The Evolutionary history of YAP and the Hippo/YAP pathway

Research Article

All authors: Dror Hilman and Uri Gat

Authors Affiliations:

Department of Cell and Developmental Biology, Silberman life Sciences Institute, The Edmond Safra Campus at Givat-Ram, The Hebrew university in Jerusalem, Israel.

Correspondence:

Uri Gat, Address: Room 3-576, Dept. of Cellular and Developmental Biology, The Alexander Silberman Institute of Life Sciences, Edmond J. Safra Campus, Givat Ram, The Hebrew University, Jerusalem, 91904 Israel. Email: gatu@vms.huji.ac.il, phone: 972-2-6585920.

Key words: YAP, yorkie, hippo/YAP pathway, organ size control, evolution, co-evolution.

Running head: Evolution of YAP and the Hippo pathway
Abstract

The Hippo/YAP pathway plays an important role in animal organ size control, which it exerts by regulating tissue proliferation and apoptosis rates as a response to developmental cues, cell contact and density. With the ever increasing advance in genome sequencing and analysis tools, our understanding of the animal world and its evolution has greatly increased in the recent years. We used bioinformatic tools to study the evolution of the Hippo/YAP pathway focusing on the transcriptional co-activator YAP, which is a pivotal effector of the pathway. The aim was to establish the origin and mode of development of YAP and its pathway in the animal world. Some pathway members can be already identified in single celled eukaryotes like the yeast that have preceded multi-cellular animals. Interestingly, we can find most of the components that are present in human in the sea-anemone *Nematostella*, which belongs to a very basal group of metazoans, the cnidarians. All of the major domains of YAP have been conserved between cnidarians and mammals, and YAP can be identified even in the more basal placozoan clade. We show a very high degree of conservation in regions such as the WW and the TEAD binding domains, TEAD being the major DNA binding partner of YAP. Remarkably, we found that the location of an intron in the WW1 genomic region has been invariant along an evolutionary span of over 700 million years. We have followed the evolutionary changes in YAP and in other main components of the pathway, from the first metazoans such as sponges, described the phylogenetic relationships between the YAP genes and indicated where YAP and other components have been secondarily lost. Evidence is provided that YAP and its binding partner TEAD demonstrate strong co-evolution. This gives further support for the importance of the TEAD-YAP association. Beyond contributing to an understanding of the evolutionary history of
this pathway, we have provided insights into the “birth” of this pathway, its functions and its mode of operation in animals with different body plans, development and life styles.

Introduction

A fundamental aspect of the development of multi-cellular organisms is the ability to control the size and shape of their tissues and organs. In recent years, the abundance and wealth of genomic data and molecular biology techniques have revealed a "toolkit" of developmental genes, which served in generating the remarkable diversity of the body plans of organisms (De Robertis 2008; King et al. 2008; Degnan et al. 2009; Srivastava et al. 2010). It appears that all the major animal groups share a subset of developmental pathways. These pathways are repeatedly employed over vast spans of evolutionary time to build ever more complex forms, using similar building blocks (De Robertis 2008). One of the most intriguing phenomena of metazoan development is their ability to precisely control organ size by coordinating cell proliferation rate during embryogenesis and maturation and in certain cases during regeneration and wound healing processes (Dong et al. 2007; Kango-Singh and Singh 2009).

In the past decade, a new pathway that controls cell proliferation and cell apoptosis in response to developmental cues has been described (reviewed in (Harvey and Tapon 2007; Kango-Singh and Singh 2009; Grusche, Richardson, and Harvey 2010; Oh and Irvine 2010; Sudol and Harvey 2010; Zhao et al. 2010)). This pathway is termed the Hippo or the Hippo/YAP pathway, after some of the major components that constitute it. The pathway is composed of proteins involved in reception of signals, signaling mediators, a transcription co-activator, and executing transcription factors (see Figure 1 for a simplified scheme) (Justice et al. 1995; Xu et al. 1995;
Tapon et al. 2002; Harvey, Pfleger, and Hariharan 2003; Pantalacci, Tapon, and Leopold 2003; Wu et al. 2003; Huang et al. 2005; Kango-Singh and Singh 2009). It has been hypothesized that several upstream regulators are involved in producing growth arrest signals upon stimulation. These include: 1) The arthropod/mammalian planar cell polarity (PCP) protocadherins Fat/Ft1-4 and dachsous/DCHS. 2) The FERM-domain proteins Expanded/FRMD6/Willin and Merlin/NF2. 3) The WW-domain protein Kibra, the Ras-familly protein dRASSF/RASSF1-10, apico-basal polarity proteins Lgl/Lgl1/2, atypical protein kinase C and Crumbs/Crb1-3. The signals elicited are transmitted through the kinases Hpo/Mst1,2 and Warts/Lats1,2, with the help of the adaptor proteins Salvador/WW45 and Mats/Mob to the co-activators Yki/YAP and TAZ. These induce the growth program by activating several transcription factors and micro RNAs. The major YAP/TAZ transcriptional partners in the program, which provide the DNA-binding function, were shown to be Scalloped/TEAD1-4, but others like Runt/Runx and p73 have been reported as well (Strano et al. 2001; Strano et al. 2005; Levy et al. 2007; Vitolo et al. 2007). The phosphorylation of Yki/YAP by Warts/Lats1,2 has been found to enhance the cytoplasmic localization of Yki/YAP preventing or terminating its co-activation activity in the nucleus (reviewed in (Oh and Irvine 2010; Zhao et al. 2010)). Importantly, it is now well-established that malfunctioning of the signaling pathway which phosphorylates YAP and prevents its proliferative and anti-apoptotic activity, can lead to disproportionate organ growth and tumorigenesis, which may ultimately develop into cancer (Hisaoka, Tanaka, and Hashimoto 2002; Tapon et al. 2002).

The conservation of this pathway from mammals to insects, suggests that its components have deep evolutionary roots. Here we examined the evolution of the Hippo/Yki/YAP pathway by following the phylogeny of its core components focusing on Yki/YAP, the crucial effector
switch of the pathway. We looked for evidence for the evolutionary origin of the pathway, examined the basal form of YAP, its most conserved components, and the changes in YAP across distant phyla. Our results shed light on the source of this pathway, its major forms and modifications along animal evolution. They also may provide novel insights into why the pathway has emerged and into the modes of its evolution.

Methods

Finding orthologs for YAP pathway proteins

To find orthologs of Hippo/YAP pathway genes, we employed a blastp (Altschul et al. 1990) search of the human proteins: FAT4, DCHS1, CSNK1E, RASSF1, FRMD6, NF2 (Merlin), KIBRA, MOB1, MST1/2, LATS1/2, YAP, TAZ and TEAD1 from refseq (Pruitt, Tatusova, and Maglott 2007) (Table 1). Other pathway associated proteins were analyzed as well. These included p73, RUNX1-3, AJUBA and c-ABL, as well as the hedgehog pathway proteins, SHH, SMO, PTCH and GLI1-3. The later were used as controls. All human sequences were searched against the refseq protein data from different organisms, representing different metazoan and non-metazoan groups. All blast hits were filtered, and only sequences with blast score > 150 and sequence length > 50 were examined. We then tested each sequence using reciprocal blast search, and assigned it as an ortholog if the best hit of one blast search matched the best hit of the other (Supp. data 1). If the sequences were not found reciprocally in two genomes, we picked the sequence with the best coverage, with score>150, relative identity>30% and relative similarity > 40%. If the parameters were lower, we assigned it as “non ortholog”. When we did not find an ortholog, we verified the lack of orthologous sequences with tBlastX against the genomic
sequence and EST libraries of the relevant organism. The existence or absence of an ortholog was also checked by blast search of the protein sequence in the Ortho-MCL4 database of orthologs (Li, Stoeckert, and Roos 2003; Chen et al. 2007). Sequence alignment of several Hippo/YAP pathway components is illustrated in Supp. data 2.

**Sequence alignment and phylogenetic analysis**

Sequence alignments were performed on orthologous proteins using MUSCLE (Edgar 2004) with default parameters (Supp. data 2), which were manually curated if necessary. Conservation of motifs and phosphorylation sites were checked by hand using data from Entrez gene (Maglott et al. 2007) and Uniprot (Consortium 2010). Phylogenetic analysis was conducted using PhyML (Guindon and Gascuel 2003; Guindon et al. 2005) and FastTree (Price, Dehal, and Arkin 2010) for ML analysis, with default parameters. Neighbor joining (NJ) trees were built using MEGA4 (Kumar et al. 2008). Phylogeny support was verified with the bootstrap consensus tree inferred from 1000 replicates. In NJ trees, the evolutionary distances were computed using the Poisson correction method and are given in units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution using default parameters. Positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons so as to keep all major protein domains in the alignment.

**Testing co-evolution**

To assess the strength of co-evolution between two proteins, we followed the methodology described in (Pazos et al. 2005; Sato et al. 2006). Pairwise distances between sequences in different taxa were measured in MEGA4 with pairwise deletion and Poisson correction was used for amino-acids substitution. Distances between different proteins in diverse organisms were
plotted, and Pearson’s correlation coefficient (product-moment r) was used to determine the overall correlation between distances of two genes (Supp. data 3).

**Retrieving genomic data**

All of the genomic data used in this analysis were retrieved from Entrez genome and Entrez nucleotide databases. Protein sequences were retrieved from NCBI Refseq database. Orthologs for comparing YAP and TAZ were obtained from OrthoMCL 4 DB to retrieve a large group of vertebrate orthologs.

**Results and Discussion**

**YAP proteins – a metazoan novelty**

Detecting the evolutionary origin of genes that belong to specific pathways may give insights into how they have contributed to the formation of novel structures and mechanisms in animals over the course of evolution. YAP orthologs have previously been explored almost exclusively in higher vertebrates and in *Drosophila*. We set out to find the origin of YAP and the Hippo/YAP pathway in the tree of life (Figure 2). We first explored the YAP protein, a critical connecting switch between the signaling pathway and the transcriptional program. We could not detect a *bona fide* YAP gene in the group that is considered to represent the most basal or “ancient” metazoans – the sponges. Searching for conserved regions of YAP in the demosponge *Amphimedon queenslandica* whole genome shotgun sequencing (Srivastava et al. 2010) and in the homoscleromorph sponge *Oscarella carmela* EST database (Nichols et al. 2006; Hemmrich and Bosch 2008) did not yield an obvious ortholog. Actually, proteins with two conserved WW
domain proteins in sponges and in the non-metazoan eukaryotes choanoflagellates *Monosiga brevicollis* (King et al. 2008), usually aligned best with the human gene WWP1 - E3 ubiquitin ligase, which seems to be conserved in nearly all eukaryotes we tested. Searching further, YAP orthologs were not found in yeasts (*Saccharomyces cerevisiae*) or in plants (*Arabidopsis thaliana*). The most basal YAP ortholog we found is that of the very basal metazoan, the placozoan *Trichoplax adhaerens*. Its existence in this organism dates the first appearance of YAP to after the emergence of the sponges in the metazoan clade, up to 1.8 billion years ago (Nichols and Wörheide 2005) (Table 2). We then proceeded to follow YAP’s evolution in the metazoans. Interestingly, we found that when compared to human YAP, the highest percentage of blast positive positions among non-vertebrates, was found in the sea-anemone *Nematostella vectensis* YAP (58% similarity). *Nematostella* belongs to the basal metazoan group of the cnidarians, estimated to have separated from the bilaterian ancestor 1.2 – 0.7 billion years ago (Otsuka and Sugaya 2003; Han et al. 2010). The *Nematostella* ortholog is more similar to human YAP than the *Drosophila* Yki (45% similarity) despite the much greater phylogenetic distance of cnidarians, as compared to arthropods, from humans (see Figure 2) (Otsuka and Sugaya 2003; Pisani et al. 2004). The least conserved YAP genes found are the putative YAP proteins of *Strongylocentrotus purpuratus* and *Trichoplax*, with 28%/29% conservation, respectively (Table 2). Surprisingly, the YAP protein seems to be absent in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*, since searching for YAP orthologs in their genomes retrieved sequences that aligned with the aforementioned WWP1. Since nematodes are animals that evolved after the separation of bilaterian animals from cnidarians (Hejnol et al. 2009) (Figure 2), it is likely that a functional loss of YAP occurred in the nematode lineage. The TAZ proteins, which are structurally similar to YAP and are regulated by the pathway in the same fashion, are
considered to be YAP paralogs and represent a late evolutionary novelty. They were not found in non-vertebrates.

**Structural comparison of YAP protein orthologs**

Next we explored changes in YAP protein structure during the course of evolution of the YAP family. The full alignment of the YAP protein sequences we retrieved is presented in Supp. Figure 1. Figure 3 shows a schematic alignment of representative YAP orthologs in a wide array of metazoans. The most noticeable YAP conserved domains are the two WW domains and the TEAD interaction domain. These are conserved from placozoans to the late bilaterians like human. The *Trichoplax* YAP seems to have only one WW domain, which aligned best with the second WW domain of vertebrates and insect YAPs. The WW2 domain is missing in TAZ proteins which contain only the first WW domain (WW1).

Another conserved region is the C-terminal trans-activation domain. It seems to consist of several separated domains and its composition varies between different clades (Figure 3). As previously shown (Oka and Sudol 2009), a conspicuous conserved PDZ-binding motif can be found in the most distal sequence of the C-termini of cnidarians, sea-urchins and chordates. However, in insects this sequence is divergent (Figure 3). The coiled-coil (c-c) domain which is important for TAZ and YAP dimerization (Murakami et al. 2005) is also conserved in vertebrates and cnidarians. The c-c domain in the insects *Tribolium castaneum* and *Apis mellifera* is present partially and it is completely absent in *Drosophila* Yki. This domain is highly divergent in *Ciona intestinalis* (Figure 3, Supp Figure 1) suggesting an altered function with respect to its dimerization and interaction activities. To conclude the analysis of the domain structure, we observed that the proline rich region in the N-terminal area of mammalian and
avian YAP is missing in other taxa (Figure 3). This indicates a yet unidentified function that arose late in evolution.

YAP activity was shown to be regulated by phosphorylation and indeed several phosphorylation sites found in humans are highly conserved in all metazoans tested, especially the five sites (S61, S109, S127, S164, S397) (Zhao et al. 2007; Hao et al. 2008) (Figure 3). These are known to be phosphorylated by the Lats kinases in mammals and three Warts kinase sites in Drosophila. The phosphoserine sites that are located prior to the WW domains are known to be important for Hpo/Wts-mediated inhibition in Yki and YAP and this is likely reflected in the high degree of conservation of the regions in which they reside (Figure 3 and Supp. Figure 1). Mammalian YAP tyrosine phosphorylation was shown to promote apoptosis in response to DNA damage (Levy et al. 2008). However, the tyrosine that is involved, Y391, seems to be conserved only in vertebrates. This tyrosine is aligned with phenylalanine in other deuterostomes (C. intestinalis, S. purpuratus) and is absent in protostomes (see Supp. Figure 1).

**The genomic composition of YAP genes in metazoans**

A genomic analysis of the YAP gene family demonstrated the overall increased complexity of YAP gene structure over evolution (Figure 4). The basal YAP genes have 2 or 3 exons, while the later forms have 8-9 exons. The second WW domain appears to reside in a distinct exon in all bilaterians, which can lead to differential splicing isoforms that contain only the WW1 domain. In both human YAP and Drosophila Yki there are isoforms documented that lack the second WW domain (Figure 4). These isoforms may mimic the ancient condition of only one WW domain in the Placozoans. The vertebrate TAZ (WWTR1) contains one WW domain, which is better aligned with the first WW domain of YAP (Figure 3, 4, Supp. Figure 1).
All of the YAP (and TAZ) genes that we analyzed, except for the *T. adhaerens* gene which does not have WW1, contain a remarkably conserved intron within the WW1 coding region (Figure 4, 5). The intron boundaries fall exactly in the same codon, which codes for a conserved asparagine (N) residue and always start in the same phase (+2) of this codon. The most diverse form of this intron boundary can be observed in the sequence of *S. purpuratus*. Here, the intron is contained in an aspartic acid codon and differs from the chordate and the arthropod clades sequences in several other positions around the intron boundaries (Figure 5A). The *N. vectensis* sequence, on the other hand, shares some features with chordate sequences and some with arthropods around the intron boundaries. This may reflect an ancient state, before the separation of the major bilaterian groups. The extreme conservation of this intron strongly suggests that the same YAP structure appeared in a eumetazoan ancestor of both cnidarians and all bilaterians. Why is this intron so well conserved over such a great span of evolutionary time? The reason is not known, but we speculate that the intron harbors an important binding site for some as yet undetermined critical factor which conferred strong conservation to its proximity. Another hypothesis may be that without the presence of this intron, this WW domain may be prone to duplication by a reverse-transcription based mechanism. The result may be evolutionary unfavorable.

**Phylogeny of YAP and TAZ proteins**

In order to examine the phylogenetic relations of the YAP family orthologs we have utilized several tree-building techniques. Both Neighbor joining (NJ) and Maximum likelihood (ML) trees, which are considered reliable for the analysis of distant sequences, were able to distinguish between three major groups, arranging the chordates, arthropods and early non-bilaterians in separate clades (Figure 6 A, B). In the ML tree the non-bilaterians *T. adhaerens* and *N. vectensis* form a clade of their own and do not form an outgroup to the bilaterians. The NJ analysis,
however, positions the *T. adhaerens* YAP as the most basal, being an outgroup to *N. vectensis* and all bilaterians. *N. vectensis* YAP, on the other hand, clustered as an outgroup to the bilaterians as could be expected from its structure. Surprisingly, the deuterostomate *S. purpuratus* tended to group closer to the protostome clade in both trees, but with low bootstrap support. Since some parts of the *S. purpuratus* YAP (spr-YAP) display more similarity to chordates and other parts to arthropods, the Spr-YAP sequence likely represents an early bilaterian form, different from both the chordates and the arthropods, as it has emerged before the separation of these major bilaterian clades. In both NJ and ML trees, the *C. intestinalis* YAP falls close to the root of the chordate clade, but with a long branch. This suggests strong divergence of its sequence after the urochordate and chordate lineage separation. A similar phenomenon can be observed in insects, where the *Drosophila* Yki sequence seems to be more diverse (Figure 3, Supp. Figure 1) and probably represents a more varied and novel sequence when compared to the more basal-like insect groups which were included in the analysis.

We went on and analyzed the later evolutionary radiations of the YAP family of proteins. The origin of the YAP paralogous gene TAZ was explored by aligning the multiple TAZ and YAP proteins that we retrieved from the databases and generating their neighbor joining cladogram (Figure 6C). The resulting plot clusters the vertebrates' YAP and TAZ proteins in different clades. Both, however, have similar distance from out groups of vertebrates. This suggests that the TAZ gene probably originated from gene duplication of YAP in early vertebrates. The conserved nature of the WW1 domain in TAZ and the identical localization of the intron in it, also alludes to a common origin. We did not find TAZ in chordates such as the cephalochordate *B. lanceolatum* or in the Urochordate *C. intestinalis* genomes. This dates the YAP/TAZ duplication to a late stage in chordates evolution, possibly associating it with the whole genome
duplication in the ancestral vertebrate lineage (Dehal and Boore 2005). A perplexing question is the reason for the emergence of the TAZ paralogs, especially since a single WW1 harboring splicing isoform seemingly of the same structure (Figure 4), exists in both vertebrates and Drosophila. Future experiments exploring the exact function of TAZ, as opposed to YAP with its splicing forms, may shed light on the rationale for this late evolutionary development and on its benefit for vertebrates.

**YAP and the Hippo/YAP pathway – developing complexity**

What is the evolutionary status of YAP in the context of the entire pathway? How did the pathway as we know it evolve? Did the components appear as a full set along with a novel animal group, or did they emerge gradually in several steps to establish and elaborate the pathway? In order to address these questions and to learn about the evolutionary history of the pathway, we searched for the orthologs of the core pathway components as well as the upstream members of the pathway in representative single celled eukaryotes and in metazoan groups (Table 1).

**Single celled Eukaryotes.** Figure 7 depicts the composition of the pathway along evolution from eukaryotes to mammals. We can find *bona fide* orthologs of TEAD and Mob in the yeast *S. cerevisiae*, which may represent the most ancient orthologs of the pathway. We did not find them in prokaryotes. Moreover, YAP orthologs were not found in plants, reducing the possibility of pathway loss during the early evolution of eukaryotes in general. Previously, it has been suggested that the roots of the pathway may be represented by the mitotic exit network (MEN) and the separation initiation network (SIN) pathways in the budding and fission yeast correspondingly (Bardin and Amon 2001; Harvey and Tapon 2007). Our results cannot,
however, categorically identify the yeast Cdc15 and Dbf2 as orthologs of Hippo and Warts, respectively, as previously suggested, since they tend to align better with other paralogs of these kinases. That said, the similarity of the genetic module and the genes involved, along with the similar interaction with Mob like protein in yeast, suggest a deep evolutionary connection that warrants further examination. Warts/Lats has been reported to be localized to the centrosome in interphase, translocate to the mitotic spindle in metaphase and anaphase and then to localize to the midbody in telophase in human cells (Nishiyama et al. 1999; Morisaki et al. 2002). Additionally, embryonic cells from Lats2 knockout mice display multiple mitosis defects including abnormal exit from mitosis (Yabuta et al. 2007). As this expression pattern and function is reminiscent of that of Dbf2 in yeast (Bardin and Amon 2001), it may represent an ancient role for the pathway in mitotic exit which has remained in metazoans up to humans (Praskova, Xia, and Avruch 2008)

Tec1, the putative yeast TEAD ortholog, regulates signaling of one of the mitogen-activated protein (MAP) kinase pathways in the budding yeast, a pathway that is involved in invasive growth (Wang et al. 2009). We could not find a direct connection between the function of the above MEN factors that are related to the Hippo/YAP pathway and that of TEAD/Tec1. However, the founding member of this family of kinases, Ste-20 in yeast (Radu and Chernoff 2009), is part of the MAP kinase network that modulates Tec1 activity. It is thus possible that Ste-20 is the ancestral Mst and that the origin of the pathway is related to the Tec1 regulated invasive response. In conclusion, there are several candidates in yeast for elements of the current Hippo/Yap pathway. These components may have joined to evolve the metazoan form of the pathway. In addition, the pathway may have consolidated along metazoan evolution due to
newly acquired roles of these progenitors as well as to the emergence and integration of new pathway members.

Analysis of the genome of the unicellular eukaryote choanoflagellate *M. brevicollis*, considered to be the closest relative of metazoans, did not yield any more orthologs of the core pathway. However, upstream elements such as the protocadherins fat/ds may have first emerged in choanoflagellates (King, Hittinger, and Carroll 2003; King et al. 2008). This would have allowed colony formation and further on in evolution they could have been utilized for the more intricate and complex tasks of tissue growth control, which necessitated formation of the Hippo/YAP pathway by recruitment of other pre-existing factors and incorporation of novel elements.

**Sponges and placozoans.** Moving along into the metazoan clade, we can identify a true ortholog of Mst/Hippo kinase in sponges and all groups beyond, but still no Sav, Lats or YAP (Figure 7). A hallmark in the evolution of this pathway is the appearance of the dynamic effector YAP. The most basal YAP ortholog we have detected is that of the placozoan *T. adhaerens*, which is a eumetazoan with simple but organized tissues (Srivastava et al. 2008). This YAP ortholog includes the TEAD-binding domain and four out of five of the Lats/Wts phosphorylation sites, but only one WW domain, which is similar to WW2 (see Figure 3 and Supp. Figure 1). Thus this YAP is different in structure from all of the following orthologs. Taking into account the lack of other components of the pathway in this organism, such as the Lats/Wts kinase, this suggests that the pathway in the placozoans is still in a primordial form and likely functions differently from that in bilaterians. It will be interesting to find out the mode of action of this rudimentary and minimal pathway in this little studied organism.
The sea-anemone *Nematostella*. The cnidarian *N. vectensis* is considered to be the closest outgroup to the bilaterians among the known extant taxa (Finnerty et al. 2004; Dunn et al. 2008; Hejnol et al. 2009; Sperling, Peterson, and Pisani 2009; Ryan et al. 2010). Our analysis clearly shows that this group of basal non-bilaterians contains most of the components of the “modern” Hippo/YAP pathway (Figure 7). This includes the complete core Hippo machinery and all of the supposed upstream signaling proteins except for the ortholog of Expanded which we could not find. Notably, the *Nematostella* YAP demonstrates almost all the features recognized in many later bilaterian groups. In addition to the conserved elements that we described above for *T. adhaerens*, these include two well conserved WW domains and a PDZ-binding C-terminal pentapeptide identical to the human form (see Figure 3, Supp. Figure 1). Intriguingly, when compared to different bilaterian groups the *Nematostella* YAP shows the highest similarity to the vertebrate YAP (see Table 2 and Supp. Figure 1). This high level of resemblance between *Nematostella* and vertebrates has been demonstrated before for many developmental genes (Ryan et al. 2007; Hejnol et al. 2009; Saina et al. 2009). It has been explained by the basal nature of both the sea-anemone genome and the vertebrate genome, which is much less derived than many of the protostome groups like insects and nematodes. The question then arises whether this similarity in protein structure is also reflected in a similar *modi operandi* among these very distant animal groups.

**The Bilaterian groups – conservation and divergence.** When we considered the status of the pathway within the bilaterian groups, for which genomic data is available, we generally found all the components that were established earlier. However, as in other major developmental pathways, the complexity increased and we were able to detect the appearances of additional factors and more paralogs added to the pre-existing pathway components. Thus, the *Mst, Lats,*
TEAD and YAP genes underwent duplication events during deuterostome evolution (Figure 7). This may have conferred more flexibility to the pathway as the various paralogs could allow modulation of function in different organs. Our phylogenetic analysis points out that the ancient YAP sequence diverged to two main groups of proteins: the arthropod Yki and the deuterostome YAP, which includes the chordates. The sequence alignments suggest that the arthropod Yki forms are much more divergent and harbor many more changes, when compared to the chordate YAP species which are more “basal” and resemble the sea-anemone ancestral YAP. Several interaction elements that are important in vertebrate YAP/TAZ, such as the SH3 binding domain (Espanel and Sudol 2001) and the Y391 like tyrosine (Levy et al. 2008), are absent in insects and early metazoans. This may point to the generation of novel functions of YAP during deuterostomes and vertebrate evolution. The new phosphorylation site that appeared on tyrosine can be associated with more elaborate apoptosis promotion abilities for which vertebrates YAPs were enlisted (Levy et al. 2008). Interestingly, while the site is a vertebrates novelty, the kinase responsible for the phosphorylation, c-Abl, is an ancient metazoan gene which is already found in the sponge.

Inspection of the Drosophila Yki shows that indeed, its structure is different in that the C-terminus is much shorter and is missing many sequences attributed to transcription activation. It also entirely lacks the coiled-coil domain found in deuterostomes and in a partial form also in other insects. Moreover the PDZ binding domain, which was recently shown to be critical for the function of YAP in vertebrates (Oka and Sudol 2009), is highly modified in Yki. The divergent nature of Yki likely indicates that a significant change in the gene function has transpired during arthropods evolution. This change may have occurred in order to accommodate a unique form of embryonic development such as the whole body segmentation mode typical to Drosophila.
Unlike many other insect and arthropod groups flies lack a growth zone that gradually gives rise to posterior segments during its embryogenesis.

The apparent lack of obvious YAP orthologs in the nematodes *C. elegans* and *C. briggasae* and the lack of other pathway components (Ex and Kibra, Figure 7), is another dramatic example of a major change in the pathway in a protostome group. A similar phenomenon of the absence of crucial and conserved developmental genes in nematodes was also previously described (Aboobaker and Blaxter 2003a; Aboobaker and Blaxter 2003b), wherein the authors portrayed the disorganization and decay of the Hox gene cluster in *C. elegans*. The Hpo/Mst orthologs Cst-1/2 in *C. elegans* are Ste-20 like kinases, which were shown to increase nematode lifespan by phosphorylating Daf-16, the mammalian FoxO ortholog which is a known substrate of Mst kinases (Lehtinen et al. 2006). The *C. elegans* Lats/Warts, Ce-Wts-1, is important to various developmental functions in *C. elegans* including body length control by the TGF-beta Sma/Mab pathway (Cai et al. 2009). It is also associated with lifespan control (Curran and Ruvkun 2007), while the putative TEAD ortholog, egl-44, is involved in mechanosensory neuron cell fate determination (Wu, Duggan, and Chalfie 2001). Further experimental data are required in order to establish the function of these and the remaining Hippo/YAP pathway elements in nematodes and to reveal the changes that the pathway has undergone in this derived group.

It will also be interesting to explore the status of the pathway in other protostome groups such as mollusks as more sequence data become available.

**Co-evolution of genes in the Hippo/YAP pathway**

We were interested in testing whether there is any evidence for co-evolution between components of the pathway, which may indicate the evolutionary origin of a functional
association between them. In recent studies an alignment of the YAP binding domain of TEAD1 and the TEAD binding domain of YAP, along with the solved structure of YAP and TEAD1 (Chen et al. 2010; Li et al. 2010), showed the putative co-evolution of these areas in YAP and TEAD1. Specifically, residues in the TEAD binding domain of YAP, and in the TEAD partner protein, which physically interact in the 3D model, are also highly conserved within metazoans (Li et al. 2010). To test whether YAP and TEAD have co-evolved we used phylogenetic evidence to compare the calculated pairwise lengths between all YAP orthologs and TEAD orthologs across distant taxa. This method was used in earlier work to demonstrate strong protein-protein interactions between two proteins (Pazos et al. 2005). The principle of this analysis is that in order to assess evolutionary connection, one examines how the evolution of one sequence, which is measured by the pairwise distances of any two orthologs, is correlated to the distances of the orthologs of another gene. We expect that if two genes interacted and thus co-evolved, we will observe a correlation between the pairwise distances (Supp. data 3).

Our co-evolution analysis indeed suggested a very strong correlation between YAP and TEAD pairwise distances, with a correlation coefficient $r = 0.85$, $p$-val = $1.1 \times 10^{-16}$ (Figure 8A). This strongly supports a functional association between these two proteins. Such co-evolution was also shown in the transcription-factor and co-activator pair Runx-CBFβ (Sullivan et al. 2008). This co-evolution suggests a strong selection pressure toward the interaction of YAP and TEAD, and strengthens the hypothesis that TEAD is a key downstream effector of the pathway (Ota and Sasaki 2008; Li et al. 2010). YAP can interact with other transcriptional partners and thus we analyzed its co-evolution with the aforementioned Runx and p73 co-factors. While YAP seems to be highly correlated to p73 ($r = 0.85$, $p$-val=$6.2 \times 10^{-7}$, Supp. data 3), which suggests that they strongly interacted during evolution, there was no indication of co-evolution with the Runx
proteins. This in turn could indicate a recent interaction, or that the interaction contact regions are very short.

Extending the analysis we showed that YAP/Yki exhibit a high degree of co-evolution with Mob1/Mats (Figure 8B) and Sav1 (Figure 8C) and a lower but positive correlation with Lats1/Wts (Figure 8D) and Mst2/Hpo (Figure 8E). All of these function in a close signaling network (Pan 2010). The lower correlations seen for the later (Figure 8 D, E) could be due to the multiple other partners of these kinases and of YAP. Some of the other components in the pathway that are known or suspected to interact with one another also demonstrate a positive correlation such as Mob1-Mst2 (Wei, Shimizu, and Lai 2007) (Figure 8F), Mob1-Lats1 (Lai et al. 2005; Wei, Shimizu, and Lai 2007) (Figure 8G), while Mer-Ex display a high correlation (McCartney et al. 2000; Hamaratoglu et al. 2006; Yu et al. 2010) (Figure 8H). As could be expected not all previously known binding partners showed co-evolution in this assay, while other pairs of proteins, such as YAP/TAZ and Merlin, exhibited unexpected high correlation, which could indicate that they may physically interact (see Supp. Data 3). Other pairs of proteins that are not known to bind or interact show a series of distances with much lower correlation. For example, the predicted correlation between YAP and FAT4 yielded an $r = 0.27$ (Figure 8I) and with DCHS1, $r = -0.02$ (8J). The correlation between FAT4 and DCHS1, that do interact in *Drosophila* (Cho et al. 2006), is much higher with $r=0.65$, (Figure 8K). β-catenin, another transcriptional co-activator that acts with the same operational logic as YAP, but in a different pathway, yields an $r= 0.10$ for YAP (Figure 8L).

To further test this technique on another signaling pathway, we analyzed the main components of the hedgehog pathway, which is also composed of an assembly of ancient eukaryotic and novel metazoan genes (Matus et al. 2008). We found an $r = 0.68$ between Hedgehog and its
receptor, the tumor suppressor Patched1, (PTCH1), an \( r = 0.32 \) of PTCH1 to the smoothened receptor and an \( r = 0.52 \) of PTCH1 to the GLI2 transcription factor of this pathway (Supp. data 3). These correlations point to a general trend of co-evolution of interacting components in developmental pathways.

Overall, our data demonstrate the deep evolutionary connections between the components of the Hippo/YAP pathway, which are manifested in their co-evolution and are presumably necessary to maintain the functionality of the network. In addition, we show the potential of this approach for defining possible protein-protein interactions that were previously not explored.

**Hypotheses regarding the origin and evolutionary role of YAP and the Hippo/YAP pathway**

From our evolutionary survey we hypothesize that the pathway as we know it in human and in *Drosophila*, with the majority of its associated machinery, can be dated to the last common ancestor of cnidarians and bilaterians, estimated to exist more than 700 million years ago (Otsuka and Sugaya 2003; Hejnol et al. 2009; Han et al. 2010). The finding of other components of the pathway that have much broader clusters of orthologs in distant phyla, well before the advent of metazoans estimated up to 1.8 billion years ago (Nichols and Wörheide 2005), indicates that the pre-existing segments of this pathway likely played roles both in mitotic regulation processes like mitotic exit and in environmental signaling. Moreover, they were redirected to add multicellular tissue growth control tasks during metazoans body plan evolution. This new recruitment was necessary to perform novel regulatory functions. These, in turn, rewired or modified some existing components, eventually adding new components such as YAP. This paradigm of using preexisting proteins and incorporating components with a new
combination of domains, or recruiting altogether novel domains on top of existing ones as well as the deployment of totally new proteins, is increasingly recognized as the “evolutionary workshop”. It created the new pathways (e.g. the Wnt pathway (Adamska et al. 2010) and the modified functional complexes (e.g. adhesion junctions (Nichols et al. 2006)), necessary for emergence of new life forms such as the metazoans and bilaterians (Srivastava et al. 2010).

While TEAD and Mob are ancient eukaryotic proteins, YAP seems to be a dynamic metazoan novelty in which an existing WW domain protein was modified and combined with novel domains such as the TEAD binding and the PDZ interacting domains. What is interesting about the evolution of YAP is that it has likely emerged to form this pathway in ancestral eumetazoans like the placozoans/cnidarians/bilaterian progenitors, which may suggest that the pathway is necessary for the growth control of more complex organized tissue as compared to sponge cells. In addition, YAP structure has been modified along the course of evolution with the putative addition of a second WW domain and the PDZ binding domain in cnidarians. Later, more minor alterations of the scheme appeared, such as diversification of the PDZ binding pentapeptide in insects and the formation of the TAZ paralogs in the vertebrates.

What new insights can be gleaned from our evolutionary study about the function of the Hippo/YAP pathway? The apparent lack of a YAP ortholog and several other pathway components in the ecdysozoan C. elegans demonstrates that the full pathway may not be essential for all types of metazoan life forms. The variant insect forms suggest that the pathway had to be modified and diversified for some others. The absence of YAP may indicate a dramatic change in the developmental mode of nematodes towards an early “deterministic” pattern (Schierenberg 2001). In this mode, which is the classical alternative to the “regulative” mode of development, a signaling pathway that controls cell growth in response to environmental cues
may no longer be needed. Thus we speculate that this pathway has primarily evolved in order to regulate tissue and organ growth during metazoan “regulative” phases of development, which are flexible and rely on cell-cell interaction and environmental signals. However, the full complement of the pathway may not be essential for the “determinative (mosaic)” like type of development which is preprogrammed in the developing organism. Exploration of the Hippo/YAP pathway’s function in developmental stages which exhibit these types of development, both of which often appear at distinct stages in the development of model organisms may show if indeed our speculation is indeed valid.

The Hippo/YAP pathway’s evolutionary tale may thus provide insight into the evolution of tissue and organ size control in multi-cellular animals. It features YAP as one of the most important tools in the metazoan developmental toolbox.

Conclusions

We explored the origin and evolution of the Hippo/YAP pathway, which controls tissue proliferation rates and organ size in animals. Some of the pathway’s elements were found already in protists like yeast and the choanoflagelates, while the first YAP progenitor was detected in the placozoan *T. adhaerens*. The full pathway as known in man can be first identified in the basal metazoan cnidarian *N. vectensis*. Following the evolution of the pathway along the bilaterian groups shows an overall increase in its complexity, such as the emergence of additional paralogs of the core kinases and the advent of the TAZ genes, the vertebrate YAP paralogs. However, some bilaterians, such as *Drosophila* show divergence of the YAP effector. This may reflect a unique mode of their embryonic development pattern. At the same time other
groups, like the nematode *C. elegans*, show a loss of YAP and several other components, which may hint at their simplified mode of cellular growth control.

Thus in reconstructing the origin of this pathway we can envision the following scenario: The evolution of organized tissues and organs in metazoans required the invention of a mechanism for tissue growth regulation. For this purpose a new genetic network that can connect inter-tissue signals with cellular proliferation control was necessary. This pathway has evolved in accordance with the developmental growth control programs of the various animal forms. Our findings indicate that the emergence of the “advanced” Hippo/YAP pathway is associated with the appearance of an organized body plan which incorporates distinct organs and tissues. Even though the pathway is highly conserved, and its genes have undergone strong co-evolution, it is adapted to the developmental program of the organism and shows plasticity that is reflected in divergence of some critical domains of its components.

**Acknowledgments**

The authors would like to thank Drs. Ariel Chipman, Liran Carmel, Rachel Green and Marshall Devor of the Hebrew University in Jerusalem for their critical reading of this manuscript and helpful suggestions. This study was supported by the Israel Science Foundation grant 825/07.
## Tables and figures

Table 1. The Hippo/YAP pathway protein components studied in this article.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Human Refseq acc.</th>
<th>Drosophila Refseq Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCHS1 /ds</td>
<td>NP_003728.1</td>
<td>NP_523446.2</td>
</tr>
<tr>
<td>FAT4/ft</td>
<td>NP_078858.4</td>
<td>NP_477497.1</td>
</tr>
<tr>
<td>CSNK1/Dco</td>
<td>NP_001885.1</td>
<td>NP_733414.1</td>
</tr>
<tr>
<td>RASSF1/Rassf</td>
<td>NP_009113.3</td>
<td>NP_651126.3</td>
</tr>
<tr>
<td>NF2 / Merlin (Mer)</td>
<td>NP_000259.1</td>
<td>NP_523413.1</td>
</tr>
<tr>
<td>WWC1 / Kibra</td>
<td>NP_001155133.1</td>
<td>NP_001034055.1</td>
</tr>
<tr>
<td>FRMD6 / expanded (ex)</td>
<td>NP_001035946.1</td>
<td>NP_476840.2</td>
</tr>
<tr>
<td>Mst1/2 / Hippo</td>
<td>NP_006273.1 / NP_006272.2</td>
<td>NP_611427.1</td>
</tr>
<tr>
<td>Salvador (Sav)</td>
<td>NP_068590.1</td>
<td>NP_788721.1</td>
</tr>
<tr>
<td>Mob1/Mats</td>
<td>NP_775739.1</td>
<td>NP_651041.3</td>
</tr>
<tr>
<td>Lats1/2 / Warts</td>
<td>NP_004681.1 / NP_055387.2</td>
<td>NP_733403.1</td>
</tr>
<tr>
<td>YAP / TAZ / Yorkie</td>
<td>NP_001123617.1 / NP_001161750.1</td>
<td>NP_001036568.2</td>
</tr>
<tr>
<td>TEAD1 / scalloped (sd)</td>
<td>NP_068780.2</td>
<td>NP_001096989.1</td>
</tr>
</tbody>
</table>

Table 2. YAP orthologs in various metazoans and their relation to human YAP.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Ortholog (refseq)</th>
<th>Length (aa)</th>
<th>Similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trichoplax adhaerens</td>
<td>XP_002108065.1</td>
<td>447</td>
<td>29%</td>
</tr>
<tr>
<td>Nematostella vectensis</td>
<td>XP_001627495.1</td>
<td>412</td>
<td>58%</td>
</tr>
<tr>
<td>Tribolium castaneum</td>
<td>XP_970492.1</td>
<td>404</td>
<td>53%</td>
</tr>
<tr>
<td>Apis mellifera</td>
<td>XP_391844.2</td>
<td>511</td>
<td>45%</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>NP_001036568.2</td>
<td>395</td>
<td>45%</td>
</tr>
<tr>
<td>Strongylocentrotus purpuratus</td>
<td>XP_789542.2</td>
<td>531</td>
<td>28%</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
<td>XP_002130260.1</td>
<td>529</td>
<td>46%</td>
</tr>
<tr>
<td>Branchiostoma floridae</td>
<td>XP_002595229.1</td>
<td>86 (partial)</td>
<td></td>
</tr>
<tr>
<td>Danio rerio</td>
<td>NP_001132952.1</td>
<td>442</td>
<td>88%</td>
</tr>
<tr>
<td>Taeniopygia guttata</td>
<td>XP_002198083.1</td>
<td>506</td>
<td>88%</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>NP_001123617.1</td>
<td>504</td>
<td>100%</td>
</tr>
</tbody>
</table>
**Figure legends**

**Figure 1.** Model of the core Hippo/YAP pathway.

The Hippo/YAP pathway controls tissue growth by regulating the sub-cellular localization of Yki/YAP/TAZ, which, when nuclear, associate with the transcription factors Sd/TEAD1-4 activating a transcriptional program that drives cell growth and proliferation and inhibits apoptosis. Other transcriptional partners such as p73, Runx and Smads have been described as well. Several upstream inputs regulate the activity of the core components Salvador (Sav), Hpo/Mts, Wts/Lats and Mats/Mob. These components control Yki/YAP/TAZ phosphorylation and localization. Rassf modulate the Mst/Hippo kinases activity. The protocadherin Ft/FAT signals to the core components by influencing the localization and levels of Expanded (Ex/FRMD6) that forms a complex with Merlin (Mer) and Kibra to facilitate and activate the Hpo/Sav/Mats/Wts complex. Fat may inhibit Dachs, which represses Lts/Warts protein levels. Fat activity is controlled by its ligand Dachsous (Ds), by the kinases Four-jointed (Fj) and Discs overgrown (Dco), and by Low Fat (Lft), Approximated (App) and Dachs (Ds).

**Figure 2.** The evolutionary tree of taxa studied in this work.

The gross estimated divergence time ranges of the major clades are mentioned near each group name. Divergence times are based on: (Otsuka and Sugaya 2003; Pisani et al. 2004; Nichols and Wörheide 2005; Hejnol et al. 2009a; Han et al. 2010)

**Figure 3.** Schematic alignment of YAP/TAZ orthologs.

The structure of the YAP proteins in the different taxa along evolution is displayed. Conservation level between the organisms shown is color coded as depicted below. Missing
areas in the alignment are blank. Domains and phosphorylation sites are marked in the figure. Protein sizes are shown to the right of each sequence.

**Figure 4.** Genomic structure of YAP in representative metazoans.

Major domains are marked with colors. The transcript coding region is drawn to scale and introns positions are illustrated but their size is not drawn to scale.

**Figure 5.** A conserved intron in the WW1 domain.

An intron is found in most YAP and TAZ orthologs within the WW1 domain. The intron position and phase inside the codon for ‘N’ is highly conserved. The illustration shows the genomic sequence flanking the intron boundaries in different organisms as well as the amino acids coded by this region.

**Figure 6.** Phylogenetics of YAP.

Maximum likelihood (A) and Neighbor joining (B) unrooted trees of YAP orthologs from distant metazoans. All sequences were downloaded from refseq, and aligned with MUSCLE. Trees were constructed as described in methods. The bootstrap consensus tree inferred from 1000 replicates. The percentage of replicate trees in which the associated ortholog clustered together in the bootstrap test is shown next to the branches. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. In the NJ tree, the distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a
gamma distribution. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. There were a total of 796 positions in the final dataset.

(C) The TAZ proteins are paralogs of vertebrate YAPs. The cladogram demonstrates the relations between the YAP and TAZ paralogs, showing that TAZ is clustered closer to the vertebrates YAP than to the invertebrates’ ortholog. The high similarity of vertebrate and Nematostella YAP (Table 2, Figure 3) relative to Drosophila Yki, puts Nematostella, closer to the vertebrates, when compared mainly with vertebrates proteins. The lack of other insects in this analysis artificially clusters Drosophila outside the bilaterians and not as presented in A and B. The cladogram was inferred using the Neighbor-Joining method. YAP and TAZ sequences were downloaded from the orthoMCL database. The tree was built with MEGA4 using default parameters. The optimal tree is shown and the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogeny.

Figure 7. Summary of the Hippo/YAP pathway evolution.

The figure shows the conservation of the core components of the Hippo/YAP signaling pathway in representative metazoans. Orthologs between different organisms are placed in the same relative position. The colors code for the proposed antiquity of the genes in evolution. When orthologs were not found, the ortholog place remained blank with a red circle. In case the ortholog identity is not sure, a question mark (?) was added, and the putative paralog that matches this sequence was written above if found. Divergence times of major metazoa groups are presented next to the relevant branches. Our analysis shows that the nearly-complete core pathway known from bilaterians is represented in the cnidarian N. vectensis. The pathway is
mostly absent in sponges and non-metazoans, and is partially present in placozoans. The nematode *C. elegans* and its relatives *C. Brigssae* and *B. malayi* are deficient in several of the pathway components (represented here by the *C. elegans* conserved components). This may indicate of a functional loss of major components in nematodes. The YAP paralogs TAZ were found only in vertebrates and most likely evolved as a result of duplication in early vertebrates or pre-vertebrate chordates.

**Figure 8.** Co-evolution of pathway components.

The co-evolution of Hippo/YAP pathway components was demonstrated as described in methods. Pairwise distances were calculated in MEGA4. A Correlation coefficient and p-values (two-tailed) are presented for each graph on the bottom right corner. Dashed lines present linear regression lines. This analysis shows a strong correlation between the amino acids substitutions of Hippo/YAP pathway interacting proteins in different organisms. The correlation is measured by Pearson’s correlation coefficient. Panels 8I and 8J show analysis of the correlation of YAP with pathway related proteins that are likely not interacting directly with YAP and thus show lower correlations. Panel 8L show the correlation of YAP with β-catenin which resembles YAP molecular function as a transcriptional co-activator, but is unrelated to the pathway. β-catenin was used as a control to examine the strength of the co-evolution forces between Hippo/YAP pathway components.
Supplementary material

Supplement 1. A list of reciprocal best hit analysis for finding orthologs of the pathway proteins.
Supplement 3. Co-evolution analysis data.

References


Figure 1

Cell proliferation, anti-apoptosis

YAP/TAZ

Cytoplasmic retention

Yki

Degradation

TEAD 1-4

Sd

Dachs

Lats1/2

Mats

Mob

Hippo

Hippo

Mst1/2

Rassf

Mer

Ex

Kibra

NF2

FRMD6?

Dco

Fj

App

P

P

14-3-3

YAP/TAZ

Yki

Yki

Yki

Yki

Yki

Yki

Runx

p73

smads

YAP

YAP

YAP

YAP
**Figure 2**

- **Eukaryota**
  - **Metazoa** (1800-900 MYA)
    - **Bilateria** (~1200-900 MYA)
      - **Deuterostomia** (~900 MYA)
        - **Chordata** (~555 MYA)
          - **Vertebrata** (~480 MYA)
            - Homo sapiens
            - Taeniopygia guttata
            - Danio rerio
            - Ciona intestinalis
            - Branchiostoma floridae
            - Strongylocentrotus purpuratus
            - Drosophila melanogaster
            - Apis mellifera
            - Tribolium castaneum
            - Caenorhabditis elegans
            - Nematostella vectensis
            - Trichoplax adhaerens
            - Amphimedon queenslandica
            - Saccharomyces cerevisiae
      - **Ecdysozoa**
        - **Arthropoda** (~555 MYA)
          - Drosophila melanogaster
          - Apis mellifera
          - Tribolium castaneum
          - Caenorhabditis elegans
          - Nematostella vectensis
          - Trichoplax adhaerens
          - Amphimedon queenslandica
          - Saccharomyces cerevisiae
  - **Placozoa** (1800-900 MYA)
  - **Cnidaria** (1200-900 MYA)
Figure 3

**T. adhaerens** YAP

**N. vectensis** YAP

**A. mellifera** Yki

**D. melanogaster** Yki

**T. castaneum** Yki

**S. purpuratus** YAP

**C. intestinalis** YAP

**D. rerio** YAP

**T. guttata** YAP

**H. sapiens** YAP

**H. sapiens** TAZ

Conservation level: 0.0 0.3 0.5 0.7 0.9
Figure 4

H. sapiens

H. sapiens YAP isoform 2

H. sapiens TAZ

C. intestinalis

S. purpuratus

D. melanogaster Yki

D. melanogaster Yki isoform D

N. vectensis

T. adhaerens

100bp
### A. YAP WW1 conserved intron:

<table>
<thead>
<tr>
<th>Species</th>
<th>Intron Conserved Region</th>
<th>Alternative Splicing Sites</th>
<th>H. sapiens Intron Sequence</th>
<th>T. guttata Intron Sequence</th>
<th>D. rerio Intron Sequence</th>
<th>C. intestinalis Intron Sequence</th>
<th>B. floridae Intron Sequence</th>
<th>S. purpuratus Intron Sequence</th>
<th>D. melanogaster Intron Sequence</th>
<th>T. castaneum Intron Sequence</th>
<th>N. vectensis Intron Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>YFLN</td>
<td>HIE</td>
<td>AC TTC TTA AA</td>
<td>GATAAGTAA ... TTGTCTTAG</td>
<td>T CAC ATC GAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAKTSSGQRYFLNHIDQTTTTWQDPRK</td>
</tr>
<tr>
<td>T. guttata</td>
<td>YFLN</td>
<td>HIE</td>
<td>AC TTT CTT AA</td>
<td>GATAAGTAAA ... TTGTCTTAG</td>
<td>T CAT ATT GAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAKTPSGQRYFLNHIDQTTTTWQDPRK</td>
</tr>
<tr>
<td>D. rerio</td>
<td>YFLN</td>
<td>HND</td>
<td>AC TTC CTA AA</td>
<td>GTGAGTTGT ... TCCCCTTGA</td>
<td>C CAT AAT GAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAKTPSGQRYFLNHIDQTTTTWQDPRK</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>YYNM</td>
<td>HND</td>
<td>AT TAT ATG AA</td>
<td>GATAAGATAA ... TTTTTTGA</td>
<td>C CAT AAC GAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAKTADGQQYYMNHNDRSTTWEDPRI</td>
</tr>
<tr>
<td>B. floridae</td>
<td>YYNM</td>
<td>HND</td>
<td>AT TAT CTA AA</td>
<td>GATAAGATAA ... CTCTGCA</td>
<td>C CAT AAC AAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAKTADGQQYYMNHNDRSTTWEDPRI</td>
</tr>
<tr>
<td>S. purpuratus</td>
<td>YYLN</td>
<td>HSN</td>
<td>GG ATC CTGAA</td>
<td>GATAAGTTGA ... TCCTAACAG</td>
<td>T CAT AGC AAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAVTPGQKYFLDHNSQQTWTWEDPRK</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>YYLN</td>
<td>HST</td>
<td>AC TAC TTG AA</td>
<td>GTGAGTTTG ... TACTTGCA</td>
<td>T CAT ACT ACA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEQAKTNDGQYYLYNHITKSTQTWEDPRI</td>
</tr>
<tr>
<td>T. castaneum</td>
<td>YYLN</td>
<td>HST</td>
<td>AC TAT CTT AA</td>
<td>GATAAGAACA ... GCTTGCAG</td>
<td>C CAC ATA ACT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEQARTPEGQLYLYNHITRTRTTTWEDPRK</td>
</tr>
<tr>
<td>N. vectensis</td>
<td>YYLN</td>
<td>HYE</td>
<td>AT TTT ATG AA</td>
<td>GATAAGTACC ... CTCTGTGA</td>
<td>C CAC TAC GAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMRTSPTGQYFMNHYEQITTWQDPRK</td>
</tr>
</tbody>
</table>

### B. TAZ WW conserved intron:

<table>
<thead>
<tr>
<th>Species</th>
<th>Intron Conserved Region</th>
<th>Alternative Splicing Sites</th>
<th>H. sapiens Intron Sequence</th>
<th>T. guttata Intron Sequence</th>
<th>D. rerio Intron Sequence</th>
<th>C. intestinalis Intron Sequence</th>
<th>B. floridae Intron Sequence</th>
<th>S. purpuratus Intron Sequence</th>
<th>D. melanogaster Intron Sequence</th>
<th>T. castaneum Intron Sequence</th>
<th>N. vectensis Intron Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>YFLN</td>
<td>HIE</td>
<td>AC TTC CTC AA</td>
<td>GATAAGTCAG ... TTCTTTGA</td>
<td>T CAC ATA GAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>WEMAFTPNGQKYFLNHIEKITTW</td>
</tr>
<tr>
<td>D. rerio TAZ</td>
<td>YFLN</td>
<td>HIE</td>
<td></td>
<td>WEMAFTPNGQKYFLNHIEKITTW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. guttata TAZ</td>
<td>YFLN</td>
<td>HIE</td>
<td></td>
<td>WEMALTQTGQRYFLNHIEKITTW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. sapiens TAZ</td>
<td>YFLN</td>
<td>HIE</td>
<td></td>
<td>WEMTFTATGQRYFLNHIEKITTW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6

A. ML tree

B. NJ tree

C. YAP and TAZ duplication