Research article

Open Access

The domain architecture of large guanine nucleotide exchange factors for the small GTP-binding protein Arf

Barbara Mouratou^{†1}, Valerie Biou^{†1}, Alexandra Joubert¹, Jean Cohen², David J Shields³, Niko Geldner^{4,5}, Gerd Jürgens⁴, Paul Melançon³ and Jacqueline Cherfils^{*1}

Address: ¹Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, avenue de la Terrasse, 91198 Gif sur Yvette cedex, France, ²Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette, France, ³Department of Cell Biology, University of Alberta, Edmonton, Canada, ⁴Center of Plant Molecular Biology, Universitaet Tuebingen, Tuebingen, Germany and ⁵Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, USA

Email: Barbara Mouratou - mouratou@lebs.cnrs-gif.fr; Valerie Biou - biou@lebs.cnrs-gif.fr; Alexandra Joubert - joubert@lebs.cnrs-gif.fr; Jean Cohen - cohen@cgm.cnrs-gif.fr; David J Shields - david.shields@ualberta.net; Niko Geldner - geldner@salk.edu; Gerd Jürgens - gerd.juergens@uni-Tuebingen.de; Paul Melançon - paul.melancon@ualberta.ca; Jacqueline Cherfils* - cherfils@lebs.cnrs-gif.fr * Corresponding author tEqual contributors

Corresponding aution (Equal contributo

Published: 17 February 2005

BMC Genomics 2005, 6:20 doi:10.1186/1471-2164-6-20

This article is available from: http://www.biomedcentral.com/1471-2164/6/20

© 2005 Mouratou et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 04 November 2004 Accepted: 17 February 2005

Abstract

Background: Small G proteins, which are essential regulators of multiple cellular functions, are activated by guanine nucleotide exchange factors (GEFs) that stimulate the exchange of the tightly bound GDP nucleotide by GTP. The catalytic domain responsible for nucleotide exchange is in general associated with non-catalytic domains that define the spatio-temporal conditions of activation. In the case of small G proteins of the Arf subfamily, which are major regulators of membrane trafficking, GEFs form a heterogeneous family whose only common characteristic is the well-characterized Sec7 catalytic domain. In contrast, the function of non-catalytic domains and how they regulate/cooperate with the catalytic domain is essentially unknown.

Results: Based on Sec7-containing sequences from fully-annotated eukaryotic genomes, including our annotation of these sequences from Paramecium, we have investigated the domain architecture of large ArfGEFs of the BIG and GBF subfamilies, which are involved in Golgi traffic. Multiple sequence alignments combined with the analysis of predicted secondary structures, non-structured regions and splicing patterns, identifies five novel non-catalytic structural domains which are common to both subfamilies, revealing that they share a conserved modular organization. We also report a novel ArfGEF subfamily with a domain organization so far unique to alveolates, which we name TBS (<u>TBC-Sec7</u>).

Conclusion: Our analysis unifies the BIG and GBF subfamilies into a higher order subfamily, which, together with their being the only subfamilies common to all eukaryotes, suggests that they descend from a common ancestor from which species-specific ArfGEFs have subsequently evolved. Our identification of a conserved modular architecture provides a background for future functional investigation of non-catalytic domains.

Background

Guanine Nucleotide Exchange Factors (GEFs) are obligatory components of signaling cascades regulated by small GTP-binding proteins (called small G proteins hereafter). Their biochemical activity is to stimulate the dissociation of the tightly bound GDP nucleotide from the small G protein in response to cellular signals. Thereby, they favor the binding of the more abundant cellular GTP, organizing the active conformation of the small G protein which can recruit its effectors (reviewed in [1]). Each small G protein family features its own ensemble of GEFs characterized by a conserved catalytic domain responsible for nucleotide exchange, which is generally combined with non-catalytic domains that define the spatio-temporal conditions of activation. In the case of small G proteins of the Arf family, which are major regulators in membrane trafficking (reviewed in [2]), the exchange domain is a conserved module of ~200 amino acids called the Sec7 domain [3]. Its biochemical (reviewed in [4]) and structural [5,6] mechanisms have been investigated in detail. Remarkably, the Sec7 domain is the only domain that is conserved in all ArfGEFs (reviewed in [7,8]) and it is to some extent interchangeable between species [9]. In contrast, little is known about the functions of the other domains, which are likely to determine intracellular localization of ArfGEFs and their responsiveness to specific signals.

As for most small G proteins, Arf family members are outnumbered by ArfGEFs in many species. In humans for instance, 5 Arf proteins have been identified, and there are at least 13 proteins carrying a Sec7 domain, of which most have been characterized as bona fide ArfGEFs (reviewed in [7,8]). Thus an individual Arf protein may be activated by more than one GEF, emphasizing that essential aspects in building up the Arf responses may be encoded by the modular architecture of their GEFs. Sequence similarity in the non-catalytic regions forms the basis for the classification of ArfGEFs into subfamilies. 8 subfamilies are currently identified in eukaryotes with sizes ranging from small (~40-80 kD including CYH, EFA6 and FBS), to medium (~100-150 kD, including BRAG/LONER, SYT1, SYT2) and large (~170-200 kD) ArfGEFs (reviewed in [7,8]). Large ArfGEFs comprise two subfamilies which we will refer to as the BIG and GBF subfamilies after the name of their human representatives. The GBF subfamily includes human GBF1 [10], Arabidopsis GNOM [11] and Saccharomyces Gea1 and Gea2 [12], the BIG subfamily human BIG1 and BIG2 [13,14] and yeast Sec7p [15]. An additional subfamily called RalF is found in Rickettsie and Legionella bacteria, likely acting on an host Arf pathway [16]. Analysis of the CYH and EFA6 subfamilies, present only in multicellular animals, and that of the large Arf-GEFs, found in all eukaryotes, have yielded most of the functional data currently available. CYH and EFA6 are



Figure I

Venn diagram of the nine Sec7-containing subfamilies sorted according to the species where each subfamily has been found. The TBS subfamily was identified in this study. The BIG and GBF subfamilies are merged in a higher order subfamily (GBG), and are the only subfamily common to all eukaryotes.

active on Arf6 at the plasma membrane where they may function in the crosstalk of membrane traffic, cytoskeleton dynamics and signalling in endosomal pathways (reviewed in [17]). Most members of the BIG and GBF subfamilies characterized so far function in vesicular trafficking at the Golgi [12,14,18], except for BIG2, which also localizes on recycling endosomes [19], and GNOM which acts in the endosomal recycling pathway [11].

The domain architecture of non-catalytic regions of Arf-GEFs, hence their contribution to specific aspects of the build-up of the Arf response, is essentially not established except for those ArfGEFs with domains found in other classes of cellular regulators. The known domains include membrane-interacting PH domains in the CYH (reviewed in [20]), EFA6 [21] and possibly BRAG/LONER[22] sub-families, and a putative F-box in the FBS subfamily [23], a protein-protein interaction domain that has been involved in the recruitment of substrates to the SCF ubiquitination machinery. Coiled-coil structures have also been predicted in the N-terminus of the CYH subfamily and in the C-terminus of the EFA6 subfamily. In CYH,

| | Species | Protein name ^a | Accession Number | Size in amino acids |
|---------------|---------|---------------------------|----------------------|---------------------|
| Metazoa | Ag | Q7PWN5 | EAA14874 | 1522 |
| | | Q7PXQ7 | EAA00837 | 1285 |
| | Ce | Q9XWG5 | NP_493386 | 1628 |
| | | Q9XTF0 | NP_499522 | 1820 |
| | Dm | Q9VJW1 | AAF53331 | 1653 |
| | | Q9V696 | AAF58532 | 1983 |
| | Hs | BIGI | Q9Y6D6 | 1849 |
| | | BIG2 | Q9Y6D5 | 1785 |
| | | GBFI | 092538 | 1859 |
| | Rn | BIGI | XP 232614 | 1987 |
| | | BIG2 | 07TSUI | 1791 |
| | | GBFI | XP_347197 | 1883 |
| Fungi | Ca | EAL04295 | EAL04295 | 1839 |
| 5 | | EAL02873 | EAL02873 | 1015 |
| | Nc | Q7SAX4 | EAA33549 | 1940 |
| | | O7SAL8 | EAA33457 | 1626 |
| | Sc | SEC7 | P11075 | 2009 |
| | | GEAL | P47102 | 1408 |
| | | GEA2 | P39993 | 1459 |
| | Sp | SC71 | O9UT02 | 1811 |
| | -F | SC72 | O9P7V5 | 1872 |
| | | Q9P7R8 | NP_596613 | 1462 |
| Viridiplantae | At | At1g01960 | Q9LPC5 | 1750 |
| • | | At3g43300 | NP 189916 | 1728 |
| | | At3g60860 | Q9LZX8 | 1793 |
| | | At4g35380 | O65490 | 1711 |
| | | At4g38200 | NP 195533 | 1698 |
| | | GNOM | O42510, Atlg13980 | 1451 |
| | | GNLI | O9FLY5. At5g39500 | 1443 |
| | | GNL2 | NP 197462. At5g19160 | 1375 |
| | 05 | 9631 m01366 | O85565 | 1789 |
| | 00 | 9630.m00920 | O9XGN9 | 1687 |
| | | 9634 m04029 | - | 1704 |
| | | 9635.m03752 | _ | 1680 |
| | | 9631.m04495 | | 1456 |
| | | 9630 m02122 | _ | 1306 |
| | | 9632.m00175 | Q7XTII | 1407 |
| Alveolata | Pt | GGGI | CR533425 | 1615 |
| | | GGG2 | CR533424 | 1628 |
| | | 6663 | CR533423 | 1598 |
| | | 6665 | CIJJJTZJ | |
| | | GGG4 | CR533422 | 1599 |

Table I: BIG and GBF protein sequences used in this study.

^a Unnamed sequences are designated by their NCBI accession number, AGI (Arabidopsis Genome Initiative) locus numbers for At and TIGR model temporary IDs for Os. BIG and GBF subfamily members are in normal and bold characters respectively, except for Pt members which have not been assigned to either subfamily (see also Figure 8). Species abbreviations are: Ag, Anopheles gambiae; Ce, Caenorhabditis. elegans; Dm, Drosophila melanogaster; Hs, Homo sapiens; Rn, Rattus norvegicus; Ca, Candida albicans; Nc, Neurospora crassa; Sc, Saccharomyces cerevisiae; Sp, Schizosaccharomyces pombe; At, Arabidopsis thaliana; Os, Oryza sativa; Pt, Parameciumtetraurelia).

they are involved in dimerization [3], recruitment of partners [24] and Golgi targeting [25], and in actin remodeling functions in the case of EFA6 [21]. On the other hand, although the functions of BIG and GBF subfamilies have been the subject of many investigations, their architecture is barely described, making it difficult to associate biochemical activities with their molecular structure. Here we investigate the domain architecture in the BIG and GBF subfamilies, including all sequences from fully annotated eukaryotic genomes and our novel annotation of Sec7-containing proteins from the *Paramecium tetraure-lia* alveolate. Sequence comparisons combined with secondary structures and splicing patterns analysis identifies five novel domains that are conserved between BIG and GBF subfamilies, thus unifying them as a higher order subfamily with a probable common ancestor. Our analysis of Sec7-domain containing sequences from *Paramecium* also introduces a novel subfamily of ArfGEFs unique to alveolates, which we call TBS (<u>TBC-Sec7</u>).

Results and discussion A conserved domain architecture in BIG and GBF subfamilies

The BIG and GBF subfamilies are the only ArfGEFs subfamilies common to all eukaryotes [8] and the sole Arf-GEFs present in plants [26] (Figure 1). They are therefore possible representatives of ancestral ArfGEF functions and may provide a model to understand the nature and implementation of activities associated with the exchange function carried by the conserved Sec7 domain. However, domain 'hunting' in BIG and GBF subfamilies was complicated by the facts that the Sec7 domain is their only domain that could be identified from known domain repertoires, and that their poorly characterized non-catalytic regions were not found outside these ArfGEF subfamilies. Alternatively, we based our search of candidate structural domains in BIGs and GBFs on the bioinformatics analysis of their own sequences, taking advantage of the growing number of sequences from fully annotated genomes from mammals, insects, plants, nematode, and fungi, to which we included our annotation of Sec7-containing proteins from the newly sequenced genome of Paramecium.

Multiple alignments of 42 sequences (listed in Table 1) revealed that the BIG and GBF subfamilies share an unexpected conserved architecture (schematized in Figure 2). Two homology domains are located in N-terminus of the Sec7 domain - the DCB (~150 aa) and HUS (Homology Upstream of Sec7, ~170 aa) domains - and three in its Cterminus -the HDS1 (Homology Downstream of Sec7, ~130 aa), HDS2 (~160 aa) and HDS3 (~120 aa) domains (Figure 3,4,5,6,7). In Arabidopsis GNOM, the DCB domain is included in an N-terminal region of ~250 residues involved in dimerization and possibly binding to cyclophilin5 and called the Dimerization/Cyclophilin Binding region [27], after which the new domain was named (Figure 3). All domains are predicted to have a high content of α -helices that co-align in the multiple sequence alignments, reinforcing the prediction of sequence similarities and suggesting that these domains form folded structural units that may share common functional features. Except for the N-terminal DCB domain



Figure 2

The common domain architecture of the BIG and **GBF** subfamilies. From N- to C-terminus : DCB , HUS, Sec7, HDS1, HDS2, HDS3. Linker regions of variable length and sequence are shown in grey, with alternate splicing sites in human GBF1, BIG1 and BIG2 in black, white and grey diamond shapes respectively. Interactions reported in the litterature are indicated in boxes of width corresponding to the mapped regions, except for myosin IXb interaction which was studied only with full-length BIG1. Arrows indicate predicted Protein kinase A-anchoring motifs. ¹ [45]; ² [27]; ³ [46]; ⁴ [47]; ⁵ [48]; ⁶ [49]; ⁷ [50]; ⁸ [51].

which is also found in the yeast protein Ysl2p [28], all of them are unique to these two ArfGEFs subfamilies within the detection limits of the BLAST search. The HUS domain features a remarkably conserved N(Y/F)DC(D/N) motif, which we call the HUS box, which is predicted to locate in a loop where it may be available for functional interactions (Figure 4). The N- and C-terminal ends of BIGs and GBFs are more variable, including an unusual enrichment in Asp/Glu or Pro residues in some members. A specific feature of BIG members is that their C-terminus is in general less variable than that of GBFs, and is predicted with a significant amount of secondary structures. In contrast to the predicted structural domains, the intervening regions are highly variable in length and do not yield aligned sequences. Analysis of their amino-acid composition reveals a paucity of hydrophobic residues which is predicted to associate with an essentially unfolded conformation, suggesting that they act as linkers to tether the functional domains together.

To further investigate the predicted organization of BIGs and GBFs in 6 conserved helical domains connected by variable linkers, splicing patterns of human BIGs and GBFs were analyzed in the large number of cDNAs and ESTs in the databases that correspond to GBF/BIG transcripts. This revealed the use of alternate splice donor and acceptor sites predicted to yield proteins with insertions and deletions ranging from 1 to 38 residues, and a number of splice variants arising from exon skipping (Table 2). Strikingly, all observed sequence variations occur in regions identified as linkers between conserved

| | عففف | يعفعه | ll | | | عففه | يععع | e | | | | |
|--|--|---|--|--|---|--|--|---|---|---|--|--|
| | i _ 1 | . <u>o</u> | 20 | | 3 <u>o</u> | | 4 | • • _ | | | | 5 <u>0</u> |
| BIG1_HUMAN/70-228 | SKTNFI <mark>E</mark> AC | KYFLP | FELACQ | KCPR | IVSTS | LDCL | OKLIA | YGHL | TGNAP | D | | .STTPG |
| BIG2_HUMAN/58-216 | PKANFIEAD | KYFLP | FELACQ | KSPR | VVSTS | LDCL | <u>KLI</u> A | YGHI | TGNAP | D | • • • • • • • • • • | .SGAPG |
| Q9VJW1_DROME/72-230 | DAASIINAE | TYFLP | FELACK | RSPR | IVVTA | | KLIA | YGHL | TGSIQ | D | ••••• | .SANPG |
| Q9XWG5_CAEEL/69-22/ | AGGTAVEAL | | FELACN | KSPK | TTEDA | | | UCVT | TGRGA | D | • • • • • • • • • • | . ISNPE |
| 1g01060 ARATH/77-233 | AFVCTAFCE | | TINACG | CVIK | | VDCT | | | RGESD | P | ••••• | - SGGAL |
| 4g35380 APATH/62_214 | SCLAASDAC | | FLIGIF | - AVCK | | TDCAF | K L F S | | RGEAD | r c | ••••• | SKO |
| 4g35560_ARATH/61-215 | FGLTTSDAD | | TILLSL | GVAK | VTEPA | LDCSI | KLFS | | RGEVC | s | ••••• | SSP |
| 3g43300 ARATH/97-252 | HTLGGAEVE | | LRLAFE | KNLK | TFDAA | LDCL | | YDHT. | EGDPG | T | | DGGKN |
| SEC7 YEAST/267-445 | NNPHYVDSI | LVFEA | LRASCR | KSSK | VOSLA | LDCL | KLFS | FRSL | DETLL | VNPPDSLAS | NDORODAADO | ITPPPK |
| GGG2 PARTE/36-189 | QIKDFY <mark>D</mark> AN | IHL <mark>L</mark> K <mark>V</mark> | YQQCIE | KQAK | LĨELA | LFDI | NIVD | QGYL | AGEQI | I | | GE |
| GBF1 HUMAN/54-215 | TELSEIEPN | IVF <mark>L</mark> RP | FLEVIR | EDTTGP | ITGL <mark>A</mark> | LTSV | KFLS | YAL <mark>I</mark> | DPTHE | G | | т |
| Q9V696_DROME/56-217 | EDLRQI <mark>E</mark> PQ | VFLAP | FLEVIR | ADATGP | LTSLA | LASV | I <mark>KLL</mark> S | YGL <mark>I</mark> | DPTSP | N | | L |
| Q9XTF0_CAEEL/56-248 | ADLADM <mark>N</mark> PQ | TY <mark>LSP</mark> | FLDVIK | QNTNGP | ITEA <mark>A</mark> | LAAV | KFLN | IYGL <mark>I</mark> | DASSI | к | | A |
| GNOM_ARATH/83-245 | QPWHTI <mark>S</mark> PM | ILY <mark>LQP</mark> | FLDVIR <mark>5</mark> | DETGAP | ITSI <mark>A</mark> | LSSV | KILN | ILNV <mark>I</mark> | DQNTA | N | | I |
| GNL1_ARATH/81-243 | S N W Q Y V <mark>D</mark> P R | LY <mark>IQP</mark> | FLDVIL | DETGAP | I T G V <mark>A</mark> | LSSV | (<mark>KIL</mark> T | LEV <mark>F</mark> | TLETV | N | | • • • • • V |
| GNL2_ARATH/67-228 | QDWRTIDPS | VYLSP | FLEVIQ | DEIPAS | ATAVA | LSSI | KILK | IEI <mark>F</mark> | DEKTP | G | • • • • • • • • • • | A |
| GEA2_YEAST/8/-248 | KNLDNIDSL | TILQP | FLLIVS | SSISGY | | | OKFF1 | | NESSQ | N | • • • • • • • • • • | ••••¥ |
| GEAI_YEAST/90-251 | KGLDSL <mark>N</mark> AL | ЕГГКБ | FLEIVS | SSVSGY | TTSLA | ГЛЗГС | <u>KVF</u> T | ткт | NKTFN | D | • • • • • • • • • • | •••• |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | 0000000000 | 00000 | Q | 0 0 | 0000. | | | | | | | 2.0 |
| | 60 | | 70 | | 80 | | | | | | 9 | 0 |
| BIG1 HUMAN/70-228 | KKT. TORTTF | TCGC | F OGI | PO. TDEG | | | | | | | TTKALLT | |
| BIG2 HUMAN/58-216 | KRLIDRIVE | TICSC | FOGI | $\mathbf{P}\mathbf{O}$. TDEG | VÕLÕ. | | | | | | . IIKALLTA | VTS.P |
| 09VJW1 DROME/72-230 | HLLIDRIVV | TIYG <mark>C</mark> | F | O. TDEA | võlõ. | | | | | | IIKALLTY | VTSO |
| 09XWG5 CAEEL/69-227 | RKLIDRIVE | AICAP | FLGC | G. TDET | VĨLÕ. | | | | | | LIKAVLAV | VLSŤ |
| 3g60860 ARATH/77-233 | SLLLFKLIC | SVCKC | HD. | L.G <mark>DE</mark> S | IELP. | | | | | | VLKTLLSA | INSI |
| 1g01960_ARATH/72-228 | ALLLSKL <mark>I</mark> E | Τ <mark>Ι</mark> СК <mark>С</mark> | не | L.D <mark>DE</mark> G | LELL. | | | | | | VLKTLLTA | VTS |
| 4g35380_ARATH/62-214 | DSILFKL <mark>V</mark> N | I A <mark>V</mark> S K <mark>V</mark> | GA | AI.A <mark>EE</mark> P | IQLA. | | | | | | VLRVLLAA | A <mark>VRS</mark> P |
| 4g38200_ARATH/61-215 | DSLLYKL <mark>I</mark> H | I A <mark>I</mark> C K <mark>V</mark> | с | GI.G <mark>EE</mark> S | IELA. | | | | | | VLRVLLA | A <mark>VRS</mark> P |
| 3g43300_ARATH/97-252 | SAPFTDI <mark>L</mark> N | IM <mark>V</mark> CS <mark>C</mark> | VDNS | SS <mark>PD</mark> S | TVLQ. | | • • • • | | | | VLKVLLTA | A <mark>VAS</mark> G |
| SEC7_YEAST/267-445 | QKIIDAA <mark>I</mark> C | TISDC | FQGI | EG.T <mark>DD</mark> R | VELQ. | • • • • • | • • • • | • • • • | • • • • • | | IVRALSSC | LEEDS |
| GGG2_PARTE/36-189 | KRAIEIALD | LVMQT | QI | E.KEET | VQIH. | • • • • • | • • • • | • • • • | • • • • • | | MIKALQAJ | MTN.K |
| GBF1_HUMAN/54-215 | AEGMENMAL | AVTHA | RFVGTDI | A.SDEV | VLMK. | • • • • • | • • • • | • • • • | • • • • • | • • • • • • • • • | · · · · · · · · · · · · · · · · · · · | |
| Q9V696_DROME/56-21/ | ADIVERIAL | | RFMGTDQ | | TFMR. | C D V V I | MEUV | | NT T E M | CONFIDEN | | |
| CNOM APATH/83_245 | FDAMHT.VVC | | REFUTDI | DA SERVI | | SKKKV | NITIN | | ИГГЕЫ | CONFIRSAN | | и и и и и и и и и и и и и и и и и и и |
| GNU1 ARATH/81-243 | GEAMHITVE | AVKSC | RFEVTDE | PA.SEEV | VI.MK | | | | | | TLOVILLAC | |
| GNL2 ARATH/67-228 | KDAMNSIVS | GITSC | RLEKTDI | V.SEDA | VMMR. | | | | | | . ILOVLTGI | MKH.P |
| GEA2 YEAST/87-248 | IGAHRATVN | ALTHC | RFEGSO | L.SDDS | VLLK. | | | | | | . VVFLLRSI | VDS.P |
| GEA1 YEAST/90-251 | OIAVRETV | ALTHC | RFEASK | JI.SDDS | VLLK. | | | | | | VVTLLRDI | ITSS |
| _ | ~ _ | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | 0.0 | | 000000 | | 0 | 0000 | | | | | | |
| | 100 | | 110 | 4 | 120 × | | 130 | | 140 | 15 | 0 | |
| BTC1 HUMAN /70 229 | UT ETUECO | | | | | | | MT NIX7 | TENDM | PNOATOFAK | OMEREDUDOI | T |
| $BIGI_HUMAN/70-228$ $BIG2_HUMAN/59-216$ | UT ETUECT | | | LAS. K | | | | | | ENQUE OF A P | DIEKEKIKUI | 1 |
| 09V.W1 DROME /72_230 | HV FTHEFT | | RTCVDI | LAS. K | | | | MT.NV | TFARM | ENOVVEL DE | DNSNDTNCS | - |
| 09XWG5 CAEEL/69-227 | HC EVHGAS | T.TT.AV | RTCFNIN | | SPINO | | | VTNT | VEGNM | EKEGNIKDD | ETTVREVVEN | 7 |
| 3g60860 APATH/77-233 | | | I I U I I I I | TTD • • 1 | | UT UT | | | | | | |
| | SL. RTHGKC | T. T. T. V V | RTCYDT | T.G.S K | NVVN <mark>O</mark> | ттаки | ST. TO | TLVT | VFRRM | EADSSTVPT | OPTVVAELME | 2 |
| 1g01960 ARATH/72-228 | SL.RIHGKC | LLLVV | R T C Y D I Y R T C Y G I Y | LGSK | N V V N <mark>O</mark> N V V N O | TT <mark>AK</mark> ATAK | SLVO | ILVI | VFRRM | EADSSTVPI | QPIVVAELME OPIVVAELME | 2 |
| 1g01960_ARATH/72-228 4g35380_ARATH/62-214 | SL.RIHGKC SL.RIHGDS CI.LIRGDC | LLLV <mark>V</mark> LLQIV LLHVV | RTCYDI RTCYGI KTCYNI | LGSK LGSR LGGL | NVVN <mark>Q</mark> NVVN <mark>Q</mark> SGTTO | TT <mark>AK</mark> AT <mark>AK</mark> IC <mark>AK</mark> S | SLIQ SLVQ VLAO | ILVI MSVI MMLV | VFRRM VFRRM IFTRS | EADSSTVPI EADSSTVPI EEDSLDVSV | QPIVVAELME QPIVVAELME KTIYVNEL. | 5 |
| 1g01960_ARATH/72-228 4g35380_ARATH/62-214 4g38200_ARATH/61-215 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC | LLLVV LLQIV LLHVV LLHLV | R TCYDIY RTCYGIY KTCYNIY RTCYNVY | LGSK LGSR LGGL LGGF | N V V N <mark>0</mark> N V V N 0 S G T T 0 N G T N 0 | TT <mark>AK</mark> ATAK ICAK ICAK | SLIQ SLVQ VLAQ VLAQ | ILVI MSVI MMLV IMLI | VFRRM VFRRM IFTRS VFTRS | EADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL | QPIVVAELME QPIVVAELME KTIYVNEL KTVNVNDLL | 5 5 4 |
| 1g01960_ARATH/72-228 4g35380_ARATH/62-214 4g38200_ARATH/61-215 3g43300_ARATH/97-252 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF | LLLVV LLQIV LLHVV LLHLV | RTCYDIY RTCYGIY KTCYNIY RTCYNVY RVCYNIZ | LGSK LGSR LGGL LGGF | NVVN <mark>Q</mark> NVVNQ SGTTQ NGTNQ .PINQ | TT <mark>AK</mark> ATAK ICAK ICAK AT <mark>SK</mark> | SLIQ SLVQ VLAQ VLAQ NLAQ MLTQ | ILVI MSVI MMLV IMLI MISI | VFRRM VFRRM IFTRS VFTRS VFRRM | EADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL ETDIVSASS | QPIVVAELMH QPIVVAELMH KTIYVNEL. KTVNVNDLLA TVSQEEHVS(| E A G |
| Jg01960_ARATH//72-228 4g35380_ARATH/62-214 4g38200_ARATH/62-214 3g43300_ARATH/61-215 3g43300_ARATH/97-252 SEC7_VEAST/267-445 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS | LLLVV LLQIV LLHVV LLHLV LLGVI LLGVI | RTCYDI RTCYGI KTCYNI RTCYNV RVCYNI RTIYN <mark>V</mark> | LGSK LGSR LGG.L LGG.F LGSF LNS VFS.L | N V V N <mark>Q</mark> N V V N Q S G T T Q N G T N Q . P I N Q N P S N Q | TT <mark>AK</mark> ATAK ICAK ICAK AT <mark>SK</mark> GI <mark>AQ</mark> | SLIQ SLVQ VLAQ VLAQ MLTQ TLTQ | ILVI MSVI MMLV IMLI MISI | VFRRM VFRRM IFTRS VFTRS VFRRM VFRRM | EADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL ETDIVSASS DLKQSTSSA | QPIVVAELMH QPIVVAELMH KTIYVNEL. KTVNVNDLLA TVSQEEHVSC VSLSTKNHQC | E E A G Q |
| Jg01960 ARATH/72-228 4g35380 ARATH/62-214 4g38200 ARATH/61-215 3g43300 ARATH/97-252 SEC7 YEAST/267-445 GGG2_PARTE/36-189 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES | LLLVV LLQIV LLHVV LLHLV LLGVI LLKAI VTRVF | RTCYDIY RTCYGIY KTCYNIY RTCYNY RVCYNI RTIYNY SLLINL | LGSK LGSR LGGL LGGF ALNS VFSL ISVSK | NVVN SGTT SGTT NGTN O .PIN O NPSN IVAI | TT <mark>AK</mark> AT <mark>AK</mark> ICAK S ICAK GIAO K NA <mark>SK</mark> | SLIQ SLVQ VLAQ VLAQ MLTQ TLTQ CLCQK | LVI MSVI MMLV IMLI MISI ISS IVST | VFRRM VFRRM IFTRS VFTRS VFRRM VYDKI YFARL | EADSSTVPI EADSSTVPI E EDSLDVSV E ANSMDASL E TDIVSASS DLKQSTSSA E DYGILAED | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLLI TVSQEEHVSC VSLSTKNHQQ EYQLAIQQQC | 2 |
| Jg01960_ARATH/72-228 4g35380_ARATH/72-228 4g38200_ARATH/62-214 4g38200_ARATH/61-215 SEC7_YEAST/267-445 GGC2_PARTE/36-189 GBF1_HUMAN/54-215 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES | LLLVV LLQIV LLHVV LLHLV LLGVI LLKAI VTRVF | RTCYDIY RTCYGIY KTCYNIY RTCYNY RTCYNY RTIYNY SLLINL QSCFRIO | LGSK LGSR LGGL LGGF LGGF LNS VFSL ISVSK CFEMR.L | NVVNQ SGTTQ NGTNQ .PINQ NPSNQ IVAII SELLR | TTAK ATAK ICAKS ICAKS ATSK GIAQ KSAE | SLIQ SLVQ VLAQ VLAQ MLTQ TLTQ ACQK ITLVD | ILVI MSVI MMLV IMLI ISI ISS IVST VQL | VFRRM VFRRM IFTRS VFTRS VFRRM VYDKI YFARL LFTRL | E ADSSTVPIE ADSSTVPIE ADSSTVPIE ADSLDVSVE ANSMDASLE TDIVSASSDLKQSTSSAE DYGILAEDP QFKEEPKN | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLLA TVSQEEHVSC VSLSTKNHQC EYQLAIQQC YVGTNMKKL | |
| Jg01960_ARATH/72-228 4g35380_ARATH/62-214 4g38200_ARATH/62-214 3g43300_ARATH/61-215 3g43300_ARATH/97-252 SEC7_YEAST/267-445 GG2_PARTE/36-189 GBF1_HUNAN/54-215 Q9V696_DROME/56-217 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES EGAAVSNVS | LLLVV LLQIV LLHVV LLHLV LLGVI LLKAI VTRVF VCEIM MCEVM | RTCYDIY RTCYGIY KTCYNIY RTCYNYY RTCYNYY RTIYNY SLLINLH QSCFRIG LSCFKIS | LGSK LGSR LGGL LGGF LGGF LSSL SVFSL SVFSK CFEMR.L SFEPR.L | NVVNQ SGTTQ NGTNQ .PINQ NPSNQ IVAII SELLR SELLR | TTAK ATAK ICAKS ICAKS ATSK GIAO KSAEH RSAEH | SLIQ SVLAQ VLAQ MLTQ ATLTQ ATLTQ KLVD SLKD | ILVI MSVI MMLV IMLI ISI ISS IVST MVQL MVLL | VFRRM IFTRS VFTRS VFTRS VFRRM VYDKI YFARL LFTRL FFMRL | EADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL ETDIVSASS DLKQSTSSA EDYGILAED PQFKEEPKN PQFAEERSD | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLLA TVSQEEHVSO VSLSTKNHQO EYQLAIQQQO YVGTNMKKLI TMLQKRFIIO | |
| Jg01960 ARATH/72-228 4g35380 ARATH/62-214 4g38200 ARATH/62-214 3g43300 ARATH/61-215 3g43300 ARATH/67-252 SEC7_YEAST/267-445 GGG2_PARTE/36-189 GBF1_HUMAN/54-215 Q9V696_DROME/56-217 Q9V560_DROME/56-217 | SL.RIHGKC SL.RIHGDS CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES EGAAVSNVS PGILLSNEA | LLLVV LLQIV LLHVV LLHLV LLGVI LLKAI VTRVF VCEIM MCEVM VCCDM | RTCYDIY RTCYGIY KTCYNIY RTCYNVY RVCYNIZ SLLINLE QSCFRIC LSCFKIS QSCFRIC | LGSK LGSR LGGL LGGF LNSL VFS.L ISVSK CFEMR.L SFEPR.L VFEQN.L | NVVNQ SGTTQ SGTTQ NGTNQ NGTNQ NPSNQ IVAII SELLR SELLR SELLR | TTAK ATAK ICAKS ICAKS ATSK GIAO KSAEH RSAEH KSAEH | SLIQ VLAQ VLAQ VLAQ ALTQ ALTQ ACQK ITLVD SLKD | ILVI MSVI MMLV IMLI ISI ISS VST MVQL MVLL MTQL | VFRRM IFTRS VFTRS VFTRS VYDKI YFARL LFTRL FFMRL IFTRL | ADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL ETDIVSASS DLKQSTSSA EDYGILAED PQFREEPKN PQFAEERSD PTVEDTRH | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLLA TVSQEEHVSC VSLSTKNHQQ EYQLAIQQQ VVGTNMKKLI TMLQKRFTIC PYIRQLVNP7 | |
| Jg01960 ARATH/72-228 4g35380 ARATH/62-214 4g38200 ARATH/61-215 3g43300 ARATH/61-215 3g43300 ARATH/97-252 SEC7 YEAST/267-445 GGG2 PARTE/36-189 GBF1_HUMAN/54-215 Q9VE96 DROME/56-217 Q9VE96 DROME/56-218 GMV 1-2020 | SL.RIHGKC SL.RIHGDC CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES EGAAVSNVS PGILLSNEA ASVMLSNEA | LLLVV LLQIV LLHVV LLHLV LLGVI LLKAI VTRVF VCEIM MCEVM VCCDMM VCCVV | R TCYDIY RTCYGIY KTCYNIY RTCYNYY RVCYNIZ RTIYNVY SLLINLE QSCFRIQ SCFRIQ QSCFRIY NTCFRV | YLGSK YLGSR YLGGL YLGGF YLSSL YVFSL YFSYSK XFEMR.L XFEQN.L YFEQN.L YHQAGMK | NVVN SGTT SGTT NGTN PIN SELL SELL SELL SELL SLLL R CELL | TTAK ATAK ICAKS ICAKS ATSK GIAQ KSAE KSAE KSAE KAAE RVAR | SLIQ VLAQ VLAQ MLTQ ATLTQ ACQK ILVD SLKD TLAD | ILVI MSVI MMLV IMLI ISS ISS MVQL MVLL MTQL | VFRRM VFRRS VFTRS VFTRS VFTRRM VYDKI YFARL LFTRL FFMRL IFFTRL IFSCH | EADSSTVPI EADSSTVPI EADSLDVSV EANSMDASL ETDIVSASS DLKQSTSSA EDYGILAED PQFKEEPKN PQFAEERSD PTFVEDTRH PDVERTETT | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLLA TVSQEEHVSG VSLSTKNHQG EYQLAIQQQQ YVGTNMKKLI TMLQKRFTIC PYIRQLVNPT | |
| Jg01960-ARATH/72-228 4g35380 ARATH/62-214 4g38200_ARATH/61-215 3g43300_ARATH/97-252 SEC7_YEAST/267-445 GGG2_PARTE/36-189 GBF1_HUMAN/54-215 09XTF0_CAEEL/56-217 09XTF0_CAEEL/56-248 GNOM_ARATH/83-245 GNU_2ARATH/81-243 | SL.RIHGKC SL.RIHGKC SL.RIHGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS VGAHLTNES EGAAVSNVS PGILLSNEA ASVMLSNQH ASVMLSNQH | LLLVV LLQIV LLHVV LLHUV LLHUV LLKAI VTRVF VCEIM VCEVM VCTVV ICTIV | R TCYDIY RTCYGIY RTCYNIY RTCYNYY RTCYNYY RTIYNY SLLINLE QSCFRIQ SCFRIQ SCFRIV NTCFRV NTCCRV | YLGSK YLGSR YLGGL YLGGL YLGSL YVFSL ISVSK CFEMR.L SFEPR.L SFEPR.L YFEQN.L YHQAGMK YHQASSSK | NVVN SGTT SGTT NGTN PIN SELL SELL SELL SELL SELL SELL SELL SEL | TTAK ATAK ICAKS ATSK GIAQ KSAEH KSAEH KSAEH KAAES RVARH RIAR | | ILVI MSVI MLV IMLI ISI ISS IVST MVQL MVLL VRC LVRC | VFRRM VFRRS VFTRS VFTRS VFTRRM VYARL LFTRL FFMRL IFSHL IFSQL | EADSSTVPI EADSSTVPI EEDSLDVSV ETDIVSASS DLKQSTSSA DLKQSTSSA EDYGILAED PQFKEEPKN PQFAEERSD PTFVEDTRH PDVERTETT PFISPLANE | QPIVVAELMH QPIVVAELMH KTIYVNEL KTVNVNDLLA TVSQEEHVSG VSLSTKNHQG EYQLAIQQQQ YVGTNMKKLH TMLQKRFTIC PYIRQLVNP7 LVNRAGSIKG CELHVDNKVC | |
| Jg01960 ARATH/72-228 4g35380 ARATH/62-214 4g38200 ARATH/61-215 3g43300 ARATH/61-215 3g43300 ARATH/97-252 SEC7_YEAST/267-445 GGS2_PARTE/36-189 GBF1 HUNAN/54-215 Q9V696 DROME/56-217 Q9XTF0_CAEEL/56-248 GNOM_ARATH/83-245 GNL1_ARATH/81-243 GNL2_ARATH/67-228 CP22_VEST/67_248 | SL.RIHGKC SL.RIHGDC CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES PGILLSNEA ASVMLSNQE SSELLEDQ SSELLEQU | LLLVV LLQIV LLHVV LLHVV LLKAI VTRVF VCEIM VCEVM VCCTVV ICTIV | R TCYGIY RTCYGIY KTCYGIY RTCYNIY RTIYNVF SLLINLF QSCFRIG QSCFRIG QSCFRIV NTCLRV NTCLRV | LLGSK LLGSR LLGGF LLGGF NLNS VFSSL ISVSK SFEPR.L VFEQN.L VFQQN.L VHQAGMK VHQASSK VQQSTGR | NVVN SGTT SGTT NGTN O NPSN SELL SELL SELL SELL SELL SELL SELL SE | TTAK ATAK ICAKS ICAKS GIAO KSAE KSAE KSAE KAAE KAAE RVAR RIAR RNAR | S L I Q S L VQ S V L A Q S V L A D S S V L A S V | LUVI MSVI MLV MLV INLI ISS ISS VST MVQL MVLL LVRC LIRC LIQI | VFRRM VFRRM VFTRS VFTRRM VFRRL LFTMRL LFFMRL LFFSQL LFSQL LFSQL | EADSSTVPI EADSSTVPI EEDSLDVS ETDIVSASS DLKQSTSSA EDYGILAED PQFKEEPKN PQFAEERSD PTFVEDTRH PDVERTETT PFISPLANE PDFEVRGDE | QPIVVAELMI QPIVVAELMI KTIYVNEL. KTVNVNDLL TVSQEEHVSG VSLSTKHQC EYQLAIQQQC YVGTNMKKLI TMLQKRFTIC PYIRQLVNPT LVNRAGSIKG CELHVDNKVC GGEDSESDT | 2 |
| Jg01960 ARATH/72-228 4g35380 ARATH/72-228 4g35380 ARATH/62-214 4g38200 ARATH/61-215 3g43300 ARATH/97-252 SEC7_YEAST/267-445 GG2_PARTE/36-189 GBF1_HUMAN/54-215 G9V696_DROME/56-217 G9V576_DROME/56-217 G9XTF0_CAEEL/56-248 GNU_ARATH/83-245 GNL1_ARATH/81-243 GNL2_ARATH/67-228 GEA1_YEAST/87-248 GEA1_YEAST/87-248 | SL.RIHGKC SL.RIHGDC CI.LIRGDC RI.LIRGDC KF.KVHGEF SS.LCHGAS KH.HIYGES VGAHLTNES EGAAVSNVS PGILLSNEA ASNGLSNOD SSELLEDQA YGDLLSNE | LLLVV LLQIV LLHVV LLHVV LLGVI LLKAI VTRVF WCEIM VCDMM VCDMM VCTVV VCTIV VCTIV | R TCYGIY RTCYGIY KTCYGIY RTCYNIY RTCYNY RTIYNY SLLINL SCFRIG SCFRIG SCFRIG SCFRIG NTCLRV NTCLRV NTCLRV NTCLQ | LLGSK LLGSR LLGGF LLGGF NFSSK SVSSK SFEPR.L SFEPR.L FEQN.L VHQSSSK VQQSTGR ACNTO | NVVNQ SGTTQ SGTTQ NGTNQ .PINQ NPSNQ SELLR SELLR SELLR GELLQ SELLQ SELLQ SELLQ | TTAK ATAK ICAKS ICAKS ATSK GIAQZ NASK KSAE RSAE RXAE RXAE RVAR RIAR RNGR NAAQS | SLIQ SLVQ SVLAQ SVLAQ LALTQ LACQKK ITLVD STLAD ITMHE ITMHE TMHE STMIA | LVI MSVI MMLVV IMLI IISS IVST MVQLL VTVK VTVK | VFRRM VFRRS VFRRS VFRRKI VFRARL VFRARL LFFMRLL IFFSS LFFSSK IFFSSK | EADSSTVPI EADSSTVPI EEDSLDVSV EANSMDASL ETDIVSASS DLKQSTSSA DLKQSTSSA DQFKEEPKN PQFAEERSD PGFAEERSD PTFVEDTRH PDVERTETT PFISPLATE PDFEVRGDE KTIEPVNVN | QPIVVAELMI QPIVVAELMI XTIYVNEL. KTVNVNDLLA TVSQEEHVSG EYQLAIQQQQ YVGTNMKKLI PYIRQLVNPT LVNRAGSIKG CELHVDNKVG GGEDSESDT QIYINDEST | 2 2 2 2 2 2 3 3 2 3 3 3 3 3 3 3 3 3 3 3 |

The conserved domains of the BIG/GBF subfamily: DCB domain. Multiple sequence alignement of the conserved domains from BIG and GBF representative sequences showing secondary structure predictions that co-align in all sequences. Colour coding is red for invariant residues, yellow for a sequence similarity score threshold of 0.15 using the BLOSUM62 matrix. The gap in helix 4 is due to an insert in the drosophila Q9V696 sequence, and may be resulting from a sequence annotation error.

| | 100000 | | 3.0 | ell | | 60 |
|-----------------------|---|--|---|---|---|--|
| | 1 10 | 20 | 30 | 4 Q | 50 | 6 Q |
| BIG1_HUMAN/411-593 | PGAKFSHILQKDAFL | VFRSLCKLSMKPI | SDGPPDI | PKSHELRSKIL | SLOLLSILO | AGPIFRTNEM |
| BIG2_HUMAN/362-544 | VAARFSHVLQKDAFL | VFRSLCKLSMKPI | | PKSHELKSKVV | | AGPVFRTHEM |
| 09XWG5 CAEEL/264-443 | DOFTEMNAYOKDAFL | VERALCILAOKEI | GG A S | ENEMSLESKUL | ALEMILLUVION | |
| 3g60860 ARATH/337-519 | LEVOIENKLRRDACL | VFRALCKLSMKAI | PKESS.A | ADPOSMRGKIL | ALELLKILLEN | IAGAVERTSEK |
| 1g01960 ARATH/328-507 | SEVOIGNKLRRDAFL | VFRALCKLSMKT | PKE | .DPELMRGKIV | ALELLKILLEN | AGAVFRTSDR |
| 4g35380 ARATH/284-466 | SETGDMSKVRQDAFL | LFKNLCKLSMRFS | SKENN.I | DDQIMVRGKTL | SLELLKVIID | IG <mark>G</mark> SVWRTNES |
| 4g38200_ARATH/263-456 | EDEGTGSKIREDGFL | <mark>LFK</mark> N <mark>LC</mark> K <mark>L</mark> SMKFS | SQENT.I | DDQILVRGKTL | <mark>S L</mark> ELLKVIID | IG <mark>G</mark> PIWLSDER.QLTLP |
| 3g43300_ARATH/324-509 | IELESMSIGQRDALL | <mark>V F R</mark> T <mark>L C</mark> K <mark>M</mark> G M K E I | S | .DEVTTKTRIL | S <mark>L</mark> E <mark>LL</mark> QGMLE(| GV <mark>S</mark> HSFTKNFH |
| SEC7_YEAST/488-678 | IAITNQDLAVKDAFL | VFRV <mark>MA</mark> KICAKPI | ETELDN | MRSHAVRSKLL | SLHIIYSIIKI | DH <mark>I</mark> DVFLSHNI.FLP |
| GGG2_PARTE/318-500 | SHSTFSEQYVKDAYE | ILEMLCQLSQRDI | PQNPQ1 | LAQMIIKCKVL | SLELIYEALAQ | OSDTTILQHKPK |
| GBF1_HUMAN/392-500 | DUTELEDVCIDETOF | LF KFLISLTNPHI | v K | ONCDCMMUT | CI CI CI TUTUALES | AP. VAUCUT |
| 09XTF0 CAEEL/377-558 | GGEEKMPYGLPCCRE | I. I. R. F. I. T. T. M. T. N. P. V. I. | R | HNTESMVIL | GINLLTVALE | TADFLENYDT |
| GNOM ARATH/305-491 | LHIMTEPYGVPSMVE | IFHFLCSLLNVVE | HVGMGSRSN | FIAFDEDVPLF | ALNLINSAIEI | GGSSIRHHPR |
| GNL1 ARATH/306-492 | ENAMMAPYGIPCMVE | I F H F L C T L L N V G F | NGEVNSRSNI | PIAFDEDVPLF | ALGLINSAIEI | GGPSFREHPK |
| GNL2_ARATH/233-414 | MSGGYGIRCCID | <mark>IFH</mark> F <mark>LC</mark> S <mark>L</mark> LNVVE | VVENLEGTN | VHTADEDVQIF | ALVLINSAIEI | S <mark>G</mark> DAIGQHPK |
| GEA2_YEAST/320-508 | QAYADDNYGLPVVRQ | <mark>YLN</mark> L <mark>LL</mark> SLIAPE. | NH | ELKHSYSTRIF | G <mark>L</mark> ELIQTALE | S <mark>G</mark> DRLQLYPR |
| GEA1_YEAST/305-490 | AENVEPNYGITVIKD | YLGLLLSLVMPE. | NI | RMKHTTSAMKL | SLQLINAA <mark>IE</mark> I | S <mark>G</mark> DKFPLYPR |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | للالالالالالالالالالالالالالالالالالالال | lll | فعفعفعفع | 200000000000 | فعفعفعفع | LELLELL |
| | 7 <u>0</u> | 80 | 9 Q | 100 | 110 1 | 20 |
| BIG1_HUMAN/411-593 | FINAIKQYL | CVA <mark>L</mark> SKNG.V <mark>S</mark> SV | PE <mark>VF</mark> ELSLS | I F L T <mark>L</mark> L S N <mark>F K</mark> T | HLKMQIEVFF | KE <mark>I</mark> FLY <mark>I</mark> LET |
| BIG2_HUMAN/362-544 | <mark>FI</mark> NA <mark>IK</mark> QY <mark>L</mark> | CVA <mark>L</mark> SKNG.V <mark>S</mark> SV | 'PD <mark>VF</mark> ELSLA <mark>I</mark> | I F L T <mark>L</mark> L S N <mark>F K</mark> M | H <mark>L K M Q I E V F F</mark> F | KE <mark>I</mark> FLN <mark>I</mark> LET |
| Q9VJW1_DROME/304-486 | FIMAIKQYL | VALSNNG.V <mark>S</mark> LV | PEVFELSLS | I F V A L L S N F K V | HLKRQIEVFF | KE <mark>I</mark> FLN <mark>I</mark> LEA |
| Q9XWG5_CAEEL/264-443 | | MALTRNA.VSNN | I LOVFEKSLA | LFVELLDKFKT | HLKASIEVFF | ISVILPMLDS |
| 1 01060 ARATH/33/-519 | FICATVOY | T GT T VNG AGNT | MITFOLSCS | LF I SLVARF KA | GLKAEIGVFFF | |
| 4g35380 ARATH/284-466 | FINAVKOVI | CLSLLKNS AVSI | MSTFOLOCA | IFMSLISKIRS | VLKAEIGTEE | |
| 4g38200 ARATH/263-456 | POKICRFLNAIKOLL | LSLLKNS.ALSV | MSIFOLOCA | IFTTLLRKYRS | GMKSEVGIFF | MLVLRVLEN |
| 3g43300 ARATH/324-509 | FIDSVKÂYL | SYALLRAS.V <mark>S</mark> QS | SVIFQYÁSG | IFSVLLLRFRD | SLKGEIGIFF | PIIVLRSLDN |
| SEC7_YEAST/488-678 | GKERVC <mark>FI</mark> DS <mark>IR</mark> QY <mark>L</mark> | <mark>R</mark> LV <mark>L</mark> SRNA.A <mark>S</mark> PI | AP <mark>VF</mark> EVTLE | IMWL <mark>L</mark> IAN <mark>LR</mark> A | DFVKEIPVFL | TE <mark>I</mark> YFP <mark>I</mark> SEL |
| GGG2_PARTE/318-500 | LISILKEQL | <mark>L</mark> ES <mark>L</mark> LKNS.L <mark>S</mark> AH | KQ <mark>LL</mark> ILTLN | IFIQ <mark>L</mark> IWR <mark>VR</mark> S | H L K K E L E A L I I | EN <mark>V</mark> YFK <mark>F</mark> LES |
| GBF1_HUMAN/392-566 | | CRHLFQLL.SIEF | LNLYAASLR | CFLLFESMRE | HLKFQMEMYI | KKLMEIITVE |
| Q9V696_DROME/356-532 | | CRNLISLL.SSEP | | CFLLFESLRG | HLKFQLEAYL | |
| CNOM APATH/305_491 | | ERNIMORG LSMS | | | FLKLOLEAFE | CVILPIACC |
| GNL1 ARATH/306-492 | LUTLIODDL | FCNLMOFG.MSMS | PLILSTVCS | IVLNLYLNLRT | ELKVOLEAFFS | VULLETAOS |
| GNL2 ARATH/233-414 | LRMVODDL | FHHLIHYG.ASSS | PLVLSMICS | LNIYHFLRK | FMRLOLEAFFS | SF <mark>V</mark> LLR <mark>V</mark> TAF |
| GEA2_YEAST/320-508 | LFTLISDPI | FKS <mark>I</mark> LFIIQN <mark>T</mark> TH | LS <mark>LL</mark> QATLQ | LFTTLVVILGN | NLQLQIELTL | R <mark>I</mark> FSI <mark>L</mark> LDDGTANNSS |
| GEA1_YEAST/305-490 | LFSL <mark>IS</mark> DPI | <mark>F</mark> KS <mark>V</mark> LFIIQS <mark>S</mark> TÇ | Y S <mark>L L</mark> Q A T L Q <mark>I</mark> | LFTSLVVILGD | YLPMQIELTL | RR <mark>I</mark> FEI <mark>L</mark> EDTTISG |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | فقفافك | lll | llll | | لفعففه | LLLLL |
| | 130 1 | 4 O | 150 | 16 Q | 170 | 180 |
| BIG1_HUMAN/411-593 | STS.SFDH <mark>K</mark> WMV | <mark>I</mark> QT <mark>I</mark> | TR <mark>IC</mark> .AD <mark>A</mark> QS | S V V D I Y V N Y D C | DLNAA <mark>NIFE</mark> RI | VNDLSKIAQGR |
| BIG2_HUMAN/362-544 | STS.SFEH <mark>R</mark> WMV | Ι QΤ <mark>Ι</mark> | TRIC.AD <mark>A</mark> QC | CVVDIYVNYDC | DLNAA <mark>NIFE</mark> RI | VNDLSKIAQGR |
| Q9VJW1_DROME/304-486 | NSS.SFEHKWMV | <u>IQ</u> AI | TRIC.ADAQS | SVVDIYVNYDC | DFSAANLFERI | |
| Q9XWG5_CAEEL/264-443 | NTC.AFEQKWIV | | GKIL ANPOS | SVVDMFVNYDC | DWISPNLFKS | |
| 1g01960 ARATH/328-507 | | | | | | |
| 4g35380 ARATH/284-466 | VLOPSYLOKMTV | | DKMS.ODPOI | LMVDIFVNYDC | DVESSNILERI | VNGLLKTALGP |
| 4g38200 ARATH/263-456 | VLQPSFVQKMTV | LSLI | ENIC.HDPNI | LIIDIFVNFDC | DVESPNIFERI | VNGLLKTALGP |
| 3g43300_ARATH/324-509 | SECPN.DQ <mark>K</mark> MGV | <mark>L</mark> RYNIFLLVQMM <mark>I</mark> | EK <mark>VC</mark> .KD <mark>P</mark> QN | M <mark>LVD</mark> V <mark>YV</mark> NYDC | D L E A P <mark>N L F E</mark> R N | IV <mark>T</mark> T <mark>LSK</mark> I <mark>A</mark> QGS |
| SEC7_YEAST/488-678 | TTS.TSQQ <mark>K</mark> RYF | <mark>[</mark> Sv <mark>]</mark> | QR <mark>IC</mark> .ND <mark>P</mark> R1 | ſ <mark>ĹVE</mark> FYĹ <mark>Ŋ</mark> Y DC | N P G M P <mark>N V M E</mark> I 1 | V D Y L T R L A L T R |
| GGG2_PARTE/318-500 | SNS.SFDH <mark>K</mark> QYT | <u> </u> | NKIL.TRPKV | VVIEIFVNYDC | SVGQNNLLKKI | IDMOCRIIOGR |
| GBF1_HUMAN/392-566 | NPKMPYEMKEMA | цеА | VQLW.RIPSH | Y TELYINYDC | DYYCSNLFEEI | TKLLSKNAFPV |
| Q7V090_DROME/350-532 | NCCCCTTOTTA | | | VTELIINYDC | DIVCONTEED | |
| GNOM ARATH/305-491 | KYGPSYOOOEVA | | VNFC. ROKST | EMVEMYAND | DITCSNVEEDI | SNTTSKSTFPV |
| GNL1 ARATH/306-492 | KHGSSYOOOEVA | MEAT | VDLC.ROHTH | FIAEVFANFDC | DITCSNVFEDV | SNLLSKNAFPV |
| GNL2_ARATH/233-414 | TGFLPLOEVA | L <mark>EG</mark> | IN <mark>FC</mark> .RQPAI | F <mark>IVE</mark> A <mark>YVNYDC</mark> | DPMCRNIFEET | GKVLCRHTFPT |
| GEA2_YEAST/320-508 | SENKNKPS.II <mark>K</mark> ELL | <mark>I</mark> EQ <mark>]</mark> | S I <mark>L W</mark> T R S <mark>P</mark> S I | F <mark>FTS</mark> T <mark>FINFDC</mark> | NLDRA <mark>DVSI</mark> NH | LKALTKLALPE |
| GEA1_YEAST/305-490 | DVSKQKPP.AI <mark>R</mark> ELI | <mark>I</mark> EQ <mark>I</mark> | SI <mark>LW</mark> IHS <mark>P</mark> AH | F <mark>FLQ</mark> L <mark>FVNFDC</mark> | NLDRS <mark>DLSI</mark> DH | 「I <mark>K</mark> E <mark>L</mark> T <mark>K</mark> F <mark>S</mark> LPA |
| | | | | | | |

The conserved domains of the BIG/GBF subfamily: HUS domain. See Figure 3 legend for alignment details. The highly conserved HUS motif is boxed in blue. The gap in helix 5 domain is due to an insert in the Arabidopsis 3g43300 sequence, and may be resulting from a sequence annotation error.

domains (Figure 2). Together with our domain analysis, this suggests that splicing at non-canonical exon/intron boundaries is only tolerated in regions of the protein where the impact upon folding of domains with essential function would be minimal.

| | 1 10 20 | مععقومه | | 222222222222222222222222222222222222222 |
|---|---|---|--|---|
| BTC1 UIMAN /015 1002 | | | | |
| BIG2 HUMAN/860-1028 | MEQMAKTAKALMEAVSHVQAP | FTSATHLDHVRPM | FKLVWTPLLAAYSIGLQDC | DDIEVASLCLEGIRCAIRIACI DDTEVASLCLEGIRCAIRIACI |
| Q9VJW1_DROME/801-969 | MEVISLTATNLMQSVSHVKSP | <mark>F</mark> TSAKHLEHV <mark>R</mark> PM | FKMAWTPFLAAFSVGLODC | DDPEIATLCLDGIRCAIRIACI |
| Q9XWG5_CAEEL/767-934 | .FFRTSKKLALMESASDADAY | FTPAQHQHHVKPM | FKICWTPCLAAFSVGVQMS | DDEEEWSLCLRGFRLGVRAACV |
| 1q01960 ARATH/835-999 | .DLIRHMQERFKEKARKSESV | Y YAASDVIILRFM | VEVCWAPMLAAFSVPLDOS | DDAVITTLCLEGFHHAIHVTSV |
| 4g35380_ARATH/793-958 | GRLIRDIQEQFQAKPEKSESV | <mark>Y</mark> HTVTDISIL <mark>R</mark> FI | LEVSWGPMLAAFSVTIDQS | DRLATSL <mark>CL</mark> Q <mark>GF</mark> RYAVH <mark>V</mark> TAV |
| 4g38200_ARATH/776-942 | GLLIKDIQEKFRSKSGKSESA | YHVVTDVAILRFM FHTVFOVDTTPPM | VEVSWGPMLAAFSVTLDQS | DDRLAAVECLRGFRYAVHVTAV |
| SEC7 YEAST/1055-1220 | ISSKTELVFKNLNKNKGGPDV | YYAASHVEHVKSI | FETLWMSFLAALTPPFKDY | DIDTTNKCLEGLKISIKIAST |
| GGG2_PARTE/791-943 | EDSLKKWFKEHP NSDA | F C Y V N S I E H M <mark>K</mark> S <mark>L</mark> | LQQTWSVIFASISVFLEQS | ED QQQILLC F E TI QAFIQ <mark>L</mark> MGR |
| GBF1_HUMAN/893-1066 09V696_DROME/839-1021 | LVRENYVWNVLLHRGATPEGI | FLRVPTASYDLDL FHVVHDASYDVET | FTMTWGPTIAALSYVFDKS | LEETIIQKAISGFRKCAMISAH |
| Q9XTF0_CAEEL/854-1051 | .VKEDYMWKVLLRRGETAEGS | FYHAPTGWNDHDL | FAVCWGPAVAALSYVFDKS | EHEQILQKALTGYRKCAKIAAY |
| GNOM_ARATH/756-930 | PEMTPSRWIDLMHKSKKTAPY | ILADSRAYLDHDM | FAIMSGPTIAAISVVFDHA | EHEDVYQTCIDGFLAIAKISAC |
| GNL1_ARATH//58-930 GNL2_ARATH/688-861 | MTASKWISVIIKSKETSPI .EMNPNRWIELMNRTKTTOPF | SLCOFDRRIGRDM | FATIAGPSIAAVSAFFEHS | DDDEVLHECVDAMISIARV.AO |
| GEA2_YEAST/777-982 | .ISSTTVITEIKKDTQSVMDK | LTPLELLNFDRAI | FKQVGPSIVSTLFNIYVVA | SDDHISTRMIT <mark>SL</mark> DKCSYISAF |
| GEA1_YEAST/773-972 | SVMTEMQRDFTNPISK | <mark>L</mark> AQIDILQYE <mark>K</mark> A <mark>I</mark> | FSNVRDIILKTLFKIFTVA | <mark>SS</mark> DQISLR <mark>IL</mark> D <mark>AI</mark> SKCTF <mark>I</mark> NYY |
| BIG1_HUMAN/915-1083 BIG2_HUMAN/860-1028 Q9YWG5_CAEEL/767-934 3g60860_ARATH/843-1008 1g01960_ARATH/843-1008 1g01960_ARATH/783-999 4g35380_ARATH/793-958 4g38200_ARATH/76-942 3g43300_ARATH/76-942 3g43300_ARATH/76-943 GBF1_HUMAN/893-1066 Q9V696_DROME/839-1021 Q9XTF0_CAEEL/854-1051 GMOM_AFATH/758-930 GNL1_ARATH/758-930 GNL1_ARATH/758-930 GNL2_ARATH/688-861 GEA2_YEAST/777-922 GEA1_YEAST/773-972 | \$0 \$0 \$0 \$0 FS \$0 \$0 FG \$0 \$0 FH \$1 \$1 FG \$0 \$1 FM \$1 \$1 FM \$2 \$1 SUTTRE \$1 \$2 SUTTRE \$2 \$2 SUTTRE \$2 <th>100 VSSGITE ASSSITK ASSSITK ASSSITK SPAD SPAD CAAD CAAD CAAD SPS SSVEQHEPAPANN SMRDGGAGGGADE NPSSV APLSA NPSTT NSSHDDELSTLAF QSSAKAVEL</th> <th>ETQ. DVDLSAAALLSHSS EYGPMPLVQIKFEDTNTEI DVDSIPLVEIFVEDTGSKI</th> <th>LOCOLOGOCOCO 110 </th> | 100 VSSGITE ASSSITK ASSSITK ASSSITK SPAD SPAD CAAD CAAD CAAD SPS SSVEQHEPAPANN SMRDGGAGGGADE NPSSV APLSA NPSTT NSSHDDELSTLAF QSSAKAVEL | ETQ. DVDLSAAALLSHSS EYGPMPLVQIKFEDTNTEI DVDSIPLVEIFVEDTGSKI | LOCOLOGOCOCO 110 |
| | | 000000000000000000000000000000000000000 | ر 150 160 | |
| BIG1_HUMAN/915-1083 | ITVAHTDGNYLGNSMHEI | LK <mark>CI</mark> SQ <mark>L</mark> KLAQ <mark>LI</mark> | GTGVKPRYISGTVRGRE | GSLTGT |
| BIG2_HUMAN/860-1028 | ITVAHTDGNYLGNSWHEI | LKCISQLELAQLI | GTGVKTRYLSGSGRERE | GSLKGH |
| 09XWG5 CAEEL/767-934 | LLIGDEDGE YLEENWVDV | MKCMSSLELVOLI | GTGVRPQFLSGAQTTLK GTGLNSAMSHDTDSSRO | YVMKAT |
| 3g60860_ARATH/843-1008 | LRLADEEGNYLQDAWEHI | LTCVSRFEQLĤ <mark>LL</mark> | GEGAPPDATFFASKQNE | SEKSKQ |
| 1g01960_ARATH/835-999 4g35380_apart#/793_958 | VKLAEEEGNYLQDAWEHI | LTCVSRFEHLHLL | GEGAPPDATFFAFPOTE | SGNSPL |
| 4g38200 ARATH/776-942 | ISIAIEDGNHLQDAWEHI | LTCLSRIEHLQLL | GEGAPSDASYFASTETE | EKKALG |
| 3g43300_ARATH/796-957 | LGLCDSEPDTLODTWNAV | LECVSRLEFII.S | TPGIAATVMHGSNQISR | DGV |
| SEC/_YEAST/1055-1220 GGG2_PARTE/791-943 | LEVALSEGNYLEGSWKDI | LUVVSQMERLQLI | SKGIDR. DTVPDVAQARV | ANPRVS |
| GBF1_HUMAN/893-1066 | FHLAHRHGDILREGWKNI | MEAMLQLFRAQLL | PKAMIEVEDFVDPNGKI | SLQREE |
| Q9V696_DROME/839-1021 | FLLVHDYGDCLRESWKHI | LDLYLQLFRLK <mark>LL</mark> | PKSLIEVEDFCEANGKA | MLILEK |
| Q9XTFU_CAEEL/854-1051 GNOM ARATH/756-930 | FTIANKYGD YTRTGWRNL | LDCILRIHKI.GT | PAELTEVEDYVDEKGWV PARVASDAADESEHSSE | OGOGKP |
| GNL1_ARATH/758-930 | FLIANKYGDYISAGWKNI | LECVLSLNKLHIL | PDHIASDAADDPELSTS | NLEQEK |
| GNL2_ARATH/688-861 | FTLANTFGDSIRGGWRNI | VDCLLKLRKLQLL | POSVIEFEINEENGGSE | SDMNNV |
| GEAZ_YEAST/777-982 GEA1_YEAST/773-972 | F KILKKNKUPKIFSKELWLNI FOTIKEISDPSTVSTRIWNOT | VNIILTLYEDL <mark>IL</mark> VOLILKIFENLIM | SPDIFPDLQKRLKLSNL EPNLPFFTNFHSLLKLPFL | PLPDPD |
| | | | | |

The conserved domains of the BIG/GBF subfamily: HDSI domain. See Figure 3 legend for alignment details.

Evolution of BIGs and GFBs from a common ancestor

Combined, our analysis reveals that the BIG and GBF subfamilies share the same overall domain organization, and are likely to descend from a common ancestor gene that duplicated first to form the BIG and GBF groups, and again within these groups to yield species-specific BIG and GBF members. These two subfamilies can therefore be unified as a higher order ArfGEF subfamily (called below GBG for <u>GBF/BIG GEFs</u>), from which unrooted phylogenetic trees can be built (Figure 8). Unlike previous phylogenetic analysis which compared ArfGEFs based on their Sec7 domains after diverging non-catalytic regions have been trimmed [8], our trees were established from the simultaneous alignment of all 6 conserved domains

| | | | ll | معمعمعم | ee | |
|---|--|---|---|---------------------------------------|--|--|
| PTG1 WWW.W/1107 1200 | | 20 7 7 7 10 7 10 | | | 50 | |
| BIG1_HUMAN/1107-1289 BIG2_HUMAN/1054-1236 | ASTOESIGETSSOSVVV ASFOESVGETSSOSVVV | /A.VDRIF /A.VDRIF | TGSTRLDGNAI | VDFVRWLC VDFVRWLC | AVSMDELLST | |
| Q9VJW1_DROME/968-1149 | PSVKEHIGETSSOSVV | /A.V <mark>D</mark> RIF | TG <mark>S</mark> MR <mark>L</mark> DGD <mark>AI</mark> | VD <mark>F</mark> VKA <mark>LC</mark> | QVSVDELQQ | QQPRM <mark>F</mark> S |
| Q9XWG5_CAEEL/943-1125 3g60860 ARATH/1053-1235 | HSLQDALGETSSQSVVV | A.IDRIF | NGSARLSAEAI | VYFVRALC TDEVKALC | AVSREELSHP | SN PRVES |
| 1g01960_ARATH/1044-1226 | EQMNNLISNLNLLEQVG. | D.MSRIF | TRSQRLNSEAI | IDFVKALC | KVSMDELRSP | SDPRVFS |
| 4g35380_ARATH/1000-1184 | EQIKSFIANLNLLDQIGNE | E.L <mark>NHVY</mark> | AN <mark>SQRL</mark> NSE <mark>AI</mark> | VS <mark>F</mark> VKA <mark>LC</mark> | KVSMSELQSP | TDPRVFS |
| 3q43300 ARATH/982-1166 | OISRDGVVOSLKELAGE | RP.AEOVF | VNSVKLPSESV | VEFFTALC | GVSAEELKO | SPARVES |
| SEC7_YEAST/1253-1438 | TLSPEISKFISSSELVV | /L.M <mark>D</mark> ÑIF | TK <mark>S</mark> SE <mark>L</mark> SGN <mark>AI</mark> | VD <mark>F</mark> IKA <mark>LT</mark> | AVSLEEIESS | ENASTPRM <mark>F</mark> S |
| GGG2_PARTE/945-1125 GBF1_HUMAN/1098-1277 | TENOEAKRVALECTKOC | DQ.I <mark>DKIF</mark> D.PEKMI | N S S I N L D S N S I T E S K F L O L E S L | LEFIRALC OELMKALV | ELSKEEIKY | |
| Q9V696_DROME/1048-1236 | YEEQDFIKLGRKCIKEC | Q.L <mark>DQML</mark> | QE <mark>S</mark> KF <mark>V</mark> QLE <mark>SL</mark> | QE <mark>L</mark> LKC <mark>VL</mark> | ALLKAPQG | H.KSIGLPYAEDQTVFW |
| Q9XTF0_CAEEL/1084-1278 | QEQLSSMKLASQVISEC | CR.P <mark>SQIV</mark> | ADSKYLTSTSL | AELLSSIA | ANSAQIVEQAEPQ | QKTASLSGEDEDALV <mark>F</mark> Y |
| GNL1 ARATH/973-1138 | EEELAAYKHARGIVKDC | CH.IDSIF | SDSKFLOAESL | OOLVNSLI | RASGK | ·····DEASSVFC |
| GNL2_ARATH/894-1084 | ALGMSEFEQNLKVIKQC | CR.I <mark>GQIF</mark> | SK <mark>S</mark> SV <mark>L</mark> PDVAV | ĹŇ <mark>L</mark> GRS <mark>LI</mark> | YAAAGKGQ | KFSTAIEEEETVK <mark>F</mark> C |
| GEA2_YEAST/1011-1194 GEA1_YEAST/1012-1184 | .EEIKSSKKAMECIKSS | SNIAASVF SHPLSSVF | GNESNITADLI ENNOLVSPKMI | KTLLDSAK ETLLSSLV | ТЕ | • KNADNSRYFEAELLFI • KTSENSPYFEOELLFI |
| | | | | | | |
| | | | | | | |
| | | | | 0 | | |
| | 50 70 UUUUU | 80 80 | 90 | 100 | 110 | 120 130 |
| BIG1_HUMAN/1107-1289 | LQKIVEISYYNMGRIRLQ | <mark>V</mark> SR <mark>IW</mark> EV <mark>I</mark> | GDHFNKVGCNP | NE.DVAIF | <mark>A</mark> VDS <mark>L</mark> R <mark>QL</mark> SM <mark>K</mark> FL | EKGELA <mark>N</mark> FR <mark>F</mark> QK <mark>DFL</mark> |
| BIG2_HUMAN/1054-1236 | | VSRIWHVI VSRIWOVI | GDHFNKVGCNP | NE.DVAIF | AVDSLRQLSMKFL | EKGELANFR <mark>F</mark> QKDFL |
| Q9XWG5_CAEEL/943-1125 | LGKVVEVAFYNMNRIRLEW | SRIWNVI | GEHFNAAGCNS | NE.AVAYF | SVDALRQLSIKFL | EKGELP <mark>N</mark> FR <mark>F</mark> QKDFL |
| 3g60860_ARATH/1053-1235 | LTKIVEIAHYNMNRIRLV | SS <mark>IWQVL</mark> | SGFFVTIGCSE | NL.SIAIF | AMDSLRQLSMKFL | EREELA <mark>N</mark> YN <mark>F</mark> QN <mark>EFM</mark> |
| 4q35380 ARATH/1044-1226 | LTKLVETAHYNMNRIRLV | SSIWHVL SRIWNVL | SDFFVSVGLSE | NL.SIAIF | VMDSLROLSMKFL | EREELANYNFUNEFM EREELANYHFOHEFL |
| 4g38200_ARATH/982-1166 | LTKLVEIAHY <mark>N</mark> MNRIRLVW | VSR <mark>IW</mark> SIL | SDFFVSVGLSE | NL.SVAIF | VMDSLRQLSMKFL | EREELA <mark>N</mark> YN <mark>F</mark> QN <mark>EFL</mark> |
| 3g43300_ARATH/951-1132 SEC7_VEAST/1253-1438 | | VARLWSVL VTPLWAVM | AEHFVSAGSHH GKAFNKTATNS | DE.KIAMY NL.AVVFF | AIDSLRQLGMKYL ATDSLROLSMRFL | ERA. ELTNFTFQNDIL DIE. ELSGFEFOHDFL |
| GGG2_PARTE/945-1125 | LSRVIDVAEFNMNRIKII | SR <mark>MW</mark> EIM | REHFLEVGCLK | NV.DVAIY | AIDQLKQLSCKFL | QQPELT <mark>N</mark> YY <mark>F</mark> QK <mark>EFL</mark> |
| GBF1_HUMAN/1098-1277 | LEMLLRIVLENRDRVGCVW | QT <mark>VRDHL</mark> | YHLCVQAQD | FC.FLVER | AVVGLLRLAIRLL | RREEISAQVL |
| Q9XTF0 CAEEL/1048-1238 | LELIVAITLENKDRLPLVV | PHVRRHL | EWLLSPRFG | RCPVLVER | AVVGLLRVANRNL | FRDNTVSDDVL |
| GNOM_ARATH/973-1158 | LELLIAITLNNRDRIVLLW | VQG <mark>VY</mark> EHI | ATIAQSTVM | PC.NLVDK | AIFG <mark>LLRI</mark> CQ <mark>R</mark> LL | PY <mark>K</mark> ES <mark>L</mark> AD <mark>ELL</mark> |
| GNL1_ARATH/973-1149 GNL2_ARATH/894-1084 | UELLIAVTLNNRDRILLIW | VPTVYEHI VPSVHEVI. | LGIVQLTLT LNVANFPL.FS | PC.TLVEK PT.PFVEK | AVFGVLKICQRLL GLPGLFRVCTKTL | PYKENLTDELL ASNTODHLPEELT |
| GEA2_YEAST/1011-1194 | IELTIALFL.FCKEEKELG | KFILQKV | FQLSHTKG | LTKRTV | RRMLTYKILLISL | CADQTEYLSKLINDELL |
| GEA1_YEAST/1012-1184 | LEISIILIS.EASYGQEFC | AL <mark>IA</mark> DH <mark>M</mark> | INISNLDG | LSKEAI | <mark>A</mark> RLA <mark>S</mark> Y <u>KM</u> FL <mark>V</mark> SR | F D N P R D I L <mark>S</mark> D L <mark>I</mark> E H <mark>D F L</mark> |
| | | | | | | |
| | | | | | | |
| | 2222222 | 22222 | 20000000000 | فقفا | 170 | 80 |
| BIG1 HUMAN/1107-1289 | | V <mark>V MC</mark> RTT | RCTAOMVNSOA | ANTRS | COKNTESVEHT. | AASDO |
| BIG2_HUMAN/1054-1236 | .RPFEHIMKKNRSI | TIRDMAI | RCIAQMVNSQA | ANIRS | GWKNIFAVFHQ | AASDU |
| Q9VJW1_DROME/968-1149 | .RPFEHIMKKNASI | PAIR <mark>DMV</mark> V | RCIAQMVNSQA | H <mark>NIRS</mark> | GWKNIFSIFHL | AAGDN |
| 3q60860 ARATH/1053-1235 | .TPFVIVMRRSNDV | ZEIRELII | RCVSOMVLSRV | NNVKS | GWWKSMFMVFTT | AAGDP AAYDD |
| 1g01960_ARATH/1044-1226 | .K <mark>P</mark> FVV <mark>V</mark> MRKSGAV | 7EIR <mark>E</mark> LII | RC <mark>V</mark> SQ <mark>MV</mark> LSRV | d <mark>nv</mark> k <mark>s</mark> | <mark>GW</mark> KSMFM <mark>I</mark> FTT | AAHDA |
| 4g35380_ARATH/1000-1184 4g38200_ARATH/982-1166 | . RPFVVVMQKSSSA | AEIRELIV AETRELIV | RCVSQMVLSRV | SNVKS | GWKNVFTVFTT GWKSVFKVFTT | AALDE AAADE |
| 3g43300_ARATH/951-1132 | .KPFVIIMRNTQSQ | TIR <mark>SLI</mark> V | DCIVQMIKSKV | G <mark>SIK</mark> S | GWRSVFMIFTA | AADDE |
| SEC7_YEAST/1253-1438 | .KPFEYTVQNSGNT | TEVQEMII I OLBEET | ECFRNFILTKS | ESIKS | GWKPILESLQY | TARSS |
| GBF1_HUMAN/1098-1277 | .L <mark>S</mark> LRI <mark>L</mark> LLMKPSVLS | SRVS <mark>HQV</mark> A | YGLHELLKTNA | ANIHS | GD <mark>DWATLFTL</mark> LEC | IGSGV |
| Q9V696_DROME/1048-1236 | .QSLKMLLMLKPALLI | RIS <mark>KQI</mark> S | IGIYELLKTSA | QNIH <mark>S</mark> | EQDWQIIFNLLEC | VGAGA |
| Q9XTFU_CAEEL/1084-1278 GNOM ARATH/973-1158 | - HSLSMLLRLSPKALF | TESROIA | F GLY EL IRANA | ANVHK | KEHWAVLFALLEA OAGWRTTTSLIST | AGAAV TARHP |
| GNL1_ARATH/973-1149 | .K <mark>S</mark> LQL <mark>V</mark> LKLKAKVAL | DAYCERIA | QE <mark>VVRLV</mark> KANA | S <mark>HV</mark> R <mark>S</mark> | RT <mark>GWR</mark> T <mark>IISL</mark> LSI | TARHP |
| GNL2_ARATH/894-1084 | FRSLTIMWKIDKEIIE | ETCYDTIT | EFVSKIIDYS | ANLHT | NIGWKSVLQLLSL | CGRHP |
| GEA1_YEAST/1011-1194 | .VKNEIFNTKYYES | SEWG <mark>KQV</mark> I | NDLFTHLNDVK | YNERALKN | VK <mark>FWN</mark> FLRILISA | KDR |
| — | | | | | | |

Figure 6 The conserved domains of the BIG/GBF subfamily: HDS2 domain. See Figure 3 legend for alignment details.

(DCB, HUS, Sec7, HDS1, HDS2 and HDS3), excluding variable linkers. The same tree topology was obtained with both neighbor-joining and maximum likelihood methods, and was retained using any one of the new conserved domains alone (data not shown). Bootstrap analy-

sis strongly supported this topology for most branches. Only a few small branches located at the base of the groups were found in less than 60% of the trials in one of the two methods, but this never occurred with both methods simultaneously.

| 1 10 20 30 40 50 60 BIG1_HUMAN/1372-1489 .APEDRVWNRCH FPTLFELSCIINCK.LDWTRCHTWFEINKTYGHTYENWKODLFRIWPETENN. BJG2_HUMAN/1372-1489 .APEDRVWNRCH FPTLFELSCIINCK.LDWTRCHTWFEINKSYGHTYENWKODLFRIWPETENN. BJG0450_ARATH/1371-1472 .AEDRVWNRCH FPTLFELSCIINCK.LDWTRSLTVWFEINKSYGHTYENWKODLFNUTPETENN. Jg06060_ARATH/1321-1452 .FLESDEHLYSN FPTLACLSCISISTEPP, PETRKSLTVWFEINKHNGHDFSLALWERVESVLFRIFUY Jg33300_ARATH/122-131 .FTDLEGOVSVTPLLCCLSCVSDP, PARKSGLEVLFNIKHNGHDFSLALWERVESVLFRIFUY Jg33300_ARATH/122-131 .FUEDRVKVKURUPPLYCCLSCVSDP, PARKSGLEVLFNIKBRGNKFSTFWIGVRSSTUPPTNY Jg33300_ARATH/122-131 .FUEDRVKVKURUPPLYCCLSCVSDP, PARKSGLEVLFNIKBRGNKFSTFWIGVRSSTUPPTNY SGC2 YEAST/155-1649 .LRCKDIFDUVWFNHECCNDITHAD DLEGKSALMVMPARVHEGOKFFDFWESIGKKUPPLEDE GB760_BORMS/1475-1522 .PGSALWSCCCLDCGLACLCCDA.RCVMC40ATYLCRALL.VHBLQCKALL.EWESCNKVFPLLNE G9760_GORMS/1475-1522 .PGSALWSCCCLDCGLACLCCDA.RCVMC40ATYLCRALL.VHBLQKLAL.EWESCNKVFPLLNE G970_GES_TATAT/1235-1349 .CKKSDFPHELGCINCCCLCCCKSCGCN, EDVTRALL, NKOSCCCCLDCOK, EDVTPLLNE G970_GES_TATAT/1235-1349 .KKSDFPHELGCINCCCCCCCCK, EDVTRALL GMMARAT/1312-1439 .KKSDFPHELGCINCCCCCCCCKCCCCCCCCCCCCCCCCCCCCCCCCCCC | | | ll | 222222222222 | | | 22222222222 | | ll | l | | 2222222222222 | | | |
|---|---|---|---|--------------------------------------|--|---|--|--|---|--|---|---|--|--|--------------------|
| BIG1_HUMAN/1372-1489 .APEDRVWRGGPTLTETESCIINCK.LDURTKCLTVWFEIKKTYGHTYEHWGDLFRIVERTENN BIG2_HUMAN/139-1436 .APEDRVWRGGPTLTETESCIINCK.LDURTKCLTVWFEIKKSYGHTSKHWGDLFRIVERTENN 09XWG5_CAERL/121-13139 .APEDRVWRGGPTLTESCIINCK.LDURTKCLTVWFEIKKSYGHTSKHWGDLFRIVERTENN 1930060_ARATH/1337-1472 EIV.NNNLISWFPTLSCISKSCUVNRCK.LDURTKSLTVWFEIKKHGSDFFFEWKKDLFNVFERTENN 1930560_ARATH/1237-1472 EIV.NNNLISWFPTLGCLSELSPDP.APE.KKSLOIMFDTLRNHGDHFSLA.LNEKVFESVLFPIFDY 193050_ARATH/1278-1410 .FUEDDEVSWIPLLGCLSELSPDP.APE.KKSLOIMFDTLRNHGDHFSLA.LNEKVFESVLFPIFUY 193300_ARATH/122-1394 .FUEDDEVSWIPLICCLSKISDS, SARKKSIEVLFNILKHGBGHFST.FFNIGVSSVTPIFTS' 193300_ARATH/122-131 DET.FDVTEHYMFPMLGCLSLISDTSN, SARKSELEVLFNILKHGBGHFST.FFNIGVSSVTPIFTS' 193400_ARATH/122-1331 DET.FDVTEHYMFPMLGCLSLISDTY, PEWRCALEVLFNLKHGBGVSKF.FFNKIIKKLFPFTFU' 19320_BRATH/122-1331 DET.FDVTEHYMFPMLGCLSLISDTY, PEWRCALEVLFNLKHGBCVKESVLFPTF.F. 1990160_BRAMM/1532-1644 .ADSRTUARCOCPLOCIACCALARDR, RCW ROALEVLFNLKHGBCVKEPTFLKK. 1990170_BRALLSSCHULLSSCHULLAUGCLARKVCLDG, RCW RAALTHEDRAK, PENNANGSCHUKVEFPLLK, SCHUPFLKK .HGGCHONGOLOGOLOGOLOGOLOGOLOGOLOGOLOGOLOGOLOGO | | 1 | 10 | | 20 | 3 | o, | 4 0 | | | 50 | | 6 | 0 | |
| BIG2_EUGAN/1315-1436 APEORVWURGNEPTLPEIGCIINNCK.LDWTRCLTVWFELTKSYGHTFERHWODDFRIVFRTPNTP 09VWIL DOWE/1233-1350 APEORVWURGNEPTLPEIGCIINNCK.LDWTRCLTVVFELTKSYGHTFERHWODDFRIVFRTPNTP 09VWIL AGMATH/1321-1422 TADQHVWLRGNEPTLPEISCIINNCK.LDWTRCLTVVFELTKSYGHTFERHWODDFRIVFRTPNTP 1g01860_ARATH/1321-1452 FLESDEHLSWIPLLAGLSELSTPPR.PERKK.LTVVFELTKHHGSDFRPEWKND_LETVFFTPTG 1g01860_ARATH/1231-1452 FLESDEHLSWIPLLAGLSELSTPPR.PERKK.LEVTFDTL.N.HHGHFSLP.L.LKEVV ESVLFTPTG 1g33300_ARATH/1262-1394 FNDADENISYWVLITGLKENSDFR.SARKSELEVLFTILMDHGHFFSTP.F.FRSTEST EST HENDENISYWVLITGLKENSDFR.SARKSELEVLFTILMDHGHFFSTP.F.FRSTEST EST HENDENISYWVLITGLKENSDFR.SARKSELEVLFTILKDHGHFFSTP.F.FRSTEST EST HENDENISYWVLLTGLKENSDFR.SARKSELEVLFNILKVHGEQGENE.FKXI.KKLPPIFDU 1g020_ARATH/1232-1331 DET.FVVENVHFMLCGINDINALGGLSLINDER.GVVMCHOLTSC.GRALDV.HTFDLLK.U.TKGGNE.FKXI.PVPIFDE 1g0262_PART/1195-1131 NKUDQLTITVCGLSLUSDER.GVVMCHCLTVLQRALLVHDLGKLDGT.EEVCSC.HVVPPLFDE 1g07636_DRMK/1475-1592 YQSALMSFGKCLU GGTARLANDR.REWTHLSCG.GVANLTSC.GVDGNE.FKXKLPPLFDE 09X65_CAEE//1355-1667 ATVAFIWNDINNLLG GTARLANDR.REWTHAGEN ALLVHLGCRGT.E.CCC.HVVPPLFDE 09X66_DRMK/1475-1592 YQSALMSFGKCLU GGTARLANDR.REWTHAGEN ALLYHDGRALLVHDLGKSGT.E.CCC HVVPPLFDE 09X66_DRMK/1475-1592 YQSALMSFGKCLU GGTARLANDR.REWTHAGEN ALLYHDGRALLVHDLGKSGT.E.CCC HVVPPLFDE 09X66_DRMK/1475-1592 YQSALMSFGKCLU GGTARLANDR.REWTHAGEN ALLYHDGRALLVHUGGTAGECC HVVPPLLNE | BIG1 HUMAN/1372-1489 | APEDRV | WVRGWF | PILFEI | SCIINRC | K.LDV | RTRGI | TVMFEI | мкт | YGHTY | ЕКН | .WWOD | FRI | VFRIFI | <mark>р</mark> имк |
| 0900011 DROME/1233-1350 AREDRWUWRGWFEMLFSIGCVUNNCK.LDWTRAITUFETVKTYGESFKPHWKDLFINVEFRIPDING 090005 TADGHVURGWFEMLFSIGCVUNNCK.LDWTRAITUFETVKTYGESFKPHWKDLFINVEFRIPDING 090005 ARATH/1231-1452 1001060 FTADGHVURGWFEMLFSIGCVUNNCK.LDWTRKCK.LDWTRK.N.HHGEPFELS.LWEKVESVLPPTFDY 4935300 ARATH/1278-1410 FTEDDEDGVSYNTFLIGGLSELSPPR.AERKTEVFFTLMHHGHEFST.FNFTGVSSTUPTFSY 4933300 ARATH/1221-1331 DET.FDVTENYWFEMLGSLSDSN.SALKKVEYFTLNHHGHEFST.FNFTGVSSTUPTSY 3943300 ARATH/1221-1331 DET.FDVTENYWFEMLGSLSDSN.SALKKVEYFLNHVEFDLNERGNKFSTP.FNSTHRIFFTFY 39630 ARATH/1212-1331 DET.FDVTENYWFEMLGSLSDLTSDY.PVENCLEVFDLNERGNKFSTP.FNFSTHRIFFTFY 39630 ARATH/1221-1331 DET.FDVTENYWFEMLGCLDOLAGENUTYCLOGIACLCOLAR.NYFDALUTY.AUXOLALL.WESCNNDD.FNFSTN.FNESTHRIFFFFGU 39630 ARATH/1237-1449 OSKALWARCKYCLUGGIALCOCAA.ROWGOCUNTTFELLX.UNGCNNDL.FNFSTPHTFYFFU 39630 ARATH/1237-1431 CMUMARCKYCLUGGIALCOCAA.ROWGOCUNTTFELLX.WHGRMALGAA.ENCYCLUGUNTTFELSK 90700 110 90700 110 908100000000000000000000000000000000000 | BIG2 HUMAN/1319-1436 | APGDRV | WVRGWF | PILFEI | SCIINRC | K.LDV | RTRGI | TVMFEI | мк | YGHTF | EKH | . WWOD | LFRI | VFRIFI | римк |
| 595W05CCAEEL/1212-1329 TADORYWLEGY FITPELSCIINNCK.LUVETREITVEFETKHBGSDFEPE WKDDLEITPETPE 1360860_ARATM/1321-1452 FIV.NNHLYFFFILSGESELSPPH.PEIKKSLOVETREITVEFETKHBGSDFEPE WKDDLEITPETPE 4935300_ARATM/1321-1452 FIV.NNHLYFFFILGGESELSPPH.PEIKKSLOVETREITVEFETKHBGSDFEPE WKDDLEITPETPE 4935300_ARATM/1262-1394 FMDADENISYVPLITGESKLESPPH.PEIKKSIEVLPTIKBEHLFRFFIGISSILDPYN. 4935300_ARATM/1262-1394 FMDADENISYVPLITGESKLESPPH.PAIKKSIEVLPTIKBEHLFRFFIGISSILDPYN. 4935300_ARATM/1262-1394 FMDADENISYVPLITGESKLESPPH.PAIKKSIEVLPTIKBEHLFRFFIGISSILPPYN. 4935300_ARATM/1262-1394 FMDADENISYVPLITGESKLESPPH.PAIKKSIEVLPTIKBEHLFRFFIGISSILPPYN. 6622_PARTM/1262-1394 FMDADENISYVPLITGESKLESPPH.PAIKKSIEVLPTIKBEHSTEFIGISSILPPYN. 6622_PARTM/1235-1643 JDSRTWAHCKCPLIOGIARLAMDRR.R.VISPENVIEPGVI 909V656 DROME/1475-1592 POSAALWSPGCCPLIOGIARLAMDRR.R.EVFHAISSCOORSNILTELKVHBOLOKLDALFESCNVPPTULDI 909V650 DROME/1475-1592 POSAALWSPGCCPLIOGIARLAMDRR.R.EVFHAISSCORCH.LVBDLOKLDALFESCNVPPTULDI 909V650 DROME/1475-1592 POSADIGEMMUKLVKVCLDOL.EVENNHAISW.CR.G.G.GVDGINLABSWSCCGVVPPTULDI 909V650 DROME/1475-1592 POSADIGEMMUKLVKVCLDOL.EVENNHAISW.CR.G.G.GVDGINLABSWSCCGVVPTVLDDI 909V650 DROME/1475-1592 POSADIGEMMUKLVKVCLDOL.EVENNHAISW.CR.G.G.GVDGINLDEFSFOSCHOVVPTVLDI 909V510 DROME/1475-1592 POSADI | 09VJW1 DROME/1233-1350 | AEEDRV | WVRGWF | PMLFST | SCVVNRC | K. T. DV | RTRAT | TVLFET | VK T | YGESE | KPH | . ww k D | I. FNV | TFRTFI | NMK |
| SigGo&GO_RATH/1317-1472EIV.NNNHLYPEPLLSGISELSTDPR.NETKSALOIKEDTRNEGHEFSLPLEKVESVDPTPTY1g01960_RATH/1217-1452FJESDEHLYSKPPLLGGISELSTDPR.AEIKKSLOVERDTYRNEGHEFSLPLEKVESVDPPTNY4g3300_RATH/1217-1452FJEDDEUXSVDPLTGGISELSTDPR.AEIKKSLOVERTHYRDEGHEFSRTFIGVISSVDPPTNY4g3300_RATH/1217-131EIV.NNNHLYPEPLLGGISELSTDPR.AEIKKSLOVERTHIKDEGHEFSRTFIGVISSVDPPTNY4g3200_RATH/1217-131EIV.PDTFHUNGGISELSTDPR.AEIKKSLOVERTHIKDEGHEFSRTFIGVISSVDPPTNY5GC1YEBST/1525-1649RKDIPDVFHYPLGCNDITMADDLEVSKSLOVERDUV.A.ERGNEFSTPFESIHRILPPTPH5GC2PARTE/195-1313KYLDGLTPVCSAUSULSDVCALCODAR.EVVCGOSNTLPELXVHGEQOSNEFKILRPPTPHDH5GC1PATE/195-1313KYLDGLTPVCSAUSULSDVCALCODAR.EVVCGOSNTLPELXVHGEQOSNEFKILRPPTPHDH5GC2PARTE/195-1313KYLDGLTPVCSAUSULSDVCALS | 09XWG5 CAEEL/1212-1329 | . TADOHV | WLRGWF | PIFFEI | SCIINRC | K.LDV | RTRSI | TVMFEI | мк.н | HGSDF | RPE | WWKD | LLEI | VFRIFI | PSK |
| IndianIndia | $3\sigma 60860$ ARATH/1337-1472 | ETV.NNN | HLYFWF | PLUSGI | SELSEDE | R. PET | RKSAT | OTMEDT | T.R N | HGHLF | SLP. | LWEKV | FESV | FPTF | VVR |
| ig3330ig330ig15300ig15300ig16300ig16300ig163000000000000000000000000000000000000 | 1001960 ARATH/1321-1452 | FLESDE | HLYSWF | PLLAGI | SELSEDE | R.AET | RKVAT | KVLFDT | LR.N | HGDHF | SLA. | LWERV | FESV | | VVR |
| ig38200_RARH/1262-1394 :PMDADENISY VPTTGCKKLTSDS:SALTRKSELEVIENTKBEGHETSRT.:PKICVESUTVPTPS; ig43200_RARH/1212-1331 DDT.FDVTEHYK FPNIAG(SUITSDYR.DEVRICALEVIENDL, L.RGNKETSDF.F.FKEIFHILPTPDH ig43200_RARH/1212-1331 DDT.FDVTEHYK FPNIAG(SUITSDYR.DEVRICALEVIENDL, L.RGNKETSDF.F.FKEIFHILPTPDH ig43200_RARH/12152-1649 LRGKDIFQDVWFPMIFCCNDTIMTAEDLEVRSRAINNFDALVAYGGKPNDDFWEKICKKLIFPIPDH ig9056_DROME/1475-1592 .PQSAALWSFGWCDPL0G(GARLANDR.REWYTHAISCLQQNALLVHDLQRAL.PANNATLGAL.EWGSCECKVLUPPLLTKI ig90760_CAEL/1555-1667 .ATVAFIWTDIWFPLQALCCLDQG.CCCCCG.GVDGINLAHSHWSGCFCFVLPPLLTKI GNOM ARATH/1225-1349 .GKMSQDIGENWLRLVQGLRKVCLDQG.EDVNHALOSLQCKCG.G.GVDGINLAHSHWSGCFCFVLPPLLTKI GNLI_RARH/1127-1341 .KKIVKNLKVVCLDQR.DEVNHALOSLQCKCG.G.GVDGINLAHSHWSGCFCFVLPPLLTKI GEA2_TASAT/1271-1375 .RGVNFVHLELKLVKNLKVVCLDQR.DEVNHALOSLQCKCG.G.GVDGINLAHSHWSGCFCFVLPPLLTKI GEA1_TEAST/1271-1375 .RSNSALNKETIAIQGLAHQCLNPC.DELGMOALQALENILC.SKTUTSDYNLEUGUL SGEA1_TEAST/1271-1375 .RSNSALNKETIAIQGLAHQCLNPC.DELGMOALQALENILCOVFTOVLEVLSDVLLDDIFAQLWCV SGEA1_TEAST/1271-1375 .RSNSALNKETIAIQGLAHQCLNPC.DELGMOALQALENILCOVFTOVLEVLSDVLLDUFAQLWCV SGEA1_TEAST/1271-1375 .RSNSALNKETIAA1QGLAHQCLNPC.DELGMOALQALENILCOVFTOVLEVLSDVLLDUFAQLWCV SGEA1_TEAST/1271-1375 .RSNSALNKETIAA1QGLAHQCLNPC.DELGMOALQALENILCOVFTOVLEVLSDVLLDUFAQLWCV SGEA1_TEAST/1271-1375 <td< th=""><th>$4\alpha 35380$ ARATH/1278-1410</th><th>FTDLDE</th><th>OVSYWT</th><th>PLUTGI</th><th>CKOVSDE</th><th>R. PAT</th><th>RKRST</th><th>EVLEHT</th><th>Т.М П</th><th>HGHLF</th><th>TRP.</th><th>FWTGT</th><th>FSST</th><th>TLPVFI</th><th>NNTR</th></td<> | $4\alpha 35380$ ARATH/1278-1410 | FTDLDE | OVSYWT | PLUTGI | CKOVSDE | R. PAT | RKRST | EVLEHT | Т.М П | HGHLF | TRP. | FWTGT | FSST | TLPVFI | NNTR |
| 3q4300_RATH/1212-1331 DET_FDUTEHY FPMIAGLSDITSPR.FEVR.CALEVIFDLDERGNKFSTPFWESIRHRILFPTPGU SEC7 YEAST/1525-1649 | $4\sigma 38200$ ARATH/1262-1394 | FMDADE | NISYWV | PLLTGI | SKLTSDS | R.SAT | RKSSI | EVLENI | L К.D | HGHIF | SRT. | FWIGV | FSSV | TYPIFI | SVW |
| SECT YEAST/1525-1649 I.RGKDIPODVWFPMLFC NVDIMAEDLEVBSRAINY MEDALV. AYGCKPNDD. FERICKKLLEPTPGVJ GGG2 PARTF/1195-1313 GGG2 PARTF/1195-1313 .NKYLDOLWIPVLSA SVLYSDEN. GVVGOGVTT FELLV. VHGDOSSE. FYRIIERGVI GBYTPG CAEEL/1555-1667 GNME/1475-1592 .PQSAALWSPGMCCPLLOG ACLCOAN. ROVRMOALTY DORALL.VHDLOKLDAT. EN SSCHWVLFPLLTKI GRUNARATH/1225-1349 GNMS/1475-1592 .PQSAALWSPGMCCPLLOG ACLCOAN. ROVRMOALTY DORALL.VHDLOKLDAT. EN SSCHWVLFPLLTKI GNMARATH/1225-1349 GNMSODIGEMWLLWVGK KVCLDOR. BVVNNALGS OKCLC.GVDGILAARS. ME SSCHWVLFPLLTKI GNMARATH/1227-1343 MKSEDIGEMWLLWVGK RKVCLDOR. BVVNNALGS OKCLC.GVDGILAARS. ME SSCHWVLFPLUDDI GNLZARATH/1177-128 GEA2_YEAST/1277-131 EDNDLRNNEIFAI QG AHQCLNPC.DEVRNALSMORAL.GADGINLAGS. ME SQCHKVVLFPLUDDI GNLZARATH/1271-1375 RSNSALNKNEIFAI QG AHQCLNPC.DELGMOALQALENIL.SRSSCHWAM DNL ETGLPIPELI GEA2_YEAST/1277-1381 EDNDLRNNEIFAI QA AHQCLNPC.DELGMOALQALENIL.SRSSCHWAM DNL ETGLPIPELI SGCA1_YEAST/1271-1375 BIG1_HUMAN/1372-1489 LPEQQTEKAEWMTTCCHALVAICDE TOVELEVLSDVLLDDIFAQUYWCV 905WG5 CAEEL/1212-1329 DO0 100 SG60G ARATH/131-1472 HSIDSGEDESADGOSGEVDELDHDAWLYENCTLAHAUSY VE TO YTOLSYALPMIYRCFGIFI 3G60SG ARATH/1321-1452 OVDPSE.DSDTDORGYNGEVDQESKLYUT VE TO YTOLSYALPMIYRCFGIFI 3G60SG ARATH/1321-1452 GUDYPSE.DSTDORGYNGEVDQESWLYENCTLAHAMARYENCH HAMLSYVEVFTOV VISUSVLLAUKVEVFI 3G40300_ARATH/1221-1451 SKTDMLFESVDS.SSSE.SSREWMTTCCHAMARYTE SLALOUVDE VN YVTNP. LLKKVLMLYVSLI 4G33300_ARATH/1278-1410 SKTDMLFESVDS.SSSESSEVENSTICHAMALVENERSVENTSULGUVNELLKVLMLYVSLI 4G33300_ARATH/1278-1 | 3q43300 ARATH/1212-1331 | DET.FDV | TEHYWF | PMLAGI | SDLTSDY | R. PEV | RNCAL | EVLFDI | | RGNKF | STP. | FWEST | FHRI | LFPIFI | DHVS |
| GGG2_PARTF/1195-1313 | SEC7 YEAST/1525-1649 | LRGKDT | FODVWF | PMLFCF | NDTTMTA | EDLEV | RSRAT | NYMEDA | T.V.A | YGGKF | NDD | FWEKT | CKKT. | FPTF | VLS |
| GBF7_HUMAR/1532-1644 ADSRTLWARCWCPLIQG ACLCCAR, ROVENO, TYLORALL.VHDLQKLDAL.EWSCCHVVLPPLITKI Q9V696_DROME/1475-1592 .PQSAALWSPGWCPLLQG ACLACHARR, REVENTHALSCLQQRALLVHDLQKLDAL.EWSCCHVVLPPLLTKI PQSAFDC CAEEL/155-1667 GNUA_RATH/1235-1349 .GKMSQDIGEMCGLUQGG ACLACHQR, REVENTHALQSLQQRALLVHDLQKLSGT.EWSCCHVVLPPLLTKI GNUA_RATH/127-1343 .KKSEDIGKMCKLVQKUKKVCLDQR.EDVRNHALQSLQCCLG.GVDGINLAHSMWSQC DKVLFVVLDDI GNL2_ARATH/127-1343 GNL2_ARATH/177-1288 .LRGVNFVHLLCLSEAL RKTTARR.EEIRNRAVTSLEKSFTMGHEDLGFTPSGCIVCCDHVLPPTIDDI GNL2_ARATH/127-1381 .BDNDLKKNELFAIQALAHQCINPC.KOISEFNVTLEQTLINKIETPTNENESVEEL EGGLUPLLNS .RSNSALNKNEITATQGLAHQCINPC.DELGMQAIQALENILL.SRASQLRTEKVAVDNLFETGLLPTFLI BIG1_HUMAN/1372-1489 LPEQQ | GGG2 PARTE/1195-1313 | NKY | T. DOT. WT | PVLSAT | SVLYSDE | R. GVV | ooosv | NTLFET | T.K.V | HGEOO | SNE. | FWKTT | LRGV | TRPLFI | DETO |
| Q9VG96_DROME/1475-1592 .PQSAALWSPGMCPLLQGLARLAMDRR.REVETHAISCLOORALLVHDLQTLSGT.EWCSCHOVLPPLINE) Q9XTPO_CAEEL/1555-1667 .ATVAFIWTDIRPLLQAIGRSCDCR.RGVRAAALTHLQRAFL.PANMATLGAA.EWQSCHGEVLPPLLK) GNOM_ARATT/1235-1349 .GKMSQDIGEM/LRLVQGLRKVCLDQR.EDVRNHALQSLQCLG.G.GVDGINLAHS.MSQCCGEVVPLLLK) GNL2_ARATT/127-1343 .MKLSEDIGKM_LKLVKNLKKVCLDQR.EDVRNHALQSLQCLG.G.GVDGINLAHS.MSQCEDSVPILLDD GNL2_ARATT/127-1343 .MKLSEDIGKM_LKLVKNLKKVCLDQR.EDVRNHALQSLQCLG.G.GVDGINLAHS.MSQCEDSVPILLDD GeL2_YEAST/1277-1381 .EDNNDLRKNE FAIIQALAHQCINPC.KQ.SEFAVVTLEQTLI.NKTEIPTNEMESVEELIEGGLIPLINS: GEL1_YEAST/1271-1375 .RSNSALNKNE FAIQGLAHQCLNPC.DELGMQALQALENILL.SRASQLRTEKVAVDNLETGLIPTPLID BIG1_HUMAN/1372-1489 LPEQQ | GBF1_HUMAN/1532-1644 | ADSRTL | WAHCWC | PLLOGI | ACLCCDA | R.ROV | RMOAL | TYLORA | LL.VH | DLOKL | DAL | EWESC | FNKV | LFPLL | TKLL |
| | 09V696 DROME/1475-1592 | POSAAL | WSPGWC | PLLOGI | ARLAMDR | R.REV | RTHAT | SCLOOR | ALLVH | DLOTL | SGT. | EWCSC | FHOV | LFPLLI | NELL |
| GRMM ARATH/1235-1349 | 09XTF0 CAFEL/1555-1667 | ATVAFT | WTDTWR | PLLOAT | GRISCOC | R. RGV | RAAAT | THLORA | FL.PA | NMATT. | GAA | EWOSC | FGEV | FPLL | TKT.T. |
| GNL1_ARATH/1227-1343 .MKLSEDIGKMMLKLVKNLKKVCLDOR.DEVENHAISMLORATA.GADGIMLPOP.LIMFOCEDSAVFILLDOV GNL2_ARATH/177-1298 .LRGVNFVHLELKLSEAFRKTTLARR.EELANRAVTSLEKSTTMGHEDLGFTPSGCIYCLDHVTFPTLDDI GEA2_YEAST/1277-1381 .EGNNDKKNEIFAIIQALAHQCLNPC.KUSEAFKTTLARR.EELANRAVTSLEKSTTMGHEDLGFTPSGCIYCLDHVTFPTLDDI GEA1_YEAST/1277-1375 .RSNSALNKNEITAQALAHQCLNPC.LDELGMQALQALENILL.SRASQLRTEKVANDNLLETGLLPIFELI BIG1_HUMAN/1372-1489 LPEQQ | GNOM ARATH/1235-1349 | GKMSOD | TGEMWI | RLVOGI | RKVCLDC | R.EDV | RNHAT | OSLOKO | LG.GV | DGINL | AHS. | MWSOC | FDKV | TFTVL | |
| GNL2_ARATH/1177-1298 .LRGVNFVHHLPLKLSPAFRKTTLARR.EEINNRAVTSLEKSTMGHEDLGFTPSGCIYCTHVIFPTIDD GEA2_YEAST/1277-1381 .EDNNDLRKNEIFAIIQALAHQCINPC.KQISFRAVVTLEQTLINKIEIPTNEMESVEELEGGLLPLINS: .RSNSALNKNEITAAIQGLAHQCINPC.DELGMQAIQALENIL.SRASQLRTEKVAVDNLETGLLPIFELI BIG1_HUMAN/1372-1489 LPEQQ | GNL1 ARATH/1227-1343 | MKLSED | T G K M W L | KT.VKNT | KKVCLDO | B.DEV | RNHAT | SMLORA | TA.GA | DGTML | POP | LWFOC | FDSA | VFTLL | |
| GEA2_YEAST/1277-1381 EDNNDLRKNEIFAIIQALAHQCINPC.KQISEFAVVTLEQTIINKIEIPTNEMESWEELTEGGLLPLINS: .RSNSALNKNEITAAIQGLAHQCLNPC.DELGMQAIQALENTLL.SRASQLRTEKVAVDNLETGLLPIFELI GEA1_YEAST/1271-1375 EDNNDLRKNEIFAIIQALAHQCINPC.KQISEFAVVTLEQTIINKIEIPTNEMESWEELTEGGLLPLINS: .RSNSALNKNEITAAIQGLAHQCLNPC.DELGMQAIQALENTLL.SRASQLRTEKVAVDNLETGLLPIFELI BIG1_HUMAN/1372-1489 LPEQQ | GNL2 ARATH/1177-1298 | LEGVNE | VHHT.FT. | KLSEAF | RKTTLAR | R.EET | RNRAV | TSLEKS | FTMGH | EDLGE | TPS. | GCTYC | TDHV | TFPTTI | DT.H |
| GEAL_VEAST/1271-1375 .RSNSALNKNETIAATOGLAHOCLNPC.DELGMQALQALENILL.SRASQLRTEKVAMDNLETGLLPIPELI GEAL_VEAST/1271-1375 .RSNSALNKNETIAATOGLAHOCLNPC.DELGMQALQALENILL.SRASQLRTEKVAMDNLETGLLPIPELI BIG1_HUMAN/1372-1489 LPEQQTEKAEWMTTICNHALYAICDVFTQYLEVLSDVLLDDIFAQLYWCV BIG1_HUMAN/1319-1436 LPEQQTEKAEWMTTICNHALYAICDVFTQYLEVLSDVLLDDIFAQLYWCV Q9VWG5_CAEEL/1212-1329 MDDHRSKRSWMTTICNHALYAICDVFTQYFDVLGHLLEELFAQLHWCV Q9XWG5_CAEEL/1212-1329 MDDHRSDKREWMSTICNHAMLSVVEVFTQFYTUNP.LLEKVLMLLVSFI 1g01960_ARATH/1321-1452 QUVDPSELDDSTDQGSGGEVDELDHDAWLYEICTLALQLVVDLFVKYTTVNP.LLKKVLMLVSFI 1g01960_ARATH/1278-1410 SKTDMLFEESVDS.PSSAS.LDTEETTWDVETSTLALQLVDLFVKFYTTVNP.LLKKVLMLVSFI 4g38200_ARATH/1278-1410 SKTDMLFEESVDS.PSSAS.LDTEETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVVSILAGLI 4g438200_ARATH/1278-1410 SKTDMLFESVDS.PSSAS.LDTEETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVVSILAGLI 4g438200_ARATH/1278-1410 SKTDMLFESVDS.PSSAS.LDTETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVSVILAGLI 4g438200_ARATH/1278-1410 SKTDMLFESVDS.PSSAS.LDTETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVSVSILAGLI 4g45200_ARATH/1278-1410 SKTDMLFESVDS.PSSAS.LDTETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVSVSILAGLI 4g45200_ARATH/1278-1444 SKTDMLFESVDS.PSAS.S.LDTETTWDVETSTLALQLLVDLFVKFFSVRS.QLPSVSILAGLI 4g45200_ARATH/1278-1444 SKTDMLSKOVT 99096_DROME/14751525-16 | GEA2_VEAST/1277-1381 | . EDNNDL | RKNETF | ATTOAT | AHOCINE | C. KOT | SEFAV | VTLEOT | T. T N | KTETP | TNEME | SVEEL | TEGG | T. T. P T. T. I | NSSE |
| Indext in the second | GEA1_YEAST/1271-1375 | RSNSAL | NKNETT | AATOGI | AHOCINE | C. DEL | GMOAT | OALENT | LT S | RASOL | RTEKV | AVDNT. | LETG | I. T. P T F I | ELDE |
| GRL2_ARATH/1177-1298 EKLLDYSRRENAEREMRSMEGTLKIAMKVLMNVFLVYLEQIVESAEFRTFW GEAZ_VEAST/1277-1381 TQEQKLISSILTIISNVYLHYLKLGKTSNETF GEAI_VEAST/1271-1375 IO | BIG1_HUMAN/1372-1489 BIG2_HUMAN/1319-1436 Q9VJW1_DROME/1233-1350 Q9XWG5_CAEEL/1212-1329 3g60860_ARATH/1321-1472 4g35380_ARATH/1278-1410 4g38200_ARATH/1278-1410 4g38200_ARATH/1278-1410 4g38200_ARATH/1212-1331 SEC7_YEAST/1525-1649 GGG2_PARTE/1195-1313 GBF1_HUMAN/1532-1644 Q9XTF0_CAEEL/1555-1667 GNOM_ARATH/1235-1349 GNL1_ARATH/127-1343 GNL2_ARATH/117-1298 | 70 LPEQQ LPEQL MDDHR HSIDPSG QDVDPSE SKTDMLF GENDLLS HAGKESL KHWEVNQ ISKLQ ENI PES EPF EFF EFS | EDESAD .DDSTD EESVDS KDEHSS. ISS. FA FA NAA A G IENS RREN. | QGSSGC QRGYNC .PSSAS FPSTFS | TEK SEK SEK SDK SDK SDK SPAE SPAE SPAE SQME SQME SQME SQME SQME SQME SQME SQME | 2000 80 AEWMT SEWMT REWMS ESWLY TTWDV VSWDA DVKFLS SVWLS SV SVWLS SVWLSV | TTCNH TTCNH TTCNH ETCSI ETSSI ETSSI ETSSI ETSI CNCKMR ETSI CNCKMR ESRIRVE DTLLI GTLLI | 9000000 90 IALYATC IALYATC IALYATC IALYATC IALYATC IALQUV ALQUV I | DVFTQ DVFTQ DVFTQ DVFTQ DLFVK DLFVK NLFVT ALFVT ALFVT ALFTT DLVVLQ KVFLQ KVFLQ NVFLV | Iooo YLEVL YFDVL YFDVL YFTTV FYTTV FFTKEV YFESL YIQQPL HLSAL QLQEL SLQDI YLEQI | SDVLL NEVLL GHLLL SVYAL NP.LL NS.QL RS.QL CF.ML NR.ML QP.CL LSL SQL SQL VES. | 0000 110 DDIFA SDVFA EELFA PMIYE KVUM KVVS SSVVS DGFLG NDLID STFA OSFFC . SEFR | QLYW QLQW QLHW QFGI LLVS LLVS LLVS LLLD LLVS LLLD LLVS LLLD LLVS LLLD LLVS TYW. | CV CV FI LI FI LA CI LV | |

Figure 7 The conserved domains of the BIG/GBF subfamily: HDS3 domain. See Figure 3 legend for alignment details.

The tree topology strongly suggests that in most organisms, GBG members sort in separate branches, corresponding to their classification in BIGs and GBFs. Remarkably, our annotation of Sec7-containing proteins in the genome of Paramecium reveals the first departure from this distribution, as all GBGs in this species are located in a single branch, which is closer to the BIGs. This unexpected tree topology may indicate that alveolates diverged from animals/fungi and plants before the duplication of an ancestral GBG into the BIG and GBF, and that GBGs in that organism are representative of this ancestral gene. Alternatively, duplication may have been followed by loss of GBF genes. Current knowledge of the phylogenetic branching of alveolates relative to the plants and animal/fungi branches does not permit resolution between those two possibilities.

GBGs in plants: refining the functional evolution of BIGs and GBFs

In fungi and mammals, BIGs and GBFs are represented by only one or two members, whose functions in vesicular trafficking at the Golgi within each group appear largely overlapping [12,14]. In contrast, plants encode a large number of GBGs in both the BIG and GBF branches but lack other ArfGEFs (Figures 1, 8). In Arabidopsis, none of the GBGs map to duplicated chromosomes where identical functions may be encoded [26,29]. In addition, comparative analysis with the rice genome nearing completion identifies at least five branches each represented by one rice and one or two Arabidopsis homologs (Figure 8). This correspondence between two highly divergent plant species indicates that GBGs diversified early during plant evolution, probably reflecting functional specialization along with the establishment of plant multicellularity. While GNOM has a plant-specific function in recycling plasma-membrane proteins needed for cell-cell communication and cell polarity establishment [11], possibly closer to the function of EFA6 or CYH subfamilies in metazoans, other plant GBGs are expected to fulfill the presumed ancient function of regulating Golgi trafficking exemplified by mammalian and yeast GBGs. Comparison of orthologous pairs in plants further reveals that they

| | Change in protein | Apparent cause of variation in transcript |
|------|--|--|
| GBFI | Extra Q at 337, 55 residues upstream of HUS domain | Insertion of 3 nucleotides (nt) resulting from use of alternate 3' |
| | New Ser and loss of 14 residues at 613, between HUS and Sec7 domains | Loss of 36 nt resulting from use of alternate 5' donor site within exon 15 during splicing with exon 16 |
| | Loss of VSQD at 1494, 38 residues upstream of HDS3 | Loss of 12 nt resulting from use of alternate 5' donor site within exon 33 during splicing with exon 34 |
| | Frame-shift at 1625 causing loss of last 19 residues of HDS3 | Intron retention between exons 36 and 37 leading to frame shift and premature termination |
| | Loss of 38 residues starting at 1784, near C-terminus | Loss of 114 nt resulting from use of novel cryptic splice donor and acceptor sites within exon 40. |
| BIGI | Frame-shift at 1340, 32 residues upstream of HDS3 | Loss of 59 nt resulting from use of alternate 5' donor site within exon 28 during splicing with exon 29 |
| | Loss of VSEKPL at 1557, 68 residues downstream of HDS3 | Loss of 18 nt resulting from use of alternate 5' donor site within exon 33 during splicing to exon 34 |
| | New T and loss of 33 residues at 1607, 118 residues downstream of HDS3 | Loss of 96 nt resulting from use of alternate 3' acceptor site within exon 35 during splicing with exon 34 |
| BIG2 | Frame-shift at 1542, 106 residues downstream of HDS3 | Loss of exon 35 resulting from splicing of 5'donor site of exon 34 with 3' acceptor site of exon 36 |

Table 2: Alternate splice variants of human GBFI, BIGI and BIG2 a,b

^a All changes were expressed relative to the reference sequence stored under accession number NM_004193 (hGBF1), NM_006421 (hBIG1) and NM_006420.1 (hBIG2).

^b All variants are supported by one or more cDNA/ESTs as detailed in the Aceview for each gene that can be obtained at [38].

have different sensitivities to Brefeldin A (a widely used fungal inhibitor of Golgi traffic) as predicted from the sequences of the binding site of the drug carried by the Sec7 domain [6]. This observation clearly illustrates that differences in outcome following BrefeldinA treatment may not reflect differences in underlying molecular mechanisms, but instead simply reflect neutral sequence differences at the Sec7 domains between species. In particular, not all BIGs may be BFA-sensitive or GBFs BFA-resistant, unlike suggested by their original nomenclature.

A novel ArfGEF subfamily in alveolates

A remarkable evolutionary feature of ArfGEFs is that while GBGs seem to be ubiquitous to all eukaryotes, fungi and animals kingdoms evolved their own ArfGEFs subfamilies unrelated to those of the other kingdoms. We thus addressed the question of whether Paramecium, which has a large number of GBGs (at least five, of which four are present as pairs as the result of recent duplications) but appears to lack the specialization into the BIG and GBF subgroups, has the same ArfGEF distribution as plants or features a second ArfGEF subfamily. We thus searched the newly sequenced genome from Paramecium tetraurelia and the available alveolate genomes from Cryptosporidium parvum and Tetrahymena thermophila for additional Sec7containing proteins. This identified a novel putative Arf-GEF subfamily characterized by the association of the Sec7 domain with a TBC (Tre/Bub2/Cdc16) domain (Figure 9), which was found only in the protists kingdom. The TBC domain is predicted to carry a GAP (GTPase activating protein) activity towards small G proteins of the Rab family [30], suggesting a potential crosstalk between Rab and Arf pathways. Such a relationship between these two small G proteins families, which are major regulators of membrane traffic, would not be unprecedented, as for example the SYT1 ArfGEF gene was identified in yeast by its genetic interactions with Rab proteins in the exocytic pathway [31]. Interestingly, alveolates have specialized exocytic pathways based on a membrane organelle lying beneath the plasma membrane, the trichocyst, where this unique ArfGEF family may potentially function.

Conclusion

A conserved scenario for the activation of Arf proteins by their GEFs?

The identification of a conserved modular architecture in all GBG subfamily members suggests that the mechanistic basis for their activation of Arf is likely to follow a similar scenario. Candidate functions for the conserved domains include oligomerization, the collection of input signals, membrane localization, regulation of the exchange activity, scaffolding of Arf proteins to their downstream effectors, not excluding signaling to partners outside the Arf pathways. Dimerization has been reported in the BIG subgroup for BIG1, which forms heterodimers with the highly homologous BIG2 ArfGEF [14], and in the GBF



Unrooted neighbour-joining phylogenetic tree of the BIG/GBF subfamily. Colour coding for the main groups is green for plants, marine blue for fungi, orange and red for animals, cyan for protists. Branches found in less than 60% bootstrap trials by either the neighbor-joining or the maximum likelihood method are in dotted lines. Species abbreviations are as in Table 1.

subgroup for GNOM, which forms homodimers [27]. The conservation of the DCB domain in GBGs, which is responsible for the dimerization of GNOM, suggests that such a dimerization function may be general to this domain in GBGs. Another unresolved issue is the conservation of the cellular partners effecting the functions associated with the conserved domains. Our identification of

an almost invariant motif in the HUS domain argues in favor of this domain interacting with a conserved partner. However, the ancient divergence into the BIG and GBF groups and their subsequent divergence into species-specific members suggest that specialized requirements are likely to have evolved in most organisms, possibly yielding less conserved partners outside the Sec7 and HUS



TBS: a novel ArfGEF subfamily in alveolates. Top: Domain structure of the TBS subfamily. Below: Sequences of the TBC domain from Paramecium TBS aligned with TBC domains from known RabGAPs. Secondary structures are from the crystal structure of yeast GYPI [30].

domains. Finally, whereas in plants all ArfGEFs are predicted to function according to the scheme defined by the conserved domains, other species have additional ArfGEF subfamilies with a modular architecture unrelated to that of the GBG subfamily. It is not known to what extent the GBG's scenario for Arf activation will also apply to non-GBGs ArfGEFs, acting alone or in association with protein partners. In the case of the GBGs, our definition of the structural homology domains as reported here should now provide a robust background for future investigations of their interactions and functions.

Methods

Protein sequence databases were searched with amino acid sequences from human BIG1, human GBF1 and Arabidopsis GNOM using the BLAST algorithm [32]. *Paramecium tetraurelia* genes were identified with the BLAST algorithm using genome sequence data from Genoscope [33] and manually annotated using Artemis [34]. Tetrahymena sequences were retrieved from the *Tetrahymena thermophila* genome sequencing project server [35]. Arabidopsis Genome Initiative database [36], rice sequences from the TIGR Rice annotation project [37].

Splice variants for hGBF1, hBIG1 and hBIG2 were identified from information provided under Aceview in the December (03) release for their respective listings at the NCBI [38]. Multiple sequence alignments were performed using ClustalW [39] with default alignment parameters or T-coffee [40,41]. Reliability of the alignments was evaluated according to the T-coffee score, and ranged from average to good for all predicted domains. Average sequence identities were respectively 24 % (DCB domain), 26 % (HUS domain), 44% (Sec7 domain), 26% (HDS1 domain), 28% (HDS2 domain) and 21% (HDS3 domain). Aligned sequences were displayed with ESPript [42] using a similarity global score of 0.15 calculated using the BLOSUM62 matrix. Unrooted phylogenetic trees were generated using the neighbor-joining algorithm of ClustalW excluding gapped regions, and with a maximum likelihood method using the PHYML package [43]. Phylogenetic trees for individual domains was performed on the subset of sequences used in Figure 3. The reliability of the trees was assessed by a bootstrap analysis (1000 replicates). Trees were drawn with TreeView version 1.6.6. Secondary structure predictions on aligned sequences were carried out with the PHD program along with the ClustalW multiple alignment [39]. Non-structured linkers poor in hydrophobic residues were predicted with the PONDR algorithm [44].

Abbreviations

GEF: Guanine nucleotide exchange factor. CYH: cytohesins/ARNO; EFA: Exchange Factor for Arf6; FBS: F-Box/ Sec7; TBS: TBC/Sec7; GBF: Golgi-associated BFA-resistant guanine nucleotide exchange Factor; BIG: BFA-Inhibited Guanine nucleotide exchange factor; GBG: GBF/BIG Gefs; SYT1: Suppressor of *ypt*. DCB: Dimerization/Cyclophilin Binding; HUS: Homology Upstream of Sec7; HDS: Homology Downstream of Sec7; TBC: Tre/Bub2/Cdc16; SCF: Skp1/Cull1/F box.

Authors' contributions

B.M. and V.B. carried out sequence and phylogenetic analysis. A.J. participated in the domain analysis. J.Co. annotated Paramecium sequences. D.S. and P.M. performed splicing pattern analysis. N.K. and G.J. analyzed the distribution of large ArfGEFs in plants. J.Ch. conceived and coordinated the study and wrote the manuscript.

Acknowledgements

This work was supported by a Human Frontiers in Science Program grant to G.J., P.M. and J.Ch. We thank Genoscope for access to Paramecium whole genome shotgun primary data and Linda Sperling (CNRS, Gif-sur-Yvette) for help with annotations.

References

 Cherfils J, Chardin P: GEFs: structural basis for their activation of small GTP-binding proteins. Trends Biochem Sci 1999, 24:306-311.

- 2. Chavrier P, Goud B: The role of ARF and Rab GTPases in membrane transport. *Curr Opin Cell Biol* 1999, 11:466-475.
- Chardin P, Paris S, Antonny B, Robineau S, Beraud-Dufour S, Jackson CL, Chabre M: A human exchange factor for ARF contains Sec7- and pleckstrin-homology domains. Nature 1996, 384:481-484.
- Pasqualato S, Renault L, Cherfils J: The GDP/GTP cycle of Arf proteins. Structural and biochemical aspects. The ARF Book Richard A Kahn, editor Kluwer Academic Publishers 2004:23-48.
- 5. Goldberg J: Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell* 1998, **95**:237-248.
- 6. Renault L, Guibert B, Cherfils J: Structural snapshots of the mechanism and inhibition of a guanine nucleotide exchange factor. *Nature* 2003, 426:525-530.
- Jackson CL, Casanova JE: Turning on ARF: the Sec7 family of guanine-nucleotide-exchange factors. Trends Cell Biol 2000, 10:60-67.
- Cox R, Mason-Gamer RJ, Jackson CL, Segev N: Phylogenetic analysis of Sec7-domain-containing Arf nucleotide exchangers. Mol Biol Cell 2004, 15:1487-1505.
- 9. Peyroche A, Paris S, Jackson CL: Nucleotide exchange on ARF mediated by yeast Geal protein. Nature 1996, 384:479-481.
- Claude A, Zhao BP, Kuziemsky CE, Dahan S, Berger SJ, Yan JP, Armold AD, Sullivan EM, Melancon P: GBF1: A novel Golgi-associated BFA-resistant guanine nucleotide exchange factor that displays specificity for ADP-ribosylation factor 5. J Cell Biol 1999, 146:71-84.
- Geldner N, Anders N, Wolters H, Keicher J, Kornberger W, Muller P, Delbarre A, Ueda T, Nakano A, Jürgens G: The Arabidopsis GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell* 2003, 112:219-230.
- 12. Peyroche A, Courbeyrette R, Rambourg A, Jackson CL: The ARF exchange factors Gealp and Gea2p regulate Golgi structure and function in yeast. J Cell Sci 2001, 114:2241-2253.
- Mansour SJ, Skaug J, Zhao XH, Giordano J, Scherer SW, Melançon P: p200 ARF-GEP1: a Golgi-localized guanine nucleotide exchange protein whose Sec7 domain is targeted by the drug brefeldin A. Proc Natl Acad Sci U S A 1999, 96:7968-7973.
- 14. Yamaji R, Adamik R, Takeda K, Togawa A, Pacheco-Rodriguez G, Ferrans VJ, Moss J, Vaughan M: Identification and localization of two brefeldin A-inhibited guanine nucleotide-exchange proteins for ADP-ribosylation factors in a macromolecular complex. Proc Natl Acad Sci U S A 2000, 97:2567-2572.
- Achstetter T, Franzusoff A, Field C, Schekman R: SEC7 encodes an unusual high molecular weight protein required fir membrane traffic from the yeast Golgi apparatus. J Biol Chem 1988, 263:11711-11717.
- Nagai H, Kagan JC, Zhu X, Kahn RA, Roy CR: A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes. *Science* 2002, 295:679-682.
 Donaldson JG: Multiple roles for Arf6: sorting, structuring, and
- Donaldson JG: Multiple roles for Arf6: sorting, structuring, and signaling at the plasma membrane. J Biol Chem 2003, 278:41573-41576.
- Zhao X, Lasell TK, Melançon P: Localization of large ADP-ribosylation factor-guanine nucleotide exchange factors to different Golgi compartments: evidence for distinct functions in protein traffic. *Mol Biol Cell* 2002, 13:119-133.
- 19. Shin HW, Morinaga N, Noda M, Nakayama K: **BIG2**, a guanine nucleotide exchange factor for ADP-ribosylation factors: its localization to recycling endosomes and implication in the endosome integrity. *Mol Biol Cell* 2004, **15**:5283-5294.
- 20. Cullen PJ, Chardin P: Membrane targeting: what a difference a G makes. Curr Biol 2000, 10:R876-8.
- Derrien V, Couillault C, Franco M, Martineau S, Montcourrier P, Houlgatte R, Chavrier P: A conserved C-terminal domain of EFA6-family ARF6-guanine nucleotide exchange factors induces lengthening of microvilli-like membrane protrusions. J Cell Sci 2002, 115:2867-2879.
- 22. Chen EH, Pryce BA, Tzeng JA, Gonzalez GA, Olson EN: Control of myoblast fusion by a guanine nucleotide exchange factor, loner, and its effector ARF6. *Cell* 2003, 114:751-762.
- 23. Ilyin GP, Rialland M, Pigeon C, Guguen-Guillouzo C: cDNA cloning and expression analysis of new members of the mammalian F-box protein family. *Genomics* 2000, **67:**40-47.

- 24. Mansour M, Lee SY, Pohajdak B: The N-terminal coiled coil domain of the cytohesin/ARNO family of guanine nucleotide exchange factors interacts with the scaffolding protein CASP. J Biol Chem 2002, 277:32302-32309.
- 25. Lee SY, Pohajdak B: N-terminal targeting of guanine nucleotide exchange factors (GEF) for ADP ribosylation factors (ARF) to the Golgi. J Cell Sci 2000, 113 (Pt 11):1883-1889.
- Jürgens G, Geldner N: Protein secretion in plants: from the trans-Golgi network to the outer space. *Traffic* 2002, 3:605-613.
- Grebe M, Gadea J, Steinmann T, Kientz M, Rahfeld JU, Salchert K, Koncz C, Jürgens G: A conserved domain of the arabidopsis GNOM protein mediates subunit interaction and cyclophilin 5 binding. *Plant Cell* 2000, 12:343-356.
- Jochum A, Jackson D, Schwarz H, Pipkorn R, Singer-Kruger B: Yeast Ysl2p, homologous to Sec7 domain guanine nucleotide exchange factors, functions in endocytosis and maintenance of vacuole integrity and interacts with the Arf-Like small GTPase Arllp. Mol Cell Biol 2002, 22:4914-4928.
- 29. Initiative AG: Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 2000, **408**:796-815.
- Rak A, Fedorov R, Alexandrov K, Albert S, Goody RS, Gallwitz D, Scheidig AJ: Crystal structure of the GAP domain of Gyp1p: first insights into interaction with Ypt/Rab proteins. Embo J 2000, 19:5105-5113.
- Jones S, Jedd G, Kahn RA, Franzusoff A, Bartolini F, Segev N: Genetic interactions in yeast between Ypt GTPases and Arf guanine nucleotide exchangers. *Genetics* 1999, 152:1543-1556.
- 32. Bork Group's WU-BLAST2 Search Service at EMBL [http:// dove.embl-heidelberg.de/Blast2/]
- 33. Paramecium Genomics [http://paramecium.cgm.cnrs-gif.fr/]
- 34. The Sanger Institute: Informatics Software: Artemis [http:// www.sanger.ac.uk/Software/Artemis/]
- 35. The TIGR Tetrahymena thermophila Genome Project [http://www.tigr.org/tdb/e2k1/ttg/]
- 36. TAIR BLAST [http://www.arabidopsis.org/Blast/]
- 37. TIGR Rice Genome Annotation [http://www.tigr.org/tdb/e2k1/ osa1/]
- 38. AceView [http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/]
- 39. Pole Bioinformatique Lyonnais [http://pbil.univ-lyon1.fr/]
- Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-217.
- 41. **Tcoffee** [http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee cgi/index.cgi]
- Easy execution of ESPript 2.x / ENDscript 1.x [<u>http://ribos</u> ome.toulouse.inra.fr/ESPript/cgi-bin/ESPript.cgi]
- 43. PHYML: fast, accurate estimation of large PHYlogenies by Maximum Likelihood [http://atgc.limm.fr/phyml/]
- Dunker AK, Brown CJ, Lawson JD, lakoucheva LM, Obradovic Z: Intrinsic disorder and protein function. Biochemistry 2002, 41:6573-6582.
- Padilla PI, Chang MJ, Pacheco-Rodriguez G, Adamik R, Moss J, Vaughan M: Interaction of FK506-binding protein 13 with brefeldin A-inhibited guanine nucleotide-exchange protein 1 (BIG1): effects of FK506. Proc Natl Acad Sci U S A 2003, 100:2322-2327.
- Li H, Adamik R, Pacheco-Rodriguez G, Moss J, Vaughan M: Protein kinase A-anchoring (AKAP) domains in brefeldin A-inhibited guanine nucleotide-exchange protein 2 (BIG2). Proc Natl Acad Sci U S A 2003, 100:1627-1632.
- Chantalat S, Park SK, Hua Z, Liu K, Gobin R, Peyroche A, Rambourg A, Graham TR, Jackson CL: The Arf activator Gea2p and the Ptype ATPase Drs2p interact at the Golgi in Saccharomyces cerevisiae. J Cell Sci 2004, 117:711-722.
- Chantalat S, Courbeyrette R, Senic-Matuglia F, Jackson CL, Goud B, Peyroche A: A novel Golgi membrane protein is a partner of the ARF exchange factors Gealp and Gea2p. Mol Biol Cell 2003, 14:2357-2371.
- Charych El, Yu W, Miralles CP, Serwanski DR, Li X, Rubio M, De Blas AL: The brefeldin A-inhibited GDP/GTP exchange factor 2, a protein involved in vesicular trafficking, interacts with the beta subunits of the GABA receptors. J Neurochem 2004, 90:173-189.
- Garcia-Mata R, Sztul E: The membrane-tethering protein p115 interacts with GBF1, an ARF guanine-nucleotide-exchange factor. EMBO Rep 2003, 4:320-325.



51. Saeki N, Tokuo H, Ikebe M: **BIGI** is a binding partner of myosin IXB and regulates its Rho gap activity. *J Biol Chem* 2005.