Running Title:
Histidinol-phosphate phosphatase activity of IMPL2

Corresponding author:
Robert A. Ingle
Department of Molecular and Cell Biology
University of Cape Town
Private Bag
Rondebosch 7701
South Africa

Tel: +27 21 650 2408
Email: robert.ingle@uct.ac.za

Journal research area:
Biochemical processes and macromolecular structures
The missing link in plant histidine biosynthesis: Arabidopsis myoinositol monophosphatase-like 2 (IMPL2) encodes a functional histidinol-phosphate phosphatase

Lindsay N. Petersen, Sandra Marineo, Salvatore Mandalà, Faezah Davids, Bryan T. Sewell and Robert A. Ingle*

Department of Molecular and Cell Biology, University of Cape Town, Private Bag, Rondebosch 7701, South Africa (L.N.P., F.D., R.A.I.); Electron Microscope Unit, University of Cape Town, Private Bag, Rondebosch 7701, South Africa (B.S.T.); and Dipartimento di Biologia Cellulare e dello Sviluppo Viale delle Scienze, Università degli Studi di Palermo, Parco d’Orleans II, 90128 Palermo, Italy (S.Mar., S.Man.)
Footnotes

1 This work was supported by a grant from the Emerging Researchers Program at the University of Cape Town to R.A.I, and by the award of an Innovation Fellowship from the National Research Foundation, South Africa to L.N.P.

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: Robert A. Ingle (robert.ingle@uct.ac.za)

* Corresponding author, email: robert.ingle@uct.ac.za
ABSTRACT

Histidine (His) plays a critical role in plant growth and development, both as one of the standard amino acids in proteins, and as a metal binding ligand. While genes encoding seven of the eight enzymes in the pathway of His biosynthesis have been characterised from a number of plant species, the identity of the enzyme catalysing the dephosphorylation of histidinol-phosphate to histidinol has remained elusive. Recently, members of a novel family of histidinol-phosphate phosphatase (HPP) proteins, displaying significant sequence similarity to known myoinositol monophosphatases (IMP) have been identified from several Actinobacteria. Here we demonstrate that a member of the IMP family from Arabidopsis thaliana, myoinositol monophosphatase-like 2 (IMPL2, At4g39120) has HPP activity. Heterologous expression of IMPL2, but not the related IMPL1 protein, was sufficient to rescue the His auxotrophy of a Streptomyces coelicolor hisN mutant. Homozygous null impl2 Arabidopsis mutants displayed embryonic lethality, which could be rescued by supplying plants heterozygous for null impl2 alleles with His. In common with the previously characterised HISN genes from Arabidopsis, IMPL2 was expressed in all plant tissues and throughout development, and an IMPL2:GFP fusion protein was targeted to the plastid, where His biosynthesis occurs in plants. Our data demonstrate that IMPL2 is the HISN7 gene product, and suggest a lack of genetic redundancy at this metabolic step in Arabidopsis, which is characteristic of the His biosynthetic pathway.
INTRODUCTION

Histidine (His) is one of the standard 20 amino acids found in proteins, and is required for plant growth and development (Muralla et al., 2007; Bikard et al., 2009). The occurrence of His in the active sites of numerous enzymes is attributable to the imidazole functional group (pK_a ~6), which can alternate between the protonated and unprotonated states under physiologically relevant conditions, allowing its participation in acid-base catalysis (Fersht, 1999). His also plays important biochemical roles as a nucleophile in phosphoryl group transfer, and as a metal-binding ligand (Fraústo da Silva and Williams, 2001; Harding, 2004).

Genetic analysis of His biosynthesis has provided insights into key regulatory mechanisms in micro-organisms, notably the discovery of operon structure and metabolic regulation through attenuation (Ames et al., 1960; Roth and Ames, 1966). However, this was the last amino acid biosynthetic pathway to be characterised in plants, and only recently has it been demonstrated that His biosynthesis follows the same route as that in micro-organisms with ten reactions catalysed by eight enzymes (Fig. 1), beginning with the condensation of ATP and 5-phosphoribosyl-1-pyrophosphate catalysed by the enzyme ATP-phosphoribosyl transferase (ATP-PRT) (Ohta et al., 2000). His biosynthesis is linked to purine metabolism through its precursors PRPP and ATP, and the release of an intermediate (5'-phosphoribosyl-4-carboxamide-5-aminoimidazole) at the branch point step catalysed by imidazole-glycerol phosphate synthase (IGPS), which enters the de novo purine synthesis pathway (Alifano et al., 1996; Ward and Ohta, 1999). While the regulation of His biosynthesis has not been as comprehensively investigated in plants as it has in micro-organisms, analysis of transgenic Arabidopsis plants over-expressing single HisN genes under the control of the CaMV 35S promoter suggests that ATP-PRT activity controls the size of the pool of free His in plants (Ingle et al., 2005; Rees et al., 2009).

Genes encoding seven of the eight enzymes in the His pathway were identified in plants during the late 1990’s (Nagai et al., 1993; Mori et al., 1995; El Malki et al., 1998;
Fujimori and Ohta, 1998a, b; Fujimori et al., 1998; Ohta et al., 2000) and, in contrast to most other amino acid biosynthetic pathways, the majority were found to be encoded by single genes in the Arabidopsis genome (Stepansky and Leustek, 2006). The missing link in the pathway has been HISN7, histidinol-phosphate phosphatase (HPP, EC 3.1.3.15). While dephosphorylation of histidinol-P to histidinol was first detected in plant extracts almost 40 years ago (Wiater et al., 1971), the identity of the enzyme(s) catalysing this reaction in plants is unknown. HPP proteins identified from micro-organisms prior to 2006 can be grouped into one of two superfamilies – the DDDD (which contain four invariant Asp residues) and PHP (polymerase and histidinol phosphatase) superfamilies (Lee et al., 2008). The DDDD superfamily contains bifunctional enzymes from enteric bacteria (such as *E. coli.*) where dephosphorylation of histidinol-P is catalysed by the N-terminal domain, while the C terminus displays IGPD activity (Alifano et al., 1996; Fani et al., 2007), and monofunctional HPP proteins such as that recently identified in *Thermococcus onnurineus* (Lee et al., 2008). In contrast, all members of the PHP family identified to date are monofunctional HPP enzymes, including those encoded by *HIS2* in *Saccharomyces cerevisiae*, and *ytvP* in *Bacillus subtilis* (Alifano et al., 1996; le Coq et al., 1999). Notably, neither the Arabidopsis nor rice genomes contain any sequences with significant sequence identity to members of either the DDDD or PHP superfamilies (Stepansky and Leustek, 2006).

A novel HPP protein, showing no sequence similarity to members of either the DDDD or PHP superfamilies, was recently isolated from *Corynebacterium glutamicum* (Mormann et al., 2006), and orthologues have subsequently been identified in other Actinobacteria, including *Streptomyces coelicolor* (Marineo et al., 2008). All members of this new family display significant sequence similarity to known myoinositol monophosphatases (IMPs) (Mormann et al., 2006), which catalyse the hydrolysis of the ester bond of D-myoinositol 1(or 3)-P (D-Ins 1-P, D-Ins 3-P) to generate myoinositol, without the production of a phospho-enzyme intermediate (Leech et al., 1993). In eukaryotes, myoinositol plays a crucial role in a number of cellular processes including phosphatidylinositol-mediated signalling and the membrane anchoring of proteins (Boss et al., 2006; Fujita and Jigami, 2008). To date, 3 putative IMP encoding sequences have been identified in the
Arabidopsis genome; VTC4 (At3g02870), IMPL1 (At1g31190) and IMPL2 (At4g39120) (Torabinejad et al., 2009). VTC4 was first reported as the L-galactose 1-P (L-Gal 1-P) phosphatase in ascorbate biosynthesis in Arabidopsis (Laing et al., 2004; Conklin et al., 2006), but was recently demonstrated to be a bifunctional enzyme also able to catalyse the dephosphorylation of D-Ins 1-P and D-Ins 3-P to myoinositol \textit{in vitro} (Torabinejad et al., 2009). vtc4 null mutants showed only a 30\% reduction in myoinositol content, suggesting genetic redundancy in the capacity to generate myoinositol, and accordingly IMPL1 and IMPL2 were shown to have \textit{in vitro} IMP activity (Torabinejad et al., 2009). Both enzymes were also able to utilise L-Gal 1-P as a substrate, suggesting some promiscuity in their substrate specificity. Here we demonstrate that the IMPL2 protein encoded by At4g39120 is also able to catalyse the dephosphorylation of histidinol-P to histidinol, completing the His biosynthetic pathway in plants.

RESULTs

Bioinformatic analysis identifies IMPL2 as a candidate HPP in Arabidopsis

To identify putative homologues in Arabidopsis of the IMP-like proteins that catalyse the dephosphorylation of histidinol-P in the Actinobacteria, we carried out BLAST analysis of the Arabidopsis protein RefSeq database with the \textit{S. coelicolor} HPP protein (NP_629355). This identified three proteins with \~30\% sequence identity to NP_629355; VTC4 (At3g02870), IMPL1 (At1g31190) and IMPL2 (At4g39120) (Supp. Fig. 1), all members of the IMP superfamily. VTC4, IMPL1 and IMPL2 have recently been shown to catalyse the dephosphorylation of D-Ins 1-P, D-Ins 3-P, and L-Gal 1-P (Torabinejad et al., 2009). However, their ability to utilise histidinol-P as a substrate has not been reported.

All known plant \textit{HISN} genes contain N-terminal plastid transit peptide encoding sequences, and it is thought that His biosynthesis is localised to the plastid (Stepansky and Leustek, 2006). It is thus highly probable that any HPP enzyme(s) in Arabidopsis would also be targeted to this organelle. Both IMPL1 and IMPL2 satisfy this criterion. These two proteins contain putative plastid transit peptide sequences as predicted by
ChloroP analysis (Emanuelsson et al., 1999), and have been detected in MS analyses of purified chloroplast preparations from Arabidopsis (Zybailov et al., 2008). In contrast, VTC4 does not contain a predicted plastid transit peptide, and is instead a membrane protein; HMMTOP analysis (Tusnady and Simon, 2001) predicts the presence of three transmembrane α-helices and VTC4 has been detected in purified plasma membrane preparations by MS analysis (Marmagne et al., 2007).

In contrast to yeast, there is little evidence for the operation of a general control of N metabolism regulatory mechanism in plants. However, the genes encoding the first 6 enzymes in Arabidopsis have GCN4 like binding sites in their upstream regions (Stepansky and Leustek, 2006), which may be indicative of transcriptional co-regulation. In addition, inhibition of IGPD activity has been shown to result in increased expression of HDH in Arabidopsis (Guyer et al., 1995). We used the Arabidopsis Coexpression Data Mining tool (Manfield et al., 2006) to identify genes coexpressed with IMPL1 and IMPL2 across 322 publicly available Affymetrix ATH1 microarray datasets. This analysis identified 2 His biosynthetic genes (in the top 5) as showing the most positive correlation in expression with IMPL2; ProFAR-I (At2g36230, HISN3) ranked 1st and HDH (At5g63890, HISN8) ranked 4th (Supp. Table I). In addition, HPA (At5g10330, HISN6A and At1g71920, HISN6B) was ranked 61st with an r-value of 0.56. In contrast, none of the known HISN genes were returned in the top 200 genes most positively correlated with IMPL1 expression in Arabidopsis.

Finally, analysis of the IMPL1 and IMPL2 amino acid sequences revealed that a Gly residue required for HPP activity in the S. coelicolor enzyme (Marineo et al., 2008) was conserved in IMPL2 but not IMPL1 where it is replaced by Glu (Supp. Fig. 1). On the basis of our bioinformatic analyses we investigated whether IMPL2 encodes a functional HPP enzyme in Arabidopsis.

Heterologous expression of IMPL2, but not IMPL1, rescues the His auxotrophy of a Streptomyces coelicolor HPP mutant
As histidinol-P is no longer commercially available to test for HPP activity of recombinant IMPL2 protein, we instead determined whether heterologous expression of IMPL2 was sufficient to complement the His auxotrophy of the S. coelicolor HPP mutant hisN (Puglia et al. 1982). hisN protoplasts transformed with the empty pIJ8600 expression vector were able to grow on minimal media (plus Phe) only if supplemented with 50 µg ml⁻¹ His, while transformation with pIJ8600:IMPL2 allowed the mutant to grow on minimal media without added His (Fig. 2), confirming that IMPL2 possesses HPP activity. In contrast, heterologous expression of IMPL1 was not sufficient to complement the His auxotrophy of the hisN strain (Fig. 2), suggesting that IMPL1 does not possess HPP activity.

**HPP activity of IMPL2 requires a conserved Gly residue at position 195**

A conserved Gly residue has previously been shown to be required for HPP activity in S. coelicolor (Marineo et al. 2008). This residue is present in all putative HPP homologues subsequently identified in the Actinobacteria, and is also present in IMPL2 but not IMPL1 (Supp. Fig 1). To determine whether Gly195 was required for the HPP activity of IMPL2, we used PCR site-directed mutagenesis to generate a mutant protein where this residue was substituted with Arg. hisN cells expressing the mutant IMPL2 protein were unable to grow on minimal media (plus Phe) without supplementation with 50 µg ml⁻¹ His, demonstrating that Gly195 is also required for HPP activity of IMPL2 (Fig. 2).

**Homology modelling suggests that substitution of Gly195 with Arg leads to disruption of the IMPL2 active site**

A search of the Protein Data Bank for homologues of IMPL2 produces 22 protein structures which, according to the criteria of pGenTHREADER (McGuffin and Jones, 2003), are certain to have similar folds. Structural alignment of several of these proteins using the Matchmaker tool of UCSF Chimera (Pettersen et al., 2004) demonstrated that there was considerable conservation of the location of most of the structural elements of the protein family. Variability in structure occurred mainly in the loops, especially in those residues corresponding to 100 to 118 in IMPL2. The two structures chosen as models were SuhB: the inositol monophosphatase and extragenic suppressor from E. coli.
(2qfl) and the inositol-1(or 4)-monophosphatase (A-chain) from *Bos taurus* (2bji). IMPL2 has sequence identities of 24.8% and 20.4% to 2qfl and 2bji respectively. The pGenTHREADER generated alignment of IMPL2 with these two proteins was slightly modified to ensure that gaps were inserted in the most appropriate loops (Supp. Fig. S2), and used to produce a homology model using MODELLER (Eswar et al., 2006) (Fig. 3). The molecule has a αβαβα layered structure and a pronounced active site cleft in which are situated three Mg2+ ions held in position by carboxylate groups of Glu and Asp residues located on loops at the side of the six-stranded sheet, and on one of the central helices (Fig. 3).

Gly195 is located in the third β strand (191 to 198) of the six-stranded sheet and is in close contact with one of the central α-helices (287 to 298) (Fig. 3). The location of this helix determines the position of Asp288, which forms a salt bridge to one of the active site Mg2+ ions. Gly195 is strongly conserved in the IMP superfamily, and the only other residue tolerated at this position is Ala, which is found in 2qfl. Replacement of Gly195 with an Arg using SCWRL4 (Shapovalov and Dunbrack, 2007) resulted in severe clashes with Pro265. Since SCWRL4 searches for an optimum packing of side chains without moving the protein backbone we conclude that it is impossible to accommodate an Arg residue at this position without separating the β-strand (191 to 198) and the α-helix (287 to 298) by at least 3Å. Such a separation would alter the relative locations of the key active site residues: Asp289, Glu147, Asp165 and Asp168. The latter three amino acids are located on other strands in the β-sheet. This, in turn, would disrupt the active site and account for the observed inactivity of the mutant IMPL2 protein.

**Analysis of impl2 mutants suggests a lack of genetic redundancy in the capacity for histidinol biosynthesis in Arabidopsis**

Homozygosity for null alleles of any of the four non-redundant genes in the His biosynthetic pathway (see Fig. 1) results in embryonic lethality in Arabidopsis (Muralla et al., 2007). If IMPL2 was the only protein with HPP activity in Arabidopsis, this phenotype should also be observed in *impl2* T-DNA insertion lines. Two T-DNA insertion lines were obtained from NASC, SAIL_146_E09 (*impl2-1*) and SALK_076930 (impl2-2).
Both *impl2-1* and *impl2-2* were supplied as segregating lines and PCR genotyping of 115 plants failed to identify a single homozygous individual in either line, consistent with homozygous lethality for null mutations in *IMPL2*. Instead $\chi^2$ analysis revealed no significant difference from a ratio of two heterozygous to one wild-type plant in either line (Table I).

To determine the developmental stage at which lethality occurs, we examined the siliques of heterozygous plants following self-pollination. Approximately a quarter of the seeds in these siliques were observed to have aborted prior to the late globular stage of development (when embryos begin to turn green), in both the *impl2-1* and *impl2-2* insertion lines. (Fig. 5, Table II). Thus lethality of the null *impl2* mutants occurred during embryo development, as previously reported for other single copy *his* mutants (Muralla et al., 2007). In contrast to several of the non-redundant *HISN* genes analysed by Muralla et al. (2007), $\chi^2$ analysis revealed no significant difference from a ratio of 3 wild-type to 1 mutant seed, and thus no evidence of reduced transmission of *impl2* mutant pollen tubes.

The supply of exogenous His to heterozygous plants can partially rescue the embryonic lethality of null *hisN* mutant seeds (Muralla et al., 2007). To determine whether His supplementation could rescue the embryonic lethality phenotype in the *impl2* mutants, we first identified heterozygous plants by PCR, and determined the frequency of seed abortion in siliques from the primary bolt. The primary bolt was then removed and the plants supplied with 1 mM His daily (added to the roots). Application of His resulted in a reduced frequency of seed abortion in the secondary bolts (compared to that observed in the primary bolts prior to the supply of His); from 24.1% to 12.1% in *impl2-1* and 27.5% to 17.8% in *impl2-2*. No such decrease was observed in plants supplied only with water (Fig. 6A). In an alternative experimental approach, 1mM His was supplied to heterozygous plants identified by PCR prior to bolting, and siliques from the primary bolt analysed 20 d after feeding began. The percentage of aborted seeds in siliques from plants supplied with His was significantly lower than that of control plants (Fig. 6B), and the majority of these plants contained several siliques with no aborted seeds.
Together, these data suggest that the embryonic lethality phenotype of impl2 null mutants is a consequence of their inability to synthesise His, and that there is no apparent genetic redundancy in the capacity for histidinol biosynthesis in Arabidopsis, at least in the developing embryo.

**IMPL2 is expressed in all plant tissues and at all developmental stages**

RNA gel blot and microarray analyses of known HISN genes from Arabidopsis has demonstrated that they are constitutively expressed in all plant tissues, and at all developmental stages (Ohta et al., 2000, Muralla et al., 2007). We analysed the temporal and spatial expression patterns of IMPL2 across multiple microarray expression profiling experiments using Genevestigator V3 (Hruz et al., 2007). IMPL2 displayed a similar pattern of expression to other known HISN genes (Supp. Fig. 3A and B); IMPL2 transcripts were detected in all tissues and throughout development, and mRNA levels were lowest in pollen, as previously reported for all other HISN genes by Muralla et al. (2007).

**IMPL2 localises to the chloroplast in planta**

All of the His biosynthetic enzymes identified to date from plants contain a putative N-terminal plastid transit peptide (Stepansky and Leustek, 2006), and two proteins (IGPD and HDH) have been immunolocalised to this organelle (Nagai et al., 1993; Tada et al., 1995). To demonstrate conclusively that the Arabidopsis HPP (IMPL2) is also targeted to the plastid in planta, we generated two C-terminal GFP fusion proteins; the first a full-length IMPL2 protein fused to GFP (IMPL2:GFP) and the second a truncated IMPL2 lacking the first 60 amino acids predicted to act as the plastid transit peptide by ChloroP (Emanuelsson et al., 1999) fused to GFP (IMPL2Δ1-60:GFP). Confocal imaging of maize leaf protoplasts transiently expressing these proteins revealed that IMPL2:GFP was targeted to the chloroplast, as evidenced by the co-localisation of GFP signal and chlorophyll autofluorescence (Fig. 7). In contrast, the truncated IMPL2Δ1-60:GFP protein remained in the cytoplasm (Fig. 7). Thus the 60 amino acid N-terminal sequence serves as a functional plastid transit peptide, and is required for transport into the chloroplast.
Phylogenetic analysis of plant IMP sequences

Querying the NCBI mRNA RefSeq database with the Arabidopsis VTC4, IMPL1 and IMPL2 cDNA sequences identified putative homologues of each gene in a range of monocot and dicot plants for which full genome sequence is available, as well as in the Chlorophyte green alga *Chlamydomonas reinhardtii*. Phylogenetic analysis of the aligned IMP cDNA sequences, using the neighbour-joining method (rooted with the Arabidopsis SAL1 gene that encodes a 3',5'-bisphosphate nucleotidase), revealed the existence of 3 well supported clades (Fig. 8), with the orthologues from the various species showing greater sequence similarity than paralogues within the same species. The phylogenetic tree is consistent with the occurrence of two gene duplication events, with subsequent sequence divergence; the first duplication event giving rise to the IMPL2 clade, and the second to the IMPL1 and VTC4 clades. These gene duplication events apparently occurred prior to the divergence of the plant and green algal lineages from their last common ancestor, which has been dated at 1000 MYA (Heckman et al., 2001).

DISCUSSION

The biochemical pathway of His biosynthesis (Fig. 1) was elucidated in plants during the late 1990’s largely as a result of efforts to identify novel targets for herbicides (Mori et al., 1995), and genes encoding 7 of the 8 enzymes identified through functional complementation of bacterial or yeast His biosynthetic mutants. The identity of the enzyme(s) catalysing the dephosphorylation of histidinol-P to histidinol was unknown, and analysis of the Arabidopsis genome sequence revealed no likely homologues of either DDDD or PHP type HPP proteins. Ward and Ohta (1999) questioned whether plants possessed a specific HPP enzyme, and suggested instead that additional phosphatase activities might function adequately to catalyse the dephosphorylation of histidinol-P. Our data suggest that this is the case in Arabidopsis; functional complementation of the *S. coelicolor* hisN mutant by IMPL2 (Fig. 2) demonstrated that in addition to galactose-P and inositol-P substrates, this enzyme can also hydrolyse histidinol-P to histidinol. While the majority of His enzymes from plants have not been biochemically characterised, they are thought to function solely in His biosynthesis, as is
the case in micro-organisms (Alifano et al., 1996). IMPL2 is thus unusual, as its previously reported in vitro activity against D-Ins 1-P and L-Gal 1-P (Torbanijedad et al., 2009) is suggestive of additional roles for IMPL2 in myoinositol and ascorbate metabolism, though this has yet to be demonstrated in vivo.

While the substrate specificity of the IMPL2 enzyme may be relatively broad, we found no evidence for genetic redundancy in histidinol biosynthesis in Arabidopsis. Null impl2 mutants displayed embryonic lethality (Fig. 5), that could be partially rescued by supplementation of plants heterozygous for impl2 mutant alleles with 1 mM His (Fig. 6A and 6B). This indicates that IMPL2 was required for His production in the developing embryo, and that endogenous His levels in the surrounding maternal tissue were insufficient to rescue the His auxotrophy of the null impl2 embryos, as previously observed for mutations in four other non-redundant HisN genes (Muralla et al., 2007). VTC4 and IMPL1, although expressed in embryonic tissue (Supp. Fig. 3C), were not sufficient to provide histidinol for the developing embryo.

While it is possible that phosphatase enzymes other than IMPL2 may also be capable of producing histidinol in other plant tissues, or at different developmental stages, it is unlikely that either VTC4 or IMPL1 play such a role in Arabidopsis. Lethality of impl1 null mutants has recently been reported (Torabinejad et al., 2009), however this phenotype does not appear to be linked to His biosynthesis, as heterologous expression of IMPL1 was not able to complement the His auxotrophy of the S. coelicolor hisN mutant (Fig. 3), indicating that IMPL1 does not possess HPP activity. While we did not test the ability of VTC4 to complement the hisN mutant in the current study, this protein is localised to the plasma membrane (Marmagne et al., 2007), and is thus unlikely to play a role in His biosynthesis. All known plant His proteins contain putative plastid transit peptides, and two (IGPD and HDH) have been immunolocalised to the chloroplast (Nagai et al., 1993; Tada et al., 1995). The observation that full-length IMPL2:GFP fusion protein was targeted to the chloroplast (Fig. 7), provides further evidence that His biosynthesis in plants is confined to this organelle.
Phylogenetic analysis of plant IMP cDNA sequences (Fig. 8) suggested that the gene duplication events giving rise to the 3 IMP clades in plants occurred prior to the split of lineages giving rise to Chlorophyta algae and plants (Fig. 8), which has been dated through analysis of 41 protein coding sequences at around 1000 MYA (Heckman et al., 2001). In light of this ancient evolutionary divergence, there is a remarkable degree of conservation in substrate specificity between the 3 Arabidopsis enzymes. VTC4, IMPL1 and IMPL2 are all capable of catalysing the dephosphorylation of D-Ins 1-P, D-Ins 3-P, and L-Gal 1-P in vitro (Torabinejad et al., 2009), although IMPL1 displays 10-fold reduced activity with D-Ins 3-P, and L-Gal 1-P compared to D-Ins 1-P. vtc4 null mutants are viable, suggesting genetic redundancy in all the enzymatic reactions catalysed by VTC4 (Torabinejad et al., 2009). In contrast, impl1 null mutants are not viable, suggesting that IMPL1 may perform additional uncharacterised reaction(s), while the lethality of the impl2 null mutants can be explained by the role of IMPL2 in His biosynthesis. The substrate specificities of the majority of putative plant IMP proteins have not been determined, though an IMP from kiwifruit (which falls in the VTC4 clade, Fig. 8) has also been shown to dephosphorylate both L-Gal 1-P and D-Ins 1-P (Laing et al., 2004), and this bifunctionality is also found in IMP proteins from mammals (Parthasarathy et al., 1997). Homology modelling of the IMPL2 protein against solved structures for IMP enzymes from *E. coli* and *B. taurus* also revealed a very high degree of structural conservation, despite the relatively low overall sequence identity (Fig. 3). The ability to utilise both galactose-P and inositol-P substrates may thus be common to all IMP proteins in plants (and beyond). Secondary substrate specificity (e.g. histidinol-P in IMPL2) may have arisen subsequent to gene duplication, with the co-aption of the duplicated gene into alternative metabolic functions. Alternatively the ability to dephosphorylate histidinol-P may be an ancestral state, subsequently lost from the VTC4/IMPL1 clade. We suggest that proteins within the IMPL2 clade (Fig. 8) are likely HISN7 candidates.

In summary, we have demonstrated that IMPL2 functions as an HPP enzyme in Arabidopsis, identifying the “missing link” in this important primary metabolic pathway. IMPL2 is thus the HISN7 gene according to the nomenclature of Muralla et al. (2007).
While the enzyme has a broad substrate range in vitro, our data suggest that it is the only protein with HPP activity in Arabidopsis, at least in the developing embryo. This lack of genetic redundancy is characteristic of the His biosynthetic pathway in Arabidopsis, with five of the eight enzymes now known to be encoded by single copy genes, in contrast to the extensive redundancy observed in other amino acid biosynthetic pathways.

MATERIALS AND METHODS

Plant growth conditions
Plants were grown on a 1:1 mix of peat (Jiffy Products International, Stange, Norway) and vermiculite in a controlled environment room with a photosynthetic flux of 60-80 μmol m⁻² s⁻¹ under a 16 h light/8 h dark cycle at 22°C and 55% relative humidity. Plants were fertilised with 0.14% (w/v) Phostrogen plant food (Bayer CropScience, Cambridge) 7 and 21 d post-germination.

Genotypic analysis of T-DNA insertion lines
Two segregating T-DNA insertion knockout lines for At4g39120, SAIL_146_E09 (impl2-1) and SALK_076930 (impl2-2) were obtained from NASC. Plants were genotyped using gene specific primers in combination with the LB3 primer (5'-GAATTTCATAACCAATCTCGATACAC-3') for impl2-1, or the LBb1.3 primer (5'-ATTTTGCCGATTTCGGAAC-3') for impl2-2. Gene specific primers were as follows, impl2-1: L (5'-TGCTGTGAAGAAACAGAGCAG-3' and R (5'-AGGAGAATGGGCAGAGGAGTAG-3') and impl2-2: L (5'-ATTATCAAACCGAACCACACC-3') and R (5'-AGCAATCAGCGTACCACACAC-3'). Two PCR reactions were set up for each DNA sample, one with the gene specific L and R primers, and a second with the left border (LB) primer and R primer, both at an annealing temperature of 60°C, and an extension time of 1 min. The exact position of the T-DNA insertions in these lines was determined by DNA sequencing of the PCR product generated using the LB and R primers, and was found to lie at the 5’ boundary of exon I (impl2-1) and 5’ boundary of exon II (impl2-2) respectively.
Phenotypic analysis of T-DNA insertion lines

The analysis of siliques for the presence of aborted seeds in heterozygous individuals was carried out as described in the tutorial section of the Seed Genes Project website (www.seedgenes.org). Rescue experiments were carried out by supplying 1 mM His (dissolved in H2O) daily to the roots of heterozygous plants prior to bolting. Siliques were analysed 14 to 20 d after supplementation with His began.

Complementation of the Streptomyces coelicolor hisN mutant

Following an initial round of PCR site directed mutagenesis (Ho et al. 1989) to remove an internal NdeI site, the IMPL2 ORF (minus putative plastid transit peptide sequence) was amplified by PCR using the primers IMPL2-F (5'-GCCATATGGCCTCAACTCAAAC-3') and IMPL2-R (5'-GCTCTAGATCAATGCCACTGAGTGAC-3'), and cloned into the conjugative and integrative expression vector pIJ8600 (Sun et al., 1999) via the XbaI and NdeI sites in the MCS. A second construct containing the IMPL2 ORF (minus plastid transit peptide) with a Gly195 to Arg195 substitution was also generated by PCR site directed mutagenesis. Two overlapping PCR products were amplified using IMPL2-F in combination with Arg-R (5'-TGATCAATCAGTCGTAGAATCG-3') and IMPL2-R with Arg-F (5'-CGATTCTACGACTGATTGATCA-3'). The resulting products were used as the template in a second round of PCR with the IMPL2-F and IMPL2-R primers and the 760 bp PCR product cloned into pIJ8600 via the XbaI and NdeI sites. The IMPL1 ORF (minus putative plastid transit peptide sequence) was amplified by PCR using the primers 5'-GCCATATGCTGTCTGAGGTTTCTGAG-3' and 5'-GCCGATCCTAAAAGCTCTGTATGATAATC-3', and cloned into pIJ8600 via the BamHI and NdeI sites in the MCS. The pIJ8600 vectors were transformed into the methylation deficient E. coli ET12567 strain containing pUZ8002 which supplies transacting functions for the mobilisation of ori-T containing plasmids (Flett et al., 1997), and introduced into S. coelicolor strains by conjugation according to Kieser et al. (2000). The S. coelicolor strains used in this study were wild-type (strain M145) and the HPP mutant (strain 316; pheA1, hisN, strA1, SCP2+) which requires supplementation with 50 µg ml⁻¹ His and 37 µg ml⁻¹ Phe for growth on minimal media (Puglia et al., 1982).
Construction of GFP fusion protein vectors
The full length IMPL2 ORF was amplified by PCR using the primers GFP-F (5′-ATGCGTGCACATGTAGCTCAGTCGCACTTC-3′) and GFP-R (5′-ATTTGCGGCCGCGAATGCCACTCAAGTGACTCCA-3′), and the ORF minus putative plastid transit peptide sequence (IMPL2Δ1-60) with the primers GFP-CTP (5′-ATGCGTGCACATGGCTTCAAACTCAAAACG-3′) and GFP-R. The amplified sequences were cloned into the Gateway pENTR1A vector (Invitrogen, www.invitrogen.com) via the SalI and NotI sites, and then recombined into binary GATEWAY destination vector pK7FWG2.0 (Karimi et al., 2002).

Localisation of IMPL2:GFP fusion proteins
Protoplasts were isolated from 0.3 g of leaf tissue from maize plants grown under standard growth conditions for 9 d, followed by 2 d in continuous darkness according to the protocol of Yoo et al. (2007). PEG-calcium mediated transfection of 1×10⁴ protoplasts with 10 μg of plasmid DNA was carried out as described by Yoo et al. (2007). Imaging was carried out using a Zeiss LSM 510 Meta confocal microscope, with an LD C-Apochromat 40×/1.1 W M27 objective. GFP was excited at 488 nm and emission detected in the 500 to 520 nm range. Chlorophyll autofluorescence was detected in the 575 to 600 nm range.

Molecular Modelling
Homologues to IMPL2 (excluding the N-terminal 60 amino acid plastid transit peptide sequence) were found in the Protein Data Bank using pGenTHREADER (McGuffin and Jones, 2003). The alignment generated by pGenTHREADER was used as the basis for the construction of a homology model using MODELLER (Eswar et al., 2006). Several iterations were necessary in order to ensure that insertions and deletions were appropriately placed in the loops connecting the secondary structural elements (Supp. Fig. S2). Side chains were placed using SCWRL4 (Shapovalov and Dunbrack, 2007). Molecular graphics images were produced using the UCSF Chimera package from the
Supplemental Data
The following materials are available in the online version of this article.

Supplemental Figure S1. Sequence alignment of Arabidopsis IMPL1, IMPL2 and VTC4 proteins with known HPP enzymes from Corynebacterium glutamicum and Streptomyces coelicolor.

Supplemental Figure S2. pGenTHREADER alignment used to generate homology model of IMPL2.

Supplemental Figure S3. Expression patterns of VTC4, IMPL1 and VTC4 in Arabidopsis.

Supplemental Table SI. Top 10 genes showing the strongest positive correlation with IMPL2 expression across 322 publically available microarray datasets.

ACKNOWLEDGEMENTS

We thank Dirk Lang (UCT) for confocal imaging of protoplasts, Nicci Illing (UCT) for donation of cellulase and macerozyme, Jacqueline Bishop (UCT) for advice on phylogenetic analyses, Marc and Heather Knight (University of Durham, UK) for assistance with Gateway cloning, and Dave Woods for critical reading of the manuscript.

LITERATURE CITED


Fujimori K, Ohta D (1998a) An Arabidopsis cDNA encoding a bifunctional glutamine amidotransferase/cyclase suppresses the histidine auxotrophy of a Saccharomyces cerevisiae his7 mutant. FEBS Lett 428: 229-234


database for the meta-analysis of transcriptomes. Advances in Bioinformatics 2008: 420747


FIGURE LEGENDS

Figure 1. The Histidine biosynthetic pathway in plants. Abbreviations used for enzyme names in this work are shown in parentheses, and enzyme commission numbers and Arabidopsis gene loci indicated. Modified after Ward and Ohta (1999), Ingle et al. (2005) and Muralla et al. (2007).

Figure 2. Functional complementation of the Streptomyces coelicolor hisN mutant. Bacteria transformed with pIJ8600:IMPL2 were able to grow in the absence of supplemental His (50 µg ml⁻¹), while those transformed with pIJ8600:IMPL1 or a mutant version of IMPL2 where Gly195 was replaced with Arg (IMPL2G►R195) were not.

Figure 3. Homology modelling of IMPL2. (A) Superposition of the homology model of IMPL2 (blue) and inositol-1(or 4)-monophosphatase (A-chain) from Bos taurus (2bji) (yellow). The location of Gly195 is indicated in red and that of Pro295 in purple. The active site carboxylates, Mg²⁺ ions (green spheres) and closely associate water molecules (red spheres) are also shown. The β-sheet at residues 191 to 198, and α-helix from 287 to 298 are indicated by black arrows. Residues 1 to 60 comprising the plastid transit peptide of IMPL2 were not included in the model. (B) Close-up view of the best packing around Arg195 achieved by SCWRL4. The two residues with the greatest overlap (Arg195 and Pro295) are represented as ball and stick models.

Figure 4. Map of At4g39120 showing T-DNA insertion sites. Schematic of At4g39120 coding and upstream sequence, indicating T-DNA insertion sites in the lines used in this study. Blocks represent exons, solid lines represent introns, and dashed lines 5' upstream sequence.
**Figure 5.** Embryonic lethality of null *impl2* mutants. Representative images of immature silique from heterozygous and wildtype plants isolated from segregating *impl2-1* and *impl2-2* T-DNA insertion populations. Scale bar indicates a distance of 1 mm.

**Figure 6.** Exogenous His partially rescues the embryonic lethality of null *impl2* mutants. (A) The frequency of seed abortion in silique from primary bolts was determined in plants heterozygous for *impl2* mutant alleles, and in wild-type plants isolated from the *impl2-1* segregating population. Following removal of the primary bolt, plants were supplied daily with 1 mM His (+His) or water (No His). Analysis of silique from secondary bolts was carried out 16 d after supplementation with His began. Five silique were analysed per plant, and values shown are the mean percentage of aborted seeds per plant ±SD (n = 4 plants). Significant differences in the percentage of aborted seeds (p <0.05) between primary and secondary bolts (determined by paired t-test analysis) are indicated by *. (B) Percentage seed abortion in primary bolts from heterozygous plants supplied with 1 mM His (+His) or water (No His) daily for 20 d prior to analysis of silique. Wild-type plants isolated from the *impl2-1* and *impl2-2* segregating populations were used as controls. Five silique were analysed per plant, and values shown are the mean percentage frequency of seed abortion per plant ±SD (n = 5 plants). Significant differences in the percentage of aborted seeds between plants supplied with His and control plants (determined by two sample t-test analysis) are indicated by ** (p <0.01) and *** (p <0.001).

**Figure 7.** Subcellular localisation of IMPL2. Confocal images of maize protoplasts transiently expressing either full-length IMPL2 protein fused to GFP (IMPL2:GFP) or a truncated IMPL2 protein lacking the first 60 amino acids predicted to act as the plastid transit peptide fused to GFP (IMPL2Δ1-60:GFP). GFP emission at 500 to 520 nm (GFP signal), chlorophyll autofluorescence at 575 to 650 nm (chlorophyll) and a merged image (overlay) are shown. Scale bars indicate a distance of 20 µm.

**Figure 8.** Neighbour joining tree depicting phylogenetic relationships between *IMP* cDNA sequences in plants. cDNA sequences obtained from the NCBI RefSeq database
(accession numbers indicated) were aligned using ClustalW. The alignment was refined manually, including the removal of sequences encoding plastid transit peptides. Phylogenetic analysis was performed with Mega 4.0 using the neighbor joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) is indicated above the nodes subtending the branches. Evolutionary distances were computed using the Tajima-Nei method. The tree is rooted on the SAL1 gene from Arabidopsis (At5g63980) which encodes a 3’,5’-bisphosphate nucleotidase.
<table>
<thead>
<tr>
<th>Insertion line</th>
<th>Wildtype</th>
<th>Heterozygote</th>
<th>Homozygote</th>
<th>$\chi^2$ value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>impl2-1</td>
<td>17</td>
<td>30</td>
<td>0</td>
<td>0.1702</td>
<td>0.68</td>
</tr>
<tr>
<td>impl2-2</td>
<td>23</td>
<td>45</td>
<td>0</td>
<td>0.0075</td>
<td>0.93</td>
</tr>
</tbody>
</table>

**Table I.** PCR genotyping of segregating *impl2-1* and *impl2-2* populations. $\chi^2$ tests were used to determine whether the ratio of heterozygous to wildtype individuals was significantly different from 2:1.
<table>
<thead>
<tr>
<th>Insertion line</th>
<th>Genotype</th>
<th>Siliques screened</th>
<th>Total no. of seeds</th>
<th>Aborted seeds</th>
</tr>
</thead>
<tbody>
<tr>
<td>impl2-1</td>
<td>Het</td>
<td>41</td>
<td>1811</td>
<td>440 (24.3%)</td>
</tr>
<tr>
<td></td>
<td>Wt</td>
<td>10</td>
<td>429</td>
<td>9 (2.1%)</td>
</tr>
<tr>
<td>impl2-2</td>
<td>Het</td>
<td>23</td>
<td>1188</td>
<td>287 (24.2%)</td>
</tr>
<tr>
<td></td>
<td>Wt</td>
<td>8</td>
<td>272</td>
<td>1 (0.4%)</td>
</tr>
</tbody>
</table>

**Table II.** Analysis of siliques from plants heterozygous for null *impl2* alleles.
**Figure 1.** The Histidine biosynthetic pathway in plants. Abbreviations used for enzyme names in this work are shown in parentheses, and enzyme commission numbers and *Arabidopsis* gene loci indicated. Modified after Ward and Ohta (1999), Ingle et al. (2005) and Muralla et al. (2007).
Figure 2. Functional complementation of the *Streptomyces coelicolor* hisN mutant. Bacteria transformed with *pIJ8600:IMPL2* were able to grow in the absence of supplemental His (50 µg ml⁻¹), while those transformed with *pIJ8600:IMPL1* or a mutant version of IMPL2 where Gly195 was replaced with Arg (IMPL2G R195) were not.
Figure 3. Homology modelling of IMPL2. (A) Superposition of the homology model of IMPL2 (blue) and inositol-1(or 4)-monophosphatase (A-chain) from Bos taurus (2bji) (yellow). The location of Gly195 is indicated in red and that of Pro295 in purple. The active site carboxylates, Mg$^{2+}$ ions (green spheres) and closely associate water molecules (red spheres) are also shown. The β-sheet at residues 191 to 198, and α-helix from 287 to 298 are indicated by black arrows. Residues 1 to 60 comprising the plastid transit peptide of IMPL2 were not included in the model. (B) Close-up view of the best packing around Arg195 achieved by SCWRL4. The two residues with the greatest overlap (Arg195 and Pro295) are represented as ball and stick models.
**Figure 4.** Map of At4g39120 showing T-DNA insertion sites. Schematic of At4g31920 coding and upstream sequence, indicating T-DNA insertion sites in the lines used in this study. Blocks represent exons, solid lines represent introns, and dashed lines 5′ upstream sequence.
Figure 5. Embryonic lethality of null impl2 mutants. Representative images of immature siliques from heterozygous and wildtype plants isolated from segregating impl2-1 and impl2-2 T-DNA insertion populations. Scale bar indicates a distance of 1 mm.
A

![Bar chart showing percentage of aborted seeds for different genotypes: impl2-1 (+His), impl2-2 (+His), impl2-1 (No His), impl2-2 (No His), and wild-type.](chart)

B

![Bar chart showing percentage of aborted seeds for different genotypes: impl2-1, impl2-2, wild-type (impl2-1), and wild-type (impl2-2).](chart)
**Figure 6.** Exogenous His partially rescues the embryonic lethality of null *impl2* mutants. (A) The frequency of seed abortion in siliques from primary bolts was determined in plants heterozygous for *impl2* mutant alleles, and in wild-type plants isolated from the *impl2-1* segregating population. Following removal of the primary bolt, plants were supplied daily with 1 mM His (+His) or water (No His). Analysis of siliques from secondary bolts was carried out 16 d after supplementation with His began. Five siliques were analysed per plant, and values shown are the mean percentage of aborted seeds per plant +SD (n = 4 plants). Significant differences in the percentage of aborted seeds (p <0.05) between primary and secondary bolts (determined by paired t-test analysis) are indicated by *. (B) Percentage seed abortion in primary bolts from heterozygous plants supplied with 1 mM His (+His) or water (No His) daily for 20 d prior to analysis of siliques. Wild-type plants isolated from the *impl2-1* and *impl2-2* segregating populations were used as controls. Five siliques were analysed per plant, and values shown are the mean percentage frequency of seed abortion per plant +SD (n = 5 plants). Significant differences in the percentage of aborted seeds between plants supplied with His and control plants (determined by two sample t-test analysis) are indicated by ** (p <0.01) and *** (p <0.001).
Figure 7. Subcellular localisation of IMPL2. Confocal images of maize protoplasts transiently expressing either full-length IMPL2 protein fused to GFP (IMPL2:GFP) or a truncated IMPL2 protein lacking the first 60 amino acids predicted to act as the plastid transit peptide fused to GFP (IMPL2Δ1-60:GFP). GFP emission at 500 to 520 nm (GFP signal), chlorophyll autofluorescence at 575 to 650 nm (chlorophyll) and a merged image (overlay) are shown. Scale bars indicate a distance of 20 µm.
**Figure 8.** Neighbour joining tree depicting phylogenetic relationships between IMP cDNA sequences in plants. cDNA sequences obtained from the NCBI RefSeq database (accession numbers indicated) were aligned using ClustalW. The alignment was refined manually, including the removal of sequences encoding plastid transit peptides. Phylogenetic analysis was performed with Mega 4.0 using the neighbor joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) is indicated above the nodes subtending the branches. Evolutionary distances were computed using the Tajima-Nei method. The tree is rooted on the SAL1 gene from Arabidopsis (At5g63980) which encodes a 3′,5′-bisphosphate nucleotidase.