

The Molecular Evolution of the p120-Catenin Subfamily and Its Functional Associations

Robert H. Carnahan¹*, Antonis Rokas²*, Eric A. Gaucher³, Albert B. Reynolds¹*

1 Department of Cancer Biology, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States of America, **3** School of Biology, Georgia Institute of Technology, Atlanta, Georgia, United States of America

Abstract

Background: p120-catenin (p120) is the prototypical member of a subclass of armadillo-related proteins that includes δ -catenin/NPRAP, ARVCF, p0071, and the more distantly related plakophilins 1–3. In vertebrates, p120 is essential in regulating surface expression and stability of all classical cadherins, and directly interacts with Kaiso, a BTB/ZF family transcription factor.

Methodology/Principal Findings: To clarify functional relationships between these proteins and how they relate to the classical cadherins, we have examined the proteomes of 14 diverse vertebrate and metazoan species. The data reveal a single ancient δ -catenin-like p120 family member present in the earliest metazoans and conserved throughout metazoan evolution. This single p120 family protein is present in all protostomes, and in certain early-branching chordate lineages. Phylogenetic analyses suggest that gene duplication and functional diversification into “p120-like” and “ δ -catenin-like” proteins occurred in the urochordate-vertebrate ancestor. Additional gene duplications during early vertebrate evolution gave rise to the seven vertebrate p120 family members. Kaiso family members (i.e., Kaiso, ZBTB38 and ZBTB4) are found only in vertebrates, their origin following that of the p120-like gene lineage and coinciding with the evolution of vertebrate-specific mechanisms of epigenetic gene regulation by CpG island methylation.

Conclusions/Significance: The p120 protein family evolved from a common δ -catenin-like ancestor present in all metazoans. Through several rounds of gene duplication and diversification, however, p120 evolved in vertebrates into an essential, ubiquitously expressed protein, whereas loss of the more selectively expressed δ -catenin, p0071 and ARVCF are tolerated in most species. Together with phylogenetic studies of the vertebrate cadherins, our data suggest that the p120-like and δ -catenin-like genes co-evolved separately with non-neural (E- and P-cadherin) and neural (N- and R-cadherin) cadherin lineages, respectively. The expansion of p120 relative to δ -catenin during vertebrate evolution may reflect the pivotal and largely disproportionate role of the non-neural cadherins with respect to evolution of the wide range of somatic morphology present in vertebrates today.

Citation: Carnahan RH, Rokas A, Gaucher EA, Reynolds AB (2010) The Molecular Evolution of the p120-Catenin Subfamily and Its Functional Associations. *PLoS ONE* 5(12): e15747. doi:10.1371/journal.pone.0015747

Editor: Michael Klymkowsky, University of Colorado, Boulder, United States of America

Received: September 22, 2010; **Accepted:** November 26, 2010; **Published:** December 31, 2010

Copyright: © 2010 Carnahan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Research in Antonis Rokas' laboratory is supported by the Searle Scholars Program and the National Science Foundation (DEB-0844968). Robert Carnahan is supported by the Vanderbilt Antibody and Protein Resource and RO1 CA055724 to Albert Reynolds. This work is supported in part by the Vanderbilt Cancer Center Support Grant (P30 CA068485), RO1 CA111947 (Albert Reynolds), and the Vanderbilt GI SPORE 50 CA95103 (Robert Coffey/Albert Reynolds). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: al.reynolds@vanderbilt.edu

† These authors contributed equally to this work.

Introduction

The integration over time of increasingly sophisticated signaling and cell-cell adhesion mechanisms has likely been an essential and ongoing process in the evolution of complex metazoan life. Interestingly, the Wnt signaling and cadherin-based adhesion functions of β -catenin have coexisted at least as far back as the origin of animals [1] (though *C. elegans* is a notable exception [2]), with coordination of these roles by a single protein perhaps facilitating evolution of the first multicellular organisms. Indeed, the evolutionary importance of β -catenin is reflected by phylogenetic analyses, which suggest a widespread and persistent stabilizing selection on each of the Armadillo (Arm) repeat sequences from Cnidarian to mouse β -catenin [3], and virtually no change in

β -catenin over the ~400 million year course of vertebrate evolution [3]. In vertebrates, β -catenin (or Plakoglobin) coexists with two other so-called “catenins” (i.e., p120-catenin and α -catenin) that together form a regulatory protein complex on the cytoplasmic tail of classical cadherins (i.e., Type I and type II cadherins). Evolutionary histories for cadherin- and β -catenin-families have been studied extensively [3,4,5,6,7,8] but similar analyses for the p120-catenin (hereafter p120) and α -catenin families have yet to be reported.

The appearance of cadherins is clearly a watershed event in metazoan evolution. While adhesion *per se* likely predates metazoans [6], the origin and diversification of the greater cadherin family has permitted an explosion in functional diversity of intercellular interactions. Interestingly, vertebrate evolution has

avored a particular paradigm, the classical cadherin, which has duplicated and reduplicated from a single vertebrate ancestor [5] to form a 26-member family. Structurally, the “classical cadherin” is comprised of five extracellular cadherin (EC) repeats and a highly conserved cytoplasmic tail containing a p120-binding juxtamembrane domain (JMD) and a C-terminal “catenin binding domain” (CBD) that interacts with β -catenin. As the predominant cadherin type in vertebrate cell-cell adhesion, the classical cadherins have also taken on fundamentally important roles in cell-cell adhesion, development and cancer, and mediate the majority of cell- and tissue-specific interactions in vertebrates.

In vertebrates, p120 behaves as a master regulator of classical cadherin stability, and is critical for proper cell-cell adhesion in most solid tissues [9,10,11]. Deletion (or knockdown) of the p120 gene in vertebrates (e.g., mouse, xenopus, zebrafish) is embryonic lethal despite the presence of ARVCF, δ -catenin, and p0071, closely related family members with at least partially overlapping functions [12,13,14,15]. Paradoxically, the single p120 family member in invertebrates (e.g., *Drosophila melanogaster*, *Caenorhabditis elegans*) is not essential for life in most species (although this point has been debated in drosophila) [16,17,18]. Thus, in vertebrates, p120 has evolved one or more essential functions relative to its invertebrate counterpart, and a critical role with respect to the classical cadherins.

p120 family members share a conserved central domain composed of 9 Arm repeats and flanking N- and C-terminal regions that diverge from one another (Figure 1). The “core” family members interact in adherens junctions with classical cadherins via Arm repeats 1–6 [19]. In contrast, the more distantly related plakophilins have evolved specialized roles in desmosomal junctions, which are mechanically and spatially distinct from the adherens junction [20]. Surprisingly, despite structural similarity to p120, their interaction with desmosomal cadherins is not mediated by the Arm domain, but occurs instead through the plakophilin N-terminal head domain [21,22,23]. Knockout studies in mice reveal that plakophilin 2 ablation is embryonic lethal [24] while plakophilin 1 and 3 can be eliminated with relatively little effect [20,25].

p120 also interacts directly with the transcription factor Kaiso [26]. Kaiso belongs to a large family of BTB/ZF proteins, most of which are important in development and cancer, and a closely related Kaiso subfamily consisting of Kaiso, ZBTB38 and ZBTB4 and [27]. Interestingly, Kaiso is bimodal in that it interacts with a conventional sequence-specific DNA motif referred to as the Kaiso Binding Site (KBS) [28] and also with methyl-CpG containing motifs [29]. The latter are high affinity interactions that have been reported to suppress the transcription of several tumor suppressors (e.g., pRb, p16, HIC) through interaction with inappropriately methylated CpG islands [30]. Kaiso has also been shown in *Xenopus* to suppress several Wnt pathway genes (e.g., Wnt 11, Siamois) by association with the KBS [31,32]. Interestingly, a third mechanism has been proposed that does not involve direct interaction with DNA. Instead, Kaiso binds TCF, a β -catenin-associated transcription factor. Kaiso and TCF associate with one another via their DNA binding motifs, thereby mutually excluding interaction with chromatin [33]. According to this scenario, p120 may interact with and/or modulate canonical Wnt signaling via regulation of Kaiso. Indeed, overexpressed p120 promotes translocation of Kaiso out of the nucleus [31,34], potentially facilitating TCF interaction with chromatin.

A feature shared by most, if not all members of the p120 family is physical and/or functional interaction with a number of Rho-GTPases, -GEFs and -GAPs [35]. For example, p120 can inhibit RhoA directly [36,37], or indirectly through p190RhoGAP [38],

and has been shown to promote Rac1 activation [39,40]. In general, these activities are thought to play critical roles in regulating the cytoplasmic interface between the various cadherin receptors and the cytoskeleton.

Here, we have analyzed proteomes from 14 diverse metazoan species to understand the evolution of the p120 protein family and the origin of its functional association with classical cadherins and Kaiso. We find that all invertebrates as well as several early-branching chordate lineages contain a single family member with a “ δ -catenin-like” set of functions, suggesting that the p120-family ancestor was “ δ -catenin-like” and highly conserved in pre-vertebrate metazoans. Gene duplications in chordate and vertebrate evolution gave rise to the six-seven family members in present day vertebrates, and provided the raw material and opportunity for functional diversification.

Together with phylogenetic studies of the classical cadherins, our data suggest that p120- and δ -catenin-like lineages split from one another in chordates and then separately co-evolved with non-neural (E- and P-cadherin) and neural (N- and R-cadherin) branches, respectively, of the vertebrate classical cadherins. A similar scenario with respect to α -catenin (also called α -catenin-1) and α -N-catenin (neural α -catenin, also called α -catenin-2) (E. Gaucher, personal communication) suggests that these distinct branches of the (vertebrate) classical cadherin family co-evolved with their own distinct subsets of both p120 (p120 vs. δ -catenin) and α -catenin (α -catenin vs. α -N-catenin) family members. Thus, the rapid expansion of p120, relative to δ -catenin, during vertebrate evolution may in large part reflect the broader spectrum of tissue and organ diversity outside of the nervous system. Other p120-specific innovations of note include the evolution of alternative splicing relevant to epithelial to mesenchymal transformation (EMT), loss of the C-terminal PDZ ligand motif, and interaction with Kaiso. The vertebrate specific appearance of Kaiso and its unique interactions with p120, TCF4 and methyl-CpG DNA suggest other p120 connections relevant to Wnt signaling and vertebrate-specific mechanisms of transcriptional regulation.

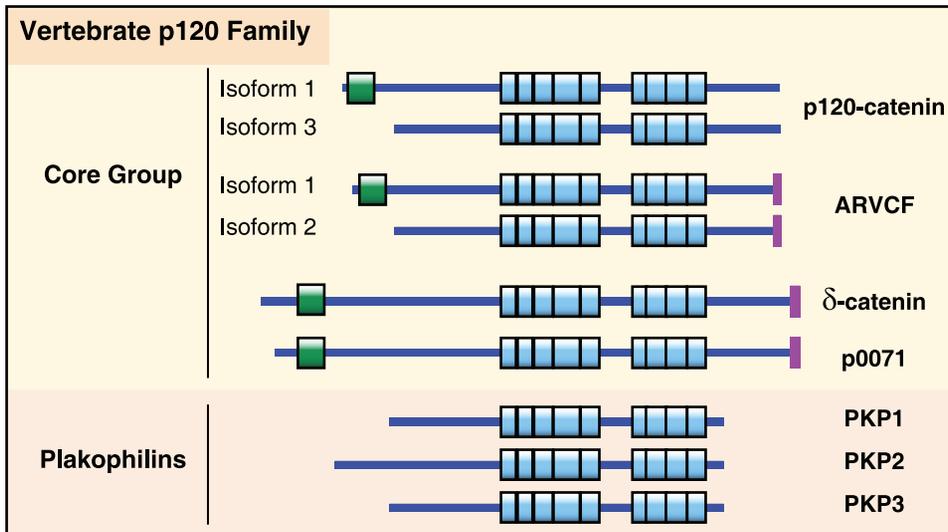
Results

p120 protein family

A total of 65 protein sequences were retrieved from the 14 species examined (Table 1). All protostome species examined and early-branching chordates (the cephalochordate *Branchiostoma floridae* and the echinoderm *Strongylocentrotus purpuratus*) contain a single member from the δ -catenin family. The urochordates (represented by *Ciona intestinalis*), the closest evolutionary relatives of vertebrates, contain two members, whereas all vertebrate species contain typically seven protein members of the p120-catenin family. The only exceptions within the vertebrates are *X. tropicalis*, whose proteome contains six proteins, and the two fish, *D. rerio* and *T. rubripes*, whose proteomes contain 10 and 13 members of the p120-catenin family, respectively, almost twice the number of members of the protein family as the non-fish vertebrates (Table 1).

Phylogenetic analysis of all protein family members identified suggests that the seven protein members typically found in vertebrates correspond to the seven delta-catenin protein subfamilies previously identified in humans (Figures 1 and 2) [35]. Specifically, these are the plakophilin 1, plakophilin 2, plakophilin 3, p0071, delta, p120 and ARVCF subfamilies. These seven subfamilies are robustly placed into three major clades: the first clade is composed of plakophilin 1, plakophilin 2, and plakophilin 3, the second clade of δ -catenin and p0071, and the third clade of

A



B

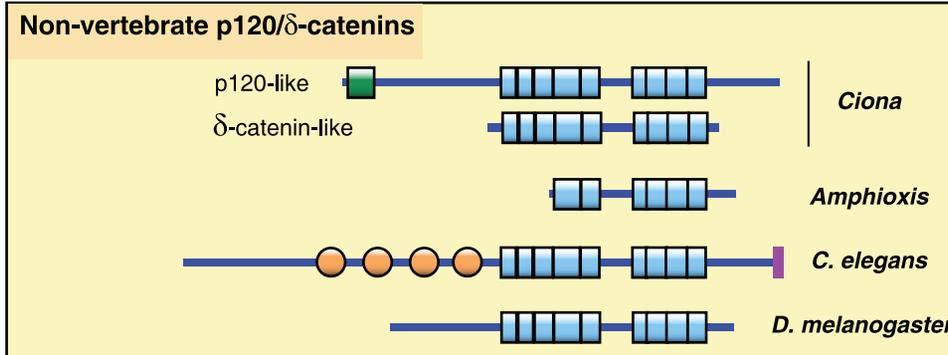


Figure 1. Domain structure of the p120-catenin family. (A) The vertebrate members of the p120-catenin family of proteins all contain a central Armadillo repeat domain consisting of 9 tandemly linked imperfect 42 amino-acid repeats (blue boxes). The four “Core” members also contain an amino-terminally located coiled-coil domain (green boxes). In the case of p120 and ARVCF, alternative splicing in this region gives rise to two major isoforms, which either contain (isoform 1) or not (isoform 3) this coiled-coil region. All “Core” members, except p120, also have a carboxy-terminally located PDZ ligand domain (purple boxes). (B) The invertebrate members of the p120-catenin family similarly possess centrally located Armadillo repeats, though in the case of *Amphioxys* this region contains 6 rather than 9 repeats. N-terminal regions show more diversity with no distinct domain structure (*D. melanogaster*, *Amphioxys*, *Ciona* δ -catenin-like), Fibronectin type III domains (orange circles, *C. elegans*), or a coiled-coil domain (*Ciona* p120-like). Similar to vertebrate members, the *C. elegans* family member also contains a carboxy-terminally located PDZ ligand domain (purple box). doi:10.1371/journal.pone.0015747.g001

p120 and ARVCF. The phylogeny of protein members within each one of these seven functional categories is consistent with the vertebrate phylogeny, suggesting that the vertebrate ancestor possessed a single protein from each of the seven functional categories. Interestingly, the proteome of *C. intestinalis*, the closest relative of vertebrates included in this study, contains only two proteins. One protein consistently groups with the p120 – ARVCF clade (Figure 2), whereas the other protein is nested within the deuterostome δ -catenin clade, but is not robustly grouped with any of the seven subfamilies or any of the three clades identified. The same is true for the single protein members identified in *B. floridae* and *S. purpuratus*, the two other deuterostome lineages included in our study. Finally, all protostomes examined contain a single member of the p120 family and likely are the outgroup of the deuterostome p120 family (Figure 2).

The increase in the number of p120 family members observed in vertebrates and the further increase in fish are consistent with studies suggesting that the ancestral vertebrate underwent two

rounds of whole-genome duplication (WGD) [41] and that actinopterygian fish underwent additional rounds of WGD [42,43]. For example, for every single non-fish vertebrate subfamily member, two or three fish subfamily members are typically identified. However, the increase in number of members is unlikely to have been solely due to the WGDs and additional gene duplications likely contributed to the generation of the current diversity of protein members of the p120 family observed today.

Kaiso protein family

A total of 17 protein sequences were retrieved from the 14 species examined (Table 2). No Kaiso protein family members were identified in protostomes and in non-vertebrate chordates. All vertebrates contain two or three of the Kaiso family proteins. All vertebrates contain Kaiso, but several are missing either ZBTB4 or ZBTB38 (Table 2). Phylogenetic analysis of all protein family members identified three major clades that correspond to the three proteins (Figure 3) [27].

Table 1. Distribution of δ -catenin protein family members across animal phyla.

Species	Phylum	Number
Human	Vertebrata	7
Mouse	Vertebrata	7
Dog	Vertebrata	7
Chicken	Vertebrata	7
Frog (<i>Xenopus laevis</i>)	Vertebrata	6
Pufferfish	Vertebrata	13
Zebrafish	Vertebrata	10 (two of the genes are identical in sequence)
Tunicate	Urochordata	2
Amphioxus	Cephalochordata	1
Echinoderm	Echinodermata	1
Fly	Arthropoda	1
Worm	Nematoda	1
Polychaete (<i>Capitella</i>)	Annelida	1
Clam (<i>Lottia</i>)	Mollusca	1

doi:10.1371/journal.pone.0015747.t001

Phylogenetic relationships of p120 family to other functionally relevant proteins

Figure 4 compares our data with respect to phylogenetic histories of the p120 and Kaiso families (Figure 4A) with published [3,4,5,27,44] and unpublished (i.e., α -catenin family, E. Gaucher, personal communication) histories of other functionally related proteins (Figure 4B). Of particular interest is the existence of a single member of each of the catenin families (i.e., δ -catenin, α -N-catenin, and β -catenin) conserved throughout pre-vertebrate metazoan evolution, a pattern that coincides with the phylogenetic origins of Wnt family proteins. The plakophilins, on the other hand, along with Kaiso and Desmosomal cadherins, are vertebrate innovations. Of note, the δ -catenin-like and p120-like ancestors of the present day p120 family arise just prior to vertebrates, as does the first common ancestor of the classical cadherins (i.e., vertebrate Type I and Type II cadherins).

Discussion

Conservation of a δ -catenin-like gene over the course of metazoan evolution suggests an ancient and evolutionarily important role, but the effects of deleting the only δ -catenin-like gene present in worms and flies is not as dire as one might expect. δ -catenin knockdown in xenopus is, in fact, embryonic lethal [45], but the effects of δ -catenin KO in mice appear to be largely cognitive [46,47,48,49]. Although fly p120/ δ -catenin associates and colocalizes with fly E-cadherin [16], the evidence overall suggests that its role is not directly comparable to that of vertebrate p120. A strong possibility is that fly p120/ δ -catenin has an ancient function that is nonessential for life but nonetheless confers a strong evolutionary advantage. For example, the significant cognitive abnormalities exhibited by δ -catenin KO mice [46,49] may not be immediately apparent in captivity but could markedly affect their ability to compete and survive in the wild. Indeed, δ -catenin is one of several genes deleted in human Cri-du-chat patients and may contribute to the mental retardation associated with the disorder [46,47].

The vertebrate p120 family consists of seven members. Four “core” members (i.e., δ -catenin, p120-catenin, ARVCF, and p0071) function in adherens junctions, and three less well conserved members function in desmosomes (Plakophilins 1, 2, and 3) (Figure 1). The phylogenetic analyses presented here show that they evolved through rounds of gene duplication and functional diversification from an ancient “ δ -catenin-like” gene that is conserved throughout metazoan evolution. The ancestral δ -catenin was probably similar in function to the gene member currently present in invertebrates, echinoderms and cephalochordates (Figures 1 and 2; Table 1). The first gene duplication took place in the urochordate-vertebrate ancestor, giving rise to “ δ -catenin-like” and “p120-catenin-like” progenitors. Additional gene duplication(s), most likely a consequence of the two rounds of whole genome duplication at the origin of vertebrates, gave rise to (1) a δ -catenin clade consisting of vertebrate δ -catenin and p0071, and (2) a p120 clade consisting of vertebrate p120 and ARVCF. The plakophilins represent a vertebrate specific offshoot of the δ -catenin-like progenitor.

The phylogeny of the p120 family is relatively straight forward, but exactly how or why p120 has evolved to become the predominant family member in vertebrates is harder to explain. One possibility is that p120 has evolved uniquely advantageous features important for cadherin function. Indeed, comparison of current structural and functional characteristics of the various family members reveals several potentially critical p120 adaptations. First, p120 is the only core family member in vertebrates that lacks a C-terminal PDZ ligand domain. This domain mediates protein-protein interactions with a number of important PDZ domain containing proteins (e.g. PSD-950, erbin, densin-180). The PDZ ligand domain itself is an ancient feature of the p120 lineage as it is present, for example, in the sole family member of various protostomes such as *C. elegans*. The p120-like progenitor of p120 and ARVCF, on the other hand, has a C-terminal sequence that differs at one residue from known consensus motif sequences (i.e., NSWV). Notably, the p120 progenitor is equally similar to p120 and ARVCF by most criteria, but a bona fide C-terminal PDZ ligand would imply that the progenitor was functionally more similar to ARVCF than p120.

Regardless, p120 is clearly the only core member of the vertebrate p120 family that lacks the C-terminal PDZ ligand domain and conceivably, certain physical and functional evolutionary constraints imposed by preexisting PDZ binding partners of ARVCF, δ -catenin and p0071. Indeed, spine (and synapse) density in mouse hippocampal neurons is significantly increased by δ -catenin ablation, but the effect is not cadherin-dependent. Instead, it clearly depends on a PDZ-ligand mediated interaction with one or more PDZ domain-containing proteins [49]. In contrast, p120 ablation in the same tissue has the opposite effect on spine density and works through a very different mechanism associated with modulation of Rho GTPases [13]. These data highlight the functional importance of the C-terminal PDZ ligand, and illustrate how it can contribute to the markedly different roles for δ -catenin and p120 in hippocampal neurons, as well as other tissue types. Overall, these observations strongly support the notion that the absence of a PDZ ligand domain may have endowed p120 (and p120 bound cadherin complexes) with significant flexibility to evolve novel physical and functional interactions that are independent of PDZ-mediated roles.

Second, a potentially critical adaptation is the evolution of alternative splicing in the amino-terminal regulatory domain of p120 and ARVCF [50,51,52], but apparently not in δ -catenin or p0071. The ability to use alternative start sites allows p120 (and

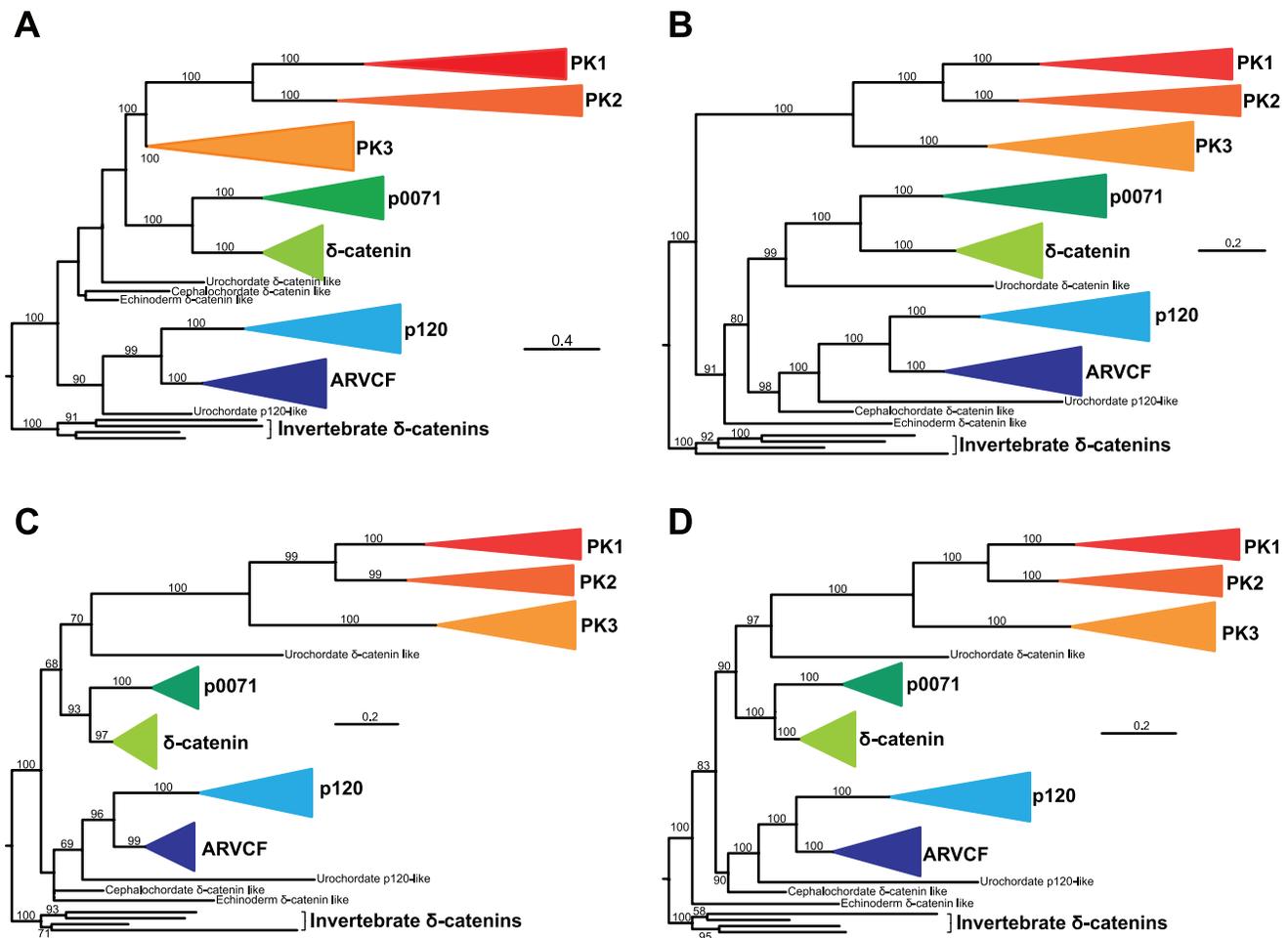


Figure 2. The evolutionary history of the p120-catenin family of proteins. (A) Maximum likelihood analysis on an alignment constructed using entire protein sequences, (B) Bayesian inference analysis on an alignment constructed using entire protein sequences, (C) Maximum likelihood analysis on an alignment constructed using the nine Arm domains, and (D) Bayesian inference analysis on an alignment constructed using the nine Arm domains. Color-coded clades correspond to the seven family members found in vertebrates (p120, ARVCF, p0071, δ -catenin, plakophilins 1–3) and each clade contains only sequences from vertebrates. Numbers near internodes indicate bootstrap (for maximum likelihood analyses)/posterior probability (for Bayesian inference analyses) clade support values. Clade support values <50% are not shown. PK1-3: plakophilins 1–3. doi:10.1371/journal.pone.0015747.g002

ARVCF) to separately express isoform 1 and/or isoform 3, forms of p120 that likely have significantly different roles. Specifically, isoform 1, but not isoform 3, contains the N-terminal coiled coil (CC) domain, a \sim 40 amino acid N-terminal domain that is presumed to be important because it is almost perfectly conserved in all core family members. p120 isoform 1 is expressed predominantly in mesenchymal (e.g., fibroblasts) and certain other non-epithelial cell types (e.g., neurons), whereas the shorter isoform 3 is preferred in epithelial and other relatively sessile cell types. Importantly, p120 isoform switching (e.g., from isoform 3 to isoform 1) is dynamic and typically coordinated with classic cadherin switching (e.g., E-cadherin to N-cadherin) that occurs during epithelial to mesenchymal transformation (EMT) [53,54]. The ability to directly modulate and/or participate in EMT is likely to be significant, as this process is critically important during development, wound healing and cancer.

Notably, p120 is the only family member possessing both of these innovations (i.e., absence of a PDZ ligand domain and presence of alternative start sites). ARVCF undergoes N-terminal alternative splicing, but contains a C-terminal PDZ ligand motif. Whether one or both of these factors substantially influenced the

adoption of p120 by classical cadherins is largely speculation. Nonetheless, if adaptive advantage did in fact play a role, the most likely determinant of such an event is p120 itself, and both factors offer plausible advantages relevant to flexibility and/or function. δ -catenin, on the other hand, is likely constrained by PDZ-mediated interactions and the inability to generate an isoform that lacks the coiled-coil domain.

Interaction with Kaiso provides a third p120 adaptation that is absent from other family members. In support of a previous study by Fillion et al [27], we find that Kaiso is vertebrate-specific, and thus coincides with both the origin of vertebrate p120 and the vertebrate specific expansion of the classical cadherins. Kaiso belongs to a unique family of transcription factors that can associate selectively, and with high affinity, to methylated CpG DNA via zinc finger domains [29]. Kaiso is actually bimodal in that it also binds with lower affinity to a conventional DNA motif [55]. A recent report shows that Kaiso can shut down the transcription of key tumor suppressors (e.g., pRb, p16, Hic1) by interaction with inappropriately methylated CpG islands. Thus, Kaiso may link p120 to epigenetic transcriptional regulation via CpG island methylation, a cancer-relevant and largely vertebrate-

Table 2. Distribution of ZBTB protein family members across animal phyla.

Species	Phylum	Kaiso	ZBTB38	ZBTB4
Human	Vertebrata	✓	✓	✓
Mouse	Vertebrata	✓	✓	✓
Dog	Vertebrata	✓	X	✓
Chicken	Vertebrata	✓	✓	X
Frog (<i>Xenopus laevis</i>)	Vertebrata	✓	✓	X
Pufferfish	Vertebrata	✓	X	✓
Zebrafish	Vertebrata	✓	✓	✓
Tunicate	Urochordata	X	X*	X
Amphioxus	Cephalochordata	X	X*	X
Echinoderm	Echinodermata	X	X*	X
Fly	Arthropoda	X	X*	X
Worm	Nematoda	X	X*	X
Polychaete (<i>Capitella</i>)	Annelida	X	X*	X
Clam (<i>Lottia</i>)	Mollusca	X	X*	X

X*: these proteomes contain many hits with e-values $\sim 10^{-30}$.
doi:10.1371/journal.pone.0015747.t002

specific mechanism associated with the use of hypomethylated CpG islands as sites of active transcription [56]. Interestingly, Kaiso and TCF are reported to associate physically via their DNA binding domains, thereby preventing one another from interacting with chromatin [33]. These data raise the possibility that p120's interaction with Kaiso modulates canonical Wnt signaling through TCF4. While it is unlikely that p120 and Kaiso are essential for Wnt signaling, their influence might be important in the context of complex developmental and regulatory vertebrate environments. Given that Kaiso is absent from non-vertebrate metazoans, the evolution of interactions with both p120 and TCF may represent a vertebrate-specific adaptation connecting cadherin complexes in

general, and p120 in particular, to canonical Wnt signaling pathways. What exactly this means for vertebrate Wnt signaling and/or related functions has yet to be determined, but in contrast to β -catenin and TCF, lessons in vertebrate p120 or Kaiso functions are unlikely to be guided by genetic studies in non-vertebrate model systems.

As mentioned, the increase in the number of p120 family members observed in vertebrates is consistent with studies suggesting that the ancestral vertebrate underwent two rounds of whole-genome duplication [41]. Evidently, these were instrumental in the evolution of at least two broad categories of classical cadherin complexes. The ancestral invertebrate forms of α -catenin and p120 were duplicated and have emerged in vertebrates as α -N-catenin and δ -catenin, both of which are found primarily in neural tissues [57,58,59,60]. Their duplicated counterparts, on the other hand, evolved to become α -catenin and p120, respectively, and are expressed in all solid tissues, including the nervous system. In parallel, the classical cadherins evolved from a single vertebrate ancestor by gene duplications that led to the evolution of at least four classical cadherins, most likely the ancestors of present day N-, R-, E- and P-cadherins [4,5]. These cadherin paralogues appear to represent early neural (N- and R-cadherins) and non-neural/epithelial (E- and P-cadherins) lineages that subsequently evolved at different rates [4]. Thus, in vertebrates, the ancestral "invertebrate counterparts" of p120 and α -catenin (i.e., δ -catenin and α -N-catenin) appear to be primarily confined to the nervous system, while p120 and α -catenin are found in essentially all solid tissues, nervous system included. The very different features of δ -catenin and p120, as discussed above, may account for the relatively restricted tissue specific expression of δ -catenin, and the subsequent emergence of p120 as the most widely expressed member of the p120 family in vertebrates. The apparently analogous co-distribution of the α -catenin isoforms (E. Gaucher, personal communication) is probably related to these events, although coordinated alterations in gene regulatory elements could contribute to such events in any of the scenarios described above.

In any event, these observations suggest that in most vertebrate tissues, the main functional unit defined by the present-day

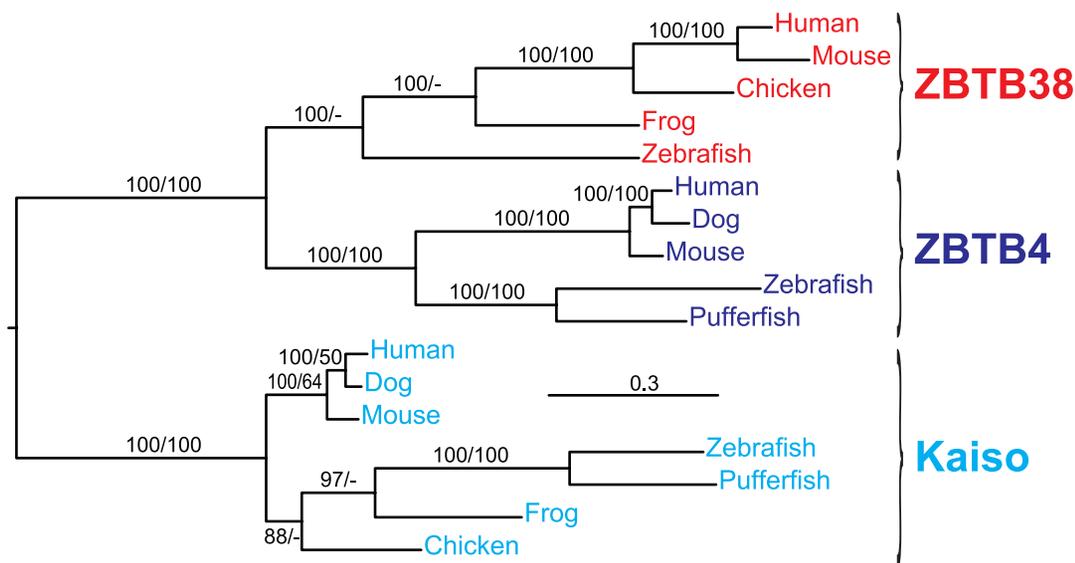


Figure 3. The evolutionary history of the Kaiso family of proteins. The majority-rule consensus tree from a Bayesian inference analysis on an alignment constructed using entire protein sequences is shown. Numbers near internodes indicate posterior probability (for Bayesian inference analyses)/bootstrap (for maximum likelihood analyses) clade support values. Clade support values <50% are not shown.
doi:10.1371/journal.pone.0015747.g003

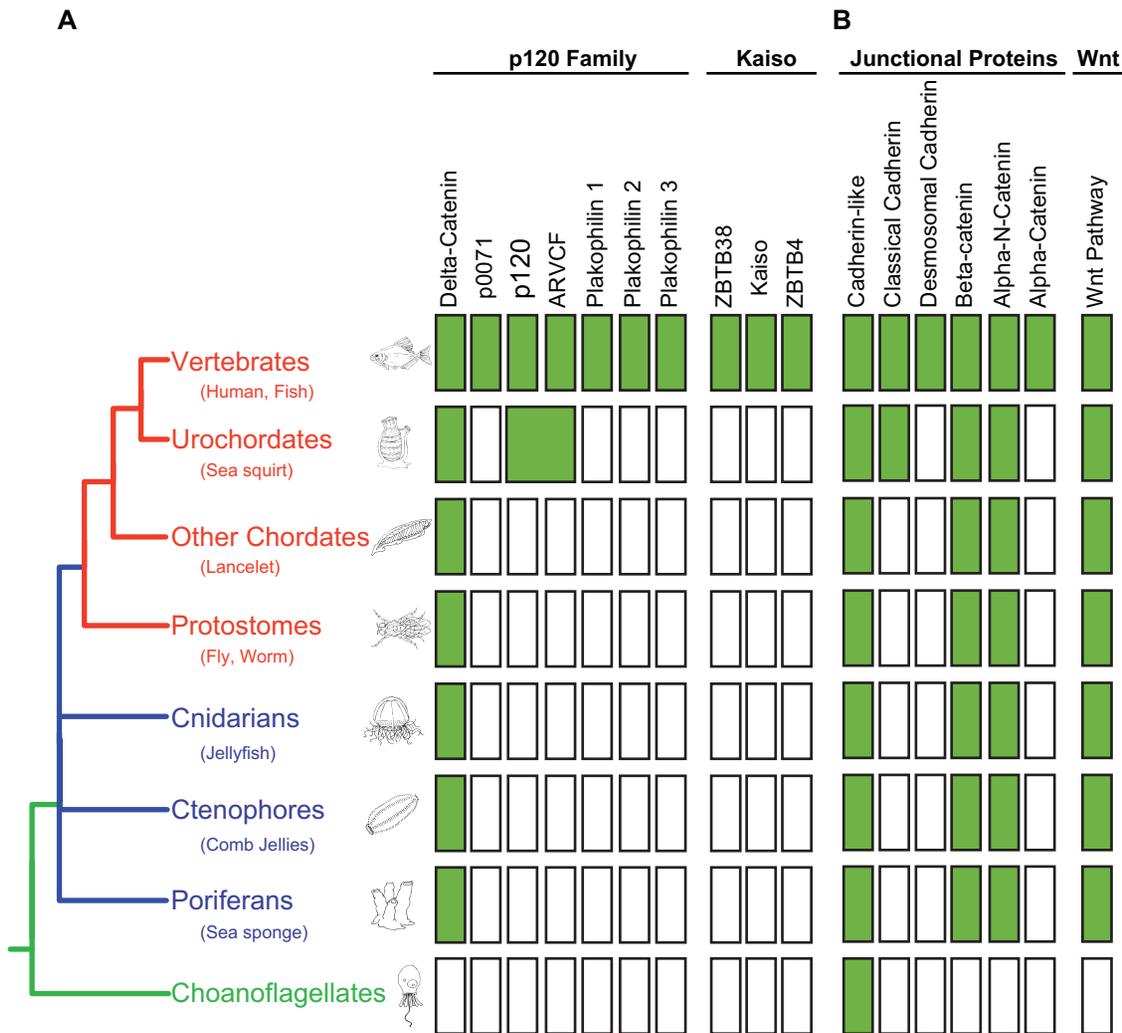


Figure 4. Presence/Absence of families involved in cell-cell adhesion and signalling. Phylogenetic distribution of p120-family proteins, kaiso-family proteins, junctional proteins, and proteins collectively required for a functional wnt pathway. A filled box indicates the presence of an orthologue from the corresponding protein family. Color coding for the phylogenetic tree is as follows: pre-metazoan in green, metazoan in blue, bilateral metazoans in red. doi:10.1371/journal.pone.0015747.g004

classical cadherin complex came together for the first time as a result of whole genome duplications that caused the ancestral catenins (δ -catenin and α -N-catenin) to partition with the neuronal lineage, (presumably in association with N-cadherin), and their vertebrate-specific derivatives (p120 and α -catenin) to form a second lineage (presumably in association with a common ancestor of non-neural cadherins - perhaps E- and/or P-cadherins), which was then favored as the raw material for diversification of most other tissues. The former was likely constrained by the need to conserve complex neuronal functions whereas the rapid evolution of the latter is consistent with a cadherin complex that is more flexible with respect to expansion of novel interactions. The extraordinary success of this ultimate classical cadherin complex is evidenced by the repeated duplication and diversification of the classical cadherins to at least 26 members, most of which use the same basic set of p120-, α - and β -catenin building blocks. This core design has thus been preserved and reutilized by classical cadherins for approximately half a billion years, while simultaneously serving as a key driver of vertebrate cell- and tissue-diversification.

Interestingly, a similar paradigm appears to extend to the desmosomal cadherins and their interaction with the more distant members of the p120 family, the plakophilins. Figure 2 shows that the plakophilins are of vertebrate origin and share a common ancestor with the vertebrate δ -catenin clade. Their appearance coincides with that of several other important components of desmosomes, which also originate in vertebrates. Plakoglobin, for example, evolved around the same time via gene duplication of β -catenin, and functions in both adherens junctions and desmosomes. Interestingly, the desmosomal cadherins later diverge from other cadherins and the family appears to expand within mammals [5], permitting evolution of the desmosome. Our analyses also show that the plakophilins are the fastest evolving members of the p120-family (Figure 2). Importantly, like p120 and β -catenin, the plakophilins also have roles in the nucleus [51,61,62,63,64], suggesting other potentially significant functions that have yet to be defined. Overall, the fastest evolving clade of the p120 family and the desmosomal cadherins appear to be recycling the evolutionary game plan of the classical cadherins.

Materials and Methods

Data matrix construction

The complete proteome sequence files of 7 vertebrates (*Homo sapiens*, *Canis familiaris*, *Mus musculus*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, and *Takifugu rubripes*), 1 urochordate (*Ciona intestinalis*), 1 cephalochordate (*Branchiostoma floridae*), 1 echinoderm (*Strongylocentrotus purpuratus*) and 4 protostomes (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Helobdella robusta*, and *Lottia gigantea*) were retrieved from the Ensembl FTP Server (<http://uswest.ensembl.org/info/data/ftp/index.html>) and JGI Genome Portal (<http://genome.jgi-psf.org/>) websites. All proteome sequence files were processed so that only the longest protein sequence product of a given gene was retained using a custom Perl script. Members of the δ -catenin protein family were identified using the BLASTP similarity search algorithm, version 2.2.16 [65]. This was done by blasting the human p120 protein (GENBANK accession number: NP_001078927.1) against each proteome and retrieving all protein sequences showing significant similarity. Similar results were obtained using other members of the δ -catenin protein family from *Homo sapiens* or from other species.

Phylogenetic analyses

Phylogenetic analyses were performed using the optimality criteria of Bayesian inference (BI) and Maximum Likelihood (ML). According to the BI optimality criterion, the tree that best explains our protein alignment is considered the best estimate of the true phylogeny of our proteins [66]. According to the ML criterion, the tree that makes our protein alignment the most probable evolutionary outcome given a specific model of protein evolution is considered the best estimate of the true phylogeny of our proteins [66]. BI and ML analyses were performed on two data matrices: the first data matrix was generated by the alignment of whole proteins, whereas the second data matrix was generated by concatenating the individual alignments of each of the nine Arm domains. All alignments were constructed using DIALIGN, version 2.2 [67]. DIALIGN is a local alignment algorithm that does not attempt to align proteins from start to finish. Instead, it only aligns the conserved protein regions between proteins and identifies all remaining (poorly conserved) regions as unaligned. This feature is particularly useful for aligning proteins like the p120 family where conserved domains are flanked by poorly conserved regions of

varying length. Importantly, DIALIGN displays all aligned residues in capital letters, and all unaligned residues in lowercase letters. In all cases, all unaligned amino acids were converted to “X”, the IUPAC symbol for unspecified amino acids, and were effectively filtered out from downstream phylogenetic analyses. Sequences belonging to each Arm domain were manually identified through careful comparison with the human p120 protein sequence. Alignments of each of the nine Arm domains were done in exactly the same fashion. BI analyses were conducted using MRBAYES, version 3.1.2 [68,69,70]. BI phylogenetic trees were constructed using a mix of empirical amino acid substitution matrices, allowing for rate of heterogeneity among sites by assuming that a certain proportion of sites were invariable and that the rates of the rest are determined according to the shape parameter alpha of the gamma distribution. Two independent analyses were run in parallel. Each analysis contained 4 chains (1 cold and 3 incrementally heated) and trees were sampled every 1,000 generations. The analyses were run for 2,000,000 generations by which time the average deviation of split frequencies was below 0.01. The trees and parameters sampled from the first 10% of generations from each of the two analyses were discarded as the burn-in. Clade support in BI analyses was assessed using posterior probabilities. ML analyses were conducted using RAXML, versions 7.2.5 and 7.2.6 [71]. ML phylogenetic trees were constructed using the WAG amino acid matrix [72], allowing for rate of heterogeneity among sites by assuming that the rates of the rest are determined according to the shape parameter alpha of the gamma distribution. Clade support in maximum likelihood analyses was assessed using non-parametric bootstrap re-sampling (100 replicates).

Acknowledgments

We thank Padmanabhan Mahadevan for providing the script for retrieval of the longest protein sequence product from each gene, and Nichole Lobdell for the drawings of representatives for various species. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University.

Author Contributions

Conceived and designed the experiments: RHC AR EAG ABR. Performed the experiments: RHC AR. Analyzed the data: RHC AR ABR. Contributed reagents/materials/analysis tools: RHC AR ABR. Wrote the paper: RHC AR ABR.

References

- Nichols SA, Dirks W, Pearse JS, King N (2006) Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA* 103: 12451–12456.
- Natarajan L, Witwer NE, Eisenmann DM (2001) The divergent *Caenorhabditis elegans* beta-catenin proteins BAR-1, WRM-1 and HMP-2 make distinct protein interactions but retain functional redundancy in vivo. *Genetics* 159: 159–172.
- Schneider SQ, Finnerty JR, Martindale MQ (2003) Protein evolution: structure-function relationships of the oncogene beta-catenin in the evolution of multicellular animals. *J Exp Zool B Mol Dev Evol* 295: 25–44.
- Gallin WJ (1998) Evolution of the “classical” cadherin family of cell adhesion molecules in vertebrates. *Mol Biol Evol* 15: 1099–1107.
- Hulpiau P, van Roy F (2009) Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol* 41: 349–369.
- Abedin M, King N (2008) The premetazoan ancestry of cadherins. *Science* 319: 946–948.
- Grimson MJ, Coates JC, Reynolds JP, Shipman M, Blanton RL, et al. (2000) Adherens junctions and beta-catenin-mediated cell signalling in a non-metazoan organism. *Nature* 408: 727–731.
- Ozawa M, Baribault H, Kemler R (1989) The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. *Embo J* 8: 1711–1717.
- Davis MA, Ireton RC, Reynolds AB (2003) A core function for p120-catenin in cadherin turnover. *J Cell Biol* 163: 525–534.
- Ireton RC, Davis MA, van Hengel J, Mariner DJ, Barnes K, et al. (2002) A novel role for p120 catenin in E-cadherin function. *J Cell Biol* 159: 465–476.
- Xiao K, Allison D, Buckley K, Kottke M, Vincent P, et al. (2003) Cellular levels of p120 catenin function as a set point for cadherin expression levels in microvascular endothelial cells. *J Cell Biol* 163: 535–545.
- Davis MA, Reynolds AB (2006) Blocked Acinar Development, E-Cadherin Reduction, and Intraepithelial Neoplasia upon Ablation of p120-Catenin in the Mouse Salivary Gland. *Dev Cell* 10: 21–31.
- Elia LP, Yamamoto M, Zang K, Reichardt LF (2006) p120 catenin regulates dendritic spine and synapse development through Rho-family GTPases and cadherins. *Neuron* 51: 43–56.
- Oas RG, Xiao K, Summers S, Wittich KB, Chiasson CM, et al. (2010) p120-Catenin Is Required for Mouse Vascular Development. *Circulation Research*.
- Smalley-Freed WG, Efimov A, Burnett PE, Short SP, Davis MA, et al. p120-catenin is essential for maintenance of barrier function and intestinal homeostasis in mice. *J Clin Invest* 120: 1824–1835.
- Myster SH, Cavallo R, Anderson CT, Fox DT, Peifer M (2003) *Drosophila* p120catenin plays a supporting role in cell adhesion but is not an essential adherens junction component. *J Cell Biol* 160: 433–449.
- Pettitt J, Cox EA, Broadbent ID, Flett A, Hardin J (2003) The *Caenorhabditis elegans* p120 catenin homologue, JAC-1, modulates cadherin-catenin function during epidermal morphogenesis. *J Cell Biol* 162: 15–22.

18. Magic C, Pinto-Santini D, Parkhurst S (2002) Rho1 interacts with p120ctn and alpha-catenin, and regulates cadherin-based adherens junction components in *Drosophila*. *Development* 129: 3771–3782.
19. Hatzfeld M, Nachtsheim C (1996) Cloning and characterization of a new armadillo family member, p0071, associated with the junctional plaque: evidence for a subfamily of closely related proteins. *J Cell Sci* 109(Pt 11): 2767–2778.
20. Hatzfeld M (2007) Plakophilins: Multifunctional proteins or just regulators of desmosomal adhesion? *Biochim Biophys Acta* 1773: 69–77.
21. Kowalczyk AP, Hatzfeld M, Bornslaeger EA, Kopp DS, Borgwardt JE, et al. (1999) The head domain of plakophilin-1 binds to desmoplakin and enhances its recruitment to desmosomes. Implications for cutaneous disease. *J Biol Chem* 274: 18145–18148.
22. Hatzfeld M, Haffner C, Schulze K, Vinzens U (2000) The function of plakophilin 1 in desmosome assembly and actin filament organization. *J Cell Biol* 149: 209–222.
23. Chen X, Bonne S, Hatzfeld M, van Roy F, Green KJ (2002) Protein binding and functional characterization of plakophilin 2. Evidence for its diverse roles in desmosomes and beta -catenin signaling. *J Biol Chem* 277: 10512–10522.
24. Grossmann KS, Grund C, Huelsken J, Behrend M, Erdmann B, et al. (2004) Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. *J Cell Biol* 167: 149–160.
25. McGrath JA, Hoeger PH, Christiano AM, McMillan JR, Mellerio JE, et al. (1999) Skin fragility and hypohidrotic ectodermal dysplasia resulting from ablation of plakophilin 1. *Br J Dermatol* 140: 297–307.
26. Daniel JM, Reynolds AB (1995) The tyrosine kinase substrate p120cas binds directly to E-cadherin but not to the adenomatous polyposis coli protein or alpha-catenin. *Mol Cell Biol* 15: 4819–4824.
27. Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhorchouk E, et al. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* 26: 169–181.
28. Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A (2002) The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res* 30: 2911–2919.
29. Prokhorchouk A, Hendrich B, Jorgensen H, Ruzov A, Wilm M, et al. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* 15: 1613–1618.
30. Lopes EC, Valls E, Figueroa ME, Mazur A, Meng F-G, et al. (2008) Kaiso Contributes to DNA Methylation-Dependent Silencing of Tumor Suppressor Genes in Colon Cancer Cell Lines. *Cancer Res* 68: 7258–7263.
31. Kim SW, Park J-i, Spring CM, Sater AK, Ji H, et al. (2004) Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat Cell Biol* 6: 1212–1220.
32. Park JI, Kim SW, Lyons JP, Ji H, Nguyen TT, et al. (2005) Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev Cell* 8: 843–854.
33. Ruzov A, Hackett JA, Prokhorchouk A, Reddington JP, Madej MJ, et al. (2009) The interaction of xKaiso with xTcf3: a revised model for integration of epigenetic and Wnt signalling pathways. *Development*.
34. Kelly KF, Spring CM, Otchere AA, Daniel JM (2004) NLS-dependent nuclear localization of p120ctn is necessary to relieve Kaiso-mediated transcriptional repression. *J Cell Sci* 117: 2675–2686.
35. Hatzfeld M (2005) The p120 family of cell adhesion molecules. *Eur J Cell Biol* 84: 205–214.
36. Yanagisawa M, Huvelde D, Kreinest P, Lohse CM, Chevillat JC, et al. (2008) A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion and predicts metastatic disease. *Journal of Biological Chemistry*. 22 p.
37. Anastasiadis PZ, Moon SY, Thoreson MA, Mariner DJ, Crawford HC, et al. (2000) Inhibition of RhoA by p120 catenin. *Nat Cell Biol* 2: 637–644.
38. Wildenberg GA, Dohn MR, Carnahan RH, Davis MA, Lobdell NA, et al. (2006) p120-catenin and p190RhoGAP regulate cell-cell adhesion by coordinating antagonism between Rac and Rho. *Cell* 127: 1027–1039.
39. Noren NK, Liu BP, Burrige K, Krefl B (2000) p120 catenin regulates the actin cytoskeleton via Rho family GTPases. *J Cell Biol* 150: 567–580.
40. Grosheva I, Shitman M, Elbaum M, Bershadsky A (2001) p120 catenin affects cell motility via modulation of activity of Rho-family GTPases: a link between cell-cell contact formation and regulation of cell locomotion. *J Cell Sci* 114: 695–707.
41. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3: e314.
42. Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* 356: 1661–1679.
43. Meyer A, Scharl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* 11: 699–704.
44. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, et al. (2010) The dynamic genome of *Hydra*. *Nature* 464: 592–596.
45. Gu D, Sater AK, Ji H, Cho K, Clark M, et al. (2009) *Xenopus* -catenin is essential in early embryogenesis and is functionally linked to cadherins and small GTPases. *J Cell Sci* 122: 4049–4061.
46. Israely I, Costa R, Xie C, Silva A, Kosik K, et al. (2004) Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. *Curr Biol* 14: 1657–1663.
47. Medina M, Marinescu RC, Overhauser J, Kosik KS (2000) Hemizygoty of delta-catenin (CTNND2) is associated with severe mental retardation in cri-du-chat syndrome. *Genomics* 63: 157–164.
48. Lu Q, Dobbs LJ, Gregory CW, Lanford GW, Revelo MP, et al. (2005) Increased expression of delta-catenin/neural plakophilin-related armadillo protein is associated with the down-regulation and redistribution of E-cadherin and p120ctn in human prostate cancer. *Hum Pathol* 36: 1037–1048.
49. Arikath J, Peng IF, Ng YG, Israely I, Liu X, et al. (2009) Delta-catenin regulates spine and synapse morphogenesis and function in hippocampal neurons during development. *J Neurosci* 29: 5435–5442.
50. Keirsebilck A, Bonn e S, Staes K, van Hengel J, Nollet F, et al. (1998) Molecular cloning of the human p120ctn catenin gene (CTNND1): expression of multiple alternatively spliced isoforms. *Genomics* 50: 129–146.
51. Mariner DJ, Wang J, Reynolds AB (2000) ARVCF localizes to the nucleus and adherens junction and is mutually exclusive with p120(ctn) in E-cadherin complexes. *J Cell Sci* 113(Pt 8): 1481–1490.
52. Mo YY, Reynolds AB (1996) Identification of murine p120 isoforms and heterogeneous expression of p120cas isoforms in human tumor cell lines. *Cancer Res* 56: 2633–2640.
53. Roura S, Dominguez D (2004) Inducible expression of p120Cas1B isoform corroborates the role for p120-catenin as a positive regulator of E-cadherin function in intestinal cancer cells. *Biochem Biophys Res Commun* 320: 435–441.
54. Ohkubo T, Ozawa M (2004) The transcription factor Snail downregulates the tight junction components independently of E-cadherin downregulation. *J Cell Sci Pt*.
55. Daniel JM, Reynolds AB (1999) The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol* 19: 3614–3623.
56. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31: 89–97.
57. Zhou J, Liyanage U, Medina M, Ho C, Simmons AD, et al. (1997) Presenilin 1 interaction in the brain with a novel member of the Armadillo family. *Neuroreport* 8: 2085–2090.
58. Ho C, Zhou J, Medina M, Goto T, Jacobson M, et al. (2000) delta-catenin is a nervous system-specific adherens junction protein which undergoes dynamic relocalization during development. *J Comp Neurol* 420: 261–276.
59. Lu Q, Paredes M, Medina M, Zhou J, Cavallo R, et al. (1999) delta-catenin, an adhesive junction-associated protein which promotes cell scattering. *J Cell Biol* 144: 519–532.
60. Hirano S, Kimoto N, Shimoyama Y, Hirohashi S, Takeichi M (1992) Identification of a neural alpha-catenin as a key regulator of cadherin function and multicellular organization. *Cell* 70: 293–301.
61. van Hengel J, Vanhoenacker P, Staes K, van Roy F (1999) Nuclear localization of the p120(ctn) Armadillo-like catenin is counteracted by a nuclear export signal and by E-cadherin expression. *Proc Natl Acad Sci U S A* 96: 7980–7985.
62. Sobolik-Delmaire T, Reddy R, Pashaj A, Roberts BJ, Wahl JK, 3rd. Plakophilin-1 Localizes to the Nucleus and Interacts with Single-Stranded DNA. *J Invest Dermatol*.
63. Schmidt A, Langbein L, Rode M, Pratzel S, Zimbelmann R, et al. (1997) Plakophilins 1a and 1b: widespread nuclear proteins recruited in specific epithelial cells as desmosomal plaque components. *Cell Tissue Res* 290: 481–499.
64. Mertens S, Kuhn C, Franke WW (1996) Plakophilins 2a and 2b: constitutive proteins of dual location in the karyoplasm and the desmosomal plaque. *J Cell Biol* 135: 1009–1025.
65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
66. Salemi M, Vandamme A-M, Lemey P (2009) The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge, UK; New York: Cambridge University Press. xxvi, 723 p.
67. Morgenstern B (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* 15: 211–218.
68. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
69. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
70. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F (2004) Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20: 407–415.
71. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
72. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18: 691–699.