Duplication of genes encoding non-clathrin coat protein γ-COP in vertebrate, insect and plant evolution

Yoonsoo Hahn[a], Young Jae Lee[a], Ji Hye Yun[a], Seung Kyoung Yang[a], Chang Won Park[a], Kazuei Mita[b], Tae-Lin Huh[c], Myungchull Rhee[d], Jae Hoon Chung[a,*]

[a] *Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Taejon 305-701, South Korea*
[b] *Genome Research Group, National Institute of Radiological Sciences, Anagawa 4-9-1, Inage-ku, Chiba 263-8555, Japan*
[c] *Department of Genetic Engineering, Kyungpook National University, Taegu 702-701, South Korea*
[d] *Department of Biology, Chungnam National University, Taejon 305-764, South Korea*

Received 6 June 2000; revised 29 August 2000; accepted 31 August 2000

Edited by Takashi Gojobori

Abstract Coatomer is a major component of COPI vesicles and consists of seven subunits. The γ-COP subunit of the coatomer is believed to mediate the binding to the cytoplasmic dilysine motifs of membrane proteins. We characterized cDNAs for *Copg* genes encoding γ-COP from mouse, zebrafish, *Drosophila melanogaster* and *Bombyx mori*. Two copies of *Copg* genes are present in vertebrates and in *B. mori*. Phylogenetic analysis revealed that two paralogous genes had been derived from a single ancestral gene by duplication independently in vertebrates and in *B. mori*. Mouse *Copg1* showed ubiquitous expression with the highest level in testis. Zebrafish *copg2* was biallelically expressed in hybrid larvae in contrast to its mammalian ortholog expressed in a parent-of-origin-specific manner. A phylogenetic analysis with partial plant cDNA sequences suggested that *copg* gene was also duplicated in the grass family (Poaceae).

Key words: Non-clathrin coat protein γ-COP; *Copg*; Gene duplication; Phylogenetic analysis

1. Introduction

In eukaryotic cells, transport of proteins and lipids between endocytic and exocytic compartments is mediated by coated vesicle carriers [1–3]. Coat complexes which participate in vesicle formation are clathrin/adaptor protein complexes [2], COPI (coat protein complex I) [4,5], COPII (coat protein complex II) [6], retromers [7], caveolin [8] and AP-3 [9]. COPI-coated vesicles are involved in protein transport in the early secretory pathway [10]. The COPI coat is composed of ADP-ribosylation factor 1 [11] and coatomer. The coatomer is a heterooligomeric protein complex consisting of seven distinct subunits, α-, β-, β′-, γ-, δ-, ε- and ζ-COP [12–14]. Coatomers interact directly with the C-terminal KKXX motif of type I transmembrane proteins and retrieve these proteins from the Golgi complex back to endoplasmic reticulum (ER) [15,16]. Photocrosslinking studies using purified coatomer suggest that γ-COP binds to the KKXX retrieval motifs and to the KKXXX motif of p23, a member of the p24 family membrane proteins enriched in COPI vesicles [17,18]. Genes encoding γ-COP were identified in *Arabidopsis thaliana*, *Bos primigenius*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and human [16,19]. Recently, the paralogous gene *COPG2* was reported to be imprinted in human [20]. In addition, we previously reported molecular cloning, genomic structure and imprinted expression of mouse *Copg2* [21].

Here, we report the identification of cDNAs for *Copg* genes encoding γ-COP proteins in vertebrates (mouse and zebrafish) and in insects (*Drosophila melanogaster* and *Bombyx mori*) by systematic searches of the expressed sequence tags database (dbEST). Tissue distribution of the mouse gene transcript, allelic expression of the zebrafish gene and the gene duplication in plants were also investigated.

2. Materials and methods

2.1. cDNA identification and sequence analyses

To obtain cDNA clones encoding γ-COP proteins from various organisms, a systematic search was performed against dbEST using amino acid sequences derived from bovine *Copg1* (GenBank accession no. X92987) or human *COPG1* (GenBank accession no. AF100756) as database queries. Sequence similarity searches were performed using BLAST programs [22] at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST). We identified several ESTs for novel γ-COP proteins from mouse (*Copg1*, AI115116), zebrafish (*copg2*, AI588473 and AI601646), *Drosophila* (*copg*, AI258296 and AI260552) and *Bombyx* (*copg1*, AU000611 and AU005766; *copg2*, AU004321). The clones were obtained from Genome Systems, Research Genetics or SilkBase (http://www.ab.a.u-tokyo.ac.jp/silkbase) [23]. The full sequences of the cDNA clones were determined by a Bigdye Terminator cycle sequencing kit (Perkin-Elmer, ABI) and an autosequencer model 373A (Perkin-Elmer, ABI). The 5′-untranslated region (UTR) and the 3′-UTR of mouse *Copg2* which were not represented in the clone AI115116 were supplemented from EST sequences. The 5′-end of *Bombyx copg1* was derived from the sequence of the clone ce–0169 which had recently been deposited in the SilkBase. Deduced amino acid sequences of *Copg* genes were aligned with ClustalW algorithm [24]. A phylogenetic tree was constructed using the amino acid sequences aligned by ClustalW algorithm with the neighbor-joining method of Saitou and Nei [25]. *S. cerevisiae* SEC21p was used as an outgroup. Support for each node was tested with the standard bootstrap analysis using 1000 replicates.

2.2. Southern and Northern hybridization analyses

15 μg of mouse genomic DNA from C57BL/6J was digested with *Bam*HI, *Stu*I, *Bgl*II, *Pvu*II, *Hin*dIII, *Eco*RI, *Xba*I or *Dra*I. Genomic DNA isolated from AB strain of zebrafish was digested with *Hin*dIII, *Bgl*II, *Eco*RI, *Pvu*II or *Sma*I. Digested genomic DNAs were electrophoresed on a 0.8% agarose gel and transferred to a Hybond-N^+ membrane (Amersham) for Southern hybridization. 30 μg of total RNAs from mouse kidney, liver, brain, testis, heart, lung, muscle,

*Corresponding author. Fax: (82)-42-869 2610.
E-mail: jhchung@mail.kaist.ac.kr

PII: S0014-5793(00)02033-0

colon, spleen and thymus was electrophoresed on a 1% agarose gel and transferred to a Hybond-N^+ membrane for Northern hybridization. The [α-^{32}P]dCTP random-primed mouse *Copg1* or zebrafish *copg2* cDNAs were used as the probes. Hybridization was carried out at 65°C using the QuickHyb solution (Stratagene). The membranes were exposed on X-ray films (Kodak).

2.3. Genomic PCR and allelic expression of zebrafish copg2 gene

A genomic PCR was performed with primers ZF1 (5′-AGGAAGAGACGTTTGCTC-3′) and ZR1 (5′-TGTCTGAACGCTCACAAG-3′). The PCR target was expected to contain an intron corresponding to the intron 22 of mouse *Copg2*. The genomic PCR products from AB and EK strains of zebrafish were sequenced to determine the exact intron position.

Nucleotide polymorphisms between AB and EK strains of zebrafish were sought to determine allelic expression of *copg2*. Genomic PCR products of the 3′-UTR from AB and EK strains were sequenced. PCR primers used were ZF2 (5′-TCTGTGGGCTAAACAAGCGATG-3′) and ZR2 (5′-TGCCATCAAATGCCAAAGAGG-3′). Three polymorphic sites, two single nucleotide polymorphisms (SNPs) and a 2 bp length polymorphism (LP) were found between two strains. One of the SNPs generated a restriction fragment LP when the PCR product was digested with restriction endonuclease *Bsm*AI. Total RNAs from the hybrid zebrafish larvae (AB×EK or EK×AB) were isolated using TRI REAGENT (Molecular Research Center) and reverse-transcribed with SuperScript II Reverse Transcriptase (Gibco BRL) using random hexamers. Reverse transcriptase (RT)-PCR products using primers ZF2 and ZR2 were digested with *Bsm*AI restriction enzyme to discriminate AB and EK alleles. Possible genomic DNA contamination was monitored with RT-negative samples.

2.4. Identification and analysis of partial cDNA sequences of plant copg genes

The sequence of *A. thaliana* γ-COP protein (database accession no. CAA18824) was used as a database query to identify plant γ-COP ESTs. ESTs derived from *Zea mays* (maize), *Oryza sativa* (rice) or *Glycine max* (soybean) showing homology with the C-terminus of *A. thaliana* γ-COP protein were retrieved and assembled using CAP program [26]. Two paralogous γ-COP partial cDNAs from *Z. mays* and single cDNAs from *O. sativa* and *G. max* could be generated. ESTs used for the generation of partial cDNA contigs were as follows: *Z. mays copg1*, AI586689, AI612452, AI622457, AI979521, AW053046, AW054123, AW060060, AW520015, AW520070, AW562936, AW566346 and BE344583; *Z. mays copg2*, AI795357, AW146649, AW506970 and AW519965; *O. sativa copg1*, AU030663, AU031418 and BE230009; *G. max copg*, AW186082, AW423375 and AW570363.

3. Results

3.1. Identification of cDNAs encoding γ-COP proteins from mouse, zebrafish, Drosophila and Bombyx

We identified full or nearly full cDNA sequences for *Copg* genes encoding γ-COP proteins of mouse, zebrafish, *Drosophila* and *Bombyx* by systematic search of dbEST and sequencing of representative EST clones. Mouse *Copg1* cDNA (GenBank accession no. AF187079) was 4 kb long and encoded 874 amino acids. Zebrafish *copg2* orthologous to human *COPG2* and mouse *Copg2* was 2.6 kb long and encoded 873 amino acids. Two zebrafish cDNA clones, one with full (GenBank accession no. AF191561) and the other with C-terminal and 3′-UTR (GenBank accession no. AF191562) were sequenced. An LP in 3′-UTR and SNPs including one that provoked amino acid substitution were found, although the two zebrafish clones were originated from the same cDNA library of Washington University Zebrafish EST Project (http://zfish.wustl.edu/). The full cDNA was 2.9 kb long and encoded 873 amino acids. *Drosophila copg* cDNA (GenBank accession no. AF191563) was 3 kb long and encoded 879 amino acids. Two *Bombyx* paralogous genes, *copg1* and *copg2*, were identified. *Bombyx copg1* (GenBank accession no. AB040669) was 4.3 kb long and encoded 861 amino acids, while *Bombyx copg2* cDNA sequence (GenBank accession no. AB040670) lacked a few residues at the N-terminus.

3.2. Sequence comparison and phylogenetic analysis of Copg genes

Deduced amino acid sequences of the *Copg* cDNAs were aligned with those of human *COPG1*, *COPG2* (GenBank accession no. AF157833) and mouse *Copg2* (GenBank accession no. AF205065) (Fig. 1A). γ-COP proteins were highly conserved between vertebrates and insects. Mouse *Copg1* shares 81% amino acid sequence identity with mouse *Copg2*, 97% with human *COPG1*, 60% with *Drosophila copg* and 52% with each of *Bombyx copg1* and *copg2*. Survey of *Drosophila* complete genome [27] found out that *Drosophila copg* gene (GenBank accession no. AE003778) consists of six exons and five introns. Two of the five intron positions of *Drosophila copg* were conserved in mammalian *Copg2* genes.

A phylogenetic tree based on the amino acid sequence alignment revealed that gene duplication event occurred in vertebrates is independent of that in *Bombyx* (Fig. 1B). Similarity search against dbEST found two paralogous genes in rat and chicken, suggesting the presence of two *Copg* genes in all the tested vertebrates. However, putative zebrafish *copg1* gene was not found in the dbEST, although the zebrafish *copg2* gene was represented by 20 ESTs.

3.3. Southern and Northern hybridization analyses of mouse Copg1 gene

Gene copy number of mouse *Copg1* gene was determined by Southern hybridization analysis. When the full cDNA was used as the probe, multiple bands were detected in all lanes (Fig. 2A), suggesting that closely related gene(s) or pseudogene(s) are present in the mouse genome. Cross-hybridization with *Copg2* gene is unlikely to explain the multiple bands since the open reading frame (ORF) of *Copg1* showed about 70% identity at the nucleotide sequence level with that of *Copg2*. Furthermore, no evidence of cross-hybridization was found in Northern hybridization analysis (see below). Alternative speculation is that the *Copg1* gene may contain many exons as does *Copg2*. Mouse *Copg2* and human *COPG2* were composed of 24 exons within the >40 kb genomic region [21]. If the exon–intron organization was established before the gene duplication, the *Copg1* gene might have as many

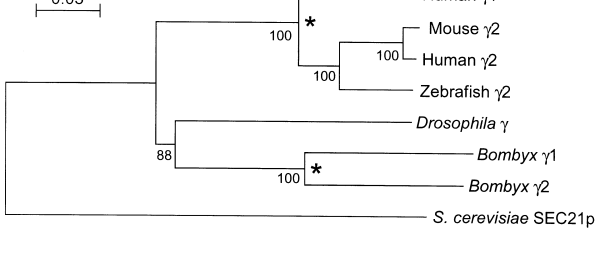
→

Fig. 1. Primary structure alignment of γ-COP proteins in this study with human homologs (A) and their phylogenetic relations (B). A: Multiple alignment of γ-COP proteins. Dots indicate amino acids identical to the top sequence, and dashes indicate gaps or lack of residues. Amino acid sequence identities with mouse *Copg1* were shown after the sequences. B: Phylogenetic relations of the γ-COP proteins. *S. cerevisiae* ortholog SEC21p was used as an outgroup. The gene duplication points are indicated by asterisks. Numbers below branches are bootstrap percentages on the basis of 1000 replicates. The scale bar indicates 0.05 substitutions per site. γ1 indicates γ-COP protein encoded by *COPG1*, *Copg1* or *copg1*; γ2 by *COPG2*, *Copg2* or *copg2*; and γ by *copg*.

A

```
Mouse γ1                M-LKKFDKKDEESGGGSNPLQHLEKSAVLQEAR-VFNETPINPRKCAHILTKILYLINQGEHLGTTEATEAFFAMTKLFQSNDPTLRRMCYLTIKEMSCI  98
Human γ1                .-.....F.....-.....                                                                                  98
Mouse γ2                .-I.....S...F.....-I.....R.L.....L....F..M.....R.....Q.....AT.                                     98
Human γ2                .-I.....S...F.....-I.....R.L.....L....F.....R.....Q.....AT.                                         98
Zebrafish γ2            .-I.....S...F.....-I.....R.L.....I..L....F.....R.....Q.....AN.                                      98
Bombyx γ1               .-KARR.G.E.D-----.VF.N.D.TTL.....-Y..S..VH....I.....L....E.T.Q...DI...T.....K.VV...LV..C...L.PM     94
Bombyx γ2               ---L.R.V...DDNS-...Y.N.D.TI....T.-E..Q.LVI....SL.....L....NFT.Q...D...T.....KEIM...V..C...L.KL     95
Drosophila γ            .GSFRRE.D...DA.P..AY.N...TS....T.-T.....V.....I.....Q.VAR...DC.....K.VV....V..G...L.S.             99
S. cerevisiae SEC21p    .-SAHTY..F.N----.TSGDLPD.MTIY.DCMNT...S.V.SKR.RLLISRL.R.LA...TFPQN...AL..SIS....HQNDP..QAV..A...L.G.  95

Mouse γ1                AEDVIIVTSSLTKDMTGKEDNYRGPAVRALCQITDSTMLQAVERYMKQAIVDKVPSVSSSALVSSLHLLKCSFDVVKRWVNEAQEAA----------SS 187
Human γ1                .....I.....-----------..                                                                            187
Mouse γ2                S.....V....I....R...G.....S..A.....MM.I.Y.....I.....----------..                                   187
Human γ2                S.....V....I....R...G.....I.....S.....MM.I.Y.....I.....----------..                                 187
Zebrafish γ2            S.....V....I....R...T.....I.....MV.M.....----------..                                                187
Bombyx γ1               .Q.....D.E..PA.I....S.....I.....N.A.G.A....A...SATVP.L.R..I.....M----------T.                       183
Bombyx γ2               .Q.....L..AA.I....S...AS....I.....N.A...A....A...SSTAP.L.R..A.....TI----------N.                    184
Drosophila γ            .....L..AA.I....S...N.....C....NAA..CA.....R.ANTAG.....A.....L----------N.                          188
S. cerevisiae SEC21p    S...LMA...IM..VQNGS.LIKPD.I.S.TYVL.ESTAFSA..LL.S.V.SRH...I..A..CT.Y...PI.EVTIR.FT..T...VLDLKQFPNQHGN. 195

Mouse γ1                -----DNIMVQYHALGLLYHVRKNDRLAVSKMISKFTRHG-LKSPFAYCMMIRVASKQLEE----EDGSRDSPLFDFIESCLRNKHEMVVYEAASAIVNL 277
Human γ1                -----.....N.....V.....-.....---.....                                                                 277
Mouse γ2                -----.....V...L.....LN...KS.-...Q.....L..I..RL.K.S---...-HE.....I.....IH.                          277
Human γ2                -----.....V...L.....LN...KS.-...Q.....L..I..RL.K.T---...-HE.....I.....IH.                          277
Zebrafish γ2            -----.....L.....T..LN...KS.-.....I...L...T---.G.-H.....HM                                           277
Bombyx γ1               -----.HV..S....AVVAGA.R....STV.L.T.LA.TP-VR..YTL.LQ..F.A.LA..D--PSE--ASE.YLE.....S.....H.....      273
Bombyx γ2               -----..A..S.....V.VLS.R..K.STV.LVTRLSKSH-I...YSL.LL..L.AQLCDGD--DSE--ASQ.YI....C...H.S...I....H.....  274
Drosophila γ            -----.....I..S.....LVN.L..GS-....Y.V..L..I.C.LI..EDIPSEELS.....T.....H.S...I...VH.....             282
S. cerevisiae SEC21p    EYYPNSTYIS.....QLK.T.KM.LL.LVRH.SENNSM.NQL.KVELVKIVNDLIYRD---PQ--LF.QFRPLLSDW.S..F.S.QL.T.KL.TSF 290

Mouse γ1                PGCSAKELAP-----AVSVLQLFCSSPKAALRYAAVRTLNKVAMKHPSAVTACNLDLENLVTDSNRSIATLAITTLLKTGSESSIDRLMKQISSFMSEIS 372
Human γ1                .....-----.....                                                                                      372
Mouse γ2                .N.T.R...-----.....P.....I.....V.....V....                                                          372
Human γ2                .N.T.R...-----.....P.....I.....V.....V....                                                          372
Zebrafish γ2            .N.T.R...-----.....I.....V.....V....                                                                372
Bombyx γ1               RKS-..D..Q-----.....I..G.S..T..L.GA...ARLTA...N..A..AV.....IS.....V....V....A..A.....E.....        367
Bombyx γ2               RKT-.RD..-----.....S....L.GA...ARLTTE..T..A..AI....IS.....V....V....A..A...V.....V...P             368
Drosophila γ            KNTNPRM.S.....F.I.....T..F.....T..A...T.....G.I.....V.....A...VE.....T.VA..                        377
S. cerevisiae SEC21p    ATRNSRLV..ELYAA.I.A..SLLTV.RVSS.F..L.I..RIS.VS.EKIVV..PE..S.INN...N.S.Y.....TSKN.SS.IST.TN.IHDV.  390

Mouse γ1                DEFKVVVVQAISALCQKYPRKHAVLMNFLFTMLRE-EGGFEYKRAIVDCIISIIEENSESKETGLSHLCEFIEDCEFTVLATRILHLLGQEGPKTNNPSK 471
Human γ1                .....-.....T.....                                                                                    471
Mouse γ2                .....H.....S.M.T..SN..D-D.....K.....V...P....A..A.....H.....K.....K...R.PV...                      471
Human γ2                .....S.M.T..SN..D-D.....V...P....A..A.....H.....K.....K...R.PV...                                   471
Zebrafish γ2            .....S.M....SN..D-D.....P.....A.....H.....K.....K...R.PT...                                         471
Bombyx γ1               ....I...R..RR..S.....QS.AA..AG...D-...LQ..T..AEAL.AL....PDA.....A.....HVT..V....V..R....ARQ..R      466
Bombyx γ2               ....II..R..RQ..S.F....QS.AA..AG...D-...L...A..A.A..ALV...PDA.....A.....H.A..V.....R....SRH..R      467
Drosophila γ            .....C..T.....T.....SG....-...L...TS...T..T.....ADA..S.....HVS..V.....K...FAAT...                   476
S. cerevisiae SEC21p    .D..IIIID.VRT.SLNF.QEWKSIL...IDV.KNS....KF.NS..EAL.D.VSFVPQ...LA.EN..D.....NEILV.....K...SAP...L  490

Mouse γ1                YIRFIYNRVVLEHEEVRAGAVSALAKFGAQN--EEMLPSILVLLKRCVMDDDNEVRDRATFYLNVLEQKQKA---LNAGYILNGLTVSIPGLEKALQQYT 566
Human γ1                .....--.....---.....R.....                                                                           566
Mouse γ2                .....F.....N.A...A.....--.SL.....Q..M..T.D.....Q.R.M.---...T..F.....M....H...                      566
Human γ2                .....F.....N.A...A.....--.SL.....Q..M..T.D.....Q.R.M.---...T..F.....V..M....H...                    566
Zebrafish γ2            .....F.....S.A...A.....--DDL...V...MQ..M..S.D.....M...Q.....---...A..F...S..VL....S.H...           566
Bombyx γ1               ...Y.....I..SGP...A....V.R...SR--P.L..N.R...S..QL.E.D.....V..SAI.DSGNPQ---.IND..I.IQVPNPVL...S.SD.L 561
Bombyx γ2               .....I..SGP...A....V.Q....V--P.L..N.K...A..QL.EED.....VY.STI..TENPQ---.INDF.V.IPKPNVVL....RDHL    562
Drosophila γ            .....I..SPI...A..T.M.Q...SC--PAL.SN.....G..Q..P.D.....Y..SI.NSERPE---.YKN..IERENC.LAL...S.VEHL    571
S. cerevisiae SEC21p    .V.H.....NSII.SA..V..S..ALTKNDPTLYE..IS....IAN.K.D.....IA.EFIDSARNKDDVIAQNL.ESKYFYD..S..SK.SS.I  590

Mouse γ1                LE--PS-EKPFDLKSVPLAT-----TPMAEQRPE-----------STATAAVKQPEKVAAT----------RQEI----FQEQLAAVPEFQGLGPLFKS 632
Human γ1                ..--..-.....-----A.....T.----------.PIT.....----------...----.....R.....                          632
Mouse γ2                ..--..-.....M..I...M-----A.VF..KS.----------I.LVTPK----...L.PS----------..D.----.....I...MN.....  630
Human γ2                ..--..-.....M..I...M-----A.VF..KA.----------I.LV.TK--...L.PS----------..D.----.....I...LNI.....  630
Zebrafish γ2            ..--..-.....M.T.....-----A.IT..HKT.----------IAPV.TS.L...L.PS----------..D.----Y....S.I.....      632
Bombyx γ1               AS--GDQSE..NIAA..T.E-----E.QEPKE---------------SPVEIETRKAPVVS----------.E.Q----YT...Q.I.GIEK.....  624
Bombyx γ2               .N--.E--E..NIL...AEE-----VVKETKK---------------PI.EIEVRKP.QMS----------IE..----YA....KI.GIEK.....T 623
Drosophila γ            N---GDVDTR..ISI..K.A-----IVKPVIAN----------DVMLVTSSA.RPPKI.----------.E.E----SAAR..QL.GI.V...IHR. 636
S. cerevisiae SEC21p    SSNTD.FATA..VNQ.RKF.EDEMKAINLKRKQ.QIFNQKSETTLD.TPE.ESV...R.DANSFAGPNLDDHQEDLLATKYADE.LSIEQIKPF.Q.VN. 690

Mouse γ1                SPEPVALTESETEYVIRCTKHTFSDHLVFQFDCTNTLNDQTLENVTVQMEPTEA----YEVLSYVPARSLPYNQPGTCYTLVALPTEDPTAVACTFSCVM 728
Human γ1                .....TN.M.....----....C.....K.....M.                                                                728
Mouse γ2                .-...Q...A....FV..V..M.T..I.....L..K.....SDS----....CCI..P.....I.....R..D.....G....T.              725
Human γ2                .-...Q...A....FV..I..M.TN.I.....L..K.....SDS----....CI..P.....I.....R..DD.....GS...T.              725
Zebrafish γ2            .-...Q...A.....V..I....AN.MI.....L.QK.L.....S.S----....H....AN...S...S..S..R..ED.....S.....T.      727
Bombyx γ1               C-.AID...P....RV..V..I.AR..IL..E.L...S..L..K.Y.RLDSIPG----.KISNV..CDE...DKQ.NIFC.LEF.-.N.IETLGN.GATL 718
Bombyx γ2               N-L.ID...A.I..RV.VV..VYAR.VI...E.V.....I..D.H.RL.VPPE----..IK.I..CQK.V..TQDNVFVV.EY.-CAFIDSLGS.MATL 717
Drosophila γ            T-A.IQ.....TVQ.I..I.GQ.V.....L...S..I....R.ELTLP.G----FTTRAVI.CPK....DLQ.TFVI.EF.-P.AANSIA..GATL 730
S. cerevisiae SEC21p    .-RAIS...P.A.F.V.GV..L.K.NV.L..NI....T.IA.D..S.VCT.EISDEAEL.E.FTLQVDR.LPSEEAA..VAFKKL--.EIVMEG-.LNNL 786

Mouse γ1                KFTVKDCDPNTGEI-D-EEGYEDEYVLEDLEVTVADHIQKVMK-VNFEAAWDEVGD--EFEKEETFTLSTIKTLEEAVGNIVKFLGMHPCERSDKVPENK 823
Human γ1                .....T..T-.-D.....-L.....--.....D..                                                                 823
Mouse γ2                ....R.....VP-.-.D..D.....S.....IL.-P..A...E....--A.....A..ST.....N..IT....Q.....                    820
Human γ2                ....R.....VP-.-.D..D.....S.....L.-P..A...E....--T.....A..ST.....N..IT....Q.....                     820
Zebrafish γ2            .YL.R.....P-.-DD..D.....L.-P..A.....--.C.....A.A.VR..D...N...S...Q.....                              822
Bombyx γ1               E.V.R....T..LP-.SN...A.T.P..EI.I.C..QFRTRAATDDW.Q..ERAAN--AA.ACD..G.TQS-DVGA.ASAVCEH...PKA-----AISGE 809
Bombyx γ2               E.V.R.....TP-.PG...A.S.P..EF.MGCS.Q.RARAGGDDW.QT.ERAAN--VP.AND..V..QT-DVGD.AKAVCEH..LPNG-----SISGD 808
Drosophila γ            R.V.....P-ES....D...M.....L...Q...TR.-N..QVS..AADSEEWLQA.D..V..AVT..QD..NT...I..LGAANL.EN...GT    828
S. cerevisiae SEC21p    T..T.EIN.D.N.PF.GD..FQ...EIDSIFLNAG.YVKSSFT-G..S.TF..LPC----.EVAV.NIQEDLSIQ.V.DK.ILNSSCL.V.STQFA.SDS 881

Mouse γ1                NTHTLLLAG-VFRGGHDILVRSRLLLLDT-VTMQVTARSSEELPVDIILASVG- 874     100%
Human γ1                .....-.....-.....L.....- 874     97%
Mouse γ2                .S.S.Y...-.Y..Y.L.....A.A.G-.....V..K.RT...V.....- 871     81%
Human γ2                .S.S.Y...-I....Y.L.....A.A.G-.....V..K.RT...V.....- 871     81%
Zebrafish γ2            .S.V.F...-.....V...A..A.A.G-.....V..TDDNV..V.....- 873     82%
Bombyx γ1               AVKEIRGG.-LW.A.APV.M.A..VAAQGT....LIT..PR.DVATLL..A..- 861     52%
Bombyx γ2               AVKEIRGG.-I....APV...A..AGSSAG...KLS...PR.DVAQLL..A..- 860     52%
Drosophila γ            HL....CS.-T....AE....AK.A.SEG-..LNL.V..TDQDVAEL.T.AI.- 879     60%
S. cerevisiae SEC21p    .S...K.F.KSALT.SKVALQIKMIKSSKGLALK.HGKGEDS.LCSDLVNGLMQ 935     32%
```

B



exons. When the membrane was reprobed with 3′-UTR of *Copg1* cDNA, only a single band was clearly detected in each lane (data not shown).

Tissue distribution of *Copg1* transcript was examined in mouse by Northern hybridization. *Copg1* was ubiquitously expressed in all the examined tissues. Its level was most abundant in testis (Fig. 2B). The overall expression profile of *Copg1* was similar to that of *Copg2* [21]. However, it turned out that there were three isoforms of *Copg1* transcripts with sizes of approximately 1.5, 3.0 and 6.0 kb, whereas *Copg2* gave rise to a single major transcript of 3.0 kb [21]. The longest transcript of *Copg1* was detected in all tissues, suggesting that it is the major functional form. Two other smaller isoforms were differentially produced in the mouse tissues. Since the sizes of *Copg1* and *Copg2* ORFs were nearly identical, size difference between major forms of *Copg1* and *Copg2* is due to different length of the UTRs. The smallest isoform of *Copg1* is shorter than the ORF and hence cannot encode a full-length protein.

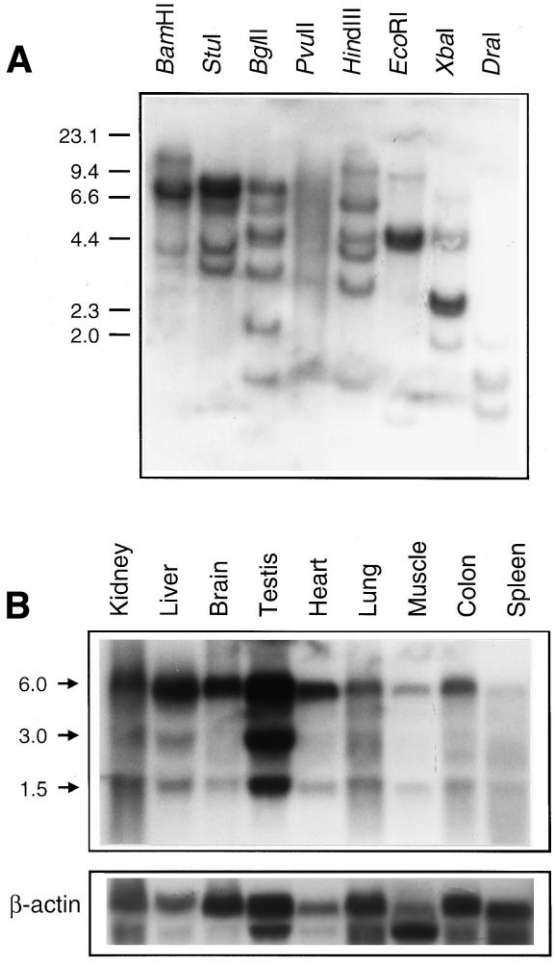


Fig. 2. Southern (A) and Northern (B) hybridization analyses of mouse *Copg1* gene. A: Southern hybridization analysis of *Copg1*. The size markers are indicated in kb. B: Expression analysis of *Copg1* in various tissues of mouse. The three forms of transcripts are marked by arrows and their estimated sizes are indicated. Reprobing of the same Southern membrane with the β-actin was shown for the loading control.

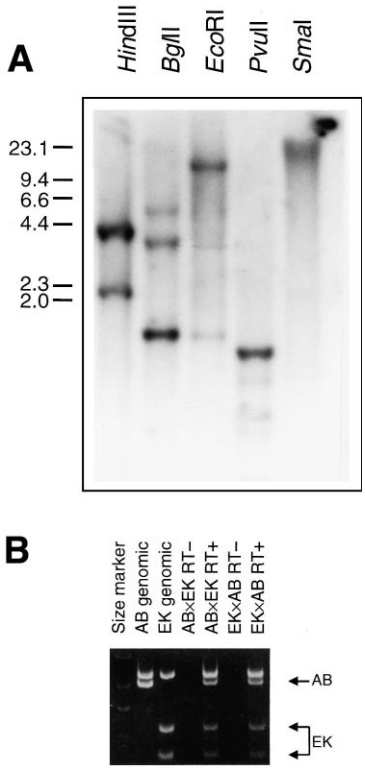


Fig. 3. Southern hybridization (A) and allelic expression (B) analyses of zebrafish *copg2* gene. A: Southern hybridization analysis of zebrafish *copg2*. The size markers are indicated in kb. B: Allelic expression of zebrafish *copg2* in hybrid larvae. Genomic or RT-PCR products were digested with *Bsm*AI. AB- and EK-specific bands were indicated. *Bsm*AI cuts the AB allele once and the EK allele twice.

3.4. Southern hybridization, genomic PCR and allelic expression analyses of zebrafish copg2 gene

Southern hybridization analysis of zebrafish *copg2* suggested the multi-exonic structure of *copg2* as the case of the mouse *Copg2* (Fig. 3A). The reduced number of bands, when compared with the Southern bands of mouse *Copg1*, predicts that the genomic size of zebrafish *copg2* is smaller than those of mouse *Copg1* and *Copg2*. Zebrafish genomic DNA digested with methyl-sensitive *Sma*I yielded high molecular weight bands. It reconciles with an earlier observation that zebrafish genome is heavily methylated in CpG sites [28].

A genomic PCR using a primer pair ZF1 and ZR1 yielded longer products than that from RT-PCR (data not shown). Sequence analysis of the genomic PCR product (GenBank accession no. AB042814) revealed that an intron corresponding to the intron 22 of mouse *Copg2* was located at the same position of mouse *Copg2*. The zebrafish intron was 112 bp long and started with GC instead of GT as the 5′-splice donor. Although the only one intron position was studied, this result supported that the genomic organization of zebrafish *copg2* was basically well-conserved with mouse *Copg2*.

Human *COPG2* and mouse *Copg2* genes were reported to be imprinted [20,21]. We tested allelic expression of zebrafish *copg2* gene in hybrid larvae. The 3′-UTR was amplified from AB and EK strains of zebrafish using the primers ZF2 and ZR2. Sequence analysis of the 3′-UTR revealed two SNPs and a 2 bp LP between AB and EK strains. Genomic sequence and the detailed information on polymorphisms were

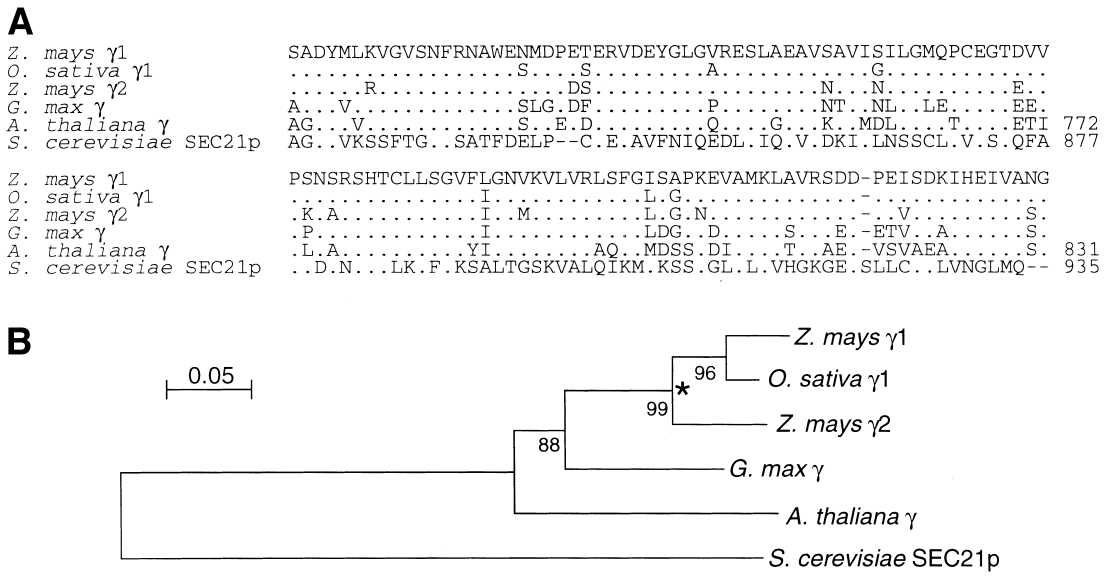


Fig. 4. Alignment of partial plant γ-COP proteins in this study with *A. thaliana* γ-COP and *S. cerevisiae* ortholog SEC21p (A) and their phylogenetic relations (B). All manipulations are the same as Fig. 1.

deposited in the GenBank database under accession no. AB042116. One of the SNPs gave rise to the second *Bsm*AI site in EK allele, providing an effective physical marker for distinguishing allelic expression. When RT-PCR products prepared from reciprocal hybrid (AB×EK and EK×AB) larvae were digested with *Bsm*AI, both AB- and EK-specific bands were detected approximately in equal amounts, indicating that zebrafish *copg2* was biallelically expressed (Fig. 3B).

3.5. Partial cDNA sequences and phylogenetic analysis of plant copg genes

Four partial cDNA sequences encompassing the C-terminal region of γ-COP from three plant species *Z. mays*, *O. sativa* and *G. max* were assembled from at least three EST sequences. From *Z. mays*, 16 EST sequences were assembled into two distinct cDNAs, 12 for *copg1* and four for *copg2*. No evidence for sequencing error or polymorphism was found. The deduced C-terminal 119 amino acid sequences from each of the four partial plant cDNA sequences were further analyzed. Two *Z. mays copg* genes shared 87% identity in amino acid sequence and 78% in nucleotide sequence, and showed no similarity in the 3′-UTR, indicating that two cDNAs were originated from paralogous genes. Multiple sequence alignment of the deduced C-terminal 119 amino acids with *A. thaliana* γ-COP and *S. cerevisiae* ortholog SEC21p was performed using ClustalW program (Fig. 4A). *Z. mays copg1* was more similar to *O. sativa copg1* than to *Z. mays copg2*. A phylogenetic analysis revealed that the duplication event occurred before the divergence of *Z. mays* and *O. sativa* (Fig. 4B).

4. Discussion

We identified several *Copg* genes encoding non-clathrin coat protein γ-COP by systematic analyses of dbEST. Duplicated gene pairs were found in human, mouse, rat, chicken, *Bombyx* and *Z. mays*. The phylogenetic analysis suggested that the independent gene duplication events occurred in an ancestral vertebrate, after the divergence of the *Bombyx* and dipteran species, and before the divergence of *Z. mays* and *O. sativa* (Figs. 1B and 4B).

The duplication of ancestral *Copg* gene long before the divergence of fishes and tetrapods suggests that zebrafish possibly possesses the *copg1* gene yet to be identified (Fig. 1B). No representation of zebrafish *copg1* in dbEST seemed to be due to scarcity of ESTs sequenced and limited sets of tissues examined. Otherwise, the zebrafish *copg1* gene might be lost during evolution. Recent studies on linkage maps for zebrafish and *hox* clusters of zebrafish and *Fugu rubripes* suggested that the teleost (bony fishes) chromosome was doubled by whole genome duplication after the divergence of the teleosts and tetrapod lineage, and that some chromosomal segments or genes were subsequently lost [29–32]. If the chromosomal segments containing *copg* genes were retained, it is possible that zebrafish has four copies of *copg* genes, two *copg1* genes and two *copg2* genes. Southern hybridization analysis of zebrafish *copg2* (Fig. 3A) seemed to support the existence of two copies of *copg2* gene. Alternatively and more plausibly, multiple bands in Southern hybridization could be explained by the multi-exonic structure of zebrafish *copg2* gene. An intron position corresponding to the intron 22 of mouse *Copg2* was exactly the same with that of mouse *Copg2*, implying the genomic organization of zebrafish *copg2* was conserved with mouse *Copg2*. Furthermore, the zebrafish *peg1* (an ortholog of mouse *Peg1/Mest*) which was found to be adjacent to *copg2* as in mammals was present as a single copy in the zebrafish genome and showed a conserved genomic organization with mouse *Peg1/Mest* (Y. Hahn, unpublished data). These results strongly indicate that one of the two paralogous chromosomal segments containing *peg1* and *copg2* loci arisen by chromosomal doubling in an ancient teleost was lost during zebrafish genome evolution.

In plants, *copg* gene duplication seemed to have occurred in the grass family (Poaceae) (Fig. 4B). The phylogenetic analysis of plant *copg* genes indicates that *Z. mays copg1* and *copg2* have diverged before the divergence of *O. sativa* and *Z. mays*, suggesting the possible presence of *copg2* gene in *O. sativa*. Segmental allotetraploid origin of *Z. mays* is unlikely to be

responsible for the duplication of *copg* genes since the allotetraploid event was estimated to have occurred after the divergence of *Z. mays* and *O. sativa* [33]. Gene duplication may have occurred in an ancient monocot flowering plant. More *copg* genes should be identified in plants to test this hypothesis, especially in monocot flowering plant clade. Two wheat ESTs, BE404005 and BE442852, recently deposited in the GenBank, showed high similarity with *Z. mays copg1* and *copg2*, respectively, strongly supporting this hypothesis.

The zebrafish *copg2* gene of which mammalian ortholog was expressed in a parent-of-origin-specific manner was biallelically expressed in the hybrid larvae (Fig. 3B). Recently, the allelic expression of imprinted genes was studied in non-eutherian species. *IGF2*, an imprinted gene which is expressed exclusively from the paternal chromosome in eutherians [34], was shown to be expressed in a paternal-specific manner in a marsupial, *Monodelphis domestica*, but biallelically expressed in chickens [35]. *M6P/IGF2R* was imprinted in a marsupial, the opossum, but not in monotremes [36]. These results together with our data confirm that genomic imprinting has evolved in the mammalian clade after the monotremes were branched out. However, the genomic imprinting phenomenon is not restricted to mammals. For the proper embryo and endosperm development in *Arabidopsis*, only the maternal *MEDEA* gene is required and hence the *MEDEA* is imprinted [37,38]. Parental effect on gene expression was also found in *Drosophila* [39–41]. In zebrafish, androgenetic embryos developed normally, indicating that none of the imprinted genes is involved in development [42]. However, delayed development and low survival rate to maturity of androgenetic or gynogenetic zebrafish suggest that genomic imprinting may be grossly dispensable but required for the proper embryo development and maturation [43]. Parental effects on embryo development may have independently emerged and selected in many taxa and hence the genes subjected to genomic imprinting would be virtually different from species to species. Brief study on allelic expression of zebrafish *peg1* and *igf2* genes showed biallelic expression as *copg2* in the zebrafish larvae (Y. Hahn, unpublished data).

Subunits of adaptor protein complexes 1, 2 and 3 and heterotetrameric subcomplex composed of β-, γ-, δ- and ζ-COP of the COPI were considered to share common ancestral genes. They were supposed to have emerged by stepwise gene duplication and functional divergence [19]. We demonstrated in this study that the duplication of γ-COP is a progressively ongoing process. Though we have no evidence at this moment whether mouse *Copg1* and *Copg2* are functionally equivalent or distinct, the similarity in transcript distribution of two genes in mouse tissues suggests that two proteins are functionally redundant to some extent. Preliminary study on *Copz* genes encoding ζ-COP revealed that two *Copz* genes are present in vertebrates and in plants, further supporting the concurrent duplication and divergence of components of the vesicle transport system.

References

[1] Schmid, S.L. (1997) Annu. Rev. Biochem. 66, 511–548.
[2] Schekman, R. and Orci, L. (1996) Science 271, 1526–1533.
[3] Allan, B.B. and Balch, W.E. (1999) Science 285, 63–66.
[4] Gaynor, E.C., Graham, T.R. and Emr, S.D. (1998) Biochim. Biophys. Acta 1404, 33–51.
[5] Lowe, M. and Kreis, T.E. (1998) Biochim. Biophys. Acta 1404, 53–66.
[6] Barlowe, C. (1998) Biochim. Biophys. Acta 1404, 67–76.
[7] Seaman, M.N., McCaffery, J.M. and Emr, S.D. (1998) J. Cell Biol. 142, 665–681.
[8] Anderson, R.G. (1998) Annu. Rev. Biochem. 67, 199–225.
[9] Dell'Angelica, E.C., Ohno, H., Ooi, C.E., Rabinovich, E., Roche, K.W. and Bonifacino, J.S. (1997) EMBO J. 16, 917–928.
[10] Rothman, J.E. and Wieland, F.T. (1996) Science 272, 227–234.
[11] Serafini, T., Orci, L., Amherdt, M., Brunner, M., Kahn, R.A. and Rothman, J.E. (1991) Cell 67, 239–253.
[12] Waters, M.G., Serafini, T. and Rothman, J.E. (1991) Nature 349, 248–251.
[13] Stenbeck, G., Harter, C., Brecht, A., Herrmann, D., Lottspeich, F., Orci, L. and Wieland, F.T. (1993) EMBO J. 12, 2841–2845.
[14] Harter, C. (1995) FEBS Lett. 369, 89–92.
[15] Cosson, P. and Letourneur, F. (1994) Science 263, 1629–1631.
[16] Letourneur, F., Gaynor, E.C., Hennecke, S., Demolliere, C., Duden, R., Emr, S.D., Riezman, H. and Cosson, P. (1994) Cell 79, 1199–1207.
[17] Harter, C. and Wieland, F.T. (1998) Proc. Natl. Acad. Sci. USA 95, 11649–11654.
[18] Harter, C., Pavel, J., Coccia, F., Draken, E., Wegehingel, S., Tschochner, H. and Wieland, F. (1996) Proc. Natl. Acad. Sci. USA 93, 1902–1906.
[19] Schledzewski, K., Brinkmann, H. and Mendel, R.R. (1999) J. Mol. Evol. 48, 770–778.
[20] Blagitko, N., Schulz, U., Schinzel, A.A., Ropers, H.H. and Kalscheuer, V.M. (1999) Hum. Mol. Genet. 8, 2387–2396.
[21] Lee, Y.J., Park, C.W., Hahn, Y., Park, J., Lee, J., Yun, J.H., Hyun, B. and Chung, J.H. (2000) FEBS Lett. 472, 230–234.
[22] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[23] Mita, K., Morimyo, M., Okano, K., Shimada, T. and Maeda, S. (1999) RIKEN Rev. 22, 63–67.
[24] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[25] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[26] Huang, X. (1992) Genomics 14, 18–25.
[27] Adams, M.D. et al. (2000) Science 287, 2185–2195.
[28] Macleod, D., Clark, V.H. and Bird, A. (1999) Nat. Genet. 23, 139–140.
[29] Amores, A. et al. (1998) Science 282, 1711–1714.
[30] Aparicio, S. et al. (1997) Nat. Genet. 16, 79–83.
[31] Aparicio, S. (2000) Trends Genet. 16, 54–56.
[32] Postlethwait, J.H. et al. (1998) Nat. Genet. 18, 345–349.
[33] Gaut, B.S. and Doebley, J.F. (1997) Proc. Natl. Acad. Sci. USA 94, 6809–6814.
[34] DeChiara, T.M., Robertson, E.J. and Efstratiadis, A. (1991) Cell 64, 849–859.
[35] O'Neill, M.J., Ingram, R.S., Vrana, P.B. and Tilghman, S.M. (2000) Dev. Genes Evol. 210, 18–20.
[36] Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., MacDonald, R.G. and Jirtle, R.L. (2000) Mol. Cell 5, 707–716.
[37] Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A. and Grossniklaus, U. (1999) Genes Dev. 13, 2971–2982.
[38] Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (1999) Plant Cell 11, 1945–1952.
[39] Lloyd, V.K., Sinclair, D.A. and Grigliatti, T.A. (1999) Genetics 151, 1503–1516.
[40] Golic, K.G., Golic, M.M. and Pimpinelli, S. (1998) Curr. Biol. 8, 1273–1276.
[41] Fitch, K.R., Yasuda, G.K., Owens, K.N. and Wakimoto, B.T. (1998) Curr. Top. Dev. Biol. 38, 1–34.
[42] Corley-Smith, G.E., Lim, C.J. and Brandhorst, B.P. (1996) Genetics 142, 1265–1276.
[43] McGowan, R.A. and Martin, C.C. (1997) Biochem. Cell Biol. 75, 499–506.