Adaptive Evolution of Signaling Partners

Daisuke Urano,[1] Taoran Dong,[2] Jeffrey L. Bennetzen,[2] and Alan M. Jones*[,1,3]
[1]Department of Biology, University of North Carolina, Chapel Hill
[2]Department of Genetics, University of Georgia
[3]Department of Pharmacology, University of North Carolina, Chapel Hill
***Corresponding author:** E-mail: alan_jones@unc.edu.
Associate editor: Csaba Pal

Abstract

Proteins that interact coevolve their structures. When mutation disrupts the interaction, compensation by the partner occurs to restore interaction otherwise counterselection occurs. We show in this study how a destabilizing mutation in one protein is compensated by a stabilizing mutation in its protein partner and their coevolving path. The pathway in this case and likely a general principle of coevolution is that the compensatory change must tolerate both the original and derived structures with equivalence in function and activity. Evolution of the structure of signaling elements in a network is constrained by specific protein pair interactions, by requisite conformational changes, and by catalytic activity. The heterotrimeric G protein-coupled signaling is a paragon of this protein interaction/function complexity and our deep understanding of this pathway in diverse organisms lends itself to evolutionary study. Regulators of G protein Signaling (RGS) proteins accelerate the intrinsic GTP hydrolysis rate of the Gα subunit of the heterotrimeric G protein complex. An important RGS-contact site is a hydroxyl-bearing residue on the switch I region of Gα subunits in animals and most plants, such as *Arabidopsis*. The exception is the grasses (e.g., rice, maize, sugarcane, millets); these plants have Gα subunits that replaced the critical hydroxyl-bearing threonine with a destabilizing asparagine shown to disrupt interaction between *Arabidopsis* RGS protein (AtRGS1) and the grass Gα subunit. With one known exception (*Setaria italica*), grasses do not encode *RGS* genes. One parsimonious deduction is that the RGS gene was lost in the ancestor to the grasses and then recently acquired horizontally in the lineage *S. italica* from a nongrass monocot. Like all investigated grasses, *S. italica* has the Gα subunit with the destabilizing asparagine residue in the protein interface but, unlike other known grass genomes, still encodes an expressed RGS gene, *SiRGS1*. SiRGS1 accelerates GTP hydrolysis at similar concentration of both Gα subunits containing either the stabilizing (AtGPA1) or destabilizing (RGA1) interface residue. SiRGS1 does not use the hydroxyl-bearing residue on Gα to promote GAP activity and has a larger Gα-interface pocket fitting to the destabilizing Gα. These findings indicate that SiRGS1 adapted to a deleterious mutation on Gα using existing polymorphism in the RGS protein population.

Key words: **signal transduction, molecular adaptation, coevolution.**

Article

Introduction

Functionally paired proteins coevolve to maintain physical and functional interactions (Sandler et al. 2014). When physical interaction is weakened by mutation in one partner, functional interaction is lost unless the other partner adapts by a stabilizing mutation. If this mutation is less fit in the partner encoded by the parent allele, then it is lost. Few examples are known for an adaptive pathway (Skerker et al. 2008; Zamir et al. 2012). We describe here a new example of adaptive molecular evolution which occurs between the Gα subunit of the heterotrimeric G protein complex and its cognate partner, Regulator of G signaling (RGS) protein.

Cytoplasmic heterotrimeric G proteins regulate various cellular responses in animals, fungi, and plants. G proteins are activated by cell surface receptors through promotion of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) exchange on the Gα subunit. GTP-bound Gα returns to its inactive state after hydrolyzing its bound GTP to GDP (Gilman 1987). This hydrolysis is accelerated by RGS (Berman et al. 1996; Chen et al. 1996; Hunt et al. 1996; Watson et al. 1996) or other GTPase-activating proteins (GAP) (Arshavsky and Bownds 1992; Berstein et al. 1992). The interaction and catalytic mechanisms in Gα and RGS proteins originated in an ancestral eukaryote and were inherited by animals, yeasts, amoeba, plants, and some protists (Dohlman et al. 1995; Berman et al. 1996; Chen et al. 2003; Bradford et al. 2013). The Gα–RGS interface is shared within those evolutionary clades (Tesmer et al. 1997; Slep et al. 2001, 2008; Johnston et al. 2007; Soundararajan et al. 2008; Urano et al. 2012; Nance et al. 2013), indicating that the binding-mode was strongly constrained under selective pressure. The core RGS domain conveying GAP activity, the so called the RGS box, is composed of approximately 120 amino acids forming nine α helices (Popov et al. 1997; Tesmer 2009). An atomic structure of rat RGS4 and Gαi1 revealed that the RGS box interacts with the three flexible switch regions of Gα and stabilizes the transition state of GTP hydrolysis by Gα (Berman et al. 1996; Tesmer et al. 1997). The RGS box mainly contacts a hydroxyl-bearing residue in switch I and a strictly conserved lysine in switch II regions (e.g., T_{182} in switch I and K_{210} in switch II of

rat Gαi1) (Tesmer et al. 1997). Site-directed mutation of these residues dramatically reduces RGS binding to the Gα subunits (Wieland et al. 2000; Day et al. 2004). When these mutations occurred in nature (Kozasa et al. 1998; Urano et al. 2012), the RGS-binding abilities were affected in those lineages (Nakamura et al. 2004; Chen et al. 2005; Urano et al. 2012).

Many plant, protist, and fungal RGS genes encode an RGS protein having an amino-terminal 7 transmembrane (TM) domain and a carboxyl-terminal domain with an RGS box, the catalytic domain of RGS proteins (Chen et al. 2003). The 7TM-RGS genes are broadly conserved in vascular plant genomes. However, known grass genomes with the exception of foxtail millet (*Setaria italica*) lack RGS genes (Urano et al. 2012). Grasses have Gα subunits with a Gα mutation of the RGS-contact site from threonine to an asparagine predicted to dramatically reduced the affinity at the Gα–RGS interface (Urano et al. 2012) and it was proposed that this disrupting mutation led to loss of the 7TM-RGS1 gene in grasses. The *S. italica* genome encodes a remnant of the 7TM domain of the RGS gene separated from the RGS box domain by extensive transposon insertions. The remnant 7TM domain of the *S. italica* RGS gene is not expressed and is in the late stage of decay but the remaining RGS domain sequence may encode a functional RGS protein (Ma et al. 2004). If so, it is unclear how the *S. italica RGS* allele remains constrained as the *S. italica* Gα subunit contains the destabilizing asparagine in the Gα–RGS protein–protein interface.

In this study, we traced the evolution of the grass Gα–RGS interface. We biochemically analyzed various RGS activities and found that *S. italica* SiRGS1 exerts GAP activity on plant Gα regardless of whether Gα possesses the stabilizing threonine or the destabilizing asparagine at the RGS contact site. Steric reduction of the SiRGS1 space in the Gα–RGS protein–protein interface, as is found on the *Arabidopsis* RGS1 (AtRGS1) surface, decreased GAP activity only on Gα subunits having the destabilizing asparagine, but not on Gα subunits having the parental threonine. The *S. italica* RGS gene became adapted in the *Setaria* lineage within the last 5 My.

Results and Discussion

Setaria italica SiRGS1 Is Expressed in *S. italica* Leaves

Plant RGS proteins have a 7TM domain as the amino-terminal half and an RGS box located centrally in the cytoplasmic carboxy-terminal half. The prototype is AtRGS1 (Chen et al. 2003). Mined expressed sequence tag (EST) data of *SiRGS1* supported expression of the cytosolic RGS box but not the 7TM region of SiRGS1. However, a gene assembly program (NCBI) predicted six isoforms for the *SiRGS1* gene including both multi-TM region and RGS domains (fig. 1*A* and supplementary fig. S1A, Supplementary Material online). Therefore, we tested this experimentally. To examine expression of the SiRGS1 gene, cDNA from 14-week-old *S. italica* leaves or roots was prepared and used as template for reverse transcribed polymerase chain reaction (PCR). The cytoplasmic region of SiRGS1 was amplified from *S. italica* leaves but not from the roots, whereas *S. italica* Gα (SiGPA1, XM_004963062.1) and the actin 1 (SiAct1, XM_004981913.1) genes were amplified equally well from leaves and roots (fig. 1*B*).

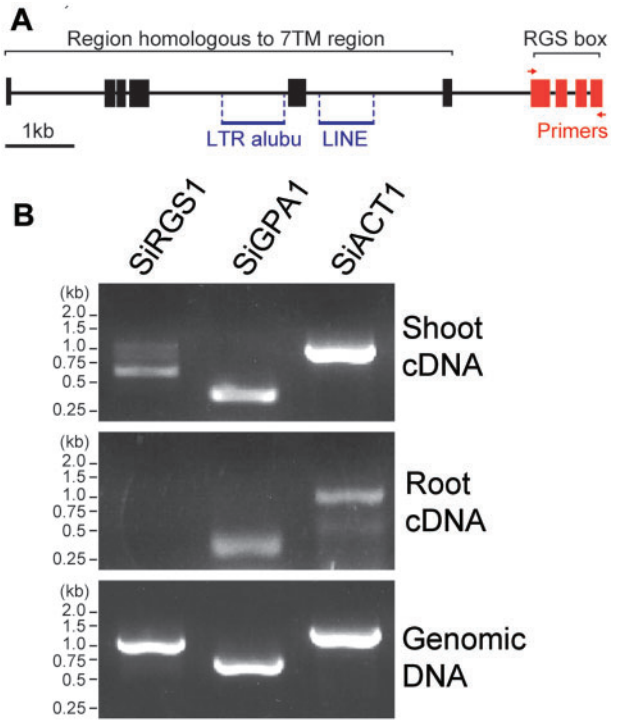

Fig. 1. A multi-TM *Setaria italica SiRGS1* gene. (*A*) SiRGS1 gene structure. Red or black boxes show DNA regions homologous to *AtRGS1* with or without experimental support of EST transcription data. Two transposon DNAs, LTR retrotransposon alubu and LINE, are indicated with blue lines. Red arrows indicate the location of the primers used to amplify the *SiRGS1* gene. Supplementary figure S1, Supplementary Material online, shows the predicted protein products of the *SiRGS1* locus. (*B*) Expression of *SiRGS1*, *SiGPA1*, and *SiACT1* genes in *S. italica* leaves and roots. The indicated cDNA or genomic DNA was prepared from 14-week-old *S. italica* leaves or roots. Using RT-PCR as described in Materials and Methods, the presence of mRNA encoding the cytoplasmic RGS box of SiRGS1 was compared with the presence of RNA encoding SiGPA1 or actin (SiACT1) was tested.

Setaria italica SiRGS1 Compensated for the Threonine to Asparagine Mutation on Gα

We next examined whether SiRGS1 exerts GAP activity and whether it has substrate specificity. Grass Gα proteins, as well as SiGPA1, possess asparagine at the RGS-contact site, whereas other plant Gα proteins have threonine (fig. 2*A*). We purified the recombinant SiRGS1 protein (1–207 aa) with a polyhistidine-tag from *Escherichia coli* and evaluated its GAP activity. We first measured the single-turnover rate of GTP hydrolysis by *Arabidopsis* and rice Gα proteins (AtGPA1 and RGA1; supplementary table S1, Supplementary Material online). SiRGS1 unexpectedly promoted GTP hydrolysis by AtGPA1 and RGA1 at similar effective ranges (half-maximal stimulation [SC_{50}] of 82 and 84 nM, respectively) (fig. 2*B* and *C* and supplementary fig. S2, Supplementary Material online), implying that SiRGS1 exerts its GAP activity without contacting the hydroxyl-bearing residue preserved on the Gα subunit

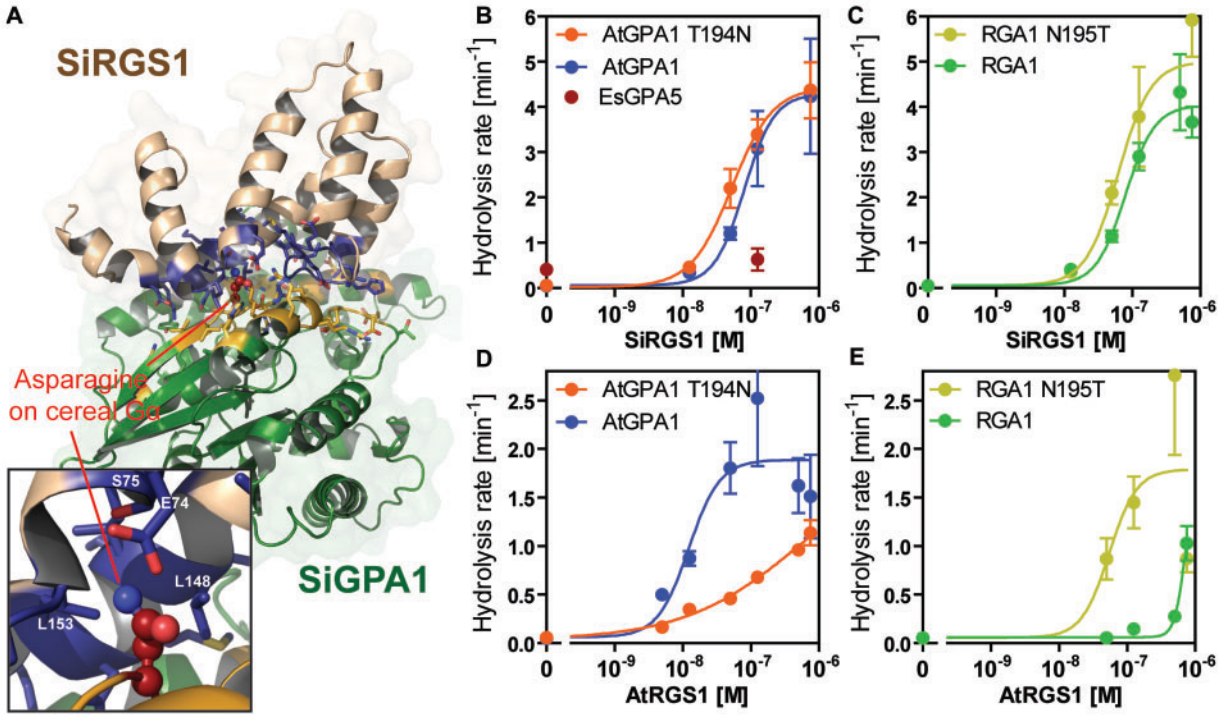

Fig. 2. In vitro GAP activity of SiRGS1 protein. (*A*) Modeled-structure of *Setaria italica* Gα–RGS complex. Gα and SiRGS1 are shown in green and light brown, respectively. The contact residues are shown with side chains in yellow (SiGPA1) and blue (SiRGS1). A close-up view shows Asn195 of SiGPA1 and the contact residues. The Asn195-contacting residues are labeled. (*B*, *C*) GAP activity of SiRGS1 on AtGPA1, AtGPA1-T194N, RGA1, and RGA1-N195T proteins. Rates for GTP hydrolysis by 200 nM Gα proteins were determined using the One-Phase association model in Prism 5.0 software with single-turnover kinetics data shown in supplementary figure S2, Supplementary Material online. The catalytic rate constant of hydrolysis (kcat, min^{-1}) was plotted against concentrations of SiRGS1 with standard errors obtained from seven or more data points of at least two experiments. (*D*, *E*) GAP activity of *Arabidopsis thaliana* AtRGS1. The dose–response curves were calculated using kinetics data published previously (Urano et al. 2012). The hydrolysis rates are shown in supplementary table S1, Supplementary Material online.

or that this hydroxyl no longer contributes to the binding affinity. To verify this, we swapped the RGS-contact residue on the *Arabidopsis* Gα subunit from threonine to the grass asparagine (AtGPA1-T194N) and on the grass (rice) from asparagine to the typical threonine (RGA1-N195T). The SiRGS1 protein promoted GTP hydrolysis on the mutated proteins, as well as the wild-type parent proteins and a distant Gα subunit from the brown alga, *Ectocarpus siliculosus*, albeit more weakly (fig. 2*B* and *C* and supplementary fig. S2, Supplementary Material online). The lack of selectivity by SiRGS1 contrasts with AtRGS1 GAP activity which is dependent on the threonine contact residue (Urano et al. 2012). The selectivity of AtRGS1 with its cognate AtGPA1 compared with rice RGA1 was approximately 100-fold (AtGPA1 SC_{50} = ~13 nM vs. RGA1 SC_{50} > 1 μM; fig. 2*D* and *E*). The T194N mutation on AtGPA1 decreased GAP activity and affinity to AtRGS1, whereas the N195T mutation on RGA1 increased those. These results indicate that in *Arabidopsis* the Gα–RGS pair utilizes the same binding mode as in animals whereas SiRGS1 evolved to compensate for the deleterious mutation that occurred on the binding partner, the Gα subunit, but accomplished this without losing catalytic activity on the parent. In other words, evolution followed a path that compensated both for the original and evolved residues.

Setaria italica SiRGS1 Has a Larger Pocket on the Gα-Binding Surface

Mammalian RGS proteins bind to Gα through three switch regions on Gα (Tesmer et al. 1997), in which switch I and switch II are directly involved in GTP hydrolysis (Coleman et al. 1994; Sondek et al. 1994). A modeled SiRGS1 structure showed only one distinct residue in their Gα-binding pocket; S_{75} on SiRGS1 and T_{321} on AtRGS1 (fig. 3*A*–*C*), whereas both SiRGS1 and AtRGS1 conserved a glutamate, (E_{74} on SiRGS1 and E_{320} on AtRGS1), which contacts a lysine on switch II (e.g., K210 of rat Gαi1) (Tesmer et al. 1997). To analyze the mode of SiRGS1 action, we examined the GAP activity of two mutant proteins, SiRGS1-S75T and SiRGS1-E74K (fig. 3*D* and supplementary fig. S2, Supplementary Material online). SiRGS1-S75T (125 nM) promoted GTP hydrolysis by AtGPA1 (K_{cat} = 2.1 min^{-1}), whereas the S75T mutation reduced the GAP activity on RGA1 (K_{cat} = 0.5 min^{-1}), suggesting that the hydroxyl-bearing residues differing by a single methyl group determine the selectivity of Gα. A surface plasmon resonance (SPR) analysis suggested that the selectivity was due to a change in the RGS-Gα affinity, as SiRGS1-S75T had high affinity for AtGPA1 (K_D = 21.9 nM) whereas an affinity for RGA1 was not calculable (table 1 and supplementary fig. S3, Supplementary Material online). The additional methyl

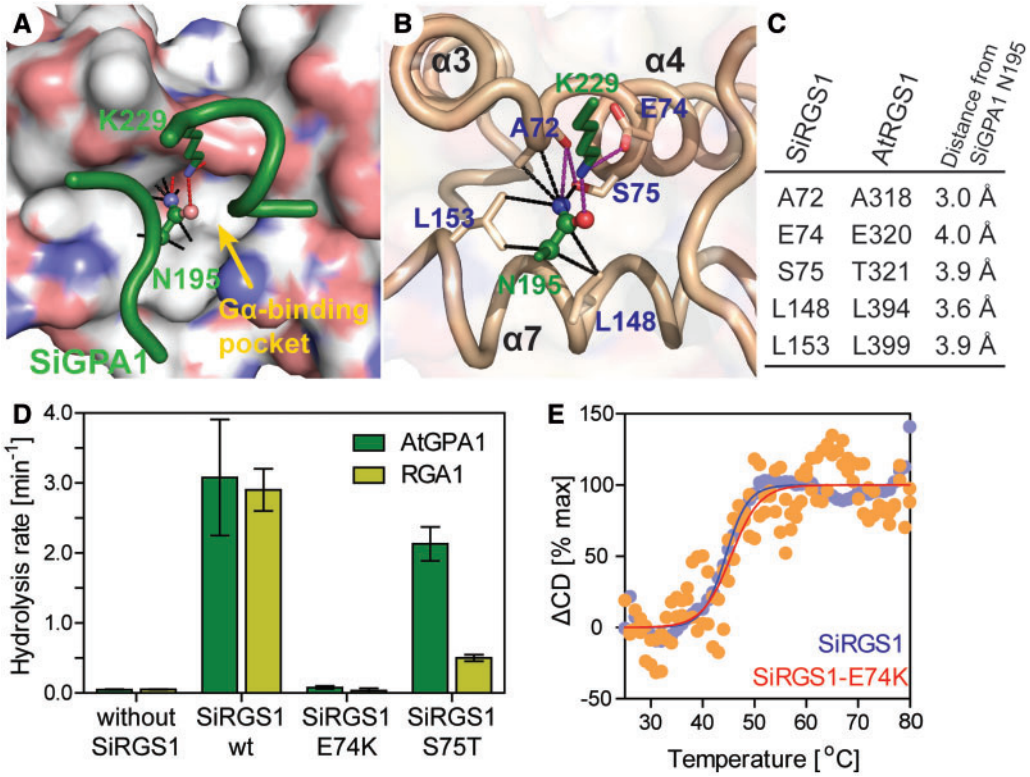

Fig. 3. Mutational analyses of SiRGS1. (*A*, *B*) A modeled structure of SiRGS1 docked with SiGPA-Asn195 and -Lys229 residues (green). (*A*) Nitrogen and oxygen on the SiRGS1 surface are shown in blue and red. A yellow arrow indicates the Gα-binding pivot of SiRGS1 that docks the Asn195 of SiGPA1. Black dot lines connect a side chain of SiGPA1–Asn195 with SiRGS1 atoms within 4 Å. Purple lines indicate potential hydrogen bonds. (*C*) List of SiRGS1 residues located within 4 Å of a side chain of SiGPA1–Asn195. Corresponding residues of AtRGS1 and distance from SiGPA1–Asn195 are shown. (*D*) Hydrolysis rates of AtGPA1 and RGA1 with 125 nM SiRGS1 or SiRGS1 S75T, or 500 nM SiRGS1 E74K mutant protein. The hydrolysis rates and standard error of mean were estimated by single-turnover [^{32}P]-GTP hydrolysis assays as shown in supplementary figure S2, Supplementary Material online. (*E*) Thermal stabilities of SiRGS1 and the SiRGS1-E74K mutant were determined from the change in CD values at 222 nm. The fraction of unfolded protein (ΔCD) was calculated as described in Materials and Methods.

Table 1. Gα-Binding and GAP Activity of *Arabidopsis thaliana* and *Setaria italica* RGS proteins.

	GTP Hydrolysis Rate (min^{-1})		Dissociation Constant (M)	
	RGA	AtGPA1	RGA	AtGPA1
AtRGS1	0.15 ± 0.01	2.52 ± 0.70	5.67×10^{-8}	1.74×10^{-8}
SiRGS1	2.90 ± 0.30	3.08 ± 0.83	—	—
SiRGS1-S75T	0.50 ± 0.05	2.13 ± 0.24	No interaction observed	2.19×10^{-8}

Note.—GTP hydrolysis rate (kcat) values were obtained from a previously published data (Urano et al. 2012) and single turnover assays (supplementary fig. S2, Supplementary Material online). Dissociation constant (K_D) values were obtained from a previous publication (Urano et al. 2012) and an SPR experiment (supplementary fig. S4, Supplementary Material online).

group potentially makes the Gα-binding pocket smaller and interferes with the asparagine in grass Gα docking in the pocket. SiRGS1-E74K protein (500 nM) lacked GTP hydrolysis activity against both AtGPA1 ($K_{cat} = 0.079\ min^{-1}$) and RGA1 ($K_{cat} = 0.034\ min^{-1}$), as was shown in *Arabidopsis* AtRGS1-E320K and mouse RGS16-E89K mutant analyses (Wieland et al. 2000; Johnston et al. 2007). Circular dichroism (CD) analyses indicated that the E74K mutation did not affect thermal stability, an intrinsic protein property (fig. 3*E* and supplementary fig. S3, Supplementary Material online). The melting temperature of SiRGS1 and SiRGS1-E74K was 44.2 or 45.5 °C, respectively. This biochemical evidence reflects evolutionary pressure for the RGS surface residues to be constrained by its partner Gα (supplementary fig. S1*B*, Supplementary Material online). Specifically, E_{74} on SiRGS1 was fixed by the switch II region on Gα since the base of the eukaryotic tree, whereas S_{75} on SiRGS1 was recently fixed and constrained by the asparagine mutation in grass Gα subunits.

The *SiRGS1* Gene Originated in Monocots

Gene expression and biochemical analyses revealed the functionally intact *S. italica* SiRGS1, whereas none of the other grasses possessed an RGS homolog (supplementary table S2, part A, Supplementary Material online). To find any remnant

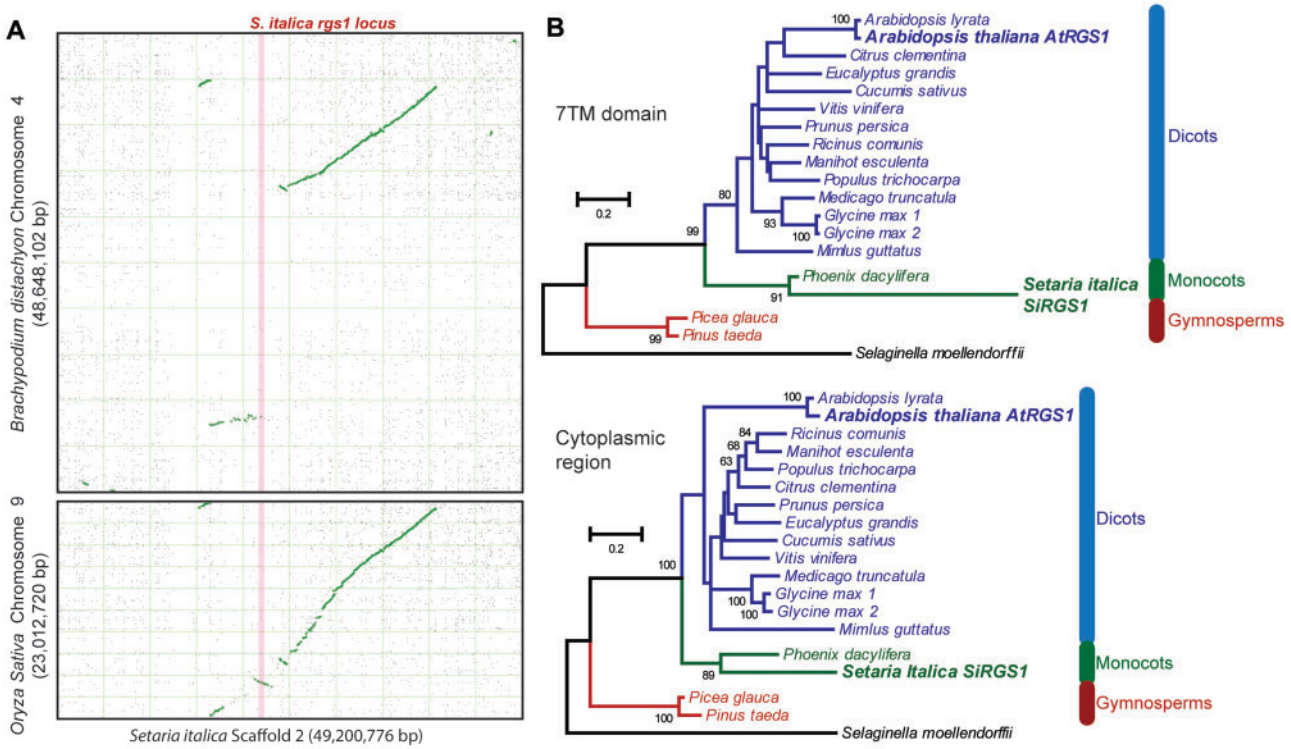

Fig. 4. Sequence analyses of *Setaria italica* RGS1. (*A*) Syntenic dot plots of *Brachypodium distachyon* chromosome 4 or *Oryza sativa* chromosome 9 (*y* axis) against *Setaria italica* chromosome 2 (*x* axis). The syntetic dot plots were generated by the SynMAP online tool at the CoGe server (https://genomevolution.org/CoGe/). Whole genomes were compared with the LastZ algorithm. Green or gray dots represent syntenic pairs or nonsyntenic matches, respectively. A light pink line represents the *S. italica rgs1* and its upstream and downstream genes (see supplementary table S2, Supplementary Material online). Note: The synteny around the *SiRGS1* locus is poor between *S. italica* and the closer relatives, *Sorghum bicolor* and *Zea mays*. (*B*) Maximum-likelihood trees of 7TM region and the RGS domain of plant 7TM-RGS proteins. Protein sequences homologous to AtRGS1 were aligned with the ClustalW algorithm. Sites having 30% or more gaps were removed from phylogenetic analyses. Maximum-likelihood trees of the 7TM region (1–248 aa of AtRGS1) and the cytoplasmic region (249–459 aa of AtRGS1) were created with JTT + F substitution model with gamma distributed rate variation across sites and analyzed by bootstrap of 100 replicates. The trees are shown with bootstrap values greater than 50. Species names in blue, green, red or black color represent sequences from dicot, monocot, gymnosperms or a spikemoss, respectively. *Arabidopsis thaliana* (AtRGS1) and *S. italica* (SiRGS1) proteins are in bold font and enlarged.

RGS sequence in other grasses, we queried grass genomes for homologous sequences using *SiRGS1* and neighboring genes (supplementary table S2, part B, Supplementary Material online) and we analyzed synteny of the genomic region (fig. 4*A*). The *SiRGS1*-neighboring genes all had the closest homologs in grasses; however, no sequence homologous to the *SiRGS1* gene was found. Most of the neighboring genes, including the ones nearest *SiRGS1*, have the closest homologs in the *Setaria* relatives, *Sorghum bicolor* or *Panicum virgatum*, indicating that the neighboring genes were vertically descended rather than horizontally transferred from other species. The synteny maps between *S. italica* and *Oryza sativa* or *Brachypodium distachyon* (purple false brome) showed synteny blocks of the *SiRGS1*-neighboring genes near a synteny breakpoint (fig. 4*A*). However, this genomic region of *S. italica* had almost no collinearity with two closer relatives; *So. bicolor* or *Zea mays*, implying that the locus was generally unstable in grass genomes. A phylogeny of plant 7TM-RGS proteins placed the SiRGS1 branch next to the *Phoenix dactylifera* (fig. 4*B*), suggesting the SiRGS1 origin in monocots rather than transferred from outside the clade.

Timeline of the Adaptive Coevolution of *S. italica* SiRGS1 Protein

To estimate when Gα mutated the threonine to a destabilizing asparagine and when SiRGS restored its function in grasses, we surveyed Gα subunit and *RGS* genes along the monocot phylogeny (fig. 5*A*). The monocot ancestor possessed the threonine Gα and 7TM-RGS genes, and so do other vascular plants (see event (1) in fig. 5*B*). There are two scenarios for a timeline.

Scenario 1

All grasses have the asparagine Gα, thus the ancestral grass Gα subunit fixed the threonine to asparagine mutation prior to 40 Ma (Sanderson 2002). Before the Gα T → N event, the Gα-binding site on 7TM-RGS (relevant to T_{321} on AtRGS1) was flexible in allowing serine, threonine, or asparagine (Urano et al. 2012) (fig. 5*A* and supplementary fig. S1*B*, Supplementary Material online). We have shown that RGS1 proteins with the stabilizing residue, AtRGS1 and the SiRGS1-S75T mutant, have weak GAP activity on grass Gα subunits (figs. 2*C* and 3*D*). The *SiRGS1* locus and the biochemical

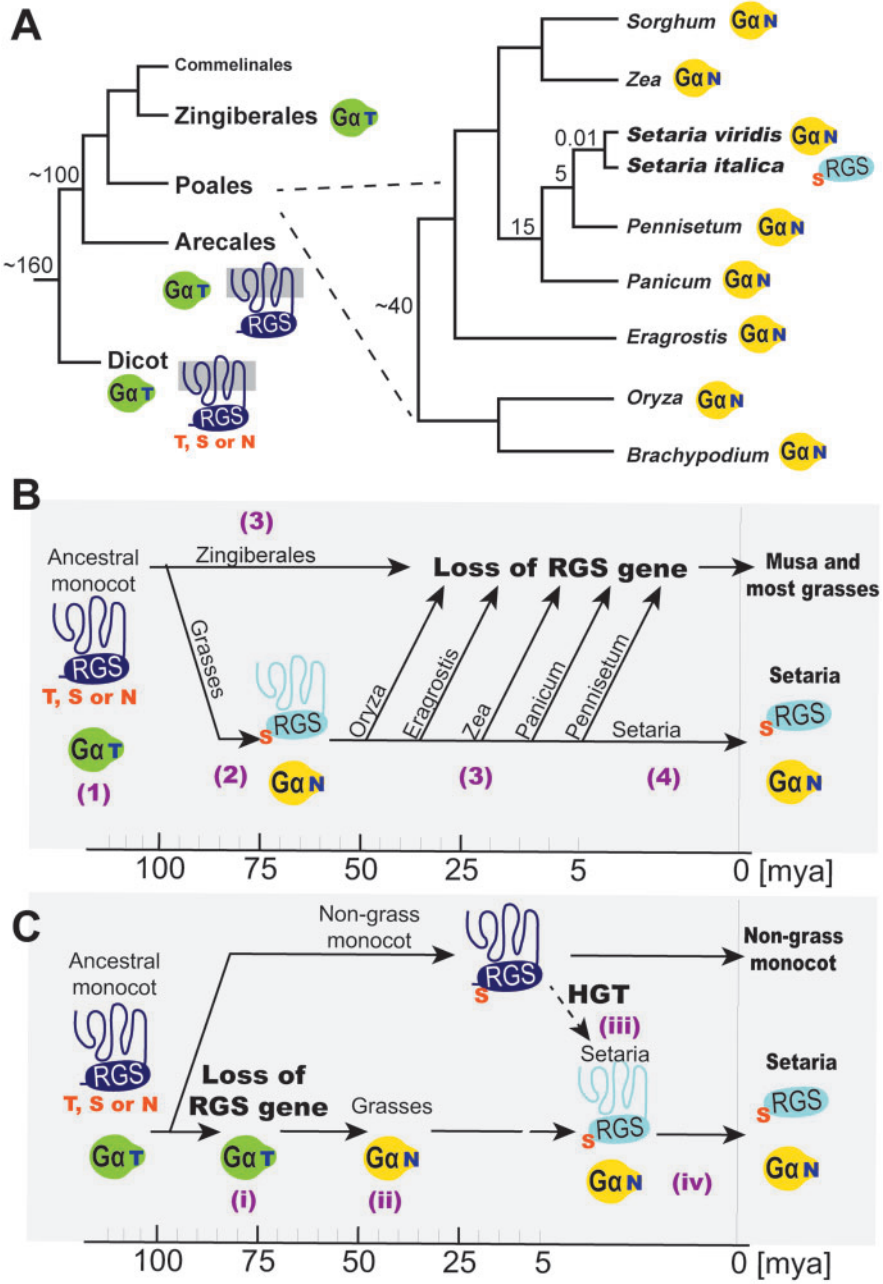

Fig. 5. Fixation of the *Setaria italica* SiRGS1 protein. (*A*) Phylogeny of grasses and their relatives is shown with estimated time of separation (Ma). Green Gα-T or yellow Gα-N represents Gα having threonine or asparagine at the RGS contact site. Dark blue 7TM-RGS or light blue RGS shows RGS with or without 7TM helices, respectively. The Gα-binding residue in dicot 7TM-RGSs is not fixed and varies between threonine (T), serine (S), and asparagine (N). Grasses, except the two *Setaria* species indicated, lack the RGS gene. (*B*) Time line of evolutionary events that have occurred on Gα and RGS genes in monocots. Key events are noted by a number in parentheses: (1) The ancestral monocot possessed the threonine Gα (green Gα-T) and 7TM-RGS. The Gα-binding residue of the 7TM-RGS varied between T, S, and N (dark blue 7TM-RGS). (2) Gα mutated the RGS-contact threonine with asparagine (yellow Gα-N). The mutation caused fixation of the Gα-contact site to serine (light blue 7TM-RGS). These events occurred between 100 and 40 Ma. (3) The *7TM-RGS* gene was deleted independently from most grass genomes. The Zingiberales *Musa acuminata* also lacks the RGS genes, implying that the gene loss event occurred regardless of whether the partner Gα had threonine or asparagine. (4) The *Setaria* RGS gene lost a functional 7TM region within the last 5 My. (*C*) A model for horizontal transfer of the *Setaria* RGS gene. Key events are noted by Latin numbers. (i) A common ancestor of Zingiberales and Poales lost the *7TM-RGS* gene. (ii) Grass Gα mutated the RGS-interacting site to asparagine. (iii) The *Setaria RGS* gene was acquired by horizontal gene transfer (HGT) from an unknown monocot in the past 5 Ma. (iv) The RGS domain, but not the 7TM domain, was fixed in the *Setaria* genome.

property of the SiRGS1 protein address how grass *RGS* genes evolved. When the Gα T → N mutation occurred, the Gα pocket on the 7TM-RGS protein had serine to accommodate the larger side chain of asparagine on Gα, then became fixed (See event 2 in fig. 5*B*). Grass RGS genes were independently lost from grass genomes at least six times after the Gα T → N event. Other RGS mutations may also have positively occurred to adapt the Gα T → N mutation. If so, the

comparison between *Setaria* and its close sister genus *Pennisetum* indicated that the RGS gene was lost as recently as 5 Ma (Bennetzen et al. 2012). This implies that the grass RGS gene probably had some function, as an unselected (i.e., unused) grass gene will become a psuedogene by genetic drift within less than a million years and largely removed from the genome within less than 5 My (Ma et al. 2004). The *Setaria RGS* locus includes 7TM and RGS sequences separated by two transposons (Urano et al. 2012), a snapshot of the gradual loss of an *RGS* gene. Grass *RGS* genes were under purifying selection and were lost independently in sublineages, but the *Setaria* RGS protein coevolved with the *Setaria* Gα (N variant) subunit to restore strong coupling that became functionally fixed in the genome less than 5 Ma.

Scenario 2

The unique presence of the RGS gene in the genus *Setaria*, compared with its absence in all other investigated grass lineages, implies frequent loss of this gene during the evolutionary descent of the Poaceae. At least six independent losses of the RGS gene would need to have occurred to generate the phylogenetic tree in figure 5*A*. Alternatively, in scenario 2, a very recent gain occurred specifically in the lineage leading to *Setaria* (fig. 5*C*). Horizontal transfer of a gene between species is expected to be rare, but well-documented cases were identified such as in the grass genus *Alloteropsis* (Christin et al. 2012). Although the current array of sequenced plant genomes does not provide the power to resolve between repeated loss or horizontal gain scenarios, we did observe that the most closely related *RGS* gene to the one in *Setaria* is from the monocot *P. dactylifera* (fig. 4*B*), a result compatible both with vertical transmission and with horizontal acquisition from a nongrass monocot.

Both the loss of 7TM-RGS genes and the evolution of the cytoplasmic *S. italica RGS* allele in the grasses have two significant implications. The first implication is that the loss of RGS proteins in a lineage that contains Gα subunits that spontaneously exchange GDP for GTP (i.e., self-activating) means that grasses, and likely all plants, have functionally redundant regulatory mechanisms to control the activation state of self-activating G proteins. Loss of RGS1-regulation through gene loss without counterselection by a major loss in fitness means that there are compensatory systems for G protein regulation. For example, GDP dissociation inhibitors (GDI) in animal cells keep the G protein in its inactive state but do so differently than RGS proteins. Plants do not have canonical GDI but we do not rule out the existence of divergent proteins that serve the GDI function. The Zingiberales genome, *Musa acuminata* (wild banana), also lacks an *RGS* gene (fig. 5*A* and supplementary table S2, part A, Supplementary Material online). This implies that the alternative GAP or GDI system emerged prior to 100 Ma and facilitated the loss of RGS loci many times independently.

The second implication is that the retention of a cytoplasmic RGS protein in *S. italica* means that this RGS protein serves a function much different than its 7TM-RGS protein counterpart in dicots, protists, and fungi. Indeed, the evolution to an extant, cytoplasmic RGS protein in *S. italica* may share steps with the evolution of GPCRs and cytoplasmic RGS proteins in animals although extensive sequencing of grass genomes for additional examples of cytoplasmic RGS proteins is needed to support this speculation. Because some genomes encode both GPCRs and 7TM-RGS genes (Bradford et al. 2013), we further speculate that splitting the ancestral *7TM-RGS* gene into genes encoding functional 7TM proteins and cytoplasmic RGS proteins may have set the stage for GPCR expansion. Splitting the 7TM genes into 7TM domains and cytoplasmic RGS proteins may have eliminated constraint on evolution and separate evolutionary trajectories. We speculate that this allowed Gα subunits to coevolve into extant receptor-activated/G protein pairs while the resulting cytoplasmic RGS proteins evolved to provide kinetic scaffolding (Yi et al. 2003; Zhong et al. 2003) and effector antagonism (Ross and Wilkie 2000; Willars 2006).

This discovery also has technological significance for future crop improvements. Plant Gα subunits have intrinsically slow GTP hydrolysis and it is this property that is regulated to control plant G protein activity (Johnston et al. 2007; Jones et al. 2011; Urano et al. 2012). SiRGS1 is a universal plant G protein regulator, and notably the sole GAP that effectively controls grass Gα subunits such as RGA1. Grass Gα subunits have critical functions in many physiological and developmental events (Urano et al. 2013), evident in the Gα-null rice *dwarf1* and maize *compact plants 2*) mutants having a short stature and abnormal fruit formation in addition to numerous physiological differences (Ashikari et al. 1999; Fujisawa et al. 1999; Bommert et al. 2013). The discovery of a grass GAP raises the new possibility to manipulate Gα activity and growth of the most widely eaten crops, namely cereals, by engineering chimeric receptor GAPs using the SiRGS1 protein.

Materials and Methods

Homology Modeling of the SiGPA1–SiRGS1 Complex

Protein sequences of SiGPA1 (sequence ID: Si022288m) and SiRGS1 (Si031119m) were obtained from the Phytozome database. Atomic structures of *Arabidopsis* GPA1 and mammalian Gα–RGS pairs were obtained from the Protein Data Bank (PDB) and used as template structures. Their PDB identifications are 2XTZ, 4EKD, 2IK8, 2ODE, 1FQK, and 2IHB. The *Arabidopsis* Gα structure was solved in its active GTPγS-bound form, whereas the mammalian Gα–RGS structures were in their hydrolysis transition state that contains a GDP, a magnesium ion (Mg^{2+}), and an aluminum fluoride (ALF). We used the GDP, Mg^{2+}, and ALF to model the *Setaria* Gα–RGS complex. Homology modeling was performed by the Modeller 9.14 program (Eswar et al. 2007). The template structures were structurally aligned with the structural alignment function in the Modeller program, and then aligned with SiGPA1 (residues from His 38 to Arg 377) and SiRGS1 (residues from Lys 52 to Lys 167) sequences. Five Gα–RGS complex structures were modeled using the *automodel* class with the default settings except molecular dynamics refinement speed of very slow. The modeled structures were assessed by the discrete optimized protein energy (DOPE) score.

The modeled structures having the lowest DOPE value were used for further analyses. The PyMOL v1.7.2 program was used for identifying potential hydrogen bonds, measuring distance between atoms, and making final images.

Phylogenetic Analyses

Protein sequences homologous to AtRGS1 were obtained as described previously (Urano et al. 2012), and aligned with the ClustalW algorism implemented in the MEGA 6.0 software (Tamura et al. 2013). The aligned sequences were separated into the 7TM region (1–248 aa of AtRGS1) and the cytoplasmic region (249–459 aa of AtRGS1). Maximum-likelihood trees of the two regions were created with JTT + F substitution model with gamma-distributed rate variation across sites, and then analyzed by bootstrap of 100 replicates. Sites having 30% or more gaps were removed from phylogenetic analyses.

Plasmids and Proteins

The cDNA encoding SiRGS1 protein (1–207 aa) was synthesized with optimization of codon usage to *E. coli*, amplified by PCR and cloned into pDEST17, which expresses N-terminal 6xHis-tagged protein in *E. coli*. The SiRGS1-E74K and SiRGS1-S75T mutants were made by the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies). Recombinant proteins were expressed and purified as described (Urano et al. 2012) with a slight modification. SiRGS1 proteins were solubilized in buffer A (50 mM Tris–HCl [pH 7.5], 100 mM NaCl, 5 mM 2-mercaptoethanol [β-ME], 1 mM PMSF, and 1 μg/ml leupeptin) with 0.25 mg/ml lysozyme and 0.2% NP-40 and affinity-purified with TALON Metal Affinity Resin (Clonetech). The resin was washed with buffer A containing 500 mM NaCl and 2 mM imidazole, and eluted with buffer A including 500 mM or 1 M imidazole. Imidazole (5mM) was included to crude extracts of *E. coli* with the aim of reducing nonspecific binding. The purified proteins were dialyzed in 12.5 mM Tris–HCl (pH 7.5), 25 mM NaCl, 10 mM β-ME, 1 mM PMSF.

Expression of SiRGS1 In Planta

Setaria italica plants were grown on soil in 4-inch square pots in a greenhouse for 14 weeks. RNAs were extracted from their roots and leaves using RNeasy Mini Kit (Qiagen), then cDNA was synthesized by SuperScript III reverse transcriptase (Invitrogen). The expression of SiRGS1, a potential 7TM region of SiRGS1, SiGPA1, and SiAct1 were examined by RT-PCR using a Taq-polymerase. Primers used were; SiRGS1-7TM fw: TTGATCAGTTGCCGCATTGTACAG, SiRGS1-7TM rv: CTACTGCTATTGTTGAAACAATAATTGC, SiRGS1-RGS box fw: ATGTCTATTTCTCAACCTCTG, SiRGS1-RGS box rv: ATTACAAGTCGCAGCGTGTT, SiGPA1 fw: AGTACGATCAGATGTTATTTGA GG, SiGPA1 full rv: TCAGGTTCCTTCTCTGGAGCG, SiAct1 fw: GCAGAAAGACGCCTACGTCG, and SiAct1 rv: AAGCATTTCCTGTGCACGAT. PCR condition consisted of 40 cycles of 94 °C for 20 s, 60 °C for 10 s, and 72 °C for 2 min. The *S. italica* genomic DNA was extracted from the 14-week-old leaf and included in PCR analyses as a control.

Single-Turn GTP Hydrolysis

Gα subunits (200 nM) were preloaded with radioactive [γ-^{32}P]GTP in Buffer B (50 mM Tris–HCl [pH 7.5], 10 mM $MgCl_2$, 10 mM β-ME, and 0.05% lubrol-PX) for 10 min on ice. The hydrolysis reaction was then started by adding 225 μl of Buffer B + 400 μM GTPγS into 600 μl of preloaded Gα. At given time points, duplicate 100 μl aliquots were taken into 1 ml of charcoal (25% [w/v] in 50 mM phosphoric acid [pH 2.0]) to remove nonhydrolyzed [γ-^{32}P]GTP and proteins. The tubes with charcoal were centrifuged, and the amount of $^{32}PO_4$ hydrolyzed was measured by scintillation counting of the centrifuged supernatants. The data were analyzed by a one-phase association model implemented in Graphpad Prizm 5.0 program. Hydrolysis data with AtRGS1 or SiRGS1 in figure 3 were taken from figure 2.

Thermal Stability of SiRGS1 Protein

CD spectra were generated with a Chirascan-plus CD Spectrometer (Applied Photophysics) to evaluate thermal stability of these proteins. The spectra were collected using 200 μg/ml SiRGS1 or SiRGS1-E74K proteins in 400 μl of 10 mM sodium phosphate (pH 7.0). The CD value at 208 nm was recorded at every 1 °C increment from 25 to 80 °C and smoothed over five data points. The fraction of unfolded proteins at temperature T (F_T) was determined using the minimum CD (CD_{min}) and maximum CD (CD_{max}) values; $F_T = (CD_T - CD_{min})/(CD_{max} - CD_{min})$. The F_T data were fitted by a four-parameter logistic function. The melting temperature was defined as the temperature at which half of the protein was denatured.

Surface Plasmon Resonance

Affinity between Gα and RGS proteins was measured by a BIAcore 3000 analyzer (GE Healthcare) as we performed previously (Urano et al. 2012). His-tagged SiRGS1-S75T protein was immobilized on sensor chip CM5 with ammine coupling. Temperature, flow rate or running buffer were 25 °C, 10 μl/min, or 10 mM Hepes, 150 mM NaCl, 3 mM ethylenediaminetetraacetic acid, 0.005% Tween-20, 100 μM GDP and 10 mM $MgCl_2$, respectively. Six different concentrations (6.25, 12.5, 25, 50, 100, and 200 nM) of His-AtGPA1 or RGA1 proteins were prepared in running buffer with 20 mM NaF and 100 μM $AlCl_3$ and flowed onto the sensor chip for 3 min. Dissociation was monitored for 5 min. The dissociation constant (K_d) value was obtained by fitting the original sensorgrams with a 1:1 Langmuir binding model.

Supplementary Material

Supplementary tables S1 and S2 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Dr Rajeev Varshney for providing pearl millet DNA sequence information, Dr Ashutosh Tripathy, Dr Jing Xi, Melissa Mathews and Ariko Urano for their technical assistance, and Dr Brenda Temple for modeling the

structures. This work was supported by the National Institute of General Medical Sciences (R01GM065989), and the National Science Foundation (MCB-0718202) to A.M.J. The Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy through the grant DE-FG02-05er15671 to A.M.J. funded technical support in this study. This research was partly supported by funding from the BioEnergy Science Center, a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science, to J. L. B. This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

References

Arshavsky V, Bownds MD. 1992. Regulation of deactivation of photoreceptor G protein by its target enzyme and cGMP. *Nature* 357: 416–417.

Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A. 1999. Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the alpha-subunit of GTP-binding protein. *Proc Natl Acad Sci U S A.* 96: 10284–10289.

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. 2012. Reference genome sequence of the model plant Setaria. *Nat Biotechnol.* 30: 555–561.

Berman DM, Wilkie TM, Gilman AG. 1996. GAIP and RGS4 are GTPase-activating proteins for the Gi subfamily of G protein alpha subunits. *Cell* 86:445–452.

Berstein G, Blank JL, Jhon DY, Exton JH, Rhee SG, Ross EM. 1992. Phospholipase C-beta 1 is a GTPase-activating protein for Gq/11, its physiologic regulator. *Cell* 70:411–418.

Bommert P, Je BI, Goldshmidt A, Jackson D. 2013. The maize Galpha gene COMPACT PLANT2 functions in CLAVATA signalling to control shoot meristem size. *Nature* 502:555–558.

Bradford W, Buckholz A, Morton J, Price C, Jones AM, Urano D. 2013. Eukaryotic G protein signaling evolved to require G protein-coupled receptors for activation. *Sci Signal.* 6:ra37.

Chen CK, Wieland T, Simon MI. 1996. RGS-r, a retinal specific RGS protein, binds an intermediate conformation of transducin and enhances recycling. *Proc Natl Acad Sci U S A.* 93:12885–12889.

Chen JG, Willard FS, Huang J, Liang J, Chasse SA, Jones AM, Siderovski DP. 2003. A seven-transmembrane RGS protein that modulates plant cell proliferation. *Science* 301:1728–1731.

Chen Z, Singer WD, Sternweis PC, Sprang SR. 2005. Structure of the p115RhoGEF rgRGS domain-Galpha13/i1 chimera complex suggests convergent evolution of a GTPase activator. *Nat Struct Mol Biol.* 12: 191–197.

Christin P, Edwards E, Besnard G, Boxall S, Gregory R, Kellogg E, Hartwell J, Osborne C. 2012. Adaptive evolution of C(4) photosynthesis through recurrent lateral gene transfer. *Curr Biol.* 22:445–449.

Coleman DE, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR. 1994. Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis. *Science* 265:1405–1412.

Day PW, Tesmer JJ, Sterne-Marr R, Freeman LC, Benovic JL, Wedegaertner PB. 2004. Characterization of the GRK2 binding site of Galphaq. *J Biol Chem.* 279:53643–53652.

Dohlman HG, Apaniesk D, Chen Y, Song J, Nusskern D. 1995. Inhibition of G-protein signaling by dominant gain-of-function mutations in Sst2p, a pheromone desensitization factor in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 15:3635–3643.

Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. 2007. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci.* Chapter 2:Unit 2.9.

Fujisawa Y, Kato T, Ohki S, Ishikawa A, Kitano H, Sasaki T, Asahi T, Iwasaki Y. 1999. Suppression of the heterotrimeric G protein causes abnormal morphology, including dwarfism, in rice. *Proc Natl Acad Sci U S A.* 96:7575–7580.

Gilman AG. 1987. G proteins: transducers of receptor-generated signals. *Annu Rev Biochem.* 56:615–649.

Hunt TW, Fields TA, Casey PJ, Peralta EG. 1996. RGS10 is a selective activator of G alpha i GTPase activity. *Nature* 383:175–177.

Johnston CA, Taylor JP, Gao Y, Kimple AJ, Grigston JC, Chen JG, Siderovski DP, Jones AM, Willard FS. 2007. GTPase acceleration as the rate-limiting step in *Arabidopsis* G protein-coupled sugar signaling. *Proc Natl Acad Sci U S A.* 104:17317–17322.

Jones JC, Duffy JW, Machius M, Temple BR, Dohlman HG, Jones AM. 2011. The crystal structure of a self-activating G protein alpha subunit reveals its distinct mechanism of signal initiation. *Sci Signal.* 4: ra8.

Kozasa T, Jiang X, Hart MJ, Sternweis PM, Singer WD, Gilman AG, Bollag G, Sternweis PC. 1998. p115 RhoGEF, a GTPase activating protein for Galpha12 and Galpha13. *Science* 280:2109–2111.

Ma J, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* 14:860–869.

Nakamura S, Kreutz B, Tanabe S, Suzuki N, Kozasa T. 2004. Critical role of lysine 204 in switch I region of Galpha13 for regulation of p115RhoGEF and leukemia-associated RhoGEF. *Mol Pharmacol.* 66:1029–1034.

Nance MR, Kreutz B, Tesmer VM, Sterne-Marr R, Kozasa T, Tesmer JJ. 2013. Structural and functional analysis of the regulator of G protein signaling 2-galphaq complex. *Structure* 21:438–448.

Popov S, Yu K, Kozasa T, Wilkie TM. 1997. The regulators of G protein signaling (RGS) domains of RGS4, RGS10, and GAIP retain GTPase activating protein activity in vitro. *Proc Natl Acad Sci U S A.* 94: 7216–7220.

Ross EM, Wilkie TM. 2000. GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. *Annu Rev Biochem.* 69:795–827.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 19:101–109.

Sandler I, Zigdon N, Levy E, Aharoni A. 2014. The functional importance of co-evolving residues in proteins. *Cell Mol Life Sci.* 71:673–682.

Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. 2008. Rewiring the specificity of two-component signal transduction systems. *Cell* 133:1043–1054.

Slep KC, Kercher MA, He W, Cowan CW, Wensel TG, Sigler PB. 2001. Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 A. *Nature* 409:1071–1077.

Slep KC, Kercher MA, Wieland T, Chen CK, Simon MI, Sigler PB. 2008. Molecular architecture of Galphao and the structural basis for RGS16-mediated deactivation. *Proc Natl Acad Sci U S A.* 105: 6243–6248.

Sondek J, Lambright DG, Noel JP, Hamm HE, Sigler PB. 1994. GTPase mechanism of Gproteins from the 1.7-A crystal structure of transducin alpha-GDP-AIF-4. *Nature* 372:276–279.

Soundararajan M, Willard FS, Kimple AJ, Turnbull AP, Ball LJ, Schoch GA, Gileadi C, Fedorov OY, Dowler EF, Higman VA, et al. 2008. Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. *Proc Natl Acad Sci U S A.* 105:6457–6462.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.

Tesmer JJ. 2009. Structure and function of regulator of G protein signaling homology domains. *Prog Mol Biol Transl Sci.* 86:75–113.

Tesmer JJ, Berman DM, Gilman AG, Sprang SR. 1997. Structure of RGS4 bound to AlF4—activated G(i alpha1): stabilization of the transition state for GTP hydrolysis. *Cell* 89:251–261.

Urano D, Chen JG, Botella JR, Jones AM. 2013. Heterotrimeric G protein signalling in the plant kingdom. *Open Biol.* 3:120186.

Urano D, Jones JC, Wang H, Matthews M, Bradford W, Bennetzen JL, Jones AM. 2012. G protein activation without a GEF in the plant kingdom. *PLoS Genet.* 8:e1002756.

Watson N, Linder ME, Druey KM, Kehrl JH, Blumer KJ. 1996. RGS family members: GTPase-activating proteins for heterotrimeric G-protein alpha-subunits. *Nature* 383:172–175.

Wieland T, Bahtijari N, Zhou XB, Kleuss C, Simon MI. 2000. Polarity exchange at the interface of regulators of G protein signaling with G protein alpha-subunits. *J Biol Chem.* 275:28500–28506.

Willars GB. 2006. Mammalian RGS proteins: multifunctional regulators of cellular signalling. *Semin Cell Dev Biol.* 17:363–376.

Yi T-M, Kitano H, Simon MI. 2003. A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proc Natl Acad Sci U S A.* 100: 10764–10769.

Zamir L, Zaretsky M, Fridman Y, Ner-Gaon H, Rubin E, Aharoni A. 2012. Tight coevolution of proliferating cell nuclear antigen (PCNA)-partner interaction networks in fungi leads to interspecies network incompatibility. *Proc Natl Acad Sci U S A.* 109: E406–E414.

Zhong H, Wade SM, Woolf PJ, Linderman JJ, Traynor JR, Neubig RR. 2003. A spatial focusing model for G protein signals: regulator of G protein signaling (RGS) protein-mediated kinetic scaffolding. *J Biol Chem.* 278:7278–7284.