Hypothesis Review: Q6YYC5 (*Oryza sativa* japonica, Os08g0135400) — protein K63-linked ubiquitination (GO:0070534)

Focus type: computational_prediction (ProtNLM2) — proposed linkage-specific refinement of GO:0016567 → GO:0070534 **Term under evaluation:** protein K63-linked ubiquitination (GO:0070534) **Investigation:** 3 iterations · 5 confirmed findings · 12 papers reviewed

Summary

Q6YYC5 is a genuine member of the RGLG (RING Domain Ligase) family of E3 ubiquitin ligases. Independent domain analysis confirms it carries the two hallmark features of the family: an N-terminal von Willebrand factor type A (vWA) domain and a C-terminal, **intact** C3HC4 RING-type zinc finger with all eight canonical metal-coordinating residues present and correctly spaced. This architecture firmly justifies the protein's **generic** annotations for protein ubiquitination (GO:0016567) and ubiquitin–protein transferase activity (GO:0004842). Its identity as a functional E3 ligase is not in doubt.

The ProtNLM2 refinement to **protein K63-linked ubiquitination (GO:0070534)** is, however, an **over-annotation** that should not be propagated. The prediction fails on two independent grounds. First and most decisively, ubiquitin-chain linkage in RING-type E3 ligases is determined by the cognate E2-conjugating enzyme, not by the E3's own sequence — the same RING E3 can build K63 chains with one E2 and K48 chains with another, so linkage specificity cannot be read out from an E3 sequence. Second, the orthology argument used to justify the refinement is misdirected: K63-chain assembly was demonstrated in vitro only for *Arabidopsis* RGLG2 (and inferred for its closest sequelog RGLG1), but Q6YYC5 is phylogenetically closest to the **RGLG3**/**RGLG4** clade (~60% identity to RGLG4). That clade uses a different E2 (UBC30) and drives **degradative** ubiquitination of its substrate GRXS17 — the opposite of an obligate K63-signaling signature.

Bottom line for curators: Retain the generic ubiquitination/E3-ligase terms (well supported by structure and family membership). Do **not** add GO:0070534 to Q6YYC5. The linkage specialization is not decidable from sequence or orthology and would require a direct in vitro chain-linkage assay with the rice cognate E2 to establish. Verdict: **generic function supported; K63 refinement over-annotated / weakly supported.**

Key Findings

F001 — Q6YYC5 has canonical RGLG-family architecture (vWA + intact C3HC4 RING)

UniProt Q6YYC5 (gene Os08g0135400; 401 amino acids) carries the two defining domains of the RGLG family. InterPro, Pfam, and SMART annotate an N-terminal **von Willebrand factor type A (vWA) domain** (IPR002035; SMART SM00327) together with a C-terminal **RING-type zinc finger** at residues 356–389 (Pfam PF13920, zf-C3HC4_3; PROSITE PS50089). Critically, the catalytic RING is intact rather than degenerate: the computed C3HC4 cross-brace contains all eight metal-coordinating ligands in canonical spacing — C356, C359, C370, H372, C374/375, C378, C385, and C388 — and the sequence matches the canonical C3HC4 RING regular expression (`C..C.{9,39}C.{1,3}H.{2,3}C..C.{4,48}C..C`). PANTHER independently classifies the protein under subfamily **PTHR45751:SF16 "E3 UBIQUITIN-PROTEIN LIGASE RGLG4."** The combination of an N-terminal vWA domain and a C-terminal foldable catalytic RING is the RGLG-family signature. This finding establishes Q6YYC5 as a bona fide E3 ubiquitin ligase and justifies the generic ubiquitination annotations — but says nothing about chain linkage.

F002 — Q6YYC5 is closest to *Arabidopsis* RGLG4/RGLG3, not to the K63-demonstrated RGLG1/RGLG2

Global pairwise alignment of Q6YYC5 against the five reviewed *Arabidopsis* RGLG proteins yields the following percent identities: **RGLG4 59.8%**, RGLG5 56.6%, RGLG3 56.1%, RGLG2 52.8%, and RGLG1 48.5%. The closest paralog, RGLG4, is also identical in length (401 aa) to Q6YYC5, consistent with the independent PANTHER subfamily assignment (SF16 = RGLG4). This distinction is the crux of the review, because the K63-chain-forming activity that motivates the hypothesis was demonstrated **in vitro only for RGLG2** (and inferred for RGLG1) in PMID: 17586653. Q6YYC5's own clade (RGLG3/RGLG4) has a distinct, characterized biochemistry: it partners with the cognate E2 **UBC30** and drives degradation of the substrate GRXS17 (PMID:

27497447) — a degradative outcome that runs counter to the non-degradative signaling role classically ascribed to K63 chains. The orthology bridge required to transfer "K63" from RGLG2 to Q6YYC5 therefore does not hold; Q6YYC5 sits in the wrong sub-branch of the family.

F003 — Ubiquitin-chain linkage specificity of RING E3s is set by the E2 partner, not by the E3 sequence

This is the central mechanistic reason the K63 refinement is not sequence-decidable. RING-type E3s do not form a catalytic thioester intermediate; they transfer ubiquitin **directly** from the E2~Ub conjugate to the substrate, so the geometry that determines which lysine of the acceptor ubiquitin is attacked — and hence the chain linkage — is contributed by the E2. As stated verbatim in PMID: 40169231: *"RING-type E3s mediate the transfer of Ub directly from the E2~Ub conjugate, implying that the specificity of Ub linkage is determined by the given E2."* This is not merely a theoretical concern: PMID: 17426036 shows experimentally that a single RING E3 (MuRF1) builds **K63 chains with the E2 UbcH13/Uev1a but K48 chains with the E2 UbcH1 (E2-25K)**. Because Q6YYC5's cognate rice E2 is unknown and no linkage assay has been performed on the rice protein, the chain type it would produce cannot be inferred from its own sequence. A RING E3 is, on its own, linkage-agnostic.

F004 — Neighbor-joining tree confirms Q6YYC5 is sister to RGLG4, separate from the K63-forming RGLG1/RGLG2 clade

An independent phylogenetic reconstruction reinforces F002. A neighbor-joining tree built from a six-sequence global-alignment distance matrix produced the following topology (Newick):

```
((RGLG3:23.7,(RGLG5:17.7,(RGLG1:16.3,RGLG2:12.4):7.1):6.9):0.6,(Q6YYC5_Os:18.9,RGLG4:21.4):0.6);
```

Q6YYC5 pairs with RGLG4 as a sister leaf, while **RGLG1 and RGLG2 — the K63-demonstrated ligases — form their own tight clade** (pairwise distance 28.6 between them, versus Q6YYC5's much larger distances of ~40–52 to that pair). Additionally, Q6YYC5 retains an N-terminal Gly2 (sequence begins "MGG…"), a potential N-myristoylation site analogous to the one that targets RGLG2 to the plasma membrane. This retained motif is a conserved family feature but does not bear on linkage specificity. The tree topology cleanly separates Q6YYC5 from the K63-forming branch, corroborating the identity-based clade assignment.

F005 — K63-chain formation in the RGLG pathway requires the K63-specific E2 UBC13, whose cognate E3s are RGLG1/RGLG2 (not the RGLG4 clade)

The genetic and biochemical partitioning of the RGLG family is explicit in the literature. PMID: 20113438 (Li & Schmidt, 2010) reports that UBC13 *"has been shown to catalyze non-canonical Lys63-linked ubiquitin chains,"* and that *"Mutations in the cognate E3 ligases RGLG2 and RGLG1 caused the constitutive formation of branched root hairs,"* tying the K63/UBC13 activity specifically to RGLG1/RGLG2 in the iron-deficiency response. In direct contrast, the RGLG3/RGLG4 clade — the clade to which Q6YYC5 belongs — uses the E2 UBC30 to drive GRXS17 degradation (PMID: 27497447). No published evidence links Q6YYC5, or the RGLG4 clade more broadly, to UBC13 or to K63 chains. The K63 module is therefore an RGLG1/RGLG2 + UBC13 property, and there is no basis to extend it to Q6YYC5.

Mechanistic Model / Interpretation

The seed hypothesis can be laid out as a three-link inference chain, and this investigation shows precisely where each link holds or fails:

```
Link 1: Q6YYC5 is an RGLG-family RING E3 ligase          → HOLDS   (F001, F004)
Link 2: RGLG-family E3s form K63 chains                  → PARTIAL (only RGLG1/2, via E2 UBC13)
Link 3: Therefore Q6YYC5 forms K63 chains                → FAILS   (F002, F003, F005)
```

The failure is best understood through how RING E3 catalysis actually works. The RING E3 is a scaffold that juxtaposes the charged E2~Ub conjugate and the substrate; it does not itself select the acceptor lysine:

```
        E2~Ub  ──────►  substrate–Ub  ──────►  poly-Ub chain
          │                                        │
   (chain linkage type                The RING E3 (Q6YYC5) positions E2~Ub
    is set HERE, by the E2)            and substrate. It does NOT dictate
                                       K48 vs K63 on its own.
```

Within the *Arabidopsis* RGLG family there are two functionally distinct modules, and Q6YYC5 maps onto the non-K63 one:

Sub-branch	Members	Cognate E2	Chain / outcome	Characterized biology
RGLG1/ RGLG2	RGLG1, RGLG2	**UBC13**	**K63**, non-degradative signaling	Apical dominance, auxin transport (PIN), Fe-deficiency root-hair branching
RGLG3/ RGLG4	RGLG3, RGLG4	**UBC30**	degradative	GRXS17 degradation; jasmonate signaling; FB1-triggered PCD

Q6YYC5 maps by both sequence identity (F002) and tree topology (F004) onto the **RGLG3/ RGLG4** row — the row associated with UBC30 and degradative ubiquitination, not with UBC13 and K63. Even granting for argument's sake that any RGLG could build K63 chains given the right E2, that E2 (UBC13) and its cognate-E3 pairing are documented specifically for RGLG1/ RGLG2. Q6YYC5's cognate rice E2 has never been identified, and no chain-linkage assay has been performed on the rice protein.

The practical consequence for GO curation is that the generic term "protein ubiquitination" captures everything the sequence and orthology can support, whereas the linkage-specific "K63-linked ubiquitination" term asserts a mechanistic property (E2-dependent linkage geometry) that the E3 sequence cannot encode. ProtNLM2's refinement is a plausible-sounding but mechanistically unfounded increase in specificity — a textbook case of a language-model annotation being **more precise than the underlying evidence permits**, likely driven by paralog frequency bias (transfer of the well-published RGLG2 K63 result onto a more distant relative).

Evidence Base / Evidence Matrix

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding
UniProt Q6YYC5 + InterPro/ Pfam/ PANTHER (database)	Structural/ evolutionary; computational	Supports (generic); Qualifies (K63)	Does Q6YYC5 have RGLG E3 architecture?	vWA (IPR002035) + C-terminal C3HC4 RING (PF13920, res 356–389); PANTHER SF16 = RGLG4
This work (computed)	Computational (motif)	Supports (E3 activity)	Is the catalytic RING intact?	Canonical C3HC4 cross-brace present: C356,C359,C370,H372,C374/5,C378,C385,C388; regex match
This work (computed)	Structural/ evolutionary	Refutes (K63 orthology)	Which At RGLG is Q6YYC5 closest to?	RGLG4 59.8% > RGLG5 56.6 > RGLG3 56.1 > RGLG2 52.8 > RGLG1 48.5%; NJ tree: Q6YYC5 sister to RGLG4; RGLG1+RGLG2 separate clade
PMID: 17586653	Direct assay + mutant phenotype	Qualifies	Do RGLGs form K63 chains?	RGLG2 forms K63-linked multiubiquitin chains in vitro; rglg1 rglg2 loses apical dominance (auxin/PIN)
PMID: 27497447	Direct assay + interaction	Refutes (K63 for RGLG3/4 clade)	What do RGLG3/ RGLG4 do?	RGLG3/4 + cognate E2 UBC30 ubiquitinate GRXS17 → degradation
PMID: 40169231	Mechanistic/ structural	Refutes (sequence-decidability)	Is linkage set by E3 or E2?	"specificity of Ub linkage is determined by the given E2" for RING E3s
PMID: 17426036	Direct assay	Refutes (sequence-decidability)	Can one RING make different linkages?	MuRF1 makes K63 with UbcH13/Uev1a but K48 with E2-25K
PMID: 20113438	Direct assay + genetics	Refutes (K63 for Q6YYC5)	Which E2/E3 make K63 in	UBC13 (K63-specific E2) has cognate E3s RGLG1/RGLG2 (Fe-deficiency root hairs)

Citation	Evidence type	Supports/ Refutes/ Qualifies	Claim tested	Key finding
			RGLG pathway?	
PMID: 40451499	Direct assay + structural	Refutes (sequence-decidability)	Does one E3 use multiple E2s/ linkages?	Arkadia/Ark2C functionally interact with several E2s (incl. Ubc13) to build different chains
PMID: 23625358	Mutant phenotype	Qualifies	RGLG1/2 substrates	rglg1 rglg2 stabilizes AtERF53 (RGLG1/2 promote its turnover)
PMID: 22898498 / PMID: 23073017 / PMID: 25788731	Mutant phenotype	Qualifies	RGLG3/4 biological role	RGLG3/4 modulate COI1-dependent jasmonate signaling & FB1-triggered PCD
PMID: 41312104	Direct assay	Qualifies (competing linkage)	Rice RGLG chain type	OsRGLG6 ubiquitinates OsOTUB1 for degradation (grain number/yield)

GO Curation Implications

Leads requiring curator verification:

- **GO:0004842 (MF, ubiquitin-protein transferase activity) — RETAIN.** Supported by intact RING + vWA architecture (F001).
- **GO:0016567 (BP, protein ubiquitination) — RETAIN** as the appropriately general term.
- **GO:0070534 (BP, protein K63-linked ubiquitination) — DO NOT ADD** on the basis of the ProtNLM2 prediction. It over-specifies a property (chain linkage) that is E2-determined and not encoded in the E3 sequence (F003); the orthology used to justify it points to RGLG4, the

degradative clade (F002, F004), not the K63-demonstrated RGLG1/2 (F005). Treat as an over-annotation / paralog-frequency-bias artifact.

- **Localization (CC):** Q6YYC5 begins "MGG…" (Gly2), a potential N-myristoylation site as in plasma-membrane–localized RGLG2. This is a lead for a membrane-localization annotation but must be verified independently before use.

The correct curation posture is to keep the annotation at the **generic granularity** the evidence supports and to flag GO:0070534 as an example of computational over-specification.

Mechanistic Scope

The immediate molecular function under test is **E3 ubiquitin-ligase activity** — Q6YYC5 acting as a RING-type scaffold that positions an E2~Ub conjugate and a substrate to transfer ubiquitin. That direct activity is well supported. The **K63 linkage** is a downstream property of the E2 partner and the reaction context, not an intrinsic property of Q6YYC5. The seed hypothesis conflates three distinct levels:

- **Direct gene-product activity:** RING-dependent ubiquitin transfer (supported).
- **E2-contributed reaction property:** chain linkage type (K48/K63/mixed) — belongs to the E2, not the E3 (not supported for K63 here).
- **Downstream/phenotypic outcomes** in relatives: apical dominance, auxin transport, jasmonate signaling, Fe-deficiency root-hair branching, GRXS17 degradation — pathway consequences of specific RGLG members, not evidence that Q6YYC5 itself builds K63 chains.

Curators should record only the direct activity at the granularity the evidence permits.

Conflicts and Alternatives

- **Paralog confusion (primary risk):** ProtNLM2 likely transferred the well-published K63 result from RGLG1/RGLG2, but Q6YYC5 clusters with RGLG4. The RGLG3/4 clade drives substrate **degradation** via UBC30 — inconsistent with obligate K63 (which is typically non-degradative/signaling).
- **Organism difference:** Rice has its own RGLG expansion; the closest *Arabidopsis* match (RGLG4, 59.8%) is a paralog-level, not strict-ortholog, relationship, weakening any

functional transfer. The one characterized rice RGLG (OsRGLG6, PMID: 41312104) is degradative, again pointing away from a K63-signaling default.

- **In-vitro-only caveat:** Even the RGLG2 K63 result is an in vitro chain-formation assay with a chosen E2; it does not license a rice-protein-specific K63 term.
- **Database carry-over risk:** ProtNLM2 is a language-model annotator prone to propagating specific-sounding terms; GO:0070534 here appears to be exactly such an over-specification.

Limitations and Knowledge Gaps

1. **Cognate E2 of Q6YYC5 (unknown).** Checked: no interaction data found. Matters because the E2 sets linkage. Resolve: yeast-two-hybrid / AlphaFold-Multimer E2-panel screen; test K63-specific (OsUBC13/UEV) vs promiscuous vs UBC30-type E2s.
2. **Actual chain product (unknown).** Checked: no biochemical assay exists for Q6YYC5. Resolve: in vitro ubiquitination with ubiquitin lysine-only mutants (K63-only, K48-only, K0) + LC-MS/MS linkage typing.
3. **True rice ortholog assignment.** Checked family identity and a 6-sequence NJ tree only. Resolve: bootstrapped phylogeny of all rice + *Arabidopsis* RGLGs; syntenic ortholog call.
4. **Substrate and degradative vs signaling role.** Resolve: RING-dead substrate-trap TAP-MS in rice.
5. **Membrane localization not experimentally verified.** The Gly2 myristoylation motif is a prediction only. Resolve: subcellular localization of tagged Q6YYC5.

Methodological limitation: pairwise identities were computed with a simple global alignment (match/mismatch scoring), adequate for ranking paralogs but not a substitute for a bootstrapped phylogeny. Domain calls are database-derived (InterPro/Pfam/PANTHER). No experimental data specific to Q6YYC5 were located; conclusions about the K63 term rest on the general RING mechanism plus clade assignment.

Proposed Follow-up Experiments / Discriminating Tests

1. **In vitro ubiquitination with ubiquitin lysine mutants** (K63-only vs K48-only vs K0) + LC-MS/MS linkage typing — directly reads out the chain linkage Q6YYC5 produces. *Most decisive test.*
2. **E2 pairing screen** — determine whether Q6YYC5 productively pairs with a K63-specific UBC13/UEV1 dimer, or with a UBC30-type E2 (predicting degradation).
3. **Substrate stability assays** — if candidate substrates are stabilized in a Q6YYC5 mutant (degradation phenotype), K63 is less likely; if a signaling readout changes without degradation, K63 is more consistent.
4. **Cross-species complementation** — test whether Q6YYC5 rescues *Arabidopsis* rglg3 rglg4 (jasmonate/GRXS17 phenotypes) versus rglg1 rglg2 (apical dominance) to assign functional clade.
5. **Phylogenomic tree** of all rice + *Arabidopsis* RGLGs with bootstrap support to confirm the RGLG3/4 vs RGLG1/2 clade placement.
6. **AlphaFold RING/vWA structural superposition** onto characterized RGLG2 and RGLG4 to confirm catalytic geometry.

Curation Leads (require curator verification)

- **Action:** Keep GO:0016567 / GO:0004842; **reject the ProtNLM2 GO:0070534 addition** as over-specific / computational-only.
- **Candidate reference + snippet (E2 determines linkage):** PMID: 40169231 — *"RING-type E3s mediate the transfer of Ub directly from the E2~Ub conjugate, implying that the specificity of Ub linkage is determined by the given E2."*
- **Candidate reference + snippet (only RGLG2 shown to form K63):** PMID: 17586653 — *"The RING domain protein RGLG2 … can form Lys-63-linked multiubiquitin chains in an in vitro reaction."*
- **Candidate reference + snippet (K63 module = RGLG1/2 + UBC13):** PMID: 20113438 — *"UBC13 has been shown to catalyze non-canonical Lys63-linked ubiquitin chains … Mutations in the cognate E3 ligases RGLG2 and RGLG1 caused the constitutive formation of branched root hairs."*

- **Candidate reference + snippet (Q6YYC5 clade = degradative, UBC30):** PMID: 27497447 — *"we used a ubiquitin-conjugating enzyme (UBC) panel screen to pinpoint UBC30 as a cognate E2 UBC … mediating … ubiquitination of GRXS17 … GRXS17 is ubiquitinated and degraded in an RGLG3- and RGLG4-dependent manner."*
- **Suggested question for curators:** Is there any rice-specific biochemical evidence (E2 partner or chain-type assay) for Q6YYC5? If not, the linkage-specific term should not be propagated.
- **Suggested experiment:** In vitro linkage-typing assay (above) to decide GO:0070534 vs generic GO:0016567.

Provenance Artifacts

Computed provenance saved under `artifacts/`: - `Q6YYC5_RGLG_distance_matrix.csv` — pairwise distances (100 − %id) + Newick NJ tree. - `Q6YYC5_evidence_matrix.csv` — machine-readable evidence table. - `Q6YYC5_GO_decision_table.csv` — GO curation recommendations.

Conclusion

Q6YYC5 is a bona fide RGLG-family RING E3 ubiquitin ligase — it carries the diagnostic vWA domain plus an intact C3HC4 RING (catalytic ligands C356/C359/C370/H372/C374-5/C378/C385/C388) — so its generic annotations GO:0016567 (protein ubiquitination) and GO:0004842 (ubiquitin-protein transferase activity) are well justified. The ProtNLM2 refinement to protein K63-linked ubiquitination (GO:0070534) is **over-annotated** and should not be added: ubiquitin-chain linkage in RING E3s is determined by the cognate E2, not by the E3 sequence, and Q6YYC5's closest *Arabidopsis* relatives are the RGLG3/RGLG4 clade (RGLG4 ~60% identity), which uses E2 UBC30 to drive degradation, rather than the K63-chain-forming RGLG1/RGLG2. No rice-specific E2 or chain-type assay exists, so the K63 term is not decidable from sequence/orthology and requires direct biochemical verification.
