Systematic functional analysis of the Ras GTPase family unveils a conserved network required for anterograde protein trafficking

Marie-Elaine Caruso, Sarah Jenna, David L. Baillie, Roger Bossé, Jeremy C. Simpson, Eric Chevet, Saïd Taouji

HAL Id: hal-01447227

https://univ-rennes.hal.science/hal-01447227

Submitted on 20 Mar 2017

Systematic functional analysis of the Ras GTPase family unveils a conserved network required for anterograde protein trafficking

Marie-Elaine Caruso[1], Sarah Jenna[1], David L Baillie[2], Roger Bossé[3], Jeremy C. Simpson[4,8] , Eric Chevet[1,5,6,8] , Saïd Taouji[7,8]

[1]Dept of Surgery, McGill University, Montreal, Qc, Canada. [2]Dept of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada. [3]PerkinElmer, Boston, MA, USA. [4]School of Biology & Environmental Science and Conway Institute of Biomolecular & Biomedical Research, University College Dublin (UCD), Dublin 4, Ireland. [5]INSERM U1242, COSS, Université de Rennes-1, CLCC Eugene Marquis Rue de la Bataille Flandres Dunkerque 35042 RENNES, France. [6]BMYscreen, Bordeaux, France. [7]INSERM U916 CLCC Institut Bergonié 229, cours de l'Argonne 33076 Bordeaux Cedex France.

[8]Equal contribution and correspondence – JS, School of Biology & Environmental Science and Conway Institute of Biomolecular & Biomedical Research, University College Dublin (UCD), Dublin 4, Ireland. Email: jeremy.simpson@ucd.ie - EC, INSERM U1242 Université de Rennes-1 CLCC Eugene Marquis Rue de la Bataille Flandres Dunkerque 35042 Rennes, France. Email: eric.chevet@inserm.fr - ST, INSERM U916 CLCC Institut Bergonié 229, cours de l'Argonne 33076 Bordeaux Cedex France. Phone : +33(0)524071957 email : said.taouji@inserm.fr

Abstract

Phylogeny is often used to compare entire families of genes/proteins. We previously showed that classification of *C. elegans* Rho GTPases on the basis of their enzymatic properties was significantly different from sequence alignments. To further develop this concept, we have developed an integrated approach to classify *C. elegans* small GTPases based on functional data comprising affinity for GTP, sub-cellular localization, tissue distribution and silencing impact. This analysis led to establishing a novel functional classification for small GTPases. To test the relevance of this classification in mammals, we focused our attention on the human orthologs of small GTPases from a specific group comprising arf-1.2, evl-20, arl-1, Y54E10BR.2, unc-108 and rab-7. We then tested their involvement in protein secretion and membrane traffic in mammalian systems. Using this approach we identify a novel network containing 18 GTPases, and 23 functionally interacting proteins, conserved between *C. elegans* and mammals, which is involved in membrane traffic and protein secretion.

Small GTPases of the Ras superfamily are major regulators of cellular and developmental events, including cell differentiation, division, membrane trafficking, nuclear assembly, and cytoskeleton remodeling. Moreover, many members of the Ras-superfamily participate in signaling pathways linking extracellular and intracellular events through transmembrane receptors [1-6]. To date, the classification of the Ras superfamily has been based on nucleotide or protein sequence data [7]. Previously, we demonstrated that when the classification of the Rho subfamily was carried out exclusively on the basis of *in vitro* biochemical properties (related to GTP/GDP), the classification tree radically differed from the sequence-based tree, thus suggesting that the biological properties of this family of proteins is not strictly reflected by the average homology of their protein sequence. In addition, proteins clustering within a subfamily (according to their amino acids) should not *de facto* be implicated in the same cellular pathway. As a consequence, we postulated that a classification based on biological data might allow protein clustering according to their functional involvement. To test this, we focused on *C. elegans* Ras GTPases. Functional data were collected in public repositories or in our experiments and then verified alongside their homologs in mammalian cells using a genetic screen. Based on this, we identified a novel functional network involving GTPase-mediated regulation of protein secretion in metazoans.

Phylogenetic and biochemical characterization of C. elegans GTPases

We specifically searched the *C. elegans* genome for genes containing Rho, Ras, Rab, Ran or Arf domains and a total of 59 *C. elegans* genes potentially encoding small G proteins were identified. These proteins clustered into five subfamilies based on their amino acid sequence, which are similar to those, described for the human small G proteins (Arf, Ras, Rho, Rab and Ran ; **Figure 1A**). Finally, six genes did not have any mammalian ortholog thus being somehow specific to *C. elegans* (4R79.2; K02E10.1; Y54E10BR.2; Y37E3.5, Y32F2B.3; W04G5.2). Furthermore, most of these genes have no biological function yet identified. To functionally characterize these proteins, we built an expression library for the 59 *C. elegans* small G proteins identified above. Full-length Open Reading Frames (ORFs; from initiation ATG to the stop codon) encoding these proteins were either retrieved from the *C. elegans* ORFeome v1.1 or amplified from a *C. elegans* cDNA library. Using this approach, we generated a library of 52 ORFs cloned in

three different expression vectors and covering all small G protein subfamilies. This expression vector library represents a valuable resource to study many aspects of small G protein biology, namely protein-protein interactions, cellular localization, RNAi phenotypes and enzymatic activities.

Since some of these genes were only identified from BLAST searches and SMART characterization, and while some were expressed here for the first time (for example, F11A5.3 and Y37E3.5 orthologs of human Rab2b and Arl13b respectively), we explored whether the proteins encoded by these genes were genuine GTP-binding proteins. Open Reading Frames (ORFs) encoding those GTPases were cloned using the Gateway technology to produce N-terminally tagged GST fusion proteins (**Figure S1**) that were use in an Alphascreen GTP binding assay [7]. IC50 value for each recombinant GTPases was determined through a competition assay using non-biotinylated GTP (**Figure 1B, Table 1**). Most of the *C. elegans* small G proteins identified (>80%) were able to bind GTP with an affinity ranging from 5nM to 360nM. This was particularly striking for the uncharacterized GTPases such as F11A5.4, which, was retrieved exclusively from sequence alignment predictions. F11A5.4 is ortholog to the yeast YPT31 GTPase involved in the exocytic pathway. Globally our approach provides information on the biochemical properties associated to *C. elegans* small GTPases.

Subcellular and tissular localization and RNAi phenotypes of C. elegans GTPases

To further investigate the functional similarity of small GTPases, we systematically generated functional data (subcellular and tissue localization, RNAi phenotype) and combined these data with the above-described biochemical information and with *C. elegans* public data relating to their co-expression profiles or their physical and genetic interactions. Some of the GTPases tested in our assay displayed a high affinity for GTP, such as Arf-6, Rap-2 and Rab-37, whereas others such as Ras-1, Crp-1 and Mig-20 exhibited a lower affinity for GTP. These affinities did not strictly match the subfamilies as defined by sequence alignment (data not shown), however this important biochemical parameter would be expected to have a significant impact on the dynamics of GTPase cycle and effector recruitment and specificity. Localization, either at the tissue or cellular level, is also a key determinant in predicting protein function. We therefore determined both the sub-cellular localization and tissue expression of small G proteins using fluorescently N-

terminally tagged proteins and microscopy analysis (**Figure 2A**). Due to the strong evolutionary conservation of GTPases sub-family cell localization, we also constructed *C. elegans* transgenic organisms expressing transcriptional fusions to identify tissue expression pattern of the GTPase (**Figure 3**). Although many of the GTPases were observed to localize to specific sub-cellular membrane structures in Vero cells using reference compartment markers (**Figure 2B**), it was notable that the majority of the Arf sub-family was only detected in a soluble cytosolic pattern. This group of proteins all rely on an N-terminal myristoylation anchor for their membrane association, and therefore the N-terminal ECFP tagging used in our approach appeared to abrogate this functionality, as previously reported [8]. To confirm the expected localization of these proteins, C-terminal fusions with a FLAG tag were expressed in Vero cells followed by immunofluorescence confocal analysis. This revealed the main Golgi localization of Arf-1.1, Arf-1.2, Arl-1 and Sar-1 (**Figure S2**). For tissular distribution in transgenic worms, except for 3 GTPases (Rho-1, Cdc-42.1 and Mig-2) which display ubiquitous expression profiles in *C. elegans*, the results presented in **Figure 3A, B and Tables S1-5** demonstrate that most small G proteins are expressed in a tissue-specific fashion. This is well illustrated by specific examples either retrieved from our data set (**Figure 3A**; Rab-6.2, Ral-1, Drn-1 or Rap-2) or from the Genome BC *C. elegans* Gene Expression Consortium (http://gfpweb.aecom.yu.edu/; **Figure 3B**; Arf-1.2, Rab-1, Crp-1). A systematic phenotyping approach was next carried out using RNA interference (RNAi)-based gene silencing of each GTPase in *C. elegans*. The silencing of 52 GTPases was carried out as previously described [9] and 49 GTPases silencing-induced phenotypes (as defined in Wormbase) were scored. Altogether 60 % of GTPases (29) with complete data sets were retained and with those we initiated an integration approach to generate a new classification of this superfamily of proteins.

C. elegans GTPase data integration

We integrated all the above-described data using a distance matrix based approach and compared the clustering pattern generated with the classification carried out exclusively on the basis of sequence alignment. The comparison of the initial sequence based matrix to the various combinations of biological information is reported in **Figure 4A**. From this comparison, it appeared that tissue (TL) and cellular localization (CL) dramatically impacted on the distance to the sequence-

based classification, thus suggesting a relative independence of these critical biological properties to the primary amino-acid sequence. We then selected the combination of data, which presented the highest distance to the sequence-based classification. Two configurations led to a similar result and both integrated the small G proteins cellular localization (CL), mRNA co-expression (coexp), affinity for GTP (gtp aff) and the tissue localization (TL) (**Figure 4A**). Interestingly, integration of the RNAi-based phenotype (phen) data did not exert a major impact on the distance of functional clustering to sequence clustering. Therefore, we constructed a hierarchical tree containing all the biological information available, with the exception of the data obtained in the *C. elegans* RNAi experiments (**Figure 4B**). The resulting tree contained five branches, designated as I to V. These branches represent five major groups of GTPases (**Figure 4B, Table 2**). Each group was then evaluated for the highest incidence of common RNAi phenotypes based on the data available in **Tables S1-5**. GTPases from group III presented the highest number of RNAi phenotypes in common (32% vs 9% (group II); 19% (group IV) and 28% (group V). As group I included only one member, it was excluded from this comparison. Group III was consequently selected for further analysis. This group contains six GTPases: Arf-1.2, Evl-20 homolog of the mammalian Arl-2), Arl-1, the unassigned Arl-3 homolog, Unc-108 (homolog of the human Rab-2) and Rab-7. As expected these GTPases did not cluster in the sequence-based tree (**Figure 4B**). To evaluate the biological relevance of our classification method, we used protein-protein interaction data and genetic interactions for these genes available in Wormbase. This analysis led to a network containing 104 nodes and 123 edges. Network analysis using GO annotation revealed that the largest number of group III GTPase interacting partners were involved in the control of membrane/protein trafficking (>40% of the total). A new network including only molecules with a known function in membrane and protein trafficking was generated and contained 41 nodes and 51 edges (**Figure 4C**). Furthermore, in addition to the six small GTPases initially contained in the test group (dark blue nodes), 12 others GTPases were contained in the new network (green nodes). The remaining 23 nodes (yellow) corresponded to proteins involved in membrane trafficking (e.g. adaptins, dynamin, COPI coat complex subunits). The GO-based (cellular process) functional annotation of GTPases interacting partners revealed that 39% of them were related to Golgi vesicle transport and mostly connected to 4 GTPases of group III.

Functional conservation in mammalian GTPases

We next tested the functional significance of this network in mammalian cells using a siRNA-based approach. Using a high-throughput RNAi platform, the impact of the human orthologs of the *C. elegans* proteins was tested on the model cargo protein tsO45G [10]. The down-regulation of genes involved in the secretory process results in the accumulation of ECFP-tagged tsO45G in intracellular compartments, with a concomitant reduction in cell surface tsO45G as detected by antibody labeling (**Figure 5A**). Results obtained with the knockdown of the COPI coat subunit β'-COP (COPB2) are shown as a typical example of an inhibitory phenotype (**Figure 5A**). The quantification methodology used to determine whether cells showed different secretion rates has been previously described (13). Twenty-eight were first selected genes for which the exact ortholog could be identified in human. To provide a qualitative representation of experimental robustness, a bar graph was generated reporting the fraction of experimental replicates in which secretion rates were measured at being either inhibitory (red) or accelerating (green) in the cells silenced for the various GTPases under test. Of the 24 genes examined, 2 of the experiments led to discordant results across the replicates (for UNC50 and STX18) and were consequently excluded from further studies. The large majority of the genes (68%) when silenced showed an inhibitory effect on tsO45G arrival at the cell surface (**Figure 5B**), with only 2 of the genes giving an accelerating phenotype. When this information was quantitatively superimposed on the network shown in **Figure 4C**, this led to the visualization of a new functional network (**Figure 5C**) (red/inhibitory, green/accelerating). Interestingly, the analysis of group III GTPase first-degree interactors now allowed us to propose an additional level of classification. Indeed when the accelerating/inhibiting role of group III GTPases first-degree interacting partners was taken into account, two major groups of GTPases appeared in group III with one mainly connected to a pro-secretory network (**Figure 5D**). Altogether, the integration strategy used in our approach led to the definition of GTPase groups with functional specificities that were proven valid using orthogonal silencing experiments in mammalian systems.

We have previously shown that sequence-based alignments do not necessarily reflect evolutionary conservation of biochemical properties [7]. Herein we

further document this concept and propose that classification of protein families based on functional data can be used to reveal new biological properties of these families. We compiled a large number of functional data for the entire family of small GTPases from *C. elegans*. Public and experimental data relative to Ras GTPases were collected, integrated, clustered into hierarchical trees and compared to the sequence-based tree. Depending on the number and the nature of the functional information integrated, novel classifications emerged and revealed new functional groups of GTPases. This would not have been provided using the sequence-based classification only. To validate the predictive power of our strategy, we selected one group (group III) of GTPases. This group is composed of Arf-1.2, Evl-20 (Arl2 homolog), Arl-1, Y54E10BR.2 (ARFRP1 human ortholog), Unc-108 and Rab-7. In *C. elegans*, these 6 GTPases participate in i) the endocytic process (*Arf-1.2, Y54E10BR.2, Arl-1, Unc-108*; [11-13]); ii) the formation of endosomes, lysosomes and phagolysosomes (Unc-108, Rab-7); and iii) cytokinesis (Evl-20; [14]). Even though the involvement of *C. elegans* proteins has never directly been show to contribute to anterograde traffic, the extreme conservation of Arf-1.2 with its human ortholog Arf-1 (98% similarity) suggests its involvement in such process [15-17]. In addition, when clustered based on their amino-acid sequences, these six proteins belong to at least two different sequence-based groups both in *C. elegans* (**Figure 1**) and in various mammalian species (not shown). Thus, prior to our study it was almost unpredictable that these six proteins would be functionally related and may participate in the regulation of protein secretion. Data integration revealed that these six *C. elegans* proteins clustered in the same functional group and shared a large number of phenotypes when silenced. In addition, when genetic or physical interacting partners were analyzed for these six proteins, a significant enrichment in genes involved in membrane trafficking was found, indicating that they could function cooperatively in a common biological process (**Figure 5**). To test our hypothesis, we took advantage of the conservation of small GTPases from *C. elegans* to mammals. Remarkably, many of the *C. elegans* GTPases are still able to localize to specific membranes in mammalian cells, thus indicating a strong evolutionary conservation (**Figure 2**). We evaluated the contribution of the human orthologs of functional network identified (**Figure 5**) to the secretion of the model cargo tsO45G using a systematic RNAi approach [10]. These experiments demonstrated that a large proportion of proteins in the network (>65%) actually participate in the cell surface delivery of tsO45G, thus validating the relevance of our approach. Moreover, genomic screens have not yet

revealed any functional connections between the members of this family. Hence not only is our integrative functional approach novel, but it also provides a useful resource within the small GTPase field. However, several potential shortcomings need to be considered if our approach is to be applied more generally. Indeed, mostly due to technical reasons, our data set remains incomplete, and so our classification (**Figure 4**) would be predicted to gain in precision and relevance when using a full data set. Indeed the classification was only carried out using GTPases for which data were complete and consequently 30 GTPases were excluded from our final analysis. Another parameter that needs to be taken into account for future studies is the weight that functional information should take in the classification process. It is currently difficult to predict whether for instance sub-cellular localization is more functionally relevant than RNAi phenotype or if the opposite is true. Finally, the number and type of data sets integrated into such studies could be relatively easily increased; for instance allowing consideration of the impact of small GTPases on the regulation of the Unfolded Protein Response [9] or the interaction networks of Rab GTPases [18] and protein secretion or membrane traffic [19]. In conclusion, our work identified a conserved cluster of GTPases that function cooperatively to promote protein secretion. Since the existence of such a functional cluster was previously unforeseen when based only on sequence alignment and literature data, we believe that a functional genomic approach could be of wider interest in the identification of novel sub-families of genes involved in the regulation of common molecular and cellular events.

Conflict of interest: None declared

Author contribution: M.E.C designed performed experiments and analyzed data ; S.J and D.L.B developed analytical tools ; R.B provided reagents and commented manuscript; J.C.S and E.C and S.T designed experiments, analyzed data and wrote the paper.

References

[1] Goldfinger, L. E., Choose your own path: specificity in Ras GTPase signaling. *Mol Biosyst* 2008, *4*, 293-299.

[2] Hakoshima, T., Shimizu, T., Maesaki, R., Structural basis of the Rho GTPase signaling. *J Biochem* 2003, *134*, 327-331.

[3] Knaus, U. G., Rho GTPase signaling in inflammation and transformation. *Immunol Res* 2000, *21*, 103-109.

[4] Rensen, W. M., Mangiacasale, R., Ciciarello, M., Lavia, P., The GTPase Ran: regulation of cell life and potential roles in cell transformation. *Front Biosci* 2008, *13*, 4097-4121.

[5] Sinha, S., Yang, W., Cellular signaling for activation of Rho GTPase Cdc42. *Cell Signal* 2008, *20*, 1927-1934.

[6] Vernoud, V., Horton, A. C., Yang, Z., Nielsen, E., Analysis of the small GTPase gene superfamily of Arabidopsis. *Plant Physiol* 2003, *131*, 1191-1208.

[7] Caruso, M. E., Jenna, S., Beaulne, S., Lee, E. H*., et al.*, Biochemical clustering of monomeric GTPases of the Ras superfamily. *Mol Cell Proteomics* 2005, *4*, 936-944.

[8] Keller, P. J., Fiordalisi, J. J., Berzat, A. C., Cox, A. D., Visual monitoring of post-translational lipid modifications using EGFP-GTPase probes in live cells. *Methods* 2005, *37*, 131-137.

[9] Caruso, M. E., Jenna, S., Bouchecareilh, M., Baillie, D. L*., et al.*, GTPase-mediated regulation of the unfolded protein response in Caenorhabditis elegans is dependent on the AAA+ ATPase CDC-48. *Mol Cell Biol* 2008, *28*, 4261-4274.

[10] Simpson, J. C., Cetin, C., Erfle, H., Joggerst, B*., et al.*, An RNAi screening platform to identify secretion machinery in mammalian cells. *J Biotechnol* 2007, *129*, 352-365.

[11] Dugani, C. B., Klip, A., Glucose transporter 4: cycling, compartments and controversies. *EMBO Rep* 2005, *6*, 1137-1142.

[12] Sahin, A., Espiau, B., Tetaud, E., Cuvillier, A*., et al.*, The leishmania ARL-1 and Golgi traffic. *PLoS One* 2008, *3*, e1620.

[13] Sato, K., Sato, M., Audhya, A., Oegema, K*., et al.*, Dynamic regulation of caveolin-1 trafficking in the germ line and embryo of Caenorhabditis elegans. *Mol Biol Cell* 2006, *17*, 3085-3094.

[14] Antoshechkin, I., Han, M., The C. elegans evl-20 gene is a homolog of the small GTPase ARL2 and regulates cytoskeleton dynamics during cytokinesis and morphogenesis. *Dev Cell* 2002, *2*, 579-591.

[15] Girod, A., Storrie, B., Simpson, J. C., Johannes, L*., et al.*, Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum. *Nat Cell Biol* 1999, *1*, 423-430.

[16] Lanoix, J., Ouwendijk, J., Lin, C. C., Stark, A*., et al.*, GTP hydrolysis by arf-1 mediates sorting and concentration of Golgi resident enzymes into functional COP I vesicles. *EMBO J* 1999, *18*, 4935-4948.

[17] Lanoix, J., Ouwendijk, J., Stark, A., Szafer, E*., et al.*, Sorting of Golgi resident proteins into different subpopulations of COPI vesicles: a role for ArfGAP1. *J Cell Biol* 2001, *155*, 1199-1212.

[18] Kanno, E., Ishibashi, K., Kobayashi, H., Matsui, T*., et al.*, Comprehensive screening for novel rab-binding proteins by GST pull-down assay using 60 different mammalian Rabs. *Traffic* 2010, *11*, 491-507.

[19] Sato, K., Norris, A., Sato, M., Grant, B. D., C. elegans as a model for membrane traffic. *WormBook, ed. The C. elegans Research Community* 2014, 1-47.

Table 1: EC50 values for GTPase binding to GTP in vitro

GTPases	Locus	EC50 nM	SD
rab11.2	W04G5.2	4.2	0.4
arf-6	Y116A8C.12	6.5	0.3
rap-2	C25D7.7	8	0.6
evl-120	F22B5.1	11.9	0.9
rab-37	Y92C3B3.b	14.7	1.7
cdc42	R07G3.1	15.05	2.9
tag-312	C33D12.6	15.4	0.4
F58C8.5	F54C8.5	17.25	5.7
arf-1.2	B0336.2	17.45	1.3
rac-2	K03D3.10b	19	3
F11A5.3	F11A5.3	19.2	2.2
rab-21	T01B7.3	19.3	5.6
rho-1	Y51H4A.3	20.2	6.1
rab-33	F43D9.2	22.7	2.2
rab-5	F26H9.6	22.95	2.8
sar-1	C08F8.7	23.7	1.6
let-60	ZK792.6	25.45	2.1
Y37E3.5	Y37E3.5	30	1.1
unc-108 (rab-2)	F53F10.4	31.45	4.7
rab-18	Y92C3B.3a	32.8	6.6
ced-10	C09G12.8a	33.25	1.4
Y57g11c.13	Y57G11C.13	35.4	5.2
crp-1	Y32F6B.3	35.8	5.3
rab11.1	F53G12.1	38.6	7.8
arl-1	F54C9.10	40.7	8.2
rab-14	K09A9.2	42.3	2.8
ral-1	Y53G8AR.3	44.65	8.1
rab-3	C18A3.6b	46.4	5.6
arl-6	C38D4.8	47.3	3.8
rab-30	Y45F3A.2	48.3	3.4
arf-3	F57H12.1	51.9	2.8
Y54E10BR.2	Y54E10BR.2	52.3	2.3
rab-27	Y87G2A.4	56.1	6.6
rab-28	Y11D7A.4	57.2	4.3
rab-6.2	T25G12.4	60.85	2.2
F11A5.4	F11A5.4	61.35	2.1
arl-5	ZK632.8	62	10.9
arl-3	F19H8.3	66.9	8.1
rab-10	T23H2.5	77.2	10.3
rab-6.1	F59B2.7	79.6	9.1
rab-39	D2013.1	86.5	6.5
rab-8	D1037.4	87.5	8.9
rab-7	W03C9.3	93.5	10.8
rab-35	Y47D3A.25	100.5	2.5
rab-1	C39F7.4	139	9.2
ras-2	F17C8.4	144.25	11.3
rab-19	Y62E10A.9	164	45.9
ras-1	C44C11.1a	360	27
arf-1.1	F45E4.1	nc	-
ora-1	F57H12.3	nc	-
ran-1	K01G5.4	nc	-
rap1	C27B7.8	nc	-

Table 2: Description of the GTPase groups as defined by data integration

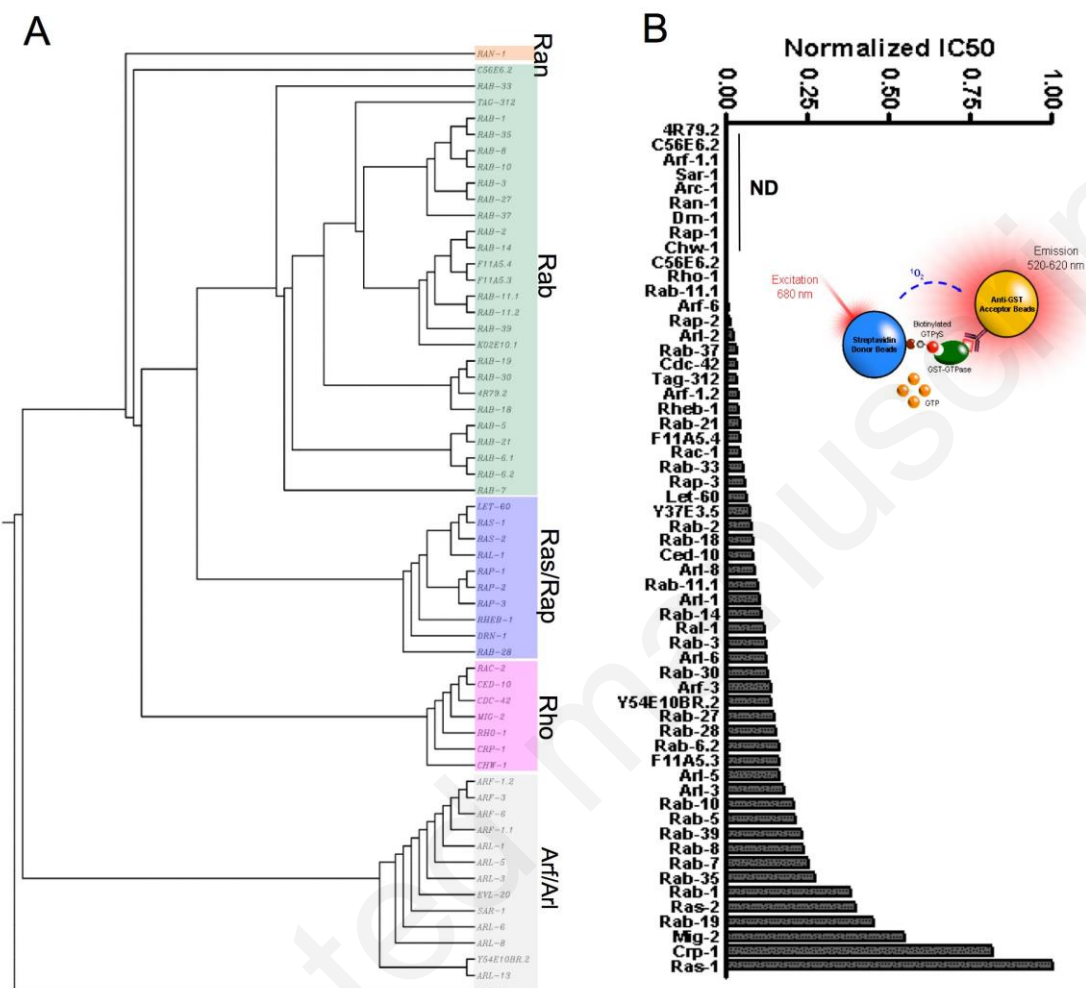| Group #	Name	Name	Function in membrane traffic in C. elegans
1	R07G3.1	cdc-42	+
2	C39F7.4	rab-1	+
	T25G12.4	rab-6.2	
	ZK792.6	let-60 (RAS)	
	F19H8.3	arl-3	+
	C44C11.1	ras-1 (RRAS2)	
3	B0336.2	arf-1.2	+
	F22B5.1	evl-20	
	F54C9.10	arl-1	+
	Y54E10BR.2	(ARFRP1)	+
	F53F10.4	unc-108	+
	W03C9.3	rab-7	+
4	C18A3.6	rab-3	+
	F17C8.4	ras-2	
	Y32F6B.3	crp-1	+
	Y87G2A.4	aex-6	+
	Y11D7A.4	rab-28	
5	T23H2.5	rab-10	+
	Y45F3A.2	rab-30	+
	C25D7.7	rap-2	Indirect
	Y53G8AR.3	ral-1	+
	D2013.1	rab-39	
	Y57G11C.13	arl-8	+
	F57H12.1	arf-3	+
	Y116A8C.12	arf-6	+
	C38D4.8	arl-6	Indirect
	K03D3.10	rac-2	

Figure Legends

Figure 1: *C. elegans* **small G proteins subfamilies**. **A**) Sequence-based hierarchical clustering of *C. elegans* small G proteins. The small G proteins amino acid sequences were retrieved for Wormbase database and hierarchical clustering was performed using ClustalW. **B**) AlphaScreen® competition assay to determine the GTP affinity of small G proteins. AlphaScreen® signal is emitted when the donor and acceptor beads are brought into close proximity by physical interaction between biotinylated GTPγS and GST-tagged small G proteins. The streptavidin Donor bead binds to b-GTPγS, and the anti-GST Acceptor bead binds to the GST-GTPase. The IC50 value corresponds to the free GTP concentration required to replace 50% of the bound biotin-GTPγS (inset). A higher IC50 corresponds to a low GTPase affinity for GTP. Each experiment was repeated twice with triplicates. ND: not determined.

Figure 2: Sub-cellular distribution *C. elegans* small G proteins. A) ECFP-tagged small G proteins were co-transfected with **B**) organelle markers into Vero cells and the cellular localization was assessed using fluorescence microscopy. **C**) Clustering analysis of small GTPases of the Ras superfamily based on their sub-cellular localization in mammalian cells. (NM= nuclear membrane; ER = endoplasmic reticulum; TGN = trans-Golgi network; ENDO = endosomes; EE = early endosomes; LE = late endosomes; M = mitochondria; Mic = microtubules; PM = plasma membrane; DC = diffuse cytoplasm; N = nucleus; FCS = faint cytoplasmic structures). Yellow indicates positive localization and black indicates negative or no localization.
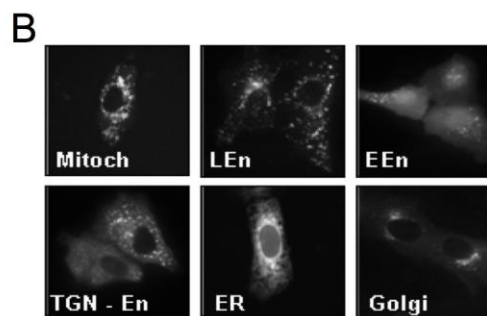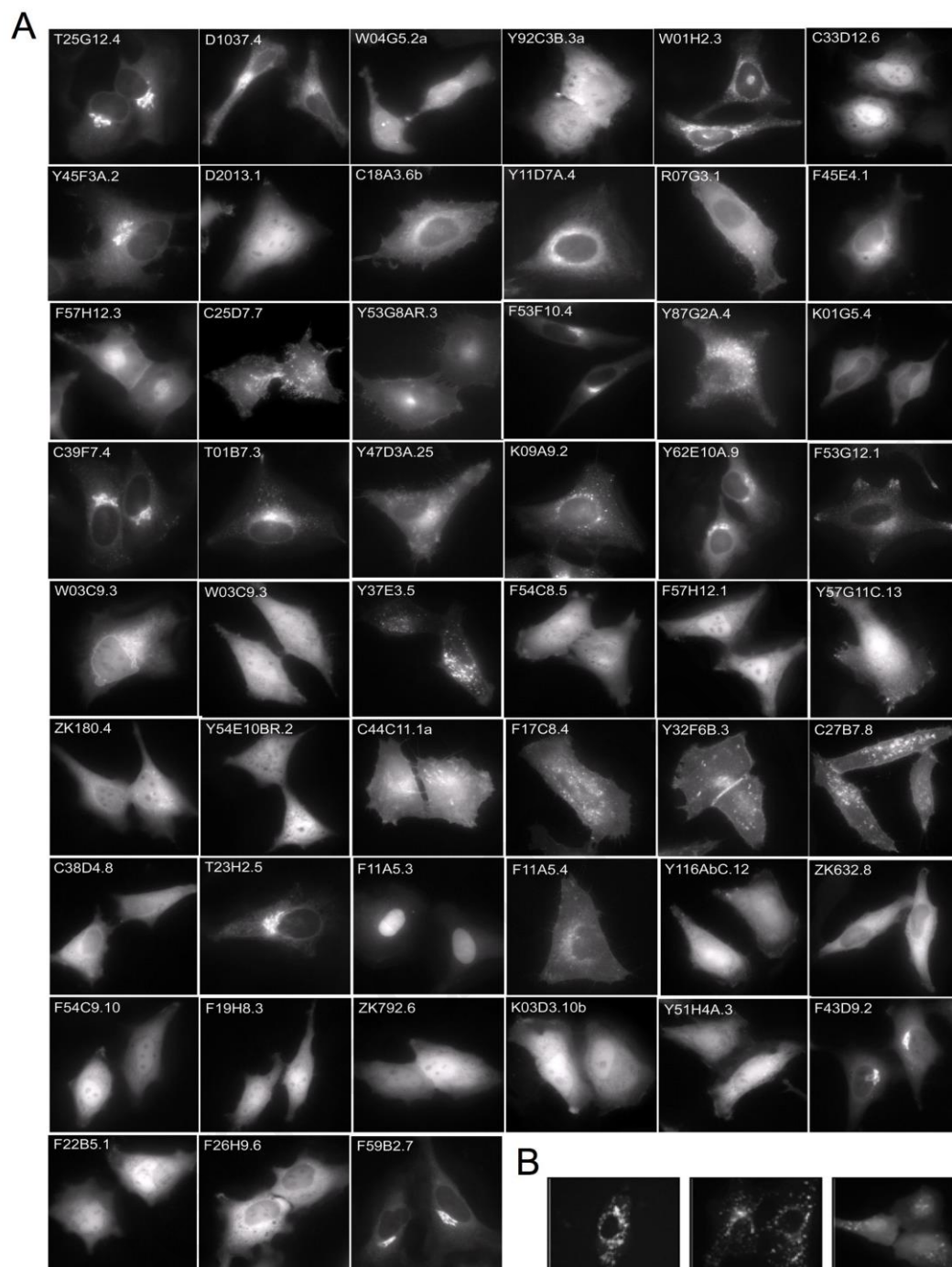
Figure 3: Tissue expression profile of small G proteins. These experiments were carried out in living *C. elegans* and performed using transgenic worms expressing *promoter::GFP* reporter genes. Tissue localization was visualized using fluorescence microscopy. Examples of localization experiments either carried out in the laboratory (**A**) or retrieved from http://gfpweb.aecom.yu.edu/index (**B**). Body wall muscle (B.W. mus); Pharynx (Pha); Head neuron (H. neur.); Tail neuron (T. neur); Hypodermis (Hyp.); Nerve ring (N. Ring); Dorsal cord (D. Cord); Ventral Cord (V. Cord); Intestine (Int.)

Figure 4: Integrative and functional analysis of C. elegans GTPase - A) Integration of the data obtained for the small G proteins cellular localization (CL), mRNA co-expression (coexp), GTP affinity (gtp aff), phenotype (phen) and tissue localization (TL). The bar graph shows the distance, calculated using Euclidean matrices, between the sequenced-based tree and different hierarchical trees integrating one or various combinations of these biochemical data. The smaller the distance is, the closer the biochemical tree is from the SB tree. **B**) Hierarchical tree obtained by integration of the five biochemical data sets. The cluster in red (sub-family III) was selected for further analyses as its members presented the highest number of RNAi phenotypes in common. **C**) Membrane trafficking network including gpIII GTPases (B) and its genetic and physical interactions as available in Wormbase. The color codes are defined as follows: dark blue = gpIII GTPases; green = other GTPases; yellow = other genes involved in membrane trafficking.

Figure 5: Conservation of the new classification in mammalian cells. A) Example images from tsO45G secretion assay showing cells treated with either non-silencing control siRNAs (upper panels) or siRNAs targeting β'-COP (COPB2) (lower panels). Bar represents 10μm. **B**) Results from the siRNA screen, indicating proportion of replicate experiments in which secretion phenotypes could be detected. Green represents an accelerating phenotype, red an inhibitory phenotype and white, no effect. **C**) Functional network involving gpIII GTPases as shown in A) that integrates inhibitory/accelerating properties of network members. Green reflects an accelerating phenotype, red an inhibiting phenotype and white a neutral phenotype. **D**) Updated classification of gpIII GTPases based on the properties of their first-degree interacting partners.
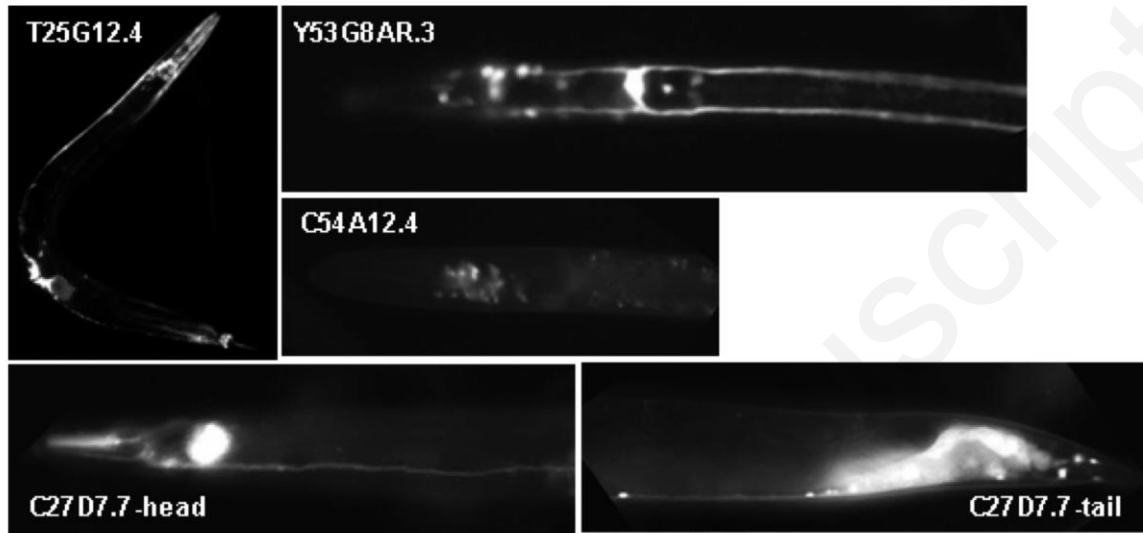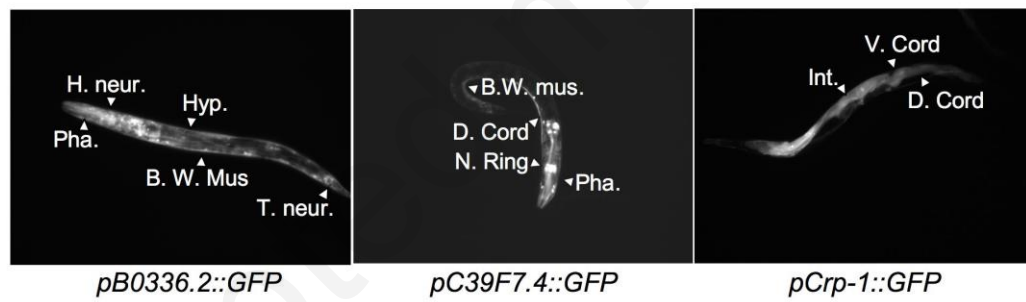
Caruso_Fig1

A

| T25G12.4 | D1037.4 | W04G5.2a | Y92C3B.3a | W01H2.3 | C33D12.6 |

| Y45F3A.2 | D2013.1 | C18A3.6b | Y11D7A.4 | R07G3.1 | F45E4.1 |

| F57H12.3 | C25D7.7 | Y53G8AR.3 | F53F10.4 | Y87G2A.4 | K01G5.4 |

| C39F7.4 | T01B7.3 | Y47D3A.25 | K09A9.2 | Y62E10A.9 | F53G12.1 |

| W03C9.3 | W03C9.3 | Y37E3.5 | F54C8.5 | F57H12.1 | Y57G11C.13 |

| ZK180.4 | Y54E10BR.2 | C44C11.1a | F17C8.4 | Y32F6B.3 | C27B7.8 |

| C38D4.8 | T23H2.5 | F11A5.3 | F11A5.4 | Y116AbC.12 | ZK632.8 |

| F54C9.10 | F19H8.3 | ZK792.6 | K03D3.10b | Y51H4A.3 | F43D9.2 |

| F22B5.1 | F26H9.6 | F59B2.7 |

B

| Mitoch | LEn | EEn |
| TGN - En | ER | Golgi |

Caruso_Fig2

A

T25G12.4

Y53G8AR.3

C54A12.4

C27D7.7 -head

C27D7.7 -tail

B

H. neur.

Hyp.

Pha.

B. W. Mus

T. neur.

pB0336.2::GFP

B.W. mus.
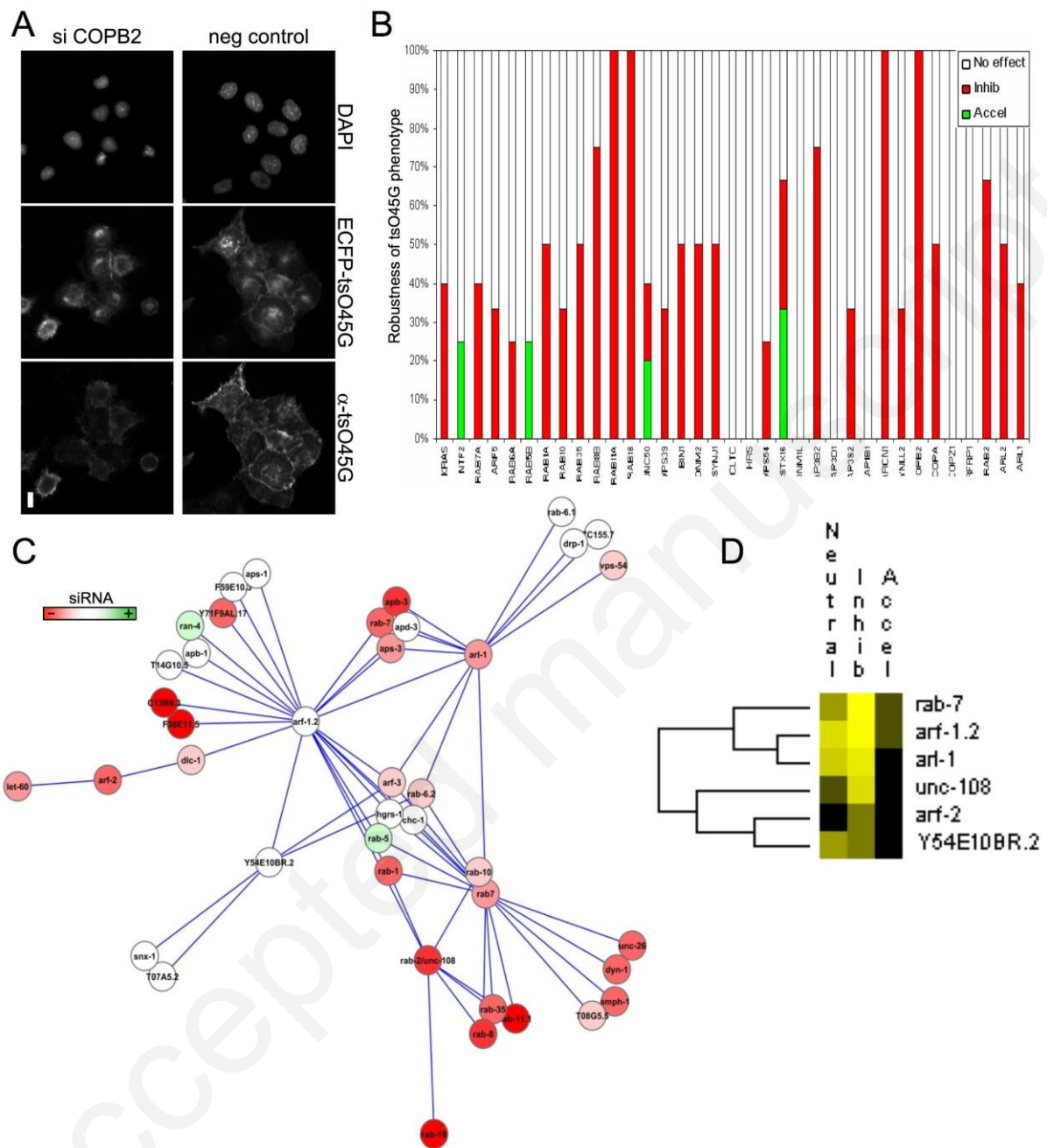
D. Cord

N. Ring

Pha.

pC39F7.4::GFP

V. Cord

Int.

D. Cord

pCrp-1::GFP

Caruso_Fig3

Caruso_Fig4

19

Caruso_Fig5