Comparative analyses of the Conserved Oligomeric Golgi (COG) complex in vertebrates

Rita Quental*1,2, Luísa Azevedo1, Rune Matthiesen1 and António Amorim1,2

Abstract

Background: The Conserved Oligomeric Golgi (COG) complex is an eight-subunit assembly that localizes peripherally to Golgi membranes and is involved in retrograde vesicular trafficking. COG subunits are organized in two heterotrimeric groups, Cog2, -3, -4 and Cog5, -6, -7, linked by a dimeric group formed by Cog1 and Cog8. Dysfunction of COG complex in humans has been associated with new forms of Congenital Disorders of Glycosylation (CDG), therefore highlighting its essential role. In the present study, we intended to gain further insights into the evolution of COG subunits in vertebrates, using comparative analyses of all eight COG proteins.

Results: We used protein distances and $d_{\text{N}}/d_{\text{S}}$ ratios as a measure of the rate of proteins evolution. The results showed that all COG subunits are evolving under strong purifying selection, although COG1 seems to evolve faster than the remaining proteins. In addition, we also tested the expression of COG genes in 20 human tissues, and demonstrate their ubiquitous nature.

Conclusions: COG complex has a critical role in Golgi structure and function, which, in turn, is involved in protein sorting and glycosylation. The results of this study suggest that COG subunits are evolutionary constrained to maintain the interactions between each other, as well with other partners involved in vesicular trafficking, in order to preserve both the integrity and function of the complex.

Background

Most cellular processes are carried out by multiprotein complexes that constitute important functional units in the cell [1]. This fact has motivated a number of studies aiming to investigate the structure, function and evolution of such multisubunit molecular machines [e.g., [1-4]].

A cellular process in which protein complexes are known to be involved is the transport of proteins between cellular compartments (vesicular trafficking) [5,6]. Proteins synthesised in the secretory pathway are transported inside vesicles that move from the endoplasmic reticulum to the Golgi apparatus, from where polypeptides are then sorted to several cellular compartments [7]. As progression through the Golgi occurs, proteins may undergo modifications like glycosylation, a necessary step for their stability and function [8]. Several large protein complexes play an important role in the fidelity of vesicle fusion, acting as tethering factors through the formation of physical links between membranes prior to fusion [5,6,9]. One of these is the Conserved Oligomeric Golgi (COG) complex [10], which localizes at the cytoplasmic surface of the Golgi apparatus [11-14].

Several studies have been performed demonstrating the involvement of COG in retrograde vesicular trafficking of Golgi resident proteins [15-18], including enzymes that participate in glycans biosynthesis [19,20]. Consequently, COG impairment results in abnormal Golgi morphology (dilated cisternae and accumulation of vesicles [10,17,21,22]) and function (glycosylation defects [reviewed in [23] and references therein]). However, its precise mechanism of action is not completely understood.

COG complex is composed by eight distinct subunits [10,12,13,24-26], Cog1 to Cog8, arranged in two lobes consisting of Cog1 to Cog4 (lobe A) and Cog5 to Cog8 (lobe B) [10]. Although several models have been advanced refining the architecture of COG [22,27-29], the most recent studies converge in suggesting that mammalian COG members are organized in two heterotri-
meric groups, Cog2-Cog3-Cog4 and Cog5-Cog6-Cog7, which are linked by the dimeric group formed by Cog1 and Cog8. In particular, Cog1 associates with Cog2-Cog3-Cog4, whereas Cog8 interacts with Cog5-Cog6-Cog7 (Figure 1) [22,29].

Previous studies have shown the possibility that COG components have different roles within the complex, since mutations or deletion of individual subunits cause sharply distinct phenotypes [25,26]. In yeast, the deletion of any one of the four lobe A subunits causes a severe growth defect, whilst disruption of the remaining genes (COG5 to COG8) does not substantially interfere with normal cell growth [25]. In mammals, COG1- and COG2-deficient cells present several dilated cisternae [10] and pleiotropic defects in the synthesis of N-, O- and lipid-linked glycans [8]. A similar phenotype is observed in COG5-deficient cells, although the alterations in glycosylation are subtle [22]. In addition, COG3 depletion entails the accumulation of vesicles distributed throughout the cytoplasm [17]. More recently, COG dysfunction caused by mutations in specific subunits has been associated with new forms of Congenital Disorders of Glycosylation (CDG) in humans [reviewed in [30,31]].

In the present study, we used a broad range of comparative sequence analyses to track the evolutionary profile of this complex in vertebrates.

Methods

In silico orthologs retrieval

Human protein sequences corresponding to each COG subunit (COG1-COG8) were retrieved from the National Center for Biotechnology Information (NCBI) [32]. These sequences were used as queries to search orthologous proteins from the reference proteins database (RefSeq) using the BLASTP algorithm [33], with an E-value cutoff of 10^-3 and using the reciprocal best-hit approach [34]. Alternatively, sequences were retrieved from Ensembl genome browser [35], release 57 from March 2010, via the orthologues option. Available sequences from the following taxa were considered in this study: mammals (Homo sapiens, Pan troglodytes, Pongo pygmaeus, Macaca mulatta, Callithrix jacchus, Mus musculus, Rattus norvegicus, Oryctolagus cuniculus, Equus caballus, Canis familiaris, Bos taurus and Monodelphis domestica), birds (Gallus gallus and Taeniopygia guttata), reptiles (Anolis carolinensis), amphibians (Xenopus tropicalis) and fishes (Danio rerio, Takifugu rubripes, Gasterosteus aculeatus, Tetraodon nigroviridis and Oryzias latipes). COG3 from G. aculeatus, COG5 from X. tropicalis and COG6 from D. rerio were assembled based on the genomic sequence through the comparison with the orthologous protein from other species, and through TBLASTN searches on ESTs database. Accession numbers are available in the additional file 1. In addition, due to probable assembling errors, residues 82 to 124 from P. troglodytes COG4, residues 57 to 146 and 436 to 470 from P. pygmaeus COG7 as well as residues 493 to 522 from M. mulatta COG7 were replaced by missing data.

Protein Distances

Protein sequences were aligned with MUSCLE [36] and manually refined by removing sites at which all sequences except one or two have alignment gaps. Additionally, the initial and terminal regions of the multiple sequences alignment had, in some cases, to be removed because they were poorly aligned. This step was particularly important in COG5 and COG8. The final alignments of each group of orthologous sequences are available at additional file 2. Protein distances were calculated using Protdist from PHYLIP 3.69 package [37], with the Jones-Taylor-Thornton (JTT) evolutionary model and gamma distribution of rates with a fixed shape parameter of 1. Divergence times between species were obtained with TimeTree [38,39].

Phylogenetic analysis and d_0/d_6 (ω) ratios

To reconstruct species phylogeny, the protein-coding sequences were aligned with ClustalW program [40] implemented in Bioedit version 7.0.9.0 [41], using protein alignments as template to avoid out-of-frame gaps. The poorly aligned positions and divergent regions in each alignment were eliminated with GBLOCKS program [42] and the resulting blocks were concatenated in a single alignment of 15195 positions. To identify the model of nucleotide substitution that best fits the data the Akaike Information Criterion (AIC) was applied, using the jModelTest 0.1.1 [43]. The selected model (GTR +I +G) was used to reconstruct the maximum likelihood phylogeny in Phyml 3.0 [44]. The tree was drawn with FigTree
program [45]. The resulting tree topology, but not branch lengths, was used to fit different models in PAML.

The number of synonymous substitutions per synonymous site \( (d_S) \) and the number of nonsynonymous substitutions per nonsynonymous sites \( (d_N) \) have been estimated with CODEML from the PAML v.4.4 package [46], using the F3 × 4 codon frequency model and treating alignment gaps as ambiguity characters \( (\text{clean data} = 0) \). The input alignments were modified by removing positions that showed evidence to represent true indels, while keeping those that appear to be missing data. Alignments are available in additional file 3. Several models that allow for different levels of heterogeneity in the \( d_S/d_N \) ratio among lineages have been applied: the one-ratio model that assumes the same \( \omega \) ratio for all branches in the phylogeny; the free-ratios model that allows \( \omega \) to vary on every lineage; and the two-ratio model, which assumes that the branch of interest has an \( \omega \) value \( (\omega_1) \) different from the ratio of the other lineages \( (\omega_0 - \text{background ratio}) \). The above models can be compared using a likelihood ratio test \( (\text{LRT}) \) to test different hypothesis, as described by Yang [47]. Because synonymous sites saturation prevents comparisons of too divergent sequences, only species from humans to birds were considered in this analysis.

Expression analyses

Expressed sequence tag (EST) profiles from several human and murine tissue samples were extracted for COG genes from the UniGene database [48] as EST counts per million transcripts and were log2 transformed \( (\text{Homo sapiens: UniGene Build } \#223 \text{ and Mus musculus: UniGene Build } \#183) \). A number of erroneously assigned ESTs for human and mouse COG8 were manually removed. For simplicity, only homologous tissues for which information was available in both organisms were included, resulting in a total of 29 tissues. The clusters on the heatmaps were made by an in house tool, using correlation as the measure of similarity.

In addition, because in several tissues some COG genes have no detectable expression \( (\text{absence of EST counts}) \), we have further tested the presence of all COG transcripts in 20 different human tissues included in a RNA panel obtained from Ambion \( (\text{FirstChoice Human Total RNA Survey Panel}) \). cDNA was synthesized from 1 μg of RNA using the First Strand cDNA synthesis Kit \( (\text{Fermentas Life Sciences, Burlington, Ontario, Canada}) \), according to the manufacturer’s instructions. Primers specific for COG transcripts, as well as for the positive control \( \text{GAPDH} \), were designed to avoid amplification of contaminating genomic DNA, either because they span an intron or the forward primer anneals with an exon/exon boundary \( (\text{additional file } 4) \). PCR amplifications were performed using the QIAGEN multiplex PCR kit \( (\text{Qiagen, Hilden, Germany}) \) at 1× Qiagen multiplex PCR master mix with 0.5 μl of cDNA in a 12.5 μl final reaction volume. Final primer concentration in the reaction was 0.4 μM. Thermocycling conditions used included pre-incubation for 15 min at 95°C, followed by 35 cycles of 30 s at 94°C, 90 s at 58°C, and 60 s at 72°C, with a final incubation for 10 min at 72°C. Amplification products ranged from 100 to 253 bp and were separated by horizontal electrophoresis in 12% polyacrylamide gels and visualized by silver staining. RT-PCR products from one sample were confirmed by direct sequencing.

Results

Evolutionary Analyses

As a first measure of the rate of COG proteins evolution, we calculated the pairwise protein distance for human proteins and each of the 20 orthologs. These values were plotted against the corresponding divergence times for the compared species, and the linear regression trend line was estimated from each group, as shown in Figure 2. From the slope of the lines and the r² values we are able to compare the rate of proteins evolution and its constancy over time, respectively. The results presented in Figure 2 show that COG1 is the subunit with the fastest rate of evolution; COG6, COG3, COG5 and COG4 have the lowest rates; while the remaining proteins \( (\text{COG7, COG8 and COG2}) \) have intermediate rates. Although being the most divergent protein, COG1 \( (\text{along with COG7}) \) is the one in which the rate of evolution has remained most constant \( (r^2 = 0.993) \).

In addition, we used a classical measure of protein evolution based on the nonsynonymous \( (d_N) \) to synonymous \( (d_S) \) substitutions rate ratio \( (d_S/d_N) \) or \( \omega \).

A \( \omega \) value higher than 1 can suggest that genes undergo positive selection, while less than 1 is indicative of purifying selection [49]. Following this approach, we tested whether \( \omega \) ratios for each COG gene are different among lineages, based on the maximum likelihood phylogeny previously inferred \( (\text{Figure } 3) \). Therefore, a likelihood ratio test \( (\text{LRT}) \) comparing the one-ratio model, that assumes the same \( \omega \) for all lineages, and the free-ratio model, which assumes independent \( \omega \) ratios for every branch, was applied. The log likelihoods obtained under each model are presented in Table 1 and indicate significant variation in \( \omega \) values among lineages in all COG genes except for COG7, suggesting relaxation of the strong selective constraints in some lineages, yet with low \( \omega \). The \( \omega \) values for branches in the phylogeny for each gene are available at additional file 5. It is interesting to note that the length of the branch that leads to modern rodents \( (\text{mouse and rat}) \) obtained for COG2 and COG6 is very long, revealing the accumulation of many substitu-
tions (additional file 5). Notwithstanding, the corresponding ω values are low, even when compared with shorter branches represented by non-rodent species. This suggests that rodent lineage has accumulated mainly synonymous rather than nonsynonymous substitutions, thus preserving the amino acid composition of the encoded protein. In fact, when the same COG2 and COG6 trees were drawn with branches lengths proportional to the expected nonsynonymous substitutions rate \( (d_{N}) \), the rodent’s branch length is more similar to all the other branches (additional file 6).

Although the ω ratio obtained under the one-ratio model does not fit every branch in the phylogeny, it represents an average over all sites and lineages \([50]\) and therefore can be used to compare the strength of constraints imposed to different COG genes. As presented in Table 1, all ω values are very low, indicating that COG genes are evolving under strong purifying selection. The highest ratio is observed for COG1, while the lowest refers to COG4 and COG6.

Expression analyses in human and mouse
Although several studies have been accumulating in the last years about COG complex, particularly in what concerns to the interaction between subunits, thus far the expression profile of different COG genes remains uncharacterized. Therefore, as a preliminary approach to study the expression of COG genes, tissue dependent expression patterns have been inferred from EST profiles accessible in UniGene database \([48]\). For comparative purposes, the analysis was performed in homologous tissues in human and mouse for which expression information was available for both organisms.

![Figure 2](image2.png)

**Figure 2** Protein distances vs divergence times between human and different species. Divergence times between humans and other species were obtained with TimeTree \([38]\) and are as follow: Ptr: 6.13 Myr; Ppy: 15.44 Myr; Mac: 29.6 Myr; Cja: 44.2 Myr; Gires (average of Mmu, Rno and Ocu): 91 Myr; Laurasatiheria (average of Eca, Bta and Cfa): 97.4 Myr; Mdo: 176.1 Myr; Sauropsida (average of Tgu, Gga and Aca): 324.5 Myr; Xtr: 361.2 Myr; Actinopterygii (average of Dre, Ola, Gac, Tni and Tru): 454.6 Myr. Linear regression trend lines were set to intercept the origin.

![Figure 3](image3.png)

**Figure 3** Maximum Likelihood phylogeny of the 14 species used in the analyses with PAML. Maximum Likelihood phylogenetic tree was constructed based on the concatenated alignment of the eight independent protein coding sequences alignment (see methods). Tree was midpoint rooted. Bootstrap values are shown at nodes and were calculated from 100 replicates. This tree topology was used to fit different models in PAML.
Figure 4 shows the clustering of gene-expression data for human and mouse. In general, COG genes appear to have a ubiquitously pattern of expression, yet differences in the level of expression can be observed. However, some of the tissues studied (e.g. adipose tissue) have no detectable expression of specific COG genes. In order to assess the presence of COG transcripts in some of those and other human tissues, we performed reverse transcription-PCR (RT-PCR). Although no quantitative inferences could be made, using this simple methodology we were able to detect the eight COG transcripts in 20 human tissues (Figure 5) thereby confirming their ubiquitous nature.

It is important to recognize, however, that from EST data and RT-PCR analysis we are not able to infer the precise pattern of expression of COG genes, revealing the need for more reliable quantitative data.

**Discussion**

COG complex is essential to establish and maintain the structure and function of the Golgi apparatus, which has itself a key role in many cellular processes, such as protein sorting and glycosylation.

In the present study, in order to better understand the evolution of COG subunits in vertebrates, we have applied distinct comparative strategies, including evolutionary and expression analyses.

We demonstrate that all COG proteins are evolving under strong evolutionary constraints, as revealed by the low $d_s/d_3$ values. This pattern of purifying selection must reflect the critical role of COG complex for Golgi function. This is well illustrated by mutations in COG-specific subunits, which give rise to different human diseases belonging to the Congenital Disorders of Glycosylation (CDG). CDGs are a genetically heterogeneous group of disorders characterized by a deficient glycosylation of glycoconjugates, such as proteins and lipids. Since 2004, defects in COG1 [51], COG4 [52], COG5 [53], COG7 [19,54-56] and COG8 [57,58] have been reported. Recently, a novel mutation in COG1 gene was detected in two patients with a cerebrosplenialdinal-like syndrome [59], showing that the impact of COG dysfunction is far from being completely known.

Being part of a multi-subunit assembly and having such an important functional role in cells must impose strong constraints on the evolution of COG proteins. On one hand, they must be constrained to maintain the structural integrity of the complex, presumably through the conservation of the residues that are involved in the interaction between subunits. This is expected to be true if we assume that COG structure is maintained by the same type of protein-protein interactions in different species. On the other hand, additional interactions with other functional partners (e.g. other protein related to trafficking, such as SNAREs or small GTPases [reviewed in [60]]) also need to be preserved. The study of the crystallographic structure of the C-terminal region of human COG4 protein, for instance, showed that distinct domains are responsible for the integration of the protein within the complex and for its function [61]. This suggests that a large proportion of the protein sequence of each member of COG complex must be constrained to be evolutionarily conserved.

In fact, the low rate of evolution of COG proteins is consistent with results from more comprehensive studies showing that evolutionary conservation increases from monomeric proteins to members of transient interactions and finally to components of stable complexes (proteins that are permanently associated with each other) [1,62]. Wong and collaborators [4] also demonstrated that as the number of unique proteins in a complex increases, the mean $d_s/d_3$ ratio of the associated genes tends to decrease.

| COG | $\omega^a$ | $\ell_{\text{r}}$ | $\ell_{\text{f}}$ | $2\Delta<2(\ell_{\text{r}}-\ell_{\text{f}})$ |
|-----|------|-----|-----|----------|
| COG1 | 0.150 | -18100.05 | -18074.99 | 50.11* |
| COG2 | 0.126 | -11122.36 | -11097.92 | 48.89* |
| COG3 | 0.084 | -10626.26 | -10606.19 | 40.14* |
| COG4 | 0.053 | -9723.53 | -9696.17 | 54.71* |
| COG5 | 0.094 | -11639.04 | -11587.57 | 102.95* |
| COG6 | 0.071 | -8462.90 | -8429.86 | 66.07* |
| COG7 | 0.080 | -11649.63 | -11632.14 | 34.99 |
| COG8 | 0.094 | -9163.24 | -9126.69 | 73.10* |

- \(\omega\): ratio obtained under the one-ratio model
- \(*\): Significant (\(P < 0.05; \chi^2_{24} = 36.415\); \(\ell\): Log likelihood values

**Table 1: Log-likelihood values under the one-ratio and free-ratio models and likelihood ratio statistics (2\(\Delta\))**
To evaluate the impact of different selective forces on COG proteins, it would be interesting to compare the rates of substitution for interacting and non-interacting residues, and also structural and functional domains. Unfortunately, the structure of fragments of only two COG subunits have been reported [61,63], hampering us to analyze with more detail the evolution of distinct regions of each protein.

Despite all COG proteins are evolving under strong selective constraints, COG1 seems to be the one with the highest rate of evolution. This subunit, together with COG8, is the bridging subunits of the mammalian COG complex, bringing together COG2-4 and COG5-7 subcomplexes [22,29]. Interestingly, a quite similar interaction map has been reported in the yeast complex, although in this case only COG1 is required for the association of the two subcomplexes [28]. COG1 from humans and COG1p from yeast share no detectable sequence homology, as happens with COG2 and COG7 [25]. Interestingly, our results revealed that COG1 and COG2 are also the least conserved subunits in vertebrates, suggesting that they are evolving under more relaxed
selective constraints. The biological implication of the higher divergence of these proteins is difficult to infer, although we can speculate that it might be related with distinct requirements of COG’s function in different species, or with the interactions established by these proteins. The remaining COG proteins (COG3, COG4, COG5, COG6 and COG8), in contrast, have related homologs in human and yeast [25].

In this study we have also demonstrated that the expression of COG genes exhibit a ubiquitous nature. These results can be taken as a starting point for more detailed quantitative expression studies that can bring additional insights into COG subunits interaction and function and, eventually, to the understanding of the phenotypic heterogeneity associated with different COG defects.

Conclusions
In the past years several studies have been focused on the evolution of protein complexes in terms of type of interactions, revealing that proteins in stable complexes are more conserved than those in transient interactions and those with no apparent interacting partners [62].

In this study, in turn, we have investigated the evolution of different subunits belonging to the same protein complex in vertebrates. Our results showed that the eight COG subunits seem to be conserved and evolving under strong purifying selection, in order to maintain the integrity and function of the complex. Finally, we confirm the ubiquitous tissue expression of the eight COG transcripts in 20 human tissues.

Additional material

Additional file 1 Protein accession numbers of COG sequences. This table provides the accession numbers of the sequences used to perform the analyses.

Additional file 2 Protein sequences alignments used to calculate the proteins distance with ProtDist

Additional file 3 Protein-coding sequences alignments used in the different analyses performed with PAML

Additional file 4 Primer pairs used in PCR reactions. This table provides the sequences of the primer pairs used to detect the presence of COG transcripts in 20 human tissues.

Additional file 5 Phylogenetic trees with branches drawn in proportion to their lengths, defined as the expected number of nucleotide substitutions per codon, for each COG gene. Values along each branch represent ω rates (for simplicity of the figure some of them were not indicated) estimated under the free-ratio model. In COG5 phylogeny the branch leading to Euarchontoglires (primates and Glires) presents a ω ratio of 4.70 (dυ = 0.0011; dσ = 0.002). A LRT comparing the one-ratio model with the two-ratio model (ωbackground, ωEuarchontoglires), revealed that the estimated ω ratio for the Euarchontoglires branch (infinite, indicating the absence of synonymous substitutions) was not significantly higher than the background ratio (0.004). Moreover, the LRT comparing the two-ratio model with and without the constraint ωEuarchontoglires ≤ 1 revealed that this ratio was not significantly higher than 1 as well.

Additional file 6 COG2 and COG6 genes phylogeny with branch lengths defined as the estimated nonsynonymous substitutions rate (dυ) or the estimated synonymous substitutions rate (dσ).

Abbreviations
COG, conserved oligomeric Golgi complex; CDG, congenital disorders of glycosylation; EST, expressed sequence tag. Taxa terminology is abbreviated using the first letter of the genus and two letters of the species name (e.g. Hsa corresponds to Homo sapiens).

Authors’ contributions
RQ, LA and AA conceived the study and main analyses. RM and RQ carried out the expression analysis. RQ, LA, RM and AA analyzed and interpreted the data. RQ and AA wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by the research project PTDC/CVT/64154/2006. LA (C2007-IPATIMUP/AA1) and RM (C2007-IPATIMUP/AA2) are supported by the Portuguese Foundation for Science and Technology (FCT) Ciência 2007. RQ (SFRH/BD/23657/2005) is a PhD grantee also from FCT. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT.

Author Details
1. IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal and 2. Faculty of Sciences, University of Porto, Porto, Portugal

Received: 27 July 2009 Accepted: 15 July 2010
Published: 15 July 2010

References
1. Pereira-Leal JB, Levy ED, Teichmann SA. The origins and evolution of functional modules: lessons from protein complexes. Philos Trans R Soc Lond B Biol Sci 2006, 361(1467):507-517.
2. Minteris J, Weng Z. Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci USA 2005, 102(31):10930-10935.
3. van Dam TJ, Snel B. Protein complex evolution does not involve extensive network rewiring. PLoS Comput Biol 2008, 4(7):e1000132.
4. Wong P, Althammer S, Hildebrand A, Kirschner A, Pagel P, Geissler B, Smialowski P, Blochl F, Oesterheld M, Schmidt T, Strack N, Theis F, Ruepp A, Frishman D. An evolutionary and structural characterization of mammalian protein complex organization. BMC Genomics 2008, 9(1):529.
5. Whyte JRC, Munro S: Vesicle tethering complexes in membrane traffic. J Cell Biol 2002, 157(3):267-267.
6. Oka T, Krieger M: Multi-Component Protein Complexes and Golgi Membrane Trafficking. J Biochem 2005, 137(2):109-114.
7. Rothman JE: Mechanisms of intracellular protein transport. Nature 1994, 372(6586):55-63.
8. Kingsley D, Kazanaki K, Segal M, Krieger M: Three types of low density lipoprotein receptor-deficient mutant have pleiotropic defects in the synthesis of N-linked, O-linked, and lipid-linked carbohydrate moieties. J Cell Biol 1986, 102(5):1576-1585.
9. Lushpich V, Szul P: Golgi tethering factors. Biochim Biophys Acta 2005, 1744(3):325-339.
10. Ungar D, Oka T, Brittle EJ, Vasile E, Lupashin VV, Chatterton JE, Heuser JE, Krieger M, Waters MG: Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. J Cell Biol 2002, 157(3):405-415.
11. Podos S, Reddy P, Ashkenas J, Krieger M: DLCC encodes a brefeldin A-sensitive, peripheral Golgi protein required for normal Golgi function. J Cell Biol 1994, 127(5):769-791.
12. Walter DM, Paul KS, Waters MG: Purification and characterization of a novel 13 S hetero-oligomeric protein complex that stimulates in vitro Golgi transport. J Biol Chem 1998, 273(45):29565-29576.
13. Chatterton JE, Hirsch D, Schwartz JJ, Bickel PE, Rosenberg RD, Lodish HF, Paul KS, Krieger M: Expression cloning of DLDB, a gene essential for normal Golgi function and assembly of the IdtCp complex. Proc Natl Acad Sci USA 1999, 96(5):915-920.
14. Suvorova ES, Kurten RC, Lupashin VV: Identification of a human orthologue of Sec34p as a component of the cis-Golgi vesicle tethering machinery. J Biol Chem 2001, 276(25):22810-22818.
15. Ungar D, Oka T, Krieger M, Hughson FM: Retrograde transport on the COG railway. Trends Cell Biol 2006, 16(2):113-120.
16. Suvorova ES, Duden R, Lupashin VV: The Sec34p-Sec35p complex, a Ypt1p effector required for retrograde intra-Golgi trafficking, interacts with Golgi SNAREs and COPI vesicle coat proteins. J Cell Biol 2002, 157(4):631-643.
17. Zolov SN, Lupashin VV: Cog3p depletion blocks vesicle-mediated Golgi retrograde trafficking in HeLa cells. J Cell Biol 2005, 168(4):774-779.
18. Oka T, Ungar D, Hughson FM, Krieger M: The COG and COPI complexes interact to control the abundance of GEARs, a subset of Golgi integral membrane proteins. Mol Biol Cell 2004, 15(5):2423-2435.
19. Wu X, Steet RA, Bohorov O, Bakker J, Ungar D, Hughson FM, Krieger M, Slaapen L, Kornfeld S, Freeze HH: Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. J Biol Chem 2004, 279(24):27329-27339.
20. Ungar D, Oka T, Vasile E, Krieger M, Hughson FM: Subunit architecture of the conserved oligomeric Golgi complex. J Cell Biol 2005, 208(38):32729-32735.
21. Foulquier F: COG defects, birth and rise! Biochim Biophys Acta 2009, 1792(9):896-902.
22. Zeevaert R, Foulquier F, Jaken J, Matthieis G: Deficiencies in subunits of the Conserved Oligomeric Golgi (COG) complex define a novel group of Congenital Disorders of Glycosylation. Mol Genet Metab 2008, 93(1):15-21.
23. NCBI- National Center for Biotechnology Information. [http://www.ncbi.nlm.nih.gov]
24. Castresana J: Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol 2000, 17(4):540-552.
25. Posada D: ModelTest: Test Phylogenetic Model Averaging. Mol Biol Evol 2002, 19(7):204-205.
26. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003, 52(5):696-704.
27. Rambaut A: FigTree-Tree Figure Drawing Tool Version 1.2. [http://tree.bio.ed.ac.uk/software/figtree/]
28. Yang Z: PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 2007, 24(8):1586-1591.
29. Yang Z: Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 1998, 15(5):568-573.
30. Pontius J, Wagner L, Schuler G: UniGene: a unified view of the transcriptome. In The NCBI Handbook Bethesda (MD): National Center for Biotechnology Information, 2003.
31. Jobling MA, Hurles M, Tyler-Smith C: Human Evolutionary Genetics: Origins, Peoples and Disease. New York: Garland Science; 2004.
32. Pons NT, Wagner L, Schulier G: UniGene: a unified view of the transcriptome. In The NCBI Handbook Bethesda (MD): National Center for Biotechnology Information, 2003.
33. Jobling MA, Hurles M, Tyler-Smith C: Human Evolutionary Genetics: Origins, Peoples and Disease. New York: Garland Science; 2004.
34. Yang Z, Swanson WJ, Vacquier VD: Maximum-Likelihood Analysis of Molecular Adaptation in Abalone Sperm Lysin Reveals Variable Selective Pressures Among Lineages and Sites. Mol Biol Evol 2000, 17(10):1446-1455.
35. Felsenstein J: Phylogeny Inference Package version 3.69. Distributed by the author Department of Genome Sciences University of Washington, Seattle.
36. Hedges SB, Dudley J, Kumar S: TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 2006, 22(23):2971-2972.
37. Hedges SB, Kumar S, (eds): The Timetree of Life, New York: Oxford University Press, 2009.
38. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc Acids Res 1994, 22(24):4673-4680.
39. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999, 41:95-96.
54. Ng BG, Kranz C, Hagebeuk EE, Duran M, Abeling NG, Wuyts B, Ungar D, Lupashin V, Hartdorff CM, Poll-The B, Freeze HH: Molecular and clinical characterization of a Moroccan Cog7 deficient patient. Mol Genet Metab 2007, 91(2):201-204.

55. Morava E, Zeevaert R, Korsch E, Huijben K, Wopereis S, Matthijs G, Keymolen K, Lefeber DJ, De Meirleir L, Wevers RA: A common mutation in the CDG7 gene with a consistent phenotype including microcephaly, adducted thumbs, growth retardation, VSD and episodes of hyperthermia. Eur J Hum Genet 2007, 15(6):638-645.

56. Zeevaert R, Foulquier F, Cheillan D, Cloix I, Guffon N, Sturiale L, Garozzo D, Matthijs G, Jaeken J: A new mutation in COG7 extends the spectrum of COG subunit deficiencies. Eur J Med Genet 2009, 52(5):303-305.

57. Foulquier F, Ungar D, Reyners E, Zeevaert R, Mills P, Garcia-Silva MT, Briones P, Winchester B, Morelle W, Krieger M, Annaert W, Matthijs G: A new inborn error of glycosylation due to a Cog8 deficiency reveals a critical role for the Cog1-Cog8 interaction in COG complex formation. Hum Mol Genet 2007, 16(7):717-730.

58. Zeevaert R, Foulquier F, Dimitrov B, Reynders E, Van Damme-Lombaerts R, Simeonov E, Annaert W, Matthijs G, Jaeken J: Cerebrocostomandibular-like syndrome and a mutation in the conserved oligomeric Golgi complex, subunit 1. Hum Mol Genet 2009, 18(3):517-524.

59. Sztul E, Lupashin V: Role of vesicle tethering factors in the ER-Golgi membrane traffic. FEBS Lett 2009, 583(23):3770-3783.

60. Richardson BC, Smith RD, Ungar D, Nakamura A, Jeffrey PD, Lupashin VV, Hughson FM: Structural basis for a human glycosylation disorder caused by mutation of the COG4 gene. Proc Natl Acad Sci USA 2009, 106(32):13329-13334.

61. Tscherny SA: The Constraints Protein-Protein Interactions Place on Sequence Divergence. J Mol Biol 2002, 324(3):399-407.

62. Cavanaugh LF, Chen X, Richardson BC, Ungar D, Pelczer I, Rizzo J, Hughson FM: Structural analysis of conserved oligomeric Golgi complex subunit 2. J Biol Chem 2007, 282(32):23418-23426.

doi: 10.1186/1471-2148-10-212

Cite this article as: Quental et al., Comparative analyses of the Conserved Oligomeric Golgi (COG) complex in vertebrates. BMC Evolutionary Biology 2010, 10:212