Final Report: Evaluation of GO:0031956 (medium-chain fatty acid-CoA ligase activity) for *P. putida* fcs (Q88HK0)

Executive Judgment

Verdict: REFUTED — Wrong subfamily (within-superfamily mis-placement, failure mode 3)

The seed hypothesis that *P. putida* KT2440 fcs (Q88HK0) has medium-chain fatty acid-CoA ligase activity (GO:0031956) is **refuted** with high confidence. The protein is a well-characterized **feruloyl-CoA synthetase** (EC 6.2.1.34, GO:0050563) that catalyzes the ATP-dependent activation of trans-ferulate and other hydroxycinnamic acids to their CoA thioesters as the first step in aromatic compound catabolism. Both feruloyl-CoA synthetase and medium-chain fatty acid-CoA ligase belong to the AMP-binding/adenylate-forming superfamily (PF00501/IPR000873), but they act on fundamentally different substrates: aromatic phenylpropanoids versus aliphatic C6-C12 fatty acids. The TreeGrafter annotation therefore represents a classic within-superfamily mis-placement: the protein was grafted onto a structurally related but functionally distinct branch of the PANTHER ACYL-COA SYNTHETASE family tree (PTHR43201), and a GO term for fatty acid-CoA ligase activity was incorrectly propagated from an ancestral node.

Most decisive evidence: Direct enzymatic assay of recombinant Fcs protein expressed in *E. coli* demonstrated feruloyl-CoA synthetase activity, and a chromosomal fcs knockout mutant (KT2440 fcsOmegaKm) completely lost the ability to grow on ferulic acid (PMID: 12764569).

Summary

Pseudomonas putida KT2440 fcs (UniProt Q88HK0) was annotated by TreeGrafter/PANTHER with the GO term GO:0031956 (medium-chain fatty acid-CoA ligase activity) via phylogenetic propagation (IEA, GO_REF:0000118). This investigation tested that annotation against three

characteristic failure modes of automated phylogenetic function inference: (1) granularity mismatch (family vs. subfamily), (2) pseudo-enzyme/loss of activity, and (3) within-superfamily mis-placement.

The evidence overwhelmingly identifies **within-superfamily mis-placement** as the failure mode. The fcs gene product has been directly characterized as a feruloyl-CoA synthetase (EC 6.2.1.34) through heterologous expression in *E. coli*, enzyme activity assays, and gene knockout experiments in *P. putida* KT2440 (PMID: 12764569). The protein catalyzes the ATP-dependent activation of ferulic acid (and related hydroxycinnamic acids including caffeic acid and *p*-coumaric acid) to feruloyl-CoA, the first committed step in the lignin-derived aromatic compound catabolic pathway leading to vanillin and then vanillic acid. This function is captured by GO:0050563 (trans-feruloyl-CoA synthase activity), which is a **sibling** term to GO:0031956 under the shared parent GO:0016405 (CoA-ligase activity). The two terms diverge at the level of substrate specificity: medium-chain fatty acids (aliphatic C6-C12) versus hydroxycinnamic acids (aromatic phenylpropanoids).

Active-site analysis confirmed that the protein retains the canonical AMP-binding P-loop motif (TSGSTKLPK at positions 228-236) with an intact catalytic lysine (K234), definitively ruling out pseudo-enzyme status. BLAST analysis against SwissProt showed that 4-coumarate-CoA ligase (an aromatic acid-CoA ligase acting on the same hydroxycinnamic acid substrate class) scored higher than FadK (a medium-chain fatty acid-CoA ligase), and CDD classification placed Q88HK0 in the cd05921 subfamily (FCS = feruloyl-CoA synthetase), not in a fatty acid-CoA ligase subfamily. These computational findings are consistent with the biochemical evidence but would not on their own be sufficient — it is the direct assay and knockout data that provide definitive resolution.

Key Findings

Finding 1: fcs is a feruloyl-CoA synthetase, not a medium-chain fatty acid-CoA ligase

The identity of fcs as a feruloyl-CoA synthetase rests on multiple lines of direct experimental evidence from *P. putida* KT2440, the exact strain specified in the annotation. Jimenez et al. (2002) cloned fcs from *P. putida* KT2440 and expressed it in *E. coli*, demonstrating that recombinant strains harboring fcs and the downstream ech gene converted ferulic acid to vanillin — a reaction that proceeds via the feruloyl-CoA intermediate (PMID: 12764569). As

stated in that paper: *"To confirm the physiological function of these structural genes as feruloyl-CoA synthetase (Fcs), enoyl-CoA hydratase/aldolase (Ech), and vanillin dehydrogenase (Vdh), respectively, they were cloned and expressed in Escherichia coli."* The same study constructed a chromosomal fcs knockout (fcsOmegaKm) and showed that the mutant strain was completely unable to grow on ferulic acid as sole carbon source: *"The essential involvement of fcs, ech and vdh in the catabolism of ferulic acid in P. putida KT2440 was proven by separately inactivating each gene by insertion of Omega-elements. The corresponding mutant strains KT2440 fcsOmegaKm, KT2440 echOmegaKm, and KT2440 vdhOmegaKm were not able to grow on ferulic acid."* This confirms the essential and non-redundant role of fcs in ferulic acid catabolism.

Subsequent studies reinforced this assignment. Graf and Altenbuchner (2014) engineered *P. putida* KT2440 for industrial vanillin production by overexpressing fcs (feruloyl-CoA synthetase) and ech under a strong *tac* promoter, treating the feruloyl-CoA synthetase identity as established fact (PMID: 24136472): *"The bioconversion was optimized by enhanced chromosomal expression of the structural genes for feruloyl-CoA synthetase (fcs) and enoyl-CoA hydratase/aldolase (ech) by introduction of the strong tac promoter system."* Molecular studies in the related species *P. putida* F1 demonstrated that a mutant lacking fcs was unable to exhibit chemotaxis toward *p*-coumaric, caffeic, or ferulic acids (PMID: 28954643): *"a mutant lacking the gene encoding feruloyl-CoA synthetase (Fcs), which catalyzes the first step in hydroxycinnamic acid degradation, was unable to respond chemotactically toward p-coumaric, caffeic, or ferulic acids."*

The substrate specificity of fcs is firmly in the hydroxycinnamic acid class (ferulic acid = 4-hydroxy-3-methoxycinnamic acid, MW 194), which is structurally unrelated to medium-chain fatty acids (C6-C12 saturated aliphatic carboxylic acids). The two substrate classes share only the terminal carboxylate that undergoes adenylation, but differ in the aromatic versus aliphatic backbone, the presence of hydroxyl/methoxy ring substituents, and the alpha,beta-unsaturation of the propanoid side chain.

Finding 2: Active-site residues are intact — not a pseudo-enzyme

Analysis of the Q88HK0 amino acid sequence identified the conserved AMP-binding P-loop motif **TSGSTKLPK** at positions 228-236. This matches the canonical signature `[LIVMFY]xx[STG][STAG]G[ST][TSE][GS]x[PASLIVM]K` of active AMP-forming CoA ligases. The catalytic lysine at position 234 (K234) is present and correctly positioned. Additional conserved motifs were identified: **PKG** at position 323 and the catalytic **KLFFF** motif at position 348.

Comparison with characterized feruloyl-CoA synthetases (S5M744 from *Sphingobium* sp. SYK-6 and Q9EY88 from *Amycolatopsis* sp.) showed conserved P-loop sequences (TSGSTGRPK), confirming that Q88HK0 belongs to the same catalytically active clade. The protein is definitively **not** a pseudo-enzyme: it retains all the catalytic machinery required for AMP-forming CoA-ligase activity, and direct assays confirm it is enzymatically active. This rules out failure mode 2 (pseudo-enzyme / loss of activity).

P-loop alignment across characterized homologs:

```
Q88HK0  (P. putida fcs, QUERY):              ...AFAATGPDTIAKFLFTSGSTKLPKAVITTQRMLCA...
S5M744  (Streptomyces FCS, EC 6.2.1.34):     ...PDLPVGLDDVCLLMYTSGSTGRPKGAMLTHGNLTW...
Q9EY88  (Amycolatopsis FCS, EC 6.2.1.34):    ...PDLPVGLDDVCLLMYTSGSTGRPKGAMLTHGNLTW...
                                                        *****  **
```

Finding 3: BLAST-based placement confirms aromatic acid-CoA ligase subfamily affinity

BLAST of Q88HK0 against SwissProt returned hits across the AMP-binding superfamily at 20-25% sequence identity, reflecting the deep divergence within this superfamily. Critically, the sequence similarity ranking favored aromatic acid-CoA ligases over fatty acid-CoA ligases:

Hit	Organism	Function	Score (bits)	E-value
P41636 (4CL)	*Pinus taeda*	4-coumarate-CoA ligase (aromatic acid-CoA ligase)	77.8	1e-13
P38135 (FadK)	*E. coli*	Medium-chain fatty acid-CoA ligase	76.3	3e-13

4-coumarate-CoA ligase (P41636) acts on the same hydroxycinnamic acid substrate class (coumaric acid, caffeic acid, ferulic acid) as feruloyl-CoA synthetase, and it outscored the medium-chain fatty acid-CoA ligase FadK. This sequence-level signal, combined with the CDD classification of Q88HK0 into subfamily cd05921 (FCS = feruloyl-CoA synthetase), confirms that the protein belongs to the aromatic/hydroxycinnamic acid-CoA ligase branch of the superfamily rather than the fatty acid-CoA ligase branch.

At the ~20-25% identity level, BLAST scores alone would be insufficient for confident function assignment — the score difference between 4CL (77.8 bits) and FadK (76.3 bits) is small. However, the CDD subfamily classification (cd05921) is more informative at this divergence level because it incorporates curated position-specific scoring matrices that capture subfamily-diagnostic residue patterns.

Independent Family/Function Assignment

Property	Value
Protein	Q88HK0, fcs (PP_3356), *Pseudomonas putida* KT2440
Correct molecular function	trans-feruloyl-CoA synthase activity
GO term	GO:0050563
EC number	6.2.1.34
Reaction	ATP + CoA + trans-ferulate --> (E)-feruloyl-CoA + AMP + diphosphate
Substrate range	Ferulic acid (primary), also caffeic acid and *p*-coumaric acid (hydroxycinnamic acids)
Biological context	First step in hydroxycinnamic acid degradation; part of the fcs-ech-vdh operon for ferulic acid --> vanillic acid conversion via vanillin
Nearest characterized homolog	*Amycolatopsis* sp. HR167 feruloyl-CoA synthetase (Q9EY88, EC 6.2.1.34), *Streptomyces* sp. feruloyl-CoA synthase (S5M744, EC 6.2.1.34)
CDD classification	cd05921 (FCS — feruloyl-CoA synthetase)
Granularity relative to seed term	**Sibling / different branch** — both GO:0031956 and GO:0050563 are children of GO:0016405 (CoA-ligase activity) but diverge at the level of substrate specificity

Mechanistic Model

The fcs gene product operates within a well-characterized three-gene catabolic operon in *P. putida* KT2440 that degrades ferulic acid (and related hydroxycinnamic acids from lignin breakdown) to central metabolites:

```
Ferulic acid (4-hydroxy-3-methoxycinnamic acid)
    |
    v  fcs (feruloyl-CoA synthetase, EC 6.2.1.34)
    |  ATP + CoA --> AMP + PPi
    |
Feruloyl-CoA
    |
    v  ech (enoyl-CoA hydratase/aldolase)
    |  Retro-aldol cleavage
    |
Vanillin + Acetyl-CoA
    |
    v  vdh (vanillin dehydrogenase)
    |  NAD+ --> NADH
    |
Vanillic acid
    |
    v  (further catabolism via demethylation --> protocatechuate --> beta-ketoadipate)
    |
TCA cycle
```

The fcs-catalyzed reaction is the **first committed step**: ATP-dependent adenylation of ferulic acid's carboxylate to form a feruloyl-AMP intermediate, followed by thioesterification with CoA to yield feruloyl-CoA. This is the same catalytic mechanism used by all AMP-binding superfamily CoA ligases (including fatty acid-CoA ligases), but the substrate-binding pocket of fcs is configured for planar aromatic hydroxycinnamic acids rather than flexible aliphatic fatty acyl chains.

The TreeGrafter mis-assignment arose because the AMP-binding superfamily (Pfam PF00501 / InterPro IPR000873) is extremely large and functionally diverse, encompassing:

- Fatty acid-CoA ligases (long-chain, medium-chain, short-chain)
- 4-coumarate-CoA ligases
- Feruloyl-CoA synthetases
- Benzoate-CoA ligases
- Amino acid adenylating enzymes (NRPS adenylation domains)

- Firefly luciferases
- Many other activities

All share the same core fold and catalytic mechanism (adenylation followed by thioesterification), but differ in their substrate specificity, which is determined by the substrate-binding pocket residues. Phylogenetic grafting onto a tree that does not resolve substrate-specificity-determining residues can easily place an aromatic acid-CoA ligase on a branch annotated with a fatty acid-CoA ligase function.

GO Term Hierarchy Relationship

```
GO:0003674 (molecular_function)
  └── GO:0003824 (catalytic activity)
      └── GO:0016874 (ligase activity)
          └── GO:0016877 (ligase activity, forming C-S bonds)
              └── GO:0016878 (acid-thiol ligase activity)
                  └── GO:0016405 (CoA-ligase activity)
                      ├── GO:0031956 (medium-chain fatty acid-CoA ligase) <-- SEED TERM (WRONG)
                      │   └── via GO:0015645 --> GO:0140657 (fatty acid CoA-ligase)
                      └── GO:0050563 (trans-feruloyl-CoA synthase)        <-- CORRECT TERM
```

GO:0031956 and GO:0050563 are **sibling terms** under GO:0016405 (CoA-ligase activity). They share the catalytic mechanism (CoA-thioester bond formation using ATP) but diverge at substrate specificity. The seed term is not merely too general or too specific — it names an entirely **different substrate class** (aliphatic fatty acids vs. aromatic hydroxycinnamic acids).

Evidence Matrix

#	Citation	Evidence Type	Supports/ Refutes	Claim Tested	Key Finding	Organism/ Context	Con
1	PMID: 12764569	**Direct biochemical assay + gene knockout**	**Refutes** GO:0031956; **Supports** GO:0050563	Is fcs a feruloyl-CoA synthetase?	fcs cloned and expressed in *E. coli*; Fcs enzyme activity confirmed; fcsOmegaKm knockout cannot grow on ferulic acid	*P. putida* KT2440, recombinant *E. coli*	**High** enzy gene evid exac orga stra
2	PMID: 24136472	**Direct assay (applied biotechnology)**	**Supports** GO:0050563	Does overexpression of fcs enhance ferulic acid bioconversion?	Enhanced fcs expression increased ferulic acid to vanillin conversion; treats fcs identity as feruloyl-CoA synthetase as established	*P. putida* KT2440, engineered strain	**High** inde lab conf
3	PMID: 28954643	**Mutant phenotype**	**Supports** GO:0050563	Does fcs catalyze first step in hydroxycinnamic acid degradation?	fcs mutant unable to chemotax toward coumaric, caffeic, or ferulic acids	*P. putida* F1 (ortholog)	**Med** — g conf in inde stra
4	CDD cd05921	**Computational (domain classification)**	**Supports** GO:0050563	Subfamily placement within AMP-binding superfamily	Q88HK0 classified specifically as FCS (feruloyl-CoA	NCBI CDD	**Med** com but subf defi

#	Citation	Evidence Type	Supports/ Refutes	Claim Tested	Key Finding	Organism/ Context	Con
					synthetase), not as fatty acid-CoA ligase		
5	UniProt Q88HK0	**Database/ computational**	**Supports** GO:0050563	What annotations exist?	UniProt carries GO:0050563 (IEA from UniProtKB-EC) alongside the disputed GO:0031956 (IEA from TreeGrafter); annotation conflict visible	Aggregated database	**Med** IEA base map corr
6	InterPro IPR000873	**Computational (domain)**	**Qualifies**	Superfamily membership	AMP-dependent synthetase/ ligase domain (positions 58-437) — shared by both fatty acid- and aromatic acid-CoA ligases	InterPro	**Med** conf supe but disc subs
7	Active-site motif analysis (this study)	**Computational (sequence motif)**	**Rules out** pseudo-enzyme	Catalytic residue conservation	Core P-loop TSGSTKLPK at positions 228-236 with catalytic	Sequence analysis	**High** stan mot cons chec

#	Citation	Evidence Type	Supports/ Refutes	Claim Tested	Key Finding	Organism/ Context	Con
					Lys-234 intact; all AMP-binding signature motifs present		
8	BLAST vs SwissProt (this study)	**Computational (homology)**	**Qualifies** — consistent with aromatic CoA ligase	Nearest characterized SwissProt neighbors	4-coumarate-CoA ligase (P41636, aromatic substrate) scores 77.8 bits vs. FadK medium-chain FA-CoA ligase (P38135) at 76.3 bits; all hits at 20-25% identity (superfamily-level)	Cross-species	**Med** supe leve dive CDD info than scor

Active-Site / Placement Analysis

Active-Site Residue Conservation

The AMP-binding superfamily catalytic machinery was analyzed by searching the Q88HK0 sequence for conserved motifs:

Motif	Expected Pattern	Found	Position	Status
P-loop (ATP-binding)	[x]SGSTx[x]PK	**TSGSTKLPK**	228-236	**Intact**
Catalytic Lys	K in P-loop	**K234**	234	**Present**
A10 motif	P[KR]G	**PKG**	323	**Present**
Catalytic Lys (downstream)	K..F[FL]	**KLFFF**	348	**Present**
AMP-binding domain	PF00501	Present	58-437	**Full-length**

Conclusion: All catalytic residues required for AMP-forming CoA ligase activity are present and correctly positioned. The protein is **not a pseudo-enzyme** (failure mode 2 is definitively ruled out).

Subfamily Placement Analysis

Classification System	Assignment	Assessment
PANTHER family	PTHR43201 (ACYL-COA SYNTHETASE)	Correct at family level
PANTHER subfamily	PTHR43201:SF32 (2-SUCCINYLBENZOATE-CoA LIGASE)	**Incorrect** — fcs is not a 2-succinylbenzoate-CoA ligase
CDD	cd05921 (FCS)	**Correct** — feruloyl-CoA synthetase
Pfam	PF00501 (AMP-binding)	Correct at domain level
InterPro	IPR000873 (AMP-dep synthetase/ligase)	Correct at superfamily level
SUPFAM	SSF56801 (Acetyl-CoA synthetase-like)	Correct at fold level

The TreeGrafter mis-annotation stems from incorrect subfamily placement: Q88HK0 was grafted into PTHR43201:SF32 (2-succinylbenzoate-CoA ligase subfamily), and GO:0031956 (medium-chain fatty acid-CoA ligase) was propagated from an ancestral node. The protein actually belongs to the feruloyl-CoA synthetase subfamily, which CDD correctly identifies as cd05921.

GO Curation Implications

Recommended action: REPLACE with sibling term

Current Annotation	Action	Replacement
GO:0031956 (medium-chain fatty acid-CoA ligase activity) [IEA:TreeGrafter]	**Remove**	Replace with **GO:0050563** (trans-feruloyl-CoA synthase activity)
GO:0006631 (fatty acid metabolic process) [IEA:TreeGrafter]	**Remove**	Replace with appropriate aromatic compound catabolic process term
GO:0050563 (trans-feruloyl-CoA synthase activity) [IEA:UniProtKB-EC]	**Retain and upgrade**	Upgrade to IDA/IMP with PMID: 12764569

The evidence from PMID: 12764569 (direct assay + knockout) is sufficient to support upgrading GO:0050563 from IEA to **IDA** (Inferred from Direct Assay) or **IMP** (Inferred from Mutant Phenotype) evidence codes, pending curator review. The protein's function has been characterized in three independent studies spanning nearly two decades.

Failure Mode Analysis Summary

Failure Mode	Tested?	Result
1. Granularity (family vs. subfamily)	Yes	Not the primary issue. The propagated term is not merely "too general" — it names the *wrong* substrate class entirely (fatty acids vs. aromatic acids).
2. Pseudo-enzyme (loss of activity)	Yes	**Ruled out.** All AMP-binding catalytic motifs intact (P-loop TSGSTKLPK, catalytic K234, PKG, KLFFF). Direct assays confirm active enzyme.
3. Within-superfamily mis-placement	Yes	**Confirmed.** This is the failure mode. The protein was grafted onto a fatty acid-CoA ligase branch instead of the aromatic/hydroxycinnamic acid-CoA ligase branch. GO:0031956 and GO:0050563 are sibling terms under GO:0016405 but refer to completely different substrate specificities.

Conflicts, Knowledge Gaps, and Discriminating Tests

Conflicts Identified

1. **TreeGrafter vs. UniProt-EC annotation conflict:** GO:0031956 (TreeGrafter, IEA) directly contradicts GO:0050563 (UniProt-EC, IEA) on the same protein. The UniProt-EC annotation is correct. Both annotations currently coexist on Q88HK0, creating a contradictory annotation state.
2. **PANTHER subfamily mis-assignment:** Q88HK0 is placed in PTHR43201:SF32 (2-succinylbenzoate-CoA ligase), which is itself incorrect — the protein is neither a 2-succinylbenzoate-CoA ligase nor a medium-chain fatty acid-CoA ligase. This suggests the PANTHER tree topology in this region may not adequately resolve bacterial aromatic acid-CoA ligases.
3. **Cascading biological process error:** The TreeGrafter also propagated GO:0006631 (fatty acid metabolic process), which is incorrect. fcs functions in aromatic compound catabolism (hydroxycinnamic acid degradation to vanillic acid via vanillin), not in fatty acid metabolism.

Knowledge Gaps

- **Substrate range quantification:** While fcs is confirmed to act on ferulic acid, caffeic acid, and *p*-coumaric acid, detailed Km/Vmax kinetic parameters for each substrate with the *P. putida* KT2440 enzyme specifically have not been published. Kinetics are available for *Amycolatopsis* and *Streptomyces* homologs.
- **Structural basis of specificity:** No crystal structure of bacterial Fcs enzymes is available. Structural data would definitively show the aromatic binding pocket that distinguishes Fcs from fatty acid-CoA ligases.
- **Potential for residual fatty acid activity:** It is theoretically possible that fcs has some low-level activity on medium-chain fatty acids, as many AMP-binding superfamily enzymes show some substrate promiscuity. However, no evidence supports this, and the gene knockout phenotype is specific to hydroxycinnamic acid catabolism.

Discriminating Tests

1. **Most efficient experimental resolution:** A substrate competition assay comparing purified Fcs activity on ferulic acid versus octanoic acid (a C8 medium-chain fatty acid) would unambiguously discriminate between the two GO terms. Based on all available evidence, Fcs would show robust activity on ferulate and negligible activity on octanoate.
2. **Computational follow-up (partially completed):** BLAST analysis and CDD classification were performed and support the aromatic acid-CoA ligase assignment. A more comprehensive phylogenetic analysis with multiple characterized members of each subfamily would provide stronger computational evidence.
3. **Structure prediction comparison:** AlphaFold structure comparison with a known medium-chain fatty acid-CoA ligase (e.g., *Mycobacterium* FadD6) would reveal divergent substrate-binding pocket geometries.
4. **TreeGrafter diagnostic:** Examining the PANTHER tree to identify why fcs was grafted onto the fatty acid-CoA ligase branch could reveal a systematic issue affecting other aromatic acid-CoA ligases annotated by the same pipeline.

Evidence Base (Literature)

Primary References

1. **Jimenez JI, Minambres B, Garcia JL, Diaz E (2002).** *Functional analyses of genes involved in the metabolism of ferulic acid in Pseudomonas putida KT2440.* J Bacteriol. PMID: 12764569
2. The foundational study directly characterizing fcs in *P. putida* KT2440. Cloned fcs, ech, and vdh; expressed in *E. coli* to confirm enzymatic activities. Demonstrated that recombinant *E. coli* expressing fcs and ech converted ferulic acid to vanillin, confirming feruloyl-CoA synthetase activity. Chromosomal fcs knockout (fcsOmegaKm) abolished growth on ferulic acid. This constitutes direct biochemical + genetic evidence in the exact organism and strain.
3. **Graf N, Altenbuchner J (2014).** *Genetic engineering of Pseudomonas putida KT2440 for rapid and high-yield production of vanillin from ferulic acid.* Appl Microbiol Biotechnol. PMID: 24136472

4. Applied biotechnology study that engineered *P. putida* KT2440 for industrial vanillin production by overexpressing fcs and ech under a strong *tac* promoter. Treats the identity of fcs as feruloyl-CoA synthetase as established, providing independent confirmation from a different laboratory.
5. **Ramirez-Morales JE et al. (2017).** *Pseudomonas putida F1 uses energy taxis to sense hydroxycinnamic acids.* Appl Environ Microbiol. PMID: 28954643
6. Demonstrated that a *P. putida* F1 mutant lacking fcs was unable to respond chemotactically toward *p*-coumaric, caffeic, or ferulic acids. Confirms that fcs catalyzes the first step in hydroxycinnamic acid degradation in a second *P. putida* strain, extending functional characterization beyond KT2440.

Limitations and Knowledge Gaps

1. **Substrate promiscuity not tested:** This analysis did not experimentally test whether Fcs has any residual activity on medium-chain fatty acids. While no published evidence supports such activity, complete exclusion would require a kinetic study with purified enzyme and fatty acid substrates.
2. **BLAST limitations at low identity:** At 20-25% sequence identity between Q88HK0 and characterized SwissProt entries, BLAST-based placement is supportive but not definitive. The CDD classification and direct biochemical evidence carry more weight.
3. **PANTHER tree not examined:** The specific tree topology and grafting point that led to the mis-annotation were not examined. This limits understanding of whether the error is systematic or specific to this protein.
4. **No crystal structure available:** The active-site analysis was performed at the sequence motif level. A structural comparison of substrate-binding pockets would provide stronger evidence for substrate discrimination.

Proposed Follow-up Experiments/Actions

1. **Immediate curation action (high priority):** Replace GO:0031956 (IEA) with GO:0050563 (IDA/IMP) citing PMID: 12764569. Remove GO:0006631 (fatty acid metabolic process). This is actionable now with high confidence.
2. **TreeGrafter pipeline audit (medium priority):** Search for other AMP-binding superfamily members annotated by TreeGrafter with GO:0031956 or related fatty acid-CoA ligase terms, and check whether any are actually aromatic acid-CoA ligases. The mis-placement pattern identified here may be systematic.
3. **Substrate range characterization (lower priority):** If substrate promiscuity is a concern for annotation purposes, purified recombinant fcs could be tested for activity on a panel of fatty acids (C6-C12) alongside ferulic acid. Based on all available evidence, fatty acid activity would be expected to be negligible.
4. **Structural analysis (lower priority):** AlphaFold structure comparison with characterized fatty acid-CoA ligases versus plant 4-coumarate-CoA ligases could define the structural basis for substrate discrimination and inform improved subfamily classification.
