BMC Evolutionary Biology

BioMed Central

Research article

Structure-function evolution of the Transforming acidic coiled coil genes revealed by analysis of phylogenetically diverse organisms

Ivan H Still*, Ananthalakshmy K Vettaikkorumakankauv, Anthony DiMatteo and Ping Liang

Address: Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263, USA

Email: Ivan H Still* - ivan.still@roswellpark.org;
Ananthalakshmy K Vettaikkorumakankauv - ananthalakshmy.vettaikkorumakankauv@roswellpark.org;
Anthony DiMatteo - 09dimatteo@cua.edu; Ping Liang - ping.liang@roswellpark.org

* Corresponding author

Published: 18 June 2004

BMC Evolutionary Biology 2004, **4**:16

Received: 12 March 2004
Accepted: 18 June 2004

This article is available from: http://www.biomedcentral.com/1471-2148/4/16

Abstract

Background: Examination of ancient gene families can provide an insight into how the evolution of gene structure can relate to function. Functional homologs of the evolutionarily conserved transforming acidic coiled coil (*TACC*) gene family are present in organisms from yeast to man. However, correlations between functional interactions and the evolution of these proteins have yet to be determined.

Results: We have performed an extensive database analysis to determine the genomic and cDNA sequences of the TACCs from phylogenetically diverse organisms. This analysis has determined the phylogenetic relationship of the TACC proteins to other coiled coil proteins, the resolution of the placement of the rabbit TACC4 as the orthologue of human TACC3, and RHAMM as a distinct family of coiled coil proteins. We have also extended the analysis of the TACCs to the interaction databases of *C. elegans* and *D. melanogaster* to identify potentially novel TACC interactions. The validity of this modeling was confirmed independently by the demonstration of direct binding of human TACC2 to the nuclear hormone receptor RXRβ.

Conclusion: The data so far suggest that the ancestral TACC protein played a role in centrosomal/mitotic spindle dynamics. TACC proteins were then recruited to complexes involved in protein translation, RNA processing and transcription by interactions with specific bridging proteins. However, during evolution, the TACC proteins have now acquired the ability to directly interact with components of these complexes (such as the LSm proteins, nuclear hormone receptors, GAS41, and transcription factors). This suggests that the function of the TACC proteins may have evolved from performing assembly or coordination functions in the centrosome to include a more intimate role in the functional evolution of chromatin remodeling, transcriptional and posttranscriptional complexes in the cell.

Background

The evolution of complex organisms has been associated with the generation of gene families by the continual duplication of an initial relatively small set of ancestral genes. Through this process, followed by subsequent mutation, reduplication and exon shuffling between gene

families, genes have evolved both discrete, and partially redundant functions with their related family members. With the completion of the genome sequencing projects of human, mouse, rat, fruit fly and nematodes, we are now in a position to ask fundamental questions in regard to how genes interact in the context of the whole organism. Thus, with the appropriate application of bioinformatics, it is now possible to trace the lineage of particular genes and gene families, with related gene families in other organisms. Furthermore, with the growing amount of large-scale proteomic and genomic data becoming publicly available, this analysis can now be extended to reveal the complex interplay between evolution of gene structure and protein function.

The first Transforming acidic coiled coil gene, *TACC1*, was identified during the development of an expression map of the proximal short arm of human chromosome 8 [1]. Two additional *TACC* family members were subsequently identified and mapped to paralogous chromosomal regions on human chromosomes 4p16 and 10q26, physically close to members of the *FGFR* gene family [1-3]. This mapping data, together with identification of a single *TACC* gene in the protostomes *Caenorhabitis elegans*, and *Drosophila melanogaster* [4-6], led to the speculation that the ancestral *FGFR* and *TACC* genes were located physically close to each other. Thus, during the evolution of vertebrates, subsequent successive duplications of the ancestral gene cluster have given rise to three *TACC* family members located close to *FGFR* genes in humans. In accordance with the proposed quadruplication of the vertebrate genome during evolution, there is a fourth *FGFR* family member in vertebrates, raising the question of whether a fourth *TACC* gene is associated with *FGFR4* in vertebrate genomes. To date, only three active *TACC* genes have been cloned in humans [1-3], one in each of mouse [7], *Xenopus laevis* [8], *D. melanogaster* [4], and *C. elegans* [5,6]. Although two additional new candidate *TACC* family members, *Oryctolagus cuniculus TACC4* [9] and human *RHAMM* [10] have been proposed, their true identity and placement in the evolution of the TACC family is under debate. Thus, the identification and functional characterization of new members of the TACC family in other organisms, alternatively spliced isoforms of each TACC and comparison of the phylogenetic relationship of these genes relative to other members of the coiled coil superfamily will resolve this issue and provide clues to the evolution of TACC function.

Results and Discussion

In silico *identification of* TACC *family members from vertebrate and invertebrate lineages*

Sequence similarity searches of the publicly available genome databases with the BLAST and TBLAST programs were performed to identify *TACC* and *RHAMM* orthologues, and other members of the coiled coil superfamily in a diverse set of species (Fig. 1). This identified the complete sequence of the *TACC* genes in representatives of five major phylogenetically distinct clades. Where possible, the construction of the TACC sequences from these organisms was also confirmed by the analysis of the cDNA databases. Several partial sequences in other vertebrate species, the echinodermate *Strongylocentrotus purpuratus* and the protostome insect *Anopheles gambiae* were also identified, suggesting an ancient conservation of the *TACC* genes in metazoan lineages. However, due to the relative infancy of the cDNA/genome projects for these latter organisms, complete characterization of these *TACC* genes could not be undertaken. No conclusion could be made about the existence of *TACC*-like sequence in non-bilaterian metazoans, such as *Cnidaria* or *Porifera*, due to the paucity of sequence information for these organisms, and additional definitive sequences with a defined TACC domain could not be found in other non-metazoan organisms.

At the base of the chordate branch of life, a single *TACC* gene was identified in the genome of the urochordate *Ciona intestinalis* [11], and a partial TACC sequence from an analysis of the *Halocynthia rortezi* EST database [12]. This confirms the original assumption that a single *TACC* gene was present in the chordate ancestor. The next major event in the evolution of the chordate genome has been suggested to have occurred 687 ± 155.7 million years ago (MYA), with the first duplication of the chordate genome, and a second duplication occurring shortly thereafter. Thus, if the *TACC* genes were duplicated at both events, we would expect to identify four *TACC* genes in the most "primitive" compact vertebrate genome sequenced to date, the pufferfish *Takifugu rubripes*, with three genes corresponding to the human *TACC1-3*, and, in keeping with the proposed model for genomic duplication of the chromosomal loci for the *TACC* genes (discussed below), a possible fourth gene deriving from the *TACC3* ancestor. Indeed, four *TACC* genes were identified in *T. rubripes*. Of these, two genes corresponded to the *T. rubripes* orthologues of human *TACC2* and *TACC3*. However, the other two genes, *trTACC1A* and *trTACC1B* are clearly most related to *TACC1* (Fig. 1). Although *trTACC1A* is highly homologous to *trTACC1B*, the latter encodes a significantly smaller predicted protein. The *trTACC1B* gene is encoded by 15 exons over approximately 7 kb of the *Takifugu* Scaffold 191 (see below). A search of this region using the *trTACC1A s*equence and gene prediction software has so far failed to identify additional exons of *trTACC1B*. However, given the intron/exon structure of this apparently complete gene, it appears likely that *trTACC1B* is active in the pufferfish, and presumably fulfils either a temporal-spatial specific function within the organism, or a distinct function from the larger *trTACC1A*

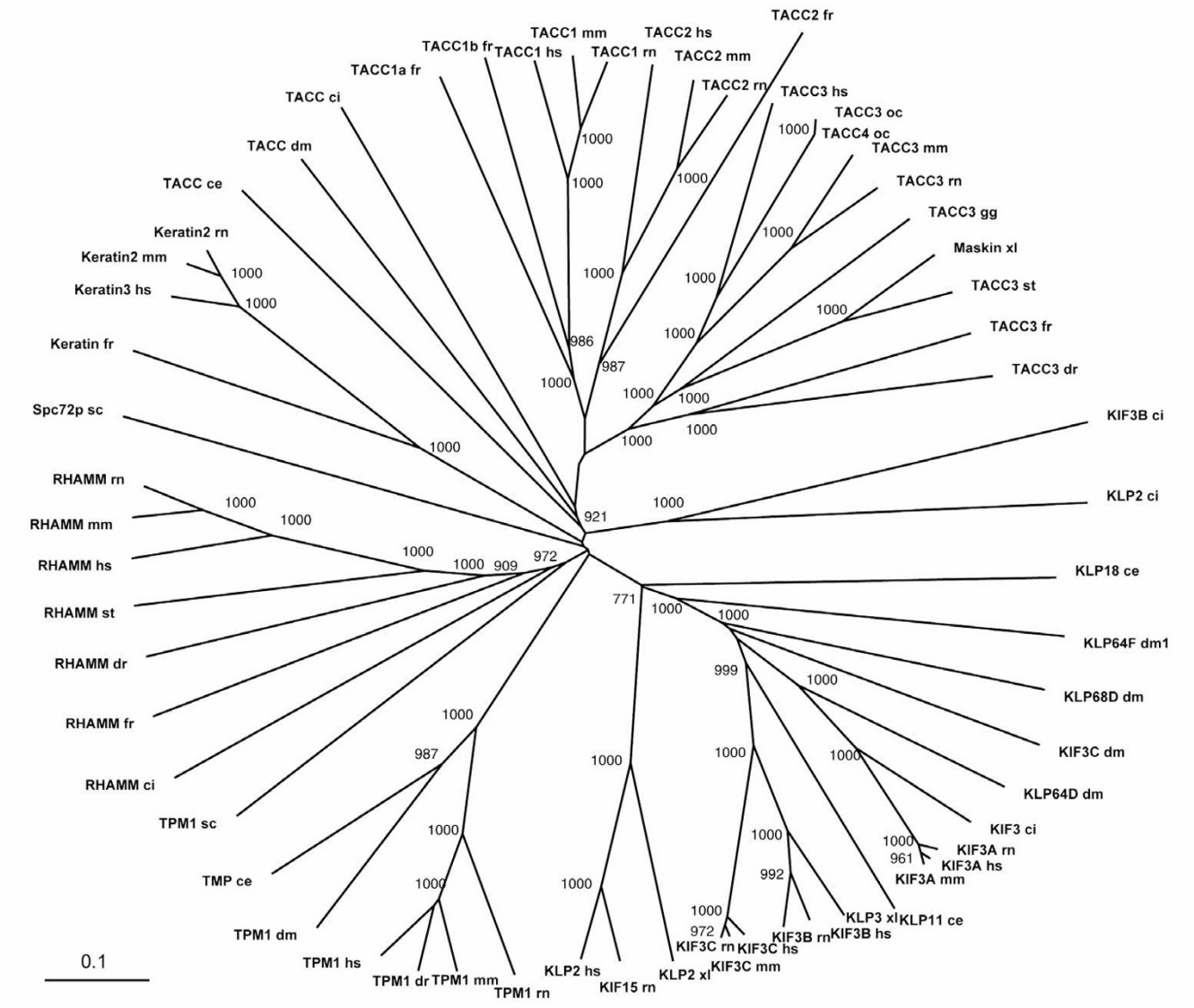

Figure 1

Phylogenetic analysis of the TACC family members compared to other coiled coil proteins. The phylogenetic tree was constructed as described in the Methods section. The TACC family defines a separate subfamily of coiled coil containing proteins, distinct from other coiled coil families such as the keratins, RHAMM and tropomyosins. Note that the RHAMM proteins form a separate branch more closely related to the tropomyosins and kinesin like proteins (KLP), than the TACC proteins.

product within the cell. Thus, based upon the surrounding chromosomal loci (see below), the *trTACC1A* and *trTACC1B* genes appear to have arisen from the duplication of the chromosomal segment containing the teleost TACC1 ancestor, during the additional partial genomic duplication that occurred in the teleost lineage. Therefore, this analysis of *T. rubripes* does not support the hypothesis that the region surrounding the *TACC3* ancestor was included in the second round of vertebrate genomic duplication.

Examination of higher vertebrates led to the identification of splice variants of *TACC1* and *TACC2* in *Mus musculus*, and the assembly of the previously unidentified orthologues of *TACC1-3* from *Rattus norvegus*. In addition, the *TACC1X* sequence was found on mouse chromosome X.

This gene is clearly related to the mouse *TACC1*, however, further examination revealed a mouse B1 repeat distributed over the length of the proposed intron. In addition, no expression of TACC1X was detected in mouse RNA by rt-PCR analysis (*data not shown*), suggesting that this sequence is a processed pseudogene. Similarly, *TACC1* pseudogenes also exist spread over 22 kb of the centromeric region of human chromosome 10 and, in 8q21, a shorter region 86% identical to the final 359 bp of the *TACC1* 3' untranslated region. No pseudogenes corresponding to *TACC2* or *TACC3* were identified in any mammalian species.

Characterization of vertebrate TACC3 *orthologues*

Based upon current functional analysis, the characterization of *TACC3* orthologues is likely to be pivotal to understanding the sequence and functional evolution of the *TACC* gene family. As indicated below, the chromosomal region containing the *TACC* gene precursors was duplicated twice during vertebrate evolution. Although the analysis of *T. rubripes*, rodents and humans so far suggests that the vertebrate *TACC3* precursor was not included in the second round of genomic duplication, it could not be excluded that a *TACC4* gene may have been lost during the evolution of these lineages. The cloning of a new member of the TACC family in *Oryctolagus cuniculus* has added to this controversy [9]. Designated TACC4, the 1.5 kb cDNA was highly related, but proposed to be distinct from TACC3. However, Northern blot data suggested that this gene produces a single 2.3 kb transcript [9], indicating that the cloned cDNA was incomplete. The degree of similarity to the published sequence of human and mouse TACC3 suggested to us that TACC4 actually represents a partial rabbit TACC3 cDNA. To test this hypothesis, we set out to clone the complete rabbit TACC3 sequence, based upon the known features of human and mouse TACC3. We have previously noted that the N-terminal and C-terminal regions of the human and mouse TACC3 proteins are highly conserved ([2], see below). Therefore, based upon the sequence identity between these genes, we designed a consensus oligonucleotide primer, T3con2, that would be suitable for the identification of the region containing the initiator methionine of the TACC3 cDNAs from primates and rodents. Using this primer, in combination with the TACC4-specific RACE primer (RACE2), initially used by Steadman et al [9], we isolated a 1.5 kb PCR product from rabbit brain cDNA by rt-PCR. In combination with 3'RACE, this generated a consensus cDNA of 2283 bp which corresponds to the transcript size of 2.3 kb detected by the "TACC4" sequence reported in Figure 4 of Steadman et al [9]. Thus, while it remains possible that the "TACC4" sequence is an alternative splice product, or is the product of reduplication of the *TACC3* gene (events that would be specific to the rabbit), the only transcript detected in rabbit RNA corresponds to the predicted transcript size of the TACC3 sequence that we have identified here. Furthermore, the string of nucleotides found at the 5' end of the "TACC4" sequence is also found at the 5' ends of a number of cDNA sequences (e.g. U82468, NM_023500), that were isolated by 5'RACE, suggesting that they may correspond to an artefact of the 5'RACE methodology used in their construction. The rabbit "TACC4" and the rabbit TACC3 sequence that we have isolated are also found on the same branch of the TACC phylogenetic tree with the other TACC3 orthologues, including maskin (*Xenopus laevis*), and the newly identified TACC3 sequences in *Rattus norvegus*, *Gallus gallus*, *Silurana tropicalis*, *Danio rerio* and *T. rubripes*, reported in this manuscript (Fig. 1). Thus, it is not in a separate branch that may be expected if the sequence was a distinct *TACC* family member.

Placement of the RHAMM *gene in the phylogeny of the coiled coil gene family*

Human *RHAMM* has also been proposed to be the missing fourth member of the *TACC* family [10]. Evidence used in support of this claim included its chromosomal location on 5q32 in humans (discussed below), its sequence similarity in its coiled coil domain to the TACC domain and the subcellular localization of the RHAMM protein in the centrosome. However, if *RHAMM* were a bona fide *TACC* family member, then we would predict its evolution would be similar to those of other *TACC* family members, and fit with the proposed evolution of the vertebrate genome. Thus, we set out to identify *RHAMM* orthologues and related genes in metazoans, so that a more complete phylogeny of the coiled coil super family could be generated. We identified a single *RHAMM* gene in all deuterostomes for which cDNA and/or genomic sequence was available, including *C. intestinalis*. No *RHAMM* gene was identified in insects or nematodes. This indicates that the *RHAMM/TACC* genes diverged after the protostome/deuterostome split 833–933 MYA, but prior to the echinodermata/urochordate divergence (>750 MYA). Significantly, sequence and phylogenetic analysis of coiled coil proteins (Fig. 1) clearly shows that RHAMM does not contain a TACC domain and instead forms a distinct family of proteins in the coiled coil superfamily, and is not a direct descendant of the ancestral *TACC* gene.

Evolution of the chromosomal segments containing the TACC *genes*

The phylogenetic tree of the *FGFR* genes closely resembles that of the vertebrate *TACC1-3* genes. Recently, detailed analyses of the chromosomal regions containing the *FGFR* gene family in humans, mouse and the arthopod *D. melanogaster* have revealed the conservation of paralogous chromosomal segments between these organisms (Fig. 2, [13], Table 1 [see Additional file 1]). This has provided further support that an ancient chromosomal segment

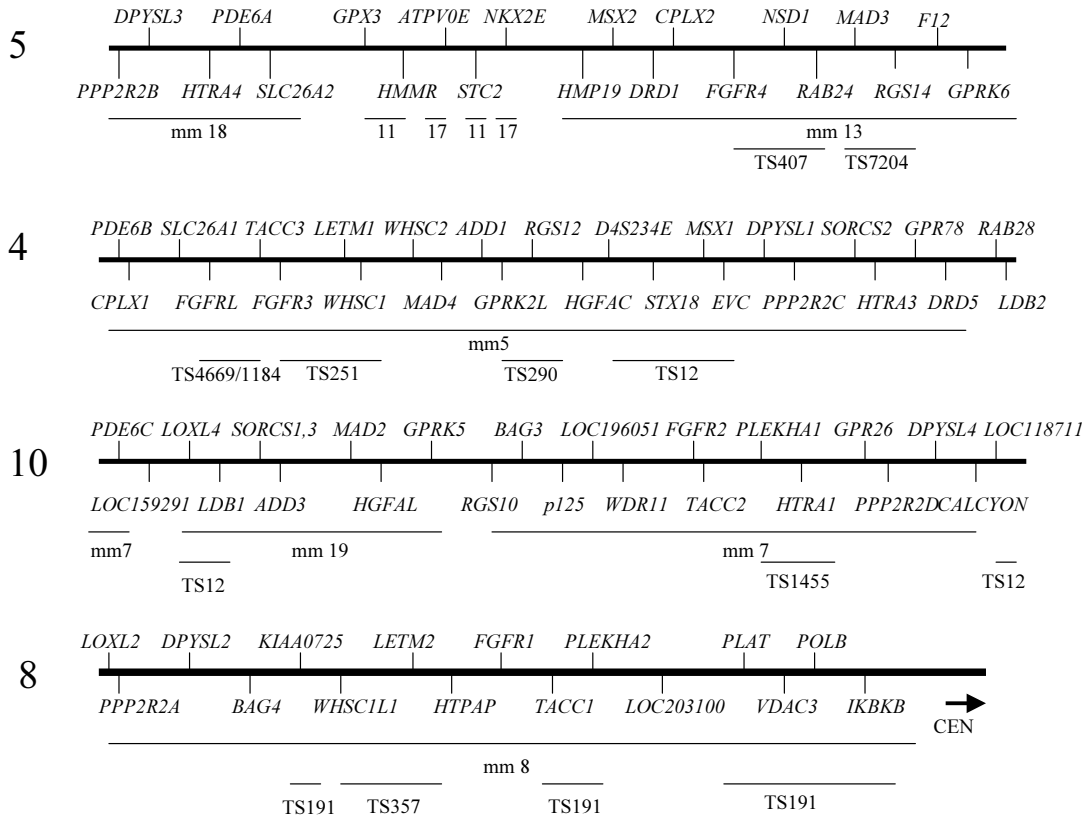

Figure 2

Linear organization of gene clusters centering upon the chromosomal loci of the *FGFR* genes in humans. Paralogous genes present in at least two of the four loci are shown, with the exception of the region between *GPX3* and *NKX2E* on chromosome 5, which appears to represent a series of intervening genes inserted after duplication of the 4p16/5q32-35 clusters, and genes mentioned in Fig. 3. Corresponding syntenic mouse chromosomal regions (mm*) are indicated. *Takifugu rubripes* scaffolds are shown (TR*) that contain more than one homologous gene from these clusters. Further details on the location of paralogous genes can be found in [see Additional file 1].

was duplicated twice during vertebrate evolution, with the first duplication that gave rise to the human chromosome 4p16/5q32-ter and human chromosome 8p/10q23-ter ancestors occurring in the early stages after the invertebrate divergence. This suggests that the ancestral *FGFR-TACC* gene pair most probably arose prior to the initial duplication and subsequent divergence of these paralogous chromosomal segments, estimated to have occurred 687 ± 155.7 MYA. This has raised the suggestion that a fourth *TACC* gene in vertebrates would reside in the same chromosomal region as *FGFR4*. Indeed this hypothesis has been used in support for the *RHAMM* gene as a member of the *TACC* family [10]. Human *RHAMM* maps to chromosome 5q32 in a region bounded by *GPX3* and *NKX2E*. These loci separate two clusters of genes on human chromosome 5 that are paralogous with 4p16. Interestingly, these three clusters are located on different chromosomes in mouse and rat (Fig. 2), further suggesting that this cluster of genes was transposed into this region after the primate/rodent divergence.

Because the conservation of gene order can also provide clues to the evolution of gene regulation, we next attempted to trace the evolution of these paralogous segments by examining the genome of the tunicate *C. intestinalis* [11] and the most "primitive" compact vertebrate genome sequenced to date, *T. rubripes* [14]. Although not fully assembled, examination of the genome of *T. rubripes* confirmed the presence of chromosomal segments paralogous to those found in higher vertebrates (Fig. 2). For instance, the orthologues of *GPRK2L* and *RGS12* are found on *T. rubripes* scaffold 290 (emb|CAAB01000290.1), and within 300 kb of each other in human 4p16. The *T. rubripes* orthologues of *FGFR3*, *LETM1* and *WHSC1* are located on the same 166 kb genomic scaffold 251 (emb|CAAB01000166.1). Significantly, the three human orthologues of these genes are also located within 300 kb of each other on 4p16. Furthermore, *TACC3* and *FGFRL* map to the overlapping scaffolds 1184/4669 (emb|CAAB01004668). Similarly, elements of these gene clusters, extending from *HMP19* to *GPRK6* in human chromosome 5q34-ter are also found in the pufferfish, with the *T. rubripes* orthologues of *NSD1*, *FGFR4* and a *RAB-like* gene mapping on scaffold 407 (emb|CAAB01000407). However, there is no evidence for a gene corresponding to a *TACC4* gene in any of these clusters.

As noted above, phylogenetic analysis of the TACC sequences indicate that there are two *TACC1* related genes in the pufferfish. *trTACC1B* is located on the 180 kb scaffold 191 (emb|CAAB01000191.1), which also contains the orthologues of several genes located in human chromosome 8p21-11. Thus, this scaffold represents the more "developed" *TACC1* chromosomal segment that is evident in higher vertebrates. On the other hand, the *trTACC1A* gene is located in the 396 kb scaffold 12 (emb|CAAB010012.1). This scaffold also contains the *T. rubripes* orthologues of *MSX1*, *STX18*, *D4S234E* and the predicted gene *LOC118711*, in addition to sequences with homology to *LOXL*, *EVC*, *LOC159291*, and the *LDB* family. Thus, scaffold 12 contains genes found in the regions of human chromosome 4 and 10 that also contain the loci for *TACC3* and *TACC2*, respectively, and may therefore more closely resemble the genomic organization resulting from the initial duplication of the ancestral paralogous chromosomal segment.

Conserved paralogous clusters may result from the initial clustering of the genes in a relatively small ancestral genomic contig. Some evidence for the existence of "protoclusters" that could correspond to the paralogous chromosomal segments noted in higher vertebrates is present in the genome of the urochordate *C. intestinalis* [11]. For instance, the orthologues of *FGFR*, and *WHSC1*, carboxypeptidase Z and *FLJ25359* cluster within an 85 kb region of the *C. intestinalis* genome and the human orthologues are still maintained in paralogous segments of 4p16, 8p and 10q (Fig. 3, [see Additional file 1]). However, it should be noted that no clusters of genes from the vertebrate paralogous segments are locate close to the *TACC* or *RHAMM* genes of *C. intestinalis*, indicating that the formation of the much larger paralogous segments encompassing the *FGFR-TACC* genes formed later in evolutionary time, or conversely have been subject to extensive rearrangement in tunicates. In combination with the examination of the *T. rubripes* genome, this also provides additional evidence that either the second round of duplication of the chromosomal segment that contained the *FGFR3/4* ancestor did not include a *TACC* gene, or that such a gene was lost very early in vertebrate evolution, prior to the divergence of the Gnanthostome lineages. However, the final resolution of the initial evolution of these paralogous segment will await the sequencing of the amphioxus and lamprey genomes, which only have one *FGFR* gene, and therefore should only contain one copy of the other corresponding genes in this conserved segment.

Comparative genomic structure of the TACC *family*

The genomic DNA sequences corresponding to the orthologous *TACC* genes of human, mouse, rat, pufferfish, *C. intestinalis*, *D. melanogaster* and *C. elegans* were extracted and analyzed by Genescan and BLAST to determine the genomic structure of each *TACC* gene. In some cases, for rat and pufferfish, exons were added or modified based on the best similarity of translated peptides to the corresponding mouse and human proteins. For regions with low sequence similarity in *T. rubripes*, genomic sequences http://www.genoscope.cns.fr/ from the fresh water pufferfish, *Tetraodon nigroviridis* were used as additional means to verify the predicted exons.

The general structure of the *TACC* genes and proteins is depicted in Fig. 4. The main conserved feature of the TACC family, the TACC domain, is located at the carboxy terminus of the protein. In the case of the *C. elegans* TAC protein, this structure comprises the majority of the protein and is encoded by two of the three exons of the gene. In the higher organisms, *D. melanogaster*, and the deuterostomes *C. intestinalis* to human, this feature is also encoded by the final exons of the gene (five in *D. melanogaster*, seven in the deuterostome genes). Outside of the TACC domain, however, TACC family members show relatively little homology. It is interesting that each *TACC* gene contains one large exon, which shows considerable variability between *TACC* orthologues, and constitutes the main difference between the *TACC3* genes in the vertebrates (see below). In deuterostomes, this exon contains the SDP repeat (or in the case of the murine TACC3's, a rodent-specific 24 amino acid repeat), which is

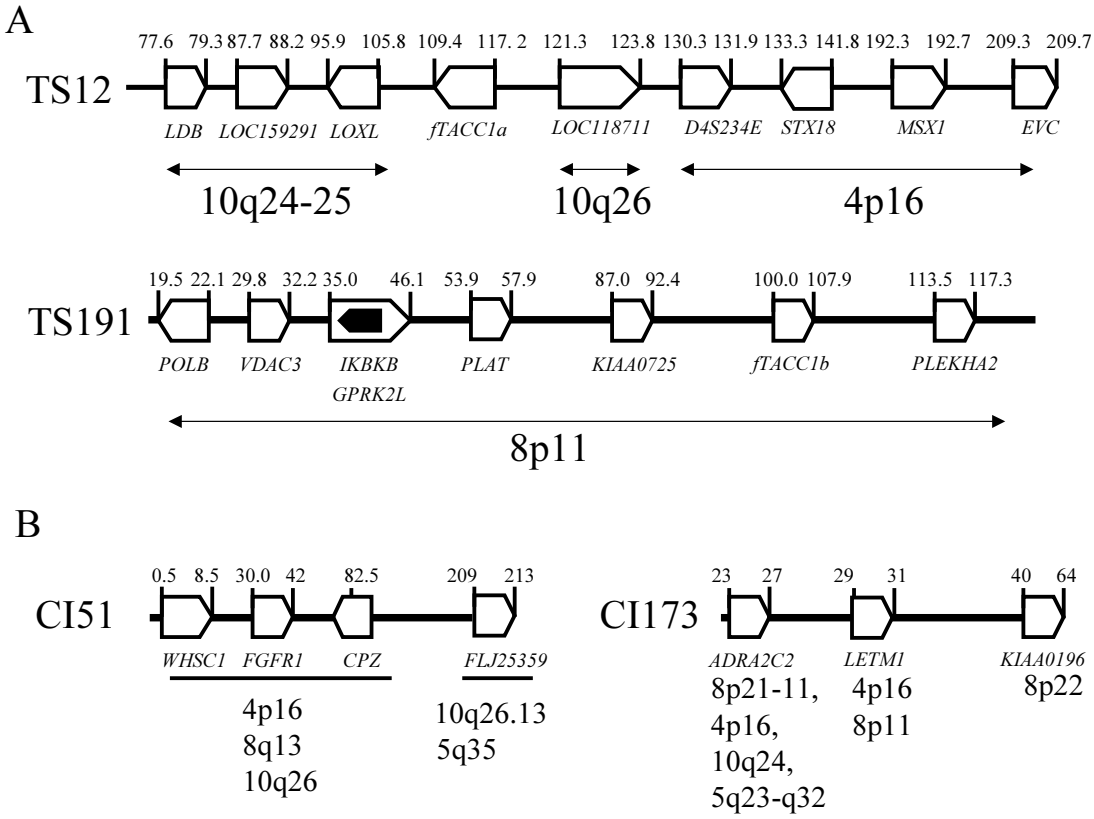

Figure 3

Formation of protoclusters in *Takifugu rubripes* and *Ciona intestinalis*: (A): Structure of the genomic scaffolds containing the *Takifugu rubripes trTACC1A* and *trTACC1B* genes. Scaffold 12, the site for the *trTACC1A* gene contains genes found with either homologues or orthologues on the distal long arm of human chromosome 10 and 4p16. This scaffold, therefore, has some of the characteristics of the predicted immediate ancestor of the *TACC1/TACC2* chromosomal segment. *trTACC1B* is found on scaffold 191, which contains orthologues of genes found in the proximal short arm of human chromosome 8. **(B):** *Ciona intestinalis* clusters containing genes found in paralogous segments on human 8, 4p16, 10q and 5q.

responsible for the binding of the SWI/SNF chromatin remodeling complex component GAS41 [15,16].

Of the vertebrate TACC proteins, the TACC3 orthologues show the greatest variability in size and sequence, ranging in size from 599 amino acids for the rat TACC3 protein, to 942 amino acids in the *Danio rerio* protein. The reasons for these differences are apparent from the genomic structure of the *TACC3* orthologues. TACC3 can be divided into three sections: a conserved N-terminal region (CNTR) of 108 amino acids, encoded by exons 2 and 3 in each vertebrate *TACC3* gene, the conserved TACC domain distributed over the final seven exons, and a highly variable central region. The lack of conservation in both size and sequence of the central portion of the TACC3 proteins of human and mouse has been previously noted, and accounts for the major difference between these two orthologues [2]. The majority of this central portion, which contains the SDP repeat motifs, is encoded by one exon in human and the pufferfish (emb|CAAB01001184). In rodents, however, this region is almost entirely composed of seven 24 amino acid

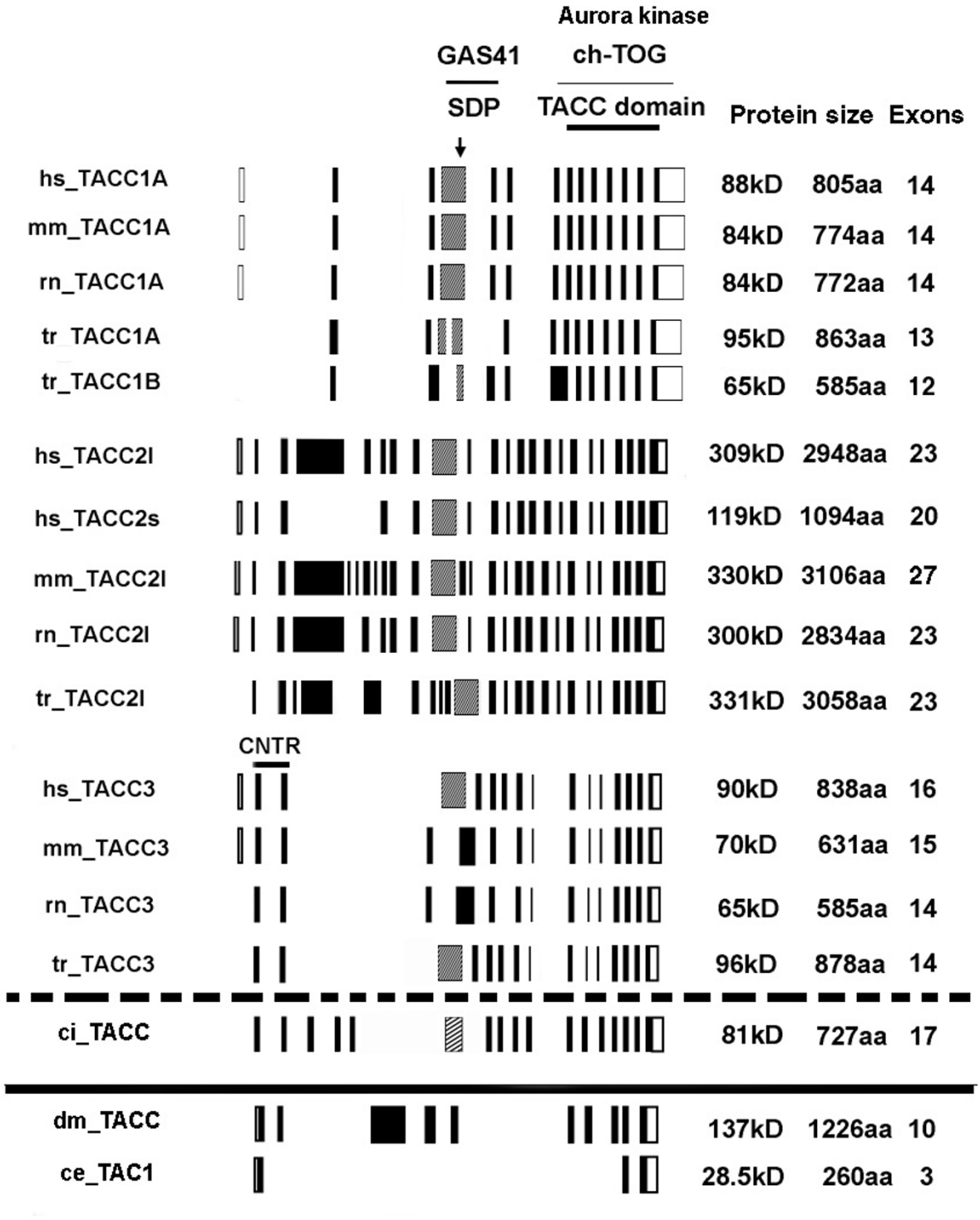

Figure 4
Genomic structure of the *TACC* genes. Conserved regions known to bind protein factors in all human TACC proteins are shown. The SDP repeat of human TACC1-3 is known to bind GAS41 ([3,15] and *data not shown*). The TACC domain binds ch-TOG and members of the Aurora kinases in all species examined to date. This motif is characteristically encoded by the 3' exons of the *TACC* genes. The size of the largest isoform for each gene is shown.

repeats, which are located in a single exon of the mouse and rat *TACC3* genes. It has been previously reported that there are four mouse TACC3 splice variants that differ in the number of these repeats [2,7,17]. As these repeats are present in a single exon, it appears likely that these different sequences may be the result of the DNA polymerases used in the cDNA synthesis and/or PCR reaction stuttering through the repeat motif. The correct sequence, reported by Sadek et al [7], is the one used throughout the entirety of this manuscript. These repeats are not evident in the rabbit protein, or any other TACC protein, and may indicate that the rodent TACC3 has evolved distinct functions, as has already been noted for the amphibian *Xenopus* TACC3, maskin [8].

Alternative splicing in vertebrate TACC *genes*

Whereas exon shuffling can drive the functional diversification of gene families over evolutionary time, the temporal and/or tissue specific alternative splicing of a gene can give rise to functional diversification of a single gene during the development of an organism. Although no alternative splicing of *TACC3* has been clearly documented, both temporal and tissue specific splicing is observed in the *TACC1* and *TACC2* genes. In the case of *TACC2*, an additional large (5 kb) exon accounts for the main difference between the major splice variants of the vertebrate *TACC2* genes [3]. The alternative splicing of this exon suggests a major functional difference between the two TACC2 isoforms, TACC2s and TACC2l [3], as well as a significant difference between TACC2 and its closest TACC family relative, TACC1. However, the function of this region of the TACC2l isoform is current unknown.

Alternative splicing, together with differential promoter usage has already been noted for the human *TACC1* gene [18,19]. In addition, as shown in Fig. 5, we have identified additional TACC1 isoforms that result from alternative splicing of exons 1b-4a. The functions of these different isoforms are unknown, however the region deleted from the shorter variants can include the binding site for LSm7 [20] (variants C, D, F-I), and/or the nuclear localization signals and binding site for GAS41 [15] and PCTAIRE2BP [20] (isoforms B-D, S). One of these isoforms, TACC1S is localized exclusively to the cytoplasm [19], suggesting that the shorter isoforms would not be able to interact with elements of the chomatin remodeling and/or RNA processing machinery in the nucleus. Thus, changes in the complement of TACC1 isoforms in the cell may result in alterations in cellular RNA metabolism at multiple levels, and may account for the observation that TACC1D and TACC1F isoforms are associated with tumorigenic changes in gastric mucosa [18].

In silico *modeling of the evolution of TACC protein function*

The protein and genomic structure of the present day TACC family members suggests that the function of the ancestral TACC protein was mediated solely through the interactions of the conserved TACC domain. Using an *in silico* protein-protein interaction model based upon known mitotic spindle and centrosomal components, we have previously predicted a number of additional interactions that could be conserved between a functional TACC homologue in yeast, spc-72, and one or more human TACC proteins [21]. Thus, it is known that all the TACC proteins examined to date interact, via the TACC domain, with the microtubule/centrosomal proteins of the stu2/msps/ch-TOG family [5,6,22-24], and with the Aurora kinases [20,21,25]. These interactions are required for the accumulation of the D-TACC, spc72, ceTAC1 and TACC3 proteins to the centrosome [5,6,22-24]. Hence, this functional interaction with the centrosome and mitotic spindle is likely to represent the ancient, conserved function of the TACC family. However, it is apparent that the human TACC proteins also differ in their ability to interact with the Aurora kinases. For instance, TACC1 and TACC3 interact with Aurora A kinase, whereas TACC2 interacts with Aurora C kinase [21], suggesting a degree of functional specialization in the derivatives of the ancestral chordate TACC, after the radiation of the vertebrate *TACC* genes.

The localization of the vertebrate TACC proteins in the interphase nucleus [15,26,27] suggests that they have additional functions outside their ancient role in the centrosome and microtubule dynamics. Thus, it seems likely that TACC family members in protostomes and deuterostomes have integrated new unique functions as the evolving *TACC* genes acquired additional exons. The results of the pilot large-scale proteomic analysis in *C. elegans* and *D. melanogaster* provide further suggestive evidence to this functional evolution. Yeast two hybrid analysis indicates that ceTAC directly binds to *C. elegans* lin15A, lin36 and lin37 [28]. These proteins bridge ceTAC to other elements of the cytoskeleton and microtubule network, as well as to components of the ribosome, the histone deacetylase chromatin remodeling machinery such as egr-1 and lin-53 (the *C. elegans* homologues of the human MTA-1 and RbAP48), and to transcription factors such as the PAL1 homeobox and the nuclear hormone receptor nhr-86 [28] (Fig. 6A). Similarly, large scale proteomics [29] has shown that *Drosophila* TACC interacts with two proteins, the RNA binding protein TBPH and CG14540 (Fig. 6B), and thus indirectly with the *Drosophila* SWI/SNF chromatin remodeling complex and DNA damage repair machinery. Significantly, the ceTAC protein has also recently been implicated in DNA repair through its direct interaction with the *C. elegans* BARD1 orthologue [30]. It should be noted that a number of interactions with the TACC

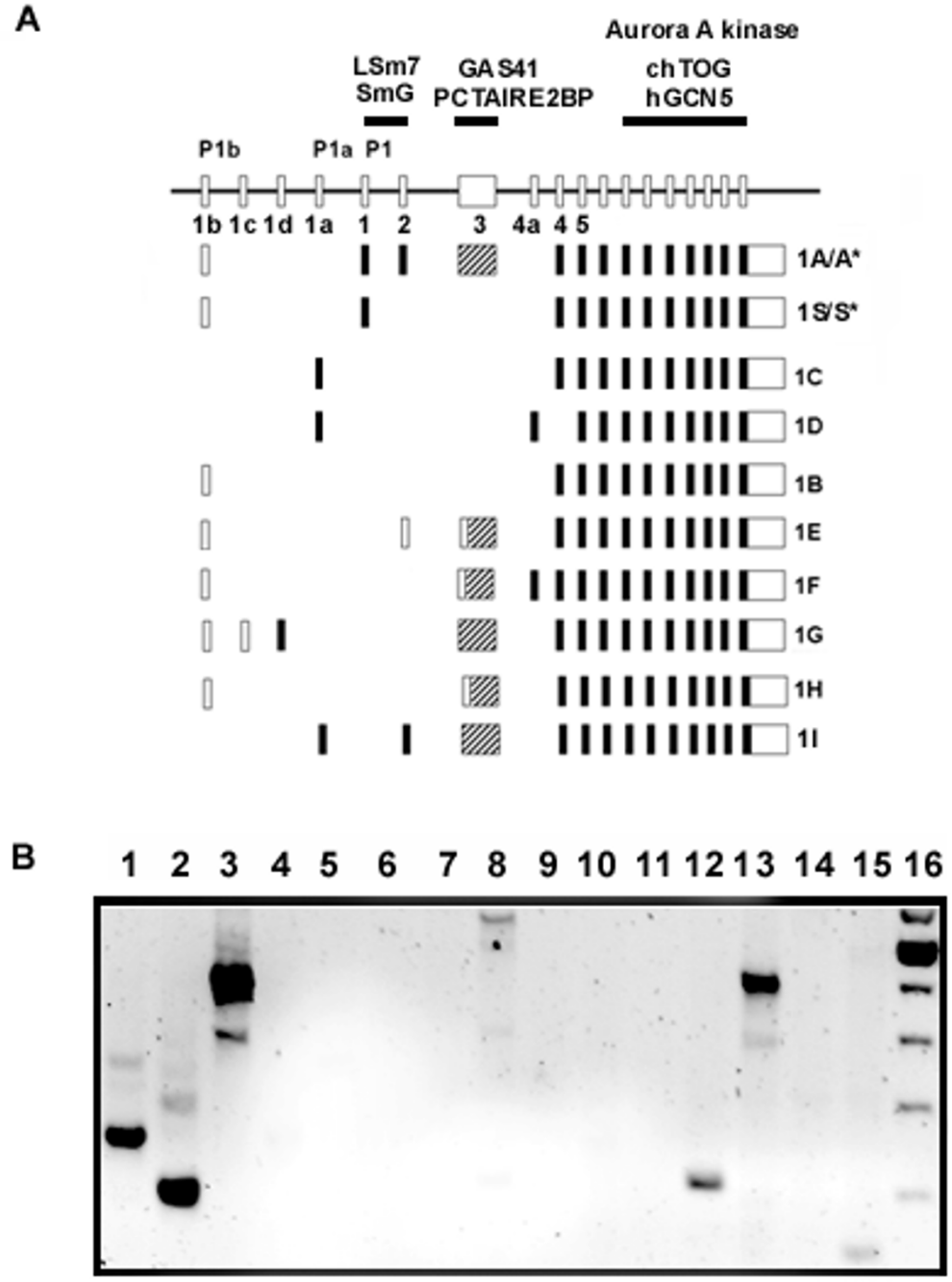

Figure 5

Alternative splicing of the human *TACC1* gene. (A): Seven splice variants have been identified for human *TACC1* (A-F and S). We have also identified additional splice variants (G-I) from database analysis and rt-PCR analysis of human brain RNA. **(B)**: Alternative splicing of *TACC1* in the human brain. rt-PCR analysis confirms splicing of the untranslated exon 1a to exon 1, with retention of the originally defined start methionine (GB:NP_006274) (Variant A*, lanes 1 and 3). Exon 1a also splices to exon 2, removing the L-Sm7 binding motif (variant I, lanes 3,12 and 13). Variants that functionally delete exons 2 and/or 3, such as variant C (lane 15) also remove the predicted nuclear localization signals, and the binding domains for GAS41 and PCTAIRE2BP. These variants would retain the TACC domain, and therefore the potential to bind to ch-TOG and Aurora A kinase in the centrosome. Lane 1: EF/X1R, Lane 2: EF/BX647R, Lane 3: EF/6SPR, Lane 4: EF/1DR, Lane 5, EF/128R, Lane 6 X3F/X1R, Lane 7: X3F/BX647R, Lane 8: X3F/6SPR, Lane 9: X3F/1DR, Lane 10, X3F/128R, Lane 11: 1DF/X1R, Lane 12: 1DF/BX647R, Lane 13: 1DF/6SPR, Lane 14: 1DF/1DR, Lane 15: 1DF/128R, Lane 16: Biorad 1 kb+ Size ladder.

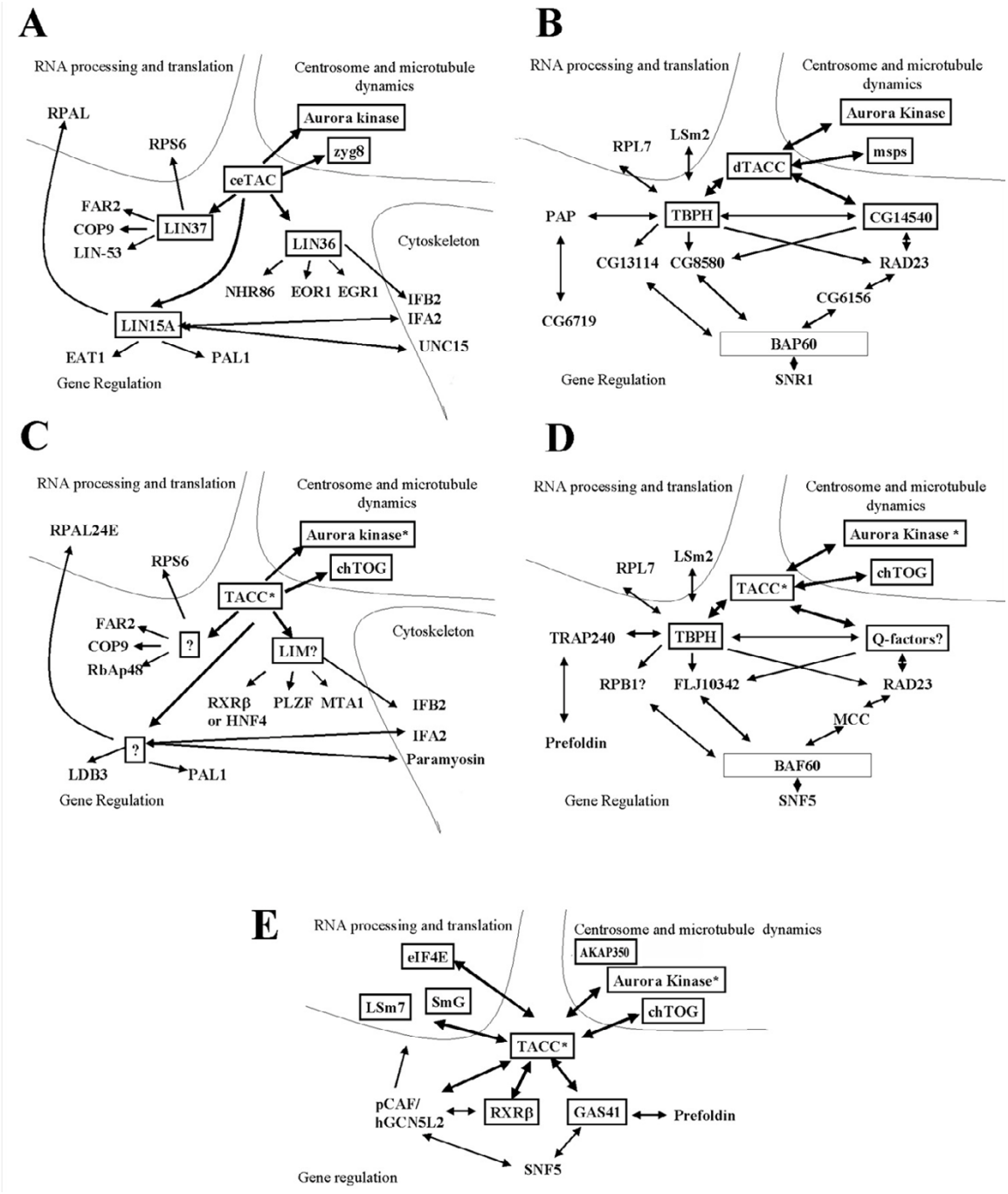

Figure 6

Functional evolution of the TACC proteins modeled in *C. elegans* and *D. melanogaster*. (**A).** *C. elegans* interaction map shows empirically defined interactions of ceTAC, and extrapolated interactions defined by [28]. (**B):** Using the BIND database [29], DTACC directly binds to TBPH and CG14540, and thus indirectly to chromatin remodeling complexes (SWI/SNF and histone acetyltransferases), DNA damage repair machinery (via RAD23), and RNA splicing, transport and translational machinery. **(C):** Predicted interaction map for vertebrate TACCs, based upon ceTAC, suggests an indirect interaction with the nuclear hormone receptor RXRβ. It is also of interest that this predicts a functional interaction with the LDB family, members of which are also found in TACC containing paralogous segments noted in Figs 2, 3 and Additional file 1. **(D):** Predicted TACC interaction map based upon DTACC. **(E):** Vertebrate TACC interactions identified to date. ? denotes uncertainty over the identity of a functional vertebrate homologue. In C, D and E, '*' denotes one or more members of the TACC or Aurora kinase family.

proteins from these organisms have probably been missed by these large scale methods, including the well documented direct interactions with the aurora kinases and the stu2/msps/ch-TOG family.

Because of the evolutionary conservation of the TACC domain, we would predict that some of the functional interactions seen in *C. elegans* and *D. melanogaster* would be observed in higher animals. Phylogenetic profiling from these interaction maps suggests two similar sets of predicted interactions for vertebrate TACCs (Fig. 6C and 6D). Strikingly, however, the *C. elegans* specific proteins lin15A, lin36 and lin37 do not have readily discernible homologues in vertebrates or *Drosophila*, although the presence of a zinc finger domain in lin36 may suggest that this protein is involved directly in transcription or perform an adaptor role similar to LIM containing proteins. For the DTACC interacting proteins, TBPH corresponds to TDP43, a protein implicated in transcriptional regulation and splicing [31,32]. However, the assignment of the human homologue of CG14540 is less clear, with the closest matches in the human databases corresponding to glutamine rich transcription factors such as CREB and the G-box binding factor.

Comparison of modeled with experimentally defined interactions of the vertebrate TACC proteins

The interaction data for the vertebrate TACCs is relatively limited; however, interaction networks are now beginning to emerge. The results of our functional analysis, as well as other published data clearly indicate that the vertebrate TACCs interact with proteins that can be divided into two broad categories: 1) proteins with roles in centrosome/mitotic spindle dynamics, and 2) proteins involved in gene regulation, either at the level of transcription, or subsequent RNA processing and translation [3,5-7,15,19-21,24,25,33,34]. Many of these proteins do not appear to interact directly with the protostome TACCs, but would be expected to be in the same protein complex (Fig. 6C,6D).

Significant analysis of the association of the TACCs with the centrosome and the dynamics of mitotic spindle assembly from yeast to humans has been published [5,6,21-24]. From this analysis, it seems likely that the vertebrate TACC3 protein has retained this direct ancestral function, based upon its location in these structures during mitosis [27], its strong interaction with Aurora Kinase A, and the observation that it is the only human TACC protein phosphorylated by this enzyme [21]. However, the variability of the central domain of the vertebrate orthologues, suggests that TACC3 may also have acquired additional, and in some instances, species-specific functions. For instance, in *X. laevis*, the maskin protein has acquired a binding site for the eIF4E protein, and thus a function in the coordinated control of polyadenylation and translation in the *Xenopus* oocyte [8,35]. A recent study has suggested that this function may be unique to maskin: although it is unclear whether the other vertebrate TACC3 proteins interact with the eIF4E/CPEB complex, the human TACC1A isoform is unable to interact with the eIF4E/CPEB complex. Instead, some TACC1 isoforms have evolved a related, but distinct function by directly interacting with elements of the RNA splicing and transport machinery [19].

To further characterize the evolving functions of the TACC proteins, we have used an unbiased yeast two hybrid screening method to identify proteins that bind to the human TACC proteins [3,34]. In a screen of a MATCHMAKER fetal brain library (BD Biosciences Clontech), in addition to isolating the histone acetyltransferase hGCN5L2 [34], we also identified the β3 isoform of retinoid-X receptor β as a protein that interacts with the TACC domain of TACC2. As shown in Fig. 7, this interaction is confirmed *in vitro* by GST-pull down analysis. Significantly, RXRβ is a close family relative of the nuclear hormone receptor, nhr-86, from *C. elegans*, which interacts with the ceTAC binding protein lin36 (Fig. 6A). This suggests that while protostome TACCs may require additional protein factors to interact with such components, the TACCs in higher organisms may have evolved the ability to directly interact with some of the proteins in the predicted interaction map (Fig. 6E). Indeed, this appears to be directly linked to the acquisition of new domains and duplication of the chordate *TACC* precursor. In fact, the first identified function of a vertebrate TACC protein was as a transcriptional coactivator acting through a direct interaction with the ARNT transcription factor [7]. It is also intriguing that the deuterostome specific SDP repeat interacts with GAS41, a component/accessory factor of the human SWI/SNF chromatin remodeling complex [3,15]. Although there is a *D. melanogaster* homologue of GAS41, dmGAS41, the large scale proteomic interaction database does not indicate a direct interaction of dmGAS41 with DTACC. This may be due to the lack of the SDP repeat region in the *Drosophila* TACC protein. This further suggests that the vertebrate TACCs have gained the specific ability to direct interact with transcriptional regulatory complexes, and that bridging protein(s) are no longer required. Thus, where the ceTAC protein is only composed of the TACC domain, the significantly larger TACC family members in higher protostomes and deuterostomes may have integrated one or more functions of the bridging protein (in this case lin15A, lin36 or lin37). This may also explain the absence of lin15A, lin36 and lin37 homologues in higher organisms, as they were no longer under selective evolutionary pressure to remain within the complex, and thus lost in the evolving genome.

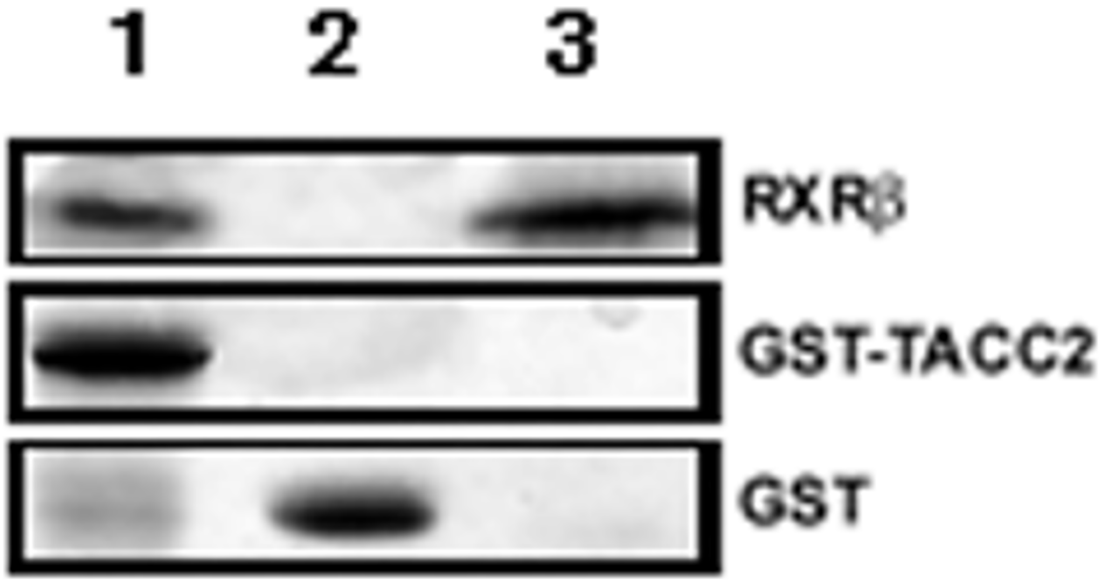

Figure 7
***In vitro* interaction of RXRβ3 and TACC2s.** Top panel: Autoradiograph of 12% SDS polyacrylamide gel with *in vitro* translated RXRβ3 construct pulled down with GST-TACC2 (Lane 1) or GST (Lane 2); Lane 3: 5% input of *in vitro* translated RXR-β protein. Bottom two panels represent Coomassie blue stained gels of pull down experiment showing loading of GST-TACC2 and GST.

Conclusion

Proposed functional evolution of the TACC family

Examination of the evolution of ancient gene families provides an insight into how gene structure relates to function. We have presented above, a detailed examination of one such gene family. The data so far suggest that the functional TACC homologue in yeast (spc72) has a specific role in centrosomal/mitotic spindle dynamics [21,22]. This ancient TACC function is conserved throughout evolution in both protostomes and deuterostomes. In addition, the TACC proteins of lower organisms appear to interact with bridging proteins that are components of several different protein complexes involved in DNA damage repair, protein translation, RNA processing and transcription. However, over the process of evolutionary time, with the acquisition of new domains and duplication of the chordate TACC precursor, the chordate TACC proteins have acquired the ability to directly interact with some of the other components of these complexes (such as the LSm proteins, nuclear hormone receptors, GAS41, accessory proteins and transcription factors), and thus evolved additional functions within these complexes. Indeed, the first assigned function of a vertebrate TACC protein, mouse TACC3, was as a transcriptional coactivator of the ARNT mediated transcriptional response to hypoxia and polyaromatic hydrocarbons [7]. Mouse TACC3 has also been reported to interact with the transcription factor STAT5 [33]. Recently, we have demonstrated that TACC2 and TACC3 can bind to nuclear histone acetyltransferases [34], further confirming a more direct role for the TACC proteins in transcriptional and chromatin remodeling events. Interestingly although all human TACC proteins can directly interact with the histone acetyltransferase pCAF *in vitro*, the TACC1 isoforms expressed in human breast cancer cells do not interact with this histone acetylase [34]. This may be attributable to the proposed function of the Exon 1 containing TACC1 variants in RNA processing, via the interaction with LSm-7 and SmG [19]. Thus, alternative splicing of the *TACC1* gene adds further diversity to TACC1 function, as the deletion of specific exons and their associated binding domains will change the potential protein complexes with which they can associate, either directly, or by redirecting the splice variants to different subcellular compartments. With the duplication of the *TACC1/TACC2* ancestor, it is apparent that an even greater functional diversity may have been introduced into the TACC family. The TACC2 protein retains the ability of TACC3 to interact with GAS41, INI1, histone acetyltransferases and transcription factors (in this case, RXRβ) (Fig. 7) [3,34]. However, the tissue specific splicing of the 5 kb exon in the TACC2l isoform [3] indicates that this

protein has several temporal and tissue specific functions yet to be identified.

Methods

Compilation and assembly of previously uncharacterized TACC cDNAs and genes

Corresponding orthologous sequences for TACC, RHAMM, KLP, KIF, TPM and keratins families were identified initially using the TBLASTN program [36] to search the published genomic and cDNA databases. For *Takifugu rubripes*, gene predictions were produced by the Ensembl automated pipeline http://www.ensembl.org/Fugu_rubripes/[37] and the JGI blast server http://bahama.jgi-psf.org/fugu/bin/fugu_search. DNA sequences covering the homology regions were extracted and analyzed by Genscan to obtain potential exons. In some cases, exons were added or modified based on the best similarity of translated peptides to the corresponding mouse and human proteins. For regions with low sequence similarity, genomic sequences http://www.genoscope.cns.fr/ from the fresh water pufferfish, *Tetraodon nigroviridis* were used as additional means to verify the predicted exons. Due to the variability of the central region of vertebrate TACC3 cDNAs (see text), to further confirm prediction of the *Takifugu rubripes* TACC3, full length cDNAs corresponding to the *Danio rerio* TACC3 (IMAGE clones 2639991, 2640369 and 3724452) were also obtained from A.T.C.C. and fully sequenced. Potential paralogous chromosomal segments and scaffold were identified by searching the public databases deposited at NCBI and at the Human Genome Mapping Project, Cambridge UK.

Cloning of vertebrate TACC cDNAs

The rabbit TACC3 was cloned by rt-PCR using the "TACC4" specific primer T4RACE2 [9] (5'-cccgaactgctccaggtaatcgatctc-3') and a consensus primer, T3con2, designed to the region encompassing the vertebrate TACC3 initiator methionine (5'-tatgagtctgcaggtcttaaacgac-3'). For cloning the mouse Tacc1X cDNA, the primers used were based upon the genomic sequence reported, and the sequence of the IMAGE cDNA clone 4933429K08: T1XF (5'-ccatgttcagtcattggcaggtc-3'), T1XF2 (5'-ctgcagaaccaacagttcaag-3'), T1XR1 (5'-agatctgtgacatcacagctc-3'), T1XR2 (5'-ctcgagtcagttagtcttatccagctt-3'), BB617F (5'-accaccaacttgagtacctg-3') and BB617R (5'-gtatcttgaactgttggttctg-3'). For analysis of TACC1 splice variants, the forward primers used were located in exon 1b: EF (5'-gagagatgcgaaatcagcg-3'), Exon 1d: X3F (5'-agtcaaagaaggcatctgcag-3'), Exon 1a: 1DF (5'-ccaagttctgcgccatggg-3'). The reverse primers used were: Exon 1: X1R (5'-ggatttggtctcggcttgcgaatc-3'), Exon 2: BX647R (5'-cttgtgattcttggctttgg-3'), Exon 3: 6SPR (5'-gtcatcgcctcgtcctggagggc-3'), Exon 4a: 1DR (5'-aatttcacttgttcagtagtc-3'), Exon 5: 128R (5'-cctgcttctgaggatgaaaacgc-3'). Rabbit brain poly A+ mRNA, mouse testis and human brain total RNA were obtained from BD Bioscience Clontech (Palo Alto, CA, U.S.A.). Reverse transcription and PCR was performed as previously described, using either 1 μg of total RNA or 50 ng of poly A+ mRNA as template for first strand cDNA synthesis. PCR products were cloned into pCR2.1 (Invitrogen, Carlsbad CA, U.S.A.) and transformed into InvαF' competent cells. Plasmid inserts were sequenced by the Roswell Park Cancer Institute Biopolymer Core Facility.

Deposition of nucleotide sequences

Sequences from this article have been deposited in the GenBank database with the following accession numbers: Homo sapiens TACC1 short isoform S (AY177411), *Mus musculus* TACC1 short isoform S (AY177412), *Mus musculus* TACC1 long isoform A (AY177413), *Mus musculus* TACC2s (AY177410), *Oryctolagus cuniculus* TACC3 (AY161270), *Danio rerio* TACC3 (AY170618). Annotations submitted to the Third party annotation database at NCBI are as follows: *Rattus norvegus* TACC1 long isoform A (BK001653), *Takifugu rubripes* TACC1A (BK000666/BK000667), *Takifugu rubripes* TACC1B (BK000664), *Mus musculus* TACC2l (BK001495), *Rattus norvegus* TACC2l (BK001658), *Rattus norvegus* TACC2s (BK001657), *Takifugu rubripes* TACC2l (BK000690), *Rattus norvegus* TACC3 (BK001491), *Gallus gallus* TACC3 (BK001482), *Silurana tropicalis* TACC3 (BK001481),*Takifugu rubripes* TACC3 (BK000649), *Takifugu rubripes* RHAMM (BK000676), *Ciona intestinalis* RHAMM (BK001479), *Takifugu rubripes* Keratin (BK000677), *Takifugu rubripes* TPM1 (BAC10576), *Ciona intestinalis* Kif3b (BK001492), *Ciona intestinalis* klp2 (BK001493).

Phylogenetic analysis

In order to examine evolutionary relationships of proteins containing coiled coil domains, protein sequences representing the major members of this super family, including TACC, RHAMM, KLP, keratin and tropomyosin from available vertebrates and their recognizable orthologues from the urochordate *Ciona intestinalis*, *Drosophila melanogaster*, *C. elegans* and *Saccharomyces cerevisiae* were either directly retrieved from NCBI sequence databases, newly predicted or isolated (see above). Species abbreviations are as follows: hs (*Homo sapiens*), mm (*Mus musculus*), rn (Rattus norvegus), oc (*Oryctolagus cuniculus*), gg (*Gallus gallus*), xl (*Xenopus laevis*), st (*Silurana tropicalis*), tr (*Takifugu rubripes*), dr (*Danio rerio*), ci (*Ciona intestinalis*), dm (*D. melanogaster*), ce (*C. elegans*), sc (*Saccharomyces cerevisiae*). The sequences identified above and the following protein or predicted translations were used for phylogenetic analysis: hsTACC1A (NP_006274), hsTACC2l (AAO62630) hsTACC2s (AAO62629), hsTACC3 (NP_006333), mmTACC3 (Q9JJ11), xlMaskin (Q9PTG8), dmTACC (AAF52099), ceTAC1 (NP_497059), scSPC72 (NP_009352), hsRHAMM

(NP_036616), mmRHAMM (NP_038580), rnRHAMM (NP_037096), drRHAMM (AAQ97980), hsKeratin (CAB76828), mmKeratin (A61368), rnKeratin (XP_235679), hsTPM1 (NP_000357), mmTPM1 (NP_077745), rnTPM1 (NP_62004, drTPM1 (NP_571180) dmTPM1 (P06754), ceTPM (NP_493540) scTPM1 (P17536), hsKLP2 (BAB03309), rnKIF15 (AAP44513), xlKLP2 (CAA08879), dmKLP2 (NP_476818), ceKLP18 (AA034669), hsKIF3A (Q9Y496), mmKIF3A (NP_032469), rnKIF3A (XP_340797), xlKIF3A (CAA08879), ceKLP11 (NP_741473), ciKIF3 (ci0100148992), hsKIF3B (NP_004789), mmKIF3B (NP_004789), rnKIF3B (XP_215883), dmKIF3B (NP_524029), hsKIF3C (NP_002245), mmKIF3C (NP_032471), rnKIF3C (NP_445938), dmKIF3C (NP_651939).

These protein sequences were initially aligned with CLUSTAL X [38]. Minor adjustments to certain regions of the alignment for optimization purposes were made based on pairwise alignments, the output saved in PHYLIP format, after which, the distances between proteins were calculated using Poisson correction and the Unrooted trees were inferred with the NJ method and then displayed using TreeView [39]. Bootstrap values above 700 for 1000 trials are shown at the node. To validate the tree, the same sequence set was analyzed with tools in the PHYLIP package [40], using PRODIST followed by FITCH or NEIGHBOR and tree displaying using TreeView. This additional method produced trees with essentially the same topology (*data not shown*).

In vitro *interaction of TACC2 and RXRβ3*

The TACC2 cDNA was cloned into GST fusion vector pGEX5X2 (Amersham Biosciences, Piscataway, NJ, USA). GST and GST-TACC2 proteins were expressed in *E. coli* BL21(DE3) plys "S" with 1 mM IPTG at 37°C shaker for 2 hrs. Cells (50 ml) were harvested and resuspended in 5 ml of 20 mM Tris-HCl pH.8.0, 200 mM NaCl, 1 mM EDTA pH8.0, Protease inhibitor set III (Calbiochem). The cells were lysed by sonication and lysate cleared by centrifugation at 7500 rpm at 4°C for 15 min. The cleared lysate was immobilized on glutathione sepharose beads in 3 ml of 20 mM Tris-HCl pH.8.0, 200 mM NaCl, 1 mM EDTA pH8.0). RXRβ3 cDNA was cloned into pET 28C(+) (Invitrogen, Carlsbad, CA, USA) and protein synthesized by TNT quick coupled transcription/translation system kit (Promega) and radiolabeled with ^{35}S methionine according to manufacturer's instructions. 100 μl of *in vitro* translated RXRβ3 protein in 1 ml of 20 mM Tris-HCl pH.8.0, 200 mM NaCl, 1 mM EDTA pH8.0 was incubated at 4°C with immobilized GST-TACC2 or GST for 90 min. Unbound RXRβ3 was removed by washing three times with 20 mM Tris-HCl pH.8.0, 200 mM NaCl, 1 mM EDTA pH8.0. Bound proteins were eluted from the beads at room temperature for 10 min in elution buffer (100 mM Tris HCl, pH8.0, 20 mM reduced glutathione). The proteins were analyzed on 12% SDS polyacrylamide gels. Coomassie blue staining verified equal loading of GST fusion protein. Dried gels were autoradiographed.

Authors' contributions

I.H.S performed most of the sequence analysis and drafted the assembly of the TACC, KLP, KIF and RHAMM sequences. I.H.S. performed the cDNA isolation of the rabbit TACC3 cDNA. AK identified potential TACC1 splice variants, and the interaction between RXRβ3 and TACC2. A.M. characterized mouse TACC orthologues. P.L. performed the identification and assembly of *T. rubripes* gene sequences, and the phylogenetic analysis. I.H.S. conceived and designed the project and drafted the complete manuscript.

Additional material

Additional file 1

Location of paralogous genes found on human 4p16, 5q31-ter, 10q23-ter, 8p and chromosome 2. Positions of genes are given in Mb from the telomere of the short arm of the chromosome.Blue highlighted genes (five copies) are found in all four paralogous segments, and the partially duplicated segment on chromosome 2.Green highlighted genes are found in either all four paralogous segments, or three of the segments and the partially duplicated region on chromosome 2.Pink highlighted genes have paralogues in three of the segments, suggesting either loss of the fourth copy or exclusion of one paralogue from the second round of duplication.Black highlighted genes (two paralogues) are found in only one of the derivatives of the second round of genome duplication, suggesting complete exclusion of both copies from the second round of duplication.Purple highlighted genes (two paralogues) were duplicated at the second round of genome duplication, or from interchromosomal recombination between paralogous clusters.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-4-16-S1.doc]

Acknowledgements

We wish to thank Dr. Sei-ichi Matsui for his assistance and input in the preparation of this manuscript. This work was supported in part by developmental funds support from the Roswell Park Cancer Institute, and Core grant 2P30CA016056-27 from the National Cancer Institute.

References

1. Still IH, Hamilton M, Vince P, Wolfman A, Cowell JK: **Cloning of TACC1, an embryonically expressed, potentially transforming coiled coil containing gene, from the 8p11 breast cancer amplicon.** *Oncogene* 1999, **18:**4032-4038.
2. Still IH, Vince P, Cowell JK: **The third member of the transforming acidic coiled coil-containing gene family, TACC3, maps in 4p16, close to translocation breakpoints in multiple myeloma, and is upregulated in various cancer cell lines.** *Genomics* 1999, **58:**165-170.
3. Lauffart B, Gangisetty O, Still IH: **Molecular cloning, genomic structure and interactions of the putative breast tumor suppressor TACC2.** *Genomics* 2003, **81:**192-201.

4. Gergely F, Kidd D, Jeffers K, Wakefield JG, Raff JW: **D-TACC: a novel centrosomal protein required for normal spindle function in the early Drosophila embryo.** *EMBO J* 2000, **19:**241-252.
5. Srayko M, Quintin S, Schwager A, Hyman AA: **Caenorhabditis elegans TAC-1 and ZYG-9 form a complex that is essential for long astral and spindle microtubules.** *Curr Biol* 2003, **13:**1506-1511.
6. Le Bot N, Tsai MC, Andrews RK, Ahringer J: **TAC-1, a regulator of microtubule length in the C. elegans embryo.** *Curr Biol* 2003, **13:**1499-1505.
7. Sadek CM, Jalaguier S, Feeney EP, Aitola M, Damdimopoulos AE, Pelto-Huikko M, Gustafsson J: **Isolation and characterization of AINT: a novel ARNT interacting protein expressed during murine embryonic development.** *Mech Dev* 2000, **97:**13-26.
8. Stebbins-Boaz B, Cao Q, de Moor CH, Mendez R, Richter JD: **Maskin is a CPEB-associated factor that transiently interacts with eIF- 4E.** *Mol Cell* 1999, **4:**1017-1027.
9. Steadman BT, Schmidt PH, Shanks RA, Lapierre LA, Goldenring JR: **Transforming Acidic Coiled-coil-containing Protein 4 Interacts with Centrosomal AKAP350 and the Mitotic Spindle Apparatus.** *J Biol Chem* 2002, **277:**30165-30176.
10. Maxwell CA, Keats JJ, Crainie M, Sun X, Yen T, Shibuya E, Hendzel M, Chan G, Pilarski LM: **RHAMM is a centrosomal protein that interacts with dynein and maintains spindle pole stability.** *Mol Biol Cell* 2003, **14:**2262-2276.
11. Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, T Shin-i, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS: **The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins.** *Science* 2002, **298:**2157-2167.
12. Kawashima T, Kawashima S, Kohara Y, Kanehisa M, Makabe KW: **Update of MAGEST: Maboya Gene Expression patterns and Sequence Tags.** *Nucleic Acids Res* 2002, **30:**119-120.
13. Vienne A, Rasmussen J, Abi-Rached L, Pontarotti P, Gilles A: **Systematic Phylogenomic Evidence of en Bloc Duplication of the Ancestral 8p11.21-8p21.3 Like Region.** *Mol Biol Evol* 2003, **20:**1290-1298.
14. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes.** *Science* 2002, **297:**1301-1310.
15. Lauffart B, Howell SJ, Tasch JE, Cowell JK, Still IH: **Interaction of the transforming acidic coiled-coil 1 (TACC1) protein with ch-TOG and GAS41/NuB1 suggests multiple TACC1-containing protein complexes in human cells.** *Biochem J* 2002, **363:**195-200.
16. Debernardi S, Bassini A, Jones LK, Chaplin T, Linder B, de Bruijn DR, Meese E, Young BD: **The MLL fusion partner AF10 binds GAS41, a protein that interacts with the human SWI/SNF complex.** *Blood* 2002, **99:**275-281.
17. Hao Z, Stoler MH, Sen B, Shore A, Westbrook A, Flickinger CJ, Herr JC, Coonrod SA: **TACC3 expression and localization in the murine egg and ovary.** *Mol Reprod Dev* 2002, **63:**291-299.
18. Line A, Slucka Z, Stengrevics A, Li G, Rees RC: **Altered splicing pattern of TACC1 mRNA in gastric cancer.** *Cancer Genet Cytogenet* 2002, **139:**78-83.
19. Conte N, Charafe-Jauffret E, Delaval B, Adelaide J, Ginestier C, Geneix J, Isnardon D, Jacquemier J, Birnbaum D: **Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators.** *Oncogene* 2002, **21:**5619-5630.
20. Conte N, Delaval B, Ginestier C, Ferrand A, Isnardon D, Larroque C, Prigent C, Seraphin B, Jacquemier J, Birnbaum D: **TACC1-chTOG-Aurora A protein complex in breast cancer.** *Oncogene* 2003, **22:**8102-8116.
21. Tien AC, Lin MH, Su LJ, Hong YR, Cheng TS, Lee YC, Lin WJ, Still IH, Huang CY: **Identification of the substrates and interaction proteins of Aurora kinases from a protein-protein interaction model.** *Mol Cell Proteomics* 2004, **3:**93-104.
22. Usui T, Maekawa H, Pereira G, Schiebel E: **The XMAP215 homologue Stu2 at yeast spindle pole bodies regulates microtubule dynamics and anchorage.** *EMBO J* 2003, **22:**4779-4793.
23. Gergely F, Draviam VM, Raff JW: **The ch-TOG/XMAP215 protein is essential for spindle pole organization in human somatic cells.** *Genes Dev* 2003, **17:**336-341.
24. Lee MJ, Gergely F, Jeffers K, Peak-Chew SY, Raff JW: **Msps/XMAP215 interacts with the centrosomal protein D-TACC to regulate microtubule behaviour.** *Nat Cell Biol* 2001, **3:**643-649.
25. Giet R, McLean D, Descamps S, Lee MJ, Raff JW, Prigent C, Glover DM: **Drosophila Aurora A kinase is required to localize D-TACC to centrosomes and to regulate astral microtubules.** *J Cell Biol* 2002, **156:**437-451.
26. Sadek CM, Pelto-Huikko M, Tujague M, Steffensen KR, Wennerholm M, Gustafsson JA: **TACC3 expression is tightly regulated during early differentiation.** *Gene Expr Patterns* 2003, **3:**203-211.
27. Gergely F, Karlsson C, Still I, Cowell J, Kilmartin J, Raff JW: **The TACC domain identifies a family of centrosomal proteins that can interact with microtubules.** *Proc Natl Acad Sci U S A* 2000, **97:**14352-14357.
28. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: **Protein interaction mapping in C. elegans using proteins involved in vulval development.** *Science* 2000, **287:**116-122.
29. Bader GD, Betel D, Hogue CW: **BIND: the Biomolecular Interaction Network Database.** *Nucleic Acids Res* 2003, **31:**248-250.
30. Boulton SJ, Martin JS, Polanowska J, Hill DE, Gartner A, Vidal M: **BRCA1/BARD1 Orthologs Required for DNA Repair in Caenorhabditis elegans.** *Curr Biol* 2004, **14:**33-39.
31. Ou SH, Wu F, Harrich D, Garcia-Martinez LF, Gaynor RB: **Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs.** *J Virol* 1995, **69:**3584-3596.
32. Wang HY, Wang IF, Bose J, Shen CK: **Structural diversity and functional implications of the eukaryotic TDP gene family.** *Genomics* 2004, **83:**130-139.
33. Piekorz RP, Hoffmeyer A, Duntsch CD, McKay C, Nakajima H, Sexl V, Snyder L, Rehg J, Ihle JN: **The centrosomal protein TACC3 is essential for hematopoietic stem cell function and genetically interfaces with p53-regulated apoptosis.** *EMBO J* 2002, **21:**653-664.
34. Gangisetty O, Lauffart B, Sondarva GV, Chelsea DM, Still IH: **The transforming acidic coiled coil proteins interact with nuclear histone acetyltransferases.** *Oncogene* 2004, **23:**2559-2563.
35. Groisman I, Huang YS, Mendez R, Cao Q, Theurkauf W, Richter JD: **CPEB, maskin, and cyclin B1 mRNA at the mitotic apparatus: implications for local translational control of cell division.** *Cell* 2000, **103:**435-447.
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.
37. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M: **The Ensembl genome database project.** *Nucleic Acids Res* 2002, **30:**38-41.
38. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with Clustal X.** *Trends Biochem Sci* 1998, **23:**403-405.
39. Page RD: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12:**357-358.
40. Felsenstein J: **Phylogenies from molecular sequences: inference and reliability.** *Annu Rev Genet* 1988, **22:**521-565.