

Research article

Open Access

Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipoprotein II/I and vertebrate apolipoprotein B

Jean-Christophe Avarre^{1,2}, Esther Lubzens¹ and Patrick J Babin^{*3}

Address: ¹Israel Oceanographic and Limnological Research, P.O. Box 8030, Haifa 31080, Israel, ²Genewave XTEC, Ecole Polytechnique, 91128 Palaiseau, France and ³Génomique et Physiologie des Poissons, Université Bordeaux 1, UMR NuAGe, 33405 Talence cedex, France

Email: Jean-Christophe Avarre - jean-christophe.avarre@genewave.com; Esther Lubzens - esther@ocean.org.il; Patrick J Babin* - p.babin@gpp.u-bordeaux1.fr

* Corresponding author

Published: 22 January 2007

Received: 24 August 2006

BMC Evolutionary Biology 2007, **7**:3 doi:10.1186/1471-2148-7-3

Accepted: 22 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/3>

© 2007 Avarre et al; licensee BioMed Central Ltd.

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

Abstract

Background: In animals, the biogenesis of some lipoprotein classes requires members of the ancient large lipid transfer protein (LLTP) superfamily, including the cytosolic large subunit of microsomal triglyceride transfer protein (MTP), vertebrate apolipoprotein B (apoB), vitellogenin (Vtg), and insect apolipoprotein II/I precursor (apoLp-II/I). In most oviparous species, Vtg, a large glycolipoprotein, is the main egg yolk precursor protein.

Results: This report clarifies the phylogenetic relationships of LLTP superfamily members and classifies them into three families and their related subfamilies. This means that the generic term Vtg is no longer a functional term, but is rather based on phylogenetic/structural criteria. In addition, we determined that the main egg yolk precursor protein of decapod crustaceans show an overall greater sequence similarity with apoLp-II/I than other LLTP, including Vtgs. This close association is supported by the phylogenetic analysis, i.e. neighbor-joining, maximum likelihood and Bayesian inference methods, of conserved sequence motifs and the presence of three common conserved domains: an N-terminal large lipid transfer module marker for LLTP, a DUF1081 domain of unknown function in their central region exclusively shared with apoLp-II/I and apoB, and a von Willebrand-factor type D domain at their C-terminal end. Additionally, they share a conserved functional subtilisin-like endoprotease cleavage site with apoLp-II/I, in a similar location.

Conclusion: The structural and phylogenetic data presented indicate that the major egg yolk precursor protein of decapod crustaceans is surprisingly closely related to insect apoLp-II/I and vertebrate apoB and should be known as apolipocrustacein (apoCr) rather than Vtg. These LLTP may arise from an ancient duplication event leading to paralogs of Vtg sequences. The presence of LLTP homologs in one genome may facilitate redundancy, e.g. involvement in lipid metabolism and as egg yolk precursor protein, and neofunctionalization and subfunctionalization, e.g. involvement in clotting cascade and immune response, of extracellular LLTP members. These protein-coding nuclear genes may be used to resolve phylogenetic relationships among the major arthropod groups, especially the Pancrustacea-major splits.

Background

In 1967, Wallace *et al.* [1] characterized a high-density lipoprotein from decapod crustaceans ovaries with similar biochemical properties to lipoproteins isolated from vertebrate eggs, and proposed the generic term "lipovitellin" for this abundant lipoprotein. Two years later, Kerr [2] identified a blood-borne protein present only in female blue crabs *Callinectes sapidus* with developing oocytes. This lipoprotein turned out to be serologically identical to oocyte lipovitellin. The term "vitellogenin" (Vtg) was proposed over thirty-five years ago [3] to describe female-specific insect hemolymph protein precursors of egg yolk, regardless of their amino acid sequences or structures. This term, based on a functional criterion, was later adopted in other egg-laying animals, including crustaceans [4], and is widely used in the scientific community and sequence databases. Molecular characterization of Vtg in numerous oviparous species has revealed that this high molecular weight glycolipoprotein is conserved among species, suggesting derivation from a common ancestor [5-8]. However, molecular data obtained in some species has revealed that the main egg yolk precursor proteins are unrelated to the Vtg protein family. For example, major egg yolk precursor protein is related to transferrin in sea urchins [9,10] and lipase in higher *Diptera* [11]. Multiple alignments of vertebrate and non-vertebrate Vtg sequences revealed five relatively well-conserved regions [6,12]. Regions I to III, located in the N-terminal part, correspond to the lipovitellin 1 subunit of vertebrate Vtg, while regions IV and V, located in the C-terminal part, correspond to the lipovitellin 2 subunit. Sequence and deduced structural homologies indicated an evolutionary relationship of Vtg with three mammalian proteins, apolipoprotein B100 (apoB), the large subunit of microsomal triglyceride transfer protein (MTP), and the von Willebrand factor [13,14]. The identification of conserved amino acid sequence motifs and ancestral exon boundaries in apoB, MTP, non-vertebrate and vertebrate Vtg, and insect apolipoprotein II/I (apoLp-II/I) indicated that large lipid transfer proteins (LLTP) are members of the same multigene superfamily and have emerged from a common ancestral molecule designed to play a pivotal role in the intracellular and extracellular transfer of lipids and liposoluble substances [15,16].

Knowledge of molecular structure and expression of Vtg in oviparous animals has increased impressively over the past two decades [6,17]. Recent molecular characterization and expression studies of the main egg yolk precursor protein, referred to as Vtg, in over ten decapod crustacean species suggests that this precursor protein is atypical in regard to Vtg from other oviparous animals [18-29]. In addition, it has been shown that the crustacean clotting protein (CP), a very high density lipoprotein (VHDL) responsible for hemolymph clot formation, is also a Vtg-

related protein [30,31]. The aim of this study was therefore to clarify the phylogenetic relationship of these crustacean Vtg-related proteins with other LLTP superfamily members. The results presented here led us to call apolipocrustacein (apoCr) rather than Vtg the major decapod crustacean egg yolk precursor protein.

Results and discussion

Crustacean apoCr sequences

Full-length apoCr cDNA sequences, annotated as Vtg in the GenBank™/EBI Data Bank (see the Methods section), are currently available for eleven decapod species: *Penaeus semisulcatus*, *Penaeus monodon*, *Metapenaeus ensis*, *Marsupenaeus japonicus*, *Litopenaeus vannamei*, *Feneropenaeus merguensis*, *Cherax quadricarinatus*, *Machrobrachium rosenbergii*, *Charybdis feriatus*, *Pandalus hypsinotus*. and *Portunus trituberculatus*. In *Penaeus semisulcatus* and *Marsupenaeus japonicus*, the same apoCr cDNA was isolated from ovary and hepatopancreas tissues [23,32]. In *Metapenaeus ensis*, two apoCr cDNAs, apoCr1 and apoCr2, were isolated from these two tissues [24,25]. They shared 56% sequence identity and showed a tissue-specific expression pattern. These two apoCr may result from a gene duplication event in the *Metapenaeus* lineage. A recent phylogeny of penaeid shrimps, performed on two mitochondrial genes, indicated that *Metapenaeus* may be representative of the ancient *Penaeus* genus but distant from the other penaeid species [33]. Alignment of the deduced amino acid sequence of *Penaeus semisulcatus* revealed a sequence identity ranging from 92% with *Feneropenaeus merguensis* to 34% with *Charybdis feriatus* [see Additional file 1]. The phylogenetic analysis resulting from this alignment (data not shown) indicated that these deduced precursor proteins may be confidently grouped according to the species tree; i.e. the penaeid shrimps of *Dendrobranchiata* suborder were grouped together and the *Pleocyemata* suborder species, crab (*Charybdis feriatus*), crayfish (*Cherax quadricarinatus*) and prawn (*Machrobrachium rosenbergii*, and *Pandalus hypsinotus*) were outside this monophyletic group. It is interesting to note that phylogenetic results correlated with the expression pattern of apoCr transcripts. *Dendrobranchiata* suborder species express apoCr transcripts in both ovaries and hepatopancreas [18,19,23,24,28,29], while the expression of paralogous apoCr2 from *Metapenaeus ensis* [24,25] and apoCr of *Pleocyemata* suborder species is apparently restricted to the hepatopancreas [20,21,26,27].

Global sequence alignment of crustacean apoCr with other LLTP

A BLASTP [34] search of the nonredundant GenBank™ database using *Penaeus semisulcatus* apoCr as a target sequence revealed high BLAST scores and expected (E) values with other decapod apoCr, from the full-length sequence of *Feneropenaeus merguensis* (score 4530, E value

0.0) to *Charybdis feriatus* (score 1134, E value 0.0). There were also similarities with other extracellular LLTP: insect apoLp-II/I, from *Locusta migratoria* (score 296, E value 7e-78) to *Anopheles gambiae* (score 173, E value 9e-41), and vertebrate apoB, from *Gallus gallus* (score 115, E value 3e-23) to *Danio rerio* (score 105, E value 4e-20), including *Homo sapiens* apoB-100 precursor protein (score 105, E value 2e-20). Lower scores and E values were retrieved with Vtg of oviparous animals, ranging from *Crassostrea gigas* mollusks (score 100, E value 1e-18) to *Samia cynthia ricini* insects (score 40, E value 1.2). BLAST scores and E values of intracellular MTP family members overlapped with Vtg family members, as shown with *Strongylocentrotus purpuratus* echinoderms (score 48.9, E value 0.003) or *Homo sapiens* (score 38.9, E value 3.4). Even if BLAST analysis represents an over simplification of reality, the results obtained by this method, i.e. so-called crustacean Vtg are more closely-related to insect apoLp-II/I than to insect Vtg, as indicated by BLAST scores, was not seriously taken into account, as the term Vtg was already used for these sequences.

Crustacean apoCr (~2,600 amino acids) is of intermediate length between insect, nematode, mollusk, and vertebrate Vtg, which are generally shorter (~1,600–1,800 amino acids), and insect apoLp-II/I, which are larger (~3,300 amino acids). Consistent with BLAST scores, crustacean apoCrs confidently align with insect apoLp-II/Is [see Additional file 1]. ApoLp-II/I is the high molecular weight apolipoprotein of lipoprotein found at high concentrations in insect hemolymph [35–37]. The best local alignments between these sequences occurred along the first 1,000 amino acid residues and in their C-terminal parts [see Additional file 1].

Domain architecture, conserved sequence motifs, and consensus cleavage site of crustacean apoCr

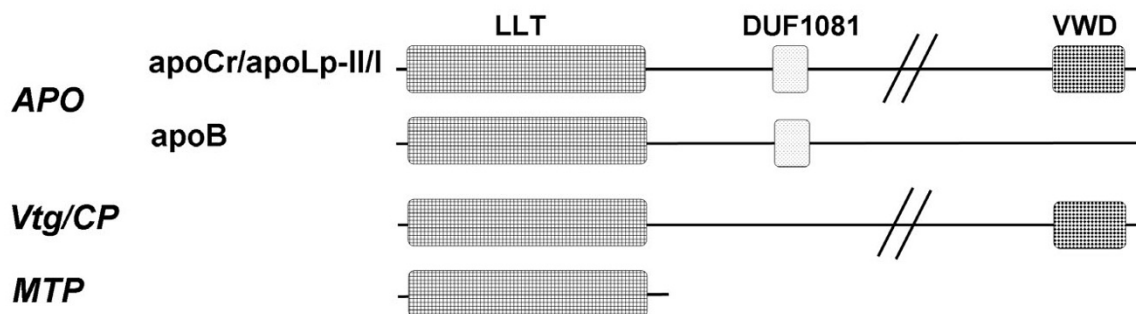
A CD-Search [38] revealed that the amino acid sequence of *Penaeus semisulcatus* and other crustacean apoCr include three conserved structural domains (Figure 1). A lipoprotein N-terminal domain (LPD-N, SMART accession number SM00638), also known as vitellogenin_N (Pfam accession number PF01347) and enclosed in the large lipid transfer (LLT) module [15] was identified in the N-terminal part of apoCr. A DUF1081 domain (Pfam accession number PF06448) of unknown function was identified in the central region, and a von Willebrand-factor type-D domain (VWD) (Pfam accession number PF00094) at the C-terminal end. While the VWD domain, initially identified in the human von Willebrand factor (vWF), is distributed over a wide range of proteins, the LLT module/LPD-N domain, part of the lipovitellin 1 subunit of vertebrate Vtg [6–12], contains twenty-two N-terminal conserved amino acid sequence motifs (N1 to N22) as a common denominator of LLTP [15]. The alignment

of these conserved sequence motifs, including those retrieved from apoCr is published as supporting information [see Additional file 2]. Similarly to apoLp-II/I, crustacean apoCr lacks polyserine tracks that correspond to the phosphotyrosine domain in the Vtg of vertebrates and some non-vertebrate species.

In insects, the apoLp-II/I is cleaved before its secretion into apoLp-II and apoLp-I (the subunits were indicated from the N-terminal to the C-terminal ends of the precursor protein) at a consensus cleavage site, RX(R/K)R, for dibasic endoprotease processing [35,37], and this cleavage is mediated by furin [39]. Likewise, crustacean apoCr is also cleaved at a similar dibasic site for subtilisin-like endoprotease processing [18,21,23,24]. This consensus site occurs at amino acid position 725/728 in *Penaeus semisulcatus* (RTRR), between conserved motifs N19 and N20, and aligns very well between crustacean apoCrs and insect apoLp-II/Is [see Additional file 1]. The N-terminal apoCr ~74-kDa subunit in *Penaeus semisulcatus* [23] and other decapod species, apoCr-II, displays high amino acid sequence similarities with the ~80-kDa apoLp-II [35,37]. For example, the *Penaeus semisulcatus* 74-kDa apoCr-II subunit displays 22%–43% identity-similarity with *Locusta migratoria* apoLp-II, while *Locusta migratoria* apoLp-II displays no more than 33%–55% and 28%–47% identity-similarity with *Manduca sexta* and *Drosophila melanogaster* apoLp-II, respectively. A similar consensus cleavage site is present in most insect Vtgs, but at a different location, in the extended region between motifs N6 and N7 [15]. However, a potential subtilisin-like convertase site is present between motifs N19 and N20 of nematode and *Crassostrea gigas* Vtgs.

Phylogenetic analysis and correlation with LLTP domain architecture and protein function

The evolutionary relationship of genes in the LLTP superfamily was evaluated after aligning the twenty-two conserved N-terminal sequence motifs of the LLT module of selected LLTP sequences [see Additional file 2] and the phylogenetic analyses were conducted by using the neighbor-joining (NJ) and the maximum likelihood (ML) (Figure 2) or the Bayesian inference (BI) (Figure 3) methods. The phylogenetic tree separated extracellular (apoCr, apoLp-II/I, apoB, Vtg, and CP) from intracellular (MTP family) LLTP sequences with confidence into separate clusters, suggesting that these extracellular proteins arose from a common ancestor (hereafter, support bootstrap confidence level values of NJ: 96 and ML: 88, and posterior probability of BI: 0.80). Significant internal branches defined two sequence groups. The first group, named the APO family, suggested that apoCr, apoLp-II/I, and apoB formed a monophyletic branch (NJ: 81; ML: 78; BI: 1.00), with two subclusters defined by highly significant internal branches, corresponding to crustacean decapod apoCr

**Figure 1**

Domain architecture of large lipid transfer protein (LLTP) superfamily members. Large lipid transfer (LLT) module [15] also referred to as Vitellogenin_N and LPD-N domain, DUF1081, and VWD domains are indicated from the N-terminal to the C-terminal ends of the proteins. LLTP families are 1) APO family including vertebrate apolipoprotein B (apoB), decapod crustacean apolipocrustacein (apoCr), and insect apolipophorin II/I (apoLp-II/I) subfamilies, 2) Vtg/CP family including vitellogenin (Vtg) and crustacean clotting protein (CP) subfamilies, and the large subunit of microsomal triglyceride transfer protein (MTP) family.

and insect apoLp-II/I subfamilies (NJ: 87; ML: 80; BI: 1.00) on one side and to vertebrate apoB subfamily (NJ: 100; ML: 100; BI: 1.00) on the other side. Internal sub-branching of the apoCr/apoLp-II/I subcluster was consistent with the species tree, i.e. crustacean and insect sequences were separated in two groups, each one with high support values (NJ: 100; ML: 100; BI: 0.98 and NJ: 100; ML: 100; BI: 1.00, respectively). Duplicated *Metapenaeus ensis* apoCr sequences were clustered together and therefore both belonged to the apoCr subfamily. The second group of sequences named the Vtg/CP family supported by a BI posterior probability of 0.82 grouped with confidence in separate clusters: vertebrate, mollusk, nematode, and insect Vtg, and crustacean CP (NJ: ≥ 99 ; ML: 100; BI: ≥ 96).

With regard to other crustacean sequences, one point must be highlighted. Vtg from *Daphnia magna*, a species belonging to the class of *Branchiopoda*, and CP from decapod crustaceans, belonging to the class of *Malacostraca*, tend to be grouped with insect Vtg (NJ: 85; ML: 55; BI: 1.00). It should be noted that *Daphnia magna* Vtg is fused with a superoxide dismutase module at the N-terminal end [40]. Therefore, these crustacean proteins are more closely related to *Arthropoda* Vtg than decapod apoCr. Vtg, CP, apoCr, and apoLp-II/I members contain a VWD domain at their C-terminal end (Figure 1) that may be implicated in the clotting cascade and multimerization, as demonstrated with human vWF [41,42]. The functional

significance of the VWD domain in Vtg requires additional elucidation. Recent data demonstrated that teleost fish Vtg exhibits an agglutinin activity and may be involved in defense reactions [43,44], and also that insect apoLp-II/I is involved in the immune response [45,46], suggesting a new meaning for the conservation of this domain during evolution. The fact that the phylogenies presented in Figures 2 and 3 were performed using sequences of the LLT module/LPD-N domain reinforces the hypothetical dual-functional nature of crustacean CP. This constitutes the only protein fraction of crustacean VLDL, and its involvement in lipid transport was suggested in earlier studies [31,47,48]. It should be noted that, in crustaceans, the most prominent lipoprotein in the hemolymph, LP1 or BGBP, is not a member of the LLTP superfamily but is also associated with defense reactions [49], supporting a link between lipid transport and immune systems in these animals.

It has been demonstrated that the LPD-N domain found in LLTP forms a "lipid pocket", enabling lipid loading of Vtg [50-53]. By comparative analysis, it was suggested that, in apoB, this domain may similarly form a lipovitelin-like "proteolipid" intermediate containing a lipid pocket that requires MTP for assembly [54]. The presence of apoCr and CP, the latter structurally related to Vtg, in the same shrimp species (*Penaeus monodon*), and of both apoLp-II/I and Vtg in honey bees (*Apis mellifera*), from separate clusters according to our phylogenetic analysis

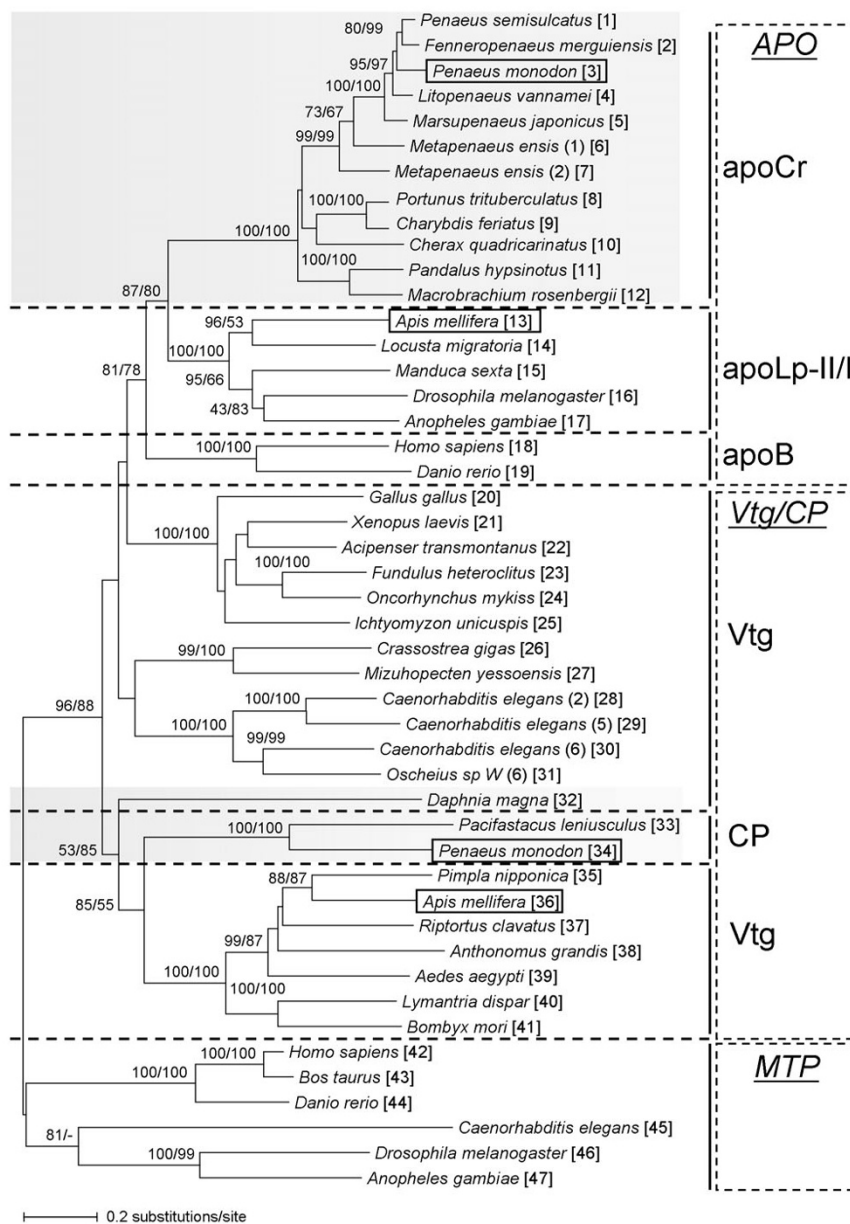


Figure 2

Neighbor-joining (NJ) and maximum likelihood (ML) phylogenetic reconstruction of the LLTP superfamily.

Tree topology was based on the alignment of the conserved amino acid sequence motifs of the LLT module [see Additional file 2]. Horizontal lengths of branches are proportional to the estimated numbers of amino acid substitutions produced by the NJ algorithm. Numerals at each node show local bootstrap confidence levels estimated by bootstrap pseudoreplicates in NJ and ML, respectively. Numbers are indicated when at least one bootstrap confidence level value at each node is ≥ 75 . The ML inference yielded a similar topology as the NJ tree at the level of significance used. However (-) indicates that this cluster was not formed in the ML analysis. Crustacean sequences are shaded in gray. *Penaeus monodon* apoCr and CP and *Apis mellifera* apoLp-II/I and Vtg are framed as an indication of the presence in one species of homologous extracellular proteins from two different LLTP families. A number in parentheses, corresponding to the extension number used in sequence databases, is attached to the species name, when more than one characterized protein was found in a defined LLTP family. The number in bracket after the species name refers to the code number used for species designation in Figure 3. Abbreviations of LLTP families and subfamilies are as in the legend to Figure 1.

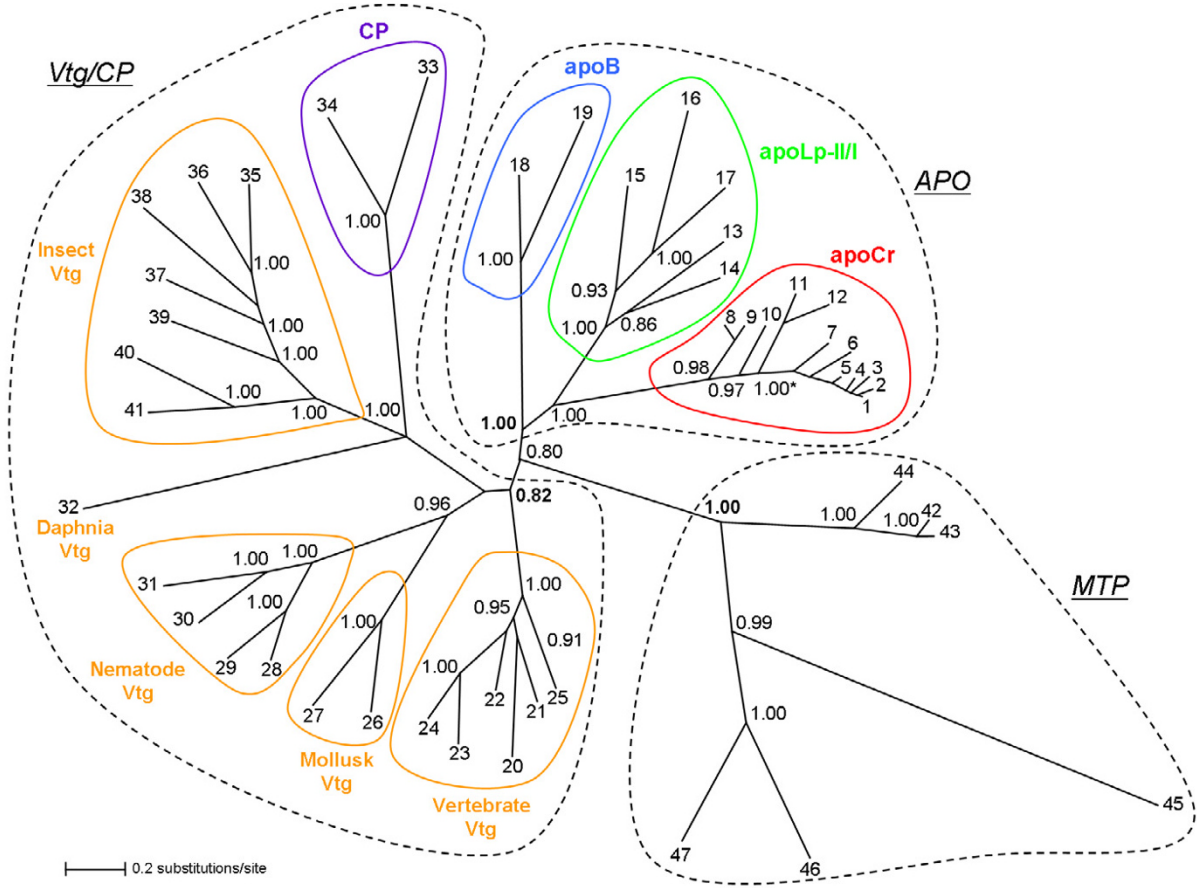


Figure 3
Bayesian inference (BI) phylogenetic reconstruction of the LLTP superfamily. The tree was based on the alignment of the conserved amino acid sequence motifs of the LLT module [see Additional file 2]. Species are indicated as numbers and are noted as in Figure 2. Branch lengths are proportionate to BI estimates of numbers of amino acid substitutions and numerals at each node indicate the posterior probability estimated by MrBayes under the model, summed over 9,000 tree samples. Posterior probabilities ≥ 0.80 are shown. Due to graphical space limitation, 1.00* indicates that the posterior probabilities values of apoCr terminal nodes were 1.00. Abbreviations of LLTP families and subfamilies are as in the legend to Figure 1. The support posterior probability values for LLTP families are bolded.

(Figures 2 and 3), strongly suggest that these sequences diverged after an ancient duplication event leading to the separation of an APO paralogous group from Vtg/CP sequences. Through their LLT module/LPD-N domain, these homologous sequences retained the capacity to bind and transport lipids, a function shared with MTP family members that is probably the ancestral function of LLTP [15,55]. The use of truncated forms of apoB in cell cultures showed that the percentage of lipid associated with apoB-100 truncated near the C-terminal end of the LPD-N domain, approached zero. Furthermore, there appeared to be a threshold in apoB size, between 1050 and 1250 amino acids from the N-terminal end, below which the polypeptide may not form a lipoprotein particle

[reviewed in ref. [56]]. However, apoB needs to reach a minimum critical length of 884 amino acids for lipoprotein assembly, an MTP-dependent process relying on a portion of apoB located between amino acids 834 and 1134 [57]. Interestingly, apoCr, apoLp-II/I and apoB sequences were found to have a common DUF1081 domain, of unknown function (Figure 1). Protein database screening revealed that this domain was specifically associated with APO sequences (data not shown), and spanned amino acids 957–1072 of human apoB. *Drosophila* MTP promotes the assembly and secretion of human apoB-41 [58], and human MTP enhances the secretion of *Xenopus laevis* Vtg [55], which lacks the DUF1081 domain. A mutant form of human MTP was

still able to promote *Xenopus* Vtg synthesis, whereas secretion of human apoB was abolished, suggesting that requirement of apoB for interacting with MTP is more stringent than that of Vtg. This may be related to the presence of the DUF1081 domain in apoB. This domain may be acquired in the APO ancestor through domain accretion and neofunctionalization potentially resulting in better biogenesis of neutral lipid-rich lipoproteins.

Gene duplication facilitates functional divergence but functional constraints enable the retention of genes with overlapping or redundant functions. In addition to their role in plasma and hemolymph lipid transport between somatic tissues, extracellular LLTP, mainly Vtg and apoCr, facilitate the massive deposition of yolk reserves inside the oocytes of most oviparous species. ApoB, apoLp-II/I, and possibly apoCr containing lipoproteins are involved in neutral lipid deposition in the oocyte after receptor-mediated endocytosis on the same or similar oocyte-specific receptor used by Vtg, which belongs to the LDL receptor superfamily [59-61]. Ovarian lipolysis of these circulating lipoproteins may also be a main source of lipids for the growing oocyte [reviewed in ref. [62]]. The low lipid load of crustacean CP and its abundance in shrimp hemolymph cannot fully account for the amount of lipid accumulated within the oocytes [63] and may explain a major role of apoCr in this process in conjunction with the BGBP/LP1.

Relations among the Arthropoda subphyla and the major groups of crustaceans are still a matter for debate. Crustacea exhibit extensive variability in body plans, compared to other arthropod groups and monophyly of crustaceans and relationships among the constituent lineages are controversial [64]. While mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic [65], molecular analyses on rRNA and protein-coding nuclear gene sequences indicate that crustaceans and hexapods form a clade (Pancrustacea or Tetraconata) and that the sister group of Hexapoda is Branchiopoda (fairy shrimps, tadpole shrimps, etc.), rather than Malacostraca (lobsters, crabs, true shrimps, isopods, etc.) [66,67], thereby making hexapods terrestrial crustaceans and the traditionally defined Crustacea, paraphyletic. Additional Branchiopoda and non Decapoda apoCr and Vtg/CP sequences, together with their role as a main egg yolk precursor protein, may help to clarify the Pancrustacea-major splits, e.g. provide support for the branchiopods being closest to hexapods and reject the alternative, e.g. malacostracan-hexapod association. Additional data will also clarify the relationships among the many malacostracan subgroups [68].

Conclusion

Observation of modular architecture and phylogenetic analyses demonstrated that the main egg yolk precursor protein of decapod crustaceans is a member of the LLTP superfamily, is homologous to insect apoLp-II/I and vertebrate apoB, and would be more appropriately called apolipocrustacein (apoCr) rather than Vtg. The presence of apoCr and a clotting protein structurally related to Vtg from the same shrimp species (e.g. *Penaeus monodon*) and both apoLp-II/I and Vtg from insect species (e.g. *Apis mellifera*), suggest that, in addition to their involvement in the lipid metabolism, extracellular LLTP may have acquired other functions during metazoan evolution, e.g. involvement in clotting cascade and immune response. The findings of this study may be also a starting point for using LLTP phylogeny and their functional role to clarify the controversial relationships among Pancrustacean constituent lineages.

Methods

Data sets

Accession numbers for LLTP sequences used from GenBank™/EBI or UniProt databases are GenBank:AY051318 (*Penaeus semisulcatus*), GenBank:AB033719 (*Marsupenaeus japonicus*), GenBank:AY103478 (Vg1) and GenBank:AY530205 (Vg2) (*Metapenaeus ensis*), GenBank:AY321153 (*Litopenaeus vannamei*), GenBank:AY499620 (*Fenneropenaeus merguensis*), and GenBank:DQ288843 (*Penaeus monodon*) for Crustacea Decapoda Penaeidae apoCr sequences; GenBank:AB117524 (*Pandalus hypsinotus*), GenBank:AF306784 (*Cherax quadricarinatus*), GenBank:AB056458 (*Macrobrachium rosenbergii*), GenBank:AY724676 (*Charybdis feriatius*), and GenBank:AAX94762 (*Portunus trituberculatus*) for Crustacea Decapoda Pleocyemata apoCr sequences; UniProt:Q9U943 for locust (*Locusta migratoria*), UniProt:Q25490 for tobacco hornworm (*Manduca sexta*), UniProt:Q9V496 for fruit fly (*Drosophila melanogaster*), GenBank:XP_392490 for honey bee (*Apis mellifera*), GenBank:XP_321226 for African malaria mosquito (*Anopheles gambiae*), and GenBank:BAB32641 for silk moth (*Samia cynthia ricini*) apoLp-II/I sequences; GenBank:XP_694827 for zebrafish (*Danio rerio*), GenBank:XP_419979 for chicken (*Gallus gallus*), and UniProt:X04714 for human (*Homo sapiens*) apoB sequences; GenBank:AAD16454 for freshwater crayfish (*Pacifastacus leniusculus*), and GenBank:AAF19002 for shrimp (*Penaeus monodon*) CP sequences; GenBank:NP_610075 for fruit fly (*Drosophila melanogaster*), GenBank:XP_319421 for African malaria mosquito (*Anopheles gambiae*), GenBank:AAR27937 for nematode (*Caenorhabditis elegans*), GenBank:XP_788526 for sea urchin (*Strongylocentrotus purpuratus*), UniProt:Q8AXV7 for zebrafish (*Danio rerio*), UniProt:X78567 for bovine (*Bos taurus*), and UniProt:P55157 for human

(*Homo sapiens*) large subunit of MTP sequences; GenBank: [BAC22716](#) for oyster (*Crassostrea gigas*), GenBank: [BAB63260](#) (partial sequence) for Yesso scallop (*Mizuhopecten yessoensis*), UniProt: P55155 (Vtg2), UniProt: P06125 (Vtg5), UniProt: P18948 (Vtg6) for nematode (*Caenorhabditis elegans*), UniProt: T18561 (Vtg6) for nematode (*Oscheius* sp., CW1 strain), UniProt: Q91062 for lamprey (*Ichthyomyzon unicuspis*), UniProt: Q90243 for sturgeon (*Acipenser transmontanus*), UniProt: U07055 for mummichog (*Fundulus heteroclitus*), UniProt: Q92093 for rainbow trout (*Oncorhynchus mykiss*), UniProt: P18709 for *Xenopus* (*Xenopus laevis*), and UniProt: P87498 for chicken (*Gallus gallus*) (Vtg1) Vtg sequences. As previously demonstrated, using the NJ method [69], insect Vtg sequences available in databases clustered with high confidence in a group separated from other Vtg sequences, including those of crustaceans apoCr. Vtg sequences representative of five insect orders (Lepidoptera, Diptera, Coleoptera, Hemiptera, and Hymenoptera) were therefore selected for analysis and were: UniProt: Q27309 for silkworm (*Bombyx mori*), UniProt: U60186 for gypsy moth (*Lymantria dispar*), UniProt: U02548 for yellow fever mosquito (*Aedes aegypti*), UniProt: Q05808 for boll weevil (*Anthonomus grandis*), UniProt: U97277 for bean bug (*Riptortus clavatus*), GenBank: [NP_001011578](#) for honey bee (*Apis mellifera*), and GenBank: [AF026789](#) for parasitoid wasp (*Pimpla nipponica*).

Sequence analyses

The deduced amino acid sequence of *Penaeus semisulcatus* [23] was subjected to a BLASTP search using the Blosum 62 matrix [70]. Amino acid sequences corresponding to the best hits were then aligned using Clustal X software, version 1.8 [71]. Sequence alignments were edited and manually corrected with GeneDoc software [72]. The twenty-two N-terminal conserved amino acid sequence motifs (N1 to N22) of the LLT module were extracted from selected LLTP and aligned as previously described [15] resulting in a concatenated sequence 351 amino acid long. Domain architecture of LLTP was examined and compared using CDART [73].

The phylogenetic tree and branch support values were estimated using three different methodologies of phylogenetic reconstruction: 1) NJ, 2) ML and 3) BI. NJ algorithm was based on the number of amino acid substitutions per site with the Poisson-correction distance method and pairwise-deletion option for gap sites, and bootstrap support values were obtained with 5,000 pseudoreplicates. The distance analysis was carried out with MEGA 3.1 [74]. ML analyses was carried out with PHYML v2.4.4 [75] starting from the BIONJ tree, and the gamma distribution for rate heterogeneity across sites (Γ) was modeled with a four-category Γ distribution and a shape parameter equal

to 2. The WAG substitution model [76] was selected by ProtTest v1.3 [77], following the Akaike information criterion, as best-fitting model among the models tested that could be used in PHYML. Bootstrap values were based on 500 pseudoreplicates to estimate support for the nodes of the ML tree. A graphical representation of the ML phylogeny was generated with MEGA 3.1 tree explorer. In the non-parametric-bootstrap trees, bootstrap confidence level values ≥ 75 were accepted as significant. BI was performed using MrBayes v3.1.2 [78] with the WAG model of amino acid substitution provided in the package. Two simultaneous runs each with four simultaneous Markov Chain Monte Carlo (MCMC) chains run initially for 1,000,000 generations, after which the average standard deviation of split frequencies was 0.004614, saving the current tree to file every 100 generations for a total of 10,000 trees in the initial sample. Default cold and heated chain parameters were used. MCMC runs were summarized and further investigated for convergence of all parameters, using the *sump* and *sumt* commands in MrBayes and the computer program Tracer version 3.1 [79]. Stationarity was determined to have occurred by the 100,000th generation. Accordingly, the first 1000 trees prior to log likelihood stabilization were discarded (as burn-in), and the following 9,000 tree samples were used to estimate topology and tree parameters. The percentage of times a node occurred within those 9,000 trees was interpreted as the posterior probability of the node. A graphical representation of the majority rule consensus tree was generated with TreeView v1.6.6 program [80]. In the Bayesian tree, clades were accepted as significant at ≥ 0.80 posterior probability.

Abbreviations

apoB, apolipoprotein B; apoCr, apolipocrustacein; apoLp-II/I, apolipoprotein II/I; APO, apolipocrustacein/apolipoprotein/apolipoprotein monophyletic group; BI, Bayesian inference; CP, clotting protein; LLT, large lipid transfer; LLTP, large lipid transfer proteins; MCMC, Markov Chain Monte Carlo; ML, maximum likelihood; MTP, large subunit of microsomal triglyceride transfer protein; neighbor-joining, NJ; very high density lipoprotein, (VHDL); Vtg, vitellogenin; VWD, von Willebrand factor type D; vWF, von Willebrand factor.

Authors' contributions

J-C.A., E.L. and P.J.B. designed research; J-C.A. and P.J.B. performed research; J-C.A., E.L. and P.J.B. analyzed data; and J-C.A., E. L. and P.J.B. wrote the paper. All authors read and approved the final manuscript.

Additional material

Additional File 1

Alignment of deduced amino acid sequences from crustacean apolipoproteins (apoCr) and insect apolipoproteins II/I (apoLp-II/I). The conservative substitutions allowed were colored and defined as follows: A, G, S, T; E, D; R, K, H; Q, N; V, I, L, M; Y, F, W; P; and C. Gaps inserted to optimize alignments are indicated by dashes. Sites of identical or conserved amino acids in all sequences are highlighted in red and gray, respectively. The conserved functional subtilisin-like endoprotease cleavage site is highlighted in yellow. The name of each domain and of each conserved motif, from N1 to N22, of the LLT module [15] is indicated above and below the alignments, respectively.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-3-S1.pdf>]

Additional File 2

Alignment of LLT module conserved protein sequence motifs extracted from LLTP superfamily members. The conservative substitutions allowed were colored and defined as in Additional file 1. Gaps inserted to optimize alignments are indicated by dashes. Single dots indicate missing data. The site of identical conserved amino acids in all sequences is highlighted in red. The name of each conserved motif, from N1 to N22, of the LLT module [15] is indicated below the alignments.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-3-S2.pdf>]

Acknowledgements

This work was supported in part by the French Ministry of Research and Education (to P.J.B.), and the European Union Grant FAIR CT-97-3660 (to E.L.).

References

- Wallace RA, Walker SL, Hauschka PV: **Crustacean lipovitellin Isolation and characterization of the major high-density lipoprotein from the eggs of decapods.** *Biochemistry* 1967, **6**:1582-1590.
- Kerr MS: **The hemolymph proteins of the blue crab, *Callinectes sapidus* II A lipoprotein serologically identical to oocyte lipovitellin.** *Dev Biol* 1969, **20**:1-17.
- Pan ML, Bell WJ, Telfer WH: **Vitellogenic blood protein synthesis by insect fat body.** *Science* 1969, **165**:393-394.
- Meusy JJ: **Vitellogenin, the extraovarian precursor of the protein yolk in Crustacea: a review.** *Reprod Nutr Dev* 1980, **20**:1-21.
- Wahli W: **Evolution and expression of vitellogenin genes.** *Trends Genet* 1988, **4**:227-232.
- Chen JS, Sappington TW, Raikhel AS: **Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry.** *J Mol Evol* 1997, **44**:440-451.
- Sappington TW, Raikhel AS: **Molecular characteristics of insect vitellogenins and vitellogenin receptors.** *Insect Biochem Mol Biol* 1998, **28**:277-300.
- Sappington TW, Oishi K, Raikhel AS: **Structural Characteristics of Insect Vitellogenins.** In *Progress in Vitellogenesis* Edited by: Adiyodi KG, Adiyodi RG. Enfield, NH: Sciences Publishers, Inc, USA; 2002:69-101.
- Brooks JM, Wessel GM: **The major yolk protein in sea urchins is a transferrin-like, iron binding protein.** *Dev Biol* 2002, **245**:1-12.
- Yokota Y, Sappington TW: **Vitellogen and Vitellogenin in Echinoderms.** In *Progress in Vitellogenesis* Edited by: Adiyodi KG, Adiyodi RG. Enfield, NH: Sciences Publishers, Inc, USA; 2002:201-221.
- Bownes M, Pathirana S: **The Yolk Proteins of Higher Diptera.** In *Progress in Vitellogenesis* Edited by: Adiyodi KG, Adiyodi RG. Enfield, NH: Sciences Publishers, Inc, USA; 2002:103-130.
- Lim EH, Teo BY, Lam TJ, Ding JL: **Sequence analysis of a fish vitellogenin cDNA with a large phosphotyrosine domain.** *Gene* 2001, **277**:175-186.
- Baker ME: **Invertebrate vitellogenin is homologous to human von Willebrand factor.** *Biochem J* 1988, **256**:1059-1061. (1988a)
- Baker ME: **Is vitellogenin an ancestor of apolipoprotein B-100 of human low-density lipoprotein and human lipoprotein lipase?** *Biochem J* 1988, **255**:1057-1060. (1988b)
- Babin PJ, Bogerd J, Kooiman FP, Van Marrewijk WJ, Van der Horst DJ: **Apolipoprotein III/I, apolipoprotein B, vitellogenin, and microsomal triglyceride transfer protein genes are derived from a common ancestor.** *J Mol Evol* 1999, **49**:150-160.
- Shelness GS, Ledford AS: **Evolution and mechanism of apolipoprotein B-containing lipoprotein assembly.** *Curr Opin Lipidol* 2005, **16**:325-332.
- Byrne BM, Gruber M, Ab G: **The evolution of egg yolk proteins.** *Prog Biophys Mol Biol* 1989, **53**:33-69.
- Tsutsui N, Kawazoe I, Ohira T, Jasmani S, Yang W, Wilder M, Aida K: **Molecular characterization of a cDNA encoding vitellogenin and its expression in the hepatopancreas and ovary during vitellogenesis in the kuruma prawn, *Penaeus japonicus*.** *Zool Sci* 2000, **17**:651-660.
- Tseng DY, Chen YN, Kou GH, Lo CF, Kuo CM: **Hepatopancreas is the extraovarian site of vitellogenin synthesis in black tiger shrimp, *Penaeus monodon*.** *Comp Biochem Physiol A Mol Integr Physiol* 2001, **129**:909-917.
- Abdu U, Davis C, Khalaila I, Sagi A: **The vitellogenin cDNA of *Cherax quadricarinatus* encodes a lipoprotein with calcium binding ability, and its expression is induced following the removal of the androgenic gland in a sexually plastic system.** *Gen Comp Endocrinol* 2002, **127**:263-272.
- Okuno A, Yang WJ, Jayasankar V, Saido-Sakanaka H, Huong do TT, Jasmani S, Atmomasoro M, Subramoniam T, Tsutsui N, Ohira T, Kawazoe I, Aida K, Wilder MN: **Deduced primary structure of vitellogenin in the giant freshwater prawn, *Macrobrachium rosenbergii*, and yolk processing during ovarian maturation.** *J Exp Zool* 2002, **292**:417-429.
- Wilder MN, Subramoniam T, Aida K: **Yolk Proteins of Crustacea.** In *Progress in Vitellogenesis* Edited by: Adiyodi KG, Adiyodi RG. Enfield, NH: Sciences Publishers, Inc, USA; 2002:131-174.
- Avarre JC, Michelis R, Tietz A, Lubzens E: **Relationship between vitellogenin and vitellin in a marine shrimp (*Penaeus semisulcatus*) and molecular characterization of vitellogenin complementary DNAs.** *Biol Reprod* 2003, **69**:355-364.
- Tsang WS, Quackenbush LS, Chow BK, Tiu SH, He JG, Chan SM: **Organization of the shrimp vitellogenin gene: evidence of multiple genes and tissue specific expression by the ovary and hepatopancreas.** *Gene* 2003, **303**:99-109.
- Kung SY, Chan SM, Hui JH, Tsang WS, Mak A, He JG: **Vitellogenesis in the sand shrimp, *Metapenaeus ensis*: the contribution from the hepatopancreas-specific vitellogenin gene (MeVg2).** *Biol Reprod* 2004, **71**:863-870.
- Tsutsui N, Saido-Sakanaka H, Yang WJ, Jayasankar V, Jasmani S, Okuno A, Ohira T, Okumura T, Aida K, Wilder MN: **Molecular characterization of a cDNA encoding vitellogenin in the constricted shrimp, *Pandalus hypsinotus* and site of vitellogenin mRNA expression.** *J Exp Zool A Comp Exp Biol* 2004, **301**:802-814.
- Yang F, Xu HT, Dai ZM, Yang WJ: **Molecular characterization and expression analysis of vitellogenin in the marine crab *Portunus trituberculatus*.** *Comp Biochem Physiol B Biochem Mol Biol* 2005, **142**:456-464.
- Phiriyangkul P, Utarabhand P: **Molecular characterization of a cDNA encoding vitellogenin in the banana shrimp, *Penaeus (Litopenaeus) merguensis* and sites of vitellogenin mRNA expression.** *Mol Reprod Dev* 2006, **73**:410-423.
- Raviv S, Parnes S, Segall C, Davis C, Sagi A: **Complete sequence of *Litopenaeus vannamei* (Crustacea: Decapoda) vitellogenin cDNA and its expression in endocrinologically induced sub-adult females.** *Gen Comp Endocrinol* 2006, **145**:39-50.
- Doolittle RF, Riley M: **The amino-terminal sequence of lobster fibrinogen reveals common ancestry with vitellogenins.** *Biochem Biophys Res Commun* 1990, **167**:16-19.
- Hall M, Wang R, van Antwerpen R, Sottrup-Jensen L, Soderhall K: **The crayfish plasma clotting protein: a vitellogenin-related protein responsible for clot formation in crustacean blood.** *Proc Natl Acad Sci USA* 1999, **96**:1965-1970.
- Tsutsui N, Kim YK, Jasmani S, Ohira T, Wilder MN, Aida K: **The dynamics of vitellogenin gene expression differs between intact and eyestalk ablated kuruma prawn *Penaeus (Marsupenaeus) japonicus*.** *Fisheries Sci* 2005, **71**:249-256.
- Voloch CM, Freire PR, Russo CA: **Molecular phylogeny of penaeid shrimps inferred from two mitochondrial markers.** *Genet Mol Res* 2005, **4**:668-674.
- McGinnis S, Madden TL: **BLAST: at the core of a powerful and diverse set of sequence analysis tools.** *Nucleic Acids Res* 2004, **32**:W20-25.

35. Ryan RO, van der Horst DJ: **Lipid transport biochemistry and its role in energy production.** *Annu Rev Entomol* 2000, **45**:233-260.
36. Arrese EL, Canavoso LE, Jouni ZE, Pennington JE, Tsuchida K, Wells MA: **Lipid storage and mobilization in insects: current status and future directions.** *Insect Biochem Mol* 2001, **31**:7-17.
37. Canavoso LE, Jouni ZE, Karnas KJ, Pennington JE, Wells MA: **Fat metabolism in insects.** *Annu Rev Nutr* 2001, **21**:23-46.
38. Marchler-Bauer A, Bryant SH: **CD-Search: protein domain annotations on the fly.** *Nucleic Acids Res* 2004, **32**:V327-331.
39. Smolenaars MM, Kasperaitis MA, Richardson PE, Rodenburg KW, Van der Horst DJ: **Biosynthesis and secretion of insect lipoprotein: involvement of furin in cleavage of the apoB homolog, apolipoprotein-III.** *J Lipid Res* 2005, **46**:412-421.
40. Kato Y, Tokishita S, Ohta T, Yamagata H: **A vitellogenin chain containing a superoxide dismutase-like domain is the major component of yolk proteins in cladoceran crustacean *Daphnia magna*.** *Gene* 2004, **334**:157-165.
41. Bonthron DT, Handin RI, Kaufman RJ, Wasley LC, Orr EC, Mitsock LM, Ewenstein B, Loscalzo J, Ginsburg D, Orkin SH: **Structure of pre-pro-von Willebrand factor and its expression in heterologous cells.** *Nature* 1986, **324**:270-273.
42. Jorieux S, Fressinaud E, Goudemand J, Gaucher C, Meyer D, Mazurier C: **Conformational changes in the D' domain of von Willebrand factor induced by CYS 25 and CYS 95 mutations lead to factor VIII binding defect and multimeric impairment.** *Blood* 2000, **95**:3139-3145.
43. Zhang S, Sun Y, Pang Q, Shi X: **Hemagglutinating and antibacterial activities of vitellogenin.** *Fish Shellfish Immunol* 2005, **19**:93-95.
44. Shi X, Zhang S, Pang Q: **Vitellogenin is a novel player in defense reactions.** *Fish Shellfish Immunol* 2006, **20**:769-772.
45. Cheon HM, Shin SW, Bian G, Park JH, Raikhel AS: **Regulation of lipid metabolism genes, lipid carrier protein lipophorin, and its receptor during immune challenge in the mosquito *Aedes aegypti*.** *J Biol Chem* 2006, **281**:8426-8435.
46. Ma G, Hay D, Li D, Asgari S, Schmidt O: **Recognition and inactivation of LPS by lipophorin particles.** *Dev Comp Immunol* 2006, **30**:619-626.
47. Hall M, van Heusden MC, Soderhall K: **Identification of the major lipoproteins in crayfish hemolymph as proteins involved in immune recognition and clotting.** *Biochem Biophys Res Commun* 1995, **216**:939-946.
48. Avarre JC, Michelis R, Hall M, Soderhall K, Khayat M, Tietz A, Lubzens E: **Lipid composition during sexual development of the noble crayfish *Astacus astacus* and effect of a fungal infection.** *Invert Reprod Dev* 2002, **41**:251-259.
49. Cerenius L, Liang Z, Duvic B, Keyser P, Hellman U, Palva ET, Iwanaga S, Soderhall K: **Structure and biological activity of a 1,3-beta-D-glucan-binding protein in crustacean blood.** *J Biol Chem* 1994, **269**:29462-29467.
50. Raag R, Appelt K, Xuong NH, Banaszak L: **Structure of the lamprey yolk lipid-protein complex lipovitellin-phosvitin at 28 Å resolution.** *J Mol Biol* 1988, **200**:553-569.
51. Banaszak L, Sharrock W, Timmins P: **Structure and function of a lipoprotein: lipovitellin.** *Annu Rev Biophys Chem* 1991, **20**:221-246.
52. Sharrock WJ, Rosenwasser TA, Gould J, Knott J, Hussey D, Gordon JI, Banaszak L: **Sequence of lamprey vitellogenin Implications for the lipovitellin crystal structure.** *J Mol Biol* 1992, **226**:903-907.
53. Anderson TA, Levitt DG, Banaszak LJ: **The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein.** *Structure* 1998, **6**:895-909.
54. Segrest JP, Jones MK, Dashti N: **N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a "lipid pocket" model for self-assembly of apob-containing lipoprotein particles.** *J Lipid Res* 1999, **40**:1401-1416.
55. Sellers JA, Hou L, Schoenberg DR, Batistuzzo de Medeiros SR, Wahli W, Shelness GS: **Microsomal triglyceride transfer protein promotes the secretion of *Xenopus laevis* vitellogenin A1.** *J Biol Chem* 2005, **280**:13902-13905.
56. Segrest JP, Jones MK, De Loof H, Dashti N: **Structure of apolipoprotein B-100 in low density lipoproteins.** *J Lipid Res* 2001, **42**:1346-1367.
57. Shelness GS, Hou L, Ledford AS, Parks JS, Weinberg RB: **Identification of the lipoprotein initiating domain of apolipoprotein B.** *J Biol Chem* 2003, **278**:44702-44707.
58. Sellers JA, Hou L, Athar H, Hussain MM, Shelness GS: **A *Drosophila* microsomal triglyceride transfer protein homolog promotes the assembly and secretion of human apolipoprotein B Implications for human and insect transport and metabolism.** *J Biol Chem* 2003, **278**:20367-20373.
59. Bujo H, Hermann M, Kaderli MO, Jacobsen L, Sugawara S, Nimpf J, Yamamoto T, Schneider WJ: **Chicken oocyte growth is mediated by an eight ligand binding repeat member of the LDL receptor family.** *EMBO J* 1994, **13**:5165-5175.
60. Sappington TW, Raikhel AS: **Ligand-binding domains in vitellogenin receptors and other LDL-receptor family members share a common ancestral ordering of cysteine-rich repeats.** *J Mol Evol* 1998, **46**:476-487.
61. Cheon HM, Seo J, Sappington TW, Raikhel AS: **Molecular characterization of the VLDL receptor homolog mediating binding of lipophorin in oocyte of the mosquito *Aedes aegypti*.** *Insect Biochem Mol Biol* 2001, **31**:753-760.
62. Ziegler R, Van Antwerpen R: **Lipid uptake by insect oocytes.** *Insect Biochem Mol Biol* 2006, **36**:264-272.
63. Ravid T, Tietz A, Khayat M, Boehm E, Michelis R, Lubzens E: **Lipid accumulation in the ovaries of a marine shrimp *Penaeus semisulcatus* (de haan).** *J Exp Biol* 1999, **202**:1819-1829.
64. Schram FR, Koenermann S: **Are the crustaceans monophyletic?** In *Assembling the Tree of Life* Edited by: Cracraft J, Donoghue MJ. Oxford University Press, USA; 2004:319-329.
65. Cook CE, Yue Q, Akam M: **Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic.** *Proc R Soc B* 2005, **272**:1295-1304.
66. Mallatt J, Giribet G: **Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch.** *Mol Phylogenet Evol* 2006, **40**:772-794.
67. Regier JC, Shultz JW, Kambic RE: **Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic.** *Proc R Soc B* 2005, **272**:395-401.
68. Martin JW, Davis GE: **An updated classification of the recent Crustacea.** In *Science Series No. 39* Los Angeles, CA: Natural History Museum of Los Angeles County; 2001.
69. Donnell DM: **Vitellogenin of the parasitoid wasp, *Encarsia formosa* (Hymenoptera: Aphelinidae): gene organization and differential use by members of the genus.** *Insect Biochem Mol Biol* 2004, **34**:951-961.
70. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
71. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
72. Nicholas KB, Nicholas HBJ, Deerfield DW: **GeneDoc: Analysis and visualization of genetic variation.** *EMBNWNEWS* 1997, **4**:14.
73. Geer LY, Domrachev M, Lipman DJ, Bryant SH: **CDART: protein homology by domain architecture.** *Genome Res* 2002, **12**:1619-1623.
74. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**:150-163.
75. Guindon S, Lethiec F, Duroux P, Gascuel O: **PHYML Online-aweb server for fast maximum likelihood-based phylogenetic inference.** *Nucleic Acids Res* 2005, **33**:W557-559.
76. Olsen R, Loomis WF: **A collection of amino acid replacement matrices derived from clusters of orthologs.** *J Mol Evol* 2005, **61**:659-665.
77. Abascal F, Zardoya R, Posada D: **ProtTest: Selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**:2104-2105.
78. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
79. Rambaut A, Drummond AJ: **Tracer: A program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes.** 2005 [<http://evolve.zoo.ox.ac.uk/software.html?id=tracer>].
80. Rod Page's Home Page: **TreeView: Tree drawing software for Apple Macintosh and Windows.** 2001 [<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>].