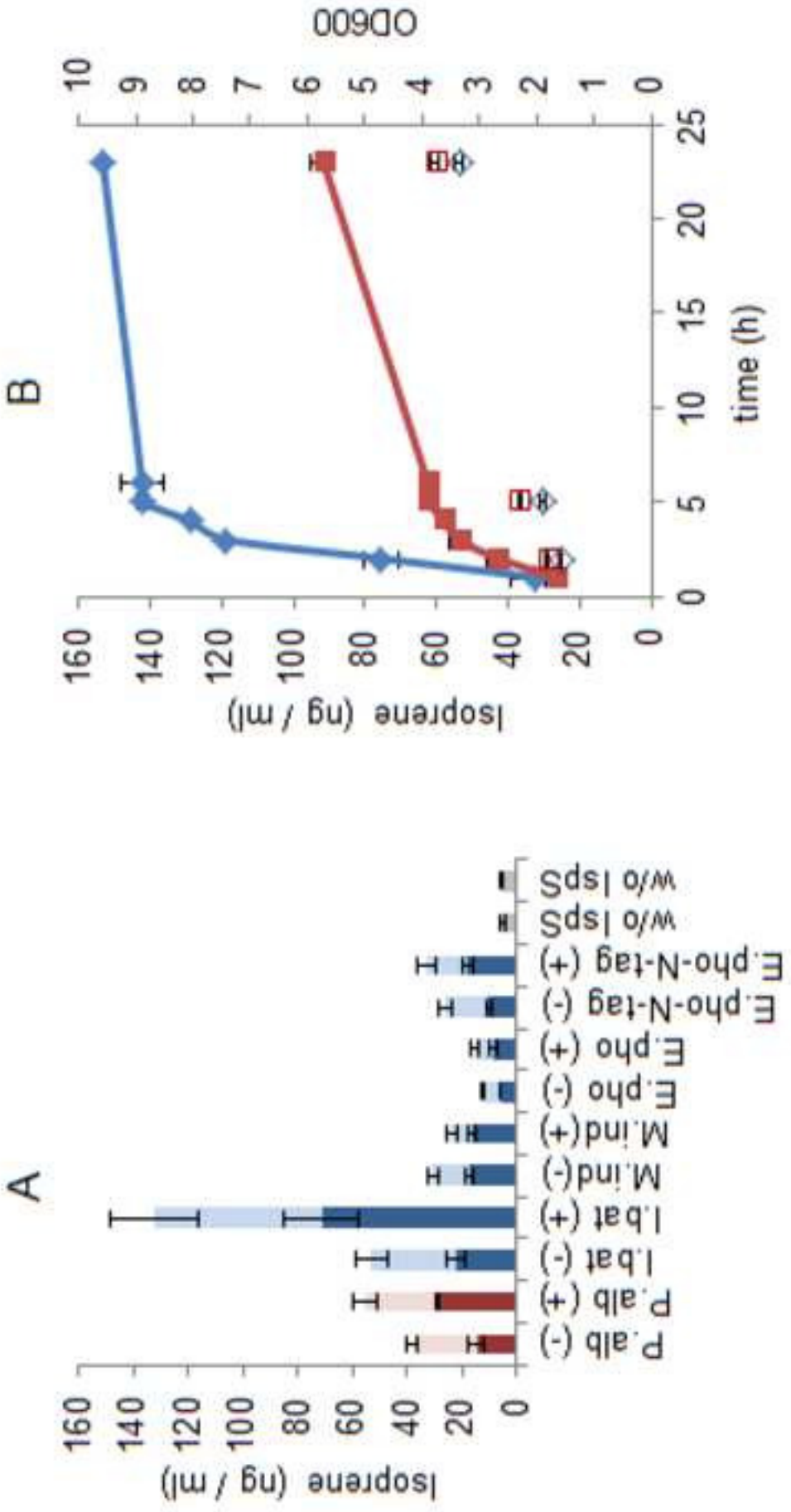
A
Isoprene (ng / ml)
0
20
40
60
80
100
120
140
160
P.alb (-)
P.alb (+)
I.bat (-)
I.bat (+)
M.ind(-)
M.ind(+)
E.pho (-)
E.pho (+)
E.pho-N-tag (-)
E.pho-N-tag (+)
w/o IspS
w/o IspS
B
Isoprene (ng / ml)
0
20
40
60
80
100
120
140
160
time (h)
0
5
10
15
20
25
OD600
0
1
2
3
4
5
6
7
8
9
10

Figure 6